

Machine learning for combustion

Lei Zhou^a, Yuntong Song^a, Weiqi Ji^b, Haiqiao Wei^{a,*}

^a State Key Laboratory of Engines, Tianjin University, Tianjin 300072, China

^b Department of Mechanical Engineering, Massachusetts Institute of Technology, Cambridge, MA 02139, USA

HIGHLIGHTS

- We introduce the fundamentals of ML and its usage in combustion field.
- The opportunities and limitations of using ML in combustion studies are also discussed.
- Applications of ML in combustion modeling and combustion measurement are presented.
- Applications of ML in engine and fuel research are presented.

ARTICLE INFO

Keywords:

Machine learning
Data-driven
Combustion modeling
Combustion diagnostic
Fuel

ABSTRACT

Combustion science is an interdisciplinary study that involves nonlinear physical and chemical phenomena in time and length scales, including complex chemical reactions and fluid flows. Combustion widely supplies energy for powering vehicles, heating houses, generating electricity, cooking food, etc. The key to study combustion is to improve the combustion efficiency with minimum emission of pollutants. Machine learning facilitates data-driven techniques for handling large amounts of combustion data, either obtained through experiments or simulations under multiple spatiotemporal scales, thereby finding the hidden patterns underlying these data and promoting combustion research. This work presents an overview of studies on the applications of machine learning in combustion science fields over the past several decades. We introduce the fundamentals of machine learning and its usage in aiding chemical reactions, combustion modeling, combustion measurement, engine performance prediction and optimization, and fuel design. The opportunities and limitations of using machine learning in combustion studies are also discussed. This paper aims to provide readers with a portrait of what and how machine learning can be used in combustion research and to inspire researchers in their ongoing studies. Machine learning techniques are rapidly advancing in this era of big data, and there is high potential for exploring the combination between machine learning and combustion research and achieving remarkable results.

1. Introduction

Combustion science is an interdisciplinary study involving fluid and chemical kinetics, which involves chemical reactions that include complex nonlinear processes on time and space scales. Its systems are widely used in our daily life; they can power cars, trucks, marines, rockets, power plants, etc. However, owing to combustion, greenhouse gasses and many other harmful emissions are produced. To attain a carbon-neutral society in the future [1], it is of utmost importance to improve the combustion efficiency of combustion systems while maintaining low harmful emissions. Current studies on combustion involve

many aspects, such as building combustion models, development of simulation approaches for combustion, and combustion optimization in systems. Many experiments and simulations generate massive amounts of data. The utilization of these data for better research on combustion has become a new challenge and research opportunity. Fortunately, machine learning (ML) provides advanced data-driven techniques to extract information from massive data and helps to reveal the underlying combustion mechanisms.

Artificial intelligence (AI) [2] is intelligence demonstrated by machines, as opposed to natural intelligence displayed by animals including humans. As a major branch of AI, ML has gained considerable attention in recent years because of its outstanding performance in many

* Corresponding authors.

E-mail address: whq@tju.edu.cn (H. Wei).

| Nomenclature | |
|---------------------|--|
| AI | artificial intelligence |
| ART | algebraic reconstruction technique |
| CA | crank angle |
| CFD | computational fluid dynamics |
| CNN | convolutional neural network |
| DNS | direct numerical simulation |
| FGM | flamelet-generated manifold |
| FRC | finite-rate chemistry |
| GAN | generative adversarial network |
| HCCI | homogenous charge compression ignition |
| IM | inert mixing |
| LES | large eddy simulation |
| LSTM | long short-term memory |
| MD | molecular dynamics |
| MLP | multilayer perceptron |
| PC | principal component |
| PDF | probability density function |
| PLIF | planar laser-induced fluorescence |
| PSO | particle swarm optimization |
| RB | reduced basis |
| RON | research octane number |
| SR | super-resolution |
| SVM | support vector regression |
| ANN | artificial neural network |
| BPNN | backpropagation neural network |
| CN | cetane number |
| DNN | deep neural network |
| FDF | filtered probability density function |
| FPV | flamelet progress variable |
| GA | genetic algorithm |
| GPU | Graphics processing unit |
| HMM | hidden Markov model |
| IMEP | indicated mean effective pressure |
| LS-SVM | least-squares support vector machine |
| MCMC | Markov chain Monte Carlo |
| ML | machine learning |
| NN | neural network |
| PCA | principal component analysis |
| PIV | particle imaging velocimetry |
| POD | proper orthogonal decomposition |
| RANS | Reynolds-averaged Navier–Stokes |
| RNN | recurrent neural network |
| SOM | self-organizing map |
| SVM | support vector machine |
| VT | volumetric tomography |

areas [3], including health care, manufacturing, education, and financial modeling, in all fields of science, technology and commerce. ML is a discipline that focuses on the automatic optimizing a model for making predictions or decisions by feeding them with data. Recently, the history, current developments, and emerging opportunities of ML for fluid mechanics were reviewed by Brunton et al. [4], and data-driven models built using ML for turbulence modeling were reviewed by Duraisamy et al. [5]. ML can handle tasks in fluid mechanics, such as dimensionality reduction, reduced-order modeling, feature extraction, and flow control. The study of combustion is interdisciplinary and involves turbulent fluid mechanics and chemical kinetics. Owing to the effects of chemical reactions, considering combustion process in turbulent fluid flows becomes more complex. The application of ML in fluid mechanics can also provide a demonstration for the combined research of combustion and ML.

Recently, many excellent works related to combustion and ML have been performed, providing a new perspective for combustion studies. As early as 2003, Kalogirou et al. [2] provided a comprehensive review of AI for the modeling and control of the combustion process, mainly focusing on the use of classical methods such as expert systems, genetic algorithms, fuzzy logic, and neural networks. In recent years, with the rapid development of ML, many novel and robust data-driven models have been proposed, including deep learning-related models, and there is a need to review the latest combustion research using the latest ML models. Shrivastava et al. [6] summarized the applications of soft computing in internal combustion engines. Soft computing refers to finding a solution to an inexact problem using artificial intelligent methods, such as fuzzy-based approaches. Raman and Hassanaly [7] introduced black-box and gray-box models for the numerical simulations of combustion systems. Zheng et al. [8] briefly reviewed the applications of ML in combustion research, while some other related topics were neglected, such as combustion modeling, combustion measurement, and fuel research, which require further research. Consequently, this paper provides a comprehensive review of the applications of ML in combustion studies. First, we provide a brief introduction to common ML concepts and approaches. The applications of ML in combustion modeling, including chemical reduction, combustion kinetic model uncertainty quantification, and subgrid modeling of combustion are then

systematically presented. Subsequently, the applications of ML in combustion measurement and diagnostics, including tackling the inverse problems in combustion measurement, diagnostics, and combustion stability monitoring are presented. In addition, the applications of ML in engine performance research, combustion control and optimization, and fuel research are presented. Finally, the challenges and outlooks of using ML for combustion studies are discussed, and the conclusions of this review are presented. It should be noted that the hardware aspect of machine learning such as high performance computing, Graphics processing unit (GPU) computing are not within the scope of this paper.

The remainder of the paper is structured as follows: A brief introduction to ML algorithms is presented in Section 2. The applications of ML in combustion modeling are systematically provided in Section 3. Then the current studies on combustion measurement and diagnostics using ML are reviewed in Section 4, and the applications of ML in engine and fuel research are showed in Section 5. Finally, the concluding remarks are given in Section 6.

2. A brief introduction to ML

ML algorithms build a model based on the given data, which originates from statistics. The development of statistics has accelerated the advancement of ML. ML aims to find patterns and mechanisms under the huge amount of data, build models based on the data for predictions, and help solve problems. Generally, ML is a very powerful tool for handling complex and nonlinear physical and chemical processes, such as combustion. Furthermore, in this era of big data, it should be noted that high-performance computing helps to deal with vast amounts of data and speed up the process of physical phenomena simulations, data mining, and AI. GPU hardware was designed for parallel processing and was originally used in graphics and video rendering. Because of the excellent parallel computing capacity of the GPU, high-performance computing is now dominated by GPU. The training of deep machine learning models with huge amounts of data can be deployed on GPUs through deep learning frameworks such as TensorFlow [9] and PyTorch [10] without much effort. Owing to the rapid development of ML, its convenient deployment, and the improvement of hardware performance, ML has spread into the area of combustion research, and many

combustion problems daunting for a long time could be tailored by adopting ML methods.

ML provides various models for handling data, which can be mainly classified into three types: supervised learning, unsupervised learning and reinforcement learning, and many specific models exist for each type [4], as shown in Fig. 1. ML can be roughly divided into three basic elements: models, learning criteria, and optimization.

In the ML procedure, the input space X containing input values x and output space Y containing output values y should be defined for the model first. In supervised learning, the output space often comprises discrete values for classification problems or continuous values for regression problems. In unsupervised learning, the output may be clusters, low-dimensional vectors, etc. The model hypothesis space F, where ML determines the best target function, contains all possible conditional probability distributions $p(y|x)$ or decision functions $y = g(x)$, which can be linear or nonlinear. Linear models can also be transformed into nonlinear models using nonlinear base kernels.

The learning process of ML proceeds based on the learning criteria or policies. The quality of the model can be measured using the expected risk $R(\theta)$, defined as Eqn 1:

$$R(\theta) = \mathbb{E}_{(x,y) \sim p_r(x,y)} [L(y, \hat{y})] \quad (1)$$

Here, $p_r(x,y)$ is the distribution of the real data, $L(y, \hat{y})$ is the loss function, quantifying the distinction between the real output value y and the model-predicted value \hat{y} , θ represents the model parameters, \mathbb{E} represents the calculation expectation. The loss function can take many forms, such as the quadratic loss function, cross-entropy loss function, and hinge loss function. The learning criterion is to minimize the expected risk to obtain a model with good prediction accuracy. Some researchers [11] tried to add physics loss to the loss function, to minimize the residuals of the governing equations for the generated data.

ML models are trained through an optimization process where the model parameters and hyperparameters are optimized. Hyperparameters are used to control the learning process, such as the number of layers of the neural network and the learning rate. Many optimization algorithms, such as gradient descent, stochastic gradient descent, adaptive moment estimation, Newton's method, and dynamic programming exist. A well-trained ML model should have good generalization ability, i.e. the ability to predict unknown data. If a trained model only fits the trained data, but performs poorly on the untrained test data, it is likely to be over-fitting. There are several approaches to improve the generalization of ML model, such as splitting datasets, cross validation, early stopping. Before training the ML model, the dataset would be split into two groups, i.e. training dataset and test dataset. The training dataset is used to train the model and the test dataset is used for testing the generalization ability of the model. Cross validation is a

resampling method that uses different portions of the data to test and train a model on different iterations. Early stopping is used to get the lowest error on the test dataset before the model is over-fitting. For deep models which has more layers, training the model may facing the problems of gradient vanishing or gradient exploding problem, there are some tricks to avoid or alleviate gradient vanishing or exploding in deep models, such as adopting ReLU activation function, gradient clipping, batch normalization, weights regularization [12], most of them can be found in [13].

2.1. Supervised learning

Supervised learning refers to the ML problem of learning predictive models from labeled data. Labeled data comprise a set of training examples, each of which comprises an input and a corresponding output or label. The labeled data are often grouped into training and test datasets. The training data were used for training the model, and the test data were used to identify whether the model was sufficiently trained. In essence, the supervised learning method establishes the relationship between the input and output, which is similar to the concept of function approximation. Once the model is well trained, predictions for new input data can be achieved based on the relationship learned from the previous datasets. Supervised learning is often used for regression and classification. Commonly used supervised learning models include neural networks, support vector machines (SVMs), decision trees, Gaussian process, k-nearest neighbor, ensemble learning, and hidden Markov model (HMM). A systematic study of statistical learning [14] involving the characteristics of various ML algorithms, related mathematical derivations, and their applications was introduced. In general, these models apply to different scenarios owing to their unique learning strategies. Detailed information can be found in [14].

2.1.1. Supervised neural networks

Inspired by the structure of the brain using the neuron as a central building element, neural networks have been the most popular supervised learning method in recent years [4] because of the capability of neural networks to approximate complex functions in a flexible form [5]. They are networks composed of layers of connected neurons and nonlinear activation functions. The architecture of neural networks comprises an input layer that receives data, several hidden layers for data processing, and an output layer for yield predictions. Neural networks can be supervised or unsupervised, depending on the nature of the objective function. In most cases, it appeared in the form of supervised learning, so it was discussed in this category. In general, three important neural network methods exist: artificial neural network (ANN), convolutional neural network (CNN), and recurrent neural network (RNN).

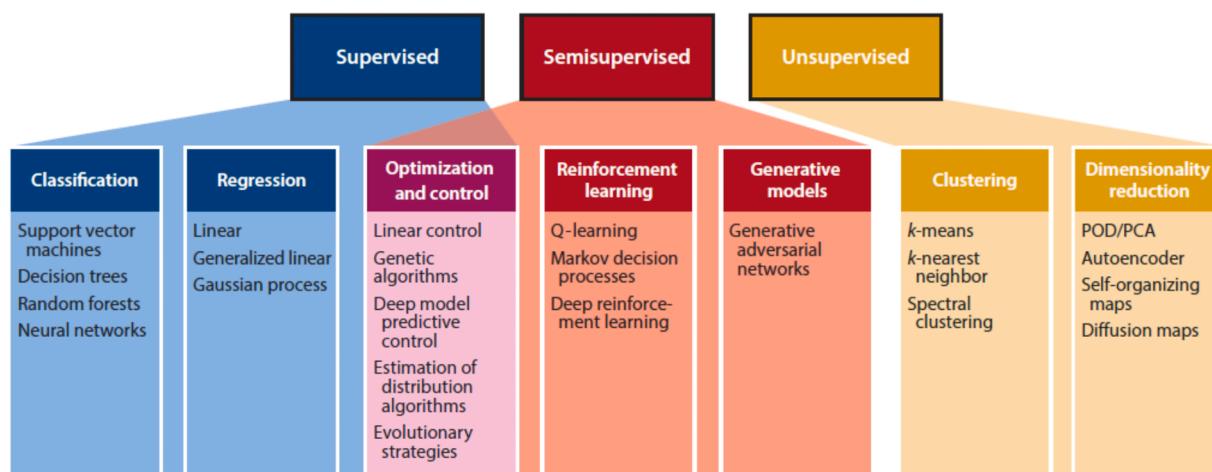


Fig. 1. Classification of ML algorithms (Abbreviations: PCA, principal component analysis; POD, proper orthogonal decomposition) [4].

Backpropagation is a nonlinear optimization method that is efficient in calculating the gradient in neural networks by implementing the chain rule of derivatives [15]. Multilayer perceptron (MLP) is a class of feed-forward neural networks utilizing backpropagation for training, sometimes also called backpropagation neural networks (BPNNs) or ANNs. Originating from mathematical neurobiology, the ANN approach and its hybrids could be promising candidates for coping with the nonlinearities and complexities of complicated biological processes [16], such as turbulent combustion. In general, an ANN is an interconnected network of artificial neurons, comprising an input layer, one or more hidden layers, and an output layer. The hidden layer contains weights and biases that amplify or attenuate the strengths of different signals coming from the preceding layer [17]. ANNs can simulate complex dynamic systems with multivariate properties with no proper understanding of the detailed underlying physical mechanisms [16]. Presently, the ANN method has become an attractive alternative strategy for the application of combustion science, especially the tabulation techniques for chemical reactions [18], owing to the outstanding smooth interpolation ability of ANNs. The idea of using an ANN as an approximator instead of directly solving nonlinear equations in a combustion simulation is shown in Fig. 2, where the formula for the calculation of the chemical source term into the ANN framework was derived. Given a dataset generated by an abstract problem, ML models such as ANN can be trained to be integrated into simulations of similar conditions. The input data for the ANN model are the mass fractions of the N species, sometimes accompanied by temperature T and time step dt , and the outputs are the corresponding chemical sources or the corresponding terms at the next time step [18].

Deep learning is ML using deep neural networks (DNNs) that involve many more layers, neurons, and more complex structures. Deep learning has the potential to extract more features from raw input data, especially images. CNN and RNN are popular deep learning models which require more data to train, only with enough data can the deep model fully learn each feature parameters. When the activation functions are expressed in terms of convolutional kernels, CNN is a powerful method that has achieved significant success in image and pattern recognition [13]. A CNN architecture is formed by a stack of distinct layers, which take advantage of the local spatial coherence of the input data, thus decreasing the number of parameters and escaping the curse of dimensionality with which the error increases with the increase in the number of parameters. CNN was originally utilized for studying visual representations [19], which has been extended to the combustion field, such as unresolved flame surface wrinkling [20], modeling of the unresolved variance [21], and chemical rate constant from shock-tube measurements [22]. In addition, RNN is naturally suitable for processing time-series or sequential data and can use its internal memory to process

sequences of input, and the output of an element in the sequence depends on the previous elements and current input. As a variant of RNN, the long short-term memory (LSTM) network [23] deploys a gating mechanism to store and forget information about the previous input, alleviating the gradient vanishing problem suffered by the RNN. Because of their ability of handling time-series data and making predictions about the future state of the system, RNN and LSTM can be used for predicting the evolution of the flame in time series [24].

2.1.2. Other supervised learning models

Among the available ML techniques, neural network technology is the most widely used method in combustion. In addition to neural networks, several other supervised learning algorithms exist. The SVM [25] is a supervised learning algorithm for two-group classification, based on the idea of maximizing the width of the gap between two categories in the data space. Nonlinear classification using SVM can be implemented using a kernel trick, which implicitly maps inputs into high-dimensional space, making the data linearly separable. After classification is completed, a hyperplane is provided for new data classification. Support vectors are data points closest to the hyperplane, and they influence the position of the hyperplane. Based on the same principle as SVM, support vector regression (SVR) [26] deals with regression problems. The objective of SVR was to gain a hyperplane to make the data points fall within the decision boundaries around the hyperplane as much as possible. In combustion fields, SVM can be used for distinguish combustion state by giving several observable variables, SVR can be used to regress fuel properties through experimental data. Decision trees [27] whose structure resembles a tree, are also a type of supervised learning technique. Decision trees are composed of decision nodes and leaves, which are continuously split at decision nodes based on a certain parameter, and the leaves are the final outcomes. Decision trees can be used for predicting the state of the engine, and behave better for categorical data than SVM. HMM [28] is a supervised ML model that can be used to describe the evolution of observable events that depend on internal factors that are not directly observable. The Gaussian process [29] is a powerful tool for classification and regression, which is a probability distribution over possible functions and can make predictions based on prior knowledge. Ensemble learning [30] is the concept of combining several learning algorithms for prediction. Ensemble learning uses multiple learning algorithms to obtain a better prediction performance than any single learning algorithm. Common ensemble learning algorithms include random forest, AdaBoost, bagging, boosting, and a mixture of experts.

2.2. Unsupervised learning

Unsupervised learning refers to an ML technique where the model does not need to be supervised. It allows the model to discover patterns and information automatically with objective functions which are not defined with target data. The unsupervised learning model was trained using unlabeled data. An unsupervised learning model can be used for clustering (grouping data points based on their similarity), anomaly detection, learning latent variables, dimension reduction (identifying a subset of dependent variables that describe the data), and generative modeling.

Clustering is the process of grouping a set of objects into several clusters based on the similarities among those objects without knowing the data label. It is useful for pattern recognition and data compression. In combustion research, clustering algorithms are used to cluster combustion data for better training of supervised learning models. Common clustering methods include K-means clustering [31] and self-organizing map (SOM) [32]. In some researches [33, 34], training data were firstly clustered before training neural networks, this can enable the trained ML model of each cluster behaves better.

Dimension reduction is the transformation of data from a high-dimensional space into a low-dimensional space while retaining the

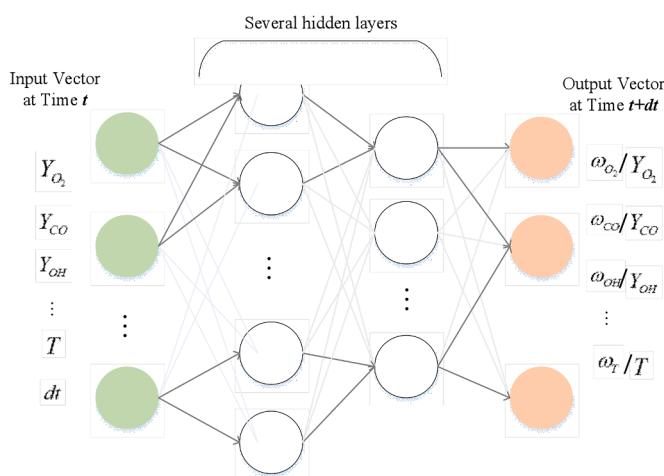


Fig. 2. Architecture of ANN training as an approximator.

useful properties of the original data, making the processing of the data less computational. It can be used for noise reduction and data visualization. Principal component analysis (PCA) [35] is a popular approach for feature extraction and dimension reduction. Proper orthogonal decomposition (POD) [36] is closely related to the PCA algorithm, it is used to decompose a physical field through eigen decomposition of data covariance matrix, it is assimilated with the PCA in the field of statistics and singular value decomposition [37] in the field of linear algebra, because it refers to eigenvalues and eigenvectors of a physical field. Although those concepts have been used in different cases, it should be noted that they are the same thing. POD was originally used to decompose the transient turbulent flow field into different modes (i.e., turbulent coherent structures) along with time-varying coefficients [38] to determine the nature of the flow. In general, the snapshot method [39] was utilized to implement POD. Flow field reconstruction using the coupling of POD and ANN can relieve the curse of the dimensionality in aerodynamics [40]. PCA can extract the most influential patterns, that is, principal components, from the original data. Thus, the original data can be represented by those principal components without resorting to a large amount of data. An autoencoder is a type of ANN for dimension reduction in an unsupervised manner. It comprises two key processes: the encoder and decoder processes. In the encoder process, the input dimensions are reduced and the input data are compressed into an encoded representation, where the model learns to reconstruct the data from the encoded representation to be as close to the original input as possible. The encoded representation is the latent attribute of the input data. Autoencoders can be used for anomaly detection and image denoising.

Generative modeling is an unsupervised learning process where the discovery of the regularities or patterns in the input data was done automatically, and after the training process, the model can be used to generate new examples with the same statistics as the training set. A generative adversarial network (GAN) is a popular generative model that has proven to be a better solution for semi-supervised learning [41] or reinforcement networks [42]. Here, semi-supervised learning is an approach that combines a small amount of labelled data with a large amount of unlabeled data, it falls between supervised and unsupervised learning, details of semi-supervised learning can be found in [41]. Detailed studies of GAN can be found in [42, 43]. A GAN is composed of two deep networks: a generator and a discriminator. During the training process, the generator learns to create images to fool the discriminator. Through alternating steps of training the generator and the discriminator, the GAN model can generate natural images, just like training data. In combustion research, GAN can generate high-resolution combustion images from low-resolution images.

