

Received 19 November 2023, accepted 30 December 2023, date of publication 5 January 2024, date of current version 12 January 2024.

Digital Object Identifier 10.1109/ACCESS.2024.3350195

RESEARCH ARTICLE

A New Class of Bessel Kernel Functions for Support Vector Machine

YEQIN SHAO^{1,2} AND QUAN JIANG^{1,3}

¹School of Transportation and Civil Engineering, Nantong University, Nantong, Jiangsu 226019, China

²School of Information Science and Technology, Nantong University, Nantong, Jiangsu 226019, China

³School of Science, Nantong University, Nantong, Jiangsu 226019, China

Corresponding author: Quan Jiang (jiang.q@ntu.edu.cn)

This work was supported by the National Natural Science Foundation of China under Grant 61671255.

ABSTRACT In this paper, we construct a Bessel-class kernels for Support Vector Machine. This new class of kernels are proved that they are continuous and satisfy Mercer's condition. The presented Bessel-class kernels can degenerate to Gaussian kernel in an infinite smooth case. Compared to the other kernels, these present Bessel kernels can be flexibly applied to classification and regression with fewer constants to be adjusted. Additionally, four simulation experiments including classifications and regressions, have been carried out to show the good performance of these Bessel-class kernels.

INDEX TERMS Machine learning, support vector machine, Bessel kernel, classification, regression.

I. INTRODUCTION

Support Vector Machine (SVM) is one of the competitive methods for classification and regression in machine learning (ML), and has been widely used in the areas of pattern recognition, speech recognition and text classification, etc., [1], [2], [3], [4]. SVM is based on the structural risk minimization principle and capacity concept with pure combinatorial definitions [5], [6], [7]. Compared with the traditional methods which minimize the empirical training error, SVM can avoid local minima by solving quadratic programming problem of convex objective function with a linear set of constraints [8], [9]. Furthermore, the quality and complexity of the SVM solution does not depend on the dimension of the input space [5], [6], [7].

Based on the statistical learning theory, SVM belongs to the class of Kernel Methods. By choosing different mapping function $\varphi(\mathbf{x})$, one can map the training data into a higher dimensional input space, and give an optimal separating hyperplane in that space. It is noted that SVM has advantage that it does not need the mapped patterns $\varphi(\mathbf{x})$, and only need a kernel function $K(\mathbf{x}_k, \mathbf{x}_l)$ which involves the dot products of these patterns instead. As mentioned before, the kernel


The associate editor coordinating the review of this manuscript and approving it for publication was Jon Atli Benediktsson .

TABLE 1. Some traditional kernels [11].

Kernel type	Kernel function
Linear kernel	$K(\mathbf{x}, \mathbf{z}) = \mathbf{x}^T \cdot \mathbf{z}$
Polynomial kernel	$K(\mathbf{x}, \mathbf{z}) = (\mathbf{x}^T \cdot \mathbf{z} + \theta)^d$
Sigmoid kernel	$K(\mathbf{x}, \mathbf{z}) = \tanh(\eta \mathbf{x}^T \cdot \mathbf{z} + \theta)$
Gaussian kernel	$K(\mathbf{x}, \mathbf{z}) = \exp\left(-\frac{\ \mathbf{x} - \mathbf{z}\ ^2}{2\sigma^2}\right)$

function can be regarded as a similarity measure between the input objects. It is also emphasized that the kernel function must satisfy Mercer's condition [10]. The mapping from the input space into the feature space is explained as well as the "Kernel Trick". With different kernel, SVM can be used to solve the specific problem. Table 1 lists some traditional kernels, which can be applied for most of general problems for SVM. Besides the linear kernel, the other kernel functions based on polynomial, sigmoid function, Gaussian radial basis function (RBF) listed in Table 1, have been successfully applied for many science and engineering problems.

The performance of SVM seriously relies on the kernel. Therefore, the choice of the kernel functions and the corresponding parameters is a key problem for an SVM [12]. To the authors' knowledge, there are no good theories to

select the best kernel function. Most of the selections for the kernel functions in SVMs depend on the experience of researchers without theoretical guidance. At present, most researches have studied the performance of SVM utilizing different kernels. Guo and Zhang et al [13] proposed a method using a kernelized multi-class support vector machine with a fast version of recursive feature elimination. Their proposed feature selection algorithm was efficient and worked well for problems of the large number features. Wang and Zhang et al [14] proposed a kernel function selection mechanism for support vector machine(SVM) under sparse representation. In their paper, the composed kernel function, which was suitable for the given data, can be selected according to the selection mechanism. Basing on the idea of enlarging the spatial resolution to increase separability between classes, Amari and Wu [15] proposed a method of modifying a kernel function. With the help of their method, the performance of a support vector machine classifier was improved. They also extended the method in Ref. [15] to improve the performance of SVMs by conformally transforming kernel functions in a data-dependent way [16]. Based on Gaussian RBF and Polynomial kernels, their simulations for two artificial data sets revealed that the method is effective.

It is noted that the most used kernels are those listed in Table 1, or the combinations of those functions. The linear kernels are suitable for linearly related data. The polynomial kernels are generally applied for the global characteristics of the sample, which have strong generalization ability. The Gaussian RBF kernels have strong local effect to training data, but they are not very good to the overall characteristics of the data. It is also mentioned that the polynomial kernels and Gaussian RBF kernels are difficult to adjust the parameters in some cases. According to the features of the above kernels, one can also establish mixed kernel for some complex situations, instead of the single kernel. If all of the above approaches can not be optimized for selection, it seems that SVM is not valid to be applied for a specific ML problem. In the authors' opinion, the poor performance is induced by the unsuitable kernel. When a good kernel is used, SVM can achieve good results. Therefore, it is necessary to search new kernel type for SVMs. In this paper, we have presented a type of Bessel-class kernels for SVM in ML problem. This new class of kernels are continuous and satisfy the satisfies Mercer's condition [17]. These Bessel-class kernels can degenerate to Gaussian kernels in an infinite case. Additionally, this class of Bessel kernels can be flexibly used with only parameter to be adjusted, which can greatly simplify the computation in ML procedure.

A brief outline of this paper is as follows. In Section II, some basic formulas on SVMs for classification and regression problems are briefly reviewed to ensure the completeness of this paper. In Section III, a new class of kernels based Bessel functions are given and proved. In Section IV, four simulation experiments including classifications and regressions have been given to verify the validation of the new presented kernels. Some conclusions are drawn in Section V.

II. SVM FORMULAS

In this Section, some basic formulas on SVM for classification and regression problems are shortly reviewed. For all further details, one can go to those references [5], [6], [7].

