Robust Multi-Linear Fuzzy SVR Designed With the Aid of Fuzzy C-Means Clustering Based on Insensitive Data Information

ZHENG WANG¹, CHENG YANG¹, SUNG-KWUN OH², (Member, IEEE), ZUNWEI FU², (Member, IEEE), AND WITOLD PEDRYCZ³,4,5, (Life Fellow, IEEE)

¹School of Electrical and Electronic Engineering, The University of Suwon, Gyeonggi-do 445-743, South Korea
²Research Center for Big Data and Artificial Intelligence, Linyi University, Linyi 276005, China
³Department of Electrical and Computer Engineering, University of Alberta, Edmonton, AB T6R 2V4, Canada
⁴Department of Electrical and Computer Engineering, Faculty of Engineering, King Abdulaziz University, Jeddah 21589, Saudi Arabia
⁵Systems Research Institute, Polish Academy of Sciences, 01447 Warsaw, Poland

Corresponding authors: Sung-Kwun Oh (ohsk@suwon.ac.kr) and Zunwei Fu (fuzunwei@lyu.edu.cn)

This work was supported in part by the Korea Electric Power Corporation under Grant R19X001-18, and in part by the Research and Development Center for valuable Recycling (Global-Top Research and Development Program) of the Ministry of Environment under Project 2016002250002.

ABSTRACT Multiple SVR based on ensemble learning could be enhanced from the viewpoint of the performance, but the performance of modeling closely depends on the initial condition of the partitioning method and they are easily affected by noise and outliers. In this study, a multi-linear fuzzy support vector regression (MFSVR) robust to noise is proposed with the aid of the composite kernel function and \( \varepsilon \)-fuzzy c-means (FCM) clustering based on insensitive data information. Here insensitive data information stands for the interval data information of ‘’\( \varepsilon \)’’ which stands for insensitive loss parameter used in the \( \varepsilon \)-insensitive loss function. The objective of this study is to reduce the effect of noise and to alleviate the overfitting problem through the synergic effect of the following methods: First, \( \varepsilon \)-FCM clustering based on insensitive data information is used for considering more impact on decision boundary and reducing the effect of noise. Second, the composite kernel based on multiple linear kernel expression is proposed for implementing multi-linear decision boundary to alleviate overfitting problem. In more detail, each training data point is assigned with corresponding membership degrees in the \( \varepsilon \)-FCM clustering. Some data which are potentially to be noise or outlier are assigned with lower membership degrees and given small contribution (compensation) considered in composite kernel function. Then, the composite kernel function for multiple local SVRs is constructed according to the distribution characteristics of \( \varepsilon \)-FCM clustering. The proposed MFSVR is tested with both synthetic and UCI data sets in order to verify the effectiveness as well as efficient performance improvement. Experimental results demonstrate that the proposed method shows the better performance when compared to other some methods studied so far.

INDEX TERMS Multi-linear fuzzy SVR, \( \varepsilon \)-FCM clustering, insensitive data information, composite kernel function, membership degree.

I. INTRODUCTION During the past two decades, an identification of nonlinear models (NM) has received much attention. One can refer here to neural networks (NN), radial Basis Function Networks (RBFNs), Wavelet Networks (WNs), Neural Fuzzy Networks (NFNs) [1]–[4]. Support Vector Regression (SVR) has attracted interest due to its high generalization ability and robustness encountered in various applications. It is based on Vapnik’s \( \varepsilon \)-insensitive loss function and structural risk minimization [5], which was subsequently proposed as the support vector machine (SVM) implementation for regression and function approximation [6]. SVR has a variety of attractive linear recognition characteristics, thus has been studied and applied successfully to identification problems. Among these characteristics, an important common point is the use of kernel techniques, which perform nonlinear mapping to a high-dimensional feature space implicitly by replacing the
inner product with a positive definite kernel function. It is well known that the performance of SVM depends on the choice of kernel function. In the SVM literature, there exist linear kernel, sigmoid kernel, radial basis function (RBF) kernel function. On the other hand, composite kernel functions produce more flexible regression mapping by combining different types of kernel functions.

Many real-world systems exhibit complicated nonlinear features that are difficult to identify through suited linear features. The reason is that there is no exact multicollinearity amongst input variables. In these tasks, nonlinear kernel SVM such as RBF and sigmoid are also confronted with the potential over-fitting issue, especially for those linearly inseparable datasets in the high dimensional feature space. As a result, this would lead to severe over-fitting in the conventional nonlinear SVR. To address this problem, “divide-and-conquer” algorithm is an important strategy and widely used as an available tool for solving conceptually difficult problems [7]. A “divide-and-conquer” algorithm works by breaking a difficult problem into two or more sub-problems. The solutions to the sub-problems are then combined to give a solution to the original problem [8]. In [9], SVR with quasi-linear kernel is proposed for nonlinear system identification, the quasi-linear can consider nonlinear system as multi-linear model.

There is also a problem with the “divide-and-conquer” strategy. Each subset is not capable of capturing the distribution information of the input data. If the number of subsets is too large, it is more likely to encounter an over-fitting phenomenon.

When the multiple SVR model are applied, the original training data are partitioned by some unsupervised learning approaches, such as hard clustering method (HCM). One shortcoming of these methods for multiple SVR is that regression results are easily affected by noise, because the multiple SVR has a small number of data located in the training subspaces [24]. Also, the partitions closely depend on the initial conditions, the performance of regression results depends on whether the partition can capture the linear characteristic of the training dataset. Due to this kind of partition, multiple SVR for local subsets could be easily imbalanced.

Vapnik–Chervonenkis (VC) theory is introduced based on a limited number of data. Structural risk minimization (SRM) is the most important research area in the VC theory. The SRM principle provides a tradeoff between the quality of an approximation and the complexity of the approximating function [26]. The way to control the approximation and complexity is to adjust the insensitivity parameter $\varepsilon$ present in the loss function. The decision boundary based on $\varepsilon$-insensitive margin is considered as the regression model. For the regression problem, more compensation on regression model has benefit for alleviating the overfitting problem on complexity regression task. Here, a FCM clustering based on $\varepsilon$-insensitive data information ($\varepsilon$–FCM) is proposed for providing more compensation on the interval data in the $\varepsilon$-insensitive margin.

In this study, a multiple fuzzy SVR (MFSVR) based on $\varepsilon$-FCM and the composite kernel function is proposed for alleviating the overfitting problem. The proposed MFSVR approximates a nonlinear separating boundary by a number of multi-local linear boundaries with interpolation, in which the local linear property of the nonlinear boundary is used to construct a composite kernel. In the composite kernel function, the membership degrees obtained by $\varepsilon$-FCM are considered to provide the regression model with more compensation. The contribution of irrelevant data which are far away from regression model will be limited. Then, the information of local subsets is merged for estimating a composite kernel, and then the composite kernel is utilized to realize the multi-linear fuzzy SVR model and implement the SRM in the same way as a single standard SVR.

The key issue and the contribution of the proposed MFSVR can be briefly outlined as follows:

(a) Fuzzy subsets obtained by $\varepsilon$–FCM based on $\varepsilon$-insensitive margin are used for limiting the effect of irrelevant data.