2.3. Reinforcement learning

Reinforcement learning [44] is a basic ML paradigms that deal with exploration and exploitation, which can be used for decision-making in many areas, such as autonomous driving [45], robotics [46], games [47], and industrial process control [48]. The four essential elements in reinforcement learning are agent, environment, action, and rewards. Reinforcement learning deals with learning through the trial-and-error exploration. During the learning process, the agent learns to perform actions on the environment to maximize the cumulative rewards. Exploration is more of a long-term benefit concept where it allows the agent to improve its knowledge about each action which could lead to long term benefit, while exploitation basically exploits the agent's current estimated value and chooses the greedy approach to get the most reward. Hence, the learned policy π can be used to make optimal actions a while given the state s of the environment, denoted as $a = \pi(s)$. The data labels for reinforcement learning are the rewards generated through the interaction between the agent and the environment; they have the characteristics of time delay. Algorithms used for reinforcement learning can be generally classified into model-based and

model-free classes, which include policy optimization methods and Q-learning methods such as deep Q-network. For model-free reinforcement learning, the agent learns by interacting directly with the true environment, whereas for model-based reinforcement learning, the agent learns by interacting with a model representing the real world. The Q-value is the value function that assumes state and action as parameters, maps state-action pairs to rewards, and is used for evaluating the long-term return of taking an action under a policy π . Reinforcement learning is suitable for optimization problems of a system through trial-and-error exploration to achieve the best performance. In combustion devices, the control of the combustion system to achieve the best performance is always challenging because the control process involves multiple control parameters and complex environments, and reinforcement learning may be suitable for handling this task.

3. ML in combustion modeling

Computational modeling and simulation of modern combustion devices are essential parts of the design process. Many combustion models [49, 50] including detailed or reduced chemical mechanisms, have been applied to canonical flows. Despite the considerable understanding of the combustion mechanism and simulation methods, the combustion modeling of different fuels in complex geometries remains a challenge because of the trade-off between computational efficiency and accuracy. With no additional combustion model, the direct numerical simulation (DNS) of combustion coupled with a detailed chemical mechanism and considering the practical working conditions is almost impossible because of the huge demand for computational resources, let alone the lack of chemical mechanisms of various real fuels, and the limited applicability of those models. With the development of AI, especially the thriving of neural networks in recent years, the exploitation of high-fidelity models and the reduction or calibration of traditional models using ML has been proven, and many related studies have been conducted. Fig. 3 shows the general method for solving the chemical kinetics calculation based on ML in computational fluid dynamics (CFD) methods.

In the past decades, ML has been used to understand, predict, optimize, and control the combustion process. The applications of ML in combustion modeling can be categorized as follows:

- 3.1 Chemical kinetics calculation acceleration.
- 3.2 Combustion kinetic model uncertainty quantification.
- 3.3 Discovery of unknown reaction pathways.
- 3.4 Subgrid combustion modeling.
- 3.5 Building surrogate solvers for simulation.

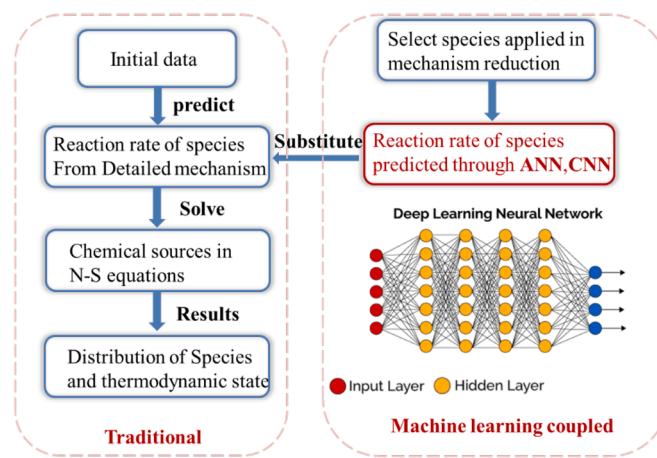


Fig. 3. Schematic of traditional CFD and CFD coupled with ML.

3.1. Chemical kinetics calculation acceleration

The chemistry of combustion is vitally important in the design of efficient and low-emission combustion engines [51]. To understand combustion chemistry and phenomena, many credible and detailed mechanisms for fuel combustion have been developed, including hundreds of species and thousands of reactions [52]. One major difficulty in combustion modeling using the detailed mechanisms in CFD methods is the high computational demand arising from the need to couple hundreds of nonlinear stiff equations of chemical kinetics, as shown in Fig. 4. The nonlinear equation for solving each species Y_i during combustion is the transport equation, as shown in Eq. (2).

$$\frac{\partial Y_i}{\partial t} + \nabla \cdot (\rho u Y_i) = \nabla \cdot (\rho D \nabla Y_i) + \omega_{Y_i} \quad (2)$$

Here, t , ρ , u , Y_N , D , and ω_{Y_N} are time, the density of the gas mixture, the velocity vector, the mass fraction of n th chemical species, molecular mass diffusivity coefficient, and the source term of species Y_i . The source term ω_{Y_i} is solved through the net production rate, whose forward and reverse rate constants are solved through the Arrhenius expression, as shown in Eq. (3).

$$k = A \exp\left(-\frac{E_A}{R_u T}\right) \quad (3)$$

where k is the rate constant, T is the absolute temperature, A is the pre-exponential factor, E_A is the activation energy, and R_u is the universal gas constant.

Solving the source term of each species at every cell, every time step incurs a very large computational cost. In the chemical mechanisms of some biofuels or much more complex fuels, the number of reactants and equations increases significantly, which makes the coupling of detailed mechanisms in combustion simulation impossible. The reduced reaction mechanism, which contains the most influential species and reactions, can be derived by minimizing the number of species and elementary reactions through many methods, including computational singular perturbation [53], rate-controlled constrained equilibrium [54], and directed relation graph [55]. Owing to the reduced number of equations that need to be solved using reduced mechanisms, the computational cost can be reduced. In addition, other on-the-fly acceleration methods, such as dynamic adaptive chemistry, dimension reduction [56], cell agglomeration methods [57, 58], and tabulation of dynamic adaptive chemistry [59] should suitably resolve this issue without compromising the prediction accuracy, as shown in Fig. 4. Nevertheless, the direct solving of these nonlinear equations for chemical species still incurs significant computational costs.

To decrease the computational cost, a tabulation method that maps high-dimensional combustion state space onto low-dimensional

manifolds was developed. In the tabulation method, the evolution of the chemical system is described by a few control variables, and the chemical state can be obtained by retrieving the lookup table. The tabulation method includes the intrinsic low-dimensional manifold method [60], flamelet-generated manifold (FGM) method [61], and flamelet progress variable (FPV) method [62]. However, the stored lookup is comprised of numbers of discrete points, which making the data retrieval from those stored sampling points is not accurate, and with the increasing of the number of data parameters, the memory occupied by the lookup table will increase exponentially, which will take a large amount of memory.

Many studies have tried to use ML methods to solve chemical reactions. Compared to traditional methods, the use of ML methods for chemical reaction calculations shows two major advantages: faster calculation speed and smaller memory occupation. Because a ML model only store its structure information, weights and activations, it usually does not take up much storage. In addition, compared to traditional tabulation methods, ML methods have higher accuracy, take ANN for example, ANN features a smooth interpolation between the sampling points computed from direct integration, which can reduce the error caused by improper interpolation, yet will sacrifice more computational time. It should be noted that the computational cost also depends on the ML architecture.

In summary, most recent studies [63–66] (See Table 1) confirmed that ML is indeed useful in efficiently predicting the evolution of species and mitigating the large storage occupation induced by tabulation methods. Three typical steps need to be considered using ML to calculate the chemical kinetics.

1 Data acquisition: generic

Ø Cover a wide range of combustion and turbulence conditions in terms of equivalence ratio, reaction progress, and enthalpy.

Ø Time interval control to avoid sub-iterations

2 ML algorithms

Ø ANN, MLP, and CNNs

Ø Label calculation. TensorFlow application; physical principle

3 Applications with the reduced mechanism

Ø Coupling of the mechanism and CFD

There are some choices for getting training dataset, either through experiments and simulations conducted by individuals, or publicly available database, such as Chemkin [67], Cantera [68], Engine Combustion Network (<https://ecn.sandia.gov/>), the NIST chemistry webbook [69]. The original data should be preprocessed before carrying out model training. Data preprocessing includes data cleaning, normalization transformation of data, data reduction. More detail information about data preprocessing can be to the book [70] written by García et al.

3.1.1. ML for predicting chemical sources

Many mathematical methods using sets of equations to derive the computation of chemical kinetics during combustion incur high computational costs. To solve this problem, ML has emerged as a surrogate model for the evolution of chemical species. For ML prediction, the source term is calculated by establishing the relationship of the thermochemical state as input and the corresponding source terms as outputs using ANNs or CNNs. The method is completely black in nature.

As early as the end of the 20th century, Christo et al. [71, 72] first attempted to use an ANN to substitute the nonlinear and complex reactions in combustion simulation. Training data are generated from small-scale probability density function (PDF)/Monte Carlo simulations using direct integration for single-step chemistry [71] and the three-step reduced reaction scheme for H_2/CO_2 flame [72]. Their ANN model used in combustion simulation achieved impressive acceleration compared to the direct integration approach and took less memory than the traditional tabulation methods. A major limitation of ANN as an approximator is that its capacity is limited by the data used for training with a

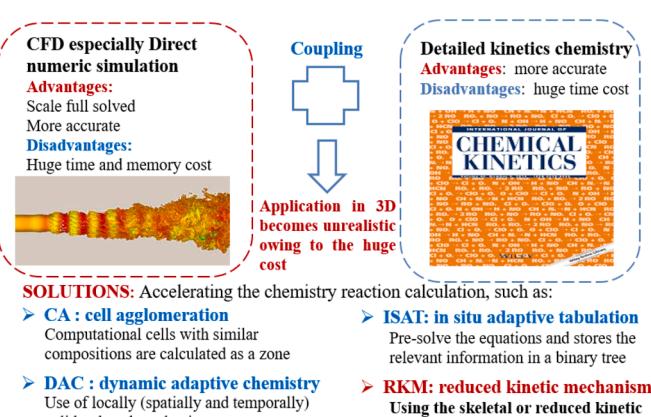


Fig. 4. General methods for chemical kinetics calculation acceleration.

Table 1

Summary of chemical calculation acceleration using ANN as an approximator.

| Authors & Refs. | Methods | Descriptions | Input | Output |
|---------------------------------------|-----------------|--|------------------------------------|-------------------------------------|
| Blasco et al. 1998, 1999 [76, 77] | ANN | ANN model for modeling the temporal evolution of species in methane combustion. | $Y_N(t)$ | $Y_N(t + dt)T(t),$ |
| Blasco et al. 2000 [79] | SOM+MLP | Modeling methane-air combustion in partially stirred reactor simulations. | $dt, T(t), Y_N(t)$ | $T(t + dt)Y_N(t + dt)$ |
| Sen and Menon 2009 [80] | ANN | Performed LES of turbulent flame-vortex interaction using ANN tabulated chemistry of syngas. | $Y_N(t)$ and T | ω_{Y_N} |
| Sen and Menon 2010 [63] | ANN | Using ANN to tabulate the evolution of chemical species based on linear eddy mixing. | $Y_N(t)$ and T | ω_{Y_N} |
| Sen et al. 2009 [64] | ANN | Using ANN-based tabulated chemistry to simulate extinction and re-ignition. | $Y_N(t)$ and T | ω_{Y_N} |
| Chatzopoulos and Rigopoulos 2013 [65] | SOM+MLP | Reduce the computational cost of the rate-controlled constrained equilibrium modeling for non-premixed $\text{CH}_4/\text{H}_2/\text{N}_2$ turbulent combustion. | $Z(t)$ and $Y_N(t)$ | $Y_N(t + dt)$ |
| Franke et al. 2017 [66] | SOM+MLP | Using MLP to tabulate the combustion chemistry to emulate ignition. | $Y_N(t), h$ | $Y_N(t + dt)$ |
| Peng and Pinkowski 2017 [75] | ANN | Using ANN as an approximator for reacting H_2/O_2 chemical systems. | $Y_N(t), T(t), P(t)$ | $Y_N(t + dt), T(t + dt), P(t + dt)$ |
| An et al. 2020 [18] | SOM-BPNN | Constructing ANN-based mechanism of hydrogen and kerosene for supersonic turbulent combustion. | $Y_N(t), T(t), P(t)$ | $Y_N(t + dt), T(t + dt), P(t + dt)$ |
| Wan et al. 2020 [73] | ANN | Tabulation of the combustion chemistry for syngas turbulent oxy-flame with side-wall effects, coupling ANN-based mechanism with DNS. | $Y_N(t), T$ | ω_{Y_N} |
| Wan et al. 2020 [78] | CNN | Detailed chemistry reduction for syngas turbulent oxy-flame based on side-wall effects using CNN. | Images built from $Y_N(t)$ and T | ω_{Y_N} |
| Sharma et al. 2020 [74] | ANN | Using ANN to build hydrogen-oxidation reaction model, and integrating it into CFD code. | $Y_N(t), T(t), P(t)$ | $Y_N(t + dt), T(t + dt), P(t + dt)$ |
| Barwey et al. 2020 [33] | K-means and ANN | K-means for combustion regimes classification and build ANN model for source term modeling in each regime. | $Y_N(t), \rho, T(t), P(t)$ | ω_{Y_N}, ω_T |

Notes: time t , mixture fraction Z , species Y_N , enthalpy h , temperature T , pressure p , chemical sources ω_{Y_N} , temperature source ω_T , density ρ , difference Δ , neural network types: BPNN, MLP, ANN, CNN.

high-dimensional dynamical system [66]. The effectiveness of a trained ANN model highly depends on the similarity between the testing and training cases. For example, the ANN model trained considering side-wall effects could be well applied to simulations with side-wall effects [73]. The high-efficiency 2-D DNS calculated results based on the ANN-DNS approach are presented in Fig. 5 [73]. The mass fractions and temperature obtained by ANN-DNS were in good agreement with the prediction using the reference detailed mechanism (GRI-3.0). More importantly, the computational cost of the chemical source was significantly decreased to a very low level compared to the convection-diffusion calculation using the ANN method.

Many trained ML models have the drawback of lacking enough scalability which limits its accuracy and wide usage. To improve the scalability of the ML model for chemical kinetics, Sharma et al. [74] using normalization and regularization methods before training the model. To ensure accurate prediction, sum of all the mass fraction for species is guaranteed to one. The input data for their model were normalized and scaled to small range using log normalization, which is useful for free radicals, whose species concentrations skew towards zero [75]. The loss function of their model has three components corresponding to the temperature, pressure and species concentration loss. L1 and L2 regularization, which has the ability of shrinking unimportant parameters of the model, are also added to the loss function to improve the robustness of the model. After their model is trained, they developed a lightweight wrapper library to integrate their trained model to CFD code for inference while calculating.

To improve the accuracy of the ANN model, classification or clustering methods are also adopted to classify datasets before ANN training; the most commonly used methods are K-means and SOM. Thus, to improve the ANN model fitness for the evolution of chemical species in a chemical system, Blasco et al. [76, 77] splits the composition domain into several subdomains and train the ANN for each subdomain. In addition, Barwey et al. [33] clustered the combustion regime in detonation waves into six groups, where the distinctions between detonation, deflagration, and intermediary regimes within the wave structure were observed, as shown in Fig. 6. For each cluster, an ANN network was built to guide the localized source term modeling from the input variables. Their results showed that the estimation accuracy of the ANN

model improved with the clustering methods.

As another effective method, An et al. [18] used SOM for classifying datasets such that the nonlinearity is reduced through the training of several ANN models, the schematic of which is shown in Fig. 7. Compared to K-means, SOM does not have to pre-set the number of clusters and could build a mapping from the N-dimensional input vector space to a 2-D space through competitive learning. By coupling the ANN-based chemical mechanism in large eddy simulation (LES) for a rocket-based combined-cycle combustor [18], the speedup of the chemical system integration was achieved.

Inspired by the applications of CNNs in image recognition, Wan et al. [78] used CNNs to predict the chemical sources (CNN output) from the 3×3 2D image inputs of species mass fractions and temperature (CNN input). A schematic of the training process is shown in Fig. 8. The training data are generated from DNS of turbulent non-premixed oxy-flame featuring side-wall effects with 80 snapshots, containing 520,000 data each. In the training process, a subset of 11 species as the input was selected to represent the major and minor species for syngas to obtain the corresponding chemical sources. Their CNN networks contain two 2D convolutional layers, two max-pooling layers, and four dense layers. The results show that the trained CNN reduced chemical model behaves better than Arrhenius in terms of CPU cost and species prediction.

Table 1 summarizes the chemical reduction works using ANN. Table 2 summarizes the memory and time occupation of ANN and traditional methods. In Table 1, they mainly focus on simplifying the chemical reaction mechanisms from the existing detailed chemical mechanisms of simple fuels. However, the oxidation chemistry for complex fuels, such as large hydrocarbon fuels, may include hundreds of thousands of species and reactions. The development and use of the detailed chemistry of these complex fuels are challenging, and it is necessary to develop a reduced chemical mechanism from experimental results. Generally, the oxidation process for complex fuels comprises two stages: fuel pyrolysis and small species oxidation. In the pyrolysis stage, the complexed fuels cracked into smaller fragments ($C_0 - C_4$) during the later oxidation process. Hybrid chemical models, such as the HyChem model introduced by Wang et al. [84], have been proposed to describe the oxidation chemistry of complex fuels. The hybrid chemistry model relies on experimental measurements of the time histories of fuel

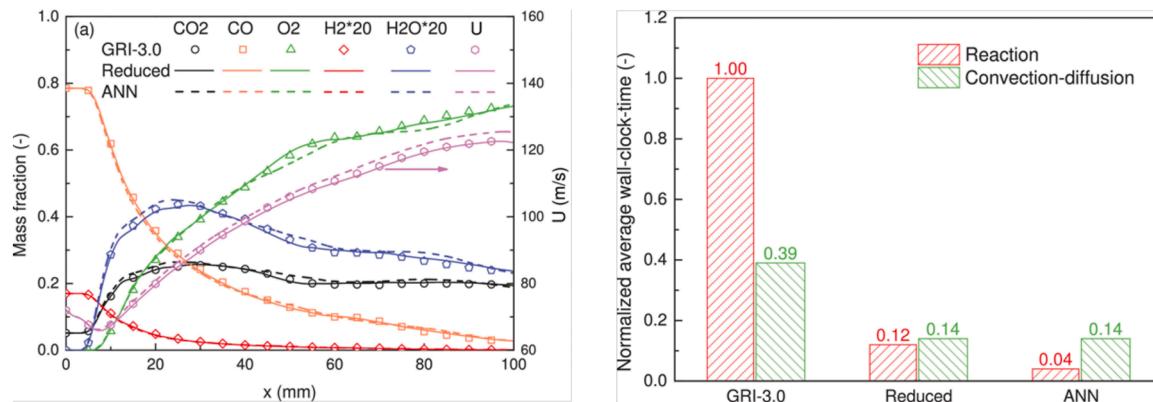


Fig. 5. Averaged distributions along the planar fuel-jet centerline and computational cost using different methods. (Reduced: the 11-species reduced mechanism) [73].

fragments to generate a reduced chemical representation of fuels. In [22], Ranade et al. built an ANN model to directly extract chemical reaction rate information from the measured temporal profiles of *n*-dodecane (*n*-C₁₂H₂₆) fragments during the pyrolysis stage. Through the combination of the ANN model with C₀–C₄ chemical mechanism, high-temperature *n*-dodecane oxidation chemistry was obtained. The ANN-based hybrid chemistry showed a good agreement with the detailed chemistry. Subsequently, an extended hybrid chemical framework combining ANN in the pyrolysis process and C₀–C₂ chemical mechanism for complex hydrocarbon fuels was developed by Ranade et al. [85], and their framework was validated using a perturbed numerical dataset. Alqahtani et al. [86] adopted the PCA dimension reduction approach to identify key species that could track the evolution of the chemical system and their reaction rates from simulations using detailed chemical mechanisms and used an ANN model to regress the reaction rates of those key species. It should be noted that their built model is suitable for a specific class of problems because it is limited by the trained data. It should be used carefully when exceeding the bounds of its corresponding composition space.

GPUs have thousands of arithmetic logic units compared with CPUs which commonly have only 4 or 8, so GPUs are good for performing compute-intensive tasks and highly parallel computing to reach high throughput efficiency. GPUs could be used for high performance scientific computing [87]. Recently, some efforts [88–90] have been paid to extracting the best performance of GPU for accelerating chemical kinetics calculation. Barwey and Raman [88] presented a GPU investigation for matrix-based formulation of chemical kinetics. In their work [88], chemical kinetics equations are organized from a matrix perspective for chemical species source term evaluation. The effects of the number of reacting cells in a domain offload to the GPU and the reaction types reflected by the number of species and reactions, on GPU saturation and speedup performance are studied. After computational analysis, they found that matrix-based methods supporting sparse

algorithms enable highly efficient GPU performance, and that designing larger mechanism with more standard, irreversible reactions is more favorable in GPU settings. Their work is meaningful in advancing the application of GPU on accelerating the calculation of chemical reactions. Besides, Buchheit [90] developed ML-hybrid algorithm for the integration of stiff chemical kinetics. In their model, a shallow NN is used for replacing polynomial fit to estimate specific thermodynamic quantities of interest. Their developed solver then deployed on GPU platform, demonstrates the results of 200 times faster than CPU calculation.

3.1.2. ML for tabulating lookup tables

Tabulation methods such as the flamelet model can accelerate the calculation of turbulent combustion reactions by indexing a lookup table, where the thermochemical scalars (temperature *T*, pressure *p*, species *Y_i*, and their source term ω_i) could be obtained through input control variables such as mixture fraction *Z* and progress variable *c*. In some combustion simulation cases, more control variables (e.g., enthalpy *h*, scalar dissipation rate χ) are required in the lookup table, thus making it much more complex. In addition, the lookup table usually requires a rather fine resolution, causing a large storage occupation. ML, especially ANN, is an efficient approach for tabulating relative to traditional lookup tables because it can significantly reduce the storage requirements.

An ANN can be trained to predict species mass fractions, density, and viscosity through the input of the mixture fraction and scalar dissipation rate in the investigation of the structure of the diffusion flame [91]. Unlike traditional lookup table methods, ANN features a smooth interpolation between the sampling points, which can minimize the erroneous velocity fluctuations. Their LES combined with the ANN-based tabulation showed good agreement with the experimental data, confirming the effectiveness of the trained ANN model. Combined with SOM for clustering data, the improved ANN was trained to mathematically relate the thermochemical and the parameterized quantities, and the

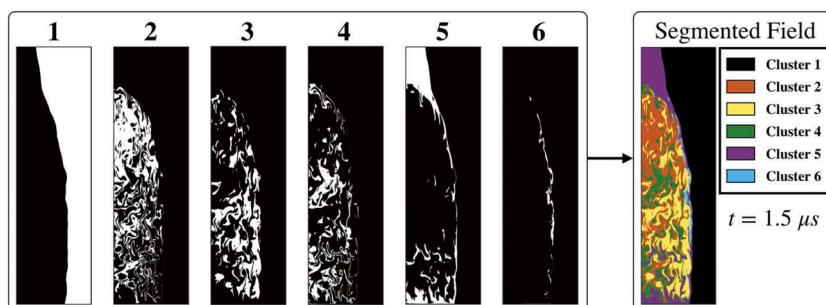


Fig. 6. Clustering results for combustion regimes within the detonation wave structure in [33].