A. FORMULAS OF CLASSIFICATION

SVM is based on a context of convex optimization theory. In the primal weight space, the nonlinear SVM classifier is defined as

$$y(\mathbf{x}) = \text{sign}[\mathbf{w}^T \varphi(\mathbf{x}) + b], \quad (1)$$

where $\mathbf{w} \in \mathbb{R}^n$, $b \in \mathbb{R}$, $\varphi(\mathbf{x})$ is map function, which maps \mathbf{x} into a higher dimensional feature space which can be infinite dimensional.

Consider a set of given training vectors $\mathbf{x}_i \in \mathbb{R}^n$ ($i = 1, \dots, N$), in two classes, and the indicator (feature) vector $y \in \mathbb{R}^N$ with $y_i \in \{-1, 1\}$, the primal SVM for classification can be written as follows

$$\begin{aligned} \min_{\mathbf{w}, b, \xi} \quad & \frac{1}{2} \mathbf{w}^T \mathbf{w} + c \sum_{k=1}^N \xi_k \\ \text{s.t.} \quad & y_k (\mathbf{w}^T \varphi(\mathbf{x}_k) + b) \geq 1 - \xi_k, \\ & \xi_k \geq 0, \quad k = 1, \dots, N, \end{aligned} \quad (2)$$

where ξ_k is slack variable, c is a positive real tradeoff constant.

By applying Lagrangian, Eq.(2) can be solved via the dual problem

$$\begin{aligned} \max_{\alpha} \quad & -\frac{1}{2} \sum_{k,l=1}^N y_k y_l K(\mathbf{x}_k, \mathbf{x}_l) \alpha_k \alpha_l + \sum_{k=1}^N \alpha_k \\ \text{s.t.} \quad & \sum_{k=1}^N y_k \alpha_k = 0, \quad \alpha_k \in [0, c], \quad k = 1, \dots, N, \end{aligned} \quad (3)$$

where α_k are Lagrange multipliers, and $K(\mathbf{x}_k, \mathbf{x}_l)$ is the kernel function as follow

$$K(\mathbf{x}_k, \mathbf{x}_l) = \varphi(\mathbf{x}_k)^T \varphi(\mathbf{x}_l). \quad (4)$$

After the dual problem is solved, the vector \mathbf{w} can then be given as

$$\mathbf{w} = \sum_{k=1}^N y_k \alpha_k \varphi(\mathbf{x}_k), \quad (5)$$

by using primal-dual relationship. Consequently, Eq.(1) can be written by

$$\begin{aligned} y(x) &= \text{sign}(\mathbf{w}^T \varphi(\mathbf{x}) + b) \\ &= \text{sign}\left(\sum_{k=1}^N y_k \alpha_k K(\mathbf{x}_k, \mathbf{x}) + b\right). \end{aligned} \quad (6)$$

B. FORMULAS OF REGRESSION

For a set of given training data $\{\mathbf{x}_k, y_k\} (k = 1, \dots, N)$, the problem for regression in the standard form of SVM can be written as

$$y(\mathbf{x}) = \mathbf{w}^T \varphi(\mathbf{x}) + b, \tag{7}$$

which is similar to Eq.(1).

The quadratic programming(QP) problem for Eq.(7) can be given by

$$\begin{aligned} \min_{\mathbf{w}, b, \xi, \xi^*} & \frac{1}{2} \mathbf{w}^T \mathbf{w} + c \sum_{k=1}^N (\xi_k + \xi_k^*) \\ \text{s.t.} & y_k - \mathbf{w}^T \varphi(\mathbf{x}_k) - b \leq \varepsilon + \xi_k, \\ & \mathbf{w}^T \varphi(\mathbf{x}_k) + b - y_k \leq \varepsilon + \xi_k^*, \\ & \xi_k, \xi_k^* \geq 0, k = 1, \dots, N, \end{aligned} \tag{8}$$

where $\varepsilon > 0$ is the tolerance of accuracy for function estimation, ξ_k and ξ_k^* are the slack variables.

The dual problem for Eq.(8) is

$$\begin{aligned} \max_{\alpha, \alpha^*} & \left\{ -\frac{1}{2} \sum_{k,l=1}^N (\alpha_k - \alpha_k^*)(\alpha_l - \alpha_l^*) x_k^T x_l \right. \\ & \left. - \varepsilon \sum_{k=1}^N (\alpha_k + \alpha_k^*) + \sum_{k=1}^N y_k (\alpha_k - \alpha_k^*) \right\} \\ \text{s.t.} & \sum_{k=1}^N (\alpha_k - \alpha_k^*) = 0, \alpha_k, \alpha_k^* \in [0, c], \end{aligned} \tag{9}$$

where α_k and α_k^* are the Lagrange multipliers.

The vector \mathbf{w} can also be given by

$$\mathbf{w} = \sum_{k=1}^N (\alpha_k - \alpha_k^*) \varphi(\mathbf{x}_k). \tag{10}$$

In the dual space, the function $y(\mathbf{x})$ can be written as follow

$$y(\mathbf{x}) = \sum_{k=1}^N (\alpha_k - \alpha_k^*) K(\mathbf{x}_k, \mathbf{x}) + b. \tag{11}$$

III. KERNELS AND KERNEL TRICK

A. CHARACTERISATION OF KERNELS

It is known that the linear SVM has been extended to a nonlinear technique by Vapnik in 1995 by introducing a kernel function $K(\mathbf{x}_k, \mathbf{x}_l)$, which is obviously the inner product of map function $\varphi(\mathbf{x})$ as shown in Eq.(4). Fortunately, it is not necessarily to know the explicit expression of the mapping, since that one can evaluate the inner products of map functions in Hilbert space by following result. According to Hilbert space theory, a symmetric, continuous function $K(\cdot, \cdot)$ satisfying Mercer’s condition [17], there exists an expansion

$$K(\mathbf{x}, \mathbf{z}) = \sum_{k=1}^{\infty} \lambda_k \varphi_k(\mathbf{x}) \varphi_k(\mathbf{z}), \tag{12}$$

where $\mathbf{x}, \mathbf{z} \in \mathbb{R}^N$, and $\lambda_k > 0$.

Mercer’s condition requires that

$$\int K(\mathbf{x}, \mathbf{z}) g(\mathbf{x}) g(\mathbf{z}) d\mathbf{x} d\mathbf{z} \geq 0, \tag{13}$$

where $g(\cdot)$ is any square integrable function.

