Artificial Intelligence Technologies and Applications
C. Chen (Ed.)

doi:10.3233/FAIA231300

Kernel Density Regularized Bayesian Learning Framework for Machining Process Anomaly Detection

Zhipeng Ma [a, b], Yue ZHANG [a, b], Xuebin DAI [a, b], Chen YANG [b], Haoning BI [a, b], Biao MA [a, b], Ming ZHAO[a1]

[a] *State Key Laboratory for Manufacturing Systems Engineering, School of Mechanical Engineering, Xi'an Jiaotong University, Xi'an 710049, Shaanxi Province, China;*
[b] *Ujoin-tech (Shanghai) Co., Ltd., Hangzhou 310051, Zhejiang Province, China.*

Abstract. Healthy and stable machining processes are critical for ensuring machining accuracy and guaranteeing machine safety. However, due to complex machining conditions and harsh service environments, machining processes inevitably suffer from abnormalities, which can lead to product defects, increased scrap rates, and even catastrophic accidents. To address this issue, a kernel density regularized Bayesian learning framework is proposed for machining process anomaly detection. In this work, an adaptive kernel density estimate is first constructed to eliminate outlier interferences and provide prior distributions to subsequent Bayesian learning for improving detection accuracy. On this basis, the Bayesian learning framework is innovatively developed for incorporating prior knowledge and multi-classification models, which presents a scientific interpretation for detection results from a probabilistic perspective. Finally, two practical engineering applications are employed to validate the effectiveness of the proposed method. The results show that the proposed method not only improves the anomaly detection accuracy under time-varying operating conditions but also provides confidence levels for detection results. By these advantages, this work may provide a useful tool for independently perceiving the health conditions of machine tools.

Keywords. Anomaly detection, Bayesian learning, Kernel density estimation, Machine tools, Support vector data description.

1. Introduction

With the development of smart manufacturing technology, the requirements for the intelligence of machine tools are rapidly increasing [1]. Machine tools should have the ability to independently perceive and self-determine. In particular, the performance of the machining process plays a pivotal role in producing high-precision and high-quality components. However, due to harsh operating environments and thermal-force coupling effects, machining processes inevitably suffer from abnormalities such as tool breakage and chips in the spindle. These abnormalities may lead to reduced productivity and product quality, or even damage to the machine tool [2]. Therefore, the

[1] Corresponding author: Ming ZHAO, State Key Laboratory for Manufacturing Systems Engineering, School of Mechanical Engineering, Xi'an Jiaotong University, Xi'an 710049, Shaanxi Province, China;email: zhaomingxjtu@mail.xjtu.edu.cn

anomaly detection of the machining process has attracted extensive attention from both industry and academia and has become an important research field in the manufacturing industry.

Machining processes are complex and dynamic, involving a myriad of parameters and variables that can affect the final product quality. Traditional quality control methods may not be sufficient to identify subtle variations or deviations in the machining process. Benefiting from advanced sensing technologies and the industrial Internet of Things (IoT), a variety of sensors are embedded in machine tools to perceive the operating conditions of machining processes [3]. For this reason, many data-driven approaches were widely used for anomaly detection of machining, such as support vector machine (SVM) [4], Gaussian mixture model (GMM) [5], and probabilistic state space model (PSSM) [6]. However, machine tools primarily work in normal conditions, so it is difficult to obtain abnormal data. As a result, the available data are usually unbalanced in quantity, which leads to the failure of supervised-based methods. In addition, although many deep learning methods have end-to-end convenience, they lack physical interpretation, which leads to the detection results lacking credibility and universality. Therefore, it is urgent to explore an interpretable anomaly detection scheme for machining processes.

It is worth mentioning that support vector data description (SVDD) is a powerful tool for anomaly detection during the machining process [7]. Even though SVDD is designed for one-class classification, it does not require the training data to be completely normal. Additionally, unlike deep neural networks, which are difficult for users to understand, SVDD provides intuitive geometric features that do not require any specific assumptions about the data distribution. These advantages make SVDD widely adopted for anomaly detection of machining processes. Chen *et al.* [8] used SVDD to detect wind turbine blade faults. Tao *et al.* [9] proposed density-regularized SVDD for anomaly detection of machine tools. Zhou *et al.* [10] constructed a hybrid model based on SVDD to further improve detection accuracy.

