Recent advances and effectiveness of machine learning models for fluid dynamics in the built environment

Tran Van Quang[1*], Dat Tien Doan[2], Geun Young Yun[1]

[1]*Department of Architectural Engineering, Kyung Hee University,*

1732, Deogyeong-daero, Giheung-gu, Yongin-si, Gyeonggi-do, 17104, Republic of Korea

[2]*Department of Built Environment Engineering, School of Future Environments, Auckland University of Technology, Auckland, 1010, New Zealand*

[*]Corresponding author: *tranvanquang2912@khu.ac.kr*

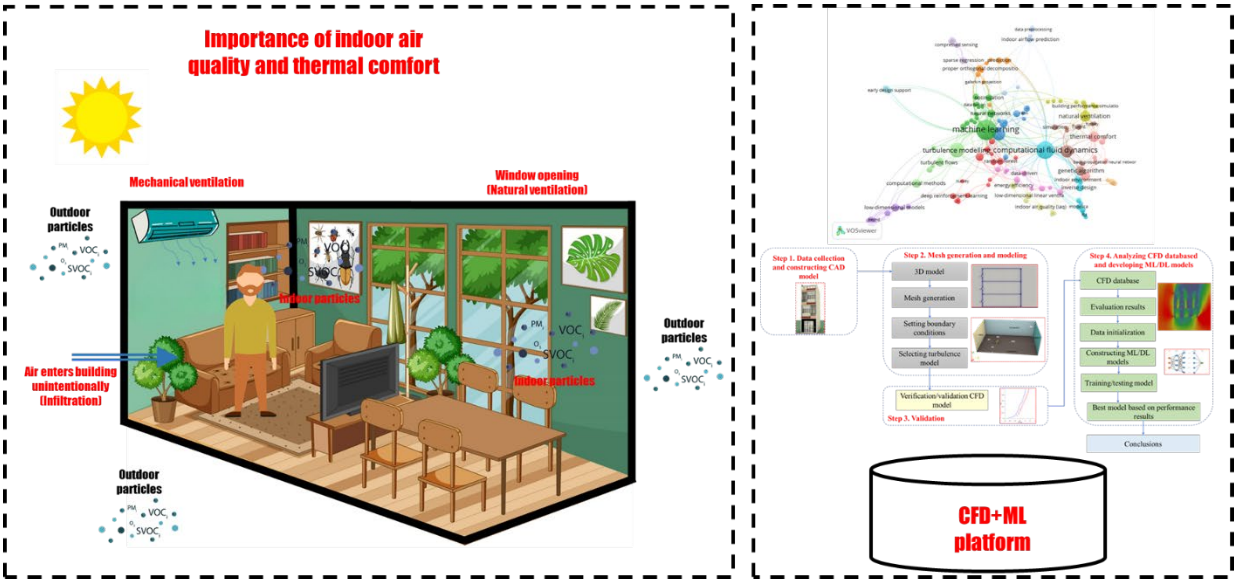
Recent advances and effectiveness of machine learning models for fluid dynamics in the built environment

Abstract

Indoor environmental quality is crucial for human health and comfort, necessitating precise and efficient computational methods to optimise indoor climate parameters. Recent advancements in machine learning (ML) and computational fluid dynamics (CFD) are promising. However, applying ML to complex building airflow presents challenges. This research aims to investigate the integration of ML with CFD in the context of built environment applications using a systematic review approach. It highlights a critical knowledge gap: the need to synthesise innovative approaches that address the limitations of indoor modelling using data-driven ML methods. The review examines contemporary literature, identifying current developments and suggesting potential future directions. It delves into the innovations in combining ML with CFD to predict thermal comfort and indoor air quality, uncovering key limitations such as the lack of high-quality experimental data for training and validation, the computational complexity of detailed CFD simulations, and the interpretability issues of 'black-box' ML models. The emergence of data-driven techniques in fluid mechanics offers promising prospects for modelling in the built environment. Future research should focus on incorporating physics-based rules in ML models, adapting turbulence closure models for indoor flows, and enhancing model validation using real-world datasets. The research emphasises the synergistic relationship between ML and CFD; it proposes pathways to overcome current limitations, aiming to enhance the precision and efficiency of indoor environment modelling through their integration.

Keywords: Built environment; Indoor environment quality; Thermal comfort; Machine learning; Computational fluid dynamics; Turbulence modeling.

Graphical abstract



Abbreviations

AI – Artificial intelligence

ACH - Air changes per hour

ANN – Artificial neural network

ARIMA – Autoregressive integrated moving average

BMSA – Bayesian model-scenario averaging

BES – Building energy simulation

CART – Classification and regression tree

CFD – Computational fluid dynamics

CNN – Convolutional neural networks

DL – Deep learning

DNN – Deep neural network

DNS – Direct numerical simulation

DSC – Differential scanning calorimetry

EUI – Energy use intensity

FFD – Fast fluid dynamics

GA – Genetic algorithm

GAN – Generative adversarial nets

HVAC – Heating, ventilation, and air conditioning

IAQ – Indoor air quality

KNN – K-nearest neighbors

LASSO – Least absolute shrinkage and selection operator

LES – Large eddy simulation

LLVM – Low-dimensional linear ventilation models

LVM – Linear ventilation models

LRWSGD – Linear regression with stochastic gradient descent

MC – Monte carlo

ML – Machine learning

MLP – Multi-layer perceptron

MSM – Multiscale modeling

MOGA – Multi-objective genetic algorithm

NN – Neural network

PMV – Predicted mean vote

PPD – Predicted percentage of dissatisfaction

PRISMA – Preferred reporting items for systematic reviews and meta-analyses

PSO – Particle swarm optimization

QMC – Quasi-Monte Carlo

RANS – Reynolds averaged Navier-Stokes simulation

RBF-ANN – Radial basis function artificial neural network

RF – Random forest

RL – Reinforcement learning

SGD – Stochastic gradient descent

SVM – Support vector machine

SBS – Sick building syndrome

1. Introduction

1.1. Importance of indoor environment

The significance of indoor quality within the built environment cannot be overstated. It is pivotal in influencing human health, comfort, and overall productivity. Various scientific studies [1], [2], [3], [4] have unequivocally demonstrated that short-term exposure to heightened pollution episodes and prolonged exposure to low-level pollutants can harm the human respiratory system.

One particularly hazardous component of air pollution is particulate matter, specifically $PM_{2.5}$, known for its ability to penetrate the alveoli [5]. Moreover, exposure to oxidizing agents like solid ozone (O_3) and acidic gaseous pollutants such as sulfur dioxide (SO_2) and nitrogen dioxide (NO_2) can lead to irritation and inflammation of the airways [6]. Additionally, restoration projects can introduce indoor air pollutants, including formaldehyde and volatile organic compounds (VOCs), posing significant health risks, particularly for children with allergies and asthma [7], [8], [9]. Understanding which indoor environmental factors and exposure timings contribute to the emergence of early symptoms and the progression of identified disorders in allergic children is paramount.

It is noteworthy that indoor spaces can exhibit substantially higher levels of air pollutants, encompassing particulate matter and VOCs, compared to outdoor environments [10], [11], [12]. This underscores the need for comprehensive investigation and management of indoor air quality. Significantly, the repercussions of air pollution extend beyond physical health, affecting mental well-being. Recent research has established a compelling link between elevated air pollution levels and an increased risk of psychiatric disorders, including schizophrenia, depression, and attention-deficit hyperactivity disorder (ADHD) among children [13], [14]. In adults, exposure to air pollution has been associated with a heightened risk of depression [15], anxiety disorders [16], and dementia [17]. Although the precise mechanisms underlying the relationship between air pollution and mental health remain a subject of ongoing inquiry, it is imperative to delve deeper into this connection to safeguard individuals from the adverse impacts of air pollution.

1.2. Computational fluid dynamics techniques in building design

Computational Fluid Dynamics (CFD) emerges as a potent tool, facilitating the meticulous simulation of crucial factors such as heat transfer, airflow velocity, and the dispersion of pollutants [18], [19]. Within indoor environment research, CFD modeling finds widespread application in providing precise predictions of airflow patterns, temperature gradients, and pollutant concentrations within defined spaces. Nevertheless, the principal drawback of the CFD methodology lies in its time-intensive computational demands, primarily rooted in the necessity for finely detailed 3D representations of the building through a dense mesh [20], [21], [22]. As a consequence, computational time escalates proportionally with mesh complexity. Additionally, the intricate nature of model implementation presents formidable challenges, particularly for those lacking prior expertise in software and fluid dynamics [23]. Furthermore, the capability of CFD to simulate turbulence remains constrained despite substantial advancements in high-performance computing resources in recent years. Consequently, the execution of CFD simulations continues to be a resource-intensive endeavor across many practical applications [24].

A novel approach has emerged to enhance turbulence models and their validation against comprehensive datasets derived from high-fidelity techniques, blending machine learning techniques with fundamental physical principles. This innovative strategy aims to bridge the gap between data obtained from high-fidelity methods and approximation models, reducing disparities between them and identifying the most fitting model [23]. Recent strides in high-performance computing have facilitated the acquisition of high-fidelity databases concerning canonical flows, while the integration of artificial intelligence-driven algorithms for big data processing has ushered in a transformative era [25]. This breakthrough is underscored by recent investigations that have harnessed numerous canonical flow tests to train these models. The foundational models encompass the conventional one and two-equation turbulence models. It is noteworthy that machine learning techniques find application in both Large Eddy Simulation (LES) and Reynolds-Averaged Navier-Stokes (RANS) models, albeit the focus has predominantly rested on the former. Fewer inquiries have explored utilizing machine learning for constructing turbulence models adjusted to complex flows featuring two-phase interactions and heat/mass transfer phenomena [26].

1.3. Machine learning in fluid dynamics

ML, a subset of AI, employs advanced algorithms to discern patterns in historical data and predict future events or outcomes under specific conditions [27]. It encompasses various learning methods, including supervised learning [27], unsupervised learning [28], and reinforcement learning [29], chosen based on the nature of available data (see Fig. 1). In supervised learning, both input and output data are accessible, allowing the construction of functions that accurately capture relationships. Unsupervised learning deals with available input data but lacks corresponding output data, aiming to uncover hidden patterns. Reinforcement learning operates without direct input-output data pairs, instead optimizing models based on estimations of output quality or "rewards."

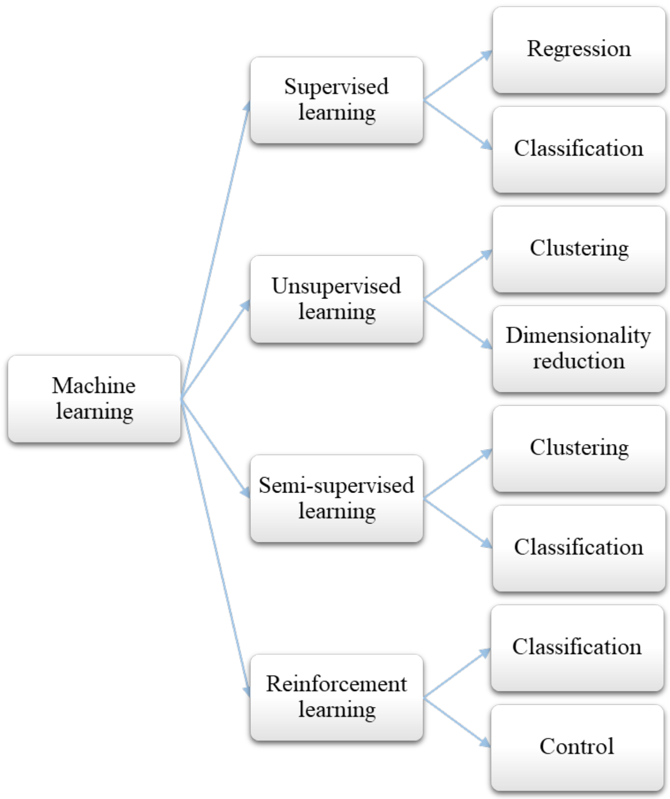


Figure 1. Machine learning algorithms

A neural network (NN), an AI model, is a potent data-driven tool that establishes connections among target variables marked by nonlinear dependencies that prove challenging to encapsulate using traditional mathematical functions. It excels at swiftly and accurately forecasting unknown data [30]. In recent times, NNs have gained substantial traction in the realm of indoor environment analysis, finding utility in diverse applications, including the evaluation of natural ventilation [31], [32], estimation of indoor carbon dioxide concentrations [33], prediction of building energy consumption [34], [35], optimization of window opening designs [36], and fine-tuning HVAC system performance [37].

A NN to identify the relationship between an aircraft's thermal comfort and cabin ventilation [38], [39], [40]. The NN was synergistically coupled with a genetic algorithm to effectively pursue reverse cabin environment engineering, accounting for thermal comfort and energy efficiency. A pioneering ventilation control strategy incorporated LLVM-based ANN as a substitute for traditional CFD [41]. The NN model exhibited prowess in predicting the low-dimensional distribution of pollutant concentrations. Furthermore, Qin et al. [42] leveraged NNs and Building Energy Simulation (BES) to probe the dynamic energy consumption patterns and temperature profiles within an atrium. The NN model adeptly characterized the intricate temperature variations within an atrium undergoing annual dynamic processes. However, the complex interplay between boundary conditions and the indoor environment necessitates further exploration to decipher how an NN model can effectively predict non-uniform indoor environments.

An illuminating demonstration of NN's capabilities materialized in indoor airflow velocity distribution prediction [43]. Nevertheless, it is imperative to acknowledge that the feasibility of an NN to concurrently predict multiple distributions of diverse variables remains an open question, especially when compared to the comprehensive capabilities of CFD modeling, which can encompass simultaneous predictions of indoor airflow, temperature profiles, and even pollutant dispersion [44].

While the overarching aim of surrogate modeling remains the maintenance of accuracy while curtailing computational expenses [45], it is commonplace for accuracy to be somewhat compromised in favor of computational efficiency. Surrogate modeling is often constructed upon and juxtaposed with CFD models, which persist as comprehensive frameworks despite their inherent uncertainties. ANNs exhibit variable performance across scenarios yet yield results that approximate accuracy in optimal circumstances. Due to its cost-effectiveness, the current research landscape in the built environments prominently spotlights surrogate modeling. However, it is worth noting that deep learning architectures hold promise for an array of potential applications beyond the confines of cost-efficient modeling.

The application of machine learning techniques in fluid dynamics poses unique challenges. Complex physics representations require extensive labeled data covering diverse conditions to train models [46] sufficiently. However, quality experimental and simulation data is often limited in availability and scope

[47]. Black-box machine learning models also suffer from interpretability issues, making integrating fluid physics difficult [48]. Additional considerations include the computational demands of high-fidelity simulations for generating training data as well as inference latency constraints for real-time applications [49]. Proper neural network design and training are critical to avoid instability and inaccuracies. These factors underscore the need for a systematic synthesis of innovations combining machine learning with computational fluid dynamics tailored to the built environment.

1.4. Aim and Objectives

The current research landscape juxtaposes surrogate modeling with CFD models, each with uncertainties. While ANNs may exhibit variability in performance, they offer reasonably comparable accuracy under optimal conditions. Due to its cost-effectiveness, the predominant focus of ongoing research in the built environment revolves around surrogate modeling, yet deep learning architectures hold promise for various applications. This study systematically reviews the progress and future directions in integrating ML with CFD for built environment applications. A critical gap exists in synthesizing recent advancements that address the unique challenges of indoor airflow modeling using data-driven machine learning. The objectives are to critically review the state-of-the-art, identify limitations of current methods, and propose opportunities for further research and model improvements. The contributions include an in-depth analysis of the latest innovations in coupling ML with CFD for thermal comfort and indoor air quality predictions, as well as a discussion of emerging fluid mechanics techniques that show promise for built environment simulations.

This study is divided into five different sections. Section 2 outlines the methodology, including the systematic review process following PRISMA guidelines. Section 3 examines recent applications of machine learning techniques in built environment fluid dynamics, focusing on thermal comfort and indoor air quality predictions. Section 4 critically analyzes innovations in coupling machine learning architectures with CFD across general fluid mechanics problems beyond buildings. Finally, Section 5 presents a comprehensive discussion synthesizing key limitations of current approaches, proposing potential paths to address identified challenges, and offering recommendations for future work to advance the merger of machine learning and CFD for enhanced modeling of built environments.