It is noted that the integral is taken over a compact subset of \mathbb{R}^N and the kernel function can be written in the form of inner product as shown in Eq.(4). Furthermore, we have the following proposition, which is equivalent to Eq.(13).

Proposition 1: Let X be a finite input space with $K(\mathbf{x}, \mathbf{z})$ a symmetric function on X . Then $K(\mathbf{x}, \mathbf{z})$ is a kernel function if and only if the matrix

$$K(\mathbf{x}_k, \mathbf{x}_l) = [K(\mathbf{x}_k, \mathbf{x}_l)]_{k,l=1}^N, \tag{14}$$

is positive semi-definite (has non-negative eigenvalues).

B. BESSEL KERNELS

It is known that the kernel trick is powerful to handle the non-linear separable problem, by mapping the inseparable data into a higher dimensional space. Although there exists many types of kernels for SVMs, as mentioned before, there is in general no best choice for all of the problems. Generally, the polynomial kernel in low order or Gaussian kernel can be chosen to be a good initial try.

The present study in this paper focuses on a new class of kernels based Bessel functions

$$K_v(\mathbf{x}, \mathbf{z}) = \frac{2^v \Gamma(v+1)}{\|\mathbf{x} - \mathbf{z}\|^v} J_v(\|\mathbf{x} - \mathbf{z}\|), \tag{15}$$

where $J_v(\cdot)$ is the J Bessel function of order v ($v = (k - 1)/2 \geq 0, k = 1, 2, \dots$), $\|\cdot\|$ denotes the standard Euclidean vector norm, $\Gamma(\cdot)$ is the Gamma function.

Lemma 1: The function given by Eq.(15) is continuous, has maximum value $K_v(\mathbf{x}, \mathbf{x}) = 1$ for $\mathbf{x} = \mathbf{z}$ ($\|\mathbf{x} - \mathbf{z}\| = 0$) and degenerates to zero as $\|\mathbf{x} - \mathbf{z}\| \rightarrow \infty$.

proof: From the Poisson integral, the Bessel function can be written as [18]

$$\begin{aligned} J_v(\|\mathbf{x} - \mathbf{z}\|) &= \frac{(\|\mathbf{x} - \mathbf{z}\|/2)^v}{\sqrt{\pi} \Gamma(v+1/2)} \\ &\times \int_0^\pi \cos(\|\mathbf{x} - \mathbf{z}\| \cos \theta) \sin^{2v} \theta d\theta. \end{aligned} \tag{16}$$

With the help of Eq.(15), one has

$$\begin{aligned} K_v(\mathbf{x}, \mathbf{z}) &= \frac{\Gamma(v+1)}{\sqrt{\pi} \Gamma(v+1/2)} \\ &\times \int_0^\pi \cos(\|\mathbf{x} - \mathbf{z}\| \cos \theta) \sin^{2v} \theta d\theta. \end{aligned} \tag{17}$$

From above equation, it is obviously $K_v(\mathbf{x}_0, \mathbf{z}_0)$ exists for arbitrary $\|\mathbf{x}_0 - \mathbf{z}_0\|$. Consider that

$$\begin{aligned} &K_v(\mathbf{x}_0 + \Delta \mathbf{x}_0, \mathbf{z}_0 + \Delta \mathbf{z}_0) \\ &= \frac{\Gamma(v+1)}{\sqrt{\pi} \Gamma(v+1/2)} \\ &\times \int_0^\pi (\cos(\|\mathbf{x}_0 - \mathbf{z}_0\| + \Delta \|\mathbf{x}_0 - \mathbf{z}_0\|) \cos \theta) \sin^{2v} \theta d\theta \end{aligned}$$

$$\begin{aligned}
 &= \frac{\Gamma(v+1)}{\sqrt{\pi}\Gamma(v+1/2)} \times \int_0^\pi (\cos(\|\mathbf{x}_0 - \mathbf{z}_0\| \cos \theta) \\
 &\quad \times \cos(\Delta\|\mathbf{x}_0 - \mathbf{z}_0\| \cos \theta)) \sin^{2v} \theta d\theta \\
 &\quad - \frac{\Gamma(v+1)}{\sqrt{\pi}\Gamma(v+1/2)} \\
 &\quad \times \int_0^\pi (\sin(\|\mathbf{x}_0 - \mathbf{z}_0\| \cos \theta) \sin(\Delta\|\mathbf{x}_0 - \mathbf{z}_0\| \cos \theta)) \\
 &\quad \times \sin^{2v} \theta d\theta. \tag{18}
 \end{aligned}$$

If $\Delta\mathbf{x}_0 \rightarrow 0$ and $\Delta\mathbf{z}_0 \rightarrow 0$, then $\Delta\|\mathbf{x}_0 - \mathbf{z}_0\| \rightarrow 0$ holds. With the help of Eq.(18), one has

$$\lim_{\Delta\mathbf{x}_0, \Delta\mathbf{z}_0 \rightarrow 0} K_v(\mathbf{x}_0 + \Delta\mathbf{x}_0, \mathbf{z}_0 + \Delta\mathbf{z}_0) = K_v(\mathbf{x}_0, \mathbf{z}_0). \tag{19}$$

Thus, $K_v(\mathbf{x}, \mathbf{z})$ is continuous.

Since that $-1 \leq \cos(\|\mathbf{x} - \mathbf{z}\| \cos \theta) \leq 1$ and $\sin^2 \theta = (\sin^2 \theta)^v \geq 0$, we have

$$\begin{aligned}
 K_v(\mathbf{x}, \mathbf{z}) &= \frac{\Gamma(v+1)}{\sqrt{\pi}\Gamma(v+1/2)} \int_0^\pi \cos(\|\mathbf{x} - \mathbf{z}\| \cos \theta) \sin^{2v} \theta d\theta \\
 &\leq \frac{\Gamma(v+1)}{\sqrt{\pi}\Gamma(v+1/2)} \int_0^\pi 1 \cdot \sin^{2v} \theta d\theta \\
 &= \frac{\Gamma(v+1)}{\sqrt{\pi}\Gamma(v+1/2)} \int_0^\pi \cos(\|\mathbf{x} - \mathbf{x}\| \cos \theta) \sin^{2v} \theta d\theta \\
 &= K(\mathbf{x}, \mathbf{x}). \tag{20}
 \end{aligned}$$

Eq.(20) implies that the $K_v(\mathbf{x}, \mathbf{z})$ has maximum value if $\mathbf{x} = \mathbf{z}$ ($\|\mathbf{x} - \mathbf{z}\| = 0$). Furthermore, we also have

$$|K_v(\mathbf{x}, \mathbf{z})| < |K_v(\mathbf{x}, \mathbf{x})| \tag{21}$$

if $0 \leq |\sin^{2v} \theta| \leq 1$, $0 \leq |\cos(\|\mathbf{x} - \mathbf{z}\| \cos \theta)| \leq 1$ and $\mathbf{x} \neq \mathbf{z}$ are considered.