Although SVDD methods have achieved satisfactory results in specific scenarios, they still encounter some challenges in anomaly detection of the machining process under time-varying conditions. Firstly, current SVDD methods assume that the trained samples are obtained from the same distribution, which is not matched with the practical machining process. Secondly, samples obtained from practical machining are inevitably contaminated by noise. These interferences may cause the current SVDD to fail to describe the actual data distribution. Finally, SVDD can only address one-class classification problems. These drawbacks greatly limit the application of SVDD for anomaly detection of machining processes.

To address the above challenges, kernel density regularized Bayesian learning is proposed for anomaly detection in machining processes under time-varying conditions. In this method, kernel density estimation (KDE) is first developed for evaluating the distribution of training data and eliminating noise interferences. Subsequently, a Bayesian learning framework is constructed for reinterpreting SVDD from a probabilistic perspective. The probabilistic estimation is used to perceive unknown anomaly conditions and to identify interferences from unknown domains. Finally, the effectiveness of the proposed method is validated by acquiring data from practical engineering applications. The results show that the developed approach may provide a useful tool for anomaly detection of machining processes.

The remainder of this paper is organized as follows. A brief description of SVDD is presented in Section 2. Next, the proposed method is elaborated in Section 3. In

Section 4, two engineering application cases are employed to validate the performance of the proposed method. Lastly, some conclusions are summarized in Section 5.

2. Theoretical Background

Mathematically, SVDD constructs a minimal hypersphere that encloses all or most of the training samples in a predetermined space $\mathbb{R}$ and identifies any other uncovered samples as outliers [11]. For any given dataset $\mathbf{D} = \{\boldsymbol{x}_i \in \mathbb{R}\}_{i=1}^{N}$, the hypersphere can be described by

$$\begin{aligned} &\min_{\boldsymbol{R},\boldsymbol{O}} F(\boldsymbol{R},\boldsymbol{O}) = \boldsymbol{R}^2 \\ &\text{s.t.} \left\|\boldsymbol{x}_i - \boldsymbol{O}\right\|_2^2 \le \boldsymbol{R}^2,\ \forall i \end{aligned} \tag{1}$$

where $\boldsymbol{O}$ and $\boldsymbol{R}$ represent the center and radius of the hypersphere, respectively.

To deal with possible interferences in training samples, an improved model is derived by adding a regularization term. This term allows that the distance from each sample $\boldsymbol{x}_i$ to the center $\boldsymbol{O}$ does not have to be strictly less than $\boldsymbol{R}$. The above model is re-expressed as follows:

$$\begin{aligned} &\min_{\boldsymbol{R},\boldsymbol{O},\xi} F(\boldsymbol{R},\boldsymbol{O},\boldsymbol{\chi}) = \boldsymbol{R}^2 + C\sum_{i=1}^{N}\chi_i \\ &\text{s.t.} \left\|\boldsymbol{x}_i - \boldsymbol{O}\right\|_2^2 \le \boldsymbol{R}^2 + \chi_i,\ \chi_i \ge 0,\ \forall i \end{aligned} \tag{2}$$

where C denotes the nonnegative penalty factor and $\boldsymbol{\chi} = [\chi_1, ..., \chi_N]^T$ represents the slack variables.

In the following, the Lagrange multiplier method is introduced to solve the above model. The Lagrange function is depicted as:

$$L(\boldsymbol{R},\boldsymbol{O},\xi,\boldsymbol{\alpha},\boldsymbol{\beta}) = \boldsymbol{R}^2 + C\sum_{i=1}^{N}\chi_i - \sum_{i=1}^{N}\alpha_i\left(\boldsymbol{R}^2 + \xi_i - \left\|\boldsymbol{x}_i - \boldsymbol{O}\right\|_2^2\right) - \sum_{i=1}^{N}\beta_i\chi_i \tag{3}$$

where $\alpha_i \ge 0,\ \beta_i \ge 0$ stand for the Lagrange multipliers.