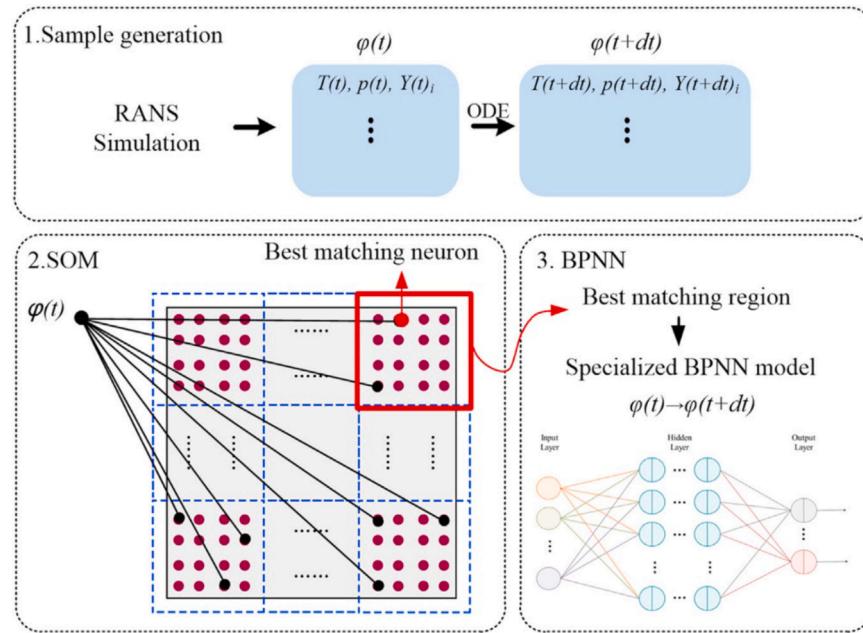


Fig. 7. The flowchart of the training process by combining SOM and BPNN (A type of ANN) in [18].

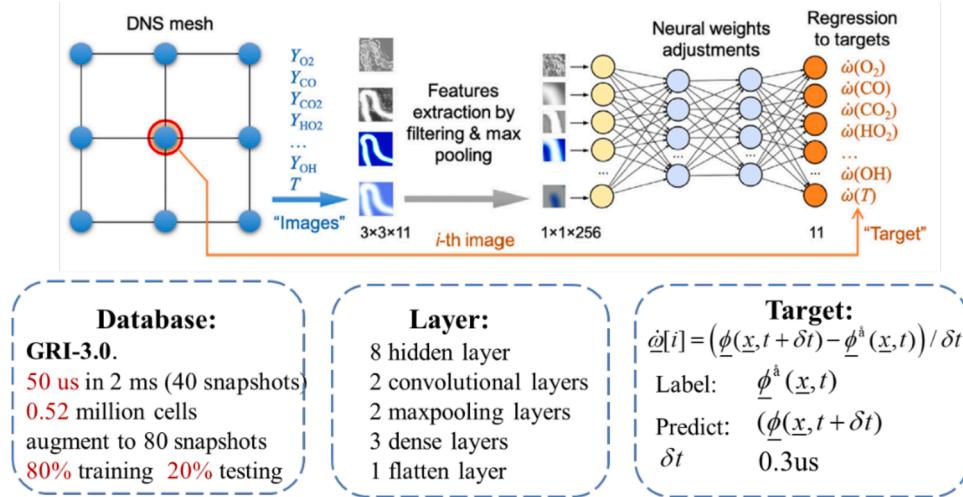


Fig. 8. The CNN training process to predict the chemical sources [78].

trained ANN model was validated in a DLR-A turbulent jet diffusion flame simulation[34]. The results show sufficient accuracy, whereas the trained ANN takes much less memory storage (less than 0.5 Mb) than the lookup table (500 Mb), in the order of 1000. For coupling with the combustion model, an ANN model was trained to mitigate the memory issues of the FGM table [81], where the input of the model is two variables, that is, Z , c , and the output is the mass species fraction Y_i . The trained model was validated in a spray H flame simulation. The results show that the ANN model can replicate the lift-off length and ignition delay time as the FGM does, and only eight times less memory storage is required.

Different simulation cases were studied by combining ML-based tabulation methods, for example, bluff-body swirl-stabilized flame simulation[92], modeling $CH_4/H_2/N_2$ jet diffusion flame[93], multi-dimensional combustion manifold modeling [83], ANN tabulation for Eulerian stochastic field FGM model [82], ANN-based tabulation simulation of rocket engines [94, 95], progress variable tabulation using deep residual networks [96], tabulation of combustion manifolds using a

deep mixture of experts [97], flamelet library tabulation for spray simulation [17], and laminar flame speed tabulation [98].

3.1.3. ML for chemical composition reduction

In the last section, the variables chosen to represent the composition space are the mixture fraction Z and progress variable c . With the increase in complexity of the research problem, more control variables should be added to the combustion model, and the joint PDFs of those variables would be very complex to obtain. To overcome this issue, an alternative strategy is dimension reduction through PCA that can extract a smaller vector of principal components (PCs), which serves as a substitute for the above-mentioned variables (e.g., Z , c).

Following the methodology in [99, 100], the principal component scores \mathbf{Z} ($n \times Q$) are derived from the original data matrix \mathbf{X} ($n \times Q$, $[T, p, Y_1, Y_2, \dots, Y_{ns}]$), where n is the number of individual observations and Q is the number of independent variables, projected on the eigenvector matrix \mathbf{A} . The formula is given by Eq. (4).

$$\mathbf{Z} = \mathbf{XA} \quad (4)$$

Table 2

A brief summary of comparing the memory and time occupation of ANN and traditional methods.

| Authors & Refs. | Methods | Memory occupation (Ratios) | Computational time (Ratios) |
|---------------------------------------|---|------------------------------|-----------------------------|
| Blasco et al. 1998 [76] | Tabulated methods | 19,934kbytes (831) | 0.03 ms (0.17) |
| | Direct Integration | 276kbytes (11) | 29 ms (165) |
| | ANN | 24kbytes (1) | 0.176 ms (1) |
| Sen and Menon 2010 [63] | tabulated methods | >67.5Mb (630) | (>200) |
| | ANN | 0.107Mb (1) | (1) |
| Chatzopoulos and Rigopoulos 2013 [65] | Tabulated methods | – | 8000 s (533) |
| | ANN | – | 15 s (1) |
| Franke et al. 2017 [66] | Tabulated methods | >>36Mb | 125 s (83) |
| | ANN | Only 36Mb | 1.5 s (1) |
| An et al. 2020 [18] | Direct Integration | – | (>8) |
| | ANN | – | (1) |
| Wan et al. 2020 [78] | Detailed chemistry | – | (>11) |
| | CNN-chemistry | – | (1) |
| Ranade et al. 2019 [34] | PDF table | 500Mb (1220) | 50 min (0.96) |
| | MLP | <0.41Mb (<1) | 52 min (1) |
| Zhang et al. 2020 [81] | FGM table:Single-core computing | 264 Mb | (0.33) |
| | 32-core parallel computing | 8448 Mb | (0.33) |
| | With an extra of dimension with 101 grids | 833 Gb | (0.33) |
| | ANN | Very small memory occupation | (1) |
| Zhang et al. 2020 [82] | P-PDF | 265.6GB | – |
| | ANN | 48.9 kb | – |
| Bhalla et al. 2020 [83] | Tabulation methods | 184.6Mb (7.6) | Serial:11ms |
| | NN | 24.2Mb (1) | Serial:55.27ms |

By truncating matrix A, the original data can be approximated by the PCs whose number is much less than the number of original variables. By solving the transport equations for PCs [100], the computational cost is dramatically reduced by chemical dimension reduction. In [101], two ML regression methods (SVM and Gaussian process regression) were used to estimate the source terms of the transport equations of the PCs in the simulation of a perfectly stirred reactor burning syngas. In their method, transport equations of two PCs are used instead of 11 original variables, and the results show great accuracy for the prediction of species and temperature. Malik et al. [99] used Gaussian process regression to map the thermal-chemical variables from PCs, and their simulation could handle the extinction and re-ignition phenomena properly. Mirgolbabaei and Echekki [102] trained an ANN model to map the original variables to PCs. When PCA and ANN are combined, superior reconstruction of thermochemical scalars in combustion can be achieved. Ranade et al. [103] constructed a framework for experimental-based turbulent combustion modeling by combining ANN-based closure terms and joint PDFs in the PC space. Dalakoti et al. [104] combined PCA and ANN for chemical tabulation under diesel engine conditions. Their approach shows great predictions for major species and their thermophysical properties.

3.1.4. Reduced-order modeling of combustion

The reduced-order modeling method is efficient in building low-dimensional and efficient models for mathematical models, and it has received significant attention in fluid mechanics [4]. Combing reduced-order modeling and machine learning could assist combustion simulation. Among the reduced-order modeling methods, the reduced basis (RB) methods [105] are a class of widely used methods; they

project high-fidelity solutions onto the reduced space. The RB method comprises two components: RB functions and reduced coefficients. The reduced space is spanned by these RB functions and is described by the related reduced coefficients. The RB functions are usually derived from snapshots of high-fidelity solutions using the POD [106] or greedy algorithm [107]. Traditional methods for solving the reduced coefficients use standard interpolation techniques, whose capabilities are limited to the number of snapshot samples available.

Recently, an ML-based regression model [108] was used to calculate the reduced coefficients in steady flows. Wang et al. [109] combined POD and ANNs for the reduced-order modeling of unsteady combustion flows. In their study, ANN was used to predicting the reduced coefficients through the input of temporal, physical, and geometrical parameters. Through the offline training of the ANN network for the reduced coefficients and the extraction of RB functions, online combustion prediction could be easily performed through ANN evaluation and the linear combination of the RB functions. Zhang et al. [110] combined POD and ANN methods for the reconstruction of the cellular surface of gaseous detonation from post-surface flow fields. In their work, POD was implemented on the post field and full field to obtain their reduced coefficients and RB functions. An ANN model was constructed to map the reduced coefficients of the post field to the full field. After the network is trained, the full pressure field can be reconstructed through its RB functions and the reduced coefficients predicted by the ANN model.

3.2. Combustion kinetic model uncertainty quantification

The combustion kinetic model representing the combustion process has many uncertainties owing to the high nonlinearity and dimensionality of the system. A need to quantify model uncertainties and their minimization exist. The uncertainty quantification of the combustion kinetic model and its propagation and minimization have been systematically reviewed by Wang and Sheen [111]. Research on uncertainty quantification can be divided into two groups: forward uncertainty propagation analysis and inverse problems. Forward uncertainty propagation is useful in evaluating the accuracy of chemical models with knowledge from experimental measurements and theoretical analysis. Without accurate knowledge of model rate parameters, constraining the models against measurements is required, which is the concept of the inverse problem.

Bayesian analysis [112] provides a framework for addressing inverse problems in uncertainty analysis. In Bayesian analysis, the posterior probability of the model input parameters $\mathbf{x} = [x_1, x_2, \dots, x_n]$, can be calibrated by a prior probability and a likelihood function based on the experimental data \mathbf{d} . The relationship between \mathbf{x} and \mathbf{d} (i.e., the Bayesian equation) is given by Eq. (5)

$$p_{post}(\mathbf{x}|\mathbf{d}) = \frac{p_{prior}(\mathbf{x}) * p(\mathbf{x})\pi(\mathbf{x}; \mathbf{d})}{\int p(\mathbf{x})\pi(\mathbf{x}; \mathbf{d})d\mathbf{x}} \quad (5)$$

where $p_{prior}(\mathbf{x})$ is the prior distribution, $p_{post}(\mathbf{x})$ is the posterior distribution, and the function $\pi(\mathbf{x}; \mathbf{d})$ is the likelihood function. The ML method, that is, the Markov chain Monte Carlo (MCMC), was used for the calculation of the equation. MCMC provides methods such as the Metropolis-Hastings algorithm [113] for sampling from a probability distribution through accept-reject sampling. When the MCMC method is used in uncertainty quantifications of the combustion kinetic model, samples used for constructing the Morkov chain are from a continuous random variable based on Monte Carlo sampling, with the conditional posterior distributions proportional to a known function. However, MCMC suffers from the curse of dimensionality. With an increase in the dimension, the volume of the sample space increases exponentially, and the computational cost of using MCMC is very high. The construction of a surrogate model for the original kinetic model to generate samples is an efficient way to accelerate the MCMC method. Consequently, a

methodology for handling high-dimensional uncertainty quantification based on the DNN surrogate model has been performed in many studies [114, 115]. It can be found that the ML-based surrogate model is suitable for high-dimensional uncertainty quantification and can be extended to solve the Bayesian analysis [114]. Recently, Wang et al. [116] built an ANN surrogate model for the approximation of combustion kinetic models, as shown in Fig. 9. The training data for the ANN model were generated from the original model simulations based on the input parameters (e.g., rate coefficient of the kinetic model) and model predictions (e.g., ignition delay time). After the ANN model is well trained, it is used to generate surrogate samples to solve the Bayesian equation using the MCMC method. By combining the ANN model, the ANN-MCMC can substantially reduce the requirement of the original samples with small relative errors. Consequently, the computational cost of the ANN-MCMC method is significantly reduced to one-tenth of that using the transitional MCMC method, and the ANN-MCMC method also enables easy application in model optimization of complicated combustion systems.

Similar to uncertainty quantifications, sensitivity analysis studies how the output of the chemical kinetic model responds to the input parameters, and it can identify those reactions that significantly affect the combustion properties. Sensitivity analysis based on global sensitivity analysis methods, such as random sampling high-dimensional model representations, has a very high computational cost. To alleviate the computational cost, Li et al. [117] trained an ANN model as a surrogate model from the samples generated from the original H₂/O₂ combustion models and used a well-trained ANN model to generate numerous samples for global sensitivity analysis. In their study, high computational efficiency was achieved. In addition, An et al. [118] built an ANN model for the sensitivity analysis of a C₂H₄ ignition model. Compared to the high-dimensional model representations method, global sensitivity analysis based on the ANN model could reduce the computational cost by two orders of magnitude.

3.3. Discovery of unknown reaction pathways

The discovery of the reaction network in the chemical processes of energy conversion, environmental engineering, and biology is

paramount to understanding the mechanisms of pollution and disease, as well as developing mitigation strategies and drugs [119]. However, revealing the reaction pathways for complex chemical processes remains challenging. However, ab initio calculations are computationally intractable, and there is limited prior knowledge about the reaction templates. Alternatively, ML can help discover unknown combustion reaction pathways. Ji and Deng [119] derived a one-neuron chemical reaction neural network for a single-step reaction through the law of mass action, and the Arrhenius formula. By stacking multiple neurons into one hidden layer, a neural network is formulated to represent multi-step reactions comprising n reactions and m species. The hidden layer neurons represent different reactions. After the network is trained with experimental data, the reaction pathways and rate constants can be derived from the weights and biases of the neural network without expert knowledge. Their build methodology has good physical interpretability.

It should be noted that reactive molecular dynamics (MD) also provide valuable insights into the chemical kinetics of combustion, and some novel reaction pathways can be found. Presently, although the empirical ReaxFF is widely used in MD simulations for the reactive process, owing to its high computational efficiency compared to the ab initio MD simulation [120], the accuracy and reliability in discovering the novel reaction pathway must be considered [121]. Recently, a study combining MD simulation and ANN for the discovery of reaction pathways for methane was performed by Zeng et al. [122]. MD simulations can discover new reactions and species with no prior knowledge. Through MD simulations, molecular species, including hundreds of intermediate radicals, were recorded. Subsequently, neural networks were trained using those from ab initio computed energies to generate the potential energy surface of MD simulations and describe the complex chemical reactions of methane combustion. Zeng et al. [122] provided a powerful method for understanding the chemical reaction process and discovering novel reaction pathways at the atomic level.

3.4. Subgrid combustion modeling

In a LES of turbulent combustion, the interaction between the flame and subgrid turbulence can significantly determine the flame behavior

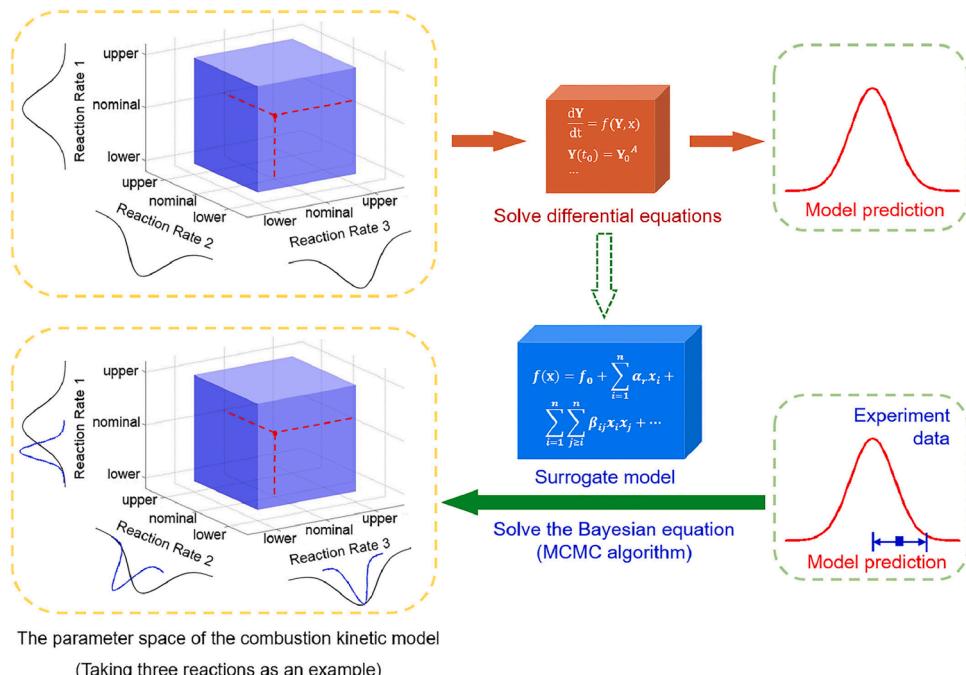


Fig. 9. The ANN-based surrogate model to solve the differential equations [116].

and emission products, which is a very important topic. In LES, the governing equations of combustion and fluid flows are filtered through a low-pass filter with a characteristic length $\bar{\Delta}$, leaving the terms that are larger than $\bar{\Delta}$ resolved and those within the filter length $\bar{\Delta}$ unsolved, resulting in many unclosed terms including filtered source term, stress tensor term, diffusion terms, scalar-flux term, etc. In the past decades, many turbulent combustion models have been proposed, such as PDF-like and flamelet-like models [123]. However, it should be noted that in the present turbulent combustion model used in LES, handling the unclosed terms involving a highly nonlinear equation is still a challenging task. In recent years, some studies [124, 125] have emerged to model turbulent combustion using ML methods. ML can be used for unclosed-term modeling, super-resolution (SR) of subgrid fields and flame surface wrinkling modeling.

3.4.1. Unclosed-term modeling

In LES, by using a low-pass filter to filter the governing equations of combustion and fluid flows, the spatial and temporal structure information under the filter length is blurred. So LES considers the less temporal and spatial structure and works with spatially filtered state variables, thus reducing the computational burden of simulating turbulent flow, regardless of being a reacting flow or non-reacting flow. However, the unsolved terms under the filtered length should be modeled using the closure models. Considerable attention [126] has been paid to the development of LES closure models for non-reacting flows. For reacting flows, such as combustion, considerable complexity arises because of the complex coupling of multi-scale turbulence and combustion. Many efforts, reviewed in [127], have focused on modeling the unclosed terms in turbulent combustion, such as unsolved source term and unsolved stress tensor. The use of ML methods to build closure models for LES combustion is an emerging research area that could build those models from the data without additional assumptions.

Progress variable source term modeling. Based on the chemical tabulation method, as mentioned before, all thermochemical parameters Φ are related to the progress variable c , which represents the combustion state of the local mixture and can be used to deduce the temperature or mass fraction of species. In the context of LES, the progress variable is solved using the transport equation shown in Eq. (6).

$$\frac{\partial \bar{\rho} \bar{c}}{\partial t} + \nabla \cdot (\bar{\rho} \bar{u} \bar{c}) = \nabla \cdot (\bar{\rho} D_c(\bar{c}) \nabla \bar{c}) + \nabla \cdot \tau + \bar{\omega} \quad (6)$$

where ρ is the density, t is the time, \mathbf{u} is the velocity vector, $D_c(c)$ is the molecular diffusion coefficient of c , τ is the subgrid scale flux, $\bar{\omega}$ is the source term of c , $\bar{\cdot}$ denotes the LES filter, and $\tilde{\cdot} = \bar{\rho}/\bar{\rho}$ denotes the Favre filter. In the equation, the terms (e.g., flux term τ and filtered source term $\bar{\omega}$) at scales smaller than the filtered width are unsolved and should be modeled.

In LES, most tabulation methods also involve filtered PDFs (FDFs) with presumed shapes, for example, β -PDF [128]. For example, the filtered source term $\bar{\omega}$ can be closed using the FDF of c or the joint FDF of $[\tilde{Z}, c]$, as shown in Eq. (7) and Eq. (8), respectively.

$$\bar{\omega} = \int \dot{\omega}(c) P(c|\bar{c}, \bar{c}'') dc \quad (7)$$

$$\bar{\omega} = \int \dot{\omega}(Z, c) P(Z, c|\tilde{Z}, \tilde{Z}'', \bar{c}, \bar{c}'') dZ dc \quad (8)$$

where $Z'' = (Z - \tilde{Z})^2$ is the square of the mixture fraction subgrid scale fluctuation, $c'' = (c - \bar{c})^2$ is the square of the progress variable subgrid scale fluctuation, $P(*)$ is the presumed FDF, $\dot{\omega}(*)$ is tabulated in chemistry libraries, and it can be obtained once the parameters are given.

However, the presumed PDF shapes may not represent the real scalar

statistical behavior across the combustion regime of a specific combustion mode, directly using the presumed PDF model without careful adaptions to a specific combustion simulation may lead to inaccurate or even wrong results.

ML can build data-driven FDF models for specific combustion configurations. Chen et al. [124] used a DNN to predict the joint FDFs of the mixture fraction Z and the reaction progress variable c . The input and output of the model contain 5 and 35×31 components respectively, where “5” stands for \tilde{Z} , \bar{c} , \tilde{Z}'' , \bar{c}'' and the covariance σ_{Zc} (corresponding to $(\tilde{Z}, \bar{c}, \tilde{Z}'', \sigma_Z^2, \sigma_{Zc}^2)$ respectively in Fig. 10, “ 35×31 ” stands for the number of $Z * c$ bin boxes. Training data were obtained from the DNS dataset of moderate intense low-oxygen dilution combustion. Combined with PCA techniques for removing the outliers in the training data, their trained DNN-predicted FDFs were used to compute $\bar{\omega}$ for a priori assessment. Their results showed that the filtered reaction rate predictions through DNN were in good agreement with the DNS data. Their predicted FDF shapes from DNNs are found to be quite different from classical FDF shapes, for example, the Gaussian model or bi-modal mostly used in the jet flame simulation. This indicates that the traditional presumed FDFs are not suitable for new combustion simulation configurations; they may be inaccurate or even wrong. With the boom of ML methods, FDFs can be modeled more accurately to perform a much more convincing LES for specific combustion configurations.

Similarly, in the work done by de Frahan [125], the joint-filter PDF model of \tilde{Z} and \bar{c} , is built through three different ML methods, that is, random forests, DNNs, and generative learning. Training data were obtained from a DNS of an experimental lean premixed turbulent low-swirl methane flame. Their trained model is evaluated on the predictions of the subfilter PDF as well as $\bar{\omega}$. Their results show that a DNN is more suitable for providing fast and accurate predictions relative to the traditional β -PDF model.