The Bessel function has the expansions as follows [18]

$$\begin{aligned}
 J_v(\|\mathbf{x} - \mathbf{z}\|) &= \frac{\|\mathbf{x} - \mathbf{z}\|^v}{2^v} \sum_{k=0}^\infty \frac{(-1)^k (\|\mathbf{x} - \mathbf{z}\|/2)^{2k}}{k! \Gamma(v+k+1)} \\
 &= \frac{\|\mathbf{x} - \mathbf{z}\|^v}{2^v \Gamma(v+1)} \\
 &\quad \times \left(1 + \Gamma(v+1) \sum_{k=1}^\infty \frac{(-1)^k (\|\mathbf{x} - \mathbf{z}\|/2)^{2k}}{k! \Gamma(v+k+1)} \right), \tag{22}
 \end{aligned}$$

$$\begin{aligned}
 (\|\mathbf{x} - \mathbf{z}\| \ll 1) J_v(\|\mathbf{x} - \mathbf{z}\|) \\
 &= \sqrt{\frac{2}{\pi \|\mathbf{x} - \mathbf{z}\|}} \cos \left(\|\mathbf{x} - \mathbf{z}\| - \frac{\pi}{4} - \frac{v\pi}{2} \right), \\
 &\quad \times (\|\mathbf{x} - \mathbf{z}\| \gg 1) \tag{23}
 \end{aligned}$$

Thus, one can obtain

$$\lim_{\|\mathbf{x}-\mathbf{z}\| \rightarrow 0} K_v(\mathbf{x}, \mathbf{z}) = 1, \quad \lim_{\|\mathbf{x}-\mathbf{z}\| \rightarrow \infty} K_v(\mathbf{x}, \mathbf{z}) = 0, \tag{24}$$

according to Eqs.(15), (22) and (23). \square

Theorem 1: The function given by Eq.(15) satisfies Mercer's condition, and is a kernel function.

There exists two ways to prove THEOREM 1.

proof: From Eq.(15), it is obviously that the function satisfies $K_v(\mathbf{x}, \mathbf{z}) = K_v(\mathbf{z}, \mathbf{x})$. Furthermore, $K_v(\mathbf{x}, \mathbf{z})$ is proved to be continuous by LEMMA 1. The function $K_v(\mathbf{x}, \mathbf{z})$ can be related to Hankel transforms as follow [18]

$$\begin{aligned}
 K_v(\mathbf{x}, \mathbf{z}) &= \frac{2^v \Gamma(v+1)}{\|\mathbf{x} - \mathbf{z}\|^v} J_v(\|\mathbf{x} - \mathbf{z}\|) \\
 &= \frac{2^v \Gamma(v+1)}{(2\pi)^{v+1}} \int_{\|\boldsymbol{\omega}\|=1} \exp(i(\mathbf{x} - \mathbf{z})^T \boldsymbol{\omega}) d\boldsymbol{\omega}, \tag{25}
 \end{aligned}$$

where the subscript $\|\boldsymbol{\omega}\| = 1$ denotes the surface integral over the unit sphere in \mathbb{R}^N .

To an arbitrary vector $\boldsymbol{\lambda} = [\lambda_1, \dots, \lambda_N]^T$, one has

$$\begin{aligned}
 &\sum_{k,l=1}^N \lambda_k \lambda_l K_v(\mathbf{x}_k, \mathbf{x}_l) \\
 &= \frac{2^v \Gamma(v+1)}{(2\pi)^{v+1}} \sum_{k,l=1}^N \lambda_k \lambda_l \int_{\|\boldsymbol{\omega}\|=1} \exp(i(\mathbf{x}_k - \mathbf{x}_l)^T \boldsymbol{\omega}) d\boldsymbol{\omega} \\
 &= \frac{2^v \Gamma(v+1)}{(2\pi)^{v+1}} \int_{\|\boldsymbol{\omega}\|=1} \sum_{k,l=1}^N \lambda_k \lambda_l \exp(i(\mathbf{x}_k - \mathbf{x}_l)^T \boldsymbol{\omega}) d\boldsymbol{\omega} \\
 &= \frac{2^v \Gamma(v+1)}{(2\pi)^{v+1}} \int_{\|\boldsymbol{\omega}\|=1} \left| \sum_{l=1}^N \lambda_l \exp(i\mathbf{x}_l^T \boldsymbol{\omega}) \right|^2 d\boldsymbol{\omega} \geq 0. \tag{26}
 \end{aligned}$$

According to Proposition 1, the theorem is proved. \square

proof: Consider a train set be $\{\mathbf{x}_1, \mathbf{x}_2\} \in \mathbb{R}^N$. The matrix $[K_v(\mathbf{x}_k, \mathbf{x}_l)]_{k,l=1}^2$ can be given as follow

$$[K_v(\mathbf{x}_k, \mathbf{x}_l)]_{k,l=1}^2 = \begin{bmatrix} K_v(\mathbf{x}_1, \mathbf{x}_1) & K_v(\mathbf{x}_1, \mathbf{x}_2) \\ K_v(\mathbf{x}_2, \mathbf{x}_1) & K_v(\mathbf{x}_2, \mathbf{x}_2) \end{bmatrix}, \tag{27}$$

From Lemma 1, one has $1 = K_v(\mathbf{x}_k, \mathbf{x}_k) = |K_v(\mathbf{x}_k, \mathbf{x}_k)| > |K_v(\mathbf{x}_k, \mathbf{x}_l)|$ ($k, l = 1, 2$). Hence, the determinant $\det([K_v(\mathbf{x}_k, \mathbf{x}_l)]_{k,l=1}^2) > 0$ and subdeterminant $\det([K_v(\mathbf{x}_k, \mathbf{x}_l)]_{k,l=1}^1) = 1 > 0$. According to Hurwitz theorem, $[K_v(\mathbf{x}_k, \mathbf{x}_l)]_{k,l=1}^2$ is positive definite.