2. Methodology

A systematic review was conducted following the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) guidelines [50] to examine the integration of machine learning and computational fluid dynamics (CFD) for built environment applications. The review aimed to map the current research landscape, critically appraise existing methods, and identify future opportunities. The search was performed across two databases – Web of Science and Google Scholar. Combinations of the following keywords were used: “machine learning”, “deep learning”, “computational fluid dynamics”, “airflow”, “indoor air quality”, “thermal comfort”, and “built environment”. Filters were applied to retrieve peer-reviewed English-language journal articles published over the past decade (2012–2022).

The search yielded 235 records after removing duplicates (see Figure 2). Two independent reviewers screened the titles, and abstract of these records against the inclusion criteria: (1) original research articles, (2) combined CFD with machine learning approaches, (3) application to indoor modeling, thermal comfort or air quality predictions, (4) focused on built environments. Through this screening, 144 records were excluded. The remaining 91 full-text articles were assessed in detail, which resulted in the further exclusion of 50 papers that did not meet all eligibility criteria. Finally, 41 studies were included in the qualitative and quantitative analysis. Data were systematically extracted into a spreadsheet on: year of publication, machine learning techniques used, CFD methods, application area, key findings, and limitations. Descriptive statistics summarized publication trends over time and by country. Thematic analysis was also conducted to synthesize common limitations and future opportunities.

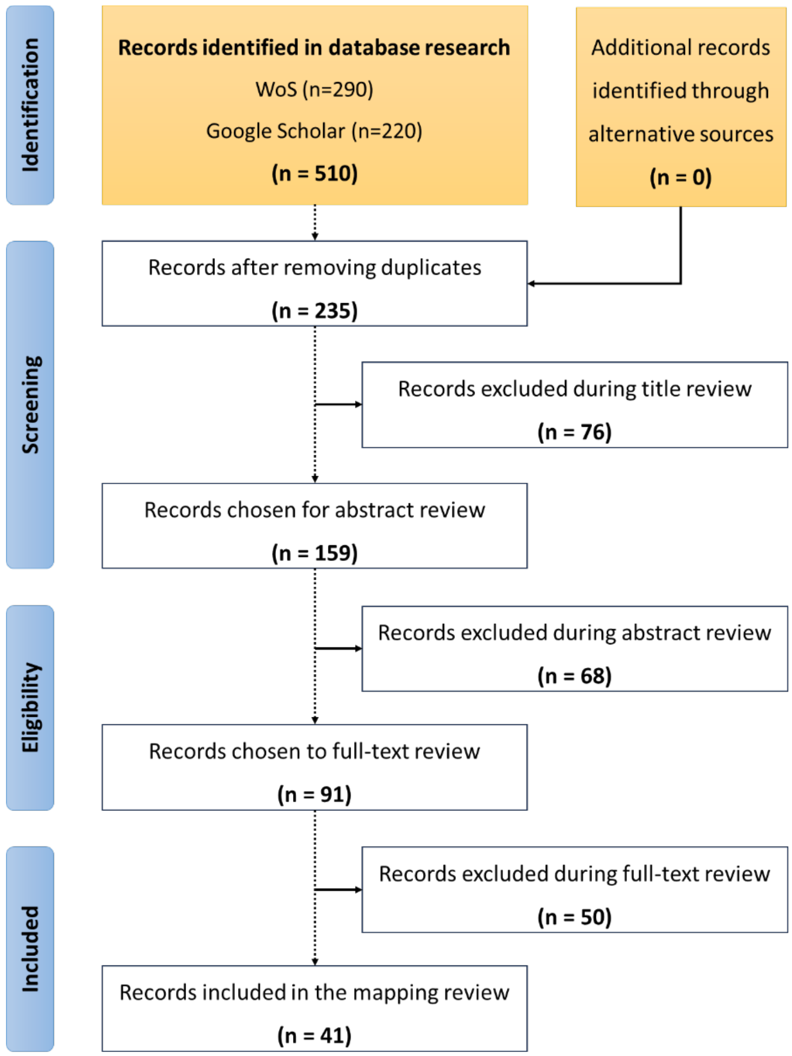


Figure 2. The PRISMA flow diagram for the systematic review

3. Application of ML and CFD in the built environment

Figure 3 shows the distribution of relevant research articles published from 2012 to 2022, focusing on integrating machine learning techniques with computational fluid dynamics for built environment applications. The results reveal a notable increase in publications in this emerging field over the past decade, with a particularly sharp uptick around 2019. This growing research interest highlights the promise of machine learning to enhance CFD simulations and indoor environment modeling. However, limitations remain regarding data availability, model interpretability, and computational complexity. The trend emphasizes the need for continued innovations to address these challenges and fully realize the potential of combining data-driven and physics-based approaches for built environments. Ongoing advances in machine learning, improved benchmark datasets, and increased computational capabilities can help accelerate progress.

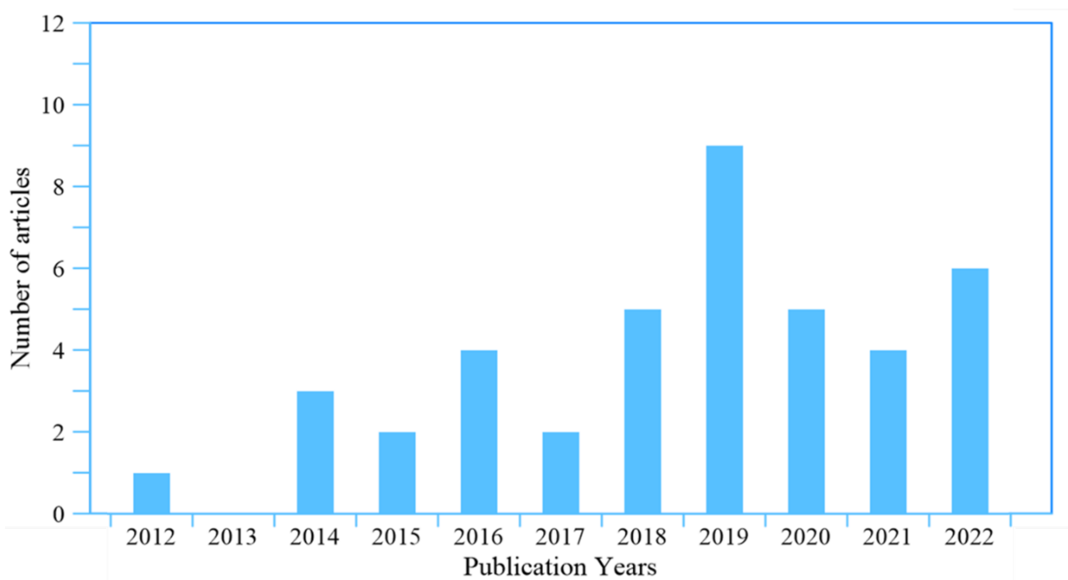


Figure 3. The number of articles published over time

To identify frequently occurring themes and relationships between keywords, a co-occurrence analysis was conducted on the extracted keywords from the literature review, as shown in Figure 4. This analytical approach determined the frequency of keyword appearances in the reviewed publications, enabling the formation of clusters of interrelated terms.

Three clusters emerged from the analysis. Cluster#1 contains keywords related to machine learning techniques, including "artificial neural network," "deep learning," "neural network," "machine learning," and associated algorithms such as "convolutional neural network," "recurrent neural network," "generative adversarial network." This cluster represents the machine learning approaches leveraged in conjunction with CFD.

Cluster#2 includes keywords about CFD methods and concepts, such as "computational fluid dynamics," "fluid dynamics," "turbulence modeling," "large eddy simulation," and "Reynolds averaged Navier-Stokes." This cluster highlights the simulation tools and methodologies employed in modeling.

Cluster#3 encompasses keywords relevant to application areas within built environments, like "indoor air quality," "ventilation," "thermal comfort," "contaminant dispersion," "airflow distribution," "temperature prediction," "built environment," and "indoor environment." This cluster signifies the application domains for the deployment of machine learning-enhanced CFD techniques.

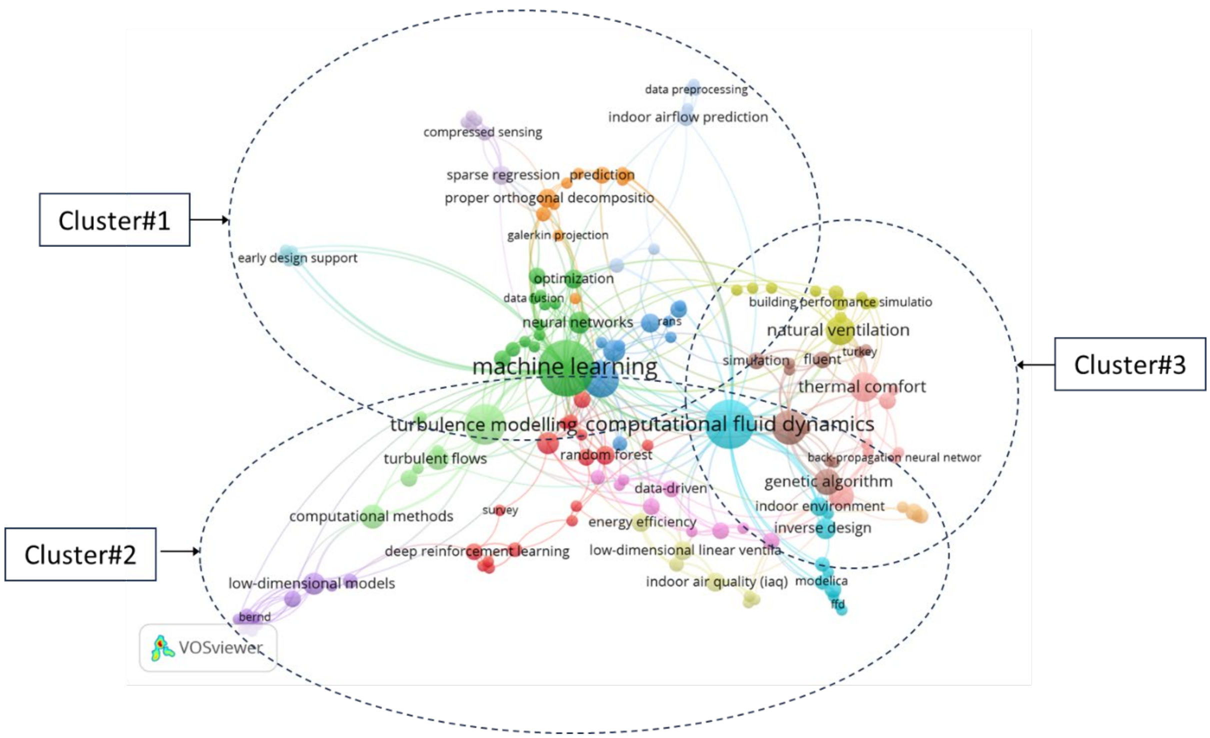


Figure 4. Visualization of keywords' co-occurrence analysis

Figure 4's visualization of the keywords' co-occurrence analysis reveals that research intersecting machine learning and CFD in indoor environments often integrates relevant techniques, simulations, and applications pertinent to built environment systems. Furthermore, the patterns of co-occurrence assist in identifying the relationships between clusters, such as the connections between machine learning algorithms and the adopted modeling approaches.

3.1. Using ML algorithms in CFD to predict thermal comfort

3.1.1. Studies employing ML-CFD integrate for thermal comfort prediction

Thermal comfort in indoor environments is affected by a variety of interrelated factors. Figure 5 illustrates the significant elements that impact occupants' thermal comfort, including 1) Environmental factors like air temperature, mean radiant temperature, air velocity, and relative humidity; 2) Personal factors such as occupants' metabolism and clothing insulation; and 3) Design factors including building layout, materials, and systems like HVAC. Carefully considering and optimizing these factors through computational fluid dynamics and machine learning can help maximize occupants' thermal comfort.

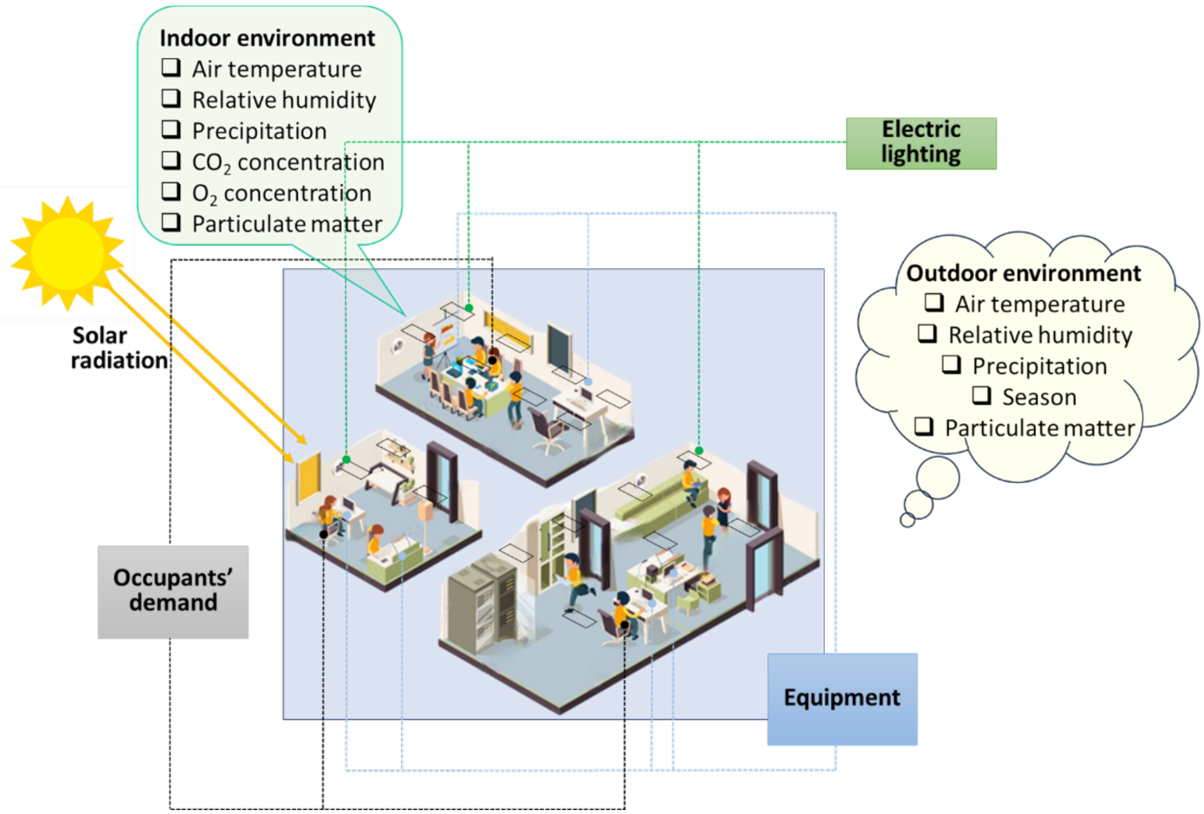


Figure 5. Factors that influence building thermal comfort

In the practical implementation of building control systems, analyzing air distribution data from airflow velocity and air temperature sensors plays a pivotal role. Data on air temperature forms the basis for calculating power consumption and crucial thermal comfort metrics, notably the Predicted Mean Vote (PMV) [51]. However, the need for additional sensors and associated infrastructure increases as the push for comprehensive indoor air quality (IAQ) metrics increases. This divergence becomes apparent when comparing indoor air quality in active areas with readings from reference sensors used in mixed ventilation scenarios. Moreover, installing sensors in what is commonly called the "occupied area" can potentially impact a significant portion of the available workspace [51].

Measuring the age of air, an essential IAQ indicator, is inherently intricate. A frequently adopted approach involves tracer gas-based techniques encompassing injection, decay, build-up, constant concentration, and pulse methodologies [52]. However, selecting the most suitable tools to address unidentified pollution sources for pollutant concentration measurements remains a formidable challenge [53]. This challenge underscores temperature distribution's vital role in shaping pollutant concentrations, indicating indoor air quality [54].

