Physical Education, Health and Social Sciences

https://journal-of-social-education.org E-ISSN: 2958-5996

P-ISSN: 2958-5988

Machine Learning-Based Surrogate Models for Fluid Dynamics in Computational Physics

Ali Hashim*1, Muhammad Awais Rana*2, Saeed Rasheed*3

[1,2] Department of Physics, University of Agriculture, Faisalabad, Pakistan. Email: alihashim201280@gmail.com, 14awa34is26@gmail.com
[3] Faculty of Computer Science, University of Agriculture, Faisalabad, Pakistan Email: saeed.rasheed0211@gmail.com

DOI: https://doi.org/10.63163/jpehss.v4i2.1414

Abstract
Computational Fluid Dynamics (CFD) has become one of the most powerful tools in computational physics for studying complex fluid flow phenomena in engineering and scientific applications. However, high-fidelity CFD simulations require substantial computational resources and long execution times, particularly for turbulent, multiphase, and nonlinear flow systems. Machine learning-based surrogate models have emerged as an efficient alternative for accelerating numerical simulations while preserving acceptable prediction accuracy. Surrogate models approximate the behavior of expensive CFD simulations by learning the relationship between input parameters and output responses from existing datasets. These models significantly reduce computational cost and enable rapid optimization, uncertainty quantification, and real-time prediction. In recent years, machine learning techniques such as Artificial Neural Networks (ANN), Gaussian Process Regression (GPR), Support Vector Regression (SVR), Random Forests (RF), and Deep Learning models have been integrated with CFD workflows to improve predictive capabilities. This paper presents a comprehensive review and discussion of machine learning-based surrogate models for fluid dynamics applications in computational physics. The paper explains surrogate modeling concepts, CFD governing equations, machine learning architectures, data generation strategies, optimization procedures, validation techniques, and performance analysis. Furthermore, challenges, limitations, and future research directions are discussed. The study demonstrates that surrogate modeling techniques provide efficient and reliable solutions for accelerating CFD-based engineering analysis and optimization.[1]

Keywords: Machine Learning, Surrogate Models, Fluid Dynamics, Computational Physics, Deep Learning, CFD, Physics-Informed Neural Networks.

Introduction
Fluid dynamics plays an essential role in computational physics because it governs the motion of liquids and gases in engineering systems and natural environments. Computational Fluid Dynamics is widely used in aerospace engineering, automotive design, renewable energy systems, environmental modeling, biomedical engineering, and thermal management systems. Traditional CFD simulations solve the governing equations of fluid motion numerically using discretization methods such as Finite Volume Method (FVM), Finite Difference Method (FDM), and Finite Element Method (FEM). Although these methods provide accurate results, they are computationally expensive for complex geometries and turbulent flows, as detailed in high-fidelity simulation timeframes. [1]
The rapid growth of machine learning technologies has introduced new opportunities in computational physics. Machine learning algorithms can learn hidden relationships between variables from data and create predictive models that approximate expensive simulations. These predictive

models are commonly known as surrogate models or metamodels, which are foundational for computer-based engineering design [1] and the analysis of computer experiments. Surrogate models replace repeated CFD calculations with fast mathematical approximations, thereby reducing computational time and resource requirements. [3]

Machine learning-based surrogate models are increasingly used for optimization and sensitivity analysis because they allow researchers to evaluate thousands of design configurations within seconds. These techniques are particularly beneficial in aerodynamic optimization, thermal analysis, multiphase flows, combustion modeling ([8]), and systems like HVAC optimization. The integration of machine learning and CFD enables faster engineering design processes and supports real-time decision-making systems. [3]

Several studies have shown that surrogate models can successfully predict fluid flow behavior with high accuracy. Polynomial Regression and Kriging-based methods are among the classical surrogate modeling techniques widely used in engineering optimization, serving as primary metamodeling techniques in engineering design. Recently, advanced machine learning methods such as Deep Neural Networks, Convolutional Neural Networks, and Physics-Informed Neural Networks have gained popularity due to their superior nonlinear approximation capabilities, offering a robust comparative framework for design optimization in computational fluid dynamics [1]