Suppose the matrix $[K_v(\mathbf{x}_k, \mathbf{x}_l)]_{k,l=1}^M$ is positive definite under the train set $\{\mathbf{x}_1, \dots, \mathbf{x}_M\} \in \mathbb{R}^N$ ($M > 2$). With the help of Hurwitz theorem, the determinant and subdeterminant $\det([K_v(\mathbf{x}_k, \mathbf{x}_l)]_{k,l=1}^P) > 0$ ($P = 1, \dots, M$) hold. When another train data \mathbf{x}_{M+1} is considered, the train set is then $\{\mathbf{x}_1, \dots, \mathbf{x}_M, \mathbf{x}_{M+1}\}$. We can force the $(M+1)$ -th vector to move from ∞ to the real situation \mathbf{x}_{M+1} . According to Lemma 1, $K_v(\mathbf{x}, \mathbf{x}_{M+1})$ is continuous, $K_v(\mathbf{x}_k, \mathbf{x}_{M+1}) = 0$ ($\|\mathbf{x}_k - \mathbf{x}_{M+1}\| \rightarrow \infty$) and $K_v(\mathbf{x}_{M+1}, \mathbf{x}_{M+1}) = 1$ ($\|\mathbf{x}_{M+1} - \mathbf{x}_{M+1}\| = 0$). Consequently, the determinant and subdeterminant $\det([K_v(\mathbf{x}_k, \mathbf{x}_l)]_{k,l=1}^P) > 0$ ($P = 1, \dots, M+1$) hold as the $(M+1)$ -th vector moving to the real situation.

Based on the above mathematical induction, matrix $[K_v(\mathbf{x}_k, \mathbf{x}_l)]_{k,l=1}^M$ is positive definite. \square

Form the above Proofs, one can see that the function given by Eq.(15) is kernel function. Furthermore, PROOF 2 shows that the matrix of function (15) is not just positive semi-definite, but indeed positive definite. It could be concluded that the proposed Bessel-class kernel function has some good properties over the other kernels.

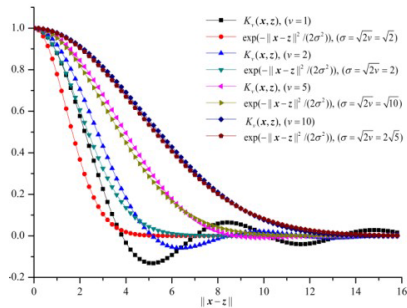


FIGURE 1. Comparison between Bessel-class kernel and Gaussian kernel with $v = 1, 2, 5, 10$.

C. RELATION BETWEEN BESSEL KERNEL AND GAUSSIAN KERNEL

Since that Gaussian kernel has good property, it is widely used in SVM. However, the presented Bessel-class kernels constructed in this paper share this property. We will prove that Gaussian kernel is a type of Bessel-class kernels in an ultimate case.

Theorem 2: The Bessel-class kernel $K_v(\mathbf{x}, \mathbf{z})$ will recover the Gaussian kernel, if $v \rightarrow \infty$, i.e.,

$$\lim_{v \rightarrow \infty} K_v(\mathbf{x}, \mathbf{z}) = \exp\left(-\frac{\|\mathbf{x} - \mathbf{z}\|^2}{2\sigma^2}\right), \quad (28)$$

where $\sigma = \sqrt{2v}$.

proof: From Eqs.(15) and (22), one has

$$\begin{aligned} K_v(\mathbf{x}, \mathbf{z}) &= 1 + \sum_{k=1}^{\infty} \frac{(-1)^k (\|\mathbf{x} - \mathbf{z}\|/2)^{2k} \Gamma(v + 1)}{k! \Gamma(v + k + 1)} \\ &= 1 + \sum_{k=1}^{\infty} \frac{(-1)^k (\|\mathbf{x} - \mathbf{z}\|/2)^{2k}}{k!} \\ &\quad \times \frac{1}{(v + k)(v + k - 1) \dots (v + 1)} \\ &= 1 + \sum_{k=1}^{\infty} \frac{(-1)^k (\|\mathbf{x} - \mathbf{z}\|/2)^{2k}}{k!} \\ &\quad \times \frac{v^{-k}}{(1 + \frac{k}{v})(1 + \frac{k-1}{v}) \dots (1 + \frac{1}{v})} \\ &\quad + \sum_{k=1}^{\infty} \frac{1}{k!} \left(\frac{-\|\mathbf{x} - \mathbf{z}\|^2}{4v}\right)^k \\ &= \sum_{k=0}^{\infty} \frac{1}{k!} \left(\frac{-\|\mathbf{x} - \mathbf{z}\|^2}{4v}\right)^k \\ &= \exp\left(-\frac{\|\mathbf{x} - \mathbf{z}\|^2}{2\sigma^2}\right), (\|\mathbf{x} - \mathbf{z}\| \ll 1), \quad (29) \end{aligned}$$

where $v \rightarrow \infty$ is applied. \square

Theorem 2 illustrates that Gaussian kernel is an infinite smooth case of Bessel-class kernels. It is also shown in Fig.1 that the parameter v control the kernel shape in such a way that the curve becomes flat when v increases.

TABLE 2. Benchmark of raisin grains classification (%).

Performance Measure	Accuracy	Sensitivity	Specificity	Precision
LR [20]	85.22	84.09	86.44	86.89
MLP [20]	86.33	84.57	88.29	88.89
SVM (Polynomial) [20]	86.44	84.17	89.05	89.78
SVM (Gaussian, Training set)	86.86	89.71	84.00	84.86
SVM (Bessel $v=0$, Training set)	85.86	86.86	84.86	85.15
SVM (Bessel $v=1/2$, Training set)	86.71	86.00	87.43	87.25
SVM (Bessel $v=1$, Training set)	87.00	85.71	88.29	87.98
SVM (Bessel $v=16$, Training set)	87.92	90.03	88.03	87.53
SVM (Bessel $v=32$, Training set)	86.86	89.71	84.00	84.86
SVM (Gaussian, Testing set)	84.50	90.00	79.00	81.08
SVM (Bessel $v=0$, Testing set)	82.50	88.00	77.00	79.28
SVM (Bessel $v=1/2$, Testing set)	84.50	88.00	81.00	83.76
SVM (Bessel $v=1$, Testing set)	78.50	98.00	59.00	70.50
SVM (Bessel $v=16$, Testing set)	85.57	90.76	86.43	85.75
SVM (Bessel $v=32$, Testing set)	84.50	90.00	79.00	81.08

IV. SIMULATION EXPERIMENTS

To evaluate the performance of the presented Bessel-class kernel in SVM, four experiments including classifications and regressions have been given in this section. In the simulations, the 5-folds cross-validation has been used, and the optimal parameters are searched by using the Grid Search (GS) and Particle Swarm Optimization (PSO) based on cross-validation. In practice, it is found that PSO is faster than GS. Furthermore, the features are normalized by

$$\bar{x}_{kl} = \frac{x_{kl} - \min(x_{kl})}{\max(x_{kl}) - \min(x_{kl})}, \quad (30)$$

for the features of every sample.

