# Statistics-Guided Accelerated Swarm Feature Selection in Data-Driven Soft Sensors for Hybrid Engine Performance Prediction

Li, Ji; Zhou, Quan; Williams, Huw; Xu, Hongming

*DOI:*
10.1109/TII.2022.3199259

# Statistics-Guided Accelerated Swarm Feature Selection in Data-Driven Soft Sensors for Hybrid Engine Performance Prediction

Ji Li, Member, *IEEE*, Quan Zhou, Member, *IEEE*, Huw Williams, Guoxiang Lu, and Hongming Xu

***Abstract*—The accurate prediction of soft sensors is essential for development of modern combustion engines to achieve better performance, lower emissions, and reduced fuel consumption. To precisely predict engine performance i.e., indicated thermal efficiency, volumetric efficiency, and fuel consumption rate of a hybrid engine, this paper proposes a novel data-driven approach of statistics-guided accelerated swarm feature selection to find the most effective features for engine soft sensors. Differing from the existing filter or wrapper feature selection approaches, this approach uses external measure information to direct velocity updates in the accelerated swarm feature selection. Several filter and wrapper methods were developed and comprehensively compared. The experimental dataset was collected from a BYD 1.5L gasoline engine. Validated by bench test, the results demonstrate that the proposed approach finds the most effective features and optimal network structure for data-driven performance prediction of the hybrid engine that was studied.**

***Index Terms*—accelerated particle swarm optimization; deep neural network; engine soft sensors; feature selection.**

## I. Introduction

With the rapid development of informatics and an increase in customer demands, the complexity of process industry is growing fast. For monitoring the operation status of systems and realizing the optimal control of products, however, their key variables or quality indices must be obtained as fast and accurately as possible [1]. Soft sensors, a kind of mathematical model with easy-to-measured auxiliary variables as input and hard-to-measure variables as output, have been widely constructed to estimate or predict important variables expediently [2].

In auto industry, engine experiment is typically complicated, costly, and time-consuming [3], due to detailed and accurate mechanism of process or a wealth of experience and knowledge. However, the increasing complexity of modern engine development makes these preconditions (e.g., emissions) difficult to meet [4]. Data-driven modelling has been proven beneficial to reduce expertise requirements, time and experimental costs [5]. Li et al. propose geometric neuro-fuzzy transfer learning for In-cylinder pressure modelling. This approach only utilizes limited experimental data obtained by geometric screening to learn a high precise transfer model, which performs superior computational and experimental efficiency [6]. Besides, Quan et al. research a transferable representation modelling routine, where two artificial intelligence technologies of deep neural network [7] and Gaussian process regression [8] are developed to cooperate with an adaptive neuro-fuzzy inference system for knowledge transfer of the energy management controller. For membership function design, statistical methods e.g., corresponding analysis can be used to measure sensitivity of lift surfaces in air-fuel ratio control [9]. Tosun et al. predicted injection duration by using different types of regression analysis [10], in which an artificial neural network (ANN) demonstrated better accuracy than other linear and non-linear regression methods. In addition, some studies on multi-input modelling by using ANNs have also performed well, such as LP-EGR flow [11] and NOx emission [12]. However, the complex nature of ANN may lead to problems of computational time, energy use, and memory. To maximize performance, network topology, deep learning, and ANN design are currently attracting attention.

Selecting the most important variables or input parameters is a key factor in the improvement of the prediction capability [13]. Feature selection approaches can be generally structured in two main paradigms [14], namely filter and wrapper selection approaches. Unlike other disciplines such as gene engineering, there is the trade-off between feature dimension and the required sensor cost in real-world engine development, and the accessible signals i.e., feature candidates are usually limited but with noises [15]. Filter approaches to feature selection problems rely on an external measure calculated from the data that must be defined to select a subset of features [13]. Mohammad et al. use the least absolute shrinkage and selection operator (Lasso) algorithm to select features used for training emission models of a diesel engine, in which the 37 variables are reduced to 25, 22, 11, and 16 inputs for NOx, CO, HC, and soot emission modeling while maintaining the accuracy [16]. Besides, the superiority of the Lasso algorithm was confirmed on predicting the fuel consumption of ship engines [17], [18]. Koprinska et al. applied different filter feature selection methods to a problem of short-term electricity load from previous samples [19]. In the work of Jurado et al. [20], hybrid methodologies combining filters based on entropy measurements are applied to forecast

[1] This work was supported by the BYD Auto Ltd, Grant: 1001636. (Corresponding authors: Guoxiang Lu and Hongming Xu).

Ji Li, Quan Zhou, Huw Williams, Hongming Xu is with the Department of Mechanical Engineering, the University of Birmingham, Birmingham, U.K. (e-mail: j.li.1@bham.ac.uk; q.zhou@bham.ac.uk; h.williams.5@bham.ac.uk; h.m.xu@bham.ac.uk).

Guoxiang Lu is with Department of New Technology Development, BYD Auto Co Ltd, Guangzhou City, China. (e-mail: lu.guoxiang@byd.com).

TABLE I
SUMMARY OF THE SELECTED STATE OF THE ART IN INDUSTRIAL FEATURE SELECTION APPLICATIONS

| FSP type | Predictor + FS algorithm | Advantages | Disadvantages |
|---|---|---|---|
| Expertise | ANN [5] [11], GNFTL [6], DDPG/ANFIS [7], ANFIS/GPR [8], RA/ANN [10], MLP/MOACO [12] | 1. Physics-explainable model<br>2. Ease to implement<br>3. Less computational effort | 1. Heavily rely on expertise<br>2. Limitations on parameter sensitivity<br>3. Low accuracy |
| Filter | SVM/FFNN + LASSO [16][17][18], NN + AC [19], FIR + Entropy-based FSP [20], SVR + PCA [23], EEMD + CFS [24] | 1. Parameter sensitivity analysis<br>2. No expertise requirement<br>3. Acceptable computational effort | 1. Inefficient removal of redundant features<br>2. Diversity of features selected among filter methods |
| Wrapper | ELM + CRO-SL [15], ELM+PSO [22], MSVR + MA [25] | 1. High accuracy<br>2. No expertise requirement | 1. Tradeoff between optimality and computational effort |
| Hybrid | EGA-CCA [26], SVM+ Pearson Correlation +PSO [27] | 1. High accuracy<br>2. Efficient search<br>3. Parameter sensitivity analysis<br>4. No expertise requirement | 1. Shortcomings in generalization |

Notes: AC: autocorrelation; ANFIS: adaptive neuro-fuzzy inference system; ANN: artificial neural network; CCA: canonical correlation analysis; CRO-SL: coral reefs optimization with substrate layer algorithm; DDPG: deep deterministic policy gradient; EGA: the elitist genetic algorithm; ELM: extreme learning machines; FFNN: feedforward neural network; FIR: fuzzy inductive reasoning; FSP: Feature selection process; GNFTL: Geometric neuro-fuzzy transfer learning; GPR: Gaussian process regression; LASSO: least absolute shrinkage and selection operator; MA: Memetic Algorithm; MLP: multilayer perceptron; MOACO: multi-objective ant colony optimization; MSVR: Multi-output Support Vector Regression; NN: neural network; PSO: particle swarm optimization; RA: regression analysis; SVM: support vector machines; SVR: support vector regression.

hourly energy consumption in buildings, wherein 20% of improvement can be achieved when considering the feature selection step. Usually, filter approaches are faster than wrapper approaches. However, their main drawback is that they completely neglect the effect of the selected feature subset on the performance of the classification/regression algorithm during the search.

Differing from filter approaches, wrapper approaches conduct a search for a good subset of features using the classifier or regressor itself as part of the evaluating function [21]. In the work of Ahila et al. [22], power system disturbances are classified with a hybrid system including extreme learning machines and particle swarm optimization (PSO). In this case, the PSO approach was used to select the best features to serve as inputs of the classifier, and also the number of hidden nodes to enhance the performance of a multi-linear regression algorithm. Hu et al. studied a case of mid-term electricity loads prediction by using a support vector machine and the firefly algorithm for feature selection [23]. Its performance can be better than filter approaches but at the cost of additional computation. Since the iteration processes of these evolutionary algorithms are mainly driven by diverse chaotic maps [24], their convergence process, particularly in feature selection problems, will become slow and unstable. Recently, hybrid filter-wrapper approaches with efficient search have been emerging for industrial practice. Ran et al. developed canonical correlation analysis to detect fault-relevant variables of diesel engines [25]. In the work of Arefnezhad et al., four different filter indexes are combined and evaluated for driver drowsiness detection via an adaptive neuro-fuzzy inference system, where Pearson correlation has the most contribution in feature selection of steering wheel data [26]. Here, advantages and disadvantages of the selected state of the art in industrial feature selection applications are summarized in Table I.

