UNIVERSITY OF
BIRMINGHAM

University of Birmingham
Research at Birmingham

Data-driven enabling technologies in soft sensors of modern internal combustion engines

Li, Ji; Zhou, Quan; He, Xu; Chen, Wan; Xu, Hongming

DOI:
10.1016/j.energy.2023.127067

License:
Creative Commons: Attribution (CC BY)

Document Version
Publisher's PDF, also known as Version of record

Citation for published version (Harvard):
Li, J, Zhou, Q, He, X, Chen, W & Xu, H 2023, 'Data-driven enabling technologies in soft sensors of modern internal combustion engines: Perspectives', *Energy*, vol. 272, 127067. https://doi.org/10.1016/j.energy.2023.127067

Link to publication on Research at Birmingham portal

General rights
Unless a licence is specified above, all rights (including copyright and moral rights) in this document are retained by the authors and/or the copyright holders. The express permission of the copyright holder must be obtained for any use of this material other than for purposes permitted by law.

•Users may freely distribute the URL that is used to identify this publication.
•Users may download and/or print one copy of the publication from the University of Birmingham research portal for the purpose of private study or non-commercial research.
•User may use extracts from the document in line with the concept of ‘fair dealing’ under the Copyright, Designs and Patents Act 1988 (?)
•Users may not further distribute the material nor use it for the purposes of commercial gain.

Where a licence is displayed above, please note the terms and conditions of the licence govern your use of this document.

When citing, please reference the published version.

Take down policy
While the University of Birmingham exercises care and attention in making items available there are rare occasions when an item has been uploaded in error or has been deemed to be commercially or otherwise sensitive.

If you believe that this is the case for this document, please contact UBIRA@lists.bham.ac.uk providing details and we will remove access to the work immediately and investigate.

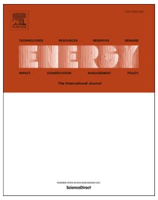
Download date: 29. May. 2023

Energy 272 (2023) 127067

Contents lists available at ScienceDirect

Energy

journal homepage: www.elsevier.com/locate/energy



Data-driven enabling technologies in soft sensors of modern internal combustion engines: Perspectives

Ji Li [a,**], Quan Zhou [a], Xu He [a], Wan Chen [b,a], Hongming Xu [a,*]

[a] Department of Mechanical Engineering, School of Engineering, University of Birmingham, Birmingham, B15 2TT, UK
[b] Hubei Key Laboratory of Advanced Technology for Automotive Components, Wuhan University of Technology, Wuhan, 430070, China

ARTICLE INFO

Handling Editor: Henrik Lund

Keywords:
Data driven
Enabling technology
Soft sensors
Internal combustion engines
Digital twin

ABSTRACT

Under the dual thrust of decarbonisation and digitalisation, data-driven enabling technologies become the most promising solutions to reducing the time, cost, and effort required in the development of modern internal combustion engines (ICEs) in which it is hard to handle high-data-cost, high-dimensional, complex nonlinear modelling problems. This paper proposes a view of data-driven enabling technologies used in ICE soft sensors with a focus on the reduction of experimental effort and model complexity to accelerate the development of ICE decarbonisation. The current progress in data-driven modelling of ICEs is briefly outlined from four aspects: data acquisition methods, data processing methods, machine learning methods and model validation methods. Moreover, the challenges of establishing ICE models with high accuracy, fast response, and strong robustness for real-time control are structured and analysed. Based on the challenges, perspectives on three aspects of versatility, practicality, and autonomy are presented. Finally, physics/data-enhanced machine learning and digital twin technology are suggested to empower soft sensors used for modern ICEs.

1. Introduction

Internal combustion engines (ICEs) are indispensable in the current power generation and transportation industries. It is critical to improve engine performance and efficiency and to reduce harmful emissions. After decades of research and design, ICEs have been now comprehensively developed but have also encountered bottlenecks. Engine modelling as the core task aims 1) to predict engine performance without having to conduct tests; 2) to deduce the performance of parameters that can be difficult to measure in tests. Such as engine-out transient emissions prediction, combustion knock and auto-ignition prediction, combustion noise and ringing intensity modelling, combustion mode transition modelling, they are still challenging due to the high nonlinearity and complexity and the highly transient and broad operating conditions of ICEs. In order to overcome these problems in the development of modern ICEs, accurate, fast-response, and low-development-cost enabling technologies are urgently needed to promote their decarbonisation and commercialisation.

Physically based modelling, although interpretable, requires immense expertise, and its solving process is time-consuming and unsuited for control purposes. Rapid development in informatics has enabled fast modelling of complex physical systems based on the measurement of real-world performance. Plenty of machine learning solutions have been proven to be able to handle complex nonlinear modelling problems in ICE development [1]. The successful implantation of data-driven models into engine control systems usually relies on a good quantity of experimental data. On the other hand, ICE experiment is typically complicated, costly, and time-consuming, due to detailed and accurate mechanisms of process or a wealth of experience and knowledge. In addition, the increasing complexity of modern ICE development makes these preconditions (e.g., emissions) difficult to meet. Therefore, it is imperative to develop data-driven modelling technologies to provide faster simulations and reduce the required expertise, interactions with the physical environment, time, and experimental costs for ICEs [2].

In view of Industry 4.0, soft sensors are efficient modelling means that are widely applied to meet the urgent development demand for different industrial applications. As a mathematical model with easy-to-measured variables as input and hard-to-measured variables as output, soft sensors estimate or predict important variables expediently [3]. Compared to the conventional physical sensors, the advantages of soft sensors are organised at each stage of the full life cycle, as shown in Fig. 1. As the unavoidable component of the digital economy, a soft

---

* Corresponding author.
** Corresponding author.
E-mail addresses: j.li.1@bham.ac.uk (J. Li), h.m.xu@bham.ac.uk (H. Xu).

https://doi.org/10.1016/j.energy.2023.127067
Received 6 January 2023; Received in revised form 22 February 2023; Accepted 24 February 2023
Available online 28 February 2023
0360-5442/© 2023 The Authors. Published by Elsevier Ltd. This is an open access article under the CC BY license (http://creativecommons.org/licenses/by/4.0/).

Abbreviations

ANN	Artificial neural network
DT	Digital twin
FCM	Fuzzy C-means
ELM	Extreme learning machine
GAN	Generative adversarial network
GB	Grey box
GMM	Gaussian mixture model
GPR	Gaussian process regression
ICE	Internal combustion engine
ML	Machine learning
NN	Neuronal network
QoS	Quality of service
SBIPV	Smart building-integrated photovoltaic
SVM	Support vector machine

sensor helps the digital economy enhance carbon emission performance at the significance level of 5% [4]. In the research of David et al. efficient and large-size soft sensors even display an outstanding ability to reduce the carbon footprint up to ~100–1000X [5]. Raul et al. [6] proposed a soft sensor for the implementation of wireless networked control systems to reduce battery consumption by up to 21%. In Norway, soft sensors reduce energy consumption in paper production, saving electricity by 10% [7]. For the ICE development, the soft sensors have great potential to reduce the dependence on experts, so that they could save a large amount of time and cost from the production and maintenance compared to conventional physical them. This further assists in promoting the rapid development of novel ICEs.

1.1. Related review work

Over the last two decades, plenty of review-type or survey-type articles related to modelling methods and specific applications of soft sensors have been published. As organzied in Table 1, the article [3] based on the current most significant advancement, presents a comprehensive review of the developments on soft sensors in process monitoring, control and optimisation. Soft sensors have been demonstrated to significantly reduce carbon emissions of power generation applications [8]. They also help use sustainable energy more effectively to reduce carbon emissions [9]. The comprehensive analysis and future

Table 1
The summary of the review papers on soft sensors.