One promising avenue involves directly predicting IAQ by leveraging insights derived from air distribution data, potentially mitigating the labor and material costs associated with IAQ measurements. However, integrating such real-time or faster-than-real-time simulations within building management systems poses a significant challenge for traditional CFD approaches [55]. On the other hand, data-driven models like ANNs are equipped to handle this task. Several endeavors have sought to extract insights into the indoor climate. For instance, Li et al. (2012) [56] successfully predicted thermal comfort indices using a constructed monotonic type 2 fuzzy neural network. In a prior study, an ambiguous PMV/predicted percentage of dissatisfaction (PPD) model was employed to provide an effective thermal comfort reference signal [57]. ANNs have been deployed to estimate local thermal comfort levels in cabins, albeit with a relative uncertainty of approximately 10% [58]. For predicting indoor pollutant concentrations for IAQ improvements, ANNs have been instrumental in providing specific concentration data based on virtual sensors [59]. Additionally, a novel approach employing a wavelet neural network, a subtype of ANN, has been developed to assess indoor air quality within buildings comprehensively [60]. However, while showcasing the predictive capabilities of ANN models for environmental variables, it is essential to note that these studies primarily focus on system optimization for logicality or correctness, potentially overlooking nuances that genuinely mirror occupants' surroundings.

Efforts to quantitatively predict air temperature and velocity within occupied spaces based on supply and exhaust conditions in stratified environments have been explored [61]. Data-driven models offer higher precision in such predictions, as their coefficients can adapt to varying magnitudes, enhancing flexibility [61]. However, it is crucial to acknowledge that the quantification of indoor air quality remains challenging in this context.

The construction of low-energy buildings and the investigation of variations in indoor thermal comfort, driven by alterations in architectural features, can be facilitated by examining various geometric configurations. This can be swiftly accomplished by conducting parametric CFD analyses that illustrate the ideal architectural design. This methodology was employed in numerous prior studies using velocity or PMV indices as objective metrics for thermal comfort [62]. However, the CFD parametric analysis domain is inherently discontinuous, limiting the precision of this approach, even though it may offer near-optimal solutions. Consequently, the optimal solution might remain concealed within the discretized domain of pre-selected design variables. Densifying the grid of design variables might reduce the divergence of near-optimal solutions, but this could incur prohibitively high computational costs.

Figure 6 illustrates the geographical distribution of publications from 2012-2022 applying CFD-ML techniques for thermal comfort predictions in buildings. The size of each country corresponds to the number of associated publications. Each country's prominence in the graphic corresponds to the volume of related journals. Notably, China emerges as the frontrunner, boasting the highest number of publications, trailed by Hong Kong, the United States, and Singapore. This pattern underscores China's pivotal position in propelling innovations at the intersection of CFD and ML, particularly in predicting thermal comfort within built environments.

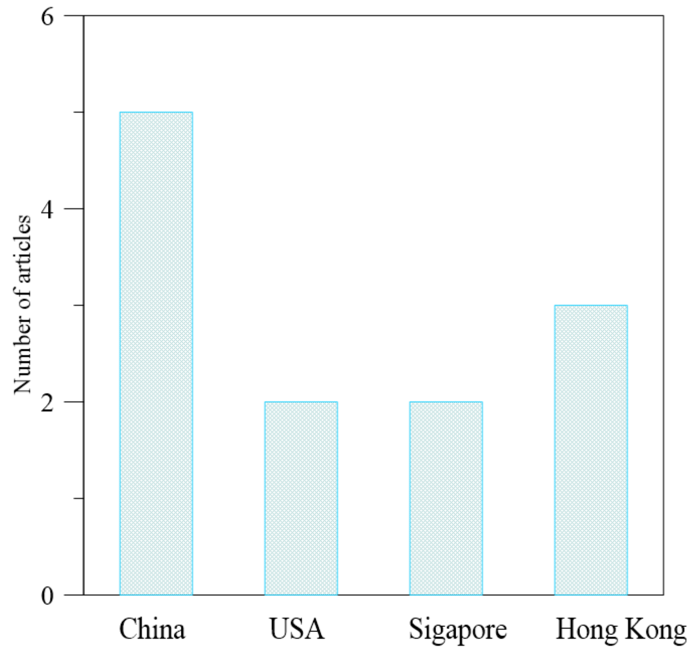


Figure 6. Publications on thermal comfort in buildings using CFD-ML techniques by country

3.1.2. Summary of the limitations

While these studies showcase the potential of ANN models in predicting environmental variables, it is essential to recognize their primary focus on optimizing system performance for consistency and correctness. This approach may not fully capture the intricacies of stratified environments and the real-world experiences of occupants. Additionally, although polynomial models show promise in forecasting indoor air temperature based on supply and exhaust conditions, as well as quantitatively estimating air temperature and velocity in stratified occupied spaces [61], their reliance on fixed coefficients can limit adaptability. In contrast, data-driven models offer greater accuracy and flexibility by allowing coefficients to vary, enhancing their precision in predicting these complex scenarios [61]. Table 1 summarizes studies employing machine learning techniques in conjunction with CFD to predict thermal comfort in buildings.

Table 1. Summary of the studies using deep learning and CFD techniques in thermal comfort in buildings

Ref	Year	DL model	CFD turbulence	Configuration	Main findings	Limitations
[63]	2020	LRWSGD, RF, ANN	RANS k-ε model	The vehicle cabin	The paper successfully combined high-fidelity CFD simulations with machine learning models, particularly ANN, to predict vehicle occupant thermal comfort across various conditions, achieving a test error of less than 5% for Equivalent Homogeneous Temperature (EHT) and enabling real-time predictions without relying on computationally expensive CFD simulations.	The limitations of this paper include potential data dependency, the need for further validation with real-world data, the computational complexity of high-fidelity CFD simulations, questions about the models' generalizability, and the absence of explicit consideration of human variability in thermal comfort responses.
[64]	2022	ANN	The RNG k-ε model	A typical consultation room	The paper's main findings indicate that a back-propagation neural network model utilizing air supply and exhaust temperatures as inputs can accurately predict indoor environment indicators, including energy utilization, thermal comfort, and indoor air quality, in stratified environments for both heating and cooling conditions, with potential improvements through the addition of specific inputs like air velocity and genetic algorithms.	The limitations of this study include its focus on stratified environments, limited consideration of relative humidity effects, and potential inapplicability to other air distribution methods and personalized ventilation.
[65]	2022	ANN, GA	RANS k-ε model	Semi-outdoor stadium (Tianjin Tuanbo tennis stadium)	This paper presents a novel approach to optimize thermal comfort in semi-outdoor stadiums using machine learning, introducing the PCave evaluation index, adjusting simulation methods, and achieving an 8.96% improvement in thermal comfort through ANN and GA-based stadium shape design while addressing challenges in data interaction and model training.	The limitations of this paper include challenges related to data transmission between climate simulation and computational fluid dynamics software, potential computational complexities when considering all weather conditions, and the specificity of findings to the Tianjin Tuanbo tennis stadium, which may not apply universally to other stadium designs.
[66]	2022	KNN, GA	N/A	Commercial building	The paper presents a methodology utilizing BIM, CFD, and GA to optimize the placement of temperature and CO_2 sensors in a multi-zone indoor environment, achieving sensor coverage requirements and addressing variations in temperature and CO_2 concentration under different scenarios, with applicability to both new and existing buildings.	The limitations of this paper include simplified manikin representations of occupants, the omission of energy-saving insulation effects, a focus on steady-state simulations, neglect of variations due to furniture and room partitions, a narrow focus on CO_2 and temperature sensors, and the need for further research to consider additional factors and building types.

[67]	2021	NN	SST k–ω	Typical apartments for residential buildings	This article proposes a BIM-enabled and data-driven framework. Results from BIM CFD simulations can be used to create a more trusted data-driven machine learning model for predicting indoor thermal comfort. The neural network represents the spatial properties. It incorporates the possible effect of interzonal airflows in the study of thermal comfort using an adjacency matrix from the field of graph theory.	The limitations of this paper include a focus primarily on natural ventilation, computational demands associated with CFD simulations, reliance on comprehensive BIM data, limited applicability beyond apartment settings, the potential margin of error in predictions, and the need for further research to enhance data-driven models.
[68]	2015	K-means clustering algorithm	N/A	The EEE living lab at NTU, Singapore	The paper presents a localized HVAC control strategy that uses CFD data and clustering algorithms, showing that an approach based on the Extreme Learning Machine (ELM) algorithm outperformed the PMV formula in predicting thermal comfort, offering the potential for energy-efficient localized HVAC control based on accurate partitioning of indoor spaces.	The limitations of this paper include the need for further extensive experiments to determine the most effective method for combining temperature and air velocity data and the potential challenges in achieving highly accurate room partitioning based on human comfort sensation for optimal localized HVAC control.
[42]	2012	ANN	Steady-state conditions, RANS k-e	Commercial buildings in Hong Kong	The paper introduces a dynamic simulation method using an ANN that effectively models the annual changes in atrium thermal environments, including complex phenomena like temperature stratification, greenhouse, and chimney effects while maintaining reasonable computational efficiency.	The limitations of this paper include the preliminary nature of the proposed method, which is still in the stage of tentative exploration, and the challenges associated with coupling different modules, defining CFD conditions, and fitting the neural network.
[38]	2014	ANN, GA	N/A	Aircraft cabin	The paper presents an inverse design approach integrating ANN, CFD, and GA for achieving preset aircraft cabin environments, demonstrating improved prediction accuracy, substantial computational cost reduction, and sensitivity to the scale of the CFD database.	One limitation of this paper is that increasing the scale of the CFD database can sometimes lead to decreased prediction accuracy.

3.2. Using ML algorithms in CFD to predict indoor air quality

3.2.1. Studies using ML-based CFD approaches for indoor air quality prediction

Indoor air quality is influenced by pollutant sources within the space. Figure 7 shows the primary sources that contribute to poor indoor air quality: 1) Outdoor air pollutants can infiltrate into indoor spaces; 2) Building materials like carpets, furniture, and insulation can emit volatile organic compounds; 3) Activities like cooking, cleaning, and smoking generate particulate matter, gases, and VOCs; 4) Humans themselves are sources of bio effluents and exhaled carbon dioxide; 5) HVAC systems can distribute and amplify pollutants if not properly maintained; and 6) Various appliances and electronic devices release chemical emissions as well. Understanding and controlling these sources through proper ventilation, filtration, and material selection is critical for indoor air quality management.

Figure 7. Indoor air pollutants are major sources

Efficient prediction of ventilation performance is essential for developing effective indoor ventilation control systems. To this end, Fast Fluid Dynamics (FFD) was introduced as a tool for efficiently forecasting indoor airflow and environmental conditions [69], [70]. Linear ventilation models (LVMs) developed by Cao and Meyers [71] enable the reconstruction of concentration fields resulting from different indoor pollutant source distributions. A reduced-order linear ventilation model was developed to handle transient regimes, facilitating online monitoring and management of ventilation systems [72], [73]. However, continuous monitoring becomes complex and costly due to the cyclical nature of indoor air properties. To address these limitations, a faster approach involves using a CFD-based ANN model, which leverages insights from CFD simulations to estimate room airflow and other parameters more efficiently. This approach has been employed to predict and optimize indoor environments, such as airplane cabins [74]. The foundation for hybrid CFD-based ANN models lies in creating a high-resolution CFD database. Subsequently, after generating LVMs, a discrete LLVM is constructed to predict interior ventilation more swiftly and accurately. Building upon LLVM, ANN techniques are integrated to enhance prediction accuracy at a relatively lower cost. These LLVM-based ANN techniques are utilized to predict indoor.

Recent advancements in deep learning have enabled the replacement, enhancement, or support of CFD simulations in construction research, offering substantial acceleration in computation for predictive, control, and design purposes [75]. NN models have been employed to swiftly generate target fields and variables, thereby predicting, controlling, and designing the built environment more efficiently [40], [41], [76], [77], [78], [79], [80], [81]. For instance, in predicting the thermal comfort of occupants within a car cabin, Warey et al. (2020) [63] employed a feedforward NN, known as a multi-layer perceptron (MLP), which demonstrated exceptional real-time accuracy under various boundary conditions. Zhang et al. (2019) [80] integrated NN with GA to regulate local indoor air quality, improving efficiency by mapping the nonlinear relationship between synthetic pollutant indices and ventilation rates, thereby circumventing resource-intensive CFD simulations. Similarly, Zhang and You (2017) [82] applied an integrated NN to perform an inverse design of a cabin environment, significantly reducing computational costs compared to the CFD-GA hybrid technique. NN has also been employed to predict specific indoor airflow distributions using a CFD database, providing immediate output based on input boundary conditions [43], [83], [84].

In Figure 8, the dispersion of publications spanning 2012 to 2022 is depicted, revealing the utilization of coupled CFD-ML models in exploring Indoor Air Quality (IAQ). The countries are arranged according to the frequency of their IAQ studies employing CFD-ML methodologies, with the size of each nation corresponding to the number of such studies conducted. China emerges as the foremost contributor, boasting the highest count of publications, succeeded by South Korea, Japan, the United States, and Taiwan. This pattern underscores China's leading role in harnessing sophisticated data-driven techniques, particularly the integration of CFD and ML, for modeling indoor air quality.

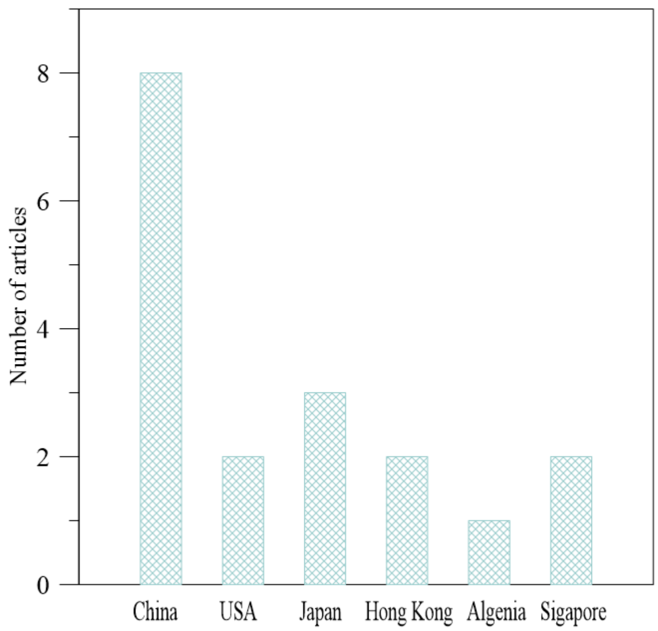


Figure 8. Publications on indoor air quality using CFD-ML techniques by country

3.2.2. Summary of the limitations

An effort has been made to integrate thermal comfort, air quality, and energy efficiency to improve ventilation systems' design configurations and operational states in office environments. However, the outcomes of earlier HVAC system control optimization research were considerable. There are three main reasons: First, it is necessary to collect comprehensive data on indoor airflow, pollutant distribution, and temperature fluctuations brought on by various ventilation systems to enhance IAQ, thermal comfort, and ventilation energy efficiency in office spaces. As a result, the problem area requires a high-resolution technique. Second, a variety of geometric and thermal factors, such as supply air conditions, office equipment and furniture configuration, ventilation strategies, air supply/return air connection configuration, outdoor and indoor temperature conditions, pollutant source location and emission rate, etc., have a significant impact on the ventilation performance of a given office. Therefore, testing a vast design space requires using a flexible modeling technique. Additionally, consideration should be given to the fact that the target function of such an optimization issue is a nonlinear mixed-integer function with several optima when choosing an optimization strategy. Table 2 overviews the examined publications using ML approaches in CFD for indoor environments.