It is clear from the literature study above, that the development of prediction systems for emerging hybrid engines needs to overcome the following research gaps: 1) there is a lack of dedicated machine learning approaches for their performance prediction; 2) existing filter and wrapper approaches to complex feature selection problems i.e., for the hybrid engine are hampered by a compromise between convergence and computing efficiency; and 3) no proper regressor structure and features used for hybrid engine performance prediction are identified in the literature. To systematically address the identified technical challenges, this paper proposes a data-driven approach of statistics-guided accelerated swarm feature selection (SGAS-FS) to find the most effective features for hybrid engine soft sensors. Differing the existing filter or wrapper feature selection approaches, this approach utilizes external measure information to direct velocity updates in the accelerated swarm feature selection optimization. The investigation's three main contributions are:

1) A dedicated machine learning approach i.e., SGAS-FS is proposed with fast convergence to find the most effective features in deep learning of the performance prediction system.
2) A statistics-guided attraction assignment policy is created that uses external measure information to direct velocity updates in the proposed approach.
3) To explore the superior optimization performance, four external measure methods are developed and compared within the proposed approach.

The paper outline is as follows. Section II details the experimental set up carried out in the laboratory showing the features of the measurement equipment. Section III analyzes four types of basis external measures. In Section IV, the proposed solution of SGAS-FS is described, which comprises three main parts of accelerated swarm feature selection, statistics-guided attraction assignment, and network training. Section V carries out a comparative study of the proposed approach. Finally, conclusions are drawn in Section VI.

## II. EXPERIMENTAL SETUP

As shown in Fig. 1, experiments were conducted with an in line 4-cylinder, 1.5 L gasoline engine. The engine specification is shown in Table II. The engine was run under steady state conditions at different operating points that covered the whole engine torque and speed range. To assess the soft sensors in real engine conditions, exhaust gas recirculation (EGR) was performed depending on the engine load at the various running points.


Fig. 1. Dedicated hybrid engine testing bench.

TABLE II
ENGINE SPECIFICATIONS

| Parameter | Value | Unit |
|---|---|---|
| Cylinder number | 4 | – |
| Bore × stroke | 72*92 | mm |
| Displacement | 1498 | $cm^3$ |
| Compression ratio | 15.5 | – |
| Injection system | PFI | – |
| Maximum power | 81/6000 | kW/rpm |
| Maximum torque | 135/4500 | Nm/rpm |
| Torque at maximum power | 129 | Nm |

Engine speed was measured through an AVL encoder with an error of 0.02 Crank Angle Degree (CAD). Engine torque was measured by a SCHENK DYNAS3 dynamometer, with an error of 0.1%. Temperatures were measured with TCA type K thermocouples, with a measurement error of 2%. Gas pressure was measured with KISTLER pressure sensors with an error of 0.3%. Air mass flow rate was measured by means of a Sensycon hot wire anemometer, with a measurement error of 1%.

TABLE III
CONTENDING FEATURES OF THE DEDICATED HYBRID ENGINE

| Parameter | Unit | Parameter | Unit |
|---|---|---|---|
| 1. Speed | rpm | 10. Throttle position | % |
| 2. Torque | Nm | 11. Exhaust pressure | kpa |
| 3. Lambda | - | 12. Intake temperature | ℃ |
| 4. Intake valve timing | CAD | 13. Relative air volume | % |
| 5. Coolant water | ℃ | 14. EGR temperature | ℃ |
| 6. MAT | ℃ | 15. Injection angle | Degree |
| 7. MAP | kpa | Ind. thermal efficiency | % |
| 8. Spark angle | Degree | Volumetric efficiency | % |
| 9.EGR position | % | Fuel consumption rate | g/h |

Considering the diversity of experiment dataset (e.g., system noise and operators), engine experiment is carried out on the two testing benches for the same type dedicated hybrid engine [27] developed by BYD Auto Ltd. 7829 samples were collected at different steady engine load conditions of the operating range of the studied engine. The range of the operating points was as follows: 1000-6000 rpm for engine speed, 1.5 -135 N m for engine torque, and 0–100% for LP EGR positions. Engine speed variables at each steady state point were obtained from the average of 600 points sampled at 10 Hz. There are 15 groups of contending features and 3 groups of soft sensors measured by the test bench have been collected and summarized in Table III to be used for modelling performance prediction system of the hybrid engine.

## III. BASIC EXTERNAL MEASURES

An external measure directly calculated from the database is widely used to select a subset of features, where its performance completely depends on the measure selected for comparing subsets. For computational simplicity and performance diversity, four types of basis external measures used in this study are Spearman rank correlation, principal component analysis (PCA), neighborhood component analysis (NCA), least absolute shrinkage and selection operator (Lasso).

### *A. Spearman Rank Correlation*

The correlation indictor reflects the linear correlation between the observed feature and soft sensors of the hybrid engine. this article uses the spearman correlation coefficient method to define the correlation indicator as

$$\rho_{ij} = \frac{6\sum_{k=1}^{K_i} d_{x_{ij}(k)}^2}{K_i(K_i^2-1)} \tag{1}$$

where $x_{ij}(k)$ is the $k$th value of the $j$th variable in the ith aeroengine sample; and $d_{x_{ij}(k)}^2$ is the difference between ranks for each $x_{ij}(k)$ and $k(k = 1,2,3,\dots,K_i)$. The range of the correlation indicator is between $-1$ and 1. A positive value means the feature is positively correlated with the soft sensors; otherwise, the feature is negatively correlated with the performance indicator.

### *B. Principal Component Analysis*

PCA is a popular and well-known transformation method, and its transformation result is not directly related to a sole feature component of the original sample. The main procedures to perform feature ranking are devised as follows:

By calculating the covariance matrix of PCA using the original training samples, all the eigenvectors and eigenvalues can be obtained. Then the eigenvectors corresponding to the first m largest eigenvalues are selected and denoted by $V_1, V_2, \dots, V_m$. The contribution, to the feature extraction result, of the $j$th feature component can be calculated as follows:

$$c_j = \sum_{p=1}^{m} |V_{pj}| \tag{2}$$

where $V_{pj}$ denotes the $j$th entry of $V_p$, $j = 1,2,\dots,N$, $p = 1,2,\dots,m$. $|V_{pj}|$ stands for the absolute value of $V_{pj}$. Finally, $c_j$ is sorted in the descending order and $d_j$ is used to store the order, where $j = 1,2,\dots,N$.

### *C. Neighborhood Component Analysis*

NCA is a simple and efficient nonlinear decision rule that uses the gradient ascent technique to maximize the expected leave-one-out classification/regression accuracy with a regularization term [28]. For feature selection purposes, NCA is modified by Yang et al. [29] to learn a Mahalanobis distance measure. Given $n$ observations: $S = \{(x_i, y_i), i = 1,2,\dots,n\}$.

The probability $P(\mathrm{Ref}(x) = x_j|S)$ that point $x_j$ is picked from $S$ as the reference point for $x$ is

$$P(\mathrm{Ref}(x) = x_j|S) = \frac{k\left(d_w(x, x_j)\right)}{\sum_{j=1}^{n} k\left(d_w(x, x_j)\right)} \tag{3}$$

Considering leave-one-out cross validation, the data in $S^{-i}$ is used for predicting the response for $x_i$.Thus, the probability that point $x_j$ is picked as the reference point for $x_i$ is

$$p_{ij} = P\left(\text{Ref}(x_i) = x_j \middle| S^{-i}\right) = \frac{k\left(d_w(x_i, x_j)\right)}{\sum_{j=1, j\neq i}^{n} k\left(d_w(x_i, x_j)\right)} \quad (4)$$

The loss function is the average value of $l(y_i, \hat{y}_i)$ that measures the disagreement between the response value, $\hat{y}_i$, and the actual response, $y_i$, for $x_i$.

$$l_i = E\left(l(y_i, \hat{y}_i) \middle| S^{-i}\right) = \sum_{j=1, j\neq i}^{n} p_{ij} l(y_i, y_j) \quad (5)$$

The objective function with the regularization term for minimization is:

$$f(w) = \frac{1}{n}\sum_{i=1}^{n} l_i + \lambda \sum_{r=1}^{P} w_r^2 \quad (6)$$

### D. Lasso Regularization

Tibshirani proposed the least absolute shrinkage and selection operator (Lasso) to estimate the parameters and select features for regression issues [30]. It combines the advantages of ridge regression and subset selection to improve predictive performance and model interpretability. The Lasso is based on the penalized least squares regression and includes L1-penalty function. The optimization factor of LASSO is denoted as $\beta_0$. The formula of $\beta_0$ is presented as:

$$\beta_0 = \arg\min_{\beta} \|\boldsymbol{y}_0 - \boldsymbol{X}_0\beta\|_2^2 + \lambda\|\beta\|_1 \quad (7)$$

where the response vector is denoted as $\boldsymbol{y}_0 = [\boldsymbol{y}_0^1, \boldsymbol{y}_0^2, \ldots, \boldsymbol{y}_0^n]^T$, the matrix of feature is denoted as $\boldsymbol{X}_0 = [x_0^1, x_0^2, \ldots, x_0^n]^T$; $||\cdot||_1$ means the $l_1$ norm; $||\cdot||_2$ means the $l_2$ norm; $\beta, \beta \in R$, $R$ means the real numbers; and $\lambda$ is the trade-off parameter for determinations of the relevant fitting goodness and sparseness of $\beta_0$. The selected features are expressed by the positions of the elements of the sparse regression coefficients with non-zero values.