A. RAISIN GRAINS CLASSIFICATION (2 CLASSES)

In this subsection, a benchmark problem for raisin grains classification is performed. The datasets used are taken from website [19]. The data consist of 900 instances with 7 attributes, which are used to make a binary decision on the variety of raisin: Kecimen(450 pieces) or Besni(450 pieces). To perform the classification by SVM, we use sequential selection of 700 training data and 200 testing data, in which the two classes of the raisin grains are equal.

In Reference [20], Cinar et al. have given the benchmark by applying the Logistic Regression (LR), Multi-Layer Perceptron (MLP) and SVM with Polynomial kernel. In this paper, we perform the classification using SVM with Gaussian kernel and Bessel-class kernels. The parameters here, are $c = 4096$, $\sigma = 8.00$ for Gaussian kernel, $c = 0.5$ for Bessel-class kernels with $v = 0, 1/2$, $c = 1000$ for Bessel-class kernel with

TABLE 3. Benchmark of accuracy for Iris classification (%).

Method	Training set	Testing set
LogitBoost NB[22]	94.87	-
Bayesian Network[22]	93.20	-
DECORATE[22]	94.87	-
LBR[22]	93.20	-
AODE[22]	93.07	-
SVM (Gaussian)	100.00	96.67
SVM (Bessel $\nu=0$)	100.00	93.33
SVM (Bessel $\nu=1/2$)	100.00	93.33
SVM (Bessel $\nu=1$)	100.00	93.33
SVM (Bessel $\nu=16$)	100.00	96.67
SVM (Bessel $\nu=32$)	100.00	96.67

$\nu = 1$ and $c = 4101$ for Bessel-class kernel with $\nu = 16, 32$, respectively.

The performance measures of accuracy, sensitivity, specificity and precision are listed in Table 2. It is seen that the performances of the SVM utilizing Bessel-class kernels (especially, $\nu = 16$) are fairly good for the training set and test set, respectively. It is noted that Gaussian kernel is an infinite smooth case of Bessel-class kernels, which can be found in Fig. 1. The Gaussian kernel ($\sigma = 8.00$) is in case of Bessel kernel with parameter $\nu = 32$ under relationship $\sigma = \sqrt{2\nu}$, if $\nu \gg 1$. Therefore, it is seen in Table 2 that, the performance of SVM with Bessel-class kernel ($\nu = 32$) is the same as one with Gaussian kernel ($\sigma = 8.00$). Additionally, it can be found that the performance of Bessel-class kernel with parameter $\nu = 16$ is the best for the specified train and test data set in this numerical experiment. Some of the performances based on the testing set are not so better than those of training set. It is interpreted that the results obtained in Reference [20] are based on the training set of 900 instances. In this paper, the instances are divided in training set and testing set, in which the number of training samples in this paper is much smaller than that of Reference [20].

B. IRIS CLASSIFICATION (3 CLASSES)

In this experiment, a flower iris classification is given to verify the validation of the presented Bessel-class kernel in SVM. The datasets from the UCI machine learning repository [21] include 150 instances with 4 attributes, which are used to make the 3 classes decision: Setosa (50 pieces), Versicolour (50 pieces) and Virginica (50 pieces). In the numerical experiment, 90 training data (30 pieces for each class) and 60 (20 pieces for each class) testing data are used to train and test the model by SVM based on “one vs one” approach.

We perform the classification using SVM with Gaussian kernel and Bessel-class kernels, comparing with the Bayesian-class classifiers [22], e.g., LogitBoost NB, Bayesian Network, Diverse Ensemble Creation by Oppositional Relabeling of Artificial Training Examples (DECORATE), Lazy Bayesian Rule (LBR) and Average One-Dependence estimation (AODE). The parameters for this experiment are $c = 32, \sigma = 5.65$ for Gaussian kernel,

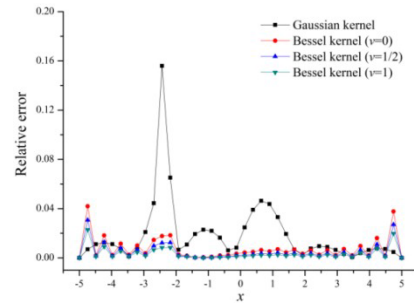


FIGURE 2. Relative errors for the regression based on different kernels.

$c = 32$ for Bessel-class kernels with $\nu = 0, 1/2, 1, 16$ and 32 , respectively.

The performance measures of accuracy are shown in Table 3. The accuracies of Bessel-class kernel are 100% for training set, 93.33% and 96.67% for testing set, respectively. It can be found in Table 3 that, Bessel-class kernel can achieve 100% accuracy in train set with different ν . Furthermore, the accuracy increases to 96.67% when the parameter ν is larger than 16. Additionally, one can see that the parameters $\sigma = 5.65$ and $\nu = 16$ satisfy the relationship $\sigma \approx \sqrt{2\nu}$. Therefore, it is illustrated in Table 3 that the performance of Gaussian kernel ($\sigma = 5.65$) is the same as that of Bessel-class kernel ($\nu = 16$), since that Gaussian kernel is a special case of Bessel-class kernel. It is seen that the SVM utilizing Bessel-class kernels give the excellent performs for all of the training set and testing set.

C. ONE-DIMENSIONAL FUNCTION REGRESSION

This numerical experiment will show the validation of the Bessel-class kernels in the one-dimensional function regression. Consider a polynomial equation as follow

$$y(x) = x^3 + x^2 - x + 5, x \in [-5, +5]. \tag{31}$$

The relative error for this function regression is defined by

$$\text{Relative error} = \left| \frac{y(x_k) - \hat{y}(x_k)}{y(x_k)} \right|, \tag{32}$$

where x_k is testing point, $\hat{y}(x_k)$ is the numerical result obtained by SVM.

In the numerical experiment, 20 training data and 50 testing data are set uniformly in the interval $x \in [-5, +5]$, respectively. To compare the performances of different kernels, we test the problem with different kernels, i.e., Gaussian kernel, Bessel-class kernels with $\nu = 0, 1/2$, and 1. The parameters here, are $c = 32768, \sigma = 8.00$ for Gaussian kernel, $c = 2200$ for Bessel-class kernels, respectively. It is shown in Fig.2 that, the Bessel-class kernels have significantly lower error than Gaussian kernel. Furthermore, the relative error decreases when the parameter ν increases. It implies that the Bessel-class kernels have advantages over the Gaussian kernel in some problems of function regression.