Review paper	Publish year	Characteristics
[3]	2021	1. Includes the procedures of soft sensors 2. Focuses on the most up-to-the-date advancement 3. Needs to display more details of real applications
[8]	2022	1. Analyses the different soft sensors for combustion in power generation applications 2. Classifies combustion processes and the related applications of soft sensors 3. Needs to compare the performance of different optimisation algorithms in the combustion applications
[9]	2023	1. Summarises the application of Machine learning in sustainable energy 2. Explains the Role of machine learning in future of multi-carrier energy systems 3. Highlight the potential of circular integration
[10]	2023	1. Analyses data driven technology comprehensively 2. Proposes the application of data driven technology in SBIPV systems 3. Needs to provide more recommendations for the specific application of data driven technology in SBIPV systems
[11]	2021	1. Focuses on the engines with biodiesel 2. Shows the details of the modelling process by using soft sensors 3. Needs to display more details of real applications
[12]	2023	1. Reviews the application of soft sensors on the ICE performance and emission prediction 2. Includes the research about different ML technologies and algorithms 3. Needs to show the outlook on ICE performance and emission prediction
[13]	2022	1. Focuses on the application of the vehicle powertrain system 2. Includes the design and control of the powertrain system 3. Proposes the outlook of soft sensors in the entire vehicle industry
[14]	2021	1. Introduces the application of deep learning 2. Needs to display the more comparison between deep learning and other pure data-driven methods
[15]	2020	1. Considers the background of Industry 4.0 2. Introduces the Gray box technology 3. Needs to display more detail of model construction
[1]	2021	1. Focuses on the specific application: ICE 2. Displays more details of Machine Learning 3. Needs to display more comparison between the grey box and other pure data-driven methods

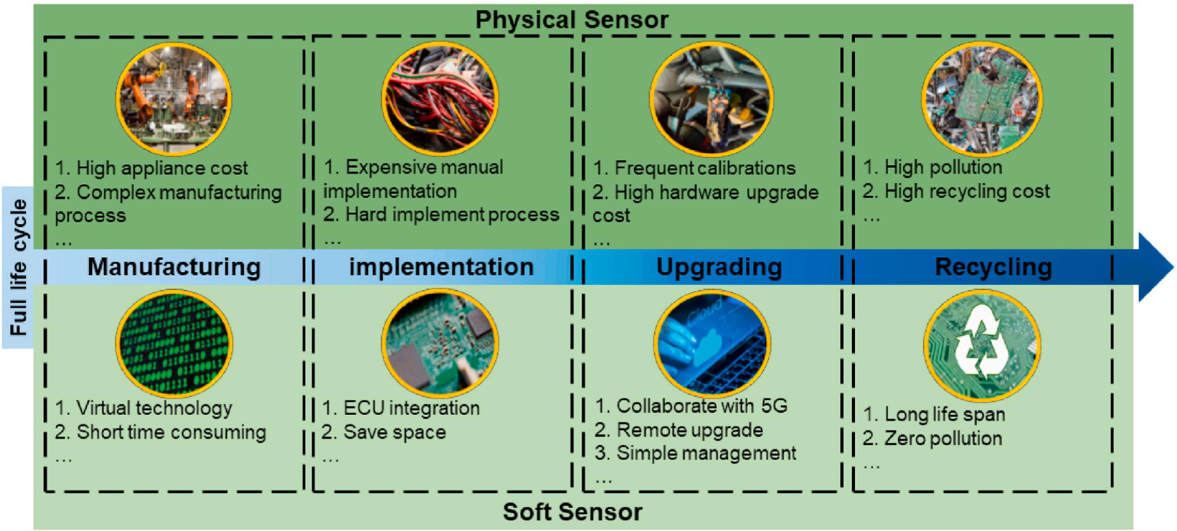

Physical Sensor
1. High appliance cost
2. Complex manufacturing process
…
1. Expensive manual implementation
2. Hard implement process
…
1. Frequent calibrations
2. High hardware upgrade cost
…
1. High pollution
2. High recycling cost
…
Full life cycle
Manufacturing
implementation
Upgrading
Recycling
1. Virtual technology
2. Short time consuming
…
1. ECU integration
2. Save space
…
1. Collaborate with 5G
2. Remote upgrade
3. Simple management
…
1. Long life span
2. Zero pollution
…
Soft Sensor

Fig. 1. Physical sensors vs soft sensors during their full life cycles.

perspectives of data driven technology in the smart building-integrated photovoltaic (SBIPV) systems are displayed in Ref. [10]. For engine development, soft sensors still play an unavoidable role. The article [11] thoroughly reviews the use of soft sensor technology in dedicated engines which use biodiesel and displays the superior potential of soft sensors for monitoring or controlling biodiesel systems in real-time. Similarly, The article [12] reviews the soft sensors used to predict ICE performance and emissions of ICEs. The soft sensors show a superior potential to extract the in-cylinder features and help to analyse the thermal process. Considering the entire vehicle powertrain system [13], summarises the application of soft sensors in the vehicle powertrain system, including the support for design and control components. Meanwhile, the outlook for the applications of soft sensors on the entire vehicle system is proposed. In Refs. [14,15], deep learning-based approaches and grey-box (GB) model-based approaches for data-driven soft sensors are specifically investigated, respectively. However, to date, few works provide some insight into the soft sensors of ICEs. The article [1] provides a critical review of the existing ICE modelling, optimisation, diagnosis, and control challenges and the promising state-of-the-art machine learning (ML) solutions for them. While this review article covers the related topic, no focus on data efficiency is put into the work.

1.2. *Scope and outline*

Based on the existing fruitful research outcomes, the focus of this paper is to observe the latest progress and common problems in data-driven soft sensors of modern ICEs and provide promising future directions for related enabling technologies.

The work presented in this paper is organised as follows. Firstly, the current progress in the data-driven enabling technology of ICE soft sensors is introduced and the related common issues are discussed. Afterwards, several future research directions are suggested to improve the versatility, practicality, and autonomy of ICE soft sensors. Based on these future research directions, some specific recommendations and implementation plans are proposed for assisting further improvement.

2. Current progress and common problems

Based on different functionalities, ICE models are generally developed for diagnostics, data analytics, optimisation, and control. To ensure the accuracy of predictive models, a large amount of data is usually obtained under diverse operating conditions. To build these models, there are four main phases, including data acquisition, data processing, model establishment, and model implementation. Its brief description is presented in Fig. 2, where each phase consists of its own main procedures to assist ICE development.

2.1. *Data acquisition*

Data acquisition refers to the use of various sensors to collect experimental samples from a target physical system. To improve sampling efficiency, Wang et al. [16] developed an efficient method for the identification of the engine volumetric efficiency map from transient condition data to reduce the map calibration time and cost compared to steady-state engine testing. The proposed method achieves the accurate identification of maps within the dynamic engine model in a short time. In the work of Li et al. [17], a Gaussian distributed resampling technique was proposed to locate and minimise the redundant data in the volumetric efficiency modelling. This allows screening a small number of samples with wide engine operations, which enables the proposed methodology to achieve superior learning efficiency with fewer samples. Though different methods have been applied to collect data efficiently, the quality of data acquisition is still unstable. The acquisition of experimental data in ICE development is a complex and expensive process. Selective collection of high-quality data needs to be taken into account in the design of experiments that could save significant experimental costs.