## IV. Statistics-Guided Accelerated Swarm Feature Selection

In order to precisely predict hybrid engine performance, this paper proposes a novel data-driven approach of SGAS-FS to find the most effective features for deep learning of the hybrid engine performance prediction system. The workflow of the SGAS-FS in the deep learning prediction system is illustrated in Fig. 2. It consists of three main parts: a) the main task of accelerated swarm feature selection; b) the subtask of statistics-guided attraction assignment; and c) the subtask of multi-layer perceptron (MLP) network training.

### A. Accelerated Swarm Feature Selection

To efficiently find the key parameters in performance prediction system of dedicated hybrid engines, accuracy of indicators as the only objective is applied to directly determine the input variables of performance prediction system. Accelerated swarm feature selection inherits widely-used accelerated particle swarm optimization [31] as the main loop of feature selection, and its velocity vector is generated by a simpler formula as

$$v_{i,j+1} = v_{i,j} + \alpha\gamma + \beta\left(g_j^* - x_{i,j}\right) \quad (8)$$

where $x_{i,j}$ is the position of $i$th particle at $j$th iteration, $i \in [1,2,\ldots,O]$ and $j \in [1,2,\ldots,P]$; $O = 20$ particles and $P = 150$ iterations are employed in APSO; $g_j^*$ denotes the current global best; $\gamma$ is drawn from the Gaussian distribution, $N(0,1)$. In order to increase the convergence even further, the update of the location in a single step can be written as

$$x_{i,j+1} = (1-\beta)x_{i,j} + \beta g_j^* + \alpha\gamma \quad (9)$$

This simpler version will give the same order of convergence. It is worth pointing out that the velocity does not appear in Eq. 8), and thus there is no need to deal with the initialization of velocity vectors. Therefore, the APSO is much simpler to implement that PSO. Here the randomization term $\alpha\gamma$ provides the ability for the system to escape any local optimum if $\alpha$ is chosen to be consistent with the scales of the problem of interest. Typically, $\alpha = 0.1L$ to $0.5L$ where $L$ is the scale of each variable, while $\beta = 0.2$ to $0.7$ is sufficient for most applications.

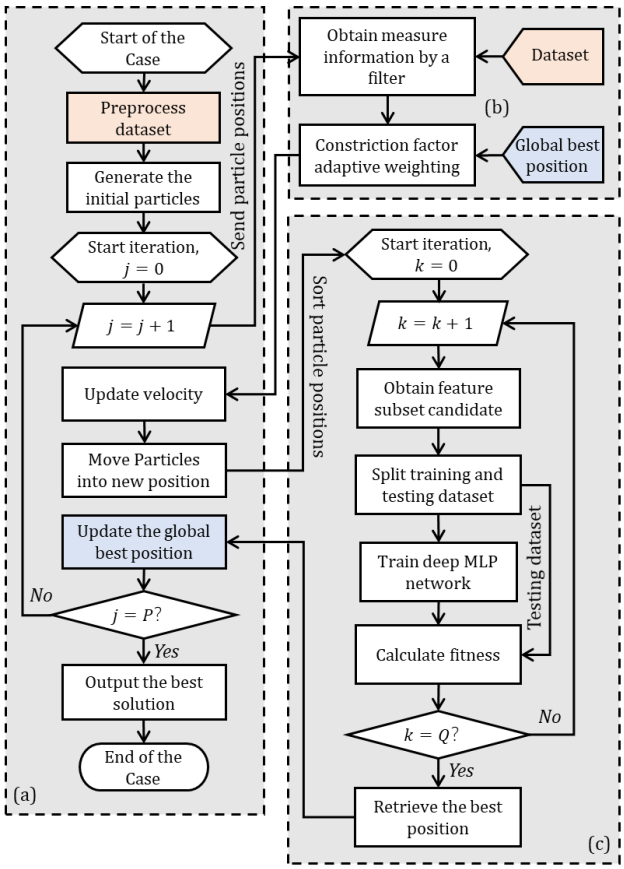


Fig. 2. Workflow of SGAS-FS in the deep learning prediction system: a) the main task of accelerated swarm feature selection; b) the subtask of statistics-guided attraction assignment; and c) the subtask of MLP network training.

A further improvement to the APSO is to reduce the randomness as iterations proceed. This means that a monotonically decreasing function can be implemented, which is given as

$$\alpha_j = \delta^j \quad (10)$$

where $0 < \delta < 1$ is an annealing-like parameter whose value can be taken as 0.7 to 0.9 for most cases. Here $j \in [0, j^*]$ and $j^*$ is the maximum of iterations.

### B. Statistics-Guided Attraction Assignment

Generally, velocity update parameter $\beta$ in APSO and its variants use constant or chaos-map-based attraction assignment policies. In order to adaptively assign attraction forces in the velocity updates, four external measure methods of Spearman rank correlation, PCA, NCA, and Lasso are developed for the accelerated swarm feature selection optimization.

To ensure convergence and optimality of APSO in feature selection, the external measure information is used to direct the velocity updates. Firstly, the feature weights $w$ , calculated by the above measure methods need to be normalized.

$$w_{st} = \frac{w - \min(w)}{\max(w) - \min(w)}, \qquad w_{st} \in [0,1] \tag{11}$$

To determine the features used for the MLP network, the current and global positions of the $i$th particle are sorted by Eq. (12).

$$\left.\begin{aligned} R_{i,j} &= \mathrm{Sort}\left(x_{i,j}\right) \\ R_{g^*} &= \mathrm{Sort}\left(g_j^*\right) \end{aligned}\right\} \tag{12}$$

where, $R_{i,j}$ is the rank of the current position $x_{i,j}$ of the $i$th particle at the $j$th iteration; $R_{g_j^*}$ is the rank of the current global $g_j^*$.

Finally, adaptive weighting is applied to the constriction factor. In the velocity update, the global position has different attractive forces to the current position. These attractive forces are determined and distinguished by the difference of the distance obtained by external measure information.

$$\beta_{i,j} = \frac{w_{st}\left(R_{i,j}\right) - w_{st}\left(R_{g_j^*}\right)}{2} + \frac{1}{2}, \qquad \beta_i \in [0,1] \tag{13}$$

*Convergence Analysis:* To ensure the convergence of the optimization process, convergence analysis is carried out. The theory of linear, discrete-time dynamic system provides a powerful set of tools and results for assessing the dynamic behaviour of a particle [32]. Using a dynamical approach with ignoring the randomness term and setting $u = g^* - x_i^t$ , a system can be simplified as

$$v_i^{t+1} = v_i^t + \beta u_i^t, \qquad u_i^{t+1} = (1-\beta)u_i^t \tag{14}$$

which can be re-written as

$$Y_{t+1} = AY_t, \qquad A = \begin{bmatrix} 1 & \beta \\ 0 & 1-\beta \end{bmatrix} \tag{15}$$

Then the characteristic equation is

$$A - \lambda \cdot I = \left| \begin{bmatrix} 1 & \beta \\ 0 & 1-\beta \end{bmatrix} - \begin{bmatrix} \lambda & 0 \\ 0 & \lambda \end{bmatrix} \right| = 0 \tag{16}$$

can be further calculated as follows

$$\lambda^2 + (\beta - 2)\lambda + 1 - \beta = 0 \tag{17}$$

and the two eigenvalues are

$$\begin{cases} \lambda_1 = 1 \\ \lambda_2 = 1 - \beta \end{cases} \tag{18}$$

Different from several conditions when solving for PSO eigenvalues, its eigenvalues are simply 1 and $1-\beta$. From theory of dynamical systems, the above iterations will always converge stably for any $\beta > 0$. In fact, the experiment of Gandomi et al. also confirm that convergence behavior of this paradigm observed is fast and stable [33].