When the training dataset is divided into several subsets, the SVR decision boundary is more susceptible to uncorrelated data with less training data. $\varepsilon$–FCM alleviates the effect of uncorrelated data through assigning more compensation (through the corresponding membership degrees) on decision boundary (regression model). In the $\varepsilon$–FCM, data that belong to the $\varepsilon$-insensitive margin are assigned with a high membership degree for fitting the regression model more accurately.

(b) The proposed MFSVR based the composite kernel function is implemented as multi-linear system for fitting the complicated system more properly.

The purpose of the composite kernel function is to alleviate the overfitting problem with multiple local linear decision boundaries. The composite kernel considers the nonlinear decision boundary as a combination of several multiple local linear decision boundaries. The local linear decision boundary obtained by local clusters can be adjusted by the number of clusters for alleviating the overfitting problem happened in the small clusters. The proposed MFSVR can alleviate over-fitting problem based on this composite kernel function.

(c) In order to carry out a fair and comprehensive comparison, statistical analysis is completed in experimental analysis.

The experiments include a comparison of the proposed model with a multi linear SVR, and a comparison of the proposed model with other general models. In order to clearly display the results of the comparison, Bonferroni-Dunn test is used to compare models run on the same datasets and to evaluate if the measured average ranks are significantly different from the mean rank.

In the sequel, the main contributions of our work can be summarized in a concise way as follows: First, The local linear decision boundary obtained by local clusters can be adjusted by the fuzzy clusters for alleviating the overfitting problem. Second, $\varepsilon$–FCM can effectively decrease the effect of uncorrelated data by assigning more compensation.
(through the corresponding membership degrees) on decision boundary (regression model). Third, the composite kernel based on multiple linear kernel expression is used through multi-linear decision boundary for better performance improvement.

The paper is organized as follows. Section 2 provides a survey of SVR models. Section 3 presents the \( \varepsilon \)-FCM method. Section 4 gives the formulation of MFSVR while section 5 outlines the procedure of MFSVR. Section 6 shows the experimental results and analysis. Finally, conclusions are covered in section 7.

II. SURVEY OF SVR

Regression is an intrinsic relationship to find a bunch of data. The traditional regression method only considers that the prediction is correct if the regression \( f(x) \) is completely equal to \( y \), and the loss is calculated. The support vector regression (SVR) considers that the deviation of the prediction is correct and there is no need to calculate the loss. Specifically, we set a threshold \( \alpha \) and calculate the loss of the data point of \( |f(x) - y| > \varepsilon \).

As shown in Fig.1, support vector regression means that as long as the value inside the dotted line can be considered as the prediction is correct. It is only necessary to calculate the loss of the value outside the dotted line.

As shown in Fig.1, support vector regression means that as long as the value inside the dotted line can be considered as the prediction is correct. It is only necessary to calculate the loss of the value outside the dotted line.

The SVR is formalized as

\[
\min_{\omega, b} \frac{1}{2} \|\omega\|^2 + \tau \sum_{i=1}^{m} \ell_\varepsilon(f(x_i) - y_i). \tag{1}
\]

Here, \( \tau \) is a regularization constant, and \( \ell_\varepsilon \) is an \( \varepsilon \)-insensitive loss function.

\[
\ell_\varepsilon(Z) = \begin{cases} 
0, & \text{if } |z| \leq \varepsilon \\
|Z| - \varepsilon, & \text{otherwise},
\end{cases} \tag{2}
\]

\[
\max_{a, \tilde{a}} \sum_{i=1}^{m} y_i(\tilde{a}_i - a_i) - \varepsilon(\tilde{a}_i - a_i), \tag{3}
\]

\[
-\frac{1}{2} \sum_{i=1}^{m} \sum_{j=1}^{m} (\tilde{a}_i - a_i)(\tilde{a}_j - a_j)x_i^T x_j, \tag{4}
\]

\[
s.t. \sum_{i=1}^{m} (\tilde{a}_i - a_i) = 0, \text{ and } 0 \leq a_i, \tilde{a}_i \leq \tau. \tag{5}
\]

The SVR diagram, gray area shows \( \varepsilon \)-spacer, and the data points falling into it are not used to calculate loss.

\[
f(x) = \sum_{i=1}^{m} (\tilde{a}_i - a_i)x_i^T x + b, \tag{6}
\]

\[
\omega = \sum_{i=1}^{m} (\tilde{a}_i - a_i)\phi(x_i), \tag{7}
\]

\[
f(x) = \sum_{i=1}^{m} (\tilde{a}_i - a_i)\kappa(x, x_i) + b, \tag{8}
\]

where \( \kappa(x_i, x_j) = \phi(x_i)^T \phi(x_j) \) is kernel function, as shown in Table 1 commonly used kernel functions are RBF, linear, and polynomial.

### TABLE 1. Commonly encountered kernel function of SVR.

| Kernel function | Expression | Parameter |
|-----------------|------------|-----------|
| Linear          | \( \kappa(x_i, x_j) = x_i^T x_j \) | \( d \geq 1 \) is coefficient of polynomial |
| Polynomial      | \( \kappa(x_i, x_j) = (x_i^T x_j)^d \) | \( \sigma > 0 \) is width of the RBF kernel |
| Gaussian (RBF)  | \( \kappa(x_i, x_j) = \exp \left( -\frac{|x_i - x_j|^2}{2\sigma^2} \right) \) | |

As shown in Fig.2, it is found that the model has the best fitting effect when using the RBF kernel function. However, the complexity of the SVR of the RBF kernel function is relatively high and it could easily lead to overfitting [20]. As a consequence, a multi-local linear SVR is proposed to solve the over-fitting problem of SVR with RBF kernel function.

III. \( \varepsilon \)-FCM BASED ON INSENSITIVE MARGIN

the nonlinear SVR model is too complex implying overfitting, and the linear SVR model is too simple to accurate enough. A “divide-and-conquer” algorithm works by breaking a difficult problem into two or more sub-problems. The solutions to the sub-problems are then combined to produce an overall a solution to the original problem.

As shown in Fig.3, the proposed solution can be described through the following steps:

Step1) Divide the training dataset into several fuzzy spaces (subsets) is completed by using \( \varepsilon \)-insensitive loss parameter based Fuzzy C-Means (\( \varepsilon \)-FCM) clustering.
Step 2) Each subset is estimated by a nonlinear mapping, and complete regression each subset by using local linear SVR.

Step 3) The composite kernel function with fuzzy inference is introduced to implement into multiple local linear model by incorporating distribution information combined from each subset.

This section illustrates the \( \varepsilon \)-FCM method. This algorithm alleviates the overfitting problem by assigning more compensation on decision boundary. As the definition of decision boundary (regression model) is realized based on training label, \( \varepsilon \)-FCM is considered as a type of supervised clustering algorithm. The motivation and procedure of \( \varepsilon \)-FCM is described as follows.

The proposed \( \varepsilon \)-FCM is a data pre-processing algorithm based on existing SVR, which limit the effect of data points away from the decision boundary [23]. When using the existing FCM, the cluster center of the clusters (subsets) are easily influenced by noisy data.

As shown in Fig. 4 the cluster center is affected by the outliers and noise data (green circles), which leads to the unsatisfactory effect of the clustering algorithm. To alleviate overfitting problem occurring in Fig. 4, \( \varepsilon \)-FCM is proposed. The procedure of the proposed \( \varepsilon \)-FCM is described as Algorithm I.