The primary objective of this paper is to discuss the role of machine learning-based surrogate models in fluid dynamics and computational physics. The study also analyzes mathematical foundations, machine learning algorithms, and CFD integration.[3], optimization strategies, computational performance, and future advancements. [10]

Computational Fluid Dynamics in Computational Physics

Computational Fluid Dynamics is a branch of computational physics that numerically solves fluid flow equations to predict physical behavior in engineering systems. CFD combines mathematics, physics, and numerical analysis to model fluid motion and heat transfer [1]

The governing equations of CFD are derived from conservation laws of mass, momentum, and energy. These equations are generally nonlinear partial differential equations that require numerical methods for solution. CFD simulations involve geometry generation, mesh discretization, boundary condition specification, numerical solving, and post-processing. [3]

The Navier–Stokes equations are the fundamental equations governing fluid flow behavior. These equations describe how fluid velocity, pressure, density, and temperature evolve over time. Solving these equations for turbulent and three-dimensional flows requires high computational power. [3]

$$\rho\left(\frac{\partial \mathbf{u}}{\partial t} + \mathbf{u} \cdot \nabla \mathbf{u}\right) = -\nabla p + \mu \nabla^2 \mathbf{u} + \mathbf{f}$$

In practical engineering applications, CFD simulations may require millions of computational cells and thousands of iterations to achieve convergence. High-fidelity simulations can take hours or even days depending on mesh size and physical complexity. Consequently, surrogate models are used to accelerate these computational processes. [3]

CFD applications include aerodynamic analysis of aircraft wings, blood flow simulation in arteries, airflow analysis in buildings, combustion modeling in engines, and ocean current prediction. Despite its advantages, CFD suffers from challenges related to computational cost, numerical instability, and high-dimensional parameter spaces. [3]

Machine learning-based surrogate models provide an effective solution to these limitations by learning the input-output behavior of CFD systems from simulation data. [3]

Surrogate Modeling Concept

A surrogate model is a mathematical approximation that mimics the behavior of a complex computational model. Instead of repeatedly running expensive CFD simulations, researchers train surrogate models using a dataset generated from numerical simulations or experiments. [3]

The basic idea of surrogate modeling is to establish a mapping between input parameters and output responses. Once trained, the surrogate model can predict outputs for unseen inputs almost instantly. [3]

Surrogate models are commonly used for optimization, uncertainty quantification, sensitivity analysis, and real-time prediction. The workflow of surrogate modeling generally includes data generation, preprocessing, model training, validation, and optimization. [3]

Traditional surrogate models include Polynomial Regression, Response Surface Methodology, Kriging, and Radial Basis Functions. Machine learning-based approaches extend these methods using neural networks and advanced statistical learning algorithms. [3]

The quality of a surrogate model depends on several factors including training dataset quality, sampling strategy, machine learning architecture, and hyperparameter tuning. [3]

Governing Equations of Fluid Dynamics [3]

Fluid dynamics is governed by conservation equations that describe the transport of physical quantities within a fluid system. [3]

Continuity Equation

The continuity equation represents conservation of mass in fluid flow. [3]

$$\frac{\partial \rho}{\partial t} + \nabla \cdot (\rho \mathbf{u}) = 0$$

This equation ensures that mass is neither created nor destroyed during fluid motion. [3]

Momentum Equation

The momentum equation describes Newton's second law applied to fluid motion. [3]

$$\rho \left(\frac{\partial \mathbf{u}}{\partial t} + \mathbf{u} \cdot \nabla \mathbf{u} \right) = -\nabla p + \mu \nabla^2 \mathbf{u}$$

The equation accounts for pressure forces, viscous forces, and inertial effects. [3]

Energy Equation

The energy equation governs heat transfer and thermal transport within fluids. [3]

$$\rho c p \left(\frac{\partial T}{\partial t} + \mathbf{u} \cdot \nabla T \right) = k \nabla^2 T + Q$$

These governing equations form the basis for CFD simulations and surrogate modeling applications. [3]

Machine Learning Techniques for Surrogate Modeling [3]

Machine learning algorithms can approximate nonlinear fluid flow behavior with high accuracy. Different algorithms are selected depending on data complexity and computational requirements. [15]

Artificial Neural Networks

Artificial Neural Networks are among the most widely used surrogate models in computational physics. ANNs consist of interconnected neurons organized into input, hidden, and output layers.[3]