### C. MLP Network and Cost Function

The last subtask in Fig. 2(c) is MLP network training. In this study, the widely used MLP network is adopted for predictive modeling of the hybrid engine soft sensors. As one of the popular deep learning structures, it is useful for solving problems stochastically, which often allows approximate solutions for extremely complex problems like fitness approximation.

Commonly, a learning algorithm owns a single objective (approximation error minimization) that is often achieved by minimizing the mean-squared error (MSE) on the learning data.

$$\mathrm{MSE} = \frac{1}{N}\sum_{i=1}^{N}(d_i - y_i)^2 \tag{19}$$

where $d$ and $y$ are the desired and the model's outputs, respectively, and $N$ is the number of data pairs in the training set. The original experimental dataset is categorized into 80% of training data and 20% of testing data via uniformly random sampling.

For preventing overestimation of the network fitness, the cost function is designed with a weighting factor $\rho = 0.2$ to balance MSE weights for training and testing data. To reduce the influence caused by network training bias, each training task will be repeated $Q = 5$ times, and its mean value will be fed back to the main task of accelerated swarm feature selection as the value of cost function. Commonly, the repeatability work can be reduced as dimension of candidate features further increased in different cases. So far, the cost function can be expressed as follows.

$$\min(\mathrm{MSE}^*) = \frac{1}{Q}\sum_{i=1}^{Q}\left(\rho \mathrm{MSE}_{\mathrm{i,train}} + (1-\rho)\mathrm{MSE}_{\mathrm{i,test}}\right) \tag{20}$$

In this paper, the selection of the best phenotype in a single objective training was solely based on a comparison of the MSEs. Once the terminal condition is reached, iterations will cease. The final hyperparameters of an MLP network will be fixed for function validation and testing at the end.

## V. Results and Discussion

Henceforward, a comprehensive comparative study is carried out from the four aspects of 1) attraction assignment policies; 2) feature selection approaches; 3) network structure and system complexity; and 4) adaptability of the proposed approach to other engine soft sensors.

### A. Evaluation on Attraction Assignment Policies

Attraction assignment policy is expectedly used to update the particle velocity which directly affects the convergence speed and final fitness of accelerated swarm feature selection optimization. For a fair comparison, all eight attraction assignment policies are embedded in the accelerated swarm feature selection with the same parameter setting. In addition, the MLP network used here has two hidden layers (in which each layer has 30 neurons using hyperbolic tangent activation functions) and is fully connected.

Fig. 3 shows the iteration process of using different attraction assignment policies under 150-episode observation. For each case, 20 repetitive experiments were performed, and the median value of MSE and the corresponding feature rank were recorded. Overall, three statistics-guided policies exhibit superior performance on both convergence and final fitness

compared to constant or map inspired policies. Lasso-guided policy performs worse convergence contrasted to other statistics-guided them. The effect of each chaotic map is quite distinct from the constant one. The authors think the reason behind this is that chaotic maps are derived from a random number, with no guidance from other measures. In terms of proposed statistics-guided policies, the rapid convergence results in a fitness gap (i.e., MSE) with other non-statistics-guided policies at the 25th iteration. Furthermore, the policy inspired by NCA continues to show better fitness after the 25th iteration up to the 70th iteration.

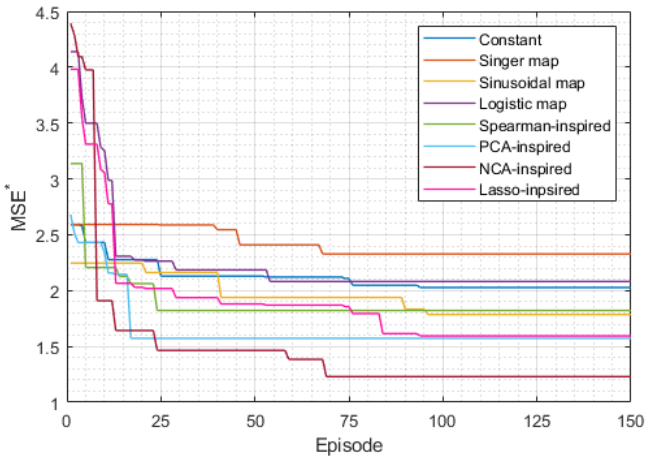


Fig. 3. Iteration process comparison of using different attraction assignment policies.

TABLE IV
FEATURE RANK AND REGRESSION PERFORMANCE OF USING SEVEN ATTRACTION ASSIGNMENT POLICIES

| Attraction assignment policy | Feature rank | Evaluation metrics | |
|---|---|---|---|
| | | MSE | $R^2$ |
| Constant [34] | [1 2 3 13 12] | 2.0267 | 0.9213 |
| Singer map [35] | [14 9 1 8 13] | 2.3284 | 0.8945 |
| Sinusoidal map [36] | [8 3 1 13 14] | 1.7860 | 0.9330 |
| Logistic map [37] | [10 3 2 8 12] | 2.0809 | 0.9113 |
| Spearman-guided | [12 11 8 3 2] | 1.8217 | 0.9277 |
| PCA-guided | [1 3 15 13 2] | 1.5728 | 0.9322 |
| NCA-guided | [3 8 12 1 11] | 1.2292 | 0.9520 |
| Lasso-guided | [4,1,8,6,9] | 1.5944 | 0.9382 |

Table IV organizes the feature top five rank and regression performance obtained by the studied seven attraction assignment policies. Compared to averaged MSE (2.07) obtained by map-guided policies, the averaged MSE (1.55) obtained by statistics-guided policies has a significant reduction of 25.1%. Moreover, the coefficient of determination $R^2$ can be improved by 2.46%. To quantitatively evaluate the contribution of each feature on regression performance, an index of the contribution score is designed and expressed as follows:

$$\text{Score}_i = \sum_{j=1}^{J} \frac{R_{i,j}}{\text{MSE}_j} \tag{21}$$

where $R_{i,j}$ is the rank score of $i$th feature in $j$th attraction assignment policy, in which the first place gets 5 scores, the fifth place gets 1 score; $\text{MSE}_j$ is the mean-squared error obtained by using $j$th attraction assignment policy.

Fig. 4 illustrates the results of contribution scores calculated by Eq. (17). Features 5 (Coolant water) and 7 (MAP) have no contribution to the regression task of the MLP network. Features 4 (Intake valve timing), 6 (MAT), and 15 (Injection angle) are adopted only by one policy. The top five features that contribute the most to the studied regression task are Feature 1, 2, 3, 8, and 12. They are Speed, Torque, Lambda, Spark angle, and intake temperature, where features Speed, Lambda, and Spark angle are selected by five policies at the same time, whereas the feature Lambda has the highest contribution score of the studied fifteen contending features.

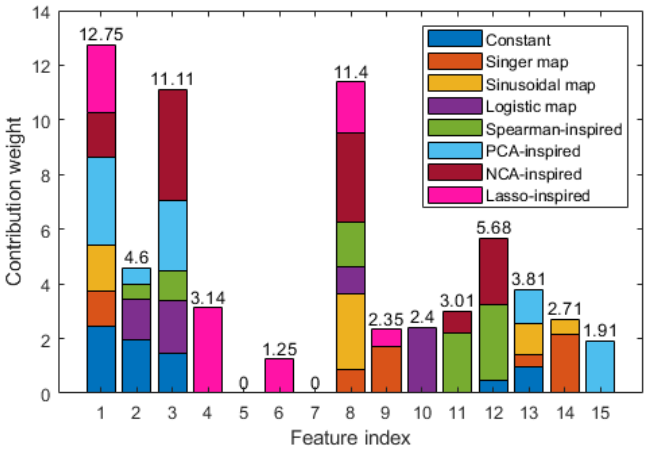


Fig. 4. Contribution scores of the studied fifteen contending features

### *B. Comparison of Feature Selection Approaches*

This section further examines the NCA-guided accelerated swarm feature selection approach elected from the last section with other feature selection approaches. Filter approaches with four types of external measures (Spearman, PCA, NCA, and Lasso) and wrapper approaches with three optimization algorithms (SA, PSO, and ACO) were used to compare with the NCA-guided accelerated swarm feature selection approach. Differing from external measures used in the attraction assignment policy as discussed in the last section, here, they are used to solely select the features before MLP network training.