2.2. *Data processing*

To further improve the quality of experimental datasets, the dataset often needs to be processed before the data can be utilised. Data cleaning, feature extraction and feature selection are typically employed for ICE development. Data cleaning refers to sifting out data that affects the quality of the dataset, such as missing values, outliers, etc. Feature extraction is to extract variables that can reflect the characteristic information of the target issue from the dataset, and feature selection is to filter out the redundant variables for feature dimensionality reduction. Effective feature extraction and feature selection approaches can guarantee model accuracy with minimum computational cost to a certain extent. A recent review summarised several feature extraction methods used for ICEs, including t-distributed stochastic neighbour embedding, time domain and frequency domain statistics, wavelet transform, and deep neural networks [18]. As for feature selection techniques used for ICEs, in addition to the common principal component analysis [19], there are other effective methods such as the least absolute shrinkage and selection operator feature selection method [20], feature ranking method [21], feature extension method based on in-feature interactions [22], elitist genetic algorithm [23], decision tree algorithm [24], and hybrid feature selection algorithm based on statistical measures [25] and Fourier transform [19]. Because of noises and complex interaction among the feature signals in ICE development, such these methods should be carefully analysed and selected for different cases to ensure the reliability of data representation.

2.3. *Model establishment*

Accurate modelling of ICE has always been a daunting task as ICE is considered as a complex physical system. Machine learning methods have been shown to be effective in predicting the highly nonlinear and complex phenomena occurring within ICE. Supervised learning is the most common machine learning approach, which is designed to learn the input-output relations through training and can be used for regression and classification, such as artificial neural network (ANN) [26,27], support vector machine (SVM) [28,29], extreme learning machine (ELM) [30] and Gaussian process regression (GPR) [29,31]. Relatively, unsupervised learning does not require labelled training data but is

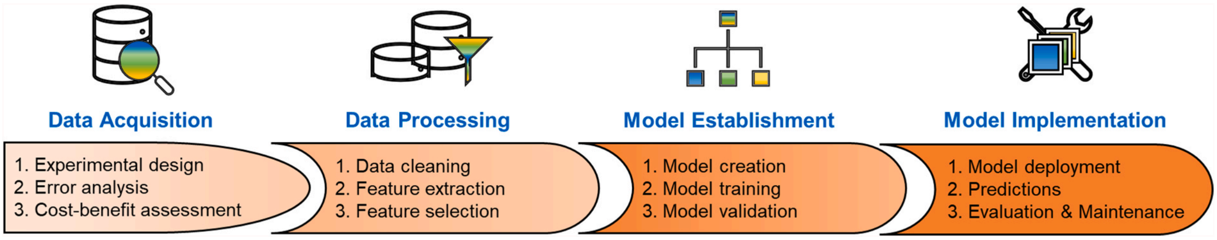


Fig. 2. A brief description of four main phases, including data acquisition, data processing, model establishment, and model implementation.

designed to recognise desired patterns from the available information set. Typical methods consist of K-means clustering, fuzzy C-means (FCM) and Gaussian mixture model (GMM), which are widely applied in industrial cases [32]. By using these machine learning technologies, however, interpretability and generalizability might be weak. Such brute force machine learning could be supplemented and improved by appropriate expert experience.

2.4. Model implementation

Data-driven models are, at this stage, widely used in various aspects. The specific model implementation includes model deployment, prediction, evaluation, and maintenance. In the development of ICE, multifaceted model implementation can replace traditional development models and save significant experimental costs. This helps to speed up the introduction of new ICEs into the marketplace. Summarising the current research, the progress of data-driven model implementation is inspiring. For instance, Siobhan et al. [33] presented a novel data-driven modelling approach to optimise the large-scale scenarios of electric vehicle charging. This data-driven model enabled rapid assessment of new electric vehicle rate designs. Meanwhile, the data-driven models have been used to predict various industrial indicators widely in the ICE industry [34]. For evaluation, data-driven models are generally validated to optimise their performance by using leave-one-out cross-validation [35] and K-fold cross-validation [36]. Based on data-driven models, various novel diagnosis approaches are proposed [1]. In the lab, these methods are proven to enhance the model prediction performance. However, the application of soft sensors should achieve the different real-time requirements in practice. The practice efficiency of the model implementation in practice should be considered and researched for the further development of soft sensors.

3. Future perspectives

Future engines require to be adaptable, cost effective, environmentally friendly and more efficient, with better fuel economy. Soft sensor technology is a promising solution to accelerate the achievement of these goals via empowering the versatility, practicality, and autonomy of ICEs.

3.1. Versatility

In order to adapt to complex environments and respond to different functional requirements, the high versatility of soft sensors needs to be met in the ICE development process. A soft sensor with high versatility owns the satisfactory ability to be compatible with different other technologies, such as digital twins and physics-informed learning to solve different kinds of issues. In ICE development, soft sensors are connected with data-driven enabling technologies to improve the efficiency of ICE development in terms of 1) modelling, 2) calibration, 3) controlling, and 4) diagnostics [1]. For instance, a digital twin auxiliary approach based on an adaptive sparse attention network is proposed by Jiang et al. to achieve the high versatility in diesel engine fault diagnosis, as shown in Fig. 3 [37]. However, the reliability of the soft sensor multifunction still needs to be improved to face different complex industrial conditions. Meanwhile, how to appropriately match the resources occupied by high versatility in soft sensors and the requirement for practical applications is still an issue which limits the further improvement of versatility in soft sensors.