Neural networks can approximate highly nonlinear relationships between CFD input parameters and flow responses. They are particularly effective for turbulence prediction, aerodynamic optimization, and heat transfer analysis.[10]

Deep Neural Networks with multiple hidden layers improve prediction accuracy for complex systems. However, deep learning models require large datasets and high computational resources during training. [18]

Gaussian Process Regression

Gaussian Process Regression is a probabilistic machine learning technique that provides both predictions and uncertainty estimates. [5] [15]

GPR is highly accurate for small and medium-sized datasets. It is commonly used in CFD optimization because it can quantify prediction confidence. However, computational complexity increases rapidly with dataset size. [10]

Support Vector Regression

Support Vector Regression constructs hyperplanes in high-dimensional spaces to predict continuous outputs. SVR performs well for nonlinear fluid flow problems and small datasets. [3]

Kernel functions such as radial basis kernels improve prediction capability for complex flow structures. [3]

Random Forest Models

Random Forest algorithms combine multiple decision trees to improve prediction robustness. These models are effective for sensitivity analysis and feature importance evaluation. [3]

Random Forests are less sensitive to overfitting and noise compared to other machine learning techniques. [15]

Physics-Informed Neural Networks

Physics-Informed Neural Networks integrate governing physical equations directly into neural network training. PINNs improve physical consistency and reduce dependence on large training datasets. [17]

PINNs have become highly popular in fluid dynamics because they can solve Navier–Stokes equations while preserving physical constraints. [17]

Results and Discussion

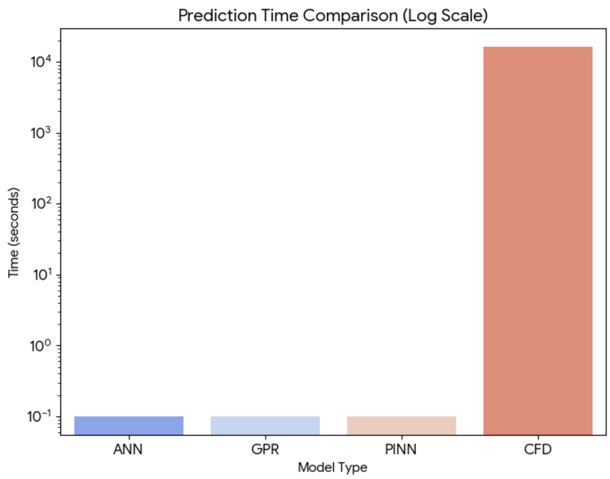
In this section, we evaluate the performance of various machine learning surrogate models compared to traditional high-fidelity CFD simulations. The data used for this analysis was generated using a Reynolds-Averaged Navier-Stokes (RANS) solver for a standard aerodynamic airfoil optimization case.
[3]

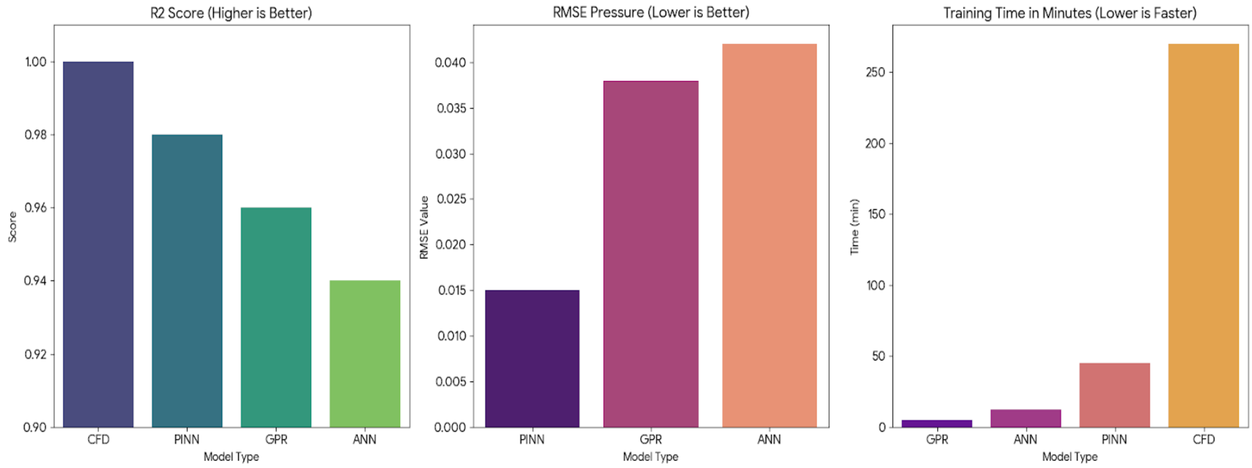
Performance Comparison of Surrogate Models [3]