Table 2. Summary of the studies using ML and CFD techniques for indoor environment

Ref.	Year	DL model	CFD turbulence	Configuration	Main Findings	Limitations
[85]	2022	DNN	RANS k-ε model	Residential high-rise buildings	The paper introduces a novel approach for predicting natural ventilation in residential high-rise buildings, combining physics-based modeling with data-driven deep neural networks, demonstrating the effectiveness of this approach by achieving a 96% accuracy in predicting air changes per hour (ACH) and highlighting its potential for optimizing building designs to maximize natural ventilation while saving energy.	The limitations of this paper include a focus on a limited set of residential high-rise building types, potential regulations in the predictive capacity of the data-driven models for other building shapes, and a need for further refinement and validation of the proposed approach for broader application in building design optimization.
[86]	2022	NN	Laminar calculation	Test room	This paper introduces a novel coupled simulation framework that combines a neural network with a Modelica-based energy simulation program to achieve fast and accurate building energy simulations for non-uniform indoor environments, significantly reducing computational time and showing the potential of neural networks in improving simulation efficiency.	The limitations of this paper include the focus on hypothetical cases for validation, the lack of exploration into real-world applications, potential challenges related to neural network training, and the need to consider the trade-off between speed and accuracy when using the neural network surrogate for CFD.
[64]	2022	BP (ANN)	RNG k-ε	Typical consultation room	The paper's main findings indicate that a back-propagation neural network model utilizing air supply and exhaust temperatures as inputs can accurately predict indoor environment indicators, including energy utilization, thermal comfort, and indoor air quality, in stratified environments for both heating and cooling conditions, with potential improvements through the addition of specific inputs like air velocity and genetic algorithms.	The limitations of this study include its focus on stratified environments, limited consideration of relative humidity effects, and potential inapplicability to other air distribution methods and personalized ventilation.
[66]	2022	GA, KNN	N/A	Commercial building	The paper presents a methodology utilizing BIM, CFD, and GA to optimize the placement of temperature and CO_2 sensors in a multi-zone indoor environment, achieving sensor coverage requirements and	The limitations of this paper include simplified representations of occupants in CFD simulations, omission of insulation effects due to limited information, a focus on steady-state simulations, and the

					addressing variations in temperature and CO_2 concentration under different scenarios, with applicability to both new and existing buildings.	exclusion of factors such as human movement, furniture placement, and consideration of additional pollutants beyond temperature and CO2.
[78]	2022	ANN	Standard k-ε model	a scaled test chamber	This paper introduces an ANN model as an efficient and accurate alternative to CFD for predicting fine particle concentration in indoor air, demonstrating lower error rates and the ability to assess the impact of factors like velocity and air exhaust positions on flow regimes.	The limitations of this paper include its reliance on a database from bibliographic literature, potential challenges in capturing complex fluid dynamics phenomena using the ANN model, and the need for further exploration of the model's effectiveness in diverse real-world environments.
[87]	2020	LLVM-based ANN	RNG k-ε model.	Three-dimensional room	The paper presents a systematic methodology using the Fuzzy C-means (FCM) algorithm for sensor deployment in indoor environments, allowing for efficient monitoring, and demonstrates accurate CO2 concentration predictions utilizing this approach, which can contribute to developing intelligent ventilation systems.	The limitations of this paper include a focus primarily on CO2 concentration predictions, limited discussion on the scalability and adaptability of the clustering methodology to different building sizes and layouts, reliance on accurate low-dimensional CFD results, and the absence of cost considerations for sensor deployment.d
[43]	2020	DNN	RANS k-ε model	A two-dimensional office room	This paper finds that DNNs are promising for rapid and accurate prediction of indoor airflow distribution, with DNN architecture significantly affecting performance and DNNs being approximately 1.9 million times faster than traditional CFD simulations. However, they show limitations for cases beyond their training data.	The limitations of this paper include a focus on a two-dimensional case, difficulties with extrapolation beyond the training data range, and a narrow exploration of specific DNN architectures, potentially limiting its applicability to more complex scenarios and architectures.
[83]	2021	NN	RNG k–ε model	Two-dimensional room	This paper demonstrates that employing neural networks with different data preprocessing methods can yield fast and accurate predictions of indoor airflow and temperature distributions, with minor variations in performance. It emphasizes the necessity of preprocessing for handling variables with varying orders of magnitude.	The limitations of this paper include its focus on two-dimensional non-isothermal cases, leaving a gap in assessing the neural network's performance in more complex three-dimensional scenarios, and its exclusive consideration of indoor airflow and temperature predictions without addressing other indoor environmental factors like pollutant distribution.

[88]	2017	ANN, GA	N/A	The geometric configuration of the exhaust hood	This paper presents an effective method that combines ANN, GA, and CFD for the inverse optimization of exhaust hood design and operation, resulting in reduced emissions and minimized wall deposition of cut tobacco.	The limitations of this paper include a lack of specific details about the exhaust hood design variables and objectives and a lack of discussion regarding the computational complexity and resource requirements for implementing the proposed method in practical industrial settings.
[74]	2016	GA, ANN	N/A	Aircraft cabin	The paper demonstrates that using logarithm-normalized ANN in the inverse design of aircraft cabin environments significantly improves prediction success rates, reducing computational costs by 23.2% while maintaining solution quality, compared to linear normalized ANN.	The limitations of this paper include the lack of detailed descriptions of the specific aircraft cabin environmental parameters, potential errors arising from using ANN as a surrogate for CFD in complex design scenarios, limited real-world validation beyond the MD-82 cabin environment, and a lack of in-depth discussion regarding the computational requirements for implementing this method.
[89]	2014	GA, ANN	N/A	Simplified first-class cabin	The paper presents an inverse design approach utilizing ANN and GA combined with CFD analysis to efficiently optimize preset aircraft cabin environments (ACE) by significantly reducing computational costs (57%) while achieving improved prediction accuracy compared to GA alone.	One limitation of this paper is that as the scale of the CFD database increases, the prediction accuracy can sometimes decrease, and the size of the CFD database influences the reduction in computational costs.
[90]	2014	GA, ANN	N/A	Cabin	The article presents an inverse design method that combines ANNs and GAs to optimize the air supply parameters of an aircraft cabin environment, achieving pre-set thermal comfort objectives while considering energy consumption, with the Bayesian regularization algorithm demonstrating superior generalization capability among ANN training methods and identifying trade-offs between thermal comfort and energy consumption.	The limitations of this article include its reliance on a simplified cabin model that may not fully represent all aircraft configurations, the need for further validation of the proposed method through real-world testing, and a lack of in-depth discussion regarding the computational resources required for scaling the approach to larger and more complex cabin environments.
[91]	2015	GA, ANN	N/A	Cabin	The paper's main findings include the identification of a superior training method, a combination of genetic algorithm and particle swarm optimization, for artificial neural networks used in predicting indoor environments based on computational fluid dynamics, as well as the	The limitations of this paper include a focus on indoor environment prediction using computational fluid dynamics, which may not cover all aspects of indoor environment design, and the evaluation of proposed

					introduction of a local logarithm normalization method to enhance prediction accuracy in scenarios where specific design space positions are more critical.	methods primarily within the context of an MD-82 aircraft cabin, potentially limiting their generalizability to other indoor spaces.
[92]	2018	N/A	N/A	Chamber of stratum ventilation	The paper presents a method to model non-uniform thermal environments in stratum ventilation systems using readily available supply and exit air conditions as inputs. It demonstrates its accuracy in predicting indoor air temperature and velocity distributions under various conditions and its applicability to dynamic control.	The limitations of this paper include the omission of relative humidity considerations, the simplicity of the proposed linear models, limited validation scope, and the assumption of homogeneous exit air properties in stratum ventilation scenarios.
[93]	2018	ANN; LLVM-based ANN	RAN k - ε	3D chamber	The paper presents a ventilation control strategy based on LLVM and ANN that effectively predicts CO_2 concentration levels, leading to substantial reductions in indoor pollutant concentration and energy consumption, with potential reductions of up to 30% and 50%, respectively.	The limitations of this paper include potential simplifications for complex geometric rooms, the need to consider additional influential factors in larger and irregularly shaped spaces, and the necessity to improve the assessment index (Ev) for more realistic weighting factors, among other factors.
[94]	2019	ANN; LLVM-based ANN	LLVM RNG k - ε	3D chamber	The main findings demonstrate that incorporating LLVM-based ANN methods with strategically positioned sensors can efficiently predict indoor pollutant concentration, enabling intelligent ventilation control.	One limitation of this paper is that it primarily focuses on indoor pollutant concentration prediction and may not fully address more complex indoor environments or multiple pollutants.
[80]	2019	ANN; GA; and integration of ANN-GA	RNG k - ε	3D room	The paper introduced a novel index for assessing local IAQ based on computational simulations, compared two ventilation models, and found that the up-inlet and up-outlet models exhibited superior pollutant resistance while also highlighting the varying efficiency of ANN, GA and their integration for IAQ control depending on the correlation between control variables and objectives.	The limitations of this paper include the absence of real-world experimental validation for the proposed IAQ index, potential challenges in applying the integrated ANN and GA control method when there is no strong correlation between control variables and objectives, and the need for further exploration of this approach in more complex indoor environments and with additional pollutants.
[95]	2019	N/A	N/A	N/A	The paper reviews various fast prediction models for indoor environmental design and control, highlighting the effectiveness of models like LLVM and Markov chain for predicting air pollutant motion	The paper does not provide empirical validation or specific case studies to demonstrate the practical applicability and performance of the

					when turbulent airflow information is available while also suggesting the potential of FFD-based Proper Orthogonal Decomposition and machine learning to reduce computing effort in indoor air quality prediction significantly.	discussed fast prediction models in real-world indoor environmental design and control scenarios.
[96]	2019	MLR model	RNG k - ε	generic high-density urban	This paper presents an urban-scale computational model for wind-driven cross ventilation in high-density cities, introduces a novel ventilation index (CIOIv) that considers indoor and outdoor wind environments, and demonstrates the effectiveness of data-driven machine learning models, particularly Gradient Boosting, in predicting CIOIv for real-time assessment of building natural ventilation potential, offering a practical and significantly faster alternative to traditional CFD simulations, particularly beneficial for early-stage design decisions in complex urban settings.	One limitation of this paper is that while it develops a data-driven model for predicting natural ventilation potential, it primarily focuses on wind-driven cross ventilation and does not consider other forms of natural ventilation or complex urban scenarios beyond high-density cities.

4. Machine learning for fluid mechanics application

4.1. Studies using ML-based CFD simulation for fluid mechanics predictions

In fluid mechanics, CFD is the predominant approach for flow analysis in wind engineering, offering established guidelines for its application. When studying airflow in urban areas, turbulence models like LES and RANS come into play [97]. While LES excels in capturing unsteady phenomena and provides high accuracy in wind velocity calculations, its extensive computational demands render it impractical for optimization tasks requiring more than 1,000 instances. Consequently, RANS emerges as the primary choice for analyzing local solid winds around structures despite its lower unsteadiness accuracy. Given the computational efficiency of RANS, it becomes a vital tool for optimization studies aiming to mitigate strong local winds around pedestrian-level buildings.

NN architectures are ingeniously employed to elicit specific physical responses in fluid mechanics applications. For instance, Ling et al. (2016) [98] designed a particular neural network layer to maintain Galilean invariance when modeling Reynolds tensors. Reynolds stress models leverage SINDy's sparse modeling method [99], [100], [101]. Hybrid models integrating nonlinear neural networks with linear system identification have effectively addressed complex aeroelastic systems [102]. The Hidden Fluid Mechanics (HFM) approach encodes Navier-Stokes equations, allowing for adaptable physics-based neural network technology to produce physically accurate flow field predictions from limited data [103]. Sensing technologies, such as light sensing, have been employed to reconstruct pressure distributions around wings [104]. Moreover, a unique operator network, the neural Fourier operator, proves adept at super-resolution upscaling and simulation modeling [105].

Equivariant convolution networks have emerged for addressing high-dimensional complex systems in fluid dynamics, enabling the imposition of symmetries [106]. Physical invariances in neural networks have also been incorporated for sub-grid-scale scalar flow modeling [107]. Recent advancements showcase the integration of deep convolutional autoencoder networks into reduced-order modeling frameworks, capitalizing on their superior dimensionality reduction capabilities [108]. Traditional RANS models often exhibit notable inaccuracies when dealing with intricate geometries, pronounced flows, strong anisotropy, and three-dimensional effects. Machine learning has emerged as a powerful tool for improving RANS turbulence modeling [109]. Techniques have focused on enhancing the accuracy of Reynolds stresses in a wide range of scenarios.

Ling et al. (2016) [98] introduced a novel architecture incorporating Galilean invariance, resulting in improved performance compared to conventional RANS models. These models have been tested for turbulent flow over undulating walls, a challenge for RANS models [110], [111]. Other machine learning-based methods, like physically-informed random forests, have enhanced RANS models [112], [113]. Jiang et al. (2021) [114] proposed an innovative RANS modeling paradigm based on a physically directed residue network, emphasizing the importance of interpretable models in turbulence modeling. The application of Sparse Identification of Nonlinear Dynamics (SINDy) has led to the development of RANS closure models [99], [100], [101]. The research highlights the significance of incorporating uncertainty quantification (UQ) into machine learning-based models [115], [116], [117].

Furthermore, machine learning has shown promise in accelerating the convergence of RANS simulations, as demonstrated by Obiols-Sales et al. 2020 [118]. Multiphase flows involving multiple thermodynamic phases have also been addressed through machine learning techniques, particularly in speeding up simulation times [119][120]. Supervised neural networks have been employed for modeling sub-grid-scale (SGS) phenomena, with significant advancements in SGS modeling in various contexts [121], [122], [123]. Machine learning techniques, such as Genetic Programming (GEP), have outperformed traditional LES models in SGS modeling [124]. Machine learning also plays a pivotal role in addressing challenges related to high Reynolds number flows, particularly in atmospheric boundary layers (ABLs) where the Reynolds number can be orders of magnitude higher than in typical turbulence modeling studies. Data-driven approaches enable the construction of maps between resolved information in the outer region and appropriate far-wall boundary conditions, offering potential solutions for turbulent channels and other complex flow scenarios [125][126], [127]. Various deep learning-based methods, including convolutional neural networks and reinforcement learning, have been explored to define unconventional boundary conditions in challenging fluid dynamics problems [128][129]. Using machine learning to determine unusual boundary conditions is a challenging but exciting study area.

Within Figure 9, the geographical representation of publications spanning 2012 to 2022 is portrayed, capturing the amalgamation of CFD and ML techniques in fluid mechanics research beyond conventional building contexts. The magnitude of each country on the map corresponds to the quantity of studies conducted. The United States and China emerged as frontrunners in the domain of CFD-ML publications for fluid mechanics applications, specifically in areas such as aerodynamics and turbulence modeling. Following this lead, the United Kingdom, France, Canada, and Germany also contribute significantly to the global research landscape at the intersection of CFD and ML, showcasing a collective pursuit of advancements in fluid mechanics.

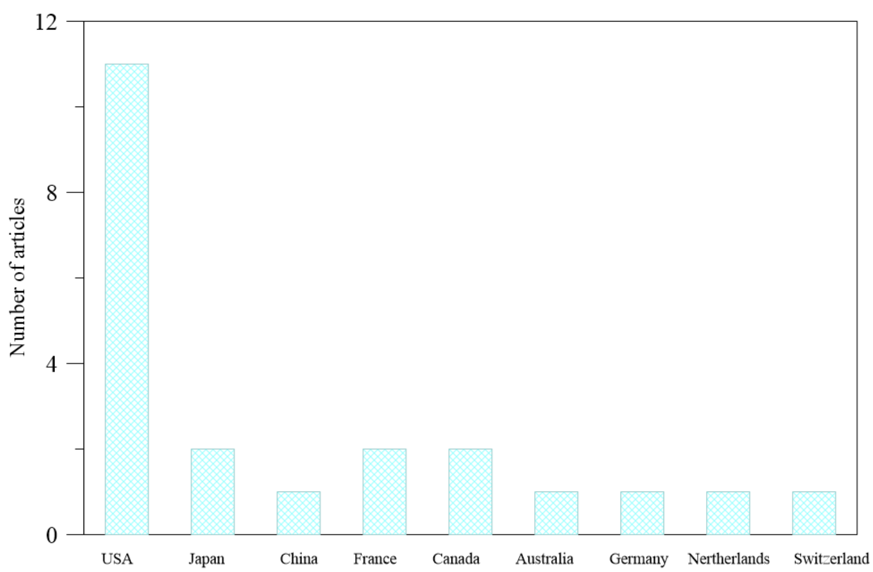


Figure 9. Publications on fluid mechanics applications using CFD-ML techniques by country

4.2. Summary of the limitations

The formalization of turbulence closure modeling as a machine learning problem is an important research area [109], [130], such as developing Reynolds stress models [98], [131] or turbulence on a smaller scale [132]. Another essential way to incorporate physics knowledge into turbulence closure modeling is to create user input features [113], [133], [134]. The problem of improving CFD solvers has recently attracted the attention of machine learning as well [135], [136], [137], [138]. The related topic of aerodynamics has made significant data-driven advances [139]. These problems involve the learning process in a broader physical framework, which immediately increases the physical relevance of the models. Fluid data is notoriously huge and highly dimensional, as each flow field must often be defined with millions of degrees of freedom. Since these flow fields often change, a time series of different snapshots is also created. Since it is time-consuming to investigate different geometries, Reynolds numbers, etc., numerically or experimentally, data in the parameter space are often sparse despite large spatial and temporal dimensions. Many fluid mechanics applications use custom neural network topologies to force physical responses. Table 3 summarizes the studies using deep learning models in fluid mechanics applications.