In the research of Seltz et al. [129], a CNN model was trained for the prediction of the filtered progress variable source term $\bar{\omega}$ and flux term $\nabla \cdot \tau$ in Eq. (5) in an LES. The training data for the model were extracted from the DNS database of a methane-air stoichiometric premixed jet flame. Once the model is trained, only given \bar{c} for the calculation of inputs to the model, the source and flux terms can be accurately predicted through the model without resorting to additional equations.

Progress variable dissipation rate modeling. The progress variable dissipation rate ($\chi_C = 2D_C|\nabla C|^2$) can be obtained from the first term on the right side of Eq. (5). The scalar dissipation rate, χ_C , is important in turbulent reacting flows. The variance c'' of c , used for the source term closing, is also related to the unresolved contribution of χ_C , i.e. $\tilde{\chi} = 2D_C|\nabla \bar{C}|^2 - 2\tilde{D}_C|\nabla \tilde{C}|^2$. Traditional algebraic progress variable dissipation rate models show relative simplicity; however, they lack sufficient accuracy. Yellapantula et al. [130] used a DNN to predict $\tilde{\chi}$, across a range of filter lengths in LES and flame conditions. Based on a few parameters, including principal rates of filtered strain, Lewis number, and several filtered variables, a data-driven prediction model for $\tilde{\chi}$ was built. Training data were obtained from the filtered DNS data of lean *n*-heptane/air premixed flames, and before the dataset was sent to train the network model, it was first clustered using a *k*-means clustering algorithm. Their trained model shows much better performance than the traditional model, and it exhibits some transferability to predict $\tilde{\chi}$ in other fuel flames.

Scalar dissipation rate modeling. For non-premixed and premixed combustion, such as spray combustion, the conditional scalar dissipation rate is an important input for the solutions of flamelet models, because spray combustion is affected by the premixed and non-premixed fraction dominated by scalar dissipation. Traditional closure models for scalar dissipation are inaccurate because of the absence of liquid phase

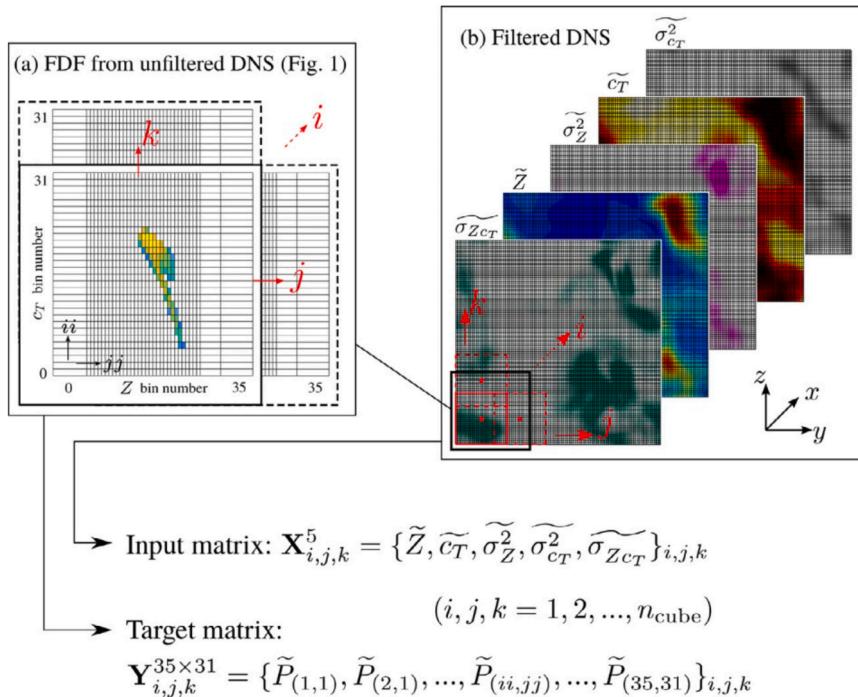


Fig. 10. Schematic demonstration of the data extraction for ML in [124].

properties in the models. To improve the accuracy of the prediction of the conditional scalar dissipation rate, Yao et al. [131] built a subgrid model for LES through ANNs with training data from carrier-phase DNS. The input for their prediction model includes liquid phase properties, such as evaporation rate. Their trained model identified the important features affecting the dissipation rates and proved the limitations of traditional models. Based on the ANN model for dissipation rate, the conditional mixing statistics (i.e., the FDF based on mixture fraction and conditional dissipation rate) were also modeled by Yao et al. in [132] through ML methods. Their results were in good agreement with the DNS data across the entire mixture fraction space.

Stress tensor modeling. In a LES, the filtered density-weighted momentum equation is given by Eq. (9),

$$\frac{\partial \tilde{\rho} \tilde{u}_i}{\partial t} + \frac{\partial \tilde{\rho} \tilde{u}_i \tilde{u}_j}{\partial x_j} = - \frac{\partial \tilde{p}}{\partial x_i} + \frac{\partial \tau_{ij}^r}{\partial x_j} - \frac{\partial \tau_{ij}}{\partial x_j} \quad (9)$$

where τ_{ij}^r is the resolved stress tensor and τ_{ij} is the unresolved tensor, which is given by Eq. (10),

$$\tau_{ij} = \bar{\rho} (\tilde{u}_i \tilde{u}_j - \tilde{u}_i \tilde{u}_j) \quad (10)$$

Many classical stress tensor models have been proposed for the closure of the unresolved tensor τ_{ij} , including the Smagorinsky model [133] and the scale-similarity model [134]. These standard models were developed and validated for incompressible and non-reacting flows, and their parameters should always be adapted to particular simulations, such as the dynamic parameter C_D in the Smagorinsky model [133]. Several studies [135–137] have been conducted to model the resolved stress in turbulent non-reacting flows using ML-based methods. However, in reacting flows, the stress tensor is affected largely by the massive variations in temperature, density, velocity, and viscosity across the flame front, which the classical stress tensor models do not account for; hence, the direct usage of those models in reacting flows may not yield excellent results.

Inspired by the model input variables evaluated in [137], Nikolaou et al. [138] modeled the unresolved stress tensor in turbulent premixed V-flames by establishing neural network-based closure models, where

the outputs are the six stress tensor components, and the inputs are normalized density and velocity gradients. Training data were filtered on the fine mesh of a high-fidelity DNS database, and mesh-coarsening effects were also studied. Compared with the classical subgrid models (e.g., the Smagorinsky model), the trained neural network model shows better fitness despite the highly demanded flow configuration. Schoepflein et al. [139] modeled the subgrid stress in LESs of premixed flames using gene expression programming an (ML method), also known as symbolic regression [140], which has also been used in the Reynolds-averaged Navier-Stokes (RANS) turbulent model formulation [141]. Symbolic regression is an evolutionary algorithm that can create models matching the input data, and it has recently been used to find expressions from the data generated from the equations from the Feynman Lectures on Physics [142]. In [139], the scalar coefficients of the basic functions in the stress equation were optimized using the gene expression programming algorithm, and the modeling of the full stress tensor was completed. Their work demonstrated the ability of evolutionary algorithms in the LES modeling framework.

3.4.2. Flame surface wrinkling modeling

In the premixed turbulent combustion simulation, the turbulence on the flame front wrinkles the flame surface. Under the flamelet assumption, the turbulence reaction rate corresponds to the flame surface area; consequently, the evaluation of the flame surface area in the subgrid scale of LES is very important. To estimate the subgrid scale contribution, Lapeyre et al. [20] built a CNN network to estimate the subgrid-scale flame surface density based on the topological information of the progress variable. Training data were obtained from the DNS database of a methane-air slot burner. The input data for the CNN model is the 3D filtered progress variable \tilde{c} , filtered from the DNS data with a filter width of eight meshes, and the output is the 3D normalized flame surface density Σ^+ , meaning the flame is wrinkled while $\Sigma^+ > 1$. By training the U-net [143]-inspired CNN model with several snapshots, the model could predict the flame surface density with great accuracy. A schematic for training the CNN network to predict the flame wrinkling factor in [20] is shown in Fig. 11.

3.4.3. Super-resolution (SR) of combustion field

The need for high-resolution flow data is a major pursuit of computational and experimental fluid dynamics. SR is the concept of reconstructing a high-resolution flow field from a grossly under-resolved flow field using some techniques. Despite the usefulness of the traditional SR method, including bicubic interpolation [144], they are not suitable for high-frequency reconstruction. Recently, ML has been proven to be efficient in handling image SR tasks by Dong et al. [145]. Fukami et al. [146] introduced an ML-based SR method for turbulent fluid flow field reconstruction. In their work, ML methods, including CNNs, have been used to reconstruct laminar flow from low-resolution data of a two-dimensional cylinder wake, as shown in Fig. 12. Subramaniam et al. [11] used GANs with a physics-informed loss function to enrich the turbulence flow field. Considering turbulent flows should satisfy the constraints including governing equations and boundary conditions which may not be obeyed by the generated data, a modification to the loss function which seeks to minimize the residuals of the governing equations for the generated data was incorporated, i.e. the physics-informed loss.

The SR used in LES needs to predict the fully resolved data from the filtered data. Here, the fully resolved data ϕ_H refers to the DNS data, and the filtered data ϕ_F is the LES data. Assuming a Gaussian filter kernel F is used in LES, the relationship between ϕ_H and ϕ_F is denoted as $\phi_F = F(\phi_H)$. The SR for the LES filtered data using ML is the reverse operation of F , denoted as $\phi_H = F^{-1}(\phi_F)$. Bode et al. [147] used GANs with a physics-informed loss function (proposed in [148]) for subgrid modeling in LES for turbulent reactive flow. Their GAN model was trained with the decaying turbulence data. Training data from lower and higher Reynolds numbers are used to train the model separately. The model trained on low Reynolds number data has poor reconstruction ability under high Reynolds number, which confirms that the flow field has different performance under different Reynolds numbers. Subsequently, their model was used for the SR of the subfilter turbulent flux and the Reynolds stress field in the turbulent reactive flows of the spray A case, and good results were obtained. For the training of a GAN model, the discriminator and generator should be both trained to maximize the classification accuracy of the discriminator and confusion ability of the generator. There are great challenges for achieving stabilization of GAN learning [13]. There are some tricks [149] for training a GAN model, such as using strided convolutions, minibatch discrimination, batch normalization, one-sided label smoothing.

In the LES method, the filtered density and the filtered density-progress variable product can be solved from the transport equation of the progress variable. Restoration of the original progress variable field has been challenging. However, the ML method can be accomplished through data training. Obtaining the original progress variable from the filtered field is similar to that of an SR task. In [21], the unfiltered progress variable field was reconstructed by deconvolution of the

filtered density and the filtered density-progress variable product through CNNs. The input for the training is the filtered data, filtered from the DNS data using Gaussian filters, and the output was the original DNS data. Based on their proposed networks on the topological features of the filtered data, good approximations of the original fields were achieved.

3.5. Building surrogate solvers for simulation

Simulating fluid flows or combustion in complex computational domains involves solving the Navier-Stokes equations and complex chemical reactions, which are computationally expensive. Although LES reduces the computational cost to some extent, a huge computational cost for complex problems is inevitable. The utilization of CFD techniques for fluid flows generates a large amount of data. ML provides various techniques for handling big data. A novel trend of utilizing big data in fluid mechanics for building ML-based surrogate solvers is emerging. Jin et al. [150] built a surrogate solver using CNNs to predict the velocity flow field around a circular cylinder based on the pressure around the cylinder. Sekar et al. [151] built a surrogate solver using CNNs and ANNs to quickly predict the flow field over airfoils. When the airfoil parameters, Reynolds number, and angle of attack are considered, the flow field around the airfoils can be predicted with sufficient accuracy.

For combustion simulation, An et al. [152] built a surrogate solver to solve the physical properties and species distribution of turbulent combustion in a cavity, which often acts as a flame stabilizer in combustion devices. They built their deep learning architectures, named CFDNN, which was inspired by U-net [143]. The training dataset for their network comes from unsteady RANS simulations of H₂, which comprises the process of flame ignition, development, and the formation of a stable turbulent flame. The input and output images of their network contain the distribution of velocity, temperature, and mass fraction of the nine species. As shown in Fig. 13, for a time sequence of the simulation data, the input is the flow information at t_0 , and the output is the information at $t_0 + \Delta t$. After the model is trained, given the inputs at t_0 , the field information at $t_0 + \Delta t$ can be predicted (as shown in Fig. 13 (b)). Their trained network could capture the flame evolution from the ignition, and it showed good agreement with the simulation results. Their exciting results demonstrate the feasibility of ML-based solvers for combustion simulation.

In summary, ML has performed well in improving the accuracy and computational cost of the combustion simulation compared to the traditional methods. Efficient and high-fidelity combustion simulations based on the ML method mainly involve 1) model reduction and tabulation for chemical acceleration, 2) dimension reduction in uncertainty quantification, 3) automatic discovery of kinetic models, and 4) flame identification for adaptive combustion modeling (discussed below). For the combustion simulation, the relationship between the input parameters involving the temperature, species, and pressure, and the output parameters involving the chemical kinetic calculation or chemical source were established based on different ML methods. It is noteworthy that the training process of neural networks considering the thermochemical state coupling with different turbulent scales can be obtained from the DNS method to further understand the turbulence-combustion interaction and establish robust and high-efficiency turbulent models.

4. ML in combustion measurement and diagnostics

Imaging techniques are the most effective methods for the measurement of complicated combustion phenomena such as turbulent combustion and supersonic reactive flow. They can help disentangle the complex interplay between flow dynamics, chemical kinetics, and heat and mass transfer. Generally, imaging techniques can be classified into two principal categories: planar imaging and tomographic imaging. Examples of planar imaging methods include planar laser-induced

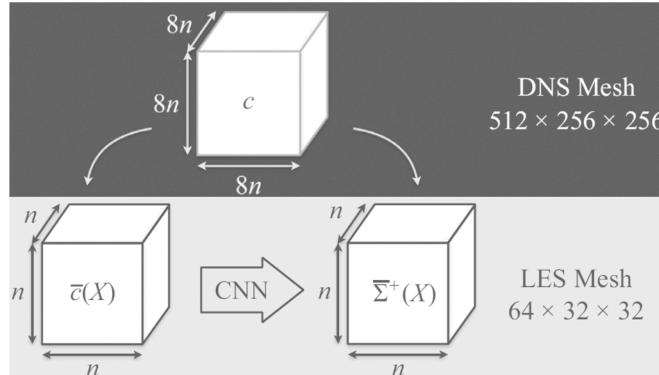


Fig. 11. The methodology of training CNN network to predict the flame wrinkling factor in [20].

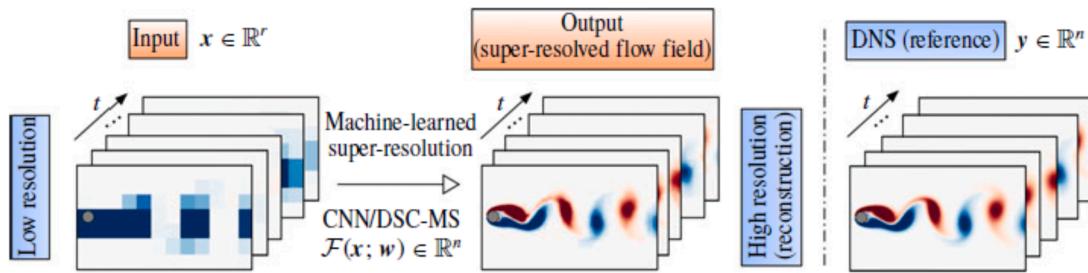


Fig. 12. SR reconstruction from low resolution based on ML [146].

fluorescence (PLIF), particle imaging velocimetry (PIV), and filtered Rayleigh scattering, where a specific plane is chosen for recording signals via 2D array detectors. In tomographic imaging methods, several projections are used to construct the flow fields. Examples of tomographic imaging methods include tomographic absorption spectroscopy, absorption-based nonlinear tomography, and flame chemiluminescence tomography. Recently, ML methods have penetrated the combustion measurement field in various ways.

4.1. Tackling the inverse problems in combustion measurement

In combustion measurement, directly measured quantities are sometimes not of interest to us. For example, in tomographic imaging, we are interested in the 3D flame structure instead of those projections, and in soot imaging, the quantity of interest is the soot volume fraction distribution rather than the measured transmissivity. Recovering the fields of interest from the measured quantities is called the inverse problem. Examples of the inverse problem in combustion measurement include the 3D reconstruction of the flame, retrieval of soot volume fraction fields, and temperature and concentration distributions from indirect measurements.

4.1.1. Reconstruction of 3D flame

Significant studies have been conducted to reconstruct 3D combustion fields from 2D projection frames. The principal task in the construction process is to solve a set of linear equations for each pixel, which can be mathematically expressed as Eqn 11:

$$W \times x = p \quad (11)$$

Here, W denotes the weight matrix, x is the vector of the target voxels to be constructed, p is the vector of the measured projections. Well-established methods for solving the equation include the algebraic

reconstruction technique (ART) [153] and the multiplicative algebraic reconstruction technique [154]. However, these algebraic methods incur great computational costs when processing thousands of projection frames. In addition, the generation of these huge amounts of frames requires a high experimental cost. Recently, some reconstruction methods based on ML have been proposed to avoid the problems encountered by these algebraic methods. For instance, several ML methods based on tomography reconstruction methods for combustion systems have been proposed [155–159].

To solve the high computational cost of nonlinear tomography for tomographic absorption spectroscopy, effective neural networks involving the extreme learning machine [155] and CNN methods [156] were employed to extract useful information from previous reconstructions, and then used for predicting reconstructions once given tomographic measurements.

The results show that the proposed neural network methods can substantially lower the computational cost with an accuracy similar to that of traditional methods. It should be noted that the inherent 3D structures of turbulent flames should be considered. Although the volumetric tomography (VT) method has proven to be able to recover complex 3D flame structures, it involves formidably high expenses owing to the limitation of using multiple high-speed cameras. Consequently, CNN-based tomographic reconstruction methods for turbulent flames were developed [157]. Eight projections of turbulent CH₄/air flames were used as the input of the trained network, and the reconstructed 3D flame structures by ART were used as the output. To avoid the problem of gradient vanishing during network training, a residual network with a skip connection was also studied in [157]. A flowchart of the 3D flame reconstruction process is shown in Fig. 14. Their trained network is immune to different classes of noise and can carry fast reconstruction, thus making online reconstruction possible. The results in Fig. 14 show that compared with the ground truth, the CNN method (VT-Net) performed better reconstruction of the 3D flame than the ART algorithm. It should also be noted that for the small-scale structure of a turbulent flame, the CNN still needs a more precise training process. Similarly, a trained CNN model was proposed to reconstruct a 3D flame from both simulative and experimental perspectives [160]. However, the difference is that their trained CNN model could accurately reconstruct the 3D flame structure of the CH* and C2* components, which are key components in flame imaging, based on the 12 projections of the flame. Similar to this work, chemical species tomography coupling with a dual-branch CNN was proposed for reactive flows [161]. Training data were obtained from an industry-oriented sensor with a few laser beams. Their trained network could accurately predict the temperature field and the distributions of H₂O concentration while bearing excellent robustness against noise. Generally, the present results demonstrate the strong capability of neural network methods for the reconstruction of 3D flame structures with credible accuracy and low computational cost compared with traditional tomographic reconstruction methods, such as ART. However, it should be noted that, to date, studies on this topic based on optical tomography using ML are still limited.

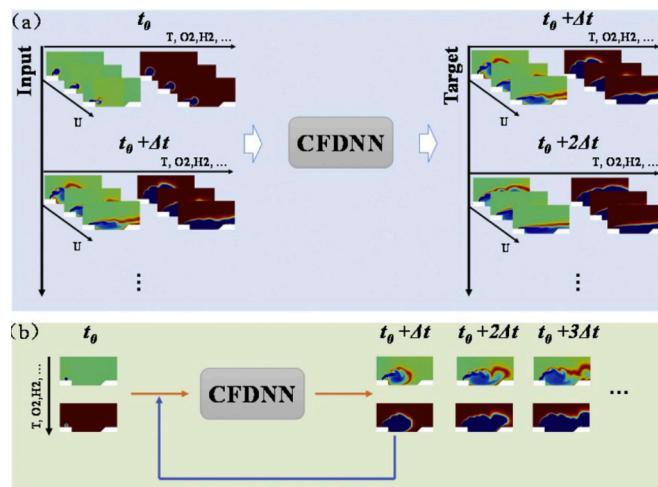


Fig. 13. The training (a) and simulation (b) process of the machine learning-based solver CFDNN [152].

4.1.2. Retrieving soot, temperature, and concentration fields

The retrievals of soot volume fraction fields, temperature, and concentration distributions from indirect measurements are important aspects in combustion diagnostics to understand the flame characteristics and combustion chemistry. For instance, the soot volume fraction is important for assessing the combustion quality. Typical soot measurement methods need to recover the soot fields from the measurements of soot transmissivity through the nonlinear Abel integral equation [162]. However, this method can lead to an ill-posed issue that small perturbations of the obtained information can cause a significant variation in recovering the soot properties, which is a common issue for the retrieval process for other scalar fields. To overcome this problem, Rodríguez et al. [163] trained a CNN model to reconstruct soot fields from projected flame images, which were generated through line-of-sight attenuation measurements in a canonical flame, and it comprises a soot volume fraction field. Their trained model was validated by constructing soot fields through real experimental measurements, and their model showed results similar to those of classical methods. Similarly, for the retrieval of temperature and concentration distributions, an ANN model was trained [164] to obtain temperature inversion and species concentration distribution from infrared emission measurements. A gas property database was used as the training dataset, which enabled the ANN model to build a map from infrared spectral emission to temperature and species concentration. The trained ANN model was experimentally validated, and a good agreement was found between the model prediction and experimental measurements.

4.2. Combustion regime diagnostics

Multi-regime combustion scenarios can be found in most combustion systems because of the partial premixing of the reactants, flow dynamics, finite chemistry effects, etc. The identification of different regimes in a combustion system at a certain moment or the classification of the combustion states of time-sequence flame images is important in combustion research and diagnosis. Many methods, such as gradient-free regime identification [166], have been proposed to identify

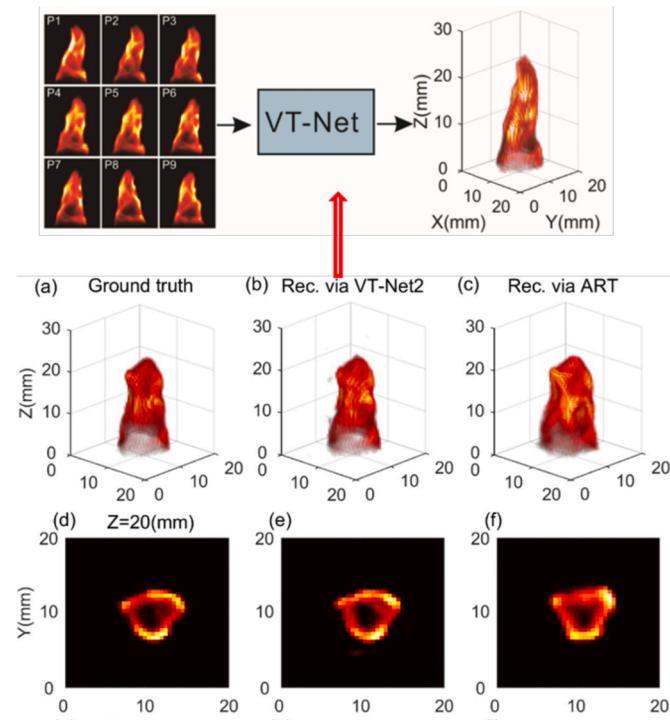


Fig. 14. 3D flame reconstruction based on the two methods VT-Net and the ART algorithm [157].

combustion regimes. Although these methods may provide accurate predictions in some cases, their computational costs are very high. Furthermore, these methods are limited in the identification of combustion regimes in complex combustion, such as detonation combustion. To overcome these problems, ML methods have been introduced. ML methods are suitable for classification and clustering problems. Studies [33, 167, 168] have been conducted to classify and cluster combustion regimes or states using ML methods. For example, Wan et al. [167] trained CNNs to identify combustion regimes from one-dimensional Raman/Rayleigh line measurements of the species mass fraction and temperature. Barwey et al. [33] clustered the combustion regime in detonation waves into six groups using k-means clustering methods before the source term modeling procedure. To classify the combustion modes in a supersonic combustor, Zhu et al. [169] trained a CNN model to extract features from the input raw pressure data to recognize its mode.