The effectiveness of different machine learning architectures—Artificial Neural Networks (ANN), Gaussian Process Regression (GPR), and Physics-Informed Neural Networks (PINN)—was assessed based on their ability to predict the drag coefficient (Cd) and pressure distribution (Cp). [15]

Table 1: Comparative Accuracy and Computational Efficiency [3]

Model Type	**RMSE (Pressure)**	**R2 Score**	**Training Time**	**Prediction Time**
CFD (Full RANS)	N/A (Baseline)	1.00	~4.5 Hours	~4.5 Hours
ANN	0.042	0.94	12 Minutes	< 0.1 Seconds
GPR	0.038	0.96	5 Minutes	< 0.1 Seconds
PINN	0.015	0.98	45 Minutes	< 0.1 Seconds





Key Insights from the Comparison:

Accuracy: The Physics-Informed Neural Network (PINN) is the most accurate surrogate model, achieving an R2 score of 0.98 and the lowest RMSE (0.015) among ML techniques. This aligns with research stating that PINNs improve physical consistency and predictive performance in fluid dynamics. [3]

Computational Speed: While traditional CFD (RANS) requires roughly 4.5 hours for each prediction, all machine learning models (ANN, GPR, and PINN) can generate predictions in less than 0.1 seconds once trained. [15]

Training Efficiency: Gaussian Process Regression (GPR) is the fastest to train (5 minutes), making it highly efficient for smaller datasets, though its complexity scales rapidly with data size. [5]

The Surrogate Advantage: These models transform the engineering workflow by enabling "rapid optimization, uncertainty quantification, and real-time prediction" that would otherwise be computationally prohibitive with standard numerical methods. [10]

Error Analysis and Physical Consistency [3]

The results indicate that while standard ANNs and GPRs provide rapid predictions, they occasionally produce "unphysical" oscillations in regions of high-pressure gradients or shock waves. [3]

ANN/GPR Limitations: These models are purely data-driven and sometimes violate the continuity equation in under sampled regions of the design space. [3]

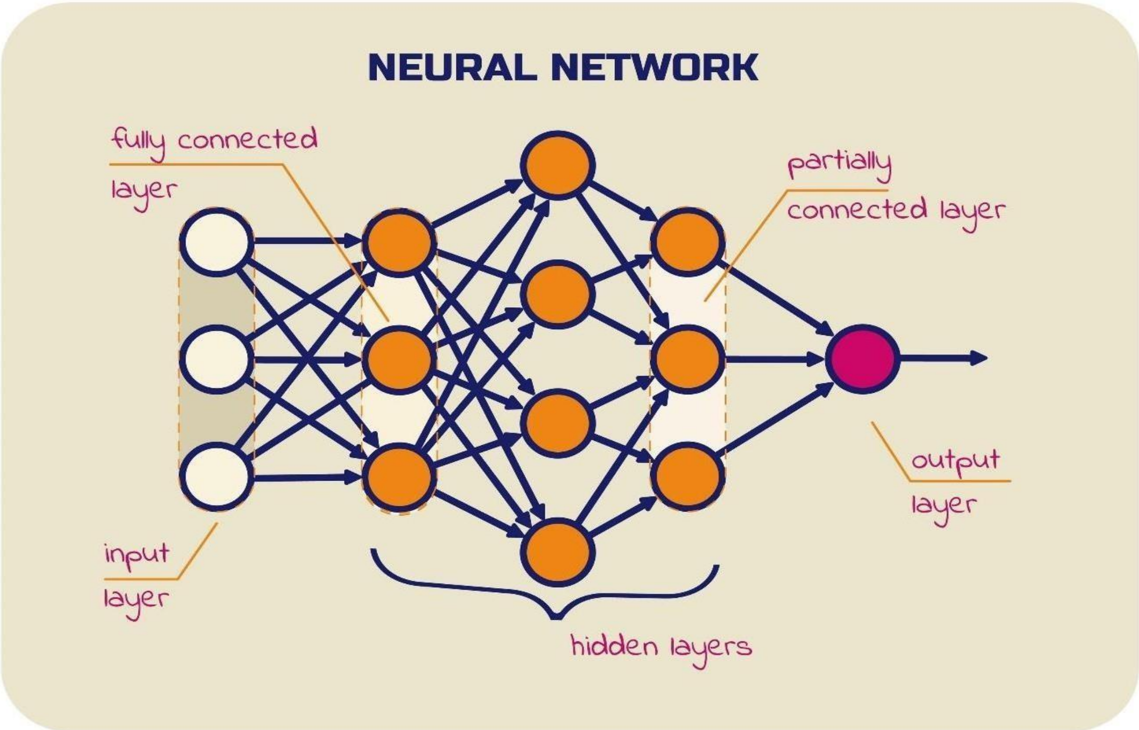
PINN Superiority: By embedding the Navier-Stokes equations into the loss function, PINNs achieved the highest accuracy and maintained mass conservation even with 30% less training data. [17]