By substituting the KKT condition into the Lagrangian function, the SVDD model can be derived via the following equation:

$$\begin{aligned} &\boldsymbol{O} = \sum_{i=1}^{N}\alpha_i\phi(\boldsymbol{x}_i) \\ &\boldsymbol{R} = \sqrt{k(\boldsymbol{x}_s,\boldsymbol{x}_s) - 2\sum_{i=1}^{N}\alpha_i k(\boldsymbol{x}_i,\boldsymbol{x}_s) + \sum_{i=1}^{N}\sum_{j=1}^{N}\alpha_i\alpha_j k(\boldsymbol{x}_i,\boldsymbol{x}_j)} \end{aligned} \tag{4}$$

where k () denotes the kernel function, and $\boldsymbol{x}_s$ is the support vector.

For an arbitrary test sample, its distance to the center of the hypersphere can be calculated by:

$$\boldsymbol{D}=\sqrt{k\left(\boldsymbol{x}^{*},\boldsymbol{x}^{*}\right)-2\sum_{i=1}^{N}\alpha_{i}k\left(\boldsymbol{x}_{i},\boldsymbol{x}_{*}\right)+\sum_{i=1}^{N}\sum_{j=1}^{N}\alpha_{i}\alpha_{j}k\left(\boldsymbol{x}_{i},\boldsymbol{x}_{j}\right)} \tag{5}$$

where $\boldsymbol{x}^*$ represents the new test sample. If $\boldsymbol{D} > \boldsymbol{R}$, $\boldsymbol{x}^*$ will be recognized as an outlier, otherwise it will be considered as a normal sample.

3. Proposed Method

In the practical machining process, machining a component requires multiple steps, which produces variable machining processes. The collected normal samples usually obey multiple distributions rather than a single distribution. As a result, multiple hyperspheres need to be constructed for anomaly detection under variable operating conditions. To address this issue, a kernel density regularized Bayesian learning is constructed for anomaly detection under time-varying operating conditions. In this work, the KDE is first proposed for estimating the distribution of training samples and removing noise interferences. On this basis, the Bayesian framework is then constructed for reinterpreting the SVDD from a probabilistic perspective. Multi-class model is integrated into the above framework to build hyperspheres for each class to contain as many observations as possible and to keep other classes as far away from the observations as possible. Finally, the superiority of the proposed method is validated by practical engineering applications.

3.1 Kernel Density Estimate

As mentioned above, some outliers will inevitably be confused with normal samples during complex machining. This will make the SVDD very sensitive to outliers in the training process, which could result in overfitting. In light of this, KDE is introduced for outlier detection and prior distribution estimation [12]. The KDE for each sample $\boldsymbol{x}_i$ is defined as:

$$\hat{p}\left(\boldsymbol{x}_{i}\right)=\frac{1}{N}\sum_{j=1}^{N}h_{i}^{-d}K\left(\boldsymbol{x}_{i}-\boldsymbol{x}_{j}\right) \tag{6}$$

where $K(\boldsymbol{x})$ is the Gaussian kernel function, and h_i indicates the width parameter.

The KDE is a measure that compares the density of each sample with its neighboring samples. It defines an outlier detection method that effectively depicts complex data structures. It is worth mentioning that the width parameter h_i determines the performance of density estimation. The width parameter h_i is denoted by:

$$h_{i}=c\left[d_{k-\max}+d_{k-\min}+\varepsilon-d_{k}\left(\boldsymbol{x}_{i}\right)\right] \tag{7}$$

where c is the scaling factor controlling the smoothing effect, ε denotes a small positive number to ensure that the width is non-zero, usually set to 10^{-5}, and $d_{k\text{-max}}$ and $d_{k\text{-min}}$ are the largest and smallest values of the distance, respectively.

To better describe the sample distribution, the Mahala Nobis distance is employed to measure the distance between individual samples. The d_k is calculated by:

$$d_k\left(\boldsymbol{x}_i\right)=\frac{1}{k-1}\sum_{j\in N_k\left(\boldsymbol{x}_i\right)} d\left(\boldsymbol{x}_i,\boldsymbol{x}_j\right) \tag{8}$$

where d () represents the Mahala Nobis distance.