Table V summarizes the feature rank and regression performance of using seven feature selection approaches under EGR on/off conditions. For each case, 20 repetitive experiments were performed, and the median value of MSE and the corresponding indexes were recorded. From a view of feature selection approaches, the proposed hybrid feature selection approach i.e., NCA-guided accelerated particle swarm optimization algorithm achieves much better regression performance in training MLP networks than the filter or wrapper approaches i.e., other six algorithms. For both EGR on/off conditions, MSE can be decreased at least by 54.3% (from PSO) and $R^2$ can be increased at least by 2.26% (from SA). Compared with the filter approaches, the performance of the wrapper approaches is closer to the performance of the proposed hybrid approaches. This is due to the interactive mechanism of heuristic exploration and real-time evaluation, rather than the segmented mechanism of independent statistical analysis and evaluation. Since the feature rank defined by Spearman correlation has a higher degree of coincidence to that defined by the proposed one, its MSE is significantly lower than those of using the other filter algorithms. From a view of the

TABLE V
FEATURE RANK AND REGRESSION PERFORMANCE OF USING SEVEN FEATURE SELECTION APPROACHES

| FS approach | Execution algorithm | Feature rank | | EGR on | | EGR off | | EGR on/off | |
|---|---|---|---|---|---|---|---|---|---|
| | | EGR on | EGR off | MSE | $R^2$ | MSE | $R^2$ | MSE | $R^2$ |
| Filter | Spearman | [15 3 14 1 11] | [3 8 2 4 1] | 3.0516 | 0.8867 | 0.5172 | 0.9894 | 1.7844 | 0.9381 |
| | PCA | [3 4 6 5 9] | [4 6 3 5 10] | 4.4358 | 0.8416 | 5.5248 | 0.8762 | 4.9803 | 0.8589 |
| | NCA | [9 10 8 15 1] | [4 3 6 12 11] | 4.0743 | 0.8973 | 4.5361 | 0.8914 | 4.3052 | 0.8944 |
| | Lasso | [4,1,8,6,9] | [4,3,9,5,13] | 4.6610 | 0.8205 | 5.9122 | 0.8491 | 5.2424 | 0.8307 |
| Wrapper | SA | [13 2 1 7 3] | [1 12 2 3 7] | 2.7327 | 0.9155 | 0.3373 | 0.9921 | 1.5350 | 0.9538 |
| | PSO | [15 9 1 8 12] | [3 12 8 9 1] | 3.4963 | 0.8827 | 0.2887 | 0.9954 | 1.8925 | 0.9391 |
| | ACO | [13 1 8 3 12] | [13 3 1 8 2] | 2.6184 | 0.9113 | 0.434 | 0.9902 | 1.5262 | 0.9508 |
| Hybrid | Proposed | [3 8 12 1 11] | [6 12 1 2 3] | 1.2292 | 0.9520 | 0.1664 | 0.9987 | 0.6978 | 0.9754 |
| Overall | Overall | [3 1 8 13 15] | [12 3 1 6 2] | 3.0912 | 0.8982 | 1.6864 | 0.9619 | 2.3888 | 0.9300 |

EGR on/off condition, the EGR-off condition would restrict engine dynamic characteristics so that contrasted to the EGR-on condition, the difficulty of indicated thermal efficiency model approximation is relatively low. Overall, averaged MSE and $R^2$ in all studied regression tasks of the MLP network can be separately decreased by 45.4% and increased by 7.09% when EGR is off.

TABLE VI
REPEATABILITY INDEX OF FEATURE RANKS UNDER THREE EGR CONDITIONS

| Execution algorithm | Repeatability index | | |
|---|---|---|---|
| | EGR on | EGR off | EGR on/off |
| SA | 12/20 | 16/20 | 13/20 |
| PSO | 15/20 | 15/20 | 14/20 |
| ACO | 16/20 | 17/20 | 16/20 |
| Proposed | 16/20 | 20/20 | 18/20 |

In order to analyze stability of the feature ranks obtained by the wrapper and hybrid approaches, a new index i.e., repeatability is introduced which describes the number of times when feature ranks match the one with the lowest MSE in 20 repetitive experiments. As shown in Table VI, the repeatability index calculated based on the proposed SGAS-FS approach takes the first place in other three wrapper approaches in all EGR conditions. Because the guidance of the particles by the filter reduces their own randomness, the results of feature ranks become stabler.

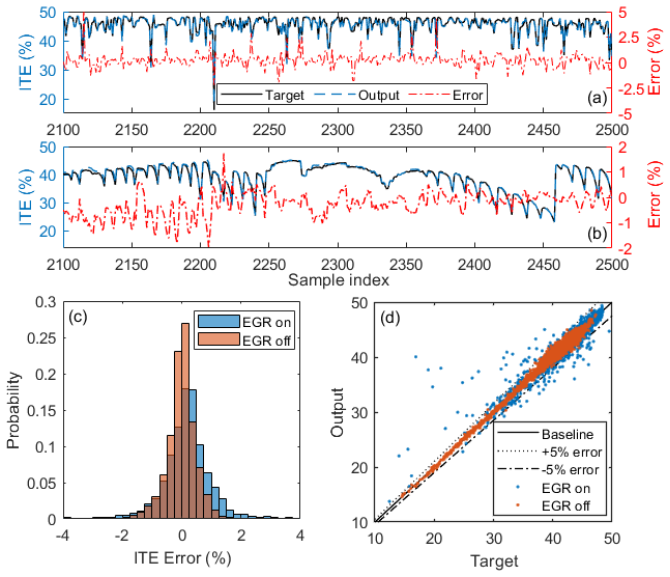


Fig. 5. Indicated thermal efficiency prediction performance comparison under EGR on/off conditions by using the proposed SGAS-FS approach: a) prediction performance under EGR-on condition; b) prediction performance under EGR-off condition; c) error distribution; and d) regression performance.

Fig. 5 displays the indicated thermal efficiency prediction performance when using the MLP network trained by the proposed hybrid feature selection approach. Fig. 5(a) and (b) present the prediction performance segment for sample indexes [2100,2500] under both EGR-on/off conditions. From the results, the maximum error of the EGR-on condition is greater than that of the EGR-off condition. The same conclusion can be found in Fig. 5(c), where the error distribution of the EGR-off condition is narrower than that of the EGR-on condition. In conjunction with Fig. 5(d), more samples are found in the outer of $\pm 5\%$ error lines under the EGR-on condition. In addition to the uncertainty caused by EGR opening, the authors believe that the relatively concentrated distribution [30,50] of EGR-on condition samples may also be the reason for increasing the difficulty of network regression.

### *C. Network Structure and Computational Complexity*

In order to implement the indicated thermal efficiency prediction system optimized by the proposed approach in practice, system complexity and computational effectiveness must be carefully considered. selected feature number is the main factor to reflect the indicated thermal efficiency prediction system complexity and applicability. Furthermore, computational effectiveness of the indicated thermal efficiency modelling is also limited by its network structure, e.g., the number of the hidden layers. Thus, this section investigates the impact of both feature and hidden layer numbers on prediction accuracy as well as the whole calculation time. For fair comparison, the total number of neurons used for hidden layers is fixed to 60. The neurons used for each hidden layer are equal and those are determined based on the number of hidden layers. For instant, the network structure with three hidden layers has 20 neurons for each.

The computational complexity of the proposed approach's improvements mainly depends on three aspects: 1) random initialization; 2) particle velocity and position updating; and 3) neural network implementation. The first two parts can all be expressed as $O(n_1 \times n_2 \times n_3)$ by the Big O notation as analyzed in [38], in that, $n_1$ is the population, $n_2$ is the problem dimensions, and $n_3$ is the main iterations. The last part includes forward propagation and backpropagation which can all be expressed as $O(n_4^3 \times n_5 \times n_6)$ by the Big O notation as analyzed in [39], in that, $n_4^3$ is an asymptotic run-time of naive matrix multiplication, $n_5$ is the number of neurons, and $n_6$ is the gradient iterations. In this paper, the neural network implementation and APSO algorithm's initialization process have not been changed, so we only compare the time complexity from particle velocity and position updating.

Compared with APSO-based feature selection, the additional time complexity that should be considered is from constriction factor adaptive weighting in particle velocity and position updating. In fact, this action has not changed the loop body of APSO algorithm, and its time complexity is still $O(n_1 \times n_2 \times n_3)$. Overall, the complexity of APSO-based feature selection and its improved one is not increased by orders of magnitude.