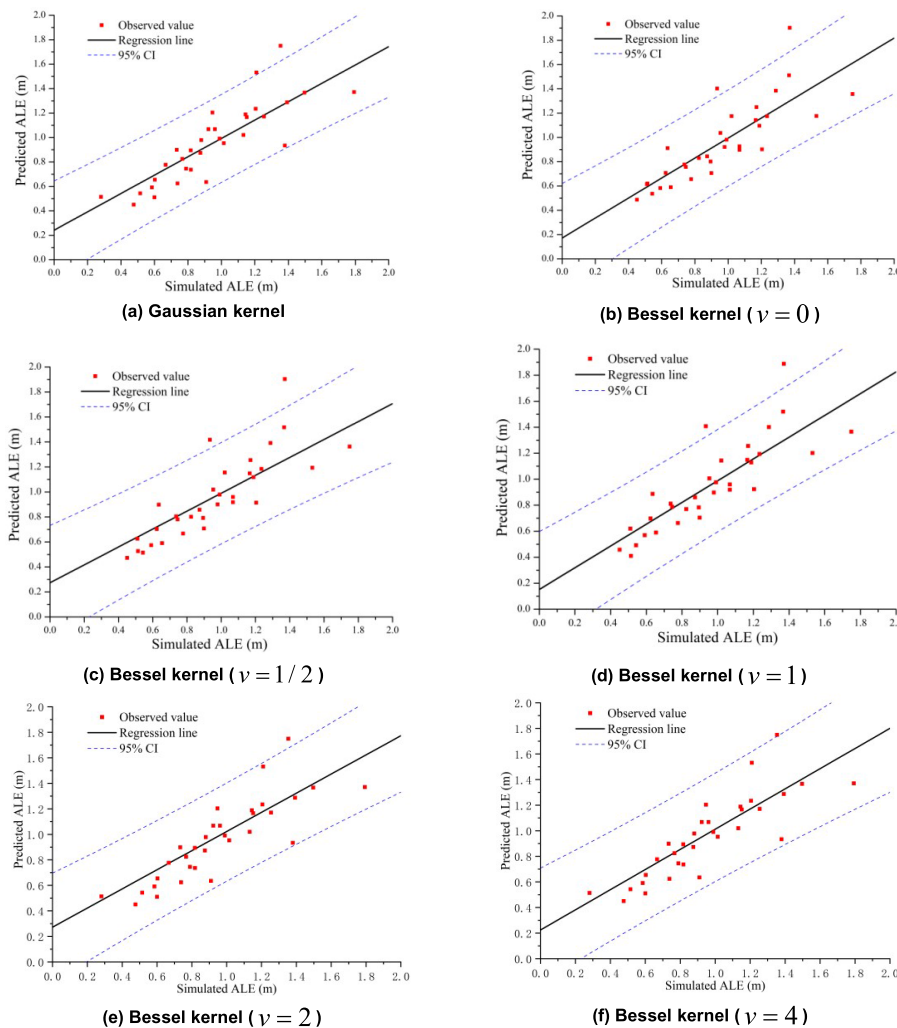


FIGURE 3. Prediction results for ALE using different kernels.

D. ALE REGRESSION IN SENSOR NODE LOCALIZATION PROCESS

The last numerical experiment including ALE regression is proposed to show the performance of the presented Bessel-class kernel in SVM. The datasets given by repository on the website [23] include 107 samples with 4 features. To compare the results with those in reference [24], we have used 75 data for training and the left 32 for testing. In the numerical computing, the Gaussian kernel and Bessel-class kernels are used to solve this problem. The parameters for this experiment are $c = 32$, $\sigma = 2.00$ for Gaussian kernel, $c = 300$ for Bessel-class kernels with $\nu = 0, 1/2, 1$, and $c = 330$ for Bessel-class kernels with $\nu = 2, 4$, respectively.

After training the model, the predicted ALE results have been given by 32 testing data with Gaussian kernel and different Bessel-class kernels. From Fig.3, it is found that predicted results agree well with the simulated results and gathered along the straight regression line with mild scattering points. The most scattering points are in 95% Confidence Interval (CI), which implies that the regression line has a strong positive correlation (R) and relative small Root Mean Square

TABLE 4. Benchmark of the results obtained by different kernels.

Kernel	R	RMSE
Polynomial (Method 1) [24]	0.80	0.234
Polynomial (Method 2) [24]	0.81	0.200
Polynomial (Method 3) [24]	0.82	0.147
Gaussian	0.85	0.167
Bessel ($\nu=0$)	0.82	0.185
Bessel ($\nu=1/2$)	0.83	0.184
Bessel ($\nu=1$)	0.84	0.184
Bessel ($\nu=2$)	0.85	0.165
Bessel ($\nu=4$)	0.87	0.145

Error (RMSE). The detailed value on R and RMSE are list in Table 4. The predicted results obtained by Bessel-class kernels have stronger positive correlation R than those of polynomial kernel. Furthermore, it is shown that the predicted results obtained by Bessel-class kernels are better than the first two results by polynomial kernel. It is illustrated in Table 4 that the performances of Bessel-class kernel with $\nu = 2$ and 4 are much more suitable to ALE regression

than polynomial kernel and Gaussian kernel. As mentioned before, the Gaussian kernel is the ultimate case of the Bessel-class kernel, the prediction given by Bessel-class kernels are nearly close to that obtained by Gaussian kernel. Therefore, it is found in Fig.3(a) and (e) that, the prediction results for ALE are similar by using Gaussian kernel ($\sigma = 2$) and Bessel-class kernel ($\nu = 2$) with the relationship $\sigma = \sqrt{2\nu}$. In these cases ($\sigma = 2$, $\nu = 2$), RMSEs are not the same, since that the parameter ν is not large enough. Furthermore, it is emphasized that R increases and RMSE decreases, when the parameter ν increases which can be found in Table 4.

V. CONCLUSION

A new type of Bessel-class kernels are presented for SVM in this paper. These new kernels are continuous and satisfy the Mercer's condition. It is proved that the Gaussian kernel is an infinite smooth case of Bessel-class kernels, if $\nu \gg 1$. This class of Bessel kernels can be flexibly used with different ν . For a fixed ν , there only exists one tradeoff parameter c to be adjusted, which greatly simplifies the computation in machine learning procedure. Four simulation experiments including classifications and regressions, have been performed to evaluate the validation of the Bessel-class kernels. These simulated results obtained by these kernels have performed significant well in those simulation experiments. Since that there is no best kernel for all the applications, the studies of Bessel-class kernels are still an open subject. The further work for this type of kernels will be done by authors in the future.