Fig. 5 shows detailed computational steps of \( \varepsilon \)-FCM.

IV. ARCHITECTURE OF PROPOSED SVR

Multiple SVR models can obtain preferred results rather than conventional SVR based on the “divide-and-conquer” strategy. However, the complexity of the model increases as the model operates the SVR algorithm multiple times [22].

In this section, a composite kernel function is formulated to alleviate overfitting problem, and refine it in a single SVR fashion for reducing model complexity [25]. According the membership degree \( u \) is (10), new composite kernel function is constructed.

In [16], Pedrycz used the FCM clustering to find antecedent variable membership functions and then identified a relational fuzzy model. In the \( \varepsilon \)-FCM, each cluster corresponds to a fuzzy IF–THEN rule in the Takagi–Sugeno–Kang form [17], [18]

\[
\text{If } x \text{ is } A_i, \text{ then } y = w_i^T x', \quad i = 1, 2, \ldots, c, \quad (11)
\]

where \( x \in R^d \) is the input variable, \( y \in R \) is the output variable, \( x' = [x^T, 1]^T \) is the augmented input vector, \( w_i = [\hat{w}_i, w_{i0}] \in R^{d+1} \) is the vector of consequent parameters of the \( i \)-th rule; \( w_{i0} \) denotes a bias of the \( i \)-th model. The antecedent fuzzy set of the \( i \)-th rule, \( A_i \) has a membership degree \( u_i(x) \in [0, 1] \). In [19], the FCM related fuzzy model is described as a weighted averaging aggregation of individual rules:

\[
y = f(x, w_g) = \sum_{i=1}^{c} u_i(x)w_i^T x'. \quad (12)
\]

where \( u_i(x) \) is the normalized firing strength of the \( i \)-th rule for input \( x, w_g = [w_1^T, w_2^T, \ldots, w_c^T] \) denotes consequent parameter vector. Define \( u = [u_1(x)x^T, u_2(x)x^T, \ldots, u_c(x)x^T] \) then the overall output of the FCM related model can be written as \( y = uw_g \).

Then, we give the formulation of the composite kernel function with respect to \( u \).
Consider a single-input-single-output (SISO) nonlinear time-invariant system whose input-output dynamics is described as

\[
y(t) = g(\varphi(t)) + e(t) \\
\varphi(t) = [y(t - 1), y(t - n_y), u(t - 1), \ldots, u(t - n_u)].
\]  \hspace{1cm} (13)

where \(u(t) \in R, y(t) \in R, e(t) \in R\) are the system input, the system output and a stochastic noise of zero-mean at time \(t\), \(n_u\) and \(n_y\) are unknown maximum delays of the input and output, respectively. \(\varphi(t) \in R^n\) is the regression vector composed of delayed input-output data. \(n\) is the number of input variables, which equals to the sum of \(n_u\) and \(n_y\).

The \(t\) can be replaced by \(u\), (13) can be rewritten as

\[
y(u) = g(\varphi(u)) + e(u) \\
\varphi(u) = [y(u - 1), y(u - n_y), u(u - 1), \ldots, u(u - n_u)].
\]  \hspace{1cm} (14)

As an example, the model is generated by performing Taylor expansion of the unknown nonlinear function \(g(\varphi(u))\) around the region \(\varphi(u) = 0\). Since \(g(\cdot)\) is assumed to be differentiable, the derivative \(g^{(i)}(0)(i = 1, 2, \ldots)\) exists. Then ignoring \(g(0)\) for simplicity and then a regression form of the system described in (14) is given with an multiple structure

\[
y(u) = \varphi^T(u)\theta(\varphi(u)) + e(u)
\]
functions. According to (12) and (16), a compact representation of the model is given as

$$y(u) = \sum_{j=0}^{M} \varphi_j^T(u)\Omega_ju_j + e(u), \quad (17)$$

in which the NNM $N_j(u)$ is represented as of $\varphi_j^T(u)u_j$ for the simplicity, and it is defined $u_0 = 1$.

By introducing $\Phi(u)$ and $\Theta$ (17) can be rewritten as follows.

$$y(u) = \Phi^T(u)\Theta + e(u), \quad (18)$$

where

$$\Phi(u) = [\varphi_0^T(u), u_1 \cdot \varphi_1^T(u), \ldots, u_C \cdot \varphi_C^T(u)]^T,$$

$$\Theta = [\Omega_0^T, \Omega_1^T, \ldots, \Omega_C^T]^T.$$

Therefore, the nonlinear system identification is reduced to a linear regression problem with respect to $\Phi(u)$, and $\Theta$ are called parameters.

The parameter vectors are estimated using the SVR approach. We Introduce the structural risk minimization (SRM) as

$$\min I = \frac{1}{2} \Theta^T \Theta + \tau \sum_{u=1}^{N} (\xi_u + \xi_u^*), \quad (19)$$

with the constraint condition

$$\begin{cases} y(u) - \Phi^T(u)\Theta \leq \varepsilon + \xi_u^*, \\
y(u) + \Phi^T(u)\Theta \leq \varepsilon + \xi_u. \end{cases}$$

where $N$ is the number of samples, $\xi_u^*$, $\xi_u$ are slack variables. The parameter $C$ controls the tradeoff between the complexity of regression model and the amount up to which errors are tolerated.

The solution can be obtained by finding a saddle point of the associated Lagrange function.

$$L(\Theta, \xi_u, \xi_u^*, \alpha, \alpha^*, \beta, \beta^*) = \frac{1}{2} \Theta^T \Theta + \tau \sum_{u=1}^{N} (\xi_u + \xi_u^*)$$

$$+ \sum_{u=1}^{N} \alpha_u(y(u) - \Phi^T(u)\Theta - \varepsilon - \xi_u)$$

$$+ \sum_{u=1}^{N} \alpha_u^*(y(u) + \Phi^T(u)\Theta - \varepsilon - \xi_u^*)$$

$$- \sum_{u=1}^{N} (\beta_u \xi_u + \beta_u^* \xi_u^*). \quad (20)$$

Then the saddle point could be acquired by minimization of $L$ with respect to $\Theta, \xi^*_u$ and $\xi$.

$$\frac{\partial L}{\partial \Theta} = 0 \Rightarrow \Theta = \sum_{u=1}^{N} (\alpha_u - \alpha_u^*)\Phi(u),$$

$$\frac{\partial L}{\partial \xi_u} = 0 \Rightarrow \beta_u = \tau - \alpha_u,$$

$$\frac{\partial L}{\partial \xi_u^*} = 0 \Rightarrow \beta_u^* = \tau - \alpha_u^*. \quad (21)$$

Thus, one can convert the primal problem (21) into an equivalent dual problem as

$$\max W(\alpha_u, \alpha_u^*) = -\frac{1}{2} \sum_{u,k=1}^{N} (\alpha_u - \alpha_u^*)(\alpha_k - \alpha_k^*)\Phi^T(u)$$

$$\times \Phi(u) + \sum_{u=1}^{N} (\alpha_u - \alpha_u^*)y(u) - \varepsilon \sum_{u=1}^{N} (\alpha_u + \alpha_u^*). \quad (22)$$

subject to:

$$\sum_{u=1}^{N} (\alpha_u - \alpha_u^*) = 0, \quad \alpha_u, \alpha_u^* \in [0, \tau], \quad (23)$$

to do this, $\alpha_i$ and $\alpha_i^*$ are reserved only in (23), which can be obtained and substituted into (18) for parameter vector $\Theta$.