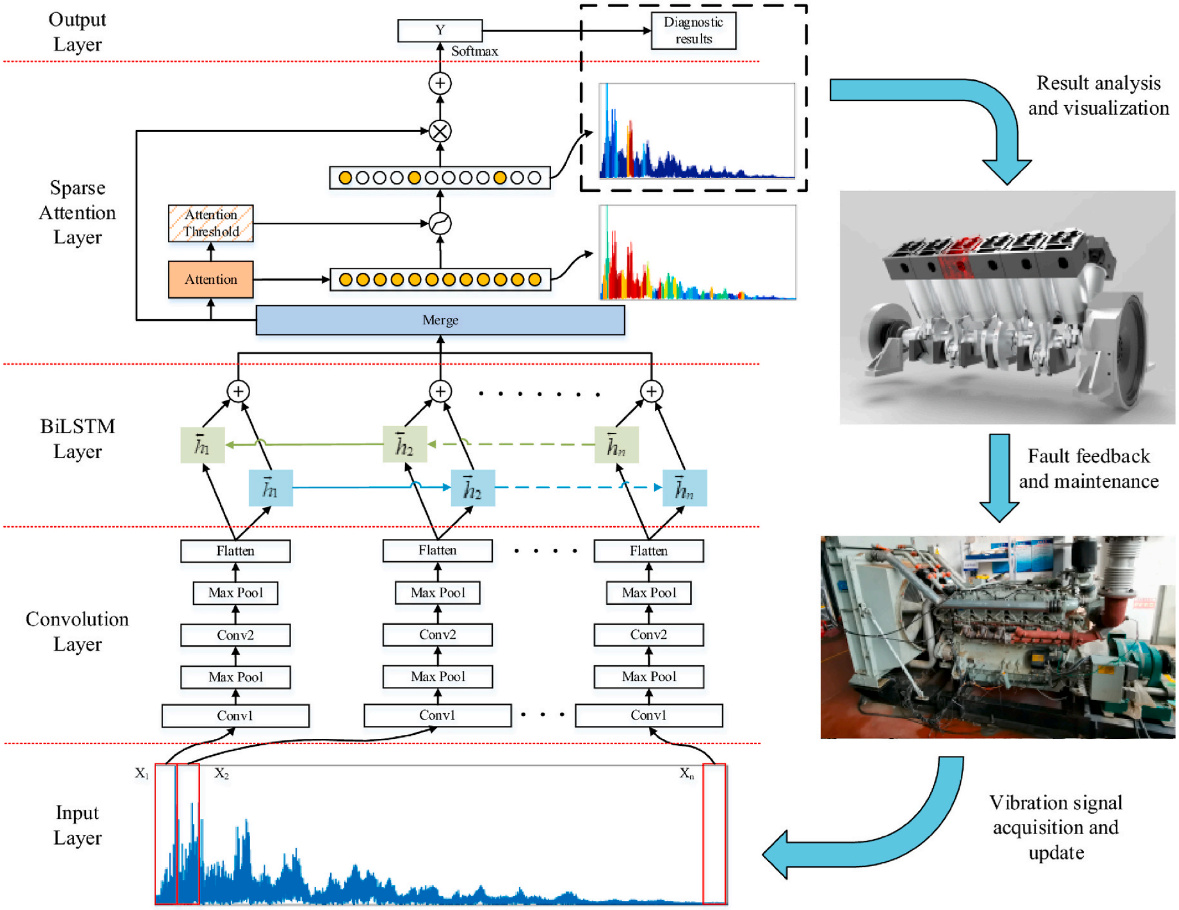

Output Layer
Y
Softmax
Diagnostic results
Result analysis and visualization
Sparse Attention Layer
Attention Threshold
Attention
Merge
BiLSTM Layer
$\overleftarrow{h}_1$
$\overleftarrow{h}_2$
$\overleftarrow{h}_n$
$\overrightarrow{h}_1$
$\overrightarrow{h}_2$
$\overrightarrow{h}_n$
Fault feedback and maintenance
Convolution Layer
Flatten
Max Pool
Conv2
Max Pool
Conv1
Input Layer
X1
X2
Xn
Vibration signal acquisition and update

Fig. 3. Fault diagnosis process of digital twin based on an adaptive sparse attention network [37].

3.2. Practicality

Early soft sensor development was obsessed with improving predictive performance, which caused the extremely complicated design of soft sensor algorithms. Inalg practice, the complex design brings a huge computational load and running time cost. Especially for the ICE development, the possible prediction delay brought by huge computational is intolerable. For ICE development, practicality is displayed in the ability of real-time prediction and acceptable implementation cost [38]. For the real-time control of ICEs, even the entire powertrain system, some attempts have been reported. For example, multiple optimisation methods have been proposed to improve the usage efficiency of the entire powertrain system, including ICEs, as shown in Fig. 4 [39]. To fill the gap between the laboratory outcome and the industrial practice, a further in-depth collaboration between academia and the industry is considered a useful approach to provide more flexible space and available opportunities to optimise the practicality of the current soft sensor algorithms.

3.3. Autonomy

Due to soft sensors being used in different application scenarios of the vehicular system, the robustness of soft sensors is considered as a key factor in measuring performance. Autonomy reflects the ability to compensate for system failures without external intervention and keep the satisfactory performance of models. Superior autonomy strengthens the robustness of models and saves the cost of manual adjustment. Therefore, it has become an urgent requirement for novel soft sensor technology. Although the direction of autonomy development has not been approached systematically and extensively today, it is inspiring to notice that there are some attempts reported. The authors in Ref. [40] proposed to conduct automated validation to improve the accuracy and reliability of soft sensors by integrating just-in-time models and relevant vector machines. Adaptive optimisation of the whole energy system is proposed for plug-in hybrid electric vehicles to improve the usage efficiency significantly [41]. The schematic diagram of this adaptive optimisation is shown in Fig. 5.

4. Recommendations

Depending on the rapid development of informatics, many emerging enabling technologies show superior potential to assist the effective application of ICE soft sensors. This section introduces useful data-driven enabling technologies for the further developing directions of soft sensors in terms of 1) physics/data-enhanced machine learning and 2) digital twin technology.

4.1. Physics/data-enhanced machine learning

The performance of a supervised machine learning model is largely influenced by the dataset that should contain abundant labelled data. However, collecting such an amount of labelled data tends to be a highly time-consuming, expensive, and complicated process in many practical applications. To reduce experimental effort and model complexity in developing accurate ICE predictive models, the emerging machine learning methods with data-directed and physical-directed enhancements are presented for ICEs in this section. Fig. 6 displays these two enhancement methods for the data-driven soft sensors.

4.1.1. Data-directed enhancement

The data-directed enhancement assists in creating trustworthy models by improving the quality of data. As shown in Fig. 6, different data-enhanced approaches, e.g., scaling-down, data augmentation and knowledge transfer, are applied to enhance the models.

Scaling down: Scaling-down sampling is applied to scale down the data size and extract the representative samples for further modelling [42]. By this sampling approach, the ICE model could be trained with fewer data and further save significant computational time. However, limited interpretability is still an unavoidable issue for the practical application of scaling-down sampling.

Data augmentation: Compared to scaling-down sampling, data augmentation is applied to generate equivalent data for expanding the training dataset. For ICE diagnostics, one of the data augmentation, generative adversarial network (GAN) generates misleading data to strengthen the judgement ability of faults and further improve the

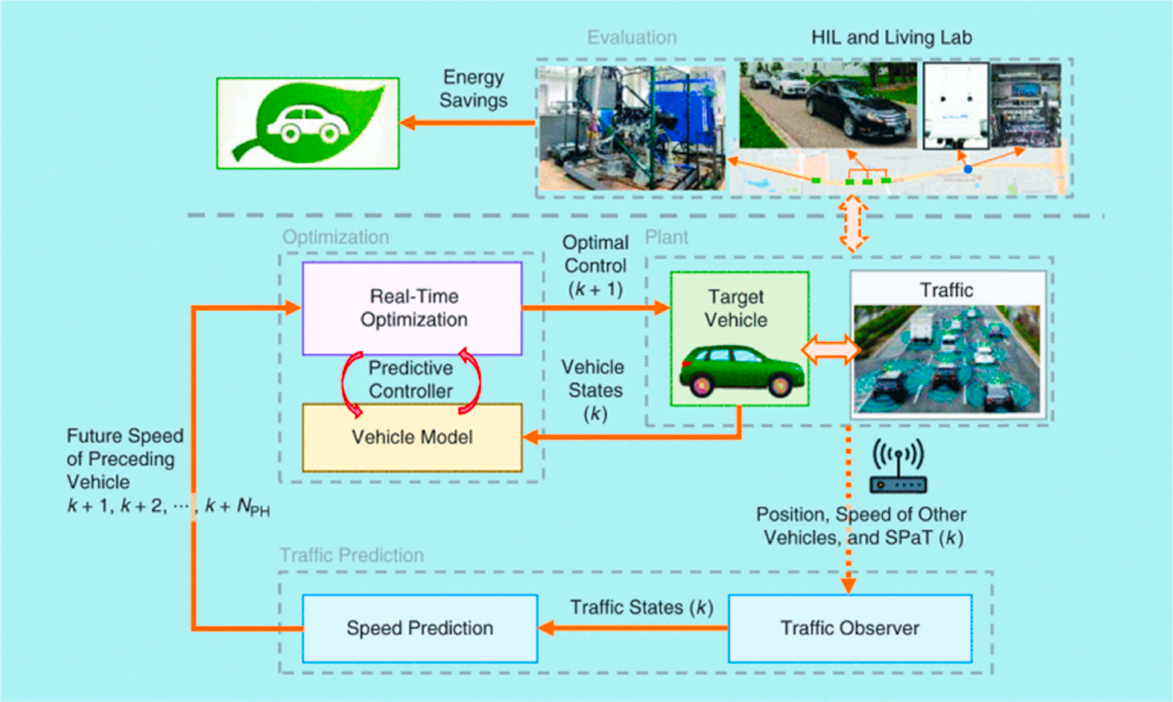


Fig. 4. A real-time control framework with superior practicality [39].