TABLE VII
NETWORK FITNESS (MSE) PERFORMANCE OF USING A VARIOUS NUMBER OF FEATURES AND HIDDEN LAYERS

| Number of features | Number of hidden layers | | | | | |
|---|---|---|---|---|---|---|
| | 1 | 2 | 3 | 4 | 5 | 6 |
| 1 | 11.80 | 11.02 | 8.56 | 8.64 | 8.72 | 8.79 |
| 2 | 7.48 | 6.83 | 5.00 | 5.95 | 5.23 | 7.30 |
| 3 | 4.65 | 4.64 | 3.04 | 3.05 | 3.12 | 3.14 |
| 4 | 2.33 | 2.05 | 1.98 | 2.56 | 2.53 | 3.67 |
| 5 | 1.78 | 1.23 | 1.66 | 2.00 | 2.40 | 1.92 |
| 6 | 1.49 | 1.37 | 1.46 | 1.69 | 2.14 | 1.90 |
| 7 | 1.45 | 1.47 | 1.46 | 1.55 | 1.80 | 1.30 |
| 8 | 1.43 | 1.47 | 1.41 | 1.48 | 1.42 | 1.60 |

Table VII and VIII organize the feature selection result (i.e., MSE and real calculation time) of MLP networks with varying numbers of features and hidden layers. For each case, 20 repetitive experiments were performed, and the minimum value of MSE and the corresponding feature selection time were recorded. Generally, the value of MSE decreases with the increase of feature numbers or hidden layer numbers. In the studied scale, the decrease of MSE value within the top 4 features and the top 3 layers were significant compared to those within the outer range. The lowest value of MSE (1.23) occurs when using the network structure with 5 features and 2 hidden layers (each layer has 30 neurons). With the increase of hidden layer numbers, the feature selection time has increased sharply (2 hidden layers) and then dropped fast (3 hidden layers) and eventually stabilized. When the number of hidden layers is set to 1, the feature number if $\leq 4$ has no effect on feature selection time. It should be noted that when the number of hidden layers is 2, the computational effectiveness is the lowest (wherein feature selection time is 13130 s to select only one feature) than those of the other numbers of hidden layers.

TABLE VIII
FEATURE SELECTION TIME ($10^3$ S) OF USING A VARIOUS NUMBER OF FEATURES AND HIDDEN LAYERS

| Number of features | Number of hidden layers | | | | | |
|---|---|---|---|---|---|---|
| | 1 | 2 | 3 | 4 | 5 | 6 |
| 1 | 0.582 | 13.130 | 2.561 | 2.558 | 2.009 | 2.174 |
| 2 | 0.582 | 5.471 | 2.589 | 2.295 | 1.926 | 1.787 |
| 3 | 0.582 | 3.371 | 2.626 | 2.159 | 1.901 | 1.715 |
| 4 | 0.584 | 4.247 | 3.029 | 2.286 | 2.025 | 1.814 |
| 5 | 0.878 | 4.455 | 3.132 | 2.468 | 2.017 | 1.752 |
| 6 | 0.874 | 5.035 | 3.320 | 2.536 | 2.023 | 1.944 |
| 7 | 0.995 | 6.620 | 3.606 | 2.857 | 2.133 | 1.838 |
| 8 | 1.126 | 8.776 | 3.818 | 2.799 | 2.143 | 2.383 |

With rapid development of informatics, real-time modelling, where data capture and modelling of a physical plant can execute at the same time, becomes a promising production solution to build surrogate models with timeliness for better real-world performance [40]. Under this condition, the number of interactions with a physical plant as defined in [41] is a key factor to reflect experimental cost. The higher the number of interactions with the physical plant, the higher the experiment cost that might be caused by operation, labor, and ageing of the physical plant.

Fig. 6 evaluates the energy-saving potential of the proposed SGAS-FS approach under the real-time modelling condition. As an increase in the number of features selected, the number of calls of all studied sensors for both brute-force search and the proposed approaches is all recorded. In order to find key features for each number of the features selected, the total number of times studied sensors was called via brute-force search is 245760 (full permutation), while that via the proposed approach is only 60600. By using the proposed approach, the number of times sensors was called is showing slow linear growth. Relatively, the number of times sensors was called by using the brute-force search over the selected feature number follows a normal distribution. The total number of times studied sensors was called reaches the top (51480) when the selected feature number is 8. Therefore, the proposed SGAS-FS approach has a great energy-saving potential in real production process, which only has 25% of physical interactions by brute-force search that could save the experiment cost caused by operation, labor, and ageing of the physical plant.

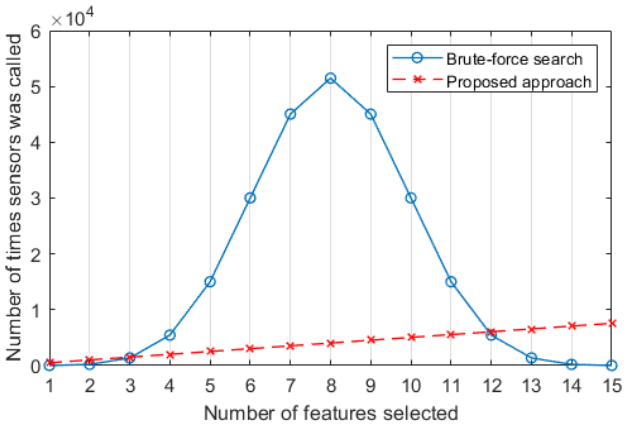

Fig. 6. The number of calls of all studied sensors in real-time modelling

## D. Adaptability to Other Engine Soft Sensors

This section discusses the adaptability of the proposed SGAS-FS approach to other engine soft sensors. In addition to indicated thermal efficiency, volumetric efficiency, and fuel consumption rate are comprehensively examined in terms of regression performance, feature contribution, and error distribution. For each case, 20 repetitive experiments were performed, and the median value of MSE and the corresponding feature rank were recorded.

TABLE IX
REGRESSION PERFORMANCE OF USING THE PROPOSED APPROACH FOR DIFFERENT ENGINE SOFT SENSORS

| FS approach | Performance indicator | Feature rank | Evaluation metrics | |
|---|---|---|---|---|
| | | | MSE | $R^2$ |
| Proposed | ITE | [3 8 12 1 11] | 0.6978 | 0.9754 |
| | VE | [14 9 1 8 13] | 1.3284 | 0.8945 |
| | FB rate | [8 3 1 13 14] | 0.7860 | 0.9530 |

Table IX organizes the feature top five rank and regression performance for three studied engine soft sensors. Generally, the proposed approach performs excellent regression performance (with up to MSE:1.3284) on these soft sensors.

Compared to the features selected for modelling fuel consumption rate, the features selected for modelling volumetric efficiency are more similar with ones selected for modelling indicated thermal efficiency. This could be caused by the strong correlation between indicated thermal efficiency and fuel consumption rate.

Fig. 7 illustrates the results of contribution scores calculated by Eq. (21). Features 2 (Torque), 4 (Intake valve timing), 5 (Coolant water), 6 (MAT),7 (MAP), 10 (Throttle position), and 15 (Injection angle) have no contribution to the regression task of the MLP network. The top three features that contribute the most to the studied regression task are Features 1, 3, and 8, where they are Speed, Lambda, and Spark angle, whereas the feature Spark angle has the highest contribution score of the studied fifteen contending features.

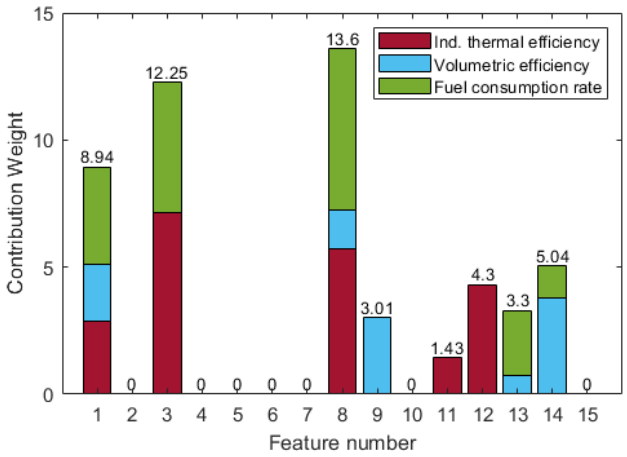


Fig. 7. Feature contribution in modelling of indicated thermal efficiency, volumetric efficiency, and fuel consumption rate

To further investigate error distribution of the deep learning prediction system established by only using the identified three key features i.e., Speed, Lambda, and Spark angle, the averaged relative error of deep learning prediction systems designed for three soft sensors are summarized in Table X. Generally, the proposed SGAS-FS approach has good adaptability to other engine where the averaged relative error is still at a lower level, 3.17%. From a view of engine torque, the averaged relative error in the low torque area (0-50 Nm) is larger than the error in the higher torque areas (>50 Nm). The authors believe the reason is that the studied three soft sensors are greatly affected by the speed in the low torque area which increases regression difficulty of the proposed approach.