With the multiplier $\alpha_i$ and $\alpha_i^*$ obtained, the model can be identified as multi SVR with the membership degrees $u$.

$$y(u) = -\frac{1}{2} \sum_{u'=1}^{N} (\alpha_{u'} - \alpha_{u'}^*)K(u, u'), \quad (24)$$

$$K(u, u') = \Phi^T(u)\Phi^T(u') = (u \cdot u') \sum_{j=1}^{C} R_j(u)R_j(u'). \quad (25)$$

Each training subset is represented by $R_j(u)$ corresponding to an RBF, which can be expressed as

$$R_j(u) = \exp[-\frac{(u - v)^2}{\lambda_j^2}]. \quad (26)$$

The $j$-th RBF function $R_j(u)$ is constructed by specifying the center of $v_j$ with width of $\sigma_j$, which are set as the center and radius of the $j$-th cluster, and $\lambda$ is a scaling constant.

In this way, the nonlinear input data become linear in a high-dimensional space formed in terms of the composite kernel function (26). In other words, the composite kernel function is learnt to reproduce an appropriate decision boundary according to data distribution.

The multi-RBF kernel is named after two folds. Firstly, it is derived from the quasi-ARX modeling method. Secondly, the nonlinearity of the kernel can fill the gap between linear and nonlinear kernel functions by adjusting the value of $C$. The proposed MFSVR as shown in Fig. 6.

V. SOME COMMON MISTAKES DESIGN PROCEDURE OF THE MFSVR

Overall, the proposed MFSVR design framework includes the following steps; refer to Figure 7.

Step 1 (Construct Training Data and Testing Data, and Set the Parameters for MFSVR Structure and Learn the Model):

The original dataset is divided into two parts: training data, and testing data. Training data is used to design MFSVR.
and testing data is used to verify the quality of the constructed MFSVR. The parameters of the MFSVR structure include: fuzzification coefficients, number of clusters, insensitive loss parameter $\varepsilon$, and penalty coefficient $C$.

**Step 2:** Carry out $\varepsilon$-FCM for input space.

**Substep 2-1:** Obtain data points in the $\varepsilon$ insensitive margin. Select the data points for $\varepsilon$ insensitive margin ($|y - f(x)| \leq \varepsilon$) by implementing the SVR model with RBF kernel function.

**Substep 2-2:** Calculate and update the cluster centers from insensitive area data points ($v_{ip}$).

Carry out FCM to estimating cluster centers $v_{ip}$ for insensitive area data points.

**Substep 2-3:** Calculate the membership degrees of all input variables ($u_{is}$). (10) is used to compute membership degrees of all input variables.

**Step 3 (Construct the Composite Kernel and Carry Out MFSVR for Modeling):** Function $K(u, u')$ is constructed according to (26). Then, the proposed MFSVR (24) model is designed with the aid of the composite kernel function.

**Step 4 (Compute Performance Index):** Root mean square error (RMSE) as a performance index on training and testing datasets is computed by designed MFSVR via steps 2~3.
Five operations were performed on the above process using five-fold cross-validation, and the average of performance indexes and its standard deviation was obtained as the output of the design model.

VI. EXPERIMENTAL STUDIES

In this section, the conventional SVR with three types of kernel functions (linear, polynomial, RBF), proposed MFSVR based on FCM and proposed MFSVR based on $\varepsilon$-FCM are used for comparing the performance. Here $\varepsilon$-FCM clustering means FCM clustering being carried out based on insensitivity margin of parameter “$\varepsilon$”.

The most important difference between the three conventional SVR and the proposed MFSVR is the kernel function. The proposed MFSVR performs fuzzy partitioning on nonlinear regression task by $\varepsilon$-FCM, and locally fits each partition with several linear kernel functions, and then integrates these partitions through the composite kernel function to alleviate overfitting problem.

We use both synthetic and machine learning datasets to compare the performance of the existing model and the proposed model. Consequently, the experiment is tested for the noise and outliers robustness of the proposed model based on 5dB, 10dB, and 15dB noise rate and 10%, 20%, and 30% outlier rate (5dB noise rate means 95% of data impacted additively with white noise). The way to add noise is designed based on the wgn (White Gaussian noise) function built in matlab toolbox, do not say about matlab but rather say how this function works and the way to add outliers is change their output value into mean value randomly [19].

All datasets are randomly divided into 80% training dataset and 20% testing dataset and all used 5-fold cross-validation. Table 2 summarizes the values of the parameters of proposed MFSVR.

| TYPE | PARAMETERS | VALUES |
|------|------------|--------|
| INPUT | Data split (Training: 80% : Testing 20%) | [2 – 10] |
| FUZZY PARTITION | Fuzzification coefficients | 1.5 |
| | Maximum iteration of FCM | 100 |
| SVR | Insensitive loss parameter | 0.001 |
| | Penalty coefficient (C) | 1 |
| | Performance index | RMSE |
| OUTPUT | k-fcv | 5 |

A. TYPES OF GRAPHICS TWO-DIMENSIONAL SYNTHETIC DATASET

The MFSVR is applied to modeling the well-known Matlab logo. This logo is a display of one of the eigenmodes for an “L” shaped membrane as shown Fig. 8. We considered 500 randomly generated input–output data pairs. From solution set of Fig. 8 (a), 500 input–output patterns are arbitrarily selected and the distribution of the selected patterns is visualized in Fig. 8 (b). The 500 pairs of input–output patterns are used for designing model and comparing performance of each model.

Table 3 shows the performance of the proposed MFSVR with different number of clusters when using existed FCM and $\varepsilon$-FCM. The performance of the proposed two models is very similar.

TABLE 3. Performance outputs of the proposed MFSVR for 2D synthetic dataset.

| NC | MFSVR with FCM | MFSVR with $\varepsilon$-FCM |
|----|----------------|-----------------------------|
|     | PI (STD) | EPI (STD) | PI (STD) | EPI (STD) |
| 2   | 0.244±0.005 | 0.251±0.018 | 0.248±0.005 | 0.248±0.024 |
| 3   | 0.248±0.004 | 0.249±0.015 | 0.247±0.005 | 0.249±0.018 |
| 4   | 0.248±0.005 | 0.253±0.020 | 0.247±0.004 | 0.248±0.017 |
| 5   | 0.247±0.005 | 0.249±0.021 | 0.247±0.006 | 0.247±0.026 |
| 6   | 0.247±0.003 | 0.248±0.012 | 0.247±0.004 | 0.246±0.015 |
| 7   | 0.246±0.007 | 0.249±0.003 | 0.246±0.005 | 0.246±0.035 |
| 8   | 0.245±0.002 | 0.249±0.008 | 0.245±0.001 | 0.246±0.011 |
| 9   | 0.245±0.004 | 0.252±0.026 | 0.245±0.008 | 0.246±0.033 |
| 10  | 0.245±0.002 | 0.254±0.012 | 0.245±0.005 | 0.247±0.025 |
| Ave.| 0.247±0.004 | 0.250±0.015 | 0.246±0.005 | 0.247±0.023 |

NC: Number of clusters, PI (STD): Training_RMSE (Standard deviation), EPI (STD): Testing_RMSE (Standard deviation) Ave.:Average

Fig. 9 shows the membership functions and local input spaces of the two proposed models with 5 clusters. Fig. 9 (a) description of partition function of each cluster and local
input spaces formed by existed FCM. Fig. 9 (b) description of partition function of each cluster and local input spaces formed by ε-FCM. It can be seen that with the general FCM, the data is divided into 5 clusters evenly (Fig. 9 a), while with ε-FCM, all five cluster centers are distributed in the ε insensitive margin (Fig. 9 b).