As expected, the KED can be utilized to measure the importance of the samples. High-density training samples are more likely to be located within the hypersphere than low-density samples. Furthermore, the KDE can also estimate the distribution of each sample, which provides a priori distributions for subsequent probability estimation.

3.2 Bayesian Learning Framework

To cope with the complex machining process, a multi-class SVDD is introduced for the health monitoring of machine tools. It constructs n hyperspheres around each normal class, where each hypersphere contains as many observations as possible in its class while keeping observations from other classes outside the boundary [13].

For given k classes $\mathbf{D}_1=\left\{\boldsymbol{x}_i^1\in\mathbb{R}\right\}_{i=1}^{N_1},\ldots,\mathbf{D}_k=\left\{\boldsymbol{x}_i^k\in\mathbb{R}\right\}_{i=1}^{N_k}$, the multi-class SVDD is constructed as follows.

$$\begin{aligned}
&\min_{\boldsymbol{R}_k,\boldsymbol{O}_k} F_k\left(\boldsymbol{R}_k,\boldsymbol{O}_k\right)=\sum\nolimits_{k=1}^{K}\boldsymbol{R}_k^2\\
&\text{s.t.}\left\|\boldsymbol{x}_i^k-\boldsymbol{O}_k\right\|_2^2\le\boldsymbol{R}_k^2,\ \forall i,k\\
&\quad\left\|\boldsymbol{x}_i^m-\boldsymbol{O}_k\right\|_2^2\ge\boldsymbol{R}_k^2,\ m\ne k
\end{aligned} \tag{9}$$

where $\boldsymbol{O}_k$ and $\boldsymbol{R}_k$ denote the center and radius of each hypersphere, respectively.

Similarly, the Lagrange multiplier method is introduced to solve the above model. The Lagrange function is presented as follows:

$$\begin{aligned}
L\left(\boldsymbol{R}_k,\boldsymbol{O}_k,\chi_i^k,\varsigma_i^{(m,k)}\right)=&\sum\nolimits_{k=1}^{K}\boldsymbol{R}_k^2+\sum\nolimits_{k=1}^{K}\left[C_k\sum_{i=1}^{N_k}\chi_i^k\right]+\sum_{m=1}^{K}\sum_{m\ne k=1}^{K}\left[B_m^k\sum_{i=1}^{N_m}\varsigma_i^{(m,k)}\right]\\
&-\sum_{k=1}^{K}\sum_{i=1}^{N_k}\alpha_i^k\left(\boldsymbol{R}_k^2+\chi_i^k-\left\|\boldsymbol{x}_i^k-\boldsymbol{O}_k\right\|_2^2\right)-\sum_{m=1}^{K}\sum_{m\ne k=1}^{K}\sum_{i=1}^{N_m}\beta_i^{(m,k)}\left(\left\|\boldsymbol{x}_i^k-\boldsymbol{O}_k\right\|_2^2-\boldsymbol{R}_k^2+\varsigma_i^{(m,k)}\right)\\
&-\sum_{k=1}^{K}\sum_{i=1}^{N_k}\gamma_i^k\chi_i^k-\sum_{m=1}^{K}\sum_{m\ne k=1}^{K}\sum_{i=1}^{N_m}\gamma_i^{(m,k)}\varsigma_i^{(m,k)}
\end{aligned} \tag{10}$$

where $\alpha_i^k\ge 0,\ \beta_i^{(m,k)}\ge 0,\gamma_i^{(m,k)}\ge 0,\varsigma_i^{(m,k)}\ge 0$ are Lagrangian dual variables.

By substituting the KKT condition into the Lagrangian function, the hypersphere model can be derived as follows:

$$\boldsymbol{O}_k = \sum_{i=1}^{N_k} \alpha_i^k \boldsymbol{x}_i^k - \sum_{k \neq m=1}^{N} \sum_{i=1}^{N_m} \beta_i^{(m,k)} \boldsymbol{x}_i^m$$
$$\boldsymbol{R}_k = \left\| \boldsymbol{x}_s - \boldsymbol{O}_k \right\|_2^2, \forall k = 1, \ldots, K \tag{11}$$

where $\boldsymbol{x}_s$ denotes the support vectors.