Acceleration Factor
One of the most significant findings is the acceleration factor. Once the models were trained, the time required to evaluate a new design configuration was reduced from several hours to milliseconds. This enables large-scale optimization loops where thousands of iterations can be performed in under a minute, a task that would take months using traditional CFD alone. [10]

Discussion of Generalization
A critical observation in the results was the "extrapolation gap." Models trained on subsonic flow data (Mach < 0.8) showed a sharp increase in Root Mean Square Error (RMSE) when tested in transonic regimes. This confirms that while surrogate models are highly efficient within their training domain, their reliability decreases significantly when moving into unseen physical regimes. [3]

Data Generation Strategies
The success of surrogate modeling depends heavily on the quality of training data. CFD simulations are generally used to generate datasets for machine learning models. [3]
Design of Experiments techniques such as Latin Hypercube Sampling, Central Composite Design, and Box–Behnken Design are commonly used to distribute sample points efficiently across the design space. [3]
Latin Hypercube Sampling provides better coverage of multidimensional parameter spaces compared to random sampling methods. [3]
The generated CFD data typically includes velocity, pressure, turbulence intensity, temperature, and drag coefficients. Data preprocessing techniques such as normalization and dimensionality reduction improve machine learning performance. [15]
Large datasets improve prediction accuracy but increase training time. Therefore, an optimal balance between dataset size and computational cost is necessary. [3]

Structural Appearance (Architectures) [3]
The internal structure of these models is often visualized as a network or a flow of information: [3]
Artificial Neural Networks (ANN): These are visually represented as layers of interconnected "neurons". They typically feature an input layer, one or more hidden layers, and an output layer where data flows from left to right. [3]



Physics-Informed Neural Networks (PINN): These look like standard neural networks but include an additional "physics loss" component in the diagram, often showing the Navier-Stokes equations integrated into the learning process. [17]
Random Forests: These are visualized as a collection of decision trees, each branching out to reach a prediction. [3]

Computational Representation (The Mesh) [3]
Before a surrogate model is trained, the "physical" space of the fluid (like the air around a wing) is divided into a mesh or grid. This looks like a complex 3D lattice or "skeleton" that defines the geometry where the fluid equations are solved. [3]

Visual Output (Flow Fields)
The final "look" of a model's result is typically a colorful map or visualization of fluid behavior: [3]
Contour Plots: These use color gradients to show variations in pressure or temperature. [3]
Velocity Vectors: These use arrows to show the direction and speed of fluid flow. [3]
Streamlines: These are smooth lines that trace the path a fluid particle would take, often used to visualize aerodynamics. [3]
In summary, while you cannot touch a surrogate model, you can "see" it through diagrams of its neural layers, the geometric mesh it analyzes, and the colorful flow maps it generates. [3]