Table 4 lists the proposed model and the existing SVR model for comparison. The approximation and generalization abilities of the proposed model are largely improved in comparison with the abilities of the existed SVR. The experiment of the proposed model is implemented on computer with the Intel Core i5-4690k 3.50 GHz CPU. It can be seen that the computing time of the proposed model is significantly shorter than that of the existing SVR model. The computing time of proposed MFSVR with ε-FCM is the same as the time of MFSVR with existing FCM, the difference of 0.003 seconds happens because of the difference of the number of clusters.

Table 5 shows the performance index of the proposed MFSVR with added different levels of white noise when using existed FCM and ε-FCM. After adding white noise to this simple synthetic dataset, the performance of MFSVR with ε-FCM is better than that of general FCM.

Table 6 lists the results obtained after adding white noise the proposed model and the existing SVR model for comparison. The approximation and generalization abilities of the
MFSVR with $\varepsilon$-FCM is largely improved in comparison with the abilities of the existing SVR after adding the same white noise.

Table 7 lists the performance of proposed model and the existing SVR model with different levels of outlier for comparison.

### Table 7. Results of comparative analysis of the existing SVR and the proposed MFSVR for 2D synthetic dataset with outliers.

| Proposed model | NC | EPI (STD) 10% | EPI (STD) 20% | EPI (STD) 30% |
|----------------|----|--------------|--------------|--------------|
| Linear         |    | 0.285±      | 0.294±       | 0.297±       |
| Existing SVR   |    | 0.031       | 0.018        | 0.020        |
| Polynomial     |    | 0.307±      | 0.310±       | 0.308±       |
| RBF            |    | 0.034       | 0.025        | 0.008        |
| Proposed MFSVR | 7 5 4 | 0.266±      | 0.277±       | 0.284±       |
| Proposed MFSVR | 5 3 6 | 0.253±      | 0.260±       | 0.272±       |
| Proposed MFSVR |      | 0.019       | 0.015        | 0.010        |
| $\varepsilon$-FCM |      | 0.025       | 0.020        | 0.006        |

NC: Number of clusters, EPI (STD): Testing_RMSE (Standard deviation)

![Fig. 10](image)

Fig. 10 shows the data has outlier points in the insensitive data points selected by using $\varepsilon$-FCM. (a) is 500 data points after adding 10% outliers including 450 original data (‘◦’) and 50 outliers (‘∗’). Fig.10(b) was shown insensitive data points selected by using $\varepsilon$-FCM. The blue period (‘◦’) are 140 insensitive data points selected from the 500 data centers with outliers added in Figure 10 (a). It can be seen that by using $\varepsilon$-FCM to select 140 insensitive data points, there are only three outliers, and the other 47 outliers have not been filtered out.

### B. ABALONE DATASET

We consider abalone datasets (http://archive.ics.uci.edu/ml/datasets/Abalone). Predicting the age of abalone from physical measurements. The age of abalone is determined by cutting the shell through the cone, staining it, and counting the number of rings through a microscope. This dataset includes 4177 input–output pairs. There are 7 input variables, and the output is the age of abalone.

Table 8 shows the performance of the proposed MFSVR with different number of clusters when using existed FCM and $\varepsilon$-FCM. The performance of the proposed two models is similar on abalone dataset.

### Table 8. Performance outputs of the proposed MFSVR for abalone dataset.

| NC | PI (STD) | EPI (STD) | EPI (STD) |
|----|----------|-----------|-----------|
| 2  | 2.251±0.016 | 2.269±0.090 | 2.250±0.021 | 2.263±0.084 |
| 3  | 2.237±0.030 | 2.267±0.132 | 2.236±0.005 | 2.263±0.104 |
| 4  | 2.232±0.020 | 2.269±0.092 | 2.233±0.010 | 2.264±0.053 |
| 5  | 2.294±0.033 | 2.269±0.126 | 2.229±0.023 | **2.263±0.076** |
| 6  | 2.264±0.022 | 2.272±0.105 | 2.227±0.006 | 2.267±0.099 |
| 7  | 2.255±0.016 | 2.277±0.067 | 2.225±0.009 | 2.271±0.044 |
| 8  | 2.223±0.022 | 2.276±0.079 | 2.222±0.016 | 2.275±0.047 |
| 9  | 2.222±0.014 | 2.277±0.046 | 2.220±0.016 | 2.281±0.115 |
| 10 | 2.220±0.025 | 2.284±0.047 | 2.220±0.015 | 2.281±0.067 |
| Ave. | 2.229±0.022 | 2.273±0.087 | 2.229±0.015 | 2.270±0.077 |

NC: Number of clusters, PI (STD): Training_RMSE (Standard deviation), EPI (STD): Testing_RMSE (Standard deviation) Ave.: Average

Table 9 lists the proposed models and the existing SVR model for comparison. The approximation and generalization abilities of the proposed two models are largely improved in comparison with the abilities of the existed SVR.

### Table 9. Results of comparative analysis of the existing SVR and the proposed MFSVR for abalone dataset.

| N | PI (STD) | EPI (STD) | CT |
|---|----------|-----------|----|
| Linear | 2.959±0.02 | 2.966±0.08 | 3m22 |
| 2 | s | 8 |
| Existed SVR | 3.267±0.04 | 3.281±0.19 | 1m41 |
| 1 | 9 | s |
| RBF | 3.165±0.01 | 3.173±0.02 | 2m43 |
| 8 | 2 | 2 |
| Proposed MFSVR | 2.237±0.03 | 2.267±0.13 | 2.19s |
| 3 | 0 | 2 |
| $\varepsilon$-FCM | 2.229±0.02 | **2.263±0.07** | 2.26s |
| 5 | 3 | 6 |

NC: Number of clusters, PI (STD): Training_RMSE (Standard deviation), EPI (STD): Testing_RMSE (Standard deviation), CT: Computing time

Table 10 shows the performance index of the proposed MFSVR with added different levels of white noise when
TABLE 10. Performance outputs of proposed MFSVR for abalone dataset with white noise.