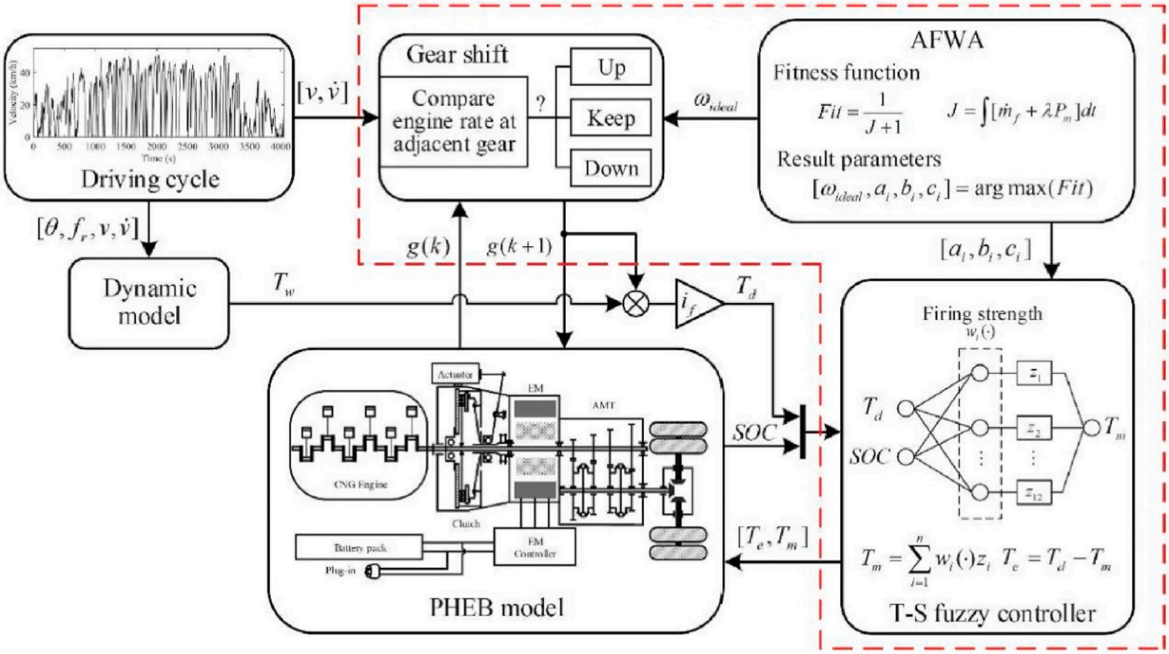

Driving cycle
$[v,\dot{v}]$
Gear shift
Compare engine rate at adjacent gear
?
Up
Keep
Down
ω_{ideal}
AFWA
Fitness function
$Fit=\frac{1}{J+1}$
$J=\int[\dot{m}_f+\lambda P_m]dt$
Result parameters
$[\omega_{ideal},a_i,b_i,c_i]=\arg\max(Fit)$
$[\theta,f_r,v,\dot{v}]$
Dynamic model
T_w
$g(k)$
$g(k+1)$
i_f
T_d
$[a_i,b_i,c_i]$
Firing strength $w_i(\cdot)$
T_d
SOC
z_1
z_2
z_{12}
T_m
SOC
$[T_e,T_m]$
$T_m=\sum_{i=1}^{n}w_i(\cdot)z_i$, $T_e=T_d-T_m$
T-S fuzzy controller
PHEB model
Actuator
EM
AMT
CNG Engine
Clutch
Battery pack
Plug-in
EM Controller

Fig. 5. The adaptive optimisation system displaying high autonomy [41].

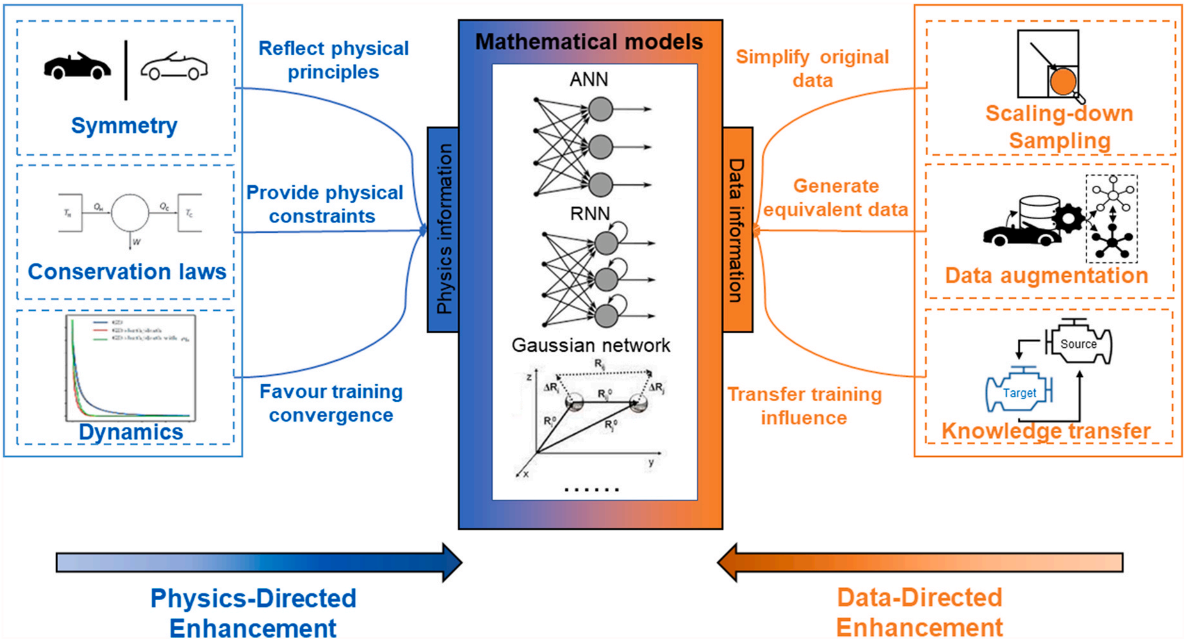

Symmetry
Conservation laws
Dynamics
Reflect physical principles
Provide physical constraints
Favour training convergence
Physics information
Mathematical models
ANN
RNN
Gaussian network
……
Data information
Simplify original data
Generate equivalent data
Transfer training influence
Scaling-down Sampling
Data augmentation
Source
Target
Knowledge transfer
Physics-Directed Enhancement
Data-Directed Enhancement

Fig. 6. The different enhancement approaches for data-driven soft sensors.

classification accuracy of the ICE models [43]. At the same time, the misleading data bring negative noise for the ICE modelling, causing unstable training and even model collapse. How to fix these issues is the urgent need for the further efficient application of GAN.

Knowledge transferring: Transfer learning [44] is a widely used method to transfer knowledge from current cases to unknown cases and further alleviate the need for large amounts of labelled data. The basic idea behind transfer learning is to transfer data information (features or models) from a source domain to a target domain to solve the problem of the lack of data in the target domain. Li et al. [45] presented a novel approach to geometric neuro-fuzzy transfer learning for a diesel engine fuelled with microalgae oil. This approach only utilises limited experimental data obtained by geometric screening to learn a high-precise transfer model of the in-cylinder pressure with different blending ratios. Though transfer learning has been proven to bring significant positive influences for modelling, the instability of the transferred performance is a major constraint to transfer learning.

4.1.2. Physics-directed enhancement

To enhance the interpretability and prediction accuracy of the models, physical information is involved in the modelling process to strengthen the connection between the physical plant and its soft sensors. Various physical information is introduced in the modelling process and forms the physics-directed enhancement to optimise the data-driven soft sensors widely [46].