In the following, Bayesian learning is introduced to reinterpret the multi-class SVDD. Each sample in the high-dimensional space is assumed to follow a multidimensional Gaussian distribution.

$$\phi\left(\boldsymbol{x}_j^k\right) \sim N\left(\sum_{i=1}^{N_k} \alpha_i^k \phi\left(\boldsymbol{x}_i^k\right) - \sum_{k \neq m=1}^{N} \sum_{i=1}^{N_m} \beta_i^{(m,k)} \phi\left(\boldsymbol{x}_i^m\right), \sigma_{kk}^2 \mathbf{I} \right) \tag{12}$$

where $\sigma_{kk}^2 \mathbf{I}$ represents the covariance matrix.

Furthermore, we assume that all Lagrange multipliers are independent and identically distributed. Thus, the prior distributions of the dual variables α_i^k, $\beta_i^{(m,k)}$, $\gamma_i^{(m,k)}$, and $\varsigma_i^{(m,k)}$ are Gaussian distributions. Among them, the conjugate prior for α_i^k is presented as follows:

$$p\left(\alpha\right) = \frac{1}{\sigma^{N_m} \left(2\pi\right)^{N_m/2}} \exp\left(-\frac{1}{2\sigma^2} \left\| \beta^{(m,k)} - m^{(m,k)} \right\|_2^2 \right) \tag{13}$$

where $\alpha = \left[\alpha_1^k, \ldots, \alpha_{N_k}^k \right]^T$.

Subsequently, the Lagrange multiplier likelihood function and the data distribution prior are substituted into the Bayesian formula to derive the posterior distribution of the dual variables.

$$p\left(\alpha \mid \mathbf{D}\right) = \frac{p\left(\mathbf{D} \mid \alpha\right) p\left(\alpha\right)}{p\left(\mathbf{D}\right)} \tag{14}$$

where $p\left(\mathbf{D} \mid \alpha\right)$ is the likelihood function, and $p\left(\mathbf{D}\right) = \hat{p}\left(\boldsymbol{x}_i^k, \mathbf{D} \subset \left\{ \boldsymbol{x}_i^k \right\}_{i=k=1}^{N_k} \right)$ defines the prior distribution of samples.

Finally, the dual variables are derived by the maximum posterior distribution. It is expressed as follows.

$$\hat{\alpha} = \arg\max_{\alpha} p\left(\alpha \mid \mathbf{D}\right) \tag{15}$$

The above optimization problem is a quadratic programming problem, which means that it has an optimal solution. Substituting the derived solution into Eq. (11), the multi-class hypersphere model can be obtained for anomaly detection. The proposed method not only performs anomaly detection for machining processes under multiple conditions but also provides probabilities for each sample. By these advantages, it may provide a useful solution for the health management of machine tools.

4. Experimental Verifications

In this section, the performance of the proposed method is validated by using experimental data collected from the compact CNC machining center. The experimental platform is shown in Fig. 1. The machining parameters are designed for a spindle speed of 10000 rpm and the workpiece material is aluminum alloy. In addition, to better monitor the machining process and to reduce signal transmission losses, an accelerometer is mounted on the spindle table to measure cutting vibrations in the X, Y, and Z directions. The output of the accelerometer is captured by the acquisition card with a frequency of 3000 Hz. The detailed parameters of the data acquisition system are shown in Table 1.

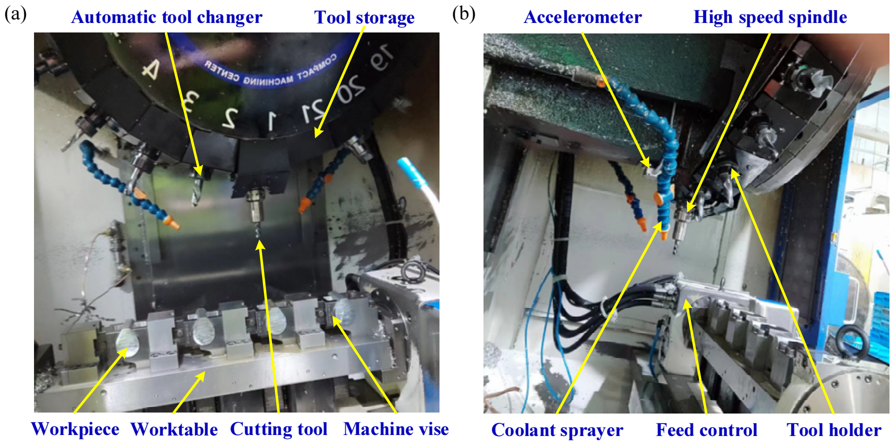

Fig. 1. (a) Compact CNC machining center; (b) The mounted position of the accelerometer.