For a set of time-sequence flame images, the classification of these images can help us understand the combustion process and identify the combustion states. To understand the complex combustion phenomena and fluid dynamics in the hybrid rocket combustion process, Petrarolo et al. [170, 171] used an unsupervised clustering algorithm, that is, K-means++, to detect the different combustion phases from an image dataset characterizing the burning process. Their clustering results demonstrate the potential of the clustering algorithm in recognizing different burning phases. Based on their clustering results, the effects of oxidizer mass flow on the burning phase were observed. Subsequently, an extended work [168] was performed to quantify anomaly detection in hybrid rocket fuel combustion by combining spectral clustering and K-means++ clustering methods, which could both identify short-term irregularities and long-term combustion phases. Recently, Pino et al. [172] used SVM-based models to evaluate soot propensity and classify burning regime conditions based on flame images. Comparisons between their model and line-of-sight attenuation measurements were performed, and the results from both approaches were in good agreement. In addition, an ML framework for the combustion state classification in a furnace was proposed [173]. In their work, a denoising encoder was first used to remove the random noise of the inputs, and then GAN was used to extract the deep features from the unlabeled flame images. Finally, a Gaussian process classifier was used to predict the combustion state from the extracted features. The proposed ML framework exhibited great robustness and generalization abilities.

In addition, a data-driven method based on the use of random forest classification [175] is proposed to assign different combustion models comprising finite-rate chemistry (FRC), FPV model, and inert mixing was proposed to improve the computational efficiency for the simulation of rocket combustors. Fig. 15 shows the entire process of the random forest classification method used in the combustion simulation, where six key thermophysical parameters were used as input features and sub-model error was employed to construct the training labels. The results show that the simulation based on the data-driven method can improve the predictions in terms of the temperature and CO mass fraction compared with the simulation using the monolithic FPV method and save an additional 20% computational cost compared with the monolithic FRC simulation.

4.3. Combustion instability monitoring

Combustion instabilities cause deterioration and safety issues in combustion engines. Detecting the transition to impending combustion instabilities is important to optimize combustion efficiency and keep the engine safe. Recently, ML-based supervised and unsupervised methods have been used to monitor combustion instabilities.

Some studies have been performed based on supervised methods involving CNN and DNN, where labeled data are needed. In [176], Choi et al. built a supervised ML model that takes ResNet [177] (a variation of CNN) architectures as the backbone and integrated fusion layers to

classify combustion instability. The labeled data for training were obtained by synchronized flame imaging and pressure measurements. Their model was trained and validated using the combustion images of a gas turbine combustor. Gangopadhyay et al. [178] trained a 2D CNN model on sequential image frames to detect the onset of instability. Training data were collected from high-speed flame imaging of the Rijke tube, and the label of the data was the measured sound pressure. The robustness of the trained model was validated through its success in the prediction of instabilities induced by a different protocol. For DNNs, Omiotek and Smolarz [179] used deep recurrent neural networks to identify undesired combustion of a mixture of pulverized coal and biomass. The time series of the flame intensity was recorded to train the network. Their proposed network could monitor combustion and identify undesired combustion states in real time with no invasions. Recently, Wang et al. [180] employed both the CNN and multilayer DNN frameworks to simultaneously monitor the combustion state and measure the heat release rate in a furnace directly from combustion images. Their framework could predict the combustion state at an input rate of 1000 images per second with very high accuracy compared with the traditional methods, showing great potential for CNN in industrial applications.

However, the use of supervised ML methods for combustion stability monitoring requires a large amount of labeled data, which are usually difficult to obtain because they require prior knowledge and precise experimental settings. To overcome the label problem, Han et al. [181] used a stacked sparse autoencoder to extract image features from flame images of ethylene under different combustion conditions. Subsequently, the extracted features were clustered into two groups (i.e., stable and unstable), and the stable index of each image was calculated. Based on previous operations for generating labeled data, direct nonlinear mappings based on a DNN are built between the image features and the labels (stability labels and index) for fast combustion state prediction. Their unsupervised combustion monitoring framework shows a higher prediction accuracy than traditional data-driven methods. Akintayo et al. [182] built end-to-end convolutional autoencoders to generate soft labels, which are between the explicit labels of “stable” and “unstable” from the grayscale images of combustion. Their network comprises convolutional encode and decode layers, which are used to extract features from combustion images. If combustion transitioned from stable to unstable, the trained model could detect early

combustion instabilities from its generated soft label. It should be noted that an HMM probabilistic reasoning for combustion monitoring was used [183]. HMM can be used to describe the evolution of observed events; thus, it is appropriate for handling time-sequence data, such as time-sequence flame features. In their study, HMM was used to represent temporal behaviors by processing the spatial behaviors of combustion images. In addition, Ref. [184] combined a convolutional autoencoder, PCA, and HMM to identify combustion conditions and avoid unstable combustion in coal-fired boilers.

Recently, Liu et al. [185] built an ML framework for online combustion quality prediction, where both unsupervised and supervised ML were considered. Because of the traditional principal component analysis, which only extracts linear features, a multilayer deep belief network was proposed to extract the nonlinear features from online flame images and simultaneously obtain the informative information from the flame images, as shown in Fig. 16. It should be noted that in this study, supervised learning based on nonlinear regression was trained to establish the relationship between the features of the flame and oxygen content.

4.4. Spatial and temporal sr

As regards the reconstruction of the three-dimensional (3-D) spatial resolution of a flame, 3D computed tomography is an efficient method for combustion diagnostics [186], especially for turbulent flames. However, it requires expensive cameras that can capture both spatial and temporal high-resolution information, which limits their application in industry. Data-driven approaches, such as ML, can help to enhance the spatial and temporal SR of images captured by industrial cameras. For the turbulent combustion process, a framework combining the CNN model and a video interpolation model was proposed, which could increase the temporal resolution of the measured low-frame-rate 2D images [187]. Their model could increase the 200 Hz frame rate of an industrial camera to 3 kHz, enabling the use of industrial cameras in the study of hypersonic combustion that originally requires high temporal resolution cameras. In addition, SR is an effective method for improving the spatial resolution of tomographic measurements [188]. Based on the knowledge of SR, a 3D SR generative adversarial network (3D-SR-GAN) was proposed to predict the high-resolution 3D structure of turbulent flames. In this method, a GAN involving the elementary

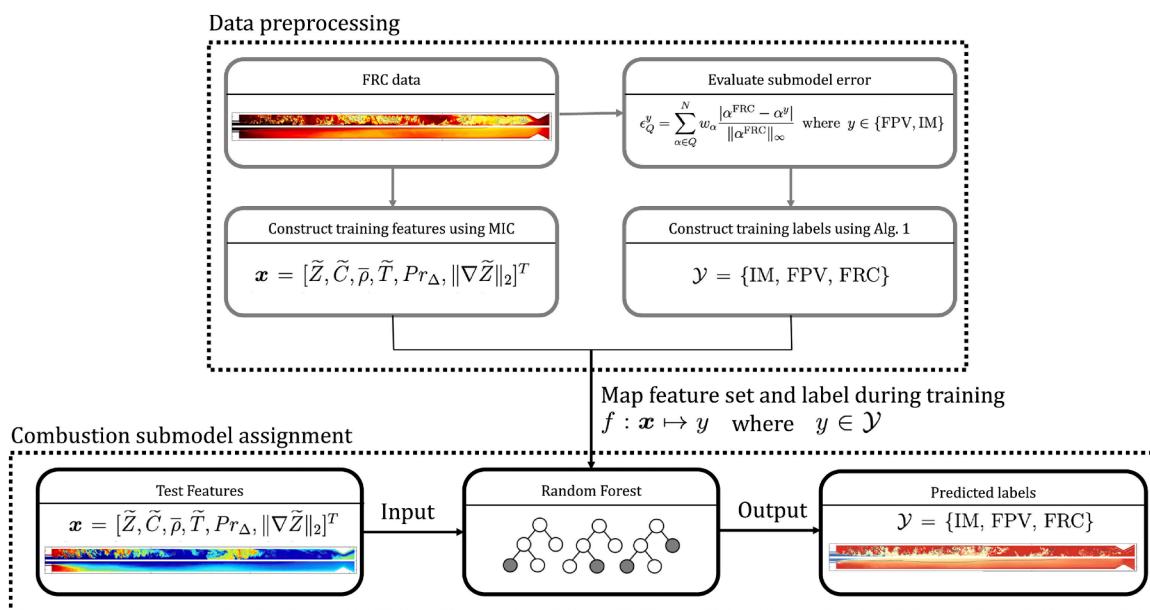


Fig. 15. Application of the data-driven method based on the use of random forest classification (finite-rate chemistry (FRC), flamelet progress variable (FPV) model, and inert mixing (IM)) [175].

layers of 3D convolution, batch normalization, and ReLU layers was trained [188] to learn topographic information, which can enhance the spatial resolution of the captured 3D low-resolution flame image. The GAN model was trained using simulated 3D turbulent jet flames. The model can increase the spatial resolution by a multiple of two with high fidelity, and the model behaves better than common direct interpolation methods.

4.5. Time-sequence prediction

In a combustion system, the combustion state evolves over time. Predicting the combustion states based on previous states is a challenging task. With the assistance of ML, the time-sequence prediction of flow fields or combustion fields is possible. RNNs [189] and LSTM [24] are popular models for time-sequence prediction. For instance, ML based on the framework has been proposed for the prediction of 3D flame evolution from its 2D projections in VT [24]. It is also related to the 3D flame reconstruction discussed above. In the framework [24], a CNN was used to extract features from the previous projections, and LSTM was trained to predict the evolution of 3D flame features, as shown in Fig. 17. Based on the predicted features (using the dataset during t_1-t_{10}), the evolution (at $t = t_{11}$) of the 3D flame structure could be reconstructed. Their framework proved to be effective in emulating the translation, rotation, scaling, and dilation of the 3D flame structure. For the RNN application, a bidirectional RNN model was used to predict the in-cylinder flow fields [189]. The dataset for model training was obtained from PIV measurements of an optical engine. Through their trained model, the velocity field in the cylinder at a moment can be predicted using the flow field at the previous and later time steps.

4.6. Build a mapping between flow field and flame field

Combustion is a complex process involving the coupling of fluid flows and chemical reactions. The combination of the velocity field measurements using PIV and species field measurements using PLIF in combustion research could help us to better understand the flow-flame interactions. Traditional qualitative research on the relationships between the flow field and species field requires expert knowledge, which may lead to the underlying interactions between the flow field and species field being ignored. ML can extract abstract features from a dataset and can learn to build a map between the flow field and species field. Barwey et al. [190] used a fully convolutional network in a

nonlinear regression framework to construct velocity fields from OH-PLIF images in a premixed swirl combustor. In this study, a global CNN model based on the entire combustion domain and a set of local CNNs based on subdomains were trained and compared. The trained models were validated for creating velocity fields for the attached and detached flame regimes. Fig. 18 shows the flame velocity field reconstructions from a single sample OH field for both the attached and detached flames in a premixed swirl combustor. It can be found that the local CNN model can accurately reconstruct the velocity field for the attached flame compared to the global CNNs and true PIV fields for PIV-x and PIV-y components. Regarding the detached flame, both local CNNs and global CNNs only obtain large-scale features; however, they cannot accurately predict the velocity fluctuation fields. Thus, in future work, other ML methods should be used to test special flames. Subsequently, Barwey et al. [191] further used ANNs to extract the overlap information of the OH-PLIF and PIV fields. The overlap is the common information in both the OH-PLIF and PIV fields, indicating the species and flow fields are not independent of each other. Their trained model can retrieve the flow field information from species fields.

In summary, the data-driven method based on ML is effective in the reconstruction and retrieval of flames, combustion diagnostics, time-sequence prediction, and distinguishing the flow field from the combustion field. Compared to traditional methods, ML can reduce the computational cost of processing thousands of projection frames and improve the accuracy (high SR). However, the reconstruction and retrieval of the small-scale structure of a turbulent flame is still a major challenge using ML to understand the in-depth mechanism, which may result from the limited number of experimental samples and noise in the measurement. If the error of the ML model trained by existing data is large, number of experimental samples for training the model are needed to be increased to reduce the error of the ML model.

5. Application of ml in engine and fuel research

5.1. Fuel properties prediction and fuel design

Owing to the environmental and sustainability concerns of petroleum-based energy, much attention has been paid to the usage and properties of alternative renewable fuels, such as biodiesel. Fuel properties can be primarily divided into two categories, that is, the physical properties (e.g., density, viscosity, heating value) and others (e.g., cetane or octane number, flash point, and ignition delay time), which are all important in the combustion process to construct physical and chemical models. In addition, compounds inside alternative fuels, such as hydrocarbons, alcohols, and esters, determine the properties of the fuels. Various quantitative structure-property relationship approaches have been proposed to predict these properties. Traditional methods are limited in their prediction accuracy, and to better understand the key factors affecting fuel properties and perform more accurate predictions of those properties, significant efforts have been made with ML methods [192–196].

Regarding the prediction of fuel properties using ML methods, especially the ANN method, studies have been performed to predict complex fuel properties, such as viscosity [192, 193], density [193, 197], and cetane number (CN) [194, 195]. Fig. 19 shows a sketch map of the ML model (neural network) to predict fuel properties. Studies [198–202] used ML methods to regress fuel properties (e.g., flash point, melting point, CN) from molecular descriptors and functional groups inside the fuels, and identify key factors affecting those properties. For instance, the application of an SVR algorithm in near-infrared data to predict flash points and CNs was performed and presented significantly better results than those of partial least squares [196]. It should be noted that SVR belongs to a supervised learning algorithm based on statistical learning theory, which can produce a global model to efficiently solve nonlinear relationships. The ignition delay time of a single fuel [203–205] or mixture fuels [206, 207] has also been studied under

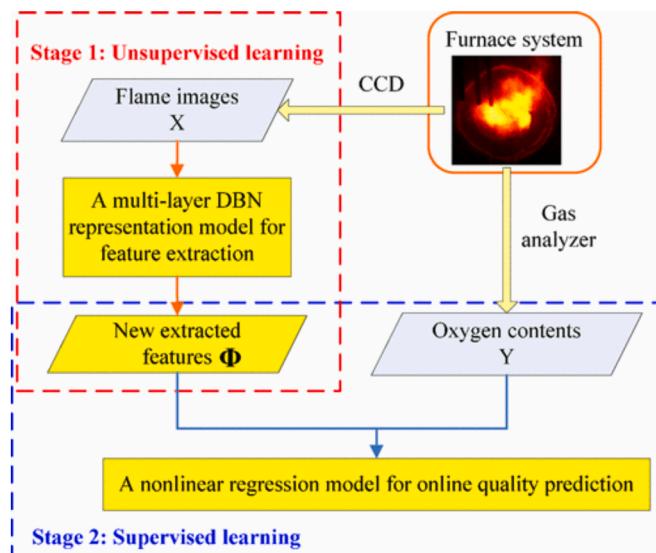


Fig. 16. The research methodology for online prediction of the oxygen content directly using unsupervised ML and supervised ML [185].

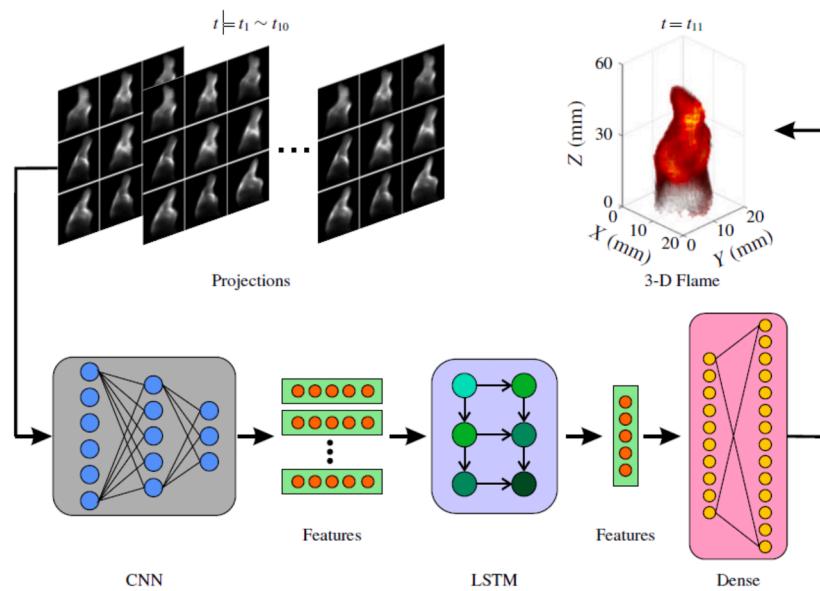


Fig. 17. Prediction of 3D flame evolution based on the proposed CNN–LSTM model [24].

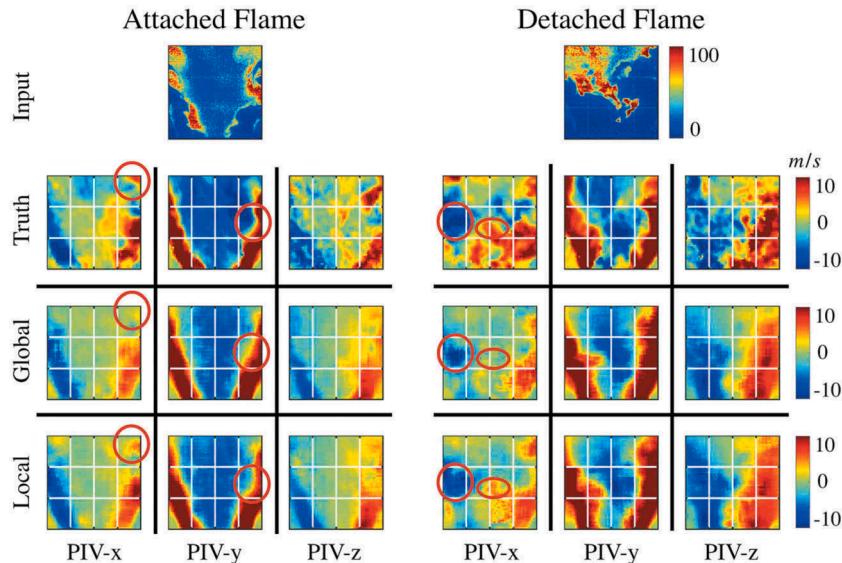


Fig. 18. Flame velocity field reconstructions. (Top) Sample OH-PLIF input from testing set. (Bottom) Corresponding PIV reconstructions for all three components using the global and local CNNs [190].

different conditions (e.g., pressures, equivalence ratios, fuel ratios, etc.). Based on the trained ANN for ignition delay time, ANN, coupled with an empirical prediction model of ignition delay time was developed to predict the start of combustion in homogenous charge compression ignition (HCCI) and was in good agreement with the results using detailed chemical mechanisms over a wide range of input parameters [203]. In addition, an ignition delay prediction using a developed ML method based on the BPNN model coupled with a genetic algorithm (GA) was presented for n-butane/hydrogen dual-fuel [207]. It should be noted that although the proposed method did not perform well in agreement, the results showed that the predictions could be improved by optimizing the weights and biases of the neural network compared with the original BPNN model. However, the above studies still lack the important feature of ignition delay: negative temperature coefficient, which should be strengthened in future studies.

In general, the octane number is measured by a cooperative fuel research engine; however, the engine is not convenient because of its

large scale [208]. Consequently, some alternative methods without direct cooperative fuel research engine tests have been proposed, such as the linear mixing rules and ignition delay times of stoichiometric fuel/air mixtures. In this method, neural network methods have demonstrated an excellent ability to predict the research octane number (RON) and motor octane number [194, 195] for single and blended fuels. For instance, a feed-forward neural network was developed, where the ignition locations and properties of various biofuel blends were used as inputs and the octane numbers as outputs [208]. Fig. 20 shows the RON predictions with absolute error for five biofuels blended using a micro-flow reactor neural network, nuclear magnetic resonant neural network, and ignition delay time correlation. It can be seen that the absolute error of the RON is less than 2 on average. In addition, training with more examples can improve the accuracy of CN predictions.

The complex real fuel biodiesel derived from plants or animals is a promising renewable alternative for diesel engines. ML is more suitable for the predictive capability of biodiesel properties and design owing to

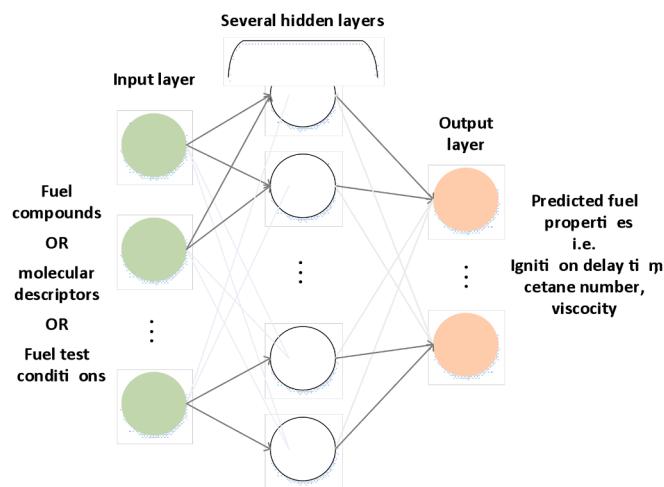


Fig. 19. Machine learning model for fuel properties prediction.

its highly complex characteristics. Compared to traditional prediction methods, ML methods can consider the interactions between individual components, thus yielding better predictions. For this topic, detailed information can be found in a recent review [16].

The design of new-generation fuels, such as high-energy-density hydrocarbon fuels, is an essential requirement to meet the urgent design of new low-carbon and high-efficiency engines. ML can be vital in assisting in the design of new-generation fuels, can also be used to build a surrogate fuel model for real fuel involving several hundred compounds. For instance, a neural network-based strategy for accelerating the design of high energy density and clean fuels has been proposed [209]. In this study, a database comprising the structures and properties of 342 hydrocarbon molecules was used for training, as shown in Fig. 21. The input is the nuclear charge information and Coulomb matrix representing the molecular structure, and the output is the fuel structure and properties. Finally, a group of 28 highly potent hydrocarbon molecules with high net heat of combustion, high specific impulse, high density, and low melting point was identified. Their trained model can be used to predict the properties of hydrocarbon molecules and search for new fuels.