| NC  | MFSVR with FCM_EPI (STD) | MFSVR with ε-FCM_EPI (STD) |
|-----|--------------------------|-----------------------------|
| 15DB| 10DB| 5DB| 15DB| 10DB| 5DB| 15DB| 10DB| 5DB|
| 2   | 2.724± 2.834± | 2.948± | 2.642± | 2.767± | 2.804± | 0.126 | 0.128 | 0.177 | 0.117 | 0.137 | 0.127 |
| 3   | 2.723± | 2.833± | 2.964± | 2.648± | 2.749± | 2.801± | 0.141 | 0.107 | 0.170 | 0.135 | 0.120 | 0.088 |
| 4   | 2.702± | 2.829± | 2.951± | 2.648± | 2.740± | 2.800± | 0.066 | 0.153 | 0.122 | 0.095 | 0.104 | 0.167 |
| 5   | 2.710± | 2.835± | 2.963± | 2.635± | 2.762± | 2.795± | 0.149 | 0.042 | 0.053 | 0.103 | 0.175 | 0.164 |
| 6   | 2.726± | 2.823± | 2.977± | 2.640± | 2.766± | 2.800± | 0.126 | 0.087 | 0.102 | 0.094 | 0.126 | 0.189 |
| 7   | 2.709± | 2.814± | 2.964± | 2.643± | 2.771± | 2.797± | 0.048 | 0.159 | 0.064 | 0.089 | 0.055 | 0.141 |
| 8   | 2.718± | 2.821± | 2.969± | 2.651± | 2.762± | 2.811± | 0.131 | 0.088 | 0.120 | 0.070 | 0.114 | 0.082 |
| 9   | 2.711± | 2.824± | 2.964± | 2.652± | 2.753± | 2.817± | 0.154 | 0.171 | 0.067 | 0.129 | 0.071 | 0.088 |
| 10  | 2.699± | 2.848± | 2.954± | 2.653± | 2.764± | 2.800± | 0.109 | 0.078 | 0.114 | 0.125 | 0.079 | 0.069 |
| Age | 2.714± | 2.829± | 2.962± | 2.646± | 2.759± | 2.803± | 0.117 | 0.113 | 0.110 | 0.106 | 0.109 | 0.124 |

NC: Number of clusters, EPI (STD): Testing RMSE (Standard deviation) Ave.: Average

TABLE 11. Results of comparative analysis of existing SVR and proposed MFSVR for abalone dataset with white noise.

| NC  | EPI (STD) |
|-----|-----------|
| 15DB| 10DB | 5DB|
| Linear | 2.982± | 3.003± | 3.038± | 0.092 | 0.143 | 0.142 |
| Polynomial | 3.285± | 3.263± | 3.307± |
| RBF | 3.183± | 3.193± | 3.342± | 0.113 | 0.149 | 0.095 |
| Proposed | FCM | 2.699± | 2.814± | 2.948± | 0.109 | 0.159 | 0.177 |
| Proposed | ε-FCM | 2.635± | 2.740± | 2.795± | 0.103 | 0.104 | 0.164 |

NC: Number of clusters, EPI (STD): Testing RMSE (Standard deviation)

TABLE 12. Results of comparative analysis of the existing SVR and the proposed MFSVR for abalone dataset with outliers.

| NC | EPI (STD) |
|-----|-----------|
| 10% | 20% | 30%|
| Linear | 3.036± | 3.088± | 3.131± | 0.076 | 0.107 | 0.078 |
| Existing SVR | Polynomial | 3.320± | 3.306± | 3.379± | 0.099 | 0.069 | 0.11 |
| RBF | 3.181± | 3.226± | 3.263± | 0.137 | 0.067 | 0.082 |
| Proposed | FCM | 2.479± | 2.654± | 2.800± | 0.122 | 0.111 | 0.126 |
| Proposed MFSVR | ε-FCM | 2.403± | 2.465± | 2.605± | 0.141 | 0.109 | 0.056 |

NC: Number of clusters, EPI (STD): Testing RMSE (Standard deviation)

C. AIR QUALITY DATASET

The dataset contains 9358 instances of hourly averaged responses from an array of 5 metal oxide chemical sensors embedded in an Air Quality Chemical Multisensor Device. Data were recorded from March 2004 to February 2005 (one year) representing the longest freely available recordings of on-field deployed air quality chemical sensor devices responses. Ground Truth hourly averaged concentrations for CO, Non Metanic Hydrocarbons, Benzene, Total Nitrogen Oxides (Nox) and Nitrogen Dioxide (NO2) and was provided by a co-located reference certified analyzer.

Table 13 shows the performance of the proposed MFSVR with different number of clusters when using existed FCM and ε-FCM. The performance of MFSVR using exciting FCM is better than that of using ε-FCM. Because air quality dataset is time series data, and the amount of data is relatively large, exciting FCM using all the data to calculate the clustering centers can obtain more comprehensive data information than the ε-FCM only using insensitive data to computing center. While the performance of MFSVR by using exciting FCM is better.

TABLE 13. Performance outputs of the proposed MFSVR for air quality dataset.

| NC | EPI (STD) |
|-----|-----------|
| PI (STD) | PI (STD) |
| MFSVR with FCM | MFSVR with ε-FCM |
| 2 | 0.072± | 0.001 | 0.078± | 0.017 | 0.081± | 0.004 | 0.092± | 0.023 |
| 3 | 0.064± | 0.002 | 0.070± | 0.023 | 0.077± | 0.006 | 0.086± | 0.018 |
| 4 | 0.061± | 0.001 | 0.069± | 0.029 | 0.079± | 0.001 | 0.090± | 0.026 |
| 5 | 0.069± | 0.004 | 0.074± | 0.012 | 0.069± | 0.007 | 0.079± | 0.044 |
| 6 | 0.070± | 0.002 | 0.077± | 0.015 | 0.075± | 0.002 | 0.084± | 0.019 |
| 7 | 0.061± | 0.002 | 0.069± | 0.037 | 0.085± | 0.002 | 0.094± | 0.033 |
| 8 | 0.065± | 0.003 | 0.072± | 0.026 | 0.074± | 0.005 | 0.082± | 0.011 |
| 9 | 0.067± | 0.001 | 0.077± | 0.014 | 0.076± | 0.003 | 0.088± | 0.021 |
| 10 | 0.059± | 0.001 | 0.065± | 0.047 | 0.080± | 0.005 | 0.089± | 0.036 |

Ave.: 0.065± 0.002 | 0.072± 0.024 | 0.071± 0.004 | 0.087± 0.026

NC: Number of clusters, PI (STD): Testing RMSE (Standard deviation), EPI (STD): Testing RMSE (Standard deviation) Ave.: Average

Table 14 lists the proposed model and the existing SVR model for comparison. The approximation and generalization abilities of the proposed two models are largely improved in comparison with the abilities of the existed SVR.
Table 15 shows the performance index of the proposed MFSVR with added different levels of white noise when using existed FCM and ε-FCM. After adding white noise to air quality dataset, the performance of the proposed two models is similar.

Table 16 and 17 shows the performance index on testing dataset including Gaussian white noise and outliers. The boldface entries stand for the better performance than existing SVR. Consequently, the performance of the proposed MFSVR with ε-FCM outperforms the proposed MFSVR with FCM, and such a tendency is much clearer as increasing the intensity of noises and outliers.

### D. OTHER DATASET

To evaluate the performance of the proposed models and the effect of the composite kernel function, different algorithms on 14 well-known benchmark datasets are compared. These datasets are obtained from the University of California at Irvine (UCI) Machine Learning Repository (http://archive.ics.uci.edu/ml/datasets.html). Table 18 presents a summary of the datasets.