Table 1. The detailed parameters of the data acquisition system.

Accelerometer type	Data acquisition card	Sampling frequency	Mounting type
IEPE 8688 A5	NI USB-9201	3000Hz	Magnetic base

To improve the detection efficiency and remove redundant information, signal features are first extracted, such as time-domain features, frequency-domain features, and time-frequency-domain features. Besides, the principal component analysis (PCA) is adopted to visualize detection results. Fig. 2 (a) presents the detection results. The proposed method not only effectively detects anomalies for the four processing conditions but also provides probability estimates for each sample. These probabilities give the confidence degree that the sample belongs to this class. It is worth mentioning that the proposed method offers possibilities for monitoring unknown faults. Fig. 2 (b) shows the anomaly detection accuracy. It can be noticed that the accuracy of the proposed method reaches 98.24% under multiple machining conditions, which further validates the effectiveness of the developed method.

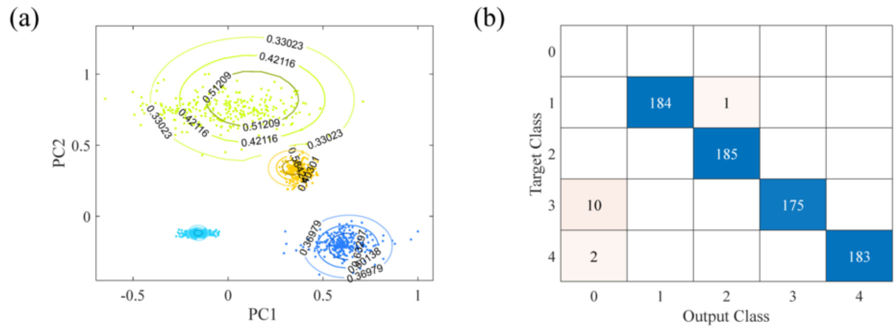

Fig. 2. Case 1 (a) Anomaly detection results obtained with proposed method; (b) The confusion matrix.

In the following, the second engineering application case is further employed to validate the performance of the developed method. The original signal is first transformed into a feature matrix. Then, the PCA is introduced to visualize the detection results. The anomaly detection results are depicted in Fig. 3 (a). It is observed from this figure that the generated decision boundary contains the normal sample space well. For each class, the probability that the sample belongs to this class is provided for recognizing outliers. Benefiting from the adaptive probability estimation strategy, the proposed method can provide a scientific interpretation for each sample from a probabilistic perspective, which improves the credibility of anomaly detection results. Additionally, Fig. 3 (b) displays the confusion matrix. It is noticed that the accuracy reaches 96.19%, which effectively demonstrates the performance of the proposed method. Therefore, the developed approach may provide a useful tool for the health monitoring of machining processes.

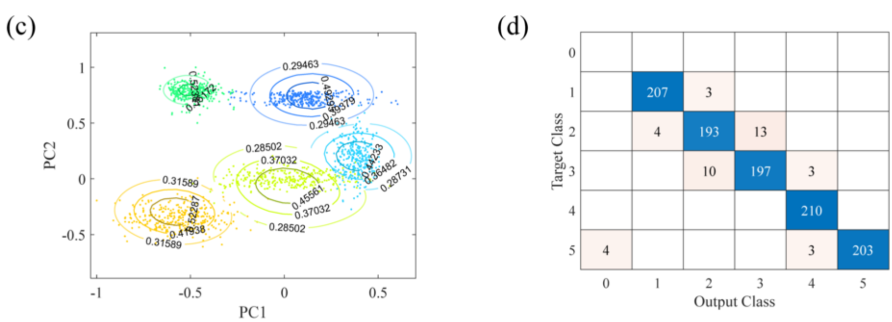

Fig. 3. Case 2 (a) Anomaly detection results obtained with proposed method; (b) The confusion matrix.