ACKNOWLEDGMENT

The authors would like to thank the kind and constructive works of Editors and reviewers for their article.

REFERENCES

- [1] M. Singla, D. Ghosh, and K. K. Shukla, "A survey of robust optimization based machine learning with special reference to support vector machines," *Int. J. Mach. Learn. Cybern.*, vol. 11, no. 7, pp. 1359–1385, Jul. 2020.
- [2] A. Ganapathiraju, J. Hamaker, and J. Picone, "Applications of support vector machines to speech recognition," *IEEE Trans. Signal Process.*, vol. 52, no. 8, pp. 2348–2355, Aug. 2004.
- [3] K. Jonsson, J. Kittler, Y. P. Li, and J. Matas, "Support vector machines for face authentication," *Image Vis. Comput.*, vol. 20, nos. 5–6, pp. 369–375, Apr. 2002.
- [4] H. Lodhi, N. Cristianini, J. Shawe-Taylor, and C. Watkins, "Text classification using string kernels," in *Proc. Adv. Neural Inf. Process. Syst.*, vol. 13, Feb. 2000, pp. 563–569.
- [5] V. Vapnik, *The Nature of Statistical Learning Theory*. New York, NY, USA: Springer-Verlag, 1995.
- [6] V. Vapnik, *Statistical Learning Theory*. Hoboken, NJ, USA: Wiley, 1998.
- [7] V. Vapnik, "The support vector method of function estimation," in *Non-linear Modeling, Advanced Black-Box Techniques*. Norwell, MA, USA: Kluwer, 1998, pp. 55–85.
- [8] C. M. Bishop, *Neural Networks for Pattern Recognition*. Oxford, U.K.: Oxford Univ. Press, 1995.
- [9] V. Cherkassky and F. Mulier, *Learning From Data: Concepts, Theory and Methods*. Hoboken, NJ, USA: Wiley, 1998.
- [10] N. Cristianini and J. Shawe-Taylor, *An Introduction to Support Vector Machines: And Other Kernel-Based Learning Methods*. New York, NY, USA: Cambridge Univ. Press, 2000.
- [11] S. Haykin, *Neural Networks: A Comprehensive Foundation*, 2nd ed. Hoboken, NJ, USA: Prentice-Hall, 1998.
- [12] B. Feizizadeh, M. S. Roodposhti, T. Blaschke, and J. Aryal, "Comparing GIS-based support vector machine kernel functions for landslide susceptibility mapping," *Arabian J. Geosci.*, vol. 10, no. 5, pp. 1–13, Mar. 2017.
- [13] Y. Guo, Z. Zhang, and F. Tang, "Feature selection with kernelized multi-class support vector machine," *Pattern Recognit.*, vol. 117, Sep. 2021, Art. no. 107988.
- [14] B. Wang, X. Zhang, S. Xing, C. Sun, and X. Chen, "Sparse representation theory for support vector machine kernel function selection and its application in high-speed bearing fault diagnosis," *ISA Trans.*, vol. 118, pp. 207–218, Dec. 2021.
- [15] S. Amari and S. Wu, "Improving support vector machine classifiers by modifying kernel functions," *Neural Netw.*, vol. 12, no. 6, pp. 783–789, Jul. 1999.
- [16] S. Wu and S. I. Amari, "Conformal transformation of kernel functions: A data-dependent way to improve support vector machine classifiers," *Neural Process. Lett.*, vol. 15, no. 1, pp. 59–67, 2002.
- [17] N. Cristianini and J. Shawe-Taylor, *An Introduction to Support Vector Machines and Other Kernel-Based Learning Methods*. Cambridge U.K.: Cambridge Univ. Press, 2000.
- [18] Z. X. Wang and D. R. Guo, *Special Functions*. Singapore: World Scientific, 1988.
- [19] M. Koklu. *Raisin Data Set*. Accessed: 2022. [Online]. Available: https://www.researchgate.net/profile/Murat-Koklu/publication/358532830_Raisin_Dataset/data/62067e16cf7c2349ca0a2089/Raisin-Dataset.zip
- [20] I. Cinar, M. Koklu, and S. Tasdemir, "Classification of raisin grains using machine vision and artificial intelligence methods," *Gazi J. Eng. Sci.*, vol. 6, no. 3, pp. 200–209, 2020.
- [21] R. A. Fisher, *Iris*. Irvine, CA, USA: UCI Machine Learning Repository, 1988.
- [22] S. B. Kotsiantis, "Logitboost of simple Bayesian classifier," *Informatica*, vol. 29, no. 1, pp. 53–59, Nov. 2004.
- [23] A. Singh. *ALE Data Set*. Accessed: 2020. [Online]. Available: https://www.researchgate.net/profile/Murat-Koklu/publication/358532830_Raisin_Dataset/data/62067e16cf7c2349ca0a2089/Raisin-Dataset.zip
- [24] A. Singh, V. Kotiyal, S. Sharma, J. Nagar, and C.-C. Lee, "A machine learning approach to predict the average localization error with applications to wireless sensor networks," *IEEE Access*, vol. 8, pp. 208253–208263, 2020.



YE QIN SHAO was born in Zhejiang, China, in 1978. He received the B.S. degree in computer communications and the M.S. degree in pattern recognition from the Nanjing University of Science and Technology, Jiangsu, China, in 2002 and 2004, respectively, and the Ph.D. degree in machine learning from Shanghai Jiao Tong University, Shanghai, China, in 2015. He is currently the Vice Dean of the School of Transportation and Civil Engineering, Nantong University, Jiangsu.

He is also an Associate Professor with Nantong University. He hosted and participated in the National Natural Science Foundation of China Grants and the Jiangsu Natural Science Foundation. His research interests include machine learning and image processing. He is a member of the Jiangsu Artificial Intelligence Society (JSAIS).



QUAN JIANG was born in Jiangsu, China, in 1974. He received the B.S., M.S., and the Ph.D. degrees from Shanghai Jiaotong University, China. He is currently a Professor with Nantong University. His research interests include machine learning for PDE, data analytics, and educational innovation. He is a member of the Solid Mechanics Committee of the Jiangsu Society of Theoretical and Applied Mechanics (JSSTAM) and the Computing Mechanics Committee of JSSTAM.

He received the Science and Technology Award for Young Scientists from the Nantong Science and Technology Bureau and the Science Award (First Class) from the Ministry of Education of China. He is the Chairperson of the Nantong Society of Theoretical and Applied Mechanics (NSTAM).