Symmetry: By introducing the observational biases in the models, the physical principles are reflected as the learning functions, vector fields and other physical data. Using this physical information, the evolution of complex phenomena in the modelling process is detected. Based on this interpretable and physically enhanced model, the efficiency of Machine learning has been enhanced in practical applications [47]. However, the physical symmetry extremely depends on the sensors to collect various physical data. This will bring a huge experimental cost, including financial and time costs. To further develop the symmetry technology in practical applications, the efficient saving of experimental cost is necessary.

Conservation laws: Physics-informed learning introduces inductive biases into the data-driven soft sensors to add tailored physical interventions. These interventions, as the derivation of the physical conservation laws, further monitor the modelling process and fix the possible errors of the models. A classic example is the application of covariant neuronal networks (NNs), which is related to the rotation and translation invariances present in the many-body system and improves the prediction accuracy of NNs [48]. Due to the limited known conservation laws in the physical world, this technology is hard to be implemented in the complex task at the current stage.

Dynamics: Because physics-informed learning keeps dynamic evaluation during the entire training process, dynamics is the unavoidable influence factor. To solve the dynamics issues and favour the training convergence rapidly, learning biases, including loss functions, constraints and inference algorithms, are introduced into the models. Compared to the conservation laws, the dynamics is considered as the soft penalty constraint to approximately satisfy the physical laws. Though the positive influence of the Dynamics is proven [49], the unsteady model performance brought from the approximate satisfaction in the different industrial applications still stops the further development of the dynamics.

4.2. *Digital twin technology*

In Industry 4.0, digital twin (DT) is gaining the ever-increasing attention of many scholars and industrial sectors. Due to the superior ability to reflect the physical asset as the virtual representation, DT has proven to be useful for the further development of the vehicle powertrain system, e.g., vehicle energy management [50] and the rapid development of ICE [51]. Model formulation, as we know, is fundamental of DT technology. Apart from this, DT technology also displays its outstanding potential in other directions. Inspired by DT's new class defined by Grieves, the further development directions of DTs for the ICE industry are drawn in Fig. 7.

4.2.1. *Data fusion*

Data fusion aims to integrate data information and improve data quality before data usage. It generally includes three procedures of data preprocessing, data mining, and data optimisation. In many industrial cases, Data fusion has been proven to bring a significant positive influence, including more comprehensive and real-time reflection of each element in the entire vehicular system [52]. To be more significant, it can provide complementary views of the same phenomenon by combining multiple interrelated datasets. The fusing information allows more accurate inferences than those from a single dataset.

4.2.2. *Interaction*

Interaction is another attractive development direction. The interaction process includes flexible connection and efficient collaboration among all DT parts. By using different connection combinations, i.e., physical-physical, physical-virtual, and virtual-virtual connection, physical entities and virtual models are flexibly connected. This flexible connection assists the interactions among each DT part to provide the entire DT system superior ability to address complex and real-time tasks. It is proven in many practical cases, such as production system optimisation [53] and fault diagnostics [54]. To maximise the efficiency of the whole vehicle system, the ICE needs to closely interact with other component parts such as motors, batteries, and drivers via different connection combinations of their DTs. To build accessible communication between each other, soft sensors are used to determine and integrate their communication signals to be readable.

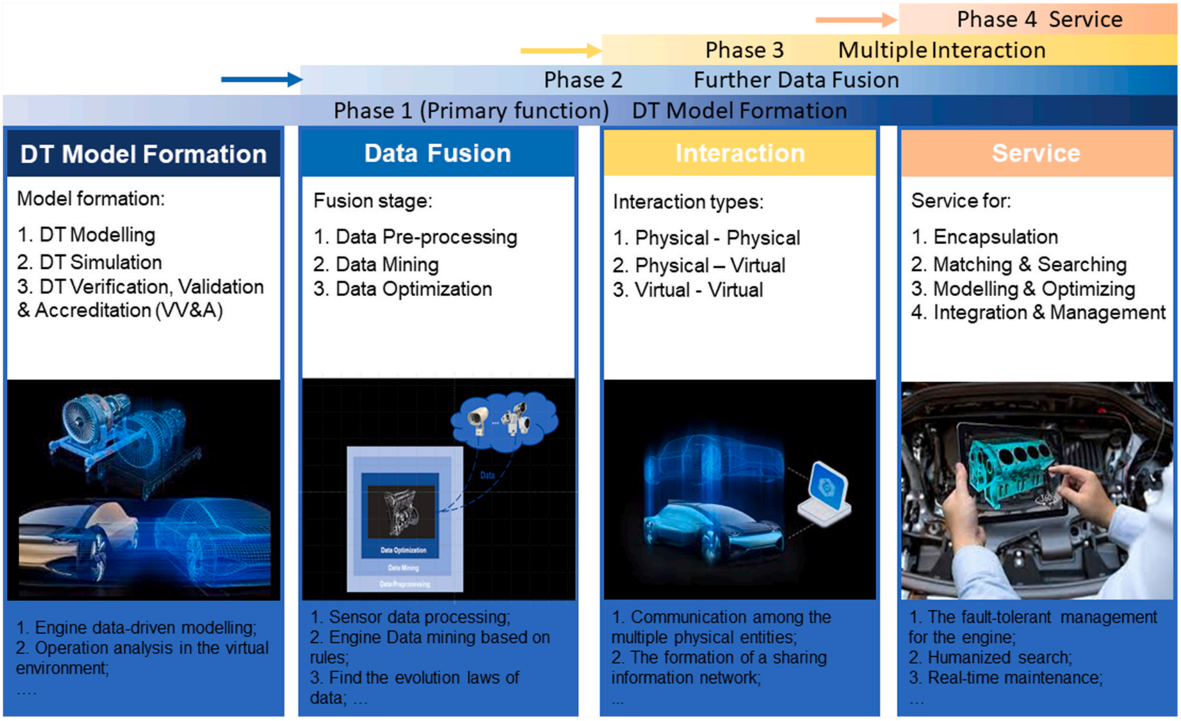

Phase 4 Service
Phase 3 Multiple Interaction
Phase 2 Further Data Fusion
Phase 1 (Primary function) DT Model Formation
DT Model Formation
Model formation:
1. DT Modelling
2. DT Simulation
3. DT Verification, Validation & Accreditation (VV&A)
1. Engine data-driven modelling;
2. Operation analysis in the virtual environment;
....
Data Fusion
Fusion stage:
1. Data Pre-processing
2. Data Mining
3. Data Optimization
Data
Data Optimization
Data Mining
Data Preprocessing
1. Sensor data processing;
2. Engine Data mining based on rules;
3. Find the evolution laws of data; ...
Interaction
Interaction types:
1. Physical - Physical
2. Physical – Virtual
3. Virtual - Virtual
1. Communication among the multiple physical entities;
2. The formation of a sharing information network;
...
Service
Service for:
1. Encapsulation
2. Matching & Searching
3. Modelling & Optimizing
4. Integration & Management
1. The fault-tolerant management for the engine;
2. Humanized search;
3. Real-time maintenance;
...

Fig. 7. DT evolution in ICE development.

4.2.3. *Service*

Services also can be reinforced by DT on many occasions such as structure monitoring, lifetime forecasting, in-time maintenance, etc. Not only can new services be enabled by DTs, but also existing services can be enhanced by the new data supplied by DTs. The topics include service encapsulation, service matching and searching, quality of service (QoS) modelling and evaluation, service optimisation and integration, and fault-tolerance management [55]. Service encapsulation helps to achieve different functions by using the same interface. In contrast, service matching and searching could meet different requirements of DT service from different clients. For the long term, QoS modelling, and service optimisation continuously update the relative services whilst the fault-tolerance management monitors the possible fault.