5.2. Engine performance

Through the combustion of fuels, a combustion engine can generate

heat energy, which can be further converted into mechanical and electrical energy for human use. The combustion inside an engine directly affects its performance. However, various factors affect combustion efficiency and engine performance. These factors may be intertwined with each other. The traditional single-variable method for studying engine performance has significant limitations. Compared to traditional models, ML models can handle high-dimensional input data, which enables the prediction of engine performance based on multiple input parameters. Through further parameter optimization of the trained ML model, the engine performance could be further optimized.

5.2.1. Effects of key factors on engine performance

The performance of an engine is influenced by many factors, such as fuel type, fuel injection strategies, combustion mode, compression ratio, fuel-air ratio, combustion chamber geometries, and engine speed. The research and optimization of engine performance across all these factors are costly. The combination of ML methods and experimental engine research could help better predict engine performance, such as exhaust emissions, engine efficiency, and indicated mean effective pressure (IMEP) of spark-ignition engines using alternative fuels, which could also reduce the reliance on experiments [210–213].

Alternative fuels. As substitutes for both petrol and diesel fuel, alternative fuels such as ethanol, biodiesel, hydrogen, and compressed natural gas are receiving increasing attention from researchers. The study of the performance of engines using alternative fuels is necessary for their wide application. Many studies have been conducted on ML models to study engine performance based on alternative fuels such as fried-oil diesel blends [214], soybean biodiesel [215], alcohol-gasoline blends [216], hydrogen dual-fueled diesel [217], biodiesel-bioethanol-diesel blends [218], butanol, and ethanol [219]. In the above-mentioned studies, a part of the experimental or simulated datasets was obtained by varying the input parameters for training the ML models, and the remaining data were used to validate the ML models. The trained models show high accuracy for engine performance prediction and can be used as a surrogate model for engine parameter optimization, thus reducing the reliance on experiments. A comprehensive review of the applications of ANNs in engine performance and exhaust emission characteristics using alternative fuels was presented [210].

Combustion state. The prediction of the combustion state in the compression ignition combustion mode is challenging because its combustion is controlled by chemical reaction kinetics and is sensitive to working conditions. Stable combustion phasing under a wide range of

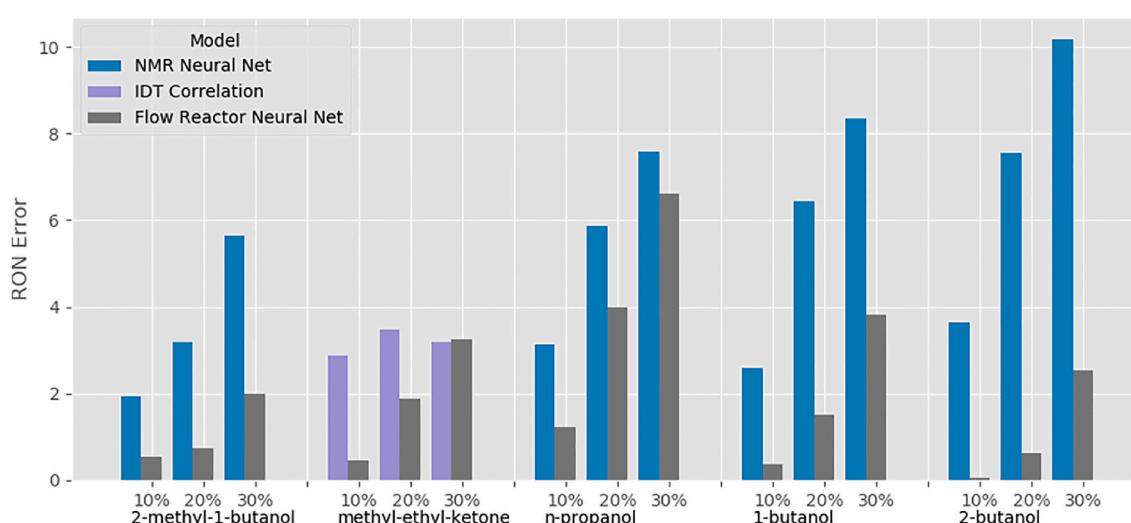


Fig. 20. Absolute error of RON predictions for five biofuels blended using different methods [208].

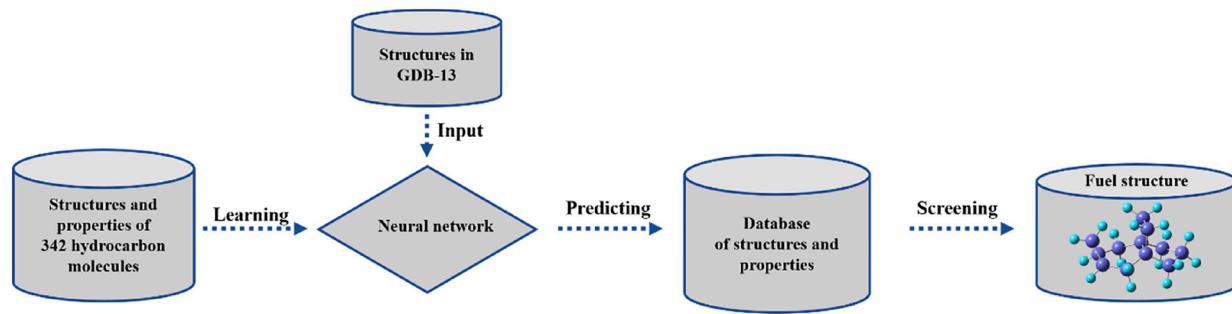


Fig. 21. The data processing methodology for discovering new fuels via ML-enabled high-throughput screening [209].

conditions for HCCI is of vital importance. Consequently, ANN methods involving a nonlinear autoregressive network with exogenous inputs, MLP, and radial basis function were employed to predict the start of ignition in an HCCI engine [220]. Meanwhile, the ANN can be optimized by a GA to obtain a high regression ratio for using network architectures, and the time required for training the network can be reduced. The input parameters systematically include engine speed, intake temperature and pressure, air-fuel ratio, exhaust gas recirculation rate. The proposed model could predict the start of combustion with good accuracy, and this provides a novel perspective for HCCI engine control, as shown in Fig. 22. Subsequently, with the help of an SVM and K-nearest neighbor techniques, Angikath et al. [221] successfully predicted the combustion state under the HCCI mode by considering several input features, such as the octane number of fuels, fueling rate, intake temperature and pressure, and compression ratios.

Effect of cycle-to-cycle variation. Cycle-to-cycle variation in engines introduces significant uncertainty in mixture formation and flame development, also leading to different turbulent flows in the cylinder in each cycle. In previous studies, a conventional linear correlation was used to identify the relationship between the flow and maximum pressure [222, 223]. In contrast, ML can solve nonlinear issues and obtain in-cylinder turbulent flows and flame topology characteristics under the effect of cycle-to-cycle variations. Thus, to qualitatively judge the effects of the flow field on the performance of an engine based on optical experiments, different ML methods (i.e., multilayer perceptron, AdaBoost, decision tree) as binary classifiers were employed to predict high IMEP or low IMEP from the cylinder flow fields based on the PIV measurements, and these methods show approximately the same results [224]. The classifiers can perform a nonlinear analysis. The results show that the in-cylinder basic flow properties in the center-plane play an important role in the classification accuracy, which can provide more information than the engineered characteristic tumble features. This implies that ML methods are less sensitive to engineering expertise. Hence, further studies regarding the prediction of combustion cycles of high energy and low energy based on the flame topology during the early flame kernel stage are required [224]. It can be found that coupling with the flame position and shape features, the cross-section of the early flame is necessary for the high-fidelity training of decision-tree-based ML. Note that the feature importance analysis can lower the requirement of the training samples without a loss of prediction accuracy [224]. Similar work was performed where an ANN with 109 experimental inputs was trained to predict the cycle-to-cycle variations of the IMEP in an optical spark-ignited propane engine [223]. The research covered many more factors such as intake and exhaust port pressures, spark energy deposition, global fuel-air ratio, dilution level, Lewis number, and air-flow-related factors. Based on simulations, a random forest model was adopted to determine factors including the flame topology and flow field, affecting the cycle-to-cycle variation of a gasoline engine [225]. In this work, training data were generated from LES calculations, the metrics of the flame topology and flow field were used as model inputs,

and the peak cylinder pressure was used as the model output. Their trained model could implicitly learn the relationships between the model inputs and outputs. Generally, ML can accurately determine the importance of in-cylinder flow features and subsequently establish the relationship between in-cylinder flow fields and engine performance.

In addition, ML can be used to study engine performance under various scenarios. For instance, 1) based on the ANN and SVM, the brake mean effective pressure, NOx, and soot formations were predicted in terms of fuel spray specifications such as crank angle and the liquid mass evaporated [226–228]. 2) Based on CNN models, combustion phases CA10 and CA50 (CA10, CA50: crank angle at 10% and 50% of fuel mass fraction burned) were predicted in an optical spark-ignited gasoline engine by providing early luminosity flame kernel images [229].

5.2.2. Combustion optimization

Inspired by the natural selection process, a GA is used to optimize problems through operations such as mutation, crossover, and selection. The GA is a subset of evolutionary algorithms. Previously, ML coupling with the GA method was used to optimize the engine performance [230, 231]. Based on the limitations (e.g., multiple local minima, large training data size, and slow learning speed) of traditional ANNs, some advanced ML methods have been proposed for engine optimization. For instance, a least-squares support vector machine (LS-SVM) model was trained to map engine control parameters to engine performance and used several optimization methods (particle swarm optimization (PSO), GA, etc.) to optimize parameters for engine control [230, 232]. The results showed that the idle-speed performance significantly improved after optimization. In addition, PSO is found to be more efficient than the GA method in terms of LS-SVMs. Similarly, a further study based on an ML model for biodiesel engine performance prediction [231] was performed. The proposed method optimized the biodiesel ratio to obtain fewer emissions and better fuel economy of the engine. Similar work was done by Bendu et al. [233], where an ANN model was trained to regress the input parameters to the output performance of an ethanol-fueled

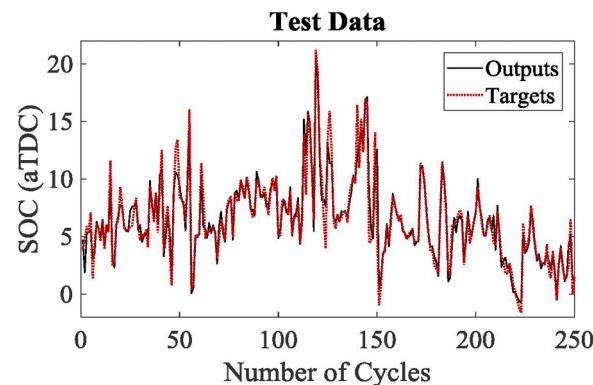


Fig. 22. Comparison of the experimental data of start of combustion and the predicted test data based on the MLP with optimization using GA [220].

HCCI engine, and then optimized the engine performance based on the generalized regression neural network coupled with PSO.

Based on the CFD method, the ANN and GA in conjunction were developed to optimize the performance of a heavy-duty engine [234]. In this work, an ANN regression model was first trained on CFD simulated data, whose samples were distributed over nine input parameters, such as initial in-cylinder chamber pressure, temperature, swirl flow, injection strategy, and engine performance and emissions related output parameters. The ANN built a map between the inputs and outputs. The GA was then used to optimize the input parameters for the ANN to achieve better engine performance while satisfying the emission constraints. Different training dataset sizes for their model were studied, and the results show that their model requires less training data than traditional methods to predict engine performance while maintaining sufficient accuracy, which is significant in making the process of engine performance optimization less computational. In addition, to solve the high computational cost of CFD, ten optimization strategies, including the micro-GA algorithm, PSO, differential evolution based on ML models with a SuperLearner, and a Gaussian process, were evaluated [235].

Generally, the present study confirmed that the hybrid method based on the combined ML and optimization methods provides a powerful tool to assist in engine optimization, such as compression ignition engine optimization [236], piston bowl geometry optimization [237], prediction of optimal combustion phasing in HCCI engines [238], and spark-ignition engine calibration [239]. Some studies using ML for engine control and optimization are briefly summarized in Table 3.

6. Concluding remarks, challenges and outlooks

6.1. Conclusions

This review presents the applications of ML techniques in combustion. With the boom in ML and continuing combustion research, the combination of these two fields has attracted renewed interest in the last several years. The present review addresses the applications of ML in chemical reduction, combustion kinetic model uncertainty quantification, subgrid modeling of combustion, reduced-order modeling of combustion, measurement and diagnosis of combustion, prediction and optimization of combustion engines, fuel design, etc. ML has a huge advantage in finding hidden patterns under large amounts of data, exploring and visualizing high-dimensional input spaces, deriving complex mapping from inputs and outputs, and reducing computational cost and memory occupation. For the combustion simulation, the relationship between the input parameters involving the temperature, species, and pressure, and the output parameters involving the chemical kinetic calculation or chemical source were established based on different ML methods. It is noteworthy that the training process of neural networks considering the flow-chemistry coupling with different turbulent scales can be obtained from the DNS database to further understand the turbulence-combustion interaction and establish robust and efficient turbulent combustion models. For the combustion diagnostics, the data-driven method based on ML is effective in the reconstruction and retrieval of flames, combustion diagnostics, time-sequence prediction, and flame regime classification. Fuel property prediction based on the neural network is more suitable for the predictive capability of physical properties (e.g., density, viscosity, heating value) and others (e.g., cetane or octane number, flash point, and ignition delay time) of fuel, and fuel design owing to its highly complex characteristics. Furthermore, the combination of ML methods and experimental engine research could help better predict engine performance based on alternative fuels, combustion state, effect of cycle-to-cycle variation and so on. The hybrid method based on the combined ML and optimization methods can also further optimize the engine-controlled parameters to achieve the best engine performance by mapping the relationship of engine-controlled parameters to engine performance.

Overall, while using ML in the combustion field, its powerful regression ability enables the transition of combustion models or research approaches from linear to nonlinear regimes, which could help to renew our insights into in-depth combustion science with the help of big data.

6.2. Existing challenges for ml in combustion

Although many successful studies on the combination between ML and combustion research have been accomplished, the framing of combustion problems in ML frameworks is a laborious task. While many studies have given us novel perspectives on combustion modeling and have drawn many interesting and valuable conclusions, some trained models are still lacks of generalizability and interpretability, and each narrow application needs to be specially trained, which makes it difficult to enrich the theory of combustion. For combustion modeling, the trained ML model needs to be cross-validated under different occasions could be convincing, if the model could be simplified to several plausible rules would be much exciting.

Due to the black box nature of ML model, it's hard to debug them to a satisfying state, except by adjusting the training data and training methods, this puts high demands on the preprocessing of training data. Take anomaly detection for example, due to the uncertainty and scarcity

Table 3

A brief summary of studies of using ML for engine control and optimization.

| Authors and Refs | Descriptions | Key words |
|----------------------|---|---|
| Agbulut, 2021 [240] | Employee ML methods to study the effects of the addition of nanoparticles of the fuel on the emission characteristics and performance of diesel engine. | Nanoparticle addition in fuel. |
| Badra, 2020 [241] | Use a ML-grid gradient ascent method to optimizing the operating conditions and the piston bowl design of a heavy-duty gasoline compression ignition engine. | Optimizing piston bowl design and operating conditions. |
| Wong, 2013 [231] | Utilize NNs and SVM to optimize biodiesel ratios to reduce emissions | Biodiesel ratios |
| Bendu, 2017 [233] | Adopt PSO model to optimizing control parameters, such as intake charge parameters, engine load, to achieve better engine performance | Operating parameters |
| Rahnama, 2020 [236] | Use ANNs and GA to optimizing start of injection timing and injection pressure of fuel for compression ignition engines. | Fuel injection parameters |
| Turkson, 2016 [239] | Giving a comprehensive review on using ANNs for engine calibration, which is used for defining engine input-output function parameters | Engine calibration |
| Cho, 2019 [242] | Building a deep ANN model to predict knock onset of a cycle, which gives insight into stochastic knocking combustion. | Knock onset prediction |
| Owoyele, 2021 [243] | Introducing a machine learning-based optimization algorithm for accelerating simulation-driven engine design, which could reduce fuel consumption maintain related constraints. | Optimization model for engine design |
| Kavuri, 2020 [244] | Combining Gaussian process and CFD simulation to reduce computational time of engine optimization | Reducing computational time |
| Shih, 2009 [245] | Introducing a reinforcement-learning-based output adaptive NN controller to control engine emissions. | Engine emission control |
| Irdmousa, 2019 [246] | Building a data-driven linear parameter-varying model to predict and control combustion phasing. | Control combustion phasing |

of the abnormal points of the combustion equipment, the ML model training will face the challenges of handling imbalanced distribution of normal and abnormal data, addressing the variations in abnormal behavior.

For the application of ML models on real engines, memory occupation and computational cost are important limitations considering the power of the control unit of engines. Besides, with the continuous aging of the engine in operation, it is also remains unclear whether the previously trained model needs to be re-calibrated.

6.3. The scope of ml in combustion

Although many successful studies on the combination between ML and combustion research have been accomplished, this research field is still in its early stages. Unlike physics-based modeling, a neural network does not understand Newton's law, or that the mass cannot be negative, there are no physical constraints with ML models, thus we don't know when and how it is violating physical laws. Therefore, developing physics-informed ML for combustion research will be critical for extending hybrid ML and combustion research from laboratory scale to large-scale industrial applications.

Different from the confirmatory analysis in which experiments are carefully designed to determine the validity of the hypothesis, ML is more like an exploratory analysis. In the case of a large amount of data, it's impossible to layout a finite set of testable hypotheses while facing hundreds of features. Meanwhile, ML algorithms are very suitable for exploratory predictive modeling, dimension reduction and classification with huge amount of data and high-dimensional feature space, like chemical reaction sources prediction, chemical composition reduction, combustion diagnostics and engine performance prediction. If the amount of data that can be used for analysis is small, it is more suitable to adopt the classic, multivariate statistical methods that are more informative than ML. Framing the combustion problems into ML framework is not an easy task. One of the most difficult steps in applying ML is to distill the problems to be researched into an ML issue, which requires a deep insight into the problem itself. Framing the problem into an ML framework is a prerequisite for the issue to be successfully solved using ML. Clarifying the combustion problem and carefully selecting and preprocessing the obtained data is vitally important. In addition, careful selection of the ML model, loss function, and training and tuning of the model are necessary components for building a predictive model.

6.4. Outlooks

ML can help solve many problems without showing us how they do it; it functions like a black or gray box; after the model is trained, the output can be obtained once given the input. However, transparency and interpretability are important for scientific research, improving the interpretability of the ML model, and using physics-informed ML could improve the generalization ability of the trained models. In different combustion systems, the operating conditions vary significantly. Embedding physics knowledge into the training of ML models and improving the interpretability of the trained model is a potential pathway to achieving more generalized models, which can be used with confidence under various conditions.

Building confidence of the trained model is very important for its widespread usage. Performance validation over the whole region of interest would be time-consuming and prohibitively expensive. More attempts, such as uncertainty quantification or sensitivity analysis, should be made to effectively verify the performance of the model, to ensure that it abides by the laws of physics and that it can accurately represent the simulated system, and thus promote its wider application.

We are in the age of big data for combustion research. Combustion studies and ML are rapidly advancing, whereas research related to the combination of ML and combustion is still in its infancy, and many research opportunities are yet to be explored. The combination of ML

and combustion will provide solutions to previous problems daunting in the combustion field and enhance our understanding of in-depth combustion mechanisms, helping us build a much cleaner future. The present review also calls for actions in the combustion community to embrace open and reproducible research works and share training data, training codes, and experiences. This will arouse the interest of researchers in related fields and driving progress. In addition, we should pay close attention to the latest developments in ML and find opportunities to introduce the latest ML techniques into the combustion field.

Declaration of interests

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

Acknowledgement

This work was supported by the National Science Fund for Distinguished Young Scholars (Grant No. 51825603), Tianjin Natural Science Foundation (Grant No. 20JCYBJC0110), and the National Natural Science Foundation of China (Grant No. 91741119). Special thanks to the valuable comments by the reviewers.

Reference

- [1] Perera ATD, Nik VM, Chen D, Scartezzini J-L, Hong T. Quantifying the impacts of climate change and extreme climate events on energy systems. *Nature Energy* 2020;5:150–9.
- [2] Kalogirou SA. Artificial intelligence for the modeling and control of combustion processes: a review. *Prog Energy Combust Sci* 2003;29:515–66.
- [3] Jordan MI, Mitchell TM. Machine learning: trends, perspectives, and prospects. *Science* 2015;349:255.
- [4] Brunton SL, Noack BR, Koumoutsakos P. Machine Learning for Fluid Mechanics. *Annu Rev Fluid Mech* 2020;52:477–508.
- [5] Duraisamy K, Iaccarino G, Xiao H. Turbulence Modeling in the Age of Data. *Annu Rev Fluid Mech* 2019;51:357–77.
- [6] Shrivastava N, Khan ZM. Application of soft computing in the field of internal combustion engines: a review. *Arch Comput Method Eng* 2018;25:707–26.
- [7] Raman V, Hassanaly M. Emerging trends in numerical simulations of combustion systems. In: Proceedings of the Combustion Institute. 37; 2019. p. 2073–89.
- [8] Zheng Z, Lin X, Yang M, He Z, Bao E, Zhang H, Tian Z. Progress in the Application of Machine Learning in Combustion Studies. *ES Energy Environ* 2020;9:1–14.
- [9] Ruiz F, Sepka S. New experiments and computations on the regenerative engine. *SAE International*; 1993. In: editor'editors.
- [10] Ferrenberg A. The single cylinder regenerated internal combustion engine. *SAE International*; 1990. In: editor'editors.
- [11] A. Subramaniam, M.L. Wong, R.D. Borker, S. Nimmagadda, S.K. Lele, Turbulence enrichment using physics-informed generative adversarial networks, 2020, pp. arXiv:2003.01907.
- [12] Loshchilov I, Hutter FJapa. Decoupled weight decay regularization. *Decoupled weight decay regularization* 2017.
- [13] Goodfellow I, Bengio Y, Courville A, Bengio Y. Deep learning. MIT press Cambridge; 2016.
- [14] Hastie T, Tibshirani R, Friedman J. The elements of statistical learning: data mining, inference, and prediction. Springer Science & Business Media; 2009.
- [15] Rumelhart DE, Hinton GE, Williams RJ. Learning representations by back-propagating errors. *Nature* 1986;323:533–6.
- [16] Aghashloho M, Peng W, Tabatabaei M, Kalogirou SA, Soltanian S, Hosseinzadeh-Bandabafha H, et al. Machine learning technology in biodiesel research: a review. *Prog Energy Combust Sci* 2021;85:100904.
- [17] Owoyele O, Kundu P, Ameen MM, Echekki T, Som S. Application of deep artificial neural networks to multi-dimensional flamelet libraries and spray flames. *Int J Engine Res* 2020;21:151–68.
- [18] An J, He G, Luo K, Qin F, Liu B. Artificial neural network based chemical mechanisms for computationally efficient modeling of hydrogen/carbon monoxide/kerosene combustion. *Int J Hydrogen Energy* 2020;45:29594–605.
- [19] Schmidhuber J. Deep learning in neural networks: an overview. *Neural Netw* 2015;61:85–117.
- [20] Lapeyre CJ, Misdariis A, Cazard N, Veynante D, Poinsot T. Training convolutional neural networks to estimate turbulent sub-grid scale reaction rates. *Combust Flame* 2019;203:255–64.
- [21] Nikolaou ZM, Chrysostomou C, Vervisch L, Cant S. Progress variable variance and filtered rate modelling using convolutional neural networks and flamelet methods. *Flow Turbul Combust* 2019;103:485–501.
- [22] Ranade R, Alqahtani S, Farooq A, Echekki T. An ANN based hybrid chemistry framework for complex fuels. *Fuel* 2019;241:625–36.