5. Conclusions

In this paper, a kernel density regularized Bayesian learning is constructed for machining process anomaly detection. The integration of kernel density estimation with Bayesian learning provides a powerful framework to effectively recognize outliers under time-varying operating conditions. By constructing the kernel density estimation, the proposed approach can adaptively model the underlying probability distribution of training samples, which not only removes the noise interference in advance but also provides a prior distribution for Bayesian estimation. Subsequently, the Bayesian learning framework further improves accuracy and robustness under different machining conditions by incorporating prior distributions and multiple classification models. Finally, the data collected from the practical machining process is employed to validate the effectiveness of the proposed method. The results show that the proposed method exhibits strong performance even with limited training data and imbalanced datasets, which may provide a useful solution for CNC machine tool health management.

Acknowledgment

This research is supported by the National Natural Science Foundation of China (Grant No. 52275127), and the National Key Research and Development Program of China (No. 2021YFB2011400). We are particularly grateful for the data sharing provided by Ujoin-tech (Shanghai) Co., Ltd.

References

[1] S. Yan, H. Shao, Y. Xiao, B. Liu, J. Wan, Hybrid robust convolutional autoencoder for unsupervised anomaly detection of machine tools under noises, Robotics, and Computer-Integrated Manufacturing. 79 (2022) 102441.

[2] S. Han, N. Mannan, D.C. Stein, K.R. Pattipati, G.M. Bollas, Classification and regression models of audio and vibration signals for machine state monitoring in precision machining systems, Journal of Manufacturing Systems. 61 (2021) 45–53.

[3] A. Majeed, Y. Zhang, S. Ren, J. Lv, T. Peng, S. Waqar, E. Yin, A big data-driven framework for sustainable and smart additive manufacturing, Robotics and Computer-Integrated Manufacturing. 67 (2021) 102026.

[4] M.C. Gomes, L.C. Brito, M. Bacci da Silva, M.A. Viana Duarte, Tool wear monitoring in micromilling using Support Vector Machine with vibration and sound sensors, Precision Engineering. 67 (2021) 137–151.

[5] N. Ding, H.X. Ma, H. Gao, Y.H. Ma, G.Z. Tan, Real-time anomaly detection based on long short-term memory and Gaussian Mixture Model, Computers and Electrical Engineering. 79 (2019).

[6] Z. Ma, M. Zhao, X. Dai, Y. Chen, A hybrid-driven probabilistic state space model for tool wear monitoring, Mechanical Systems, and Signal Processing. 200 (2023) 110599.

[7] Y.P. Zhao, Y.L. Xie, Z.F. Ye, A new dynamic radius SVDD for fault detection of aircraft engine, Engineering Applications of Artificial Intelligence. 100 (2021) 104177.

[8] B. Chen, S. Yu, Y. Yu, Y. Zhou, Acoustical damage detection of wind turbine blade using the improved incremental support vector data description, Renewable Energy. 156 (2020) 548–557.

[9] X. Tao, W. Chen, X. Li, X. Zhang, Y. Li, J. Guo, The ensemble of density-sensitive SVDD classifier based on maximum soft margin for imbalanced datasets, Knowledge-Based Systems. 219 (2021) 106897.

[10] Y. Zhou, X. Liang, W. Zhang, L. Zhang, X. Song, VAE-based Deep SVDD for anomaly detection, Neurocomputing. 453 (2021) 131–140.

[11] T. Hu, Q. Guo, H. Sun, T.E. Huang, J. Lan, Nontechnical Losses Detection through Coordinated BiWGAN and SVDD, IEEE Transactions on Neural Networks and Learning Systems. 32 (2021) 1866–1880.

[12] P. Lv, Y. Yu, Y. Fan, X. Tang, X. Tong, Layer-constrained variational autoencoding kernel density estimation model for anomaly detection, Knowledge-Based Systems. 196 (2020) 105753.

[13] M. Turkoz, S. Kim, Y. Son, M.K. Jeong, E.A. Elsayed, Generalized support vector data description for anomaly detection, Pattern Recognition. 100 (2020) 107119.