5. Conclusion

Soft sensors are proven to own the superior ability to accelerate the development of modern internal combustion engines. To further enhance the performance of soft sensors, data-driven enabling technologies have been commonly applied in both labs and practice. Currently, the data-driven enabling technologies in soft sensors are limited by four major issues: 1) the instability of data acquisition's quality; 2) the weak robustness of feature selection methods in different cases; 3) the poor interpretability of the established ML models and 4) the limited practical efficiency of data-driven enabling technologies in the engineering application. To address these issues, future perspectives are proposed in terms of versatility, practicality, and autonomy. Based on these perspectives, physics/data-enhanced machine learning and digital twins are considered the main assistance for the further development of soft sensors in ICEs. With these two advanced technologies, the performance of soft sensors in ICEs could be improved significantly, including less experimental cost, higher prediction accuracy and stronger practical robustness.

Based on the comprehensive analysis of the current progress, further perspectives, and future recommendation, we hope this paper will be useful to inspire the academic and the industry which focus on soft sensors in the development of internal combustion engines and inspirit them in further research.

Credit author statement

Ji Li: Conceptualization, Methodology, Writing original draft, Writing – review & editing. Quan Zhou: Investigation, Project administration, Writing – review & editing. Xu He: Literature analysis, Visualisation, Writing original draft. Wan Chen: Literature analysis, Writing original draft. Hongming Xu: Supervision, Project administration, Writing – review & editing.

Declaration of competing interest

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

Data availability

No data was used for the research described in the article.

References

[1] Aliramezani M, Koch CR, Shahbakhti M. Modeling, diagnostics, optimisation, and control of internal combustion engines via modern machine learning techniques: a review and future directions. Prog Energy Combust Sci 2022;88(September 2021): 100967. https://doi.org/10.1016/j.pecs.2021.100967.

[2] The rise of data-driven modelling. Nat. Rev. Phys. 2021;3(6):383. https://doi.org/10.1038/s42254-021-00336-z.

[3] Jiang Y, Yin S, Dong J, Kaynak O. A review on soft sensors for monitoring, control, and optimization of industrial processes. IEEE Sensor J 2021;21(11):12868–81. https://doi.org/10.1109/JSEN.2020.3033153.

[4] Zhang W, Liu X, Wang D, Zhou J. Digital economy and carbon emission performance: evidence at China's city level. Energy Pol 2022;165:112927. https://doi.org/10.1016/j.enpol.2022.112927. March.

[5] Patterson D, et al. Carbon emissions and large neural network training. 2021. p. 1–22 [Online]. Available:.

[6] Mansano RK, Godoy EP, Porto AJV. The benefits of soft sensor and multi-rate control for the implementation of wireless networked control systems. Sensors 2014;14(12):24441–61. https://doi.org/10.3390/s141224441.

[7] The Research Council of Norway, "SoftSens - soft sensor technology and advanced modeling for reduced energy consumption in paper production." https://prosjektbanken.forskningsradet.no/project/FORISS/310135?Kilde=FORISS&distribution=Ar&chart=bar&calcType=funding&Sprak=no&sortBy=date&sortOrder=desc&resultCount=30&offset=0&Prosjektleder=Lars+Johansson.

[8] K. A., Kasra Mohammadi KMP, , Jake Immonen, Blackburn Landen D, Tuttle Jacob F. A review on the application of machine learning for combustion in power generation applications. Rev Chem Eng 2022:618. https://doi.org/10.1109/I-SMAC55078.2022.9987365.

[9] Ifaei P, Nazari-Heris M, Tayerani Charmchi AS, Asadi S, Yoo CK. Sustainable energies and machine learning: an organised review of recent applications and challenges. Energy 2023;266:126432. https://doi.org/10.1016/j.energy.2022.126432.

[10] Liu Z, et al. A review of data-driven smart building-integrated photovoltaic systems: challenges and objectives. Energy 2023;263:126082. https://doi.org/10.1016/j.energy.2022.126082.

[11] Aghbashlo M, et al. Machine learning technology in biodiesel research: a review. Prog Energy Combust Sci 2021;85:100904. https://doi.org/10.1016/j.pecs.2021.100904.

[12] Karunamurthy K, Janvekar AA, Palaniappan PL, Adhitya V, Lokeswar TTK, Harish J. Prediction of IC engine performance and emission parameters using machine learning: a review. J Therm Anal Calorim 2023. https://doi.org/10.1007/s10973-022-11896-2.

[13] Zhou Q, Li J, Xu H. Artificial intelligence and its roles in the R & D of vehicle powertrain products. 2022.

[14] Sun Q, Ge Z. A survey on deep learning for data-driven soft sensors. IEEE Trans Ind Inf 2021;17(9):5835–66. https://doi.org/10.3934/qfe.2021032.

[15] Ahmad I, Ayub A, Kano M, Cheema II. Gray-box soft sensors in process industry: current practice, and future prospects in era of big data. Processes 2020;8(2):1–20. https://doi.org/10.3390/pr8020243.

[16] Wang C, Liang M, Chai Y. Finite-time identification algorithm for volumetric efficiency map in SI gasoline engines. IEEE Trans Ind Electron 2020;67(12): 10702–12. https://doi.org/10.1109/TIE.2019.2962481.

[17] Li J, Zhou Q, Williams H, Xu P, Xu H, Lu G. Fuzzy-tree-constructed data-efficient modelling methodology for volumetric efficiency of dedicated hybrid engines. Appl Energy 2021;310:2022. https://doi.org/10.1016/j.apenergy.2022.118534.

[18] Shi Q, Hu Y. Review on intelligent diagnosis technology of electronically controlled fuel injection system of ME diesel engine. Acad J Sci Technol 2022;1(2):69–75. https://doi.org/10.54097/ajst.v1i2.351.

[19] Ping X, Yang F, Zhang H, Xing C, Yao B, Wang Y. An outlier removal and feature dimensionality reduction framework with unsupervised learning and information theory intervention for organic Rankine cycle (ORC). Energy 2022;254:124268. https://doi.org/10.1016/j.energy.2022.124268.

[20] Shahpouri S, Norouzi A, Hayduk C, Rezaei R, Shahbakhti M, Koch CR. Soot emission modeling of a compression ignition engine using machine learning. IFAC-PapersOnLine 2021;54(20):826–33. https://doi.org/10.1016/j.ifacol.2021.11.274.

[21] Shaw H, Vijay SK. Diagnosis and detection of IC engine fault at end of line engine vibration measurement system with machine learning model. Proc Int Conf Ind Eng Oper Manag 2021:216.

[22] Gordon D, et al. Support vector machine based emissions modeling using particle swarm optimisation for homogeneous charge compression ignition engine. Int J Engine Res 2021. https://doi.org/10.1177/14680874211055546.

[23] Ran Q, Song Y, Du W, Du W, Peng X. Fault detection of diesel engine air and after-treatment systems with high-dimensional data: a novel fault-relevant feature selection method. Processes 2021;9(2):1–15. https://doi.org/10.3390/pr9020259.

[24] Arockia Dhanraj J, et al. Implementation of K* classifier for identifying misfire prediction on spark ignition four-stroke engine through vibration data. SAE Tech Pap 2021;17323. https://doi.org/10.4271/2021-28-0282.

[25] Li J, Zhou Q, Williams H, Lu G, Xu H. Statistics-guided accelerated swarm feature selection in data-driven soft sensors for hybrid engine performance prediction. IEEE Trans Ind Inf 2022:1. https://doi.org/10.1109/TII.2022.3199259. –11.