TABLE X
ERROR DISTRIBUTION OF USING THE PROPOSED APPROACH

| Speed | Torque | | | | |
|---|---|---|---|---|---|
| | 0-50 | 50-75 | 75-100 | 100-135 | Overall |
| 0-2000 | 8.55% | 1.58% | 5.33% | 0.90% | 4.55% |
| 2000-3000 | 2.90% | 0.87% | 0.80% | 0.75% | 1.17% |
| 3000-4000 | 9.56% | 1.52% | 1.00% | 0.79% | 2.92% |
| 4000-5000 | 10.68% | 2.52% | 0.90% | 1.42% | 3.32% |
| 5000-6000 | 7.85% | 4.39% | 3.72% | 1.87% | 4.41% |
| Overall | 7.93% | 1.99% | 2.22% | 1.12% | 3.17% |

## VI. CONCLUSIONS

This paper proposes the novel data-driven approach of SGAS-FS for hybrid engine soft sensors that has an ability to find the most effective features for engine soft sensors. Validated by bench experiment, the proposed approach has been comprehensively evaluated in four aspects of: 1) attraction assignment policies; 2) feature selection approaches; 3) network structure and system complexity; and 4) adaptability to other engine soft sensors. The conclusions drawn from the investigation are as follows:

1) Compared to averaged MSE (2.07) obtained by the map-guided approach, the averaged MSE (1.54) obtained by statistics-guided policies has a significant reduction of 25.6%, in which the NCA-guided method is the biggest contributor.
2) When EGR is off, the averaged MSE and $R^2$ in all studied regression tasks of the MLP network can be separately decreased by 45.4% and increased by 7.09%.
3) The optimal network structure with the lowest value of MSE (1.23) for indicated thermal efficiency modelling of the studied hybrid engine has been detected which has 5 features and 2 hidden layers (each layer has 30 neurons).
4) The proposed SGAS-FS approach has a great energy-saving potential in real production process, which only has 25% of physical interactions by brute-force search under the real-time modelling condition.
5) The proposed approach has good adaptability to the studied engine soft sensors i.e., indicated thermal efficiency, volumetric efficiency, and fuel consumption rate, where the three key features of Speed, Lambda, and Spark angle are identified which results in the averaged relative error by 3.17%.

## VII. LIMITATION AND FUTURE WORK

The work presents a holistic solution to efficiently find the key parameters for hybrid engine soft sensors. However, it is only a first step and neglects some aspects of a complete solution, and these will be the subject of future work. In terms of the generalization ability of the proposed approach, it shows great potential to be applied to other engine parameters, different engine configurations, and biofuels with various physicochemical characteristics. But those need to be further investigated. Besides, reducing experimental cost of map calibration by using transient data is an alternative way to collaboratively improve the experimental efficiency of this approach. These all are worthy to be studied in the future work.

## REFERENCES

[1] Q. Sun and Z. Ge, "A Survey on Deep Learning for Data-Driven Soft Sensors," *IEEE Trans. Ind. Informatics*, vol. 17, no. 9, pp. 5835–5866, 2021, doi: 10.3934/qfe.2021032.

[2] Y. Jiang, S. Yin, J. Dong, and O. Kaynak, "A Review on Soft Sensors for Monitoring, Control, and Optimization of Industrial Processes," *IEEE Sens. J.*, vol. 21, no. 11, pp. 12868–12881, 2021, doi: 10.1109/JSEN.2020.3033153.

[3] W. Wen, S. Yang, P. Zhou, and S. Z. Gao, "Impacts of COVID-19 on the electric vehicle industry: Evidence from China," *Renew. Sustain. Energy Rev.*, vol. 144, no. January, p. 111024, 2021, doi: 10.1016/j.rser.2021.111024.

[4] P. Kadlec, B. Gabrys, and S. Strandt, "Data-driven Soft Sensors in the process industry," *Comput. Chem. Eng.*, vol. 33, no. 4, pp. 795–814, 2009, doi: 10.1016/j.compchemeng.2008.12.012.

[5] A. N. Bhatt and N. Shrivastava, "Application of Artificial Neural Network for Internal Combustion Engines: A State of the Art Review,"

*Arch. Comput. Methods Eng.*, no. 0123456789, 2021, doi: 10.1007/s11831-021-09596-5.

[6] J. Li, D. Wu, M. Attar, and H. Xu, "Geometric Neuro-Fuzzy Transfer Learning for In-Cylinder Pressure Modelling of a Diesel Engine Fuelled with Raw Microalgae Oil," *Appl. Energy*, 2021.

[7] Q. Zhou, D. Zhao, B. Shuai, Y. Li, H. Williams, and H. Xu, "Knowledge Implementation and Transfer With an Adaptive Learning Network for Real-Time Power Management of the Plug-in Hybrid Vehicle," *IEEE Trans. Neural Networks Learn. Syst.*, vol. PP, pp. 1–11, 2021, doi: 10.1109/TNNLS.2021.3093429.

[8] Q. Zhou *et al.*, "Transferable Representation Modelling for Real-time Energy Management of the Plug-in Hybrid Vehicle based on K-fold Fuzzy Learning and Gaussian Process Regression," *Appl. Energy*, 2021.

[9] J. Li, Z. Li, Q. Zhou, Y. Zhang, and H. Xu, "Improved scheme of membership function optimisation for fuzzy air-fuel ratio control of GDI engines," *IET Intell. Transp. Syst.*, vol. 13, no. 1, pp. 209–217, 2018, doi: 10.1049/iet-its.2018.5013.

[10] E. Tosun, K. Aydin, S. S. Merola, and A. Irimescu, "Estimation of operational parameters for a direct injection turbocharged spark ignition engine by using regression analysis and artificial neural network," *Therm. Sci.*, vol. 21, no. 1, pp. 401–412, 2017, doi: 10.2298/TSCI160302151T.

[11] Y. Jo, K. Min, D. Jung, M. Sunwoo, and M. Han, "Comparative study of the artificial neural network with three hyper-parameter optimization methods for the precise LP-EGR estimation using in-cylinder pressure in a turbocharged GDI engine," *Appl. Therm. Eng.*, vol. 149, no. December 2018, pp. 1324–1334, 2019, doi: 10.1016/j.applthermaleng.2018.12.139.

[12] J. Martínez-Morales, H. Quej-Cosgaya, J. Lagunas-Jiménez, E. Palacios-Hernández, and J. Morales-Saldaña, "Design optimization of multilayer perceptron neural network by ant colony optimization applied to engine emissions data," *Sci. China Technol. Sci.*, vol. 62, no. 6, pp. 1055–1064, 2019, doi: 10.1007/s11431-017-9235-y.

[13] S. Salcedo-Sanz, L. Cornejo-Bueno, L. Prieto, D. Paredes, and R. García-Herrera, "Feature selection in machine learning prediction systems for renewable energy applications," *Renew. Sustain. Energy Rev.*, vol. 90, no. April, pp. 728–741, 2018, doi: 10.1016/j.rser.2018.04.008.

[14] B. Xue, M. Zhang, W. N. Browne, and X. Yao, "A Survey on Evolutionary Computation Approaches to Feature Selection," *IEEE Trans. Evol. Comput.*, vol. 20, no. 4, pp. 606–626, 2016, doi: 10.1109/TEVC.2015.2504420.

[15] M. Aliramezani, C. R. Koch, and M. Shahbakhti, "Modeling, diagnostics, optimization, and control of internal combustion engines via modern machine learning techniques: A review and future directions," *Prog. Energy Combust. Sci.*, vol. 88, no. September 2021, p. 100967, 2022, doi: 10.1016/j.pecs.2021.100967.

[16] M. Mohammad, A., Rezaei, R., Hayduk, C., Delebinski, T., Shahpouri, S., & Shahbakhti, "Physical-oriented and machine learning-based emission modeling in a diesel compression ignition engine: Dimensionality reduction and regression," *Int. J. Engine Res.*, 2022, [Online]. Available: https://doi.org/10.1177/14680874211070736.

[17] S. Wang, B. Ji, J. Zhao, W. Liu, and T. Xu, "Predicting ship fuel consumption based on LASSO regression," *Transp. Res. Part D Transp. Environ.*, vol. 65, no. October 2017, pp. 817–824, 2018, doi: 10.1016/j.trd.2017.09.014.