Table 16 and 17 shows the performance index on testing dataset including Gaussian white noise and outliers. The boldface entries stand for the better performance than existing SVR. Consequently, the performance of the proposed MFSVR with ε-FCM outperforms the proposed MFSVR with FCM, and such a tendency is much clearer as increasing the intensity of noises and outliers.

### D. OTHER DATASET

To evaluate the performance of the proposed models and the effect of the composite kernel function, different algorithms on 14 well-known benchmark datasets are compared. These datasets are obtained from the University of California at Irvine (UCI) Machine Learning Repository (http://archive.ics.uci.edu/ml/datasets.html). Table 18 presents a summary of the datasets.

To further analyze whether the proposed model is statistically significantly better than the other comparative models, we use the Bonferroni-Dunn test in Table 20, which fit for situations where all models are only compared to the control model and not between them. If the
corresponding average rank differs by at least the critical difference (CD), the performance of any two models is significantly different. At $p = 0.10$ (significance level), the CD value is $1.70$. Table 20 covers the difference of average rank between the five comparative models (weka) and the two proposed models, as well as the comparison results with CD. Since the difference between the average rank of all comparative model and the proposed MFSVR with $\varepsilon$-FCM is greater than CD ($2.96 > 1.70$), we can conclude that the prediction accuracy of MFSVR is statistically superior to the five comparative models. Furthermore, the proposed MFSVR with $\varepsilon$-FCM is slightly better than MFSVR with existing FCM.

Table 19 provides comparison on the proposed models and existing SVR of weka on 14 datasets. Obviously, compared with the three kinds of SVR models, the proposed models have a great improvement in prediction accuracy. 9 in 14 datasets the MFSVR based on $\varepsilon$-FCM obtained the best performance.

According to Table 21 and 22, the experiments are carried out under noise and with the existence of outliers. Consequently, the performance of the proposed MFSVR with $\varepsilon$-FCM outperforms the MFSVR with FCM, and such a tendency gets much clearer as the increase of the noise and outliers.

### Table 19. Results comparison by the exciting SVR and the proposed MFSVR.

| Data | Linear | Existing SVR | MFSVR(Proposed) |
|------|--------|--------------|-----------------|
| BH   | 4.89   | 5.79         | 5.28            |
| CCS  | 9.88   | 13.57        | 13.33           |
| EE   | 2.08   | 3.71         | 3.20            |
| FF   | 51.67  | 50.90        | 52.75           |
| F1   | 0.85   | 0.89         | 0.85            |
| MIS  | 10.22  | 10.37        | 9.58            |
| Nox  | 8.67   | 8.67         | 11.77           |
| TR   | 5.45   | 6.20         | 2.33            |
| CV   | 8.28   | 11.77        | 11.99           |
| F4   | 1.33   | 1.00         | 1.07            |
| Pu   | 0.05   | 0.03         | 0.03            |
| WR   | 0.66   | 0.72         | 0.65            |
| MO   | 0.14   | 0.13         | 0.13            |
| St   | 2.40   | 2.41         | 2.80            |

### Table 20. Results of comparative analysis.

| Data | A | B | C | D | E | Proposed |
|------|---|---|---|---|---|----------|
| BH   | 4.88| 4.99| 4.89| 4.85| 5.30| 3.88      |
| C    | 10.48| 8.16| 10.57| 10.81| 7.08| 6.50      |
| EE   | 1.96| 2.67| 2.02| 1.95| 1.72| 1.65      |
| FF   | 52.58| 55.13| 51.87| 66.87| 100.86| 49.50    |
| F1   | 0.84| 0.88| 0.88| 1.04| 1.06| 0.80      |
| MIS  | 6.50| 8.84| 7.91| 8.25| 10.24| 6.55      |
| TR   | 10.48| 8.16| 10.57| 10.81| 7.08| 6.50      |
| CV   | 1.96| 2.67| 2.02| 1.95| 1.72| 1.65      |
| F4   | 52.58| 55.13| 51.87| 66.87| 100.86| 49.50    |
| Pu   | 0.84| 0.88| 0.88| 1.04| 1.06| 0.80      |
| WR   | 6.50| 8.84| 7.91| 8.25| 10.24| 6.55      |
| MO   | 1.96| 2.67| 2.02| 1.95| 1.72| 1.65      |
| St   | 10.48| 8.16| 10.57| 10.81| 7.08| 6.50      |

### Table 21. Performance outputs of the proposed MFSVR for 14 datasets with white noise.

| Data  | 15DB    | 16DB    | 17DB    | 18DB    | 19DB    |
|-------|---------|---------|---------|---------|---------|
| BH    | 3.955   | 3.996   | 4.051   | 4.017   | 4.103   |
| C    | 0.30 + 0.27 + 0.62 + 0.48 + 0.56 + 0.52 |
| EE    | 6.601   | 6.473   | 6.287   | 6.632   | 6.523   |
| F1    | 4.907   | 4.951   | 4.917   | 4.931   | 4.927   |

VOLUME 8, 2020
outliers. It means that the effect of \( \varepsilon \)-FCM increases with the increase of the noise and outliers.

### VII. CONCLUSION

In this study, a multiple fuzzy support vector regression (MFSVR) with the aid of composite kernel function and \( \varepsilon \)-FCM is introduced. First, \( \varepsilon \)-FCM is used to partition the training dataset into several subsets as preprocessing of proposed modeling. Second, the composite kernel based on multiple linear kernel expression is considered to avoid overfitting problem. In more detail, each training data point is assigned with corresponding membership degrees used in preprocessing. Some data which are far away from the decision boundary are assigned with lower membership degrees and given lower contribution to composite kernel function. Then, the composite kernel function for multiple local SVRs is constructed according to the distribution structure of training data.

The performance of the proposed two models is very similar on original data. The approximation and generalization abilities of the MFSVR with \( \varepsilon \)-FCM is largely improved in comparison with the abilities of the MFSVR with FCM. Consequently, the performance of the proposed MFSVR with \( \varepsilon \)-FCM outperforms the MFSVR with existing FCM when the noise and outliers was improved increasing. It can be drawn to a conclusion that the effect of \( \varepsilon \)-FCM gets more increasing according to the increase of the noise and outliers. In future works, type-2 fuzzy sets could be investigated and their impact on the robustness of the model could be investigated.

### REFERENCES

[1] J. Sjöberg, *Nonlinear Black-Box Modeling in System Identification: A Unified Overview*. Linköping, Sweden: Linköping Univ., 1995.

[2] L. Ljung, *System Identification-Theory for the User*, 2nd ed. Upper Saddle River, NJ, USA: Prentice-Hall, 1999.

[3] Z. Zhang, “Using wavelet network in nonparametric estimation,” *IEEE Trans. Neural Netw.*, vol. 8, no. 2, pp. 227–236, Mar. 1997.

[4] H. Chris, H. Xia, and G. Qiang, *Adaptive Modelling, Estimation and Fusion From Data*, a Neuro Fuzzy Approach. Berlin, Germany: Springer-Verlag, 2002.

[5] V. Vapnik, *The Nature of Statistical Learning Theory*. Springer, 2013, pp. 156–167.