- [23] Hochreiter S, Schmidhuber J. Long short-term memory. *Neural Comput* 1997;9: 1735–80.
- [24] Huang J, Liu H, Cai W. Online in situ prediction of 3-D flame evolution from its history 2-D projections via deep learning. *J Fluid Mech* 2019;875:R2.
- [25] Suykens JA, Vandewalle J. Least squares support vector machine classifiers. *Neural processing letters* 1999;9:293–300.
- [26] Smola AJ, Schölkopf B. A tutorial on support vector regression. *Stat Comput* 2004;14:199–222.
- [27] Quinlan JR. Induction of decision trees. *Mach Learn* 1986;1:81–106.
- [28] Eddy SR. What is a hidden Markov model? *Nature Biotechnol* 2004;22:1315–6.
- [29] Quinonero-Candela J, Rasmussen CE. A unifying view of sparse approximate Gaussian process regression. *J Mach Learn Res* 2005;6:1939–59.
- [30] Polikar R. Ensemble learning, ensemble machine learning. Springer; 2012. p. 1–34.
- [31] Kaiser E, Noack BR, Cordier L, Spohn A, Segond M, Abel M, et al. Cluster-based reduced-order modelling of a mixing layer. *J Fluid Mech* 2014;754:365–414.
- [32] Kohonen T. Essentials of the self-organizing map. *Neural Netw* 2013;37:52–65.
- [33] Barwey S, Prakash S, Hassanaly M, Raman V. Data-driven classification and modeling of combustion regimes in detonation waves. *Flow Turbul Combust* 2020. <https://doi.org/10.1007/s10494-020-00176-4>.
- [34] Ranade R, Li G, Li S, Echekki T. An efficient machine-learning approach for PDF tabulation in turbulent combustion closure. *Combust Sci Technol* 2019;1–20. <https://doi.org/10.1080/00102202.2019.1686702>.
- [35] Wold S, Esbensen K, Geladi PJC. i.l. systems. Principal component analysis 1987; 2:37–52.
- [36] Schmid PJ. Dynamic mode decomposition of numerical and experimental data. *J Fluid Mech* 2010;656:5–28.
- [37] Golub GH, Reinsch C. Singular value decomposition and least squares solutions, linear algebra. Springer; 1971. p. 134–51.
- [38] Taira K, Brunton SL, Dawson ST, Rowley CW, Colonius T, McKeon BJ, et al. Modal analysis of fluid flows: an overview. *AIAA J* 2017;55:4013–41.
- [39] Rowley CW, Dawson ST. Model reduction for flow analysis and control. *Annu Rev Fluid Mech* 2017;49:387–417.
- [40] Yu J, Hesthaven JS. Flowfield reconstruction method using artificial neural network. *Aiaa J* 2019;57:482–98.
- [41] X. Zhu, AB.J.S.I.o.a.i. Goldberg, m. learning, Introduction to semi-supervised learning, 3 (2009) 1–130.
- [42] Aggarwal A, Mittal M, Battineni G. Generative adversarial network: an overview of theory and applications. *Int J Inf Manag Data Insight* 2021;1:100004.
- [43] Creswell A, White T, Dumoulin V, Arulkumaran K, Sengupta B, Bharath AA. Generative adversarial networks: an overview. *IEEE Signal Process Mag* 2018;35: 53–65.
- [44] Y. Li, Deep Reinforcement Learning: An Overview, 2017, pp. [arXiv:1701.07274](https://arxiv.org/abs/1701.07274).
- [45] B.R. Kiran, I. Sobh, V. Talpaert, P. Mannion, A.A. Al Sallab, S. Yogamani, P. Pérez, Deep Reinforcement Learning for Autonomous Driving: A Survey, 2020, pp. [arXiv:2002.00444](https://arxiv.org/abs/2002.00444).
- [46] Kober J, Peters J. Reinforcement learning in robotics: a survey. In: Kober J, Peters J, editors. Learning motor skills: from algorithms to robot experiments. Cham: Springer International Publishing; 2014. p. 9–67.
- [47] Silver D, Schrittwieser J, Simonyan K, Antonoglou I, Huang A, Guez A, Hubert T, Baker L, Lai M, Bolton A, Chen Y, Lillicrap T, Hui F, Sifre L, van den Driessche G, Graepel T, Hassabis D. Mastering the game of Go without human knowledge. *Nature* 2017;550:354–9.
- [48] Nian R, Liu J, Huang B. A review On reinforcement learning: introduction and applications in industrial process control. *Comput Chem Eng* 2020;139:106886.
- [49] Hochgreb S. Mind the gap: turbulent combustion model validation and future needs. In: Proceedings of the Combustion Institute. 37; 2019. p. 2091–107.
- [50] Poinsot T, Veynante D. Theoretical and numerical combustion. RT Edwards, Inc; 2005.
- [51] Kohse-Höninghaus K. Combustion in the future: the importance of chemistry. In: Proceedings of the Combustion Institute; 2020. <https://doi.org/10.1016/j.proci.2020.06.375>.
- [52] Curran HJ. Developing detailed chemical kinetic mechanisms for fuel combustion. In: Proceedings of the Combustion Institute. 37; 2019. p. 57–81.
- [53] Lam SH, Goussis DA. Understanding complex chemical kinetics with computational singular perturbation. In: Symposium (International) on Combustion. 22; 1989. p. 931–41.
- [54] Jones WP, Rigopoulos S. Rate-controlled constrained equilibrium: formulation and application to nonpremixed laminar flames. *Combust Flame* 2005;142: 223–34.
- [55] Lu T, Law CK. A directed relation graph method for mechanism reduction. In: Proceedings of the Combustion Institute. 30; 2005. p. 1333–41.
- [56] Ren Z, Goldin GM, Hiremath V, Pope SB. Reduced description of reactive flows with tabulation of chemistry. *Combust Theor Model* 2011;15:827–48.
- [57] Goldin GM, Ren Z, Zahirovic S. A cell agglomeration algorithm for accelerating detailed chemistry in CFD. *Combust Theory Modell* 2009;13:721–39.
- [58] Liang L, Stevens JG, Farrell JT. A dynamic multi-zone partitioning scheme for solving detailed chemical kinetics in reactive flow computations. *Combust Sci Technol*. 2009;181:1345–71.
- [59] Contino F, Jeanmart H, Lucchini T, D'Errico G. Coupling of in situ adaptive tabulation and dynamic adaptive chemistry: an effective method for solving combustion in engine simulations. *Proc Combust Inst* 2011;33:3057–64.
- [60] Maas U, Pope SB. Simplifying chemical kinetics: intrinsic low-dimensional manifolds in composition space. *Combust Flame* 1992;88:239–64.
- [61] Van Oijen J, De Goey L. Modelling of premixed laminar flames using flamelet-generated manifolds. *Combust Sci Technol* 2000;161:113–37.
- [62] Pierce CD, Moin P. Progress-variable approach for large-eddy simulation of non-premixed turbulent combustion. *J Fluid Mech* 2004;504:73–97.
- [63] Sen BA, Menon S. Linear eddy mixing based tabulation and artificial neural networks for large eddy simulations of turbulent flames. *Combust Flame* 2010; 157:62–74.
- [64] Sen BA, Hawkes ER, Menon S. Large eddy simulation of extinction and reignition with artificial neural networks based chemical kinetics. *Combust Flame* 2009; 157:566–78.
- [65] Chatzopoulos AK, Rigopoulos S. A chemistry tabulation approach via Rate-Controlled Constrained Equilibrium (RCCE) and Artificial Neural Networks (ANNs), with application to turbulent non-premixed CH₄/H₂/N₂ flames. In: Proceedings of the Combustion Institute. 34; 2013. p. 1465–73.
- [66] Franke LLC, Chatzopoulos AK, Rigopoulos S. Tabulation of combustion chemistry via Artificial Neural Networks (ANNs): methodology and application to LES-PDF simulation of Sydney flame L. *Combust Flame* 2017;185:245–60.
- [67] Kee RJ, Rupley FM, Miller JA. Chemkin-II: a fortran chemical kinetics package for the analysis of gas-phase chemical kinetics. Livermore, CAUnited States: Sandia National Lab.(SNL-CA); 1989.
- [68] Goodwin DG, Moffat H, Speth RL, Cantera: An Object-oriented Software Toolkit for Chemical Kinetics, Thermodynamics, and Transport Processes. Version 2.2.1, 2016.
- [69] P.J. Linstrom, W.G.J.J.O.C. Mallard, E. Data, The NIST Chemistry WebBook: a chemical data resource on the internet, 46 (2001) 1059–1063.
- [70] Garcia S, Luengo J, Herrera F. Data preprocessing in data mining. Springer; 2015.
- [71] Christo FC, Masri AR, Nebot EM, Pope SB. An integrated PDF/neural network approach for simulating turbulent reacting systems. In: Symposium (International) on Combustion. 26; 1996. p. 43–8.
- [72] Christo FC, Masri AR, Nebot EM. Artificial neural network implementation of chemistry with pdf simulation of H₂/CO₂ flames. *Combust Flame* 1996;106: 406–27.
- [73] Wan K, Barnaud C, Vervisch L, Domingo P. Chemistry reduction using machine learning trained from non-premixed micro-mixing modeling: application to DNS of a syngas turbulent oxy-flame with side-wall effects. *Combust Flame* 2020;220: 119–29.
- [74] Sharma AJ, Johnson RF, Kessler DA, Moses A. Deep Learning for Scalable Chemical Kinetics. *AIAA Scitech 2020 Forum* 2020.
- [75] W.Y. Peng, N.H. Pinkowski, Efficient and accurate time-integration of combustion chemical kinetics using artificial neural networks, (2017).
- [76] Blasco JA, Fueyo N, Dopazo C, Ballester J. Modelling the temporal evolution of a reduced combustion chemical system with an artificial neural network. *Combust Flame* 1998;113:38–52.
- [77] Blasco JA, Fueyo N, Larroja JC, Dopazo C, Chen YJ. A single-step time-integrator of a methane-air chemical system using artificial neural networks. *Comput Chem Eng* 1999;23:1127–33.
- [78] Wan K, Barnaud C, Vervisch L, Domingo P. Machine learning for detailed chemistry reduction in DNS of a syngas turbulent oxy-flame with side-wall effects. In: Proceedings of the Combustion Institute; 2020. <https://doi.org/10.1016/j.proci.2020.06.047>.
- [79] Blasco JA, Fueyo N, Dopazo C, Chen YJ. A self-organizing-map approach to chemistry representation in combustion applications. *Combust Theory Modell* 2000;4:61–76.
- [80] Sen BA, Menon S. Turbulent premixed flame modeling using artificial neural networks based chemical kinetics. In: Proceedings of the Combustion Institute. 32; 2009. p. 1605–11.
- [81] Zhang Y, Xu S, Zhong S, Bai X-S, Wang H, Yao M. Large eddy simulation of spray combustion using flamelet generated manifolds combined with artificial neural networks. *Energy and AI* 2020;2:100021.
- [82] Zhang J, Huang H, Xia Z, Ma L, Duan Y, Feng Y, Huang J. Artificial neural networks for chemistry representation in numerical simulation of the Flamelet-based models for turbulent combustion. *IEEE Access* 2020;8:80020–9.
- [83] S. Bhalla, M. Yao, J.-P. Hickey, M. Crowley. Compact representation of a multi-dimensional combustion manifold using deep neural networks. In: U. Brefeld, E. Fromont, A. Hotho, A. Knobbe, M. Maathuis, C. Robardet, editor*editors. Machine learning and knowledge discovery in databases; 2020, 2020; Cham: Springer International Publishing. p. 602–617.
- [84] Wang K, Xu R, Parise T, Shao J, Lee D, Movaghari A, Davidson D, Hanson R, Wang H, Bowman C. Combustion kinetics of conventional and alternative jet fuels using a hybrid chemistry (HyChem) approach. In: editor*editors. In: 10th US National Combustion Meeting; 2017. p. 23–6.
- [85] Ranade R, Alqahtani S, Farooq A, Echekki T. An extended hybrid chemistry framework for complex hydrocarbon fuels. *Fuel* 2019;251:276–84.
- [86] Alqahtani S, Echekki T. A data-based hybrid model for complex fuel chemistry acceleration at high temperatures. *Combust Flame* 2021;223:142–52.
- [87] M. Papadrakakis, G. Stavroulakis, A.J.C.M.i.A.M. Karatarakis, Engineering, A new era in scientific computing: Domain decomposition methods in hybrid CPU-GPU architectures, 200 (2011) 1490–508.
- [88] Barwey S, Raman V. A Neural network-inspired matrix formulation of chemical kinetics for acceleration on GPUs. *Energies* 2021;14.
- [89] Barwey S, Raman V. A neural network inspired formulation of chemical kinetics. arXiv preprint 2020. [arXiv:2008.08483](https://arxiv.org/abs/2008.08483).
- [90] K. Buchheit, O. Owoyele, T. Jordan, D.J.a.e.-p. Van Essendelft, The stabilized explicit variable-load solver with machine learning acceleration for the rapid solution of stiff chemical kinetics, 2019, pp. [arXiv:1905.09395](https://arxiv.org/abs/1905.09395).
- [91] Kempf A, Flemming F, Janicka J. Investigation of lengthscales, scalar dissipation, and flame orientation in a piloted diffusion flame by LES. In: Proceedings of the Combustion Institute. 30; 2005. p. 557–65.

- [92] Ihme M, Schmitt C, Pitsch H. Optimal artificial neural networks and tabulation methods for chemistry representation in LES of a bluff-body swirl-stabilized flame. In: Proceedings of the Combustion Institute. 32; 2009. p. 1527–35.
- [93] Emami MD, Fard AEshghinejad. Laminar flamelet modeling of a turbulent CH₄/H₂/N₂ jet diffusion flame using artificial neural networks. *Appl Math Model* 2012;36:2082–93.
- [94] Shadram Z, Nguyen TM, Sideris A, Sirignano WA. Neural network closure models for estimating flame variables in a liquid-propellant rocket engine. *AIAA Scitech 2019 Forum* 2019.
- [95] Shadram Z, Nguyen TM, Sideris A, Sirignano WA. Neural network flame closure for a turbulent combustor with unsteady pressure. *arXiv preprint* 2020. [arXiv:2005.14167](https://arxiv.org/abs/2005.14167).
- [96] Hansinger M, Ge Y, Pfitzner M. Deep residual networks for flamelet/progress variable tabulation with application to a piloted flame with inhomogeneous inlet. *Combust Sci Technol* 2020;1–27. <https://doi.org/10.1080/00102202.2020.1822826>.
- [97] Owoyele O, Kundu P, Pal P. Efficient bifurcation and tabulation of multi-dimensional combustion manifolds using deep mixture of experts: an a priori study. In: Proceedings of the Combustion Institute; 2020. <https://doi.org/10.1016/j.proci.2020.09.006>.
- [98] Pulga L, Bianchi GM, Falfari S, Forte C. A machine learning methodology for improving the accuracy of laminar flame simulations with reduced chemical kinetics mechanisms. *Combust Flame* 2020;216:72–81.
- [99] Malik MR, Obando Vega P, Coussement A, Parente A. Combustion modeling using Principal Component Analysis: a posteriori validation on Sandia flames D, E and F. In: Proceedings of the Combustion Institute; 2020. <https://doi.org/10.1016/j.proci.2020.07.014>.
- [100] Sutherland JC, Parente A. Combustion modeling using principal component analysis. In: Proceedings of the Combustion Institute. 32; 2009. p. 1563–70.
- [101] Isaac BJ, Thorneck JN, Sutherland J, Smith PJ, Parente A. Advanced regression methods for combustion modelling using principal components. *Combust Flame* 2015;162:2592–601.
- [102] Mirgolbabaei H, Echebki T. The reconstruction of thermo-chemical scalars in combustion from a reduced set of their principal components. *Combust Flame* 2015;162:1650–2.
- [103] Ranade R, Echebki T. A framework for data-based turbulent combustion closure: a posteriori validation. *Combust Flame* 2019;210:279–91.
- [104] Dalakoti DK, Wehrfritz A, Savard B, Day MS, Bell JB, Hawkes ER. An a priori evaluation of a principal component and artificial neural network based combustion model in diesel engine conditions. In: Proceedings of the Combustion Institute; 2020. <https://doi.org/10.1016/j.proci.2020.06.263>.
- [105] Benner P, Gugercin S, Willcox K. A survey of projection-based model reduction methods for parametric dynamical systems. *SIAM Rev* 2015;57:483–531.
- [106] Chatterjee A. An introduction to the proper orthogonal decomposition. *Curr Sc* 2000;78:808–17.
- [107] Hesthaven JS, Stamm B, Zhang S. Efficient greedy algorithms for high-dimensional parameter spaces with applications to empirical interpolation and reduced basis methods*. *ESAIM: M2AN* 2014;48:259–83.
- [108] Hesthaven JS, Ubbiali S. Non-intrusive reduced order modeling of nonlinear problems using neural networks. *J Comput Phys* 2018;363:55–78.
- [109] Wang Q, Hesthaven JS, Ray D. Non-intrusive reduced order modeling of unsteady flows using artificial neural networks with application to a combustion problem. *J Comput Phys* 2019;384:289–307.
- [110] Zhang Y, Zhou L, Meng H, Teng H. Reconstructing cellular surface of gaseous detonation based on artificial neural network and proper orthogonal decomposition. *Combust Flame* 2020;212:156–64.
- [111] Wang H, Sheen DA. Combustion kinetic model uncertainty quantification, propagation and minimization. *Prog Energy Combust Sci* 2015;47:1–31.
- [112] Gelman A, Carlin JB, Stern HS, Dunson DB, Vehtari A, Rubin DB. Bayesian data analysis. CRC press; 2013.
- [113] Hastings WK. Monte carlo sampling methods using Markov chains and their applications. *Biometrika* 1970;57:97–109.
- [114] Tripathy RK, Bilionis I, Deep UQ. Learning deep neural network surrogate models for high dimensional uncertainty quantification. *J Comput Phys* 2018;375:565–88.
- [115] Freitas RSM, Rochinha FA, Mira D, Jiang X. Parametric and model uncertainties induced by reduced order chemical mechanisms for biogas combustion. *Chem Eng Sci* 2020;227:115949.
- [116] Wang J, Zhou Z, Lin K, Law CK, Yang B. Facilitating Bayesian analysis of combustion kinetic models with artificial neural network. *Combust Flame* 2020;213:87–97.
- [117] Li S, Yang B, Qi F. Accelerate global sensitivity analysis using artificial neural network algorithm: case studies for combustion kinetic model. *Combust Flame* 2016;168:53–64.
- [118] An J, He G, Qin F, Li R, Huang Z. A new framework of global sensitivity analysis for the chemical kinetic model using PSO-BPNN. *Comput Chem Eng* 2018;112:154–64.
- [119] Ji W, Deng S. Autonomous Discovery of Unknown Reaction Pathways from Data by Chemical Reaction Neural Network. *J Phys Chem A* 2021;125:1082–92.
- [120] Car R, Parrinello M. Unified approach for molecular dynamics and density-functional theory. *Phys. Rev. Lett.* 1985;55:2471.
- [121] Bertels LW, Newcomb LB, Alaghemandi M, Green JR, Head-Gordon M. Benchmarking the performance of the ReaxFF reactive force field on hydrogen combustion systems. *J Phys Chem A* 2020;124:5631–45.
- [122] Zeng J, Cao L, Xu M, Zhu T, Zhang JZH. Complex reaction processes in combustion unraveled by neural network-based molecular dynamics simulation. *Nat Commun* 2020;11:5713.
- [123] Pope SB. Small scales, many species and the manifold challenges of turbulent combustion. *Proc Combust Inst* 2013;34:1–31.
- [124] Chen ZX, Iavarone S, Ghiasi G, Kannan V, D'Alessio G, Parente A, et al. Application of machine learning for filtered density function closure in MILD combustion. *Combust Flame* 2021;225:160–79.
- [125] Henry de Frahan MT, Yellapantula S, King R, Day MS, GROUT RW. Deep learning for presumed probability density function models. *Combust Flame* 2019;208:436–50.
- [126] Durbin PA. Some recent developments in turbulence closure modeling. *Annu Rev Fluid Mech* 2018;50:77–103.
- [127] van Oijen JA, Domini A, Bastiaans RJM, ten Thije Boonkamp JHM, de Goey LPH. State-of-the-art in premixed combustion modeling using flamelet generated manifolds. *Prog Energy Combust Sci* 2016;57:30–74.
- [128] Cook AW, Riley JJ. A subgrid model for equilibrium chemistry in turbulent flows. *Phys Fluids* 1994;6:2868–70.
- [129] Seltz A, Domingo P, Vervisch L, Nikolaou ZM. Direct mapping from LES resolved scales to filtered-flame generated manifolds using convolutional neural networks. *Combust Flame* 2019;210:71–82.
- [130] Yellapantula S, Perry BA, Grout RW. Deep learning-based model for progress variable dissipation rate in turbulent premixed flames. In: Proceedings of the Combustion Institute; 2020. <https://doi.org/10.1016/j.proci.2020.06.205>.
- [131] Yao S, Wang B, Kronenburg A, Stein OT. Conditional scalar dissipation rate modeling for turbulent spray flames using artificial neural networks. In: Proceedings of the Combustion Institute; 2020. <https://doi.org/10.1016/j.proci.2020.06.135>.
- [132] Yao S, Wang B, Kronenburg A, Stein OT. Modeling of sub-grid conditional mixing statistics in turbulent sprays using machine learning methods. *Phys Fluids* 2020;32:115124.
- [133] Moin P, Squires K, Cabot W, Lee S. A dynamic subgrid-scale model for compressible turbulence and scalar transport. *Phys Fluids A* 1991;3:2746–57.
- [134] J. Bardino, J.H. Ferziger, W.C. Reynolds, Improved turbulence models based on large eddy simulation of homogeneous, incompressible turbulent flows, (1983).
- [135] Prat A, Sautory T, Navarro-Martinez S. A Priori Sub-grid Modelling Using Artificial Neural Networks. *Int J Comput Fluid Dyn* 2020;34:397–417.
- [136] Xie Chenyue, Yuan Zelong, Wang Jianchun. Artificial neural network-based nonlinear algebraic models for large eddy simulation of turbulence. *Phys Fluids* 2020;32:115101.
- [137] Wang Z, Luo K, Li D, Tan J, Fan J. Investigations of data-driven closure for subgrid-scale stress in large-eddy simulation. *Phys Fluids* 2018;30:125101.
- [138] Nikolaou ZM, Chrysostomou C, Minamoto Y, Vervisch L. Evaluation of a neural network-based closure for the unresolved stresses in turbulent premixed V-Flames. *Flow Turbul Combust* 2020. <https://doi.org/10.1007/s10494-020-0170-w>.
- [139] Schoeppelein M, Weatheritt J, Sandberg R, Talei M, Klein M. Application of an evolutionary algorithm to LES modelling of turbulent transport in premixed flames. *J Comput Phys* 2018;374:1166–79.
- [140] Koza JR. Genetic programming as a means for programming computers by natural selection. *Stat Comput* 1994;4:87–112.
- [141] Weatheritt J, Sandberg R. A novel evolutionary algorithm applied to algebraic modifications of the RANS stress–strain relationship. *J Comput Phys* 2016;325:22–37.
- [142] Udrescu S-M, Tegmark M, Feynman AI. A physics-inspired method for symbolic regression. *Sci Adv* 2020;6:eaay2631.
- [143] Ronneberger O, Fischer P, Brox T. U-Net: Convolutional Networks for Biomedical Image Segmentation 2015. [arXiv:1505.04597](https://arxiv.org/abs/1505.04597).
- [144] Keys R. Cubic convolution interpolation for digital image processing. *IEEE Trans Acoust* 1981;29:1153–60.
- [145] Dong C, Loy CC, He K, Tang X. Image super-resolution using deep convolutional networks. *IEEE Trans Pattern Anal Mach Intell* 2015;38:295–307.
- [146] Fukami K, Fukagata K, Taira K. Super-resolution reconstruction of turbulent flows with machine learning. *Journal of Fluid Mechanics/J Fluid Mech* 2019;870:106–20.
- [147] Bode M, Gauding M, Lian Z, Denker D, Davidovic M, Kleinheinz K, Jitsev J, Pitsch H. Using physics-informed super-resolution generative adversarial networks for subgrid modeling in turbulent reactive flows. *arXiv preprint* 2019. [arXiv:1911.11380](https://arxiv.org/abs/1911.11380).
- [148] Raissi M, Perdikaris P, Karniadakis GE. Physics-informed neural networks: a deep learning framework for solving forward and inverse problems involving nonlinear partial differential equations. *J Comput Phys* 2019;378:686–707.
- [149] A. Creswell, T. White, V. Dumoulin, K. Arulkumaran, B. Sengupta, A.A.J.I.S.P.M. Bharath, Generative adversarial networks: an overview, 35 (2018) 53–65.
- [150] Jin X, Cheng P, Chen W-L, Li H. Prediction model of velocity field around circular cylinder over various Reynolds numbers by fusion convolutional neural networks based on pressure on the cylinder. *Phys Fluids* 2018;30:047105.
- [151] Sekar V, QinghuaJiang ChangShu, Khoo BC. Fast flow field prediction over airfoils using deep learning approach. *Phys Fluids* 2019;31:057103.
- [152] An J, Wang H, Liu B, Luo KH, Qin F, He GQ. A deep learning framework for hydrogen-fueled turbulent combustion simulation. *Int J Hydrogen Energy* 2020;45:17992–8000.
- [153] Floyd J, Geipel P, Kempf AM. Computed Tomography of Chemiluminescence (CTC): instantaneous 3D measurements and Phantom studies of a turbulent opposed jet flame. *Combust Flame* 2011;158:376–91.