[18] C. Gkerekos, I. Lazakis, and G. Theotokatos, "Machine learning models for predicting ship main engine Fuel Oil Consumption: A comparative study," *Ocean Eng.*, vol. 188, no. January, p. 106282, 2019, doi: 10.1016/j.oceaneng.2019.106282.

[19] I. Koprinska, M. Rana, and V. G. Agelidis, "Correlation and instance based feature selection for electricity load forecasting," *Knowledge-Based Syst.*, vol. 82, pp. 29–40, 2015, doi: 10.1016/j.knosys.2015.02.017.

[20] S. Jurado, À. Nebot, F. Mugica, and N. Avellana, "Hybrid methodologies for electricity load forecasting: Entropy-based feature selection with machine learning and soft computing techniques," *Energy*, vol. 86, pp. 276–291, 2015, doi: 10.1016/j.energy.2015.04.039.

[21] G. H. John, R. Kohavi, and K. Pfleger, "Irrelevant Features and the Subset Selection Problem," in *Machine Learning Proceedings 1994*, 1994, pp. 121–129, doi: 10.1016/b978-1-55860-335-6.50023-4.

[22] R. Ahila, V. Sadasivam, and K. Manimala, "An integrated PSO for parameter determination and feature selection of ELM and its application in classification of power system disturbances," *Appl. Soft Comput. J.*, vol. 32, pp. 23–37, 2015, doi: 10.1016/j.asoc.2015.03.036.

[23] Z. Hu, Y. Bao, R. Chiong, and T. Xiong, "Mid-term interval load forecasting using multi-output support vector regression with a memetic algorithm for feature selection," *Energy*, vol. 84, pp. 419–431, 2015, doi: 10.1016/j.energy.2015.03.054.

[24] H. Lu, X. Wang, Z. Fei, and M. Qiu, "The effects of using chaotic map on improving the performance of multiobjective evolutionary algorithms," *Math. Probl. Eng.*, vol. 2014, 2014, doi: 10.1155/2014/924652.

[25] Q. Ran, Y. Song, W. Du, W. Du, and X. Peng, "Fault detection of diesel engine air and after-treatment systems with high-dimensional data: A novel fault-relevant feature selection method," *Processes*, vol. 9, no. 2, pp. 1–15, 2021, doi: 10.3390/pr9020259.

[26] S. Arefnezhad, S. Samiee, A. Eichberger, and A. Nahvi, "Driver drowsiness detection based on steering wheel data applying adaptive neuro-fuzzy feature selection," *Sensors (Switzerland)*, vol. 19, no. 4, 2019, doi: 10.3390/s19040943.

[27] J. Li, Q. Zhou, H. Williams, P. Xu, H. Xu, and G. Lu, "Fuzzy-tree-constructed data-efficient modelling methodology for volumetric efficiency of dedicated hybrid engines," *Appl. Energy*, vol. 310, no. December 2021, p. 118534, 2022, doi: 10.1016/j.apenergy.2022.118534.

[28] R. R. Goldberger, J., Hinton, G. E., Roweis, S., & Salakhutdinov, "Neighbourhood components analysis," *Adv. Neural Inf. Process. Syst.*, vol. 17, 2004.

[29] W. Yang, K. Wang, and W. Zuo, "Neighborhood component feature selection for high-dimensional data," *J. Comput.*, vol. 7, no. 1, pp. 162–168, 2012, doi: 10.4304/jcp.7.1.161-168.

[30] R. Shrinkage, "Regression Shrinkage and Selection via the Lasso," *J. R. Stat. Soc. Ser. B*, vol. 58, no. 1, pp. 267–288, 1996.

[31] X. S. Yang, *Engineering Optimization: An Introduction with Metaheuristic Applications*. John Wiley & Sons, 2010.

[32] M. Clerc and J. Kennedy, "The particle swarm-explosion, stability, and convergence in a multidimensional complex space," *IEEE Trans. Evol. Comput.*, vol. 6, no. 1, pp. 58–73, 2002, doi: 10.1109/4235.985692.

[33] A. Hossein, G. Jin, X. Yang, and S. Talatahari, "Chaos-enhanced accelerated particle swarm optimization," *Commun. Nonlinear Sci. Numer. Simul.*, vol. 18, no. 2, pp. 327–340, 2013, doi: 10.1016/j.cnsns.2012.07.017.

[34] S. Fong, R. Wong, and A. V. Vasilakos, "Accelerated PSO Swarm Search Feature Selection for Data Stream Mining Big Data," *IEEE Trans. Serv. Comput.*, vol. 9, no. 1, pp. 33–45, 2016, doi: 10.1109/TSC.2015.2439695.

[35] D. Peitgen, H. O., Jürgens, H., Saupe, *Chaos and fractals: new frontiers of science*. Springer Science & Business Media, 2006.

[36] Y. Li, S. Deng, and D. Xiao, "A novel Hash algorithm construction based on chaotic neural network," *Neural Comput. Appl.*, vol. 20, no. 1, pp. 133–141, 2011, doi: 10.1007/s00521-010-0432-2.

[37] A. Erramilli, R. Singh, and P. Pruthi, "Modeling packet traffic with chaotic maps," *Stock. KTH*, pp. 1–33, 1994, [Online]. Available: http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.46.8102&rep=rep1&type=pdf%5Cnhttp://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.46.8102&rep=rep1&type=pdf.

[38] X. Zhang, D. Zou, and X. Shen, *A novel simple particle swarm optimization algorithm for global optimization*, vol. 6, no. 12. 2018.

[39] B. A. Pearlmutter, "Gradient Calculations for Dynamic Recurrent Neural Networks: A Survey," *IEEE Trans. Neural Networks*, vol. 6, no. 5, pp. 1212–1228, 1995, doi: 10.1109/72.410363.

[40] F. Tao and Q. Qi, "Make more digital twins," *Nature*, vol. 573, no. 7775, pp. 490–491, 2019, doi: 10.1038/d41586-019-02849-1.

[41] J. Li, Q. Zhou, H. Williams, H. Xu, and C. Du, "Cyber-Physical Data Fusion in Surrogate-assisted Strength Pareto Evolutionary Algorithm for PHEV Energy Management Optimization," *IEEE Trans. Ind. Informatics*, 2021.

**Ji Li** (M'19) awarded the Ph.D. degree in mechanical engineering from the University of Birmingham, U.K, in 2020. He is currently a Research Fellow and works on the Connected and Autonomous Systems for Electrified Vehicles at the Vehicle Technology Research Centre. His current research interests include computational intelligence, informatic fusion, data-driven modelling, man-machine system, and their applications on connected and autonomous vehicles.

**Quan Zhou** (M'17) received the Ph.D. degree in mechanical engineering from the University of Birmingham in 2019 that was distinguished by being the school's solo recipient of the University of Birmingham's Ratcliffe Prize of the year. He is currently a Research Fellow and leads the Connected and Autonomous Systems for Electrified Vehicles Research at University of Birmingham. His research interests include evolutionary computation, fuzzy logic, reinforcement learning, and their application in vehicular systems.

**Huw Williams** received the B.A. and M.A. degrees in mathematics from the University of Oxford, Oxford, U.K., in 1978 and 1983, respectively, and the Ph.D. degree in theoretical mechanics from the University of East Anglia, Norwich, U.K. He is an Honorary Professor of Energy and Automotive Engineering at the UoB, Birmingham, U.K. He is a professional mathematician with excellent skills in Lean, Six Sigma, Engineering Physics and Statistics. Huw has over 20 years' experience in the automotive industry. He also developed statistical skills through TQM in the 1980's culminating in his accreditation as Ford's top-scoring Master Black Belt in 2005.

**Guoxiang Lu** received the Ph.D. degree in control engineering from South China University of Technology, China, in 2010. He started a research fellow and led the development of modern engine control at the UoB, U.K., in 2015. He is currently senior manager at department of research and development of renewable and sustainable vehicles, BYD Auto Ltd, Guangzhou, China. His research interests include GDI engine combustion control, system modeling, real-time control, engine control strategy development, and intelligent vehicle system.

**Hongming Xu** received the Ph.D. degree in mechanical engineering from Imperial College London, London, U.K., in 1995. He is a Professor of Energy and Automotive Engineering at the University of Birmingham, Birmingham, U.K., and Head of Vehicle and Engine Technology Research Centre. He has six years of industrial experience with Jaguar Land Rover. He has authored and co-authored more than 400 journal and conference publications on advanced vehicle powertrain systems involving both experimental and modeling studies.