[26] Zandie M, Ng HK, Gan S, Muhamad Said MF, Cheng X. Multi-input multi-output machine learning predictive model for engine performance and stability, emissions, combustion and ignition characteristics of diesel-biodiesel-gasoline blends. Energy 2023;262:125425. https://doi.org/10.1016/j.energy.2022.125425.

[27] Can Ö, Baklacioglu T, Özturk E, Turan O. Artificial neural networks modeling of combustion parameters for a diesel engine fueled with biodiesel fuel. Energy 2022; 247. https://doi.org/10.1016/j.energy.2022.123473.

[28] Ağbulut Ü, Gürel AE, Sarıdemir 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. https://doi.org/10.1016/j.energy.2020.119076.

[29] Wang H, et al. Comparison and evaluation of advanced machine learning methods for performance and emissions prediction of a gasoline Wankel rotary engine. Energy 2022;248:123611. https://doi.org/10.1016/j.energy.2022.123611.

[30] Cocco Mariani V, Hennings Och S, dos Santos Coelho L, Domingues E. Pressure prediction of a spark ignition single cylinder engine using optimised extreme learning machine models. Appl Energy 2019;249:204–21. https://doi.org/10.1016/j.apenergy.2019.04.126.

[31] Huang H, Song Y, Peng X, Ding S, Zhong W, Du W. A sparse nonstationary trigonometric Gaussian process regression and its application on nitrogen oxides prediction of the diesel engine. IEEE Trans Ind Inf 2021;17(12):8367–77. https://doi.org/10.1109/TII.2021.3068288.

[32] Vanem E, Brandsæter A. Unsupervised anomaly detection based on clustering methods and sensor data on a marine diesel engine. J Mar Eng Technol 2021;20(4):217–34. https://doi.org/10.1080/20464177.2019.1633223.

[33] Powell S, Vianna Cezar G, Apostolaki-Iosifidou E, Rajagopal R. Large-scale scenarios of electric vehicle charging with a data-driven model of control. Energy 2022;248. https://doi.org/10.1016/j.energy.2022.123592.

[34] Simsek S, Uslu S, Simsek H. Proportional impact prediction model of animal waste fat-derived biodiesel by ANN and RSM technique for diesel engine. Energy 2022;239:122389. https://doi.org/10.1016/j.energy.2021.122389.

[35] de Carvalho RN, Machado GB, Colaço MJ. Estimating gasoline performance in internal combustion engines with simulation metamodels. Fuel 2017;193:230–40. https://doi.org/10.1016/j.fuel.2016.12.057.

[36] Wang H, et al. Comparison and implementation of machine learning models for predicting the combustion phases of hydrogen-enriched Wankel rotary engines. Fuel 2022;310:122371. https://doi.org/10.1016/j.fuel.2021.122371.

[37] Jiang J, et al. A digital twin auxiliary approach based on adaptive sparse attention network for diesel engine fault diagnosis. Sci Rep 2022;12(1):1–18. https://doi.org/10.1038/s41598-021-04545-5.

[38] zhao An Y, et al. Development of a PAH (polycyclic aromatic hydrocarbon) formation model for gasoline surrogates and its application for GDI (gasoline direct injection) engine CFD (computational fluid dynamics) simulation. Energy 2016;94:367–79. https://doi.org/10.1016/j.energy.2015.11.014.

[39] Shao Yunli, Sun Z. Energy-Efficient connected and automated vehicles: real-time traffic prediction-enabled co-optimisation of vehicle motion and powertrain operation. IEEE Veh Technol Mag 2016;138(12). https://doi.org/10.1115/1.2016-dec-2. S5–S11.

[40] Liu Y. Adaptive just-in-time and relevant vector machine based soft-sensors with adaptive differential evolution algorithms for parameter optimisation. Chem Eng Sci 2017;172:571–84. https://doi.org/10.1016/j.ces.2017.07.006.

[41] Yang C, Liu K, Jiao X, Wang W, Chen R, You S. An adaptive firework algorithm optimisation-based intelligent energy management strategy for plug-in hybrid electric vehicles. Energy 2022;239:122120. https://doi.org/10.1016/j.energy.2021.122120.

[42] ElRafey A, Wojtusiak J. Recent advances in scaling-down sampling methods in machine learning. Wiley Interdiscip Rev Comput Stat 2017;9(6). https://doi.org/10.1002/wics.1414.

[43] Qin C, et al. DTCNNMI: a deep twin convolutional neural networks with multi-domain inputs for strongly noisy diesel engine misfire detection. Meas J Int Meas Confed 2021;180(May):109548. https://doi.org/10.1016/j.measurement.2021.109548.

[44] Bozinovski S. Reminder of the first paper on transfer learning in neural networks. Inform 2020;44(3):291–302. https://doi.org/10.31449/INF.V44I3.2828. 1976.

[45] Li J, Wu D, Mohammadsami Attar H, Xu H. Geometric neuro-fuzzy transfer learning for in-cylinder pressure modelling of a diesel engine fuelled with raw microalgae oil. Appl Energy April 2021;306:2022. https://doi.org/10.1016/j.apenergy.2021.118014.

[46] Karniadakis GE, Kevrekidis IG, Lu L, Perdikaris P, Wang S, Yang L. Physics-informed machine learning. Nat. Rev. Phys. 2021;3(6):422–40. https://doi.org/10.1038/s42254-021-00314-5.

[47] Kashefi A, Rempe D, Guibas LJ. A point-cloud deep learning framework for prediction of fluid flow fields on irregular geometries. Phys Fluids 2021;33. https://doi.org/10.1063/5.0033376.

[48] Kondor R, Hy TS, Pan H, Anderson BM, Trivedi S. Covariant compositional networks for learning graphs. 6th Int Conf Learn Represent ICLR 2018 - Work Track Proc 2018:1–19.

[49] Geneva N, Zabaras N. Modeling the dynamics of PDE systems with physics-constrained deep auto-regressive networks. J Comput Phys 2020;403:109056. https://doi.org/10.1016/j.jcp.2019.109056.

[50] Zhang C, et al. Dedicated Adaptive Particle Swarm Optimisation Algorithm for Digital Twin Based Control Optimization of the Plug-in Hybrid Vehicle 2022:1–12. https://doi.org/10.1109/TTE.2022.3219290.

[51] Zheng Y, Chen L, Lu X, Sen Y, Cheng H. Digital twin for geometric feature online inspection system of car body-in-white. Int J Comput Integrated Manuf 2021;34(7–8):752–63. https://doi.org/10.1080/0951192X.2020.1736637.

[52] Tao F, Cheng Y, Cheng J, Zhang M, Xu W, Qi Q. Theories and technologies for cyber-physical fusion in digital twin shop-floor. Comput Integr Manuf Syst 2017;23:1603–11.

[53] Vachalek J, Bartalsky L, Rovny O, Sismisova D, Morhac M, Loksik M. The digital twin of an industrial production line within the industry 4.0 concept. Proc 2017 21st Int Conf Process Control PC 2017:258–62. https://doi.org/10.1109/PC.2017.7976223. 2017.

[54] Huang Y, Tao J, Sun G, Wu T, Yu L, Zhao X. A novel digital twin approach based on deep multimodal information fusion for aero-engine fault diagnosis. Energy 2023;270(September 2022):126894. https://doi.org/10.1016/j.energy.2023.126894.

[55] T. C. F.. In: Combustion, fuels, materials, design. MIT Press; 1985. https://doi.org/10.1109/TII.2018.2873186. Internal combustion engine in theory and practice, revised.