[6] A. J. Smola and B. Schölkopf, “A tutorial on support vector regression,” *Statist. Comput.*, vol. 14, no. 3, pp. 199–222, Aug. 2004.

[7] T. H. Cormen, C. E. Leiserson, R. L. Rivest, and C. Stein, *Introduction to Algorithms*. Cambridge, MA, USA: MIT Press, 2001.

[8] B. Li, Q. Wang, and J. Hu, “Multi-SVM classifier system with piecewise interpolation,” *IEEE Trans. Electron. Electron. Eng.*, vol. 8, no. 2, pp. 132–138, Mar. 2013.

[9] Y. Cheng and J. Hu, “Nonlinear system identification based on SVR with quasi-linear kernel,” in *Proc. Int. Joint Conf. Neural Netw. (IJCNN)*, Jun. 2012, pp. 1–8.

[10] J. L. Bentley, “Multidimensional divide-and-conquer,” *Commun. ACM*, vol. 23, no. 4, pp. 214–229, Apr. 1980.

[11] E. Vazquez and E. Walter, “Multi-output support vector regression,” *IFAC Proc. Volumes*, vol. 36, no. 16, pp. 1783–1788, Sep. 2003.

[12] X.-S. Tan, Z.-M. Yuan, T.-J. Zhou, C.-J. Wang, and J.-Y. Xiong, “Multi-KNN-SVR combinatorial forecast and its application to QSAR of fluorine-containing compounds,” *Chem. J. Chin. Univ.*, vol. 29, pp. 95–99, Jan. 2008.

[13] Z.-Y. Chen and Z.-P. Fan, “Dynamic customer lifetime value prediction using longitudinal data: An improved multiple kernel SVR approach,” *Knowl.-Based Syst.*, vol. 43, pp. 123–134, May 2013.

[14] C.-F. Lin and S.-D. Wang, “Fuzzy support vector machines,” *IEEE Trans. Neural Netw.*, vol. 13, no. 2, pp. 464–471, Mar. 2002.

[15] C.-Y. Zhang, Z. Wang, C.-W. Fei, Z.-S. Yuan, J.-S. Wei, and W.-Z. Tang, “Fuzzy multi-SVR learning model for reliability-based design optimization of turbine blades,” *Fuzzy Sets Syst.*, vol. 13, no. 2, pp. 153–167, Jul. 2019.

[16] W. Pedrycz, “An identification algorithm in fuzzy relational systems,” *IEEE Trans. Syst., Man, Cybern.*, vol. 15, no. 2, pp. 153–167, Jul. 1984.

[17] M. Sugeno and G. T. Kang, “Structure identification of fuzzy model,” *Fuzzy Sets Syst.*, vol. 23, no. 4, pp. 231–239, Apr. 1980.

[18] T. Takagi and M. Sugeno, “Fuzzy identification of systems and its application to modeling and control,” *IEEE Trans. Systems, Man, Cybern.*, vol. SMC-15, no. 1, pp. 116–132, Jan. 1985.

[19] J. M. Leski, “TSK-fuzzy modeling based on \( \varepsilon \)-insensitive learning,” *Fuzzy Syst. Syst.*, vol. 13, no. 2, pp. 181–193, Apr. 2005.

[20] Z. Wang, C. Yang, S.-K. Oh, and Z. Fu, “Multi-radial basis function SVM classifier: Design and analysis,” *J. Electr. Eng. Technol.*, vol. 13, no. 6, pp. 2511–2520, 2018.

[21] T. Tao, E. Zio, and W. Zhao, “A novel support vector regression method for online reliability prediction under multi-state varying operating conditions,” *Rel. Eng. Syst. Saf.*, vol. 177, pp. 35–49, Sep. 2018.
ZHENG WANG received the B.Sc. degree in communication engineering and the M.Sc. degree from Suwon University, Suwon, South Korea, in 2012 and 2016, respectively, where he is currently pursuing the Ph.D. degree with the Department of Electrical Engineering. His research interests include fuzzy set and fuzzy systems, evolution computing, and neural networks.

CHENG YANG received the M.Sc. degree in information science from Waseda University, Japan, in 2013, and the M.Sc. degree in computer engineering from Shanghai University, China, in 2013. He is currently pursuing the Ph.D. degree in electrical engineering with Suwon University, South Korea. His research interests include fuzzy systems, support vector machine, swarm intelligence, and machine learning.

SUNG-KWUN OH (Member, IEEE) received the B.Sc., M.Sc., and Ph.D. degrees in electrical engineering from Yonsei University, Seoul, South Korea, in 1981, 1983, and 1993, respectively. From 1983 to 1989, he was a Senior Researcher with the Research and Development Laboratory, Lucky-Goldstar Industrial Systems Company Ltd. From 1996 to 1997, he was a Postdoctoral Fellow with the Department of Electrical and Computer Engineering, University of Manitoba, Winnipeg, MB, Canada. He is currently a Professor with the School of Electrical and Electronic Engineering, The University of Suwon, Suwon, South Korea. He is also with the Research Center of Big Data and Artificial Intelligence, Linyi University, China. His research interests include fuzzy systems, fuzzy-neural networks, automation systems, advanced computational intelligence, and intelligent control. He currently serves as an Associate Editor for the IEEE Transactions on Systems and Control, the International Journal of Fuzzy Logic and Intelligent Systems of the KFIS, and Information Sciences.

ZUNWEI FU (Member, IEEE) received the M.Sc. degree in applied mathematics from the China University of Mining and Technology, Beijing, China, in 2005, and the Ph.D. degree in mathematical science from Beijing Normal University, Beijing, in 2008. Since 2011, he has been a Professor with Linyi University (LU). He was the Dean of the Research Center of Big Data and Artificial Intelligence, LU. His current research interests include fractional integral operators and the functional space theory in artificial intelligence. He is a Reviewer of Mathematical Reviews (American Mathematical Society). He is also a leading expert in Key Disciplines at the Provincial Level and National Outstanding Teacher. He has served as an Associate Editor for the Journal of LU.

WITOLD PEDRZYC (Life Fellow, IEEE) received the M.Sc., Ph.D., and D.Sci. from the Silesian University of Technology, Gliwice, Poland. He is currently a Professor and a Canada Research Chair (CRC) in computational intelligence with the Department of Electrical and Computer Engineering, University of Alberta, Edmonton, AB, Canada. He is also with the Systems Research Institute, Polish Academy of Sciences, Warsaw, Poland. He has published numerous articles in these areas. He is also the author of 15 research monographs. His research interests include computational intelligence, fuzzy modeling, knowledge discovery and data mining, fuzzy control, including fuzzy controllers, pattern recognition, knowledge-based neural networks, granular and relational computing, and software engineering. He has been a member of numerous program committees of IEEE conferences in the area of fuzzy sets and neurocomputing. He currently serves on the Advisory Board of the IEEE Transactions on Fuzzy Systems. He served as the President of IFSA. He is also an Editor-in-Chief of Information Sciences (Elsevier), WIREs Data Mining and Knowledge Discovery (Wiley), and the International Journal of Granular Computing (Springer). He served as the Editor-in-Chief for the IEEE Transactions on Systems Man and Cybernetics Part A.