Table 3. Summary of the studies using deep learning model in fluid dynamics application

Ref	Year	DL model	CFD turbulence	Configuration	Main findings	Limitations
[140]	2019	CNN, parameter sharing, the gradient sharpening	RANS	Airfoil	The paper demonstrates that CNNs can predict aerodynamic flow fields, such as velocity and pressure, with remarkable speed and accuracy compared to traditional methods, enabling near-real-time simulation-based design and optimization for various airfoil shapes, albeit with the limitation of data diversity in training.	The limitations of this paper include its reliance on a limited training dataset, a focus on qualitative velocity field estimates rather than precise aerodynamic characteristics, and the need for further exploration and improvement of the model's predictive capabilities.
[141]	2018	Wasserstein GANs	DNS	N/A	This paper demonstrates the successful use of Wasserstein Generative Adversarial Networks (WGANs) to generate small-scale turbulent structures with qualitative agreement to DNS data, marking an important step toward predicting turbulent flows at different Reynolds numbers.	The limitations of this paper include a limited exploration of network architectures and hyperparameters, a focus on small-scale turbulence without addressing more significant flows, dependency on high-fidelity DNS data, and the absence of discussions regarding computational resource requirements for practical applications.
[142]	2019	MLP, CNN	N/A	N/A	This paper classifies five machine learning frameworks for data-driven thermal fluid models. It demonstrates that Type III models, known as Physics-Integrated Machine Learning (PIML), have the highest potential for leveraging "big data" in thermal fluid research despite the technical challenges they pose.	The paper primarily introduces innovative machine learning frameworks for data-driven thermal fluid models but lacks an in-depth exploration of practical implementation challenges and does not extensively address potential limitations or uncertainties associated with using these frameworks in diverse thermal fluid scenarios.
[143]	2018	MLP, SGD	N/A	Airfoil	The paper demonstrates that neural networks can effectively classify different vortex wake patterns using local vorticity measurements, even when the distinguishing features of these patterns are not easily recognizable, with potential for future improvements using more advanced network architectures and complex sensor data.	The limitations of this paper include the neural network's lack of shift-independence, the simplicity of relying on vorticity time series as input, and the absence of practical implementations or applications in real-world scenarios.
[45]	2019	CNN, hybrid DSC/MS	DNS	Two-dimensional cylinder	This paper demonstrates that machine learning models, including CNN and hybrid Downsampled Skip-Connection Multi-Scale models, can	The limitations of this paper include a focus on image-like representations of flow fields, limited exploration of the models'

					effectively perform super-resolution reconstruction of laminar and turbulent flow fields from low-resolution data, with significant potential to reveal subgrid-scale physics in fluid flows.	performance in diverse flow scenarios, and the absence of a discussion on real-world implementation challenges and uncertainties in experimental or computational fluid dynamics applications.
[144]	2019	CNN, MLP	DNS	Synthetic turbulent inflow generators	This paper presents a machine-learned turbulence generator that accurately reproduces turbulence statistics and spatio-temporal flow development. It provides a cost-effective alternative to conventional methods while acknowledging its sensitivity to machine learning parameters and the ongoing need for data availability during training.	The study's limitations include the requirement for spatiotemporal data of the target flow for training, sensitivity to machine learning parameters, and further exploration and optimization of network structures, potentially limiting its applicability and generalization to different flow scenarios.
[145]	2020	Monte Carlo (MC) and Quasi-Monte Carlo (QMC)	RAN	RAE 2822 Airfoil	The paper presents a machine learning algorithm using DNN for predicting input parameters to observable maps in computational fluid dynamics, achieving low prediction errors with significantly reduced computational costs, and demonstrates its application for efficient uncertainty propagation.	This paper's limitations include the careful selection of network architectures and hyperparameters, potential challenges related to generalization in data-poor scenarios, and the need for further validation across a broader range of fluid dynamics problems.
[146]	2021	ANN and GA	LES/RAN	The three-dimensional porous media	The paper presents a hybrid machine learning approach using ANN and GA to predict permeability in porous media based on pore structure parameters, achieving robust predictions with high accuracy.	The limitation of this paper is that it focuses on predicting permeability based on pore structure parameters, which may not capture all the factors influencing permeability in real-world porous media.
[147]	2021	CNN	RAN	2D randomly shaped obstacles	The paper presents DeepCFD, a CNN model that efficiently approximates solutions for non-uniform steady laminar flows described by the Navier-Stokes equations, achieving significant computational speedup while maintaining low error rates compared to traditional CFD approaches.	The paper's limitations include a focus on 2D flow configurations, a need for future work to address 3D flows and turbulence, and a lack of incorporation of physical constraints into the neural network training procedure.
[148]	2016	Markov process, Q-learning	N/A	2D cylinder	The paper presents a data-driven closed-loop control strategy for complex fluid flow systems that leverages hash functions to create a discrete state space from limited sensor measurements, enabling reinforcement learning-based control without prior knowledge, which is demonstrated	One limitation of this paper is that the proposed control strategy may face challenges when dealing with systems with highly complex dynamics or requiring fine-grained control due to the potential need for many state space entries.

					effectively in controlling transitions in a Lorenz'63 system and reducing drag in a cylinder flow.	
[149]	2018	BMSA	RANS	Transonic flow over a 3D wing	The paper introduces Bayesian model-scenario averaging (BMSA) to address uncertainty in RANS simulations, offering quantified uncertainty estimates and demonstrating its potential for predicting complex flow scenarios with trustworthiness insights.	The paper's computational cost for propagating posterior distributions in the BMSA approach makes it impractical for industrial-relevant test cases, necessitating Dirac-δ function approximations to reduce computational demands.
[150]	2016	CNN	N/A	Signed Distance Function (SDF) sampled on a Cartesian grid as the geometry representation	The paper demonstrates that CNN can provide real-time predictions of non-uniform steady laminar flow in 2D or 3D domains with a significant speed advantage over traditional CFD solvers while maintaining low error rates, making them suitable for immediate feedback and design space exploration.	The paper's applicability is primarily demonstrated for steady-state laminar flows, limiting its scope regarding more complex flow scenarios.
[151]	2017	MLP, RF regression	N/A	Lid-driven cubic	The paper demonstrates the feasibility of using ML algorithms, specifically ANN and RFR, to construct surrogate models that effectively correct errors in Coarse Grid Computational Fluid Dynamics (CG-CFD) simulations, providing computationally efficient and accurate predictions for a range of new cases with different Reynolds numbers and grid sizes.	The paper's approach relies on the assumption that coarsening the grid-primarily introduces discretization errors while neglecting model discrepancies, potentially limiting its applicability in scenarios where model errors play a more significant role.
[152]	2019	MLP	RANS	A two-dimensional S809 airfoil	The main findings of this paper are that integrating ML training into the field inversion process for RANS modeling improves the consistency of model discrepancy corrections, resulting in learnable model discrepancies and demonstrating feasibility for data-augmented modeling of turbulent flow over a 2D airfoil.	The limitations of this paper include the complexity and computational cost of neural network training, especially with thousands of iterations, and the need for appropriate training data, which can be limited for high Reynolds number flows and complex geometries.
[153]	2016	GP, ensemble Kalman method,	RANS	Periodic Hills	The paper presents an open-box, physics-informed Bayesian framework for quantifying and reducing model-form uncertainties in RANS simulations, effectively improving the agreement between RANS model	A notable limitation of the paper is that the inferred Reynolds stress field may not be accurate due to the high dimension of the Reynolds stress uncertainty space, sparseness of velocity observation data, and nonlinear

		Bayesian inference			predictions and benchmark data by incorporating empirical prior knowledge, physical constraints, and sparse observation data.	mapping between Reynolds stresses and velocities, but the potential for extrapolation mitigates this limitation to cases with similar characteristics.
[154]	2020	MLP, batch normalization, ensemble	N/A	Multi-dimensional combustion manifold	The paper introduces a novel approach using deep neural networks to efficiently model complex multi-dimensional combustion manifolds, significantly improving combustion simulations and reducing computational requirements.	One limitation of the paper is that the proposed neural network model does not adequately perform for predicting a specific variable (HR) with an extensive output range, which suggests room for improvement in handling such high-dimensional outputs.

5. Discussion and Future Directions

5.1. Challenges in machine learning for fluid dynamics

A key challenge in ML is recognizing the relative merits of different methods. The reliance on hyperparameter tuning, where hyperparameters are adjusted to specific problems, further compounds this challenge by making broad comparisons across various problems more difficult [155]. Therefore, the advantages of physics-informed features, labels, architectures, and loss functions over other machine learning approaches remain unclear, especially since some physics-informed machine learning methods have shown limited effectiveness or unintended consequences in specific test cases [156].

Fortunately, with the increasing application of ML in fluid mechanics, progress is being made in addressing these challenges. Established areas of machine learning, such as computer vision and natural language processing, have developed a culture of benchmarking different approaches through organized comparative evaluations using consensus data sets [157], [158]. Adopting a similar approach in fluid mechanics research could be helpful by allowing more standardized benchmark studies across multiple studies. However, two key developments are still required to realize the benefits of this approach entirely. First, representative, community-accepted datasets must be created to serve as benchmarks. Secondly, organized community benchmarking initiatives must be implemented at a community-wide level. Achieving these two goals would represent significant progress in standardizing assessment methods and addressing the challenge of reliably comparing different machine learning techniques for fluid mechanics applications [159].

5.2. Adapting turbulence models for the built environment

A key difference when simulating engineered environments is that the flows are typically low velocities, as opposed to the high velocity flows for which most RANS turbulence models were initially developed for aviation applications. As a result, the standard coefficients used in these models may not capture the full range of scenarios encountered in built environment simulations [160]. Addressing this potential shortcoming represents an opportunity for improved accuracy.

A practical approach to mitigating errors on important variables is fine-tuning the model coefficients using ANN [161]. However, it is essential to recognize the limitations of this approach. The optimized coefficients are usually explicitly tailored to the particular flow field under study and may not translate well to other cases. There is also a risk that neural networks may be inadvertently used to compensate for underlying analytical deficiencies rather than improving model accuracy [162]. A more promising research direction is to pursue methods to explicitly improve turbulence models for applications in low-velocity built environments, building on the advances made in other flow regimes. Adapting models to capture the unique properties of indoor flows better represents a potentially more effective long-term strategy.

5.3. Machine learning applications in computational fluid dynamics

One of the most common CFD machine learning applications is deriving RANS turbulence inferences using higher fidelity DNS and LES data. One of the most prevalent applications of machine learning in CFD is the derivation of RANS turbulence closures by utilizing higher fidelity DNS and LES databases. Weatheritt and Sandberg [163] introduced a gene expression programming (GEP) approach for model development to demonstrate this. This technique can regress tensor expressions and generate explicit algebraic formulations for quantities such as anisotropy through symbolic regression. They used this method within the clear algebraic Reynolds stress framework proposed by Pope [164] to refine nonlinear expressions for anisotropy. This data-driven modeling approach showed potential for predicting scenarios with analogous physical properties. Using large-scale turbulence databases represents a promising way to improve RANS model accuracy across different flow regimes.

The utility of the GEP methodology was subsequently extended to more complex 3D flows and those within turbomachinery applications, highlighting its capabilities for data-driven modeling in fluid dynamics. It is worthwhile noting that the functional forms achievable through GEP can be somewhat limiting. Alternatively, ANNs offer enhanced flexibility and the capacity to explore a broader parameter space. Frey et al. [165] built upon the work of Ling et al. [161] and harnessed very shallow neural networks trained on DNS data. Their approach demonstrated accurately predicting turbulent mixing within specific flow regions, particularly emphasizing its importance in these zones.

5.4. Advancing Reynolds-Averaged models using machine learning

Using machine learning techniques, Reynolds-averaged models can be improved with LES and DNS data [162]. These methods include various ANN architectures and GEP, as described by Weatheritt et al. [166].

However, exploring alternative approaches for assessing model accuracy during RANS applications is essential. Zhao et al. [167] proposed a methodology involving the utilization of the updated flow field derived from a fresh RANS calculation, employing the latest constitutive flow candidate to validate the trained $\tau_{i,j}^{ML}$. While this approach mitigates issues arising from such nonlinearities, it does require additional computational resources. Continued examination of robust yet resource-efficient validation strategies remains an area of ongoing research to fully leverage the capabilities of data-driven modeling for turbulence closures.

5.5. Future directions and emerging applications

Exploring the applications of machine learning in CFD for both fluid mechanics and built environments revealed promising directions. Incorporating physical insights into modeling continually improves performance in these areas. Notable new areas include non-contact flow measurement based on wall data, which is critical to closed-loop control strategies. Furthermore, techniques such as high-resolution predictions and tools such as CNNs, autoencoders, and GANs have shown significant potential, particularly in the study of wall-bound turbulence.

Furthermore, integrating physical invariances and symmetries into machine learning models, a practice widely used in sub-grid scale modeling (SGS), reduced order models (ROMs), and geophysical flow analysis, is promising. Another research avenue lies in physics-informed neural networks (PINNs), which use deep learning to efficiently solve partial differential equations (PDEs), often outperforming traditional numerical techniques. The application of PINNs goes beyond fluid dynamics and encompasses various biomedical areas. To overcome the significant challenges of CFD, novel methods are essential, especially in the search for precise simulations with coarse resolution within three-dimensional wall-confined turbulent flows. These challenges play a significant role in high Reynolds number flows, where traditional meshing strategies can distort the generation of turbulent kinetic energy.

High-fidelity scientific computing will benefit significantly from machine learning, driven by advances in AI and machine learning as well as the rapid growth in computing power through on-demand cloud computing. Machine learning-assisted fluid dynamics facilitates data collection through CFD simulations within a structured framework (see Figure 10). The data set is typically divided into two halves. The first subset forms the training data set, which serves as the basis for model training. The test data set's second subset evaluates the model's performance. The data allocation to training and testing sets is usually expressed as a percentage between 0 and 1. The remaining part is allocated to the test set. For example, if the training set size is 0.7 (70%) and the test set size is 0.3 (30%), the set consists of 30% of the data.