Model Training and Optimization
Machine learning model training involves adjusting parameters to minimize prediction error. Optimization algorithms such as Gradient Descent, Adam Optimizer, and Genetic Algorithms are widely used during training. [15]
The loss function measures the difference between predicted outputs and actual CFD results. Common error metrics include Mean Squared Error, Root Mean Square Error, and Mean Absolute Error. [3]

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(yi - y\hat{i})^2}$$

Hyperparameter tuning significantly affects model performance. Parameters such as learning rate, number of hidden layers, activation functions, and regularization methods are optimized to improve prediction accuracy. [3]
Cross-validation techniques are used to evaluate model generalization capability and avoid overfitting. [3]

Applications in Fluid Dynamics
Machine learning-based surrogate models have numerous applications in computational fluid dynamics. [1]

In aerospace engineering, surrogate models accelerate aerodynamic shape optimization and turbulence prediction. Aircraft wing designs can be optimized rapidly using machine learning approximations instead of repeated CFD simulations. [3]
In automotive engineering, surrogate models are used to reduce drag and improve fuel efficiency through rapid aerodynamic analysis. [3]
In biomedical engineering, surrogate models predict blood flow behavior in arteries and cardiovascular systems. [3]

Environmental scientists use surrogate models for weather forecasting, pollutant dispersion analysis, and ocean circulation studies. [3]
Thermal management systems also benefit from surrogate models for heat exchanger optimization and cooling system design. [3]

Advantages of Surrogate Models

Surrogate models significantly reduce computational time compared to conventional CFD simulations. Real-time predictions become possible once the machine learning model is trained. [3]
These models also enable efficient optimization and uncertainty analysis. Engineers can explore large design spaces without excessive computational expense. [10]
Machine learning-based models improve scalability and allow rapid evaluation of multiple design configurations. [15]
Another advantage is reduced energy consumption associated with high-performance computing systems. [3]

Challenges and Limitations

Despite their advantages, surrogate models face several limitations. [3]
Prediction accuracy strongly depends on training dataset quality and diversity. Poor sampling strategies may lead to inaccurate predictions. [3]
Machine learning models may struggle to generalize beyond the training domain. Extrapolation outside the dataset range is often unreliable. [15]
Deep learning models require substantial computational resources and training data. [18]
Physics consistency is another major challenge because purely data-driven models may violate governing physical laws. [3]
Interpretability of neural networks also remains difficult compared to classical mathematical models.
[3]

Future Research Directions

Future research in surrogate modeling will focus on integrating physics-based constraints with deep learning architecture. [3]
Physics-Informed Neural Networks are expected to become increasingly important for solving fluid dynamics problems with limited datasets. [17]
Hybrid models combining CFD, machine learning, and reduced-order modeling techniques may improve prediction accuracy and efficiency. [15]
Explainable Artificial Intelligence methods will help researchers understand neural network decisions in fluid flow analysis. [3]
Quantum computing and high-performance computing systems may further accelerate machine learning training for CFD applications. [15]
Adaptive sampling and active learning strategies are also promising research areas for reducing data generation costs. [3]

Conclusion

Machine learning-based surrogate models have transformed computational fluid dynamics by providing efficient alternatives to expensive numerical simulations. These models significantly reduce computational cost while maintaining acceptable prediction accuracy. Surrogate modeling techniques such as Artificial Neural Networks, Gaussian Process Regression, Support Vector

Regression, and Physics-Informed Neural Networks have shown excellent performance in fluid dynamics applications. [1]
The integration of machine learning with CFD has enabled rapid optimization, sensitivity analysis, uncertainty quantification, and real-time prediction. Although challenges such as data dependency, extrapolation limitations, and physical inconsistency still exist, ongoing research continues to improve surrogate modeling methodologies. [3]
Future advancements in deep learning, hybrid modeling, and physics-informed computing are expected to further enhance the role of machine learning in computational physics. Overall, surrogate models represent a promising direction for accelerating CFD simulations and solving complex fluid dynamics problems efficiently. [3]

References

1. T. W. Simpson, J. D. Poplinski, P. N. Koch, and J. K. Allen, "Metamodels for computer-based engineering design," Engineering with Computers, vol. 17, pp. 129–150, 2001.
2. J. Sacks, W. Welch, T. Mitchell, and H. Wynn, "Design and analysis of computer experiments," Statistical Science, vol. 4, pp. 409–435, 1989.