- [154] Wang J, Song Y, Li Z-h, Kempf A, He A-z. Multi-directional 3D flame chemiluminescence tomography based on lens imaging. *Opt. Lett.* 2015;40: 1231–4.
- [155] Yu T, Cai W, Liu Y. Rapid tomographic reconstruction based on machine learning for time-resolved combustion diagnostics. *Rev Sci Instrum* 2018;89:043101.
- [156] Huang J, Liu H, Dai J, Cai W. Reconstruction for limited-data nonlinear tomographic absorption spectroscopy via deep learning. *J Quant Spectrosc Radiat Transfer* 2018;218:187–93.
- [157] Huang J, Liu H, Wang Q, Cai W. Limited-projection volumetric tomography for time-resolved turbulent combustion diagnostics via deep learning. *Aerospace Sci Technol* 2020;106:106123.
- [158] Deng A, Huang J, Liu H, Cai W. Deep learning algorithms for temperature field reconstruction of nonlinear tomographic absorption spectroscopy. *Measurement: Sensors* 2020;10:12:100024.
- [159] Wei C, Schwarm KK, Pineda DI, Spearin RM. Deep neural network inversion for 3D laser absorption imaging of methane in reacting flows. *Opt. Lett.* 2020;45: 2447–50.
- [160] Jin Y, Zhang W, Song Y, Qu X, Li Z, Ji Y, He A. Three-dimensional rapid flame chemiluminescence tomography via deep learning. *Opt. Express* 2019;27: 27308–34.
- [161] Jiang Y, Si J, Zhang R, Enemali G, Zhou B, McCann H, et al. CSTNet: a dual-branch convolutional network for imaging of reactive flows using chemical species tomography. *arXiv preprint 2020*. arXiv:2010.03868.
- [162] Dreyer JA, Slavchov RI, Rees EJ, Akroyd J, Salamanca M, Mosbach S, et al. Improved methodology for performing the inverse Abel transform of flame images for color ratio pyrometry. *Appl Opt* 2019;58:2662–70.
- [163] Rodríguez A, Escudero F, Cruz JJ, Carvajal G, Fuentes A. Retrieving soot volume fraction fields for laminar axisymmetric diffusion flames using convolutional neural networks. *Fuel* 2021;285:119011.
- [164] Ren T, Modest MF, Fateev A, Sutton G, Zhao W, Rusu F. Machine learning applied to retrieval of temperature and concentration distributions from infrared emission measurements. *Appl Energy* 2019;252:113448.
- [165] Butz D, Hartl S, Popp S, Walther S, Barlow RS, Hasse C, et al. Local flame structure analysis in turbulent CH₄/air flames with multi-regime characteristics. *Combust Flame* 2019;210:426–38.
- [166] Wan K, Hartl S, Vervisch L, Domingo P, Barlow RS, Hasse C. Combustion regime identification from machine learning trained by Raman/Rayleigh line measurements. *Combust Flame* 2020;219:268–74.
- [167] Debus C, Ruettgers A, Petrarolo A, Kobald M, Sigel M. High-performance data analytics of hybrid rocket fuel combustion data using different machine learning approaches. *AIAA Scitech 2020 Forum* 2020.
- [168] Zhu X, Cai Z, Wu J, Cheng Y, Huang Q. Convolutional neural network based combustion mode classification for condition monitoring in the supersonic combustor. *Acta Astronaut* 2019;159:349–57.
- [169] Petrarolo A, Ruettgers A, Kobald M. Data Clustering of Hybrid Rocket Combustion Flame. *AIAA Propulsion and Energy 2019*. Forum.
- [170] Rüttgers A, Petrarolo A, Kobald M. Clustering of paraffin-based hybrid rocket fuels combustion data. *Exp Fluids* 2019;61:4.
- [171] Pino J, Garcés HO, Cuevas J, Arias LE, Rojas AJ, Fuentes A. Soot propensity by image magnification and artificial intelligence. *Fuel* 2018;225:256–65.
- [172] Han Z, Li J, Zhang B, Hossain MM, Xu C. Prediction of combustion state through a semi-supervised learning model and flame imaging. *Fuel* 2020;119745. <https://doi.org/10.1016/j.fuel.2020.119745>.
- [173] Chung WT, Mishra AA, Perakis N, Ihme M. Data-assisted combustion simulations with dynamic submodel assignment using random forests. *Combust. Flame* 2021; 227:172–85.
- [174] Choi O, Choi J, Kim N, Lee MC. Combustion instability monitoring through deep-learning-based classification of sequential high-speed flame images. *Electronics (Basel)* 2020;9:848.
- [175] He K, Zhang X, Ren S, Sun J. Deep residual learning for image recognition. In: editor“editors. In: Proceedings of the IEEE conference on computer vision and pattern recognition; 2016. p. 770–8.
- [176] Gangopadhyay T, Locurto A, Boor P, Michael JB, Sarkar S. Characterizing Combustion Instability Using Deep Convolutional Neural Network. In: editor“editors. In: *ASME 2018 Dynamic Systems and Control Conference*; 2018.
- [177] Omiotek Z, Smolarz A. Combustion process monitoring based on flame intensity time series. In: Proceedings of the Institution of Mechanical Engineers, Part I: Journal of Systems and Control Engineering; 2020, 0959651820965475. <https://doi.org/10.1177/0959651820965475>.
- [178] Wang Z, Song C, Chen T. Deep learning based monitoring of furnace combustion state and measurement of heat release rate. *Energy* 2017;131:106–12.
- [179] Han Z, Hossain MM, Wang Y, Li J, Xu C. Combustion stability monitoring through flame imaging and stacked sparse autoencoder based deep neural network. *Appl Energy* 2020;259:114159.
- [180] Akintayo A, Lore KG, Sarkar S, Sarkar S. Prognostics of combustion instabilities from hi-speed flame video using a deep convolutional selective autoencoder. *Int J Prognostic Health Manag* 2016;7:1–14.
- [181] Chen J, Hsu T-Y, Chen C-C, Cheng Y-C. Monitoring combustion systems using HMM probabilistic reasoning in dynamic flame images. *Appl Energy* 2010;87: 2169–79.
- [182] Qiu T, Liu M, Zhou G, Wang L, Gao K. An unsupervised classification method for flame image of pulverized coal combustion based on convolutional auto-encoder and hidden Markov model. *Energies* 2019;12:2585.
- [183] Liu Y, Fan Y, Chen J. Flame Images for Oxygen Content Prediction of Combustion Systems Using DBN. *Energy Fuels* 2017;31:8776–83.
- [184] Cai W, Kaminski CF. Tomographic absorption spectroscopy for the study of gas dynamics and reactive flows. *Prog Energy Combust Sci* 2017;59:1–31.
- [185] Cai W, Liu H, Huang J, Zhang J. Reconstruction of kHz-rate 3-D flame image sequences from a low-rate 2-D recording via a data-driven approach. *J. Opt. Soc. Am. B* 2020;37:3564–71.
- [186] Xu W, Luo W, Wang Y, You Y. Data-driven three-dimensional super-resolution imaging of a turbulent jet flame using a generative adversarial network. *Appl Opt* 2020;59:5729–36.
- [187] Zhao F, Ruan Z, Yue Z, Hung DLS, Som S, Xu M. Time-sequenced flow field prediction in an optical spark-ignition direct-injection engine using bidirectional recurrent neural network (bi-RNN) with long short-term memory. *Appl Therm Eng* 2020;173:115253.
- [188] Barwey S, Hassanalay M, Raman V, Steinberg A. Using machine learning to construct velocity fields from OH-PLIF images. *Combust Sci Technol* 2019;1–24. <https://doi.org/10.1080/00102202.2019.1678379>.
- [189] Barwey S, Raman V, Steinberg A. Extracting information overlap in simultaneous OH-PLIF and PIV fields with neural networks. In: *Proceedings of the Combustion Institute*; 2020. <https://doi.org/10.1016/j.proci.2020.06.180>.
- [190] Meng X, Jia M, Wang T. Neural network prediction of biodiesel kinematic viscosity at 313K. *Fuel* 2014;121:133–40.
- [191] Saldana DA, Starck L, Mougin P, Rousseau B, Ferrando N, Creton B. Prediction of Density and Viscosity of Biofuel Compounds Using Machine Learning Methods. *Energy Fuels* 2012;26:2416–26.
- [192] Piloto-Rodríguez R, Sánchez-Borroto Y, Lapuerta M, Goyos-Pérez L, Verhelst S. Prediction of the cetane number of biodiesel using artificial neural networks and multiple linear regression. *Energy Convers Manage* 2013;65:255–61.
- [193] Hosseinpour S, Aghbashlo M, Tabatabaei M, Khalife E. Exact estimation of biodiesel cetane number (CN) from its fatty acid methyl esters (FAMEs) profile using partial least square (PLS) adapted by artificial neural network (ANN). *Energy Convers Manage* 2016;124:389–98.
- [194] Alves JCL, Henriques CB, Poppi RJ. Determination of diesel quality parameters using support vector regression and near infrared spectroscopy for an in-line blending optimizer system. *Fuel* 2012;97:710–7.
- [195] Rocabruno-Valdés CI, Ramírez-Verduzco LF, Hernández JA. Artificial neural network models to predict density, dynamic viscosity, and cetane number of biodiesel. *Fuel* 2015;147:9–17.
- [196] Saldana DA, Starck L, Mougin P, Rousseau B, Pidol L, Jeuland N, Creton B. Flash Point and Cetane Number Predictions for Fuel Compounds Using Quantitative Structure Property Relationship (QSPR) Methods. *Energy Fuels* 2011;25:3900–8.
- [197] Saldana DA, Starck L, Mougin P, Rousseau B, Creton B. On the rational formulation of alternative fuels: melting point and net heat of combustion predictions for fuel compounds using machine learning methods. *SAR QSAR Environ Res* 2013;24:259–77.
- [198] Kessler T, Sacia ER, Bell AT, Mack JH. Artificial neural network based predictions of cetane number for furanic biofuel additives. *Fuel* 2017;206:171–9.
- [199] Sun X, Krakauer NJ, Politowicz A, Chen W-T, Li Q, Li Z, et al. Assessing Graph-based Deep Learning Models for Predicting Flash Point. *Mol Inform* 2020;39: 1900101.
- [200] Li R, Herreros JM, Tsolakis A, Yang W. Machine learning regression based group contribution method for cetane and octane numbers prediction of pure fuel compounds and mixtures. *Fuel* 2020;280:118589.
- [201] Choi Y, Chen JY. Fast prediction of start-of-combustion in HCCI with combined artificial neural networks and ignition delay model. In: *Proceedings of the Combustion Institute*. 30; 2005. p. 2711–8.
- [202] Zhao Z, Chen Z, Chen S. Correlations for the ignition delay times of hydrogen/air mixtures. *Chin Sci Bull* 2011;56:215–21.
- [203] N. Shah, P. Zhao, D. DelVescovo, H. Ge, Prediction of autoignition and flame properties for multicomponent fuels using machine learning techniques, Report No. 0148-7191, SAE Technical Paper, 2019.
- [204] Han W, Sun Z, Scholtissek A, Hasse C. Machine Learning of ignition delay times under dual-fuel engine conditions. *Fuel* 2020;119650. <https://doi.org/10.1016/j.fuel.2020.119650>.
- [205] Cui Y, Wang Q, Liu H, Zheng Z, Wang H, Yue Z, Yao M. Development of the ignition delay prediction model of n-butane/hydrogen mixtures based on artificial neural network. *Energy and AI* 2020;2:100033.
- [206] Druzgalski CL, Lapointe S, Whitesides R, McEneny MJ. Predicting octane number from microscale flame dynamics. *Combust Flame* 2019;208:5–14.
- [207] Li G, Hu Z, Hou F, Li X, Wang L, Zhang X. Machine learning enabled high-throughput screening of hydrocarbon molecules for the design of next generation fuels. *Fuel* 2020;265:116968.
- [208] Yusri IM, Abdul Majeed APP, Mamat R, Ghazali MF, Awad OI, Azmi WH. A review on the application of response surface method and artificial neural network in engine performance and exhaust emissions characteristics in alternative fuel. *Renew Sustain Energy Rev* 2018;90:665–86.
- [209] Najafi G, Ghobadian B, Tavakoli T, Buttsworth DR, Yusaf TF, Faizollahnejad M. Performance and exhaust emissions of a gasoline engine with ethanol blended gasoline fuels using artificial neural network. *Appl Energy* 2009;86:630–9.
- [210] Çay Y, Korkmaz I, Çiçek A, Kara F. Prediction of engine performance and exhaust emissions for gasoline and methanol using artificial neural network. *Energy* 2013; 50:177–86.
- [211] Liu J, Ullishney C, Dumitrescu CE. Random Forest Machine Learning Model for Predicting Combustion Feedback Information of a Natural Gas Spark Ignition Engine. *J. Energy Resour Technol* 2020;143.
- [212] Sharon H, Jayaprakash R, Karthigai selvan M, Soban kumar DR, Sundaresan A, Karuppasamy K. Biodiesel production and prediction of engine performance using SIMULINK model of trained neural network. *Fuel* 2012;99:197–203.

- [215] Kenanoğlu R, Baltacıoğlu MK, Demir MH, Özdemir M Erkinay. Performance & emission analysis of HHO enriched dual-fuelled diesel engine with artificial neural network prediction approaches. *Int J Hydrogen Energy* 2020;45: 26357–69.
- [216] Kapusuz M, Ozcan H, Yamin JA. Research of performance on a spark ignition engine fueled by alcohol-gasoline blends using artificial neural networks. *Appl Therm Eng* 2015;91:525–34.
- [217] Syed J, Baig RU, Algarni S, Murthy YVVS, Masood M, Inamurrahman M. Artificial Neural Network modeling of a hydrogen dual fueled diesel engine characteristics: an experiment approach. *Int J Hydrogen Energy* 2017;42:14750–74.
- [218] Silitonga AS, Masjuki HH, Ong HC, Sebayang AH, Dharma S, Kusumo F, Siswantoro J, Milano J, Daud K, Mahlia TMI, Chen W-H, Sugiyanto B. Evaluation of the engine performance and exhaust emissions of biodiesel-bioethanol-diesel blends using kernel-based extreme learning machine. *Energy* 2018;159:1075–87.
- [219] Rezaei J, Shahbakhti M, Bahri B, Aziz AA. Performance prediction of HCCI engines with oxygenated fuels using artificial neural networks. *Appl Energy* 2015; 138:460–73.
- [220] Taghavi M, Gharehgani A, Nejad FB, Mirsalim M. Developing a model to predict the start of combustion in HCCI engine using ANN-GA approach. *Energy Convers Manage* 2019;195:57–69.
- [221] Angikath Shamsudheen F, Yalamanchi K, Yoo KH, Voice A, Boehman A, Sarathy M. Machine learning techniques for classification of combustion events under homogeneous charge compression ignition (HCCI) conditions. *SAE International*; 2020.
- [222] Bode J, Schorr J, Krüger C, Dreizler A, Böhm B. Influence of the in-cylinder flow on cycle-to-cycle variations in lean combustion DISI engines measured by high-speed scanning-PIV. In: Proceedings of the Combustion Institute. 37; 2019. p. 4929–36.
- [223] Di Mauro A, Chen H, Sick V. Neural network prediction of cycle-to-cycle power variability in a spark-ignited internal combustion engine. In: Proceedings of the Combustion Institute. 37; 2019. p. 4937–44.
- [224] Hanuschkin A, Schober S, Bode J, Schorr J, Böhm B, Krüger C, Peters S. Machine learning-based analysis of in-cylinder flow fields to predict combustion engine performance. *Int J Engine Res* 2019;1468087419833269.
- [225] Kodavasal J, Abdul Moiz A, Ameen M, Som S. Using machine learning to analyze factors determining Cycle-to-Cycle variation in a spark-ignited gasoline engine. *J Energy Resour Technol* 2018;140.
- [226] Taghavifar H, Taghavifar H, Mardani A, Mohebbi A. Exhaust emissions prognostication for DI diesel group-hole injectors using a supervised artificial neural network approach. *Fuel* 2014;125:81–9.
- [227] Cocco Mariani V, Hennings Och S, dos Santos Coelho L, Domingues E. Pressure prediction of a spark ignition single cylinder engine using optimized extreme learning machine models. *Appl Energy* 2019;249:204–21.
- [228] Norouzi A, Aliramezani M, Koch CR. A correlation-based model order reduction approach for a diesel engine NOx and brake mean effective pressure dynamic model using machine learning. *Int J Engine Res* 2020. <https://doi.org/10.1177/1468087420936949>.
- [229] Johnson R, Kaczynski D, Zeng W, Warey A, Grover R, Keum S. Prediction of combustion phasing using deep convolutional neural networks. *SAE International*; 2020.
- [230] Wong P, Tam L, Li K, Vong C. Engine idle-speed system modelling and control optimization using artificial intelligence. In: Proceedings of the Institution of Mechanical Engineers, Part D: Journal of Automobile Engineering. 224; 2010. p. 55–72.
- [231] Wong KI, Wong PK, Cheung CS, Vong CM. Modeling and optimization of biodiesel engine performance using advanced machine learning methods. *Energy* 2013;55: 519–28.
- [232] Wong PK, Tam LM, Ke L. Automotive engine power performance tuning under numerical and nominal data. *Control Eng Pract* 2012;20:300–14.
- [233] Bendu H, Deepak BBVL, Murugan S. Multi-objective optimization of ethanol fuelled HCCI engine performance using hybrid GRNN-PSO. *Appl Energy* 2017; 187:601–11.
- [234] Moiz AA, Pal P, Probst D, Pei Y, Zhang Y, Som S, Kodavasal J. A machine learning-genetic algorithm (ML-GA) approach for rapid optimization using high-performance computing. *SAE Int J Commer Veh* 2018;11:291–306.
- [235] Probst DM, Raju M, Senecal PK, Kodavasal J, Pal P, Som S, et al. Evaluating Optimization Strategies for Engine Simulations Using Machine Learning Emulators. *J Eng Gas Turbine Power* 2019;141.
- [236] Rahnama P, Arab M, Reitz RD. A time-saving methodology for optimizing a compression ignition engine to reduce fuel consumption through machine learning. *SAE International*; 2020.
- [237] Badra J, khaled F, Sim J, Pei Y, Viollet Y, Pal P, Futterer C, Brenner M, Son S, Farooq A, Chang J. Combustion system optimization of a light-duty GCI engine using CFD and machine learning. *SAE International*; 2020.
- [238] Nazoktabar M, Jazayeri SA, Parsa M, Ganji DD, Arshabbar K. Controlling the optimal combustion phasing in an HCCI engine based on load demand and minimum emissions. *Energy* 2019;182:82–92.
- [239] Turkson RF, Yan F, Ali MKA, Hu J. Artificial neural network applications in the calibration of spark-ignition engines: an overview. *Engineering Science and Technology, an International Journal* 2016;19:1346–59.
- [240] Ağbulut Ü, Gürel AE, Saridemir S. Experimental investigation and prediction of performance and emission responses of a CI engine fuelled with different metal-oxide based nanoparticles-diesel blends using different machine learning algorithms. *Energy* 2021;215:119076.
- [241] Badra JA, Khaled F, Tang M, Pei Y, Kodavasal J, Pal P, et al. Engine combustion system optimization using computational fluid dynamics and machine learning: a methodological approach. *J Energy Resour Technol* 2020;143.
- [242] Cho S, Park J, Song C, Oh S, Lee S, Kim M, Min K. Prediction modeling and analysis of knocking combustion using an improved OD RGF model and supervised deep learning. *Energies* 2019;12.
- [243] O. Owoyele, P.J.A.E. Pal, A novel machine learning-based optimization algorithm (ActivO) for accelerating simulation-driven engine design, 285 (2021) 116455.
- [244] C. Kavuri, S.L.J.I.J.o.E.R. Kokjohn, Exploring the potential of machine learning in reducing the computational time/expense and improving the reliability of engine optimization studies, 21 (2020) 1251-1270.
- [245] Shih P, Kaul BC, Jagannathan S, Drallmeier JA. Reinforcement-learning-based output-feedback control of nonstrict nonlinear discrete-time systems with application to engine emission control. *IEEE Trans Syst Man Cybernetic Part B (Cybernetics)* 2009;39:1162–79.
- [246] Irdmousa BK, Rizvi SZ, Veini JM, Nabert JD, Shahbakhti M. Data-driven modeling and predictive control of combustion phasing for RCCI Engines. In: editor*editors. In: 2019 American Control Conference (ACC). 2019; 10-12 July 2019. p. 1617–22.