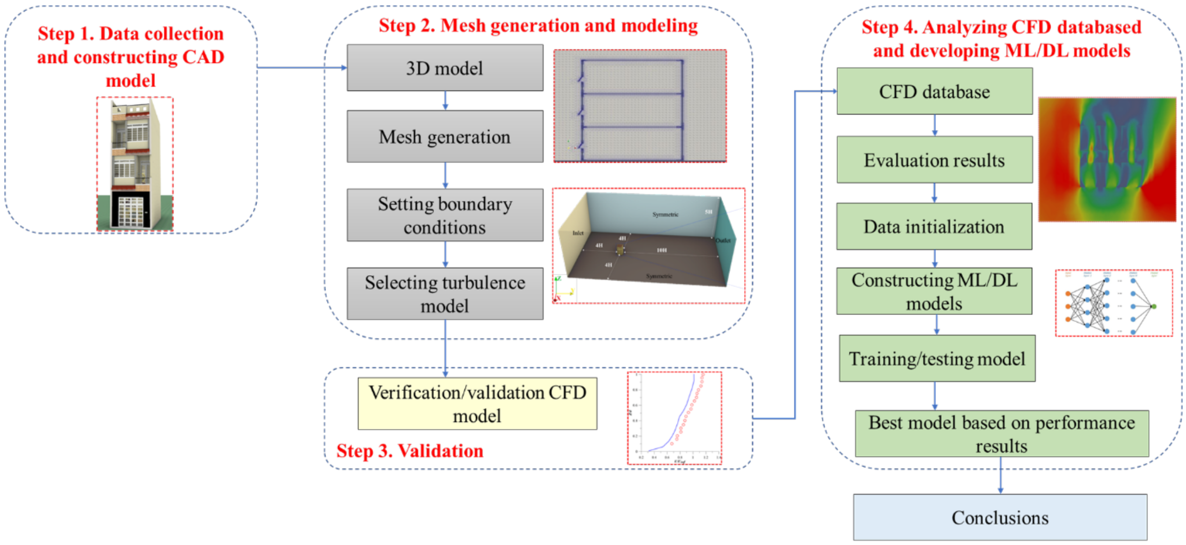


Figure 10. Framework for applying machine learning in fluid dynamics

6. Conclusion

This research contributes significantly to the nascent domain of ML applications in CFD for built environment modeling. The systematic literature review examined the integration of ML and CFD techniques for indoor thermal comfort and air quality prediction, analyzing cutting-edge innovations in the field. The discussion extends to advanced methods in fluid mechanics, highlighting promising techniques such as physics-informed neural networks for simulations relevant to the built environment.

Despite these advancements, challenges persist, primarily due to the intricate task of encapsulating fluid physics within ML models and the need for high-quality experimental data for model training and validation. Issues with the interpretability of black-box models and the computational intensity of high-fidelity CFD simulations also pose significant hurdles. This review suggests that adapting turbulence closures, like RANS models, could be a beneficial approach to represent indoor airflow characteristics better. It also advocates for enhanced model validation using real-world datasets and the implementation of physics-based regularization to refine these models.

The research clarified elucidates the interdependent relationship between ML and CFD, while offering strategies to surmount existing limitations and enhance the modeling of indoor environments. It spotlights emerging data-driven techniques in fluid mechanics that have potential applicability in the built environment sector. Future research directions should include the collaborative creation of benchmark datasets, conducting comparative assessments of ML models, and exploring innovative approaches like digital twins. This points to exciting prospects at the convergence of artificial intelligence and physics-based modeling, aimed at augmenting simulations' precision, efficiency, and scalability to optimise indoor spaces.

Acknowledgments

This research has been supported with a Presidential Scholarship from Kyung Hee University, Republic of Korea.

The authors would like to express our sincere appreciation to The Human Behaviour and Energy Laboratory (HuBEL) at the Kyung Hee University, the Republic of Korea, supported facilitates research.

References

[1] F. C. Goldizen, P. D. Sly, and L. D. Knibbs, "Respiratory effects of air pollution on children," *Pediatric pulmonology*, vol. 51, no. 1. John Wiley and Sons Inc., pp. 94–108, Jan. 01, 2016. doi: 10.1002/ppul.23262.

[2] Z. Liu, Q. Liang, H. Liao, W. Yang, and C. Lu, "Effects of short-term and long-term exposure to ambient air pollution and temperature on long recovery duration in COVID-19 patients," *Environ Res*, vol. 216, Jan. 2023, doi: 10.1016/j.envres.2022.114781.

[3] Y. He, W. Jiang, X. Gao, C. Lin, J. Li, and L. Yang, "Short-term effects and economic burden of air pollutants on acute lower respiratory tract infections in children in Southwest China: a time-series study," *Environ Health*, vol. 22, no. 1, Dec. 2023, doi: 10.1186/s12940-023-00962-3.

[4] M. Shen *et al.*, "Declining Pulmonary Function in Populations with Long-term Exposure to Polycyclic Aromatic Hydrocarbons-Enriched PM2.5," *Environ Sci Technol*, vol. 52, no. 11, pp. 6610–6616, Jun. 2018, doi: 10.1021/acs.est.8b00686.

[5] M. Svartengren, R. Falk, L. Linnman, K. Philipson, and P. Camner, "Deposition of Large Particles in Human Lung," 1987.

[6] Y. Yoda, N. Otani, S. Sakurai, and M. Shima, "Acute effects of summer air pollution on pulmonary function and airway inflammation in healthy young women," *J Epidemiol*, vol. 24, no. 4, pp. 312–320, 2014, doi: 10.2188/jea.JE20130155.

[7] K. Rumchev, J. Spickett, M. Bulsara, M. Phillips, and S. Stick, "Association of domestic exposure to volatile organic compounds with asthma in young children," *Thorax*, vol. 59, no. 9, pp. 746–751, Sep. 2004, doi: 10.1136/thx.2003.013680.

[8] C. J. Weschler, "Changes in indoor pollutants since the 1950s," *Atmos Environ*, vol. 43, no. 1, pp. 153–169, Jan. 2009, doi: 10.1016/j.atmosenv.2008.09.044.

[9] M. H. Garrett, M. A. Hooper, B. M. Hooper, P. R. Rayment, and M. J. Abramson, "Increased risk of allergy in children due to formaldehyde exposure in homes," *Allergy*, vol. 54, no. 4, pp. 330–337, 1999.

[10] B. Pradhan *et al.*, "Utility of outdoor central site monitoring in assessing exposure of school children to ultrafine particles," *Science of The Total Environment*, vol. 859, p. 160162, Feb. 2023, doi: 10.1016/J.SCITOTENV.2022.160162.

[11] L. Morawska *et al.*, "Indoor aerosols: From personal exposure to risk assessment," *Indoor Air*, vol. 23, no. 6. pp. 462–487, Dec. 2013. doi: 10.1111/ina.12044.

[12] I. Myers and R. L. Maynard, "Polluted air - Outdoors and indoors," *Occupational Medicine*, vol. 55, no. 6. pp. 432–438, Sep. 2005. doi: 10.1093/occmed/kqi137.

[13] S. Antonsen *et al.*, "Exposure to air pollution during childhood and risk of developing schizophrenia: a national cohort study," *Lancet Planet Health*, vol. 4, no. 2, pp. e64–e73, Feb. 2020, doi: 10.1016/S2542-5196(20)30004-8.

[14] M. Thygesen *et al.*, "Exposure to air pollution in early childhood and the association with Attention-Deficit Hyperactivity Disorder," *Environmental Research*, vol. 183. Academic Press Inc., Apr. 01, 2020. doi: 10.1016/j.envres.2019.108930.

[15] M. A. Kioumourtzoglou *et al.*, "The association between air pollution and onset of depression among middle-aged and older women," *Am J Epidemiol*, vol. 185, no. 9, pp. 801–809, May 2017, doi: 10.1093/aje/kww163.

[16] M. C. Power, M. A. Kioumourtzoglou, J. E. Hart, O. I. Okereke, F. Laden, and M. G. Weisskopf, "The relation between past exposure to fine particulate air pollution and prevalent anxiety: Observational cohort study," *BMJ (Online)*, vol. 350, Mar. 2015, doi: 10.1136/bmj.h1111.

[17] H. Chen *et al.*, "Living near major roads and the incidence of dementia, Parkinson's disease, and multiple sclerosis: a population-based cohort study," *The Lancet*, vol. 389, no. 10070, pp. 718–726, Feb. 2017, doi: 10.1016/S0140-6736(16)32399-6.

[18] Y. Yin, J. K. Gupta, X. Zhang, J. Liu, and Q. Chen, “Distributions of respiratory contaminants from a patient with different postures and exhaling modes in a single-bed inpatient room,” *Build Environ*, vol. 46, no. 1, pp. 75–81, 2011, doi: 10.1016/j.buildenv.2010.07.003.

[19] Z. Zhang, W. Zhang, Z. J. Zhai, and Q. Y. Chen, “Evaluation of various turbulence models in predicting airflow and turbulence in enclosed environments by CFD: Part 2—comparison with experimental data from literature,” *HVAC and R Research*, vol. 13, no. 6, pp. 871–886, 2007, doi: 10.1080/10789669.2007.10391460.

[20] D. Zhao, J. Chen, Y. Zheng, Z. Huang, and J. Zheng, “Fine-grained parallel algorithm for unstructured surface mesh generation,” *Comput Struct*, vol. 154, pp. 177–191, Jul. 2015, doi: 10.1016/j.compstruc.2015.04.004.

[21] V. Haroutunian, M. S. Engelman, and I. Hasbani, “SEGREGATED FINITE ELEMENT ALGORITHMS FOR THE INCOMPRESSIBLE FLOW PROBLEMS NUMERICAL SOLUTION OF LARGE-SCALE,” 1993.

[22] G. Hauke, A. Landaberea, I. Garmendia, and J. Canales, “A segregated method for compressible eow computation. Part II: General divariant compressible eows,” *INTERNATIONAL JOURNAL FOR NUMERICAL METHODS IN FLUIDS Int. J. Numer. Meth. Fluids*, vol. 49, pp. 183–209, 2005, doi: 10.1002//d.999.

[23] T. Norton and D. W. Sun, “Computational fluid dynamics (CFD) - an effective and efficient design and analysis tool for the food industry: A review,” *Trends in Food Science and Technology*, vol. 17, no. 11. pp. 600–620, Nov. 2006. doi: 10.1016/j.tifs.2006.05.004.

[24] P. Sagaut, *homogeneous turbulence dynamics, 2nd edition*. 2018.

[25] Y. Zhao, H. D. Akolekar, J. Weatheritt, V. Michelassi, and R. D. Sandberg, “RANS turbulence model development using CFD-driven machine learning,” *J Comput Phys*, vol. 411, Jun. 2020, doi: 10.1016/j.jcp.2020.109413.

[26] Y. Zhang, D. Zhang, and H. Jiang, “Review of Challenges and Opportunities in Turbulence Modeling: A Comparative Analysis of Data-Driven Machine Learning Approaches,” *Journal of Marine Science and Engineering*, vol. 11, no. 7. Multidisciplinary Digital Publishing Institute (MDPI), Jul. 01, 2023. doi: 10.3390/jmse11071440.

[27] R. Caruana and A. Niculescu-Mizil, “An Empirical Comparison of Supervised Learning Algorithms,” in *Proceedings of the 23rd International Conference on Machine Learning*, in ICML ’06. New York, NY, USA: Association for Computing Machinery, 2006, pp. 161–168. doi: 10.1145/1143844.1143865.

[28] A. Coates, A. Ng, and H. Lee, “An Analysis of Single-Layer Networks in Unsupervised Feature Learning,” in *Proceedings of the Fourteenth International Conference on Artificial Intelligence and Statistics*, G. Gordon, D. Dunson, and M. Dudík, Eds., in Proceedings of Machine Learning

Research, vol. 15. Fort Lauderdale, FL, USA: PMLR, Oct. 2011, pp. 215–223. [Online]. Available: https://proceedings.mlr.press/v15/coates11a.html

[29] L. P. Kaelbling, M. L. Littman, and A. W. Moore, “Reinforcement learning: A survey,” *Journal of artificial intelligence research*, vol. 4, pp. 237–285, 1996.

[30] J. Schmidhuber, “Deep Learning in neural networks: An overview,” *Neural Networks*, vol. 61, pp. 85–117, 2015, doi: 10.1016/j.neunet.2014.09.003.

[31] T. Ayata, E. Arcaklioğlu, and O. Yildiz, “Application of ANN to explore the potential use of natural ventilation in buildings in Turkey,” *Appl Therm Eng*, vol. 27, no. 1, pp. 12–20, 2007, doi: 10.1016/j.applthermaleng.2006.05.021.

[32] T. Van Quang, D. T. Doan, N. L. Phuong, and G. Y. Yun, “Data-driven prediction of indoor airflow distribution in naturally ventilated residential buildings using combined CFD simulation and machine learning (ML) approach,” *J Build Phys*, vol. 47, no. 4, pp. 439–471, Jan. 2024, doi: 10.1177/17442591231219025.

[33] B. Khazaei, A. Shiehbeigi, and A. R. Haji Molla Ali Kani, “Modeling indoor air carbon dioxide concentration using artificial neural network,” *International Journal of Environmental Science and Technology*, vol. 16, no. 2, pp. 729–736, 2019, doi: 10.1007/s13762-018-1642-x.

[34] Z. Li, J. Dai, H. Chen, and B. Lin, “An ANN-based fast building energy consumption prediction method for complex architectural form at the early design stage,” *Build Simul*, vol. 12, no. 4, pp. 665–681, 2019, doi: 10.1007/s12273-019-0538-0.

[35] T. Van Quang and N. L. Phuong, “USING DEEP LEARNING TO OPTIMIZE HVAC SYSTEMS IN RESIDENTIAL BUILDINGS,” *Journal of Green Building*, vol. 19, no. 1, pp. 29–50, Feb. 2024, doi: 10.3992/jgb.19.1.29.

[36] G. M. Stavrakakis, P. L. Zervas, H. Sarimveis, and N. C. Markatos, “Optimization of window-openings design for thermal comfort in naturally ventilated buildings,” *Appl Math Model*, vol. 36, no. 1, pp. 193–211, 2012, doi: 10.1016/j.apm.2011.05.052.

[37] A. Afram, F. Janabi-Sharifi, A. S. Fung, and K. Raahemifar, “Artificial neural network (ANN) based model predictive control (MPC) and optimization of HVAC systems: A state of the art review and case study of a residential HVAC system,” *Energy Build*, vol. 141, pp. 96–113, 2017, doi: 10.1016/j.enbuild.2017.02.012.

[38] T. hu Zhang and X. yi You, “A simulation-based inverse design of preset aircraft cabin environment,” *Build Environ*, vol. 82, pp. 20–26, 2014, doi: 10.1016/j.buildenv.2014.08.002.

[39] T. H. Zhang and X. Y. You, “Applying neural networks to solve the inverse problem of indoor environment,” *Indoor and Built Environment*, vol. 23, no. 8, pp. 1187–1195, 2014, doi: 10.1177/1420326X13499596.

[40] T.-H. Zhanga; and X.-Y. You;, “Inverse design of aircraft cabin environment by coupling artificial neural network and genetic algorithm,” *HVAC and R Research*, vol. 20, no. 7, pp. 836–843, 2014, doi: 10.1080/10789669.2014.950895.

[41] S. Cao and C. Ren, “Ventilation control strategy using low-dimensional linear ventilation models and arti fi cial neural network,” *Build Environ*, vol. 144, no. July, pp. 316–333, 2018, doi: 10.1016/j.buildenv.2018.08.032.

[42] R. Qin, D. Yan, X. Zhou, and Y. Jiang, “Research on a dynamic simulation method of atrium thermal environment based on neural network,” *Build Environ*, vol. 50, pp. 214–220, 2012, doi: 10.1016/j.buildenv.2011.11.001.

[43] Q. Zhou and R. Ooka, “Comparison of different deep neural network architectures for isothermal indoor airflow prediction,” *Build Simul*, vol. 13, no. 6, pp. 1409–1423, 2020, doi: 10.1007/s12273-020-0664-8.

[44] Z. Song, B. T. Murray, and B. Sammakia, “Airflow and temperature distribution optimization in data centers using artificial neural networks,” *Int J Heat Mass Transf*, vol. 64, pp. 80–90, 2013, doi: 10.1016/j.ijheatmasstransfer.2013.04.017.

[45] K. Fukami, K. Fukagata, and K. Taira, “Super-resolution reconstruction of turbulent flows with machine learning,” *J Fluid Mech*, vol. 870, pp. 106–120, 2019, doi: 10.1017/jfm.2019.238.

[46] M. Raissi, P. Perdikaris, and G. E. Karniadakis, “Physics-informed neural networks: A deep learning framework for solving forward and inverse problems involving nonlinear partial differential equations,” *J Comput Phys*, vol. 378, pp. 686–707, Feb. 2019, doi: 10.1016/j.jcp.2018.10.045.