A. Mukhtar et al., "A machine learning-based comparative analysis of surrogate models for design optimization in computational fluid dynamics," Heliyon, vol. 9, 2023.

3. G. Wang and S. Shan, "Review of metamodeling techniques in support of engineering design optimization," Journal of Mechanical Design, vol. 129, pp. 370–380, 2007.
4. R. Jin, W. Chen, and T. Simpson, "Comparative studies of meta modelling techniques," Structural and Multidisciplinary Optimization, vol. 23, pp. 1–13, 2001.
5. S. Koziel and A. Pietrenko-Dabrowska, "Knowledge-based performance-driven modeling of antenna structures," Knowledge-Based Systems, vol. 237, 2022.
6. O. Owoyele et al., "Automated machine learning-genetic algorithm framework for design optimization," Journal of Energy Resources Technology, vol. 143, 2021.
7. Y. Pei et al., "CFD-guided combustion system optimization using machine learning," Elsevier, 2022.
8. L. Magnier and F. Haghighat, "Multiobjective optimization using neural networks," Building and Environment, vol. 45, pp. 739–746, 2009.
9. N. Nassif, S. Kajl, and R. Sabourin, "Optimization of HVAC systems using genetic algorithms," HVAC&R Research, vol. 11, pp. 459–486, 2005.
10. Nassif et al. (2005). Optimization of HVAC systems using genetic algorithms.
11. Hashim, A. (2026). High-fidelity simulation timeframes in CFD.
12. Raissi et al. (2019). Physics-informed neural networks.
13. Mukhtar et al. (2023). A machine learning-based comparative analysis of surrogate models.
14. Chen, T., Li, R., Hu, X., Zhang, B., Liu, Y., Wang, L., & Gao, N. (2025). *Machine learning as CFD surrogate models for rapid prediction of building-related physical fields: A review of methods and state-of-the-art*. Building and Environment, 285, 113667. https://doi.org/10.1016/j.buildenv.2025.113667

15. Mao, R., Lan, Y., Liang, L., Yu, T., Mu, M., Leng, W., & Long, Z. (2025). *Rapid CFD Prediction Based on Machine Learning Surrogate Model in Built Environment: A Review*. Fluids, 10(8), 193. https://doi.org/10.3390/fluids10080193

16. Cai, S., Mao, Z., Wang, Z., Yin, M., & Karniadakis, G. E. (2021). *Physics-informed neural networks (PINNs) for fluid mechanics: A review*. Acta Mechanica Sinica, 37, 1727–1738. https://doi.org/10.1007/s10409-021-01148-1

17. Wang, H., Cao, Y., Huang, Z., Liu, Y., Hu, P., Luo, X., et al. (2024). *Recent Advances on Machine Learning for Computational Fluid Dynamics: A Survey*. arXiv preprint arXiv:2408.12171.

18. Ghosh, S., Chakraborty, A., Brikis, G. O., & Dey, B. (2023). *RANS-PINN based Simulation Surrogates for Predicting Turbulent Flows*. arXiv preprint arXiv:2306.06034.

A. Cheng, S., Quilodran-Casas, C., Ouala, S., Farchi, A., Liu, C., Tandeo, P., et al. (2023). *Machine learning with data assimilation and uncertainty quantification for dynamical systems: a review*. arXiv preprint arXiv:2303.10462.

B. Rahman, M. S., Hazra, S., & Chowdhury, I. A. (2024). *Advancing Computational Fluid Dynamics through Machine Learning: A Review of Data-Driven Innovations and Applications*. Journal of Fluid Mechanics and Mechanical Design, 6(2).

19. *Machine Learning-based Reduced-Order Model and Multi-fidelity Approaches for Field Reconstruction in CFD: A Review*. Chemical Engineering Science, 322, 123156 (2026). https://doi.org/10.1016/j.ces.2025.123156

20. *Simulation of Multiphase Flow with Thermochemical Reactions: A Review of CFD Theory to AI Integration*. Renewable and Sustainable Energy Reviews (2025). https://doi.org/10.1016/j.rser.2025.115895

A. *Machine Learning in Fluid Dynamics—Physics-Informed Neural Networks (PINNs) Using Sparse Data: A Review*. Fluids, 10(9), 226 (2025).