[47] J. Ling and J. Templeton, “Evaluation of machine learning algorithms for prediction of regions of high Reynolds averaged Navier Stokes uncertainty,” *Physics of Fluids*, vol. 27, no. 8, Aug. 2015, doi: 10.1063/1.4927765.

[48] R. Wang, K. Kashinath, M. Mustafa, A. Albert, and R. Yu, “Towards Physics-informed Deep Learning for Turbulent Flow Prediction,” in *Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, Association for Computing Machinery, Aug. 2020, pp. 1457–1466. doi: 10.1145/3394486.3403198.

[49] K. Duraisamy, G. Iaccarino, and H. Xiao, “Turbulence Modeling in the Age of Data,” *Fluid review*, pp. 1–23, 2019.

[50] D. Moher *et al.*, “Preferred reporting items for systematic reviews and meta-analyses: The PRISMA statement,” *PLoS Med*, vol. 6, no. 7, 2009, doi: 10.1371/journal.pmed.1000097.

[51] G. Cao, K. Sirén, and S. Kilpeläinen, "Modelling and experimental study of performance of the protected occupied zone ventilation," *Energy Build*, vol. 68, pp. 515–531, 2014, doi: 10.1016/j.enbuild.2013.10.008.

[52] G. Remion, B. Moujalled, and M. El Mankibi, "Review of tracer gas-based methods for the characterization of natural ventilation performance : comparative analysis of their accuracy," *Build Environ*, vol. 106, 2019.

[53] T. Salthammer, "Emerging indoor pollutants," *Int J Hyg Environ Health*, vol. 224, no. July 2019, p. 113423, 2020, doi: 10.1016/j.ijheh.2019.113423.

[54] S. H. Ho, L. Rosario, and M. M. Rahman, "Three-dimensional analysis for hospital operating room thermal comfort and contaminant removal," *Appl Therm Eng*, vol. 29, no. 10, pp. 2080–2092, 2009, doi: 10.1016/j.applthermaleng.2008.10.016.

[55] Z. Feng, C. W. Yu, and S. Cao, "Fast prediction for indoor environment : Models assessment," *Indoor and Built Environment*, vol. 28, no. 6, pp. 727–730, 2019, doi: 10.1177/1420326X19852450.

[56] C. Li, J. Yi, and M. Wang, "Monotonic type-2 fuzzy neural network and its application to thermal comfort prediction," *Neural Comput Appl*, 2012, doi: 10.1007/s00521-012-1140-x.

[57] R. Z. Homod, K. Salleh, M. Sahari, H. A. F. Almurib, and F. Ha, "RLF and TS fuzzy model identi fi cation of indoor thermal comfort based on PMV / PPD," *Build Environ*, vol. 49, pp. 141–153, 2012, doi: 10.1016/j.buildenv.2011.09.012.

[58] T. Zhang and X. You, "Improvement of the training and normalization method of artificial neural network in the prediction of indoor environment," *Procedia Eng*, vol. 121, pp. 1245–1251, 2015, doi: 10.1016/j.proeng.2015.09.152.

[59] J. Ren and S. Cao, "Incorporating online monitoring data into fast prediction models towards the development of arti fi cial intelligent ventilation systems," *Sustain Cities Soc*, vol. 47, no. February, p. 101498, 2019, doi: 10.1016/j.scs.2019.101498.

[60] L. Lei, W. Chen, Y. Xue, and W. Liu, "A comprehensive evaluation method for indoor air quality of buildings based on rough sets and a wavelet neural network," *Build Environ*, vol. 162, no. July, p. 106296, 2019, doi: 10.1016/j.buildenv.2019.106296.

[61] S. Zhang, Y. Cheng, C. Huan, and Z. Lin, "Modeling non-uniform thermal environment of stratum ventilation with supply and exit air conditions," *Build Environ*, vol. 144, no. August, pp. 542–554, 2018, doi: 10.1016/j.buildenv.2018.08.063.

[62] A. Bastide, P. Lauret, F. Garde, and H. Boyer, "Building energy efficiency and thermal comfort in tropical climates: Presentation of a numerical approach for predicting the percentage of

wellventilated living spaces in building...,” *Energy Build*, no. August, 2006, doi: 10.1016/j.enbuild.2005.12.005.

[63] A. Warey, S. Kaushik, B. Khalighi, M. Cruse, and G. Venkatesan, “Data-driven prediction of vehicle cabin thermal comfort: using machine learning and high-fidelity simulation results,” *Int J Heat Mass Transf*, vol. 148, p. 119083, 2020, doi: 10.1016/j.ijheatmasstransfer.2019.119083.

[64] X. Tian, Y. Cheng, and Z. Lin, “Modelling indoor environment indicators using artificial neural network in the stratified environments,” *Build Environ*, vol. 208, no. November 2021, p. 108581, 2022, doi: 10.1016/j.buildenv.2021.108581.

[65] R. Zhang, D. Liu, and L. Shi, “Thermal-comfort optimization design method for semi-outdoor stadium using machine learning,” *Build Environ*, vol. 215, no. January, p. 108890, 2022, doi: 10.1016/j.buildenv.2022.108890.

[66] J. C. P. Cheng, H. H. L. Kwok, A. T. Y. Li, J. C. K. Tong, and A. K. H. Lau, “BIM-supported sensor placement optimization based on genetic algorithm for multi-zone thermal comfort and IAQ monitoring,” *Build Environ*, vol. 216, no. December 2021, p. 108997, 2022, doi: 10.1016/j.buildenv.2022.108997.

[67] V. J. L. Gan, H. Luo, Y. Tan, M. Deng, and H. L. Kwok, “Bim and data-driven predictive analysis of optimum thermal comfort for indoor environment,” *Sensors*, vol. 21, no. 13, 2021, doi: 10.3390/s21134401.

[68] H. Zhou, Y. C. Soh, and X. Wu, “Integrated analysis of CFD data with K-means clustering algorithm and extreme learning machine for localized HVAC control,” *Appl Therm Eng*, vol. 76, pp. 98–104, 2015, doi: 10.1016/j.applthermaleng.2014.10.004.

[69] W. Liu;, M. Jin;, C. Chen;, R. You;, and Q. Chen, “Implementation of a fast fluid dynamics model in OpenFOAM for simulating indoor airflow,” *Numeri Heat Transf A Appl*, vol. 69, no. 7, pp. 748–762, 2016, doi: 10.1080/10407782.2015.1090780.

[70] W. Zuo and Q. Chen, “Real-time or faster-than-real-time simulation of airflow in buildings,” *Indoor Air*, vol. 19, no. 1, pp. 33–44, 2009, doi: 10.1111/j.1600-0668.2008.00559.x.

[71] S. J. Cao and J. Meyers, “On the construction and use of linear low-dimensional ventilation models,” *Indoor Air*, vol. 22, no. 5, pp. 427–441, 2012, doi: 10.1111/j.1600-0668.2012.00771.x.

[72] L. Vervecken, J. Camps, and J. Meyers, “Stable reduced-order models for pollutant dispersion in the built environment,” *Build Environ*, vol. 92, pp. 360–367, 2015, doi: 10.1016/j.buildenv.2015.05.008.

[73] S. Cao and J. Meyers, “Fast prediction of indoor pollutant dispersion based on reduced-order ventilation models,” *Build Simul*, pp. 415–420, 2015, doi: 10.1007/s12273-015-0240-9.

[74] T. Zhang; and X. You;, “Comparing the linear and logarithm normalized artificial neural networks in inverse design of aircraft cabin environment,” *Build Simul*, pp. 729–734, 2016, doi: 10.1007/s12273-016-0301-8.

[75] G. Calzolari and W. Liu, “Deep learning to replace, improve, or aid CFD analysis in built environment applications: A review,” *Build Environ*, vol. 206, no. June, 2021, doi: 10.1016/j.buildenv.2021.108315.

[76] W. Cao; and X. You;, “The inverse optimization of exhaust hood by using intelligent algorithms and CFD simulation,” *Powder Technol*, vol. 315, pp. 282–289, 2017, doi: 10.1016/j.powtec.2017.04.019.

[77] Q. Fang, Z. Li, Y. Wang, M. Song, and J. Wang, “A neural-network enhanced modeling method for real-time evaluation of the temperature distribution in a data center,” *Neural Comput Appl*, vol. 31, no. 12, pp. 8379–8391, 2019, doi: 10.1007/s00521-019-04508-y.

[78] A. Gheziel, S. Hanini, and B. Mohamedi, “Artificial neural network (ANN) for prediction indoor airborne particle concentration,” *International Journal of Ventilation*, vol. 21, no. 1, pp. 74–87, 2022, doi: 10.1080/14733315.2021.1876408.

[79] A. Warey, S. Kaushik, B. Khalighi, M. Cruse, and G. Venkatesan, “Data-driven prediction of vehicle cabin thermal comfort: using machine learning and high-fidelity simulation results,” *Int J Heat Mass Transf*, vol. 148, p. 119083, 2020, doi: 10.1016/j.ijheatmasstransfer.2019.119083.

[80] T. Zhang;, X. Li;, Q. Zhao;, and Y. Rao;, “Control of a novel synthetical index for the local indoor air quality by the artificial neural network and genetic algorithm,” *Sustain Cities Soc*, vol. 51, no. May, p. 101714, 2019, doi: 10.1016/j.scs.2019.101714.

[81] T. Zhang; and X. You;, “The use of genetic algorithm and self-updating artificial neural network for the inverse design of cabin environment,” *Indoor and Built Environment*, vol. 26, no. 3, pp. 347–354, 2017, doi: 10.1177/1420326X15609772.

[82] T. H. Zhang and X. Y. You, “The use of genetic algorithm and self-updating artificial neural network for the inverse design of cabin environment,” *Indoor and Built Environment*, vol. 26, no. 3, pp. 347–354, 2017, doi: 10.1177/1420326X15609772.

[83] Q. Zhou and R. Ooka, “Influence of data preprocessing on neural network performance for reproducing CFD simulations of non-isothermal indoor airflow distribution,” *Energy Build*, vol. 230, p. 110525, 2021, doi: 10.1016/j.enbuild.2020.110525.

[84] Q. Zhou and R. Ooka, “Neural network for indoor airflow prediction with CFD database,” *J Phys Conf Ser*, vol. 2069, no. 1, 2021, doi: 10.1088/1742-6596/2069/1/012154.

[85] V. J. L. Gan, B. Wang, C. M. Chan, A. U. Weerasuriya, and J. C. P. Cheng, "Physics-based, data-driven approach for predicting natural ventilation of residential high-rise buildings," *Build Simul*, vol. 15, no. 1, pp. 129–148, 2022, doi: 10.1007/s12273-021-0784-9.

[86] Q. Zhou and R. Ooka, "Implementation of a coupled simulation framework with neural network and Modelica for fast building energy simulation considering non-uniform indoor environment," *Build Environ*, vol. 211, no. October 2021, p. 108740, 2022, doi: 10.1016/j.buildenv.2021.108740.

[87] S. J. Cao, J. Ding, and C. Ren, "Sensor deployment strategy using cluster analysis of Fuzzy C-Means Algorithm: Towards online control of indoor environment's safety and health," *Sustain Cities Soc*, vol. 59, no. April, p. 102190, 2020, doi: 10.1016/j.scs.2020.102190.

[88] W. xue Cao and X. yi You, "The inverse optimization of exhaust hood by using intelligent algorithms and CFD simulation," *Powder Technol*, vol. 315, pp. 282–289, 2017, doi: 10.1016/j.powtec.2017.04.019.

[89] T.-H. Zhang and X.-Y. You, "A simulation-based inverse design of preset aircraft cabin environment," *Build Environ*, vol. 82, pp. 20–26, 2014, doi: 10.1016/j.buildenv.2014.08.002.

[90] T.-H. Zhang and X.-Y. You, "Inverse design of aircraft cabin environment by coupling artificial neural network and genetic algorithm Inverse design of aircraft cabin environment by coupling artificial neural network and genetic algorithm," *HVAC and R Research*, vol. 9669, 2014, doi: 10.1080/10789669.2014.950895.

[91] T.-H. Zhang and X.-Y. You, "Improvement of the training and normalization method of artificial neural network in the prediction of indoor environment," *Procedia Eng*, vol. 121, pp. 1245–1251, 2015, doi: 10.1016/j.proeng.2015.09.152.

[92] S. Zhang, Y. Cheng, C. Huan, and Z. Lin, "Modeling non-uniform thermal environment of stratum ventilation with supply and exit air conditions," *Build Environ*, vol. 144, no. August, pp. 542–554, 2018, doi: 10.1016/j.buildenv.2018.08.063.

[93] S. Cao and C. Ren, "Ventilation control strategy using low-dimensional linear ventilation models and artificial neural network," *Build Environ*, vol. 144, no. July, pp. 316–333, 2018, doi: 10.1016/j.buildenv.2018.08.032.

[94] J. Ren and S. J. Cao, "Incorporating online monitoring data into fast prediction models towards the development of artificial intelligent ventilation systems," *Sustain Cities Soc*, vol. 47, May 2019, doi: 10.1016/j.scs.2019.101498.

[95] Z. Feng, C. W. Yu, and S. J. Cao, "Fast prediction for indoor environment: Models assessment," *Indoor and Built Environment*, vol. 28, no. 6, pp. 727–730, 2019, doi: 10.1177/1420326X19852450.

[96] C. Ding and K. Poh, “Data-driven model for cross ventilation potential in high-density cities based on coupled CFD simulation and machine learning,” *Build Environ*, vol. 165, no. August, p. 106394, 2019, doi: 10.1016/j.buildenv.2019.106394.

[97] Y. Tominaga *et al.*, “AIJ guidelines for practical applications of CFD to pedestrian wind environment around buildings,” *Journal of Wind Engineering and Industrial Aerodynamics*, vol. 96, no. 10–11, pp. 1749–1761, 2008, doi: 10.1016/j.jweia.2008.02.058.

[98] J. Ling, A. Kurzawski, and J. Templeton, “Reynolds averaged turbulence modelling using deep neural networks with embedded invariance,” *J Fluid Mech*, vol. 807, pp. 155–166, 2016, doi: 10.1017/jfm.2016.615.

[99] S. Beetham and J. Capecelatro, “Formulating turbulence closures using sparse regression with embedded form invariance,” *Phys Rev Fluids*, vol. 5, no. 8, pp. 1–34, 2020, doi: 10.1103/PhysRevFluids.5.084611.

[100] S. Beetham, R. O. Fox, and J. Capecelatro, “Sparse identification of multiphase turbulence closures for coupled fluid-particle flows,” *J Fluid Mech*, vol. 914, pp. 1–23, 2021, doi: 10.1017/jfm.2021.53.

[101] M. Schmelzer, R. P. Dwight, and P. Cinnella, “Discovery of Algebraic Reynolds-Stress Models Using Sparse Symbolic Regression,” *Flow Turbul Combust*, vol. 104, no. 2–3, pp. 579–603, 2020, doi: 10.1007/s10494-019-00089-x.

[102] J. Kou and W. Zhang, “A hybrid reduced-order framework for complex aeroelastic simulations,” *Aerosp Sci Technol*, vol. 84, pp. 880–894, 2019, doi: 10.1016/j.ast.2018.11.014.

[103] M. Raissi;, A. Yazdani;, and G. E. Karniadakis;, “Hidden Fluid Mechanics,” *Science (1979)*, vol. 4741, no. January, pp. 1–8, 2020.

[104] X. Zhao, L. Du, X. Peng, Z. Deng, and W. Zhang, “Research on refined reconstruction method of airfoil pressure based on compressed sensing,” *Theoretical and Applied Mechanics Letters*, vol. 11, no. 2, p. 100223, 2021, doi: 10.1016/j.taml.2021.100223.

[105] Z. Li *et al.*, “Fourier Neural Operator for Parametric Partial Differential Equations,” *conference paper at ICLR*, pp. 1–16, 2021.

[106] R. Wang, R. Walters, and R. Yu, “Incorporating Symmetry into Deep Dynamics Models for Improved Generalization,” *conference paper at ICLR*, 2021.

[107] H. Frezat, G. Balarac, J. Le Sommer, R. Fablet, and R. Lguensat, “Physical invariance in neural networks for subgrid-scale scalar flux modeling,” *Phys Rev Fluids*, vol. 6, no. 2, pp. 1–32, 2021, doi: 10.1103/PhysRevFluids.6.024607.

[108] K. Lee and K. T. Carlberg, *Model reduction of dynamical systems on nonlinear manifolds using deep convolutional autoencoders*, vol. 404. 2020. doi: 10.1016/j.jcp.2019.108973.

[109] K. Duraisamy, G. Iaccarino, and H. Xiao, “Turbulence Modeling in the Age of Data,” *Fluid review*, pp. 1–23, 2019.

[110] O. Marin, R. Vinuesa, A. V. Obabko, and P. Schlatter, “Characterization of the secondary flow in hexagonal ducts,” *Physics of Fluids*, vol. 28, no. 12, 2016, doi: 10.1063/1.4968844.

[111] A. Vidal, H. M. Nagib, P. Schlatter, and R. Vinuesa, “Secondary flow in spanwise-periodic in-phase sinusoidal channels,” *J Fluid Mech*, vol. 851, pp. 288–316, 2018, doi: 10.1017/jfm.2018.498.

[112] J. X. Wang, J. L. Wu, and H. Xiao, “Physics-informed machine learning approach for reconstructing Reynolds stress modeling discrepancies based on DNS data,” *Phys Rev Fluids*, vol. 2, no. 3, pp. 1–36, 2017, doi: 10.1103/PhysRevFluids.2.034603.

[113] J. L. Wu, H. Xiao, and E. Paterson, “Physics-informed machine learning approach for augmenting turbulence models: A comprehensive framework,” *Phys Rev Fluids*, vol. 7, no. 3, pp. 1–42, 2018, doi: 10.1103/PhysRevFluids.3.074602.

[114] C. Jiang, R. Vinuesa, R. Chen, J. Mi, S. Laima, and H. Li, “An interpretable framework of data-driven turbulence modeling using deep neural networks,” *Physics of Fluids*, vol. 33, no. 5, 2021, doi: 10.1063/5.0048909.

[115] S. Rezaeiravesh, R. Vinuesa, and P. Schlatter, “On numerical uncertainties in scale-resolving simulations of canonical wall turbulence,” *Comput Fluids*, vol. 227, p. 105024, 2021, doi: 10.1016/j.compfluid.2021.105024.

[116] M. Emory, J. Larsson, and G. Iaccarino, “Modeling of structural uncertainties in Reynolds-averaged Navier-Stokes closures,” *Physics of Fluids*, vol. 25, no. 11, pp. 0–20, 2013, doi: 10.1063/1.4824659.

[117] A. A. Mishra and G. Iaccarino, “Uncertainty estimation for reynolds-averaged navier-stokes predictions of high-speed aircraft nozzle jets,” *AIAA Journal*, vol. 55, no. 11, pp. 3999–4004, 2017, doi: 10.2514/1.J056059.

[118] O. Obiols-Sales, A. Vishnu, N. Malaya, and A. Chandramowliswharan, “CFDNet: A deep learning-based accelerator for fluid simulations,” *Proceedings of the International Conference on Supercomputing*, 2020, doi: 10.1145/3392717.3392772.

[119] F. Gibou, D. Hyde, and R. Fedkiw, “Sharp interface approaches and deep learning techniques for multiphase flows,” *J Comput Phys*, vol. 380, no. May, pp. 442–463, 2019, doi: 10.1016/j.jcp.2018.05.031.

[120] M. Ma, J. Lu, and G. Tryggvason, “Using statistical learning to close two-fluid multiphase flow equations for a simple bubbly system,” *Physics of Fluids*, vol. 27, no. 9, p. 092101, 2015, doi: 10.1063/1.4930004.

[121] A. Vollant, G. Balarac, and C. Corre, “Subgrid-scale scalar flux modelling based on optimal estimation theory and machine-learning procedures,” *Journal of Turbulence*, vol. 18, no. 9, pp. 854–878, 2017, doi: 10.1080/14685248.2017.1334907.

[122] M. Gamahara and Y. Hattori, “Searching for turbulence models by artificial neural network,” *Phys Rev Fluids*, vol. 2, no. 5, pp. 1–20, 2017, doi: 10.1103/PhysRevFluids.2.054604.

[123] R. Maulik and O. San, “A neural network approach for the blind deconvolution of turbulent flows,” *J Fluid Mech*, vol. 831, pp. 151–181, 2017, doi: 10.1017/jfm.2017.637.

[124] M. Reissmann, J. Hasslberger, R. D. Sandberg, and M. Klein, “Application of Gene Expression Programming to a-posteriori LES modeling of a Taylor Green Vortex,” *J Comput Phys*, vol. 424, p. 109859, 2021, doi: 10.1016/j.jcp.2020.109859.

[125] R. E. Britter and S. R. Hanna, “Flow and dispersion in urban areas,” *Annu Rev Fluid Mech*, vol. 35, pp. 469–496, 2003, doi: 10.1146/annurev.fluid.35.101101.161147.

[126] Y. Mizuno and J. Jiménez, “Wall turbulence without walls,” *J Fluid Mech*, vol. 723, pp. 429–455, 2013, doi: 10.1017/jfm.2013.137.

[127] M. P. Encinar, R. García-Mayoral, and J. Jiménez, “Scaling of velocity fluctuations in off-wall boundary conditions for turbulent flows,” *J Phys Conf Ser*, vol. 506, no. 1, 2014, doi: 10.1088/1742-6596/506/1/012002.

[128] N. Moriya, K. Fukami, Y. Nabae, M. Morimoto, T. Nakamura, and K. Fukagata, “Inserting machine-learned virtual wall velocity for large-eddy simulation of turbulent channel flows,” *Physics of Fluids*, pp. 1–18, 2021.

[129] H. J. Bae and P. Koumoutsakos, “Scientific multi-agent reinforcement learning for wall-models of turbulent flows,” *Nat Commun*, vol. 13, no. 1, pp. 1–9, 2022, doi: 10.1038/s41467-022-28957-7.

[130] S. E. Ahmed, S. Pawar, O. San, A. Rasheed, T. Iliescu, and B. R. Noack, “On closures for reduced order models—A spectrum of first-principle to machine-learned avenues,” *Physics of Fluids*, vol. 33, no. 9, pp. 1–38, 2021, doi: 10.1063/5.0061577.

[131] J. N. Kutz, “Deep learning in fluid dynamics,” *J Fluid Mech*, vol. 814, pp. 1–4, 2017, doi: 10.1017/jfm.2016.803.

[132] R. Maulik, O. San, A. Rasheed, and P. Vedula, “Subgrid modelling for two-dimensional turbulence using neural networks,” *J Fluid Mech*, vol. 858, pp. 122–144, 2019, doi: 10.1017/jfm.2018.770.

[133] L. Zhu, W. Zhang, J. Kou, and Y. Liu, “Machine learning methods for turbulence modeling in subsonic flows around airfoils,” *Physics of Fluids*, vol. 31, no. 1, 2019, doi: 10.1063/1.5061693.

[134] L. Zhu, W. Zhang, X. Sun, Y. Liu, and X. Yuan, “Turbulence closure for high Reynolds number airfoil flows by deep neural networks,” *Aerosp Sci Technol*, vol. 110, p. 106452, 2021, doi: 10.1016/j.ast.2020.106452.

[135] Y. Bar-sinai, S. Hoyer, J. Hickey, and M. P. Brenner, “Learning data driven discretizations for partial differential equations,” 2019.

[136] S. Beetham, R. O. Fox, and J. Capecelatro, “Sparse identification of multiphase turbulence closures for coupled fluid-particle flows,” *J Fluid Mech*, vol. 914, pp. 1–23, 2021, doi: 10.1017/jfm.2021.53.

[137] B. Stevens and T. Colonius, “Enhancement of shock-capturing methods via machine learning,” 2002.

[138] D. Kochkov, J. A. Smith, A. Alieva, Q. Wang, M. P. Brenner, and S. Hoyer, “Machine learning–accelerated computational fluid dynamics,” *Proc Natl Acad Sci U S A*, vol. 118, no. 21, pp. 1–13, 2021, doi: 10.1073/pnas.2101784118.

[139] J. Kou and W. Zhang, “Data-driven modeling for unsteady aerodynamics and aeroelasticity,” *Progress in Aerospace Sciences*, vol. 125, no. June, p. 100725, 2021, doi: 10.1016/j.paerosci.2021.100725.

[140] S. Bhatnagar, Y. Afshar, S. Pan, K. Duraisamy, and S. Kaushik, “Prediction of aerodynamic flow fields using convolutional neural networks,” *Comput Mech*, vol. 64, no. 2, pp. 525–545, 2019, doi: 10.1007/s00466-019-01740-0.

[141] M. Bode;, M. Gauding;, J. H. Gobbert;, B. Liao;, J. Jitsev;, and H. Pitsch;, “Towards Prediction of Turbulent Flows at High Reynolds Numbers Using High Performance Computing Data and Deep Learning,” *International Conference on High Performance Computing*, vol. 11203 LNCS, pp. 614–623, 2018, doi: 10.1007/978-3-030-02465-9_44.

[142] C. W. Chang and N. T. Dinh, “Classification of machine learning frameworks for data-driven thermal fluid models,” *International Journal of Thermal Sciences*, vol. 135, no. August 2018, pp. 559–579, 2019, doi: 10.1016/j.ijthermalsci.2018.09.002.

[143] B. Colvert;, M. Alsalman;, and E. Kanso;, “Classifying vortex wakes using neural networks,” *Bioinspir Biomim*, 2018.

[144] K. Fukami, Y. Nabae, K. Kawai, and K. Fukagata, “Synthetic turbulent inflow generator using machine learning,” *Phys Rev Fluids*, vol. 4, no. 6, pp. 1–18, 2019, doi: 10.1103/PhysRevFluids.4.064603.

[145] K. O. Lye, S. Mishra, and D. Ray, “Deep learning observables in computational fluid dynamics,” *J Comput Phys*, vol. 410, p. 109339, 2020, doi: 10.1016/j.jcp.2020.109339.

[146] J. Tian, C. Qi, Y. Sun, Z. M. Yaseen, and B. T. Pham, “Permeability prediction of porous media using a combination of computational fluid dynamics and hybrid machine learning methods,” *Eng Comput*, vol. 37, no. 4, pp. 3455–3471, 2021, doi: 10.1007/s00366-020-01012-z.

[147] M. D. Ribeiro, A. Rehman, S. Ahmed, and A. Dengel, “DeepCFD : Efficient Steady-State Laminar Flow Approximation with Deep Convolutional Neural,” *Computational Physics*, 2021.

[148] F. Guéniat, L. Mathelin, and M. Y. Hussaini, “A statistical learning strategy for closed-loop control of fluid flows,” *Theor Comput Fluid Dyn*, vol. 30, no. 6, pp. 497–510, 2016, doi: 10.1007/s00162-016-0392-y.

[149] W. N. Edeling, M. Schmelzer, R. P. Dwight, and P. Cinnella, “Bayesian predictions of reynolds-averaged navier–stokes uncertainties using maximum a posteriori estimates,” *AIAA Journal*, vol. 56, no. 5, pp. 2018–2029, 2018, doi: 10.2514/1.J056287.

[150] X. Guo, W. Li, and F. Iorio, “Convolutional neural networks for steady flow approximation,” *Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, vol. 13-17-August-2016, pp. 481–490, 2016, doi: 10.1145/2939672.2939738.

[151] B. N. Hanna, N. T. Dinh, R. W. Youngblood, and I. A. Bolotnov, “Coarse-Grid Computational Fluid Dynamic (CG-CFD) Error Prediction using Machine Learning,” 2017.

[152] J. R. Holland, J. D. Baeder, and K. Duraisamy, “Towards integrated field inversion and machine learning with embedded neural networks for rans modeling,” *AIAA Scitech 2019 Forum*, no. January, pp. 1–14, 2019, doi: 10.2514/6.2019-1884.

[153] H. Xiao, J. L. Wu, J. X. Wang, R. Sun, and C. J. Roy, “Quantifying and reducing model-form uncertainties in Reynolds-averaged Navier–Stokes simulations: A data-driven, physics-informed Bayesian approach,” *J Comput Phys*, vol. 324, pp. 115–136, 2016, doi: 10.1016/j.jcp.2016.07.038.

[154] S. Bhalla, M. Yao, J. P. Hickey, and M. Crowley, “Compact Representation of a Multi-dimensional Combustion Manifold Using Deep Neural Networks,” *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, vol. 11908 LNAI, pp. 602–617, 2020, doi: 10.1007/978-3-030-46133-1_36.

[155] L. Yang and A. Shami, “On hyperparameter optimization of machine learning algorithms: Theory and practice,” *Neurocomputing*, vol. 415, pp. 295–316, Nov. 2020, doi: 10.1016/j.neucom.2020.07.061.

[156] A. S. Krishnapriyan, A. Gholami, S. Zhe, R. M. Kirby, and M. W. Mahoney, “Characterizing possible failure modes in physics-informed neural networks,” in *35th Conference on Neural Information Processing Systems (NeurIPS 2021)*, 2021.

[157] W. Hu *et al.*, “Open Graph Benchmark: Datasets for Machine Learning on Graphs,” in *34th Conference on Neural Information Processing Systems (NeurIPS 2020)*, 2020. [Online]. Available: https://ogb.stanford.edu.

[158] O. Russakovsky *et al.*, “ImageNet Large Scale Visual Recognition Challenge,” *Int J Comput Vis*, vol. 115, no. 3, pp. 211–252, Dec. 2015, doi: 10.1007/s11263-015-0816-y.

[159] W. T. Chung, K. S. Jung, J. H. Chen, and M. Ihme, “BLASTNet: A call for community-involved big data in combustion machine learning,” *Applications in Energy and Combustion Science*, vol. 12, Dec. 2022, doi: 10.1016/j.jaecs.2022.100087.

[160] S. Yarlanki, B. Rajendran, and H. Hamann, “Estimation of turbulence closure coefficients for data centers using machine learning algorithms,” in *13th InterSociety Conference on Thermal and Thermomechanical Phenomena in Electronic Systems*, 2012, pp. 38–42. doi: 10.1109/ITHERM.2012.6231411.

[161] J. Ling, A. Kurzawski, and J. Templeton, “Reynolds averaged turbulence modelling using deep neural networks with embedded invariance,” *J Fluid Mech*, vol. 807, pp. 155–166, Nov. 2016, doi: 10.1017/jfm.2016.615.

[162] H. Gao, L. Sun, and J.-X. Wang, “Super-resolution and denoising of fluid flow using physics-informed convolutional neural networks without high-resolution labels,” *Physics of Fluids*, Nov. 2021, doi: 10.1063/5.0054312.

[163] J. Weatheritt and R. Sandberg, “A novel evolutionary algorithm applied to algebraic modifications of the RANS stress–strain relationship,” *J Comput Phys*, vol. 325, pp. 22–37, Nov. 2016, doi: 10.1016/j.jcp.2016.08.015.

[164] S. B. Pope, “A more general effective-viscosity hypothesis,” *J Fluid Mech*, vol. 72, no. 2, pp. 331–340, Nov. 1975, doi: 10.1017/S0022112075003382.

[165] Y. F. Marioni, E. A. De Toledo Ortiz, A. Cassinelli, F. Montomoli, P. Adami, and R. Vazquez, “A Machine Learning Approach to Improve Turbulence Modelling from DNS Data Using Neural Networks,” in *International Journal of Turbomachinery Propulsion and Power*, European Conference on Turbomachinery (ETC), 2021. doi: 10.3390/ijtpp6020017.

[166] J. Weatheritt, R. D. Sandberg, J. Ling, G. Saez, and J. Bodart, “A COMPARATIVE STUDY OF CONTRASTING MACHINE LEARNING FRAMEWORKS APPLIED TO RANS MODELING OF JETS IN CROSSFLOW,” 2017. [Online]. Available: http://www.asme.org/about-asme/terms-of-use

[167] Y. Zhao, H. D. Akolekar, J. Weatheritt, V. Michelassi, and R. D. Sandberg, “RANS turbulence model development using CFD-driven machine learning,” *J Comput Phys*, vol. 411, Jun. 2020, doi: 10.1016/j.jcp.2020.109413.