Journal of Intelligent Manufacturing
https://doi.org/10.1007/s10845-020-01631-9

DyS-IENN: a novel multiclass imbalanced learning method for early warning of tardiness in rocket final assembly process

Zilong Zhuang[1] · Liangxun Guo[1] · Zizhao Huang[1] · Yanning Sun[1] · Wei Qin[1] · Zhao-Hui Sun[1]

Received: 14 November 2019 / Accepted: 12 July 2020

Abstract

Establishing an effective early warning mechanism for the rocket final assembly process (RFAP) is crucial for the timely delivery of rockets and the reduction of additional production costs. To solve the unsystematic design of warning indicators and warning levels in RFAP and address the problem of low warning accuracy caused by imbalanced data distribution, this paper redesigns the warning indicators and warning levels in a systematic way, and develops a novel multiclass imbalanced learning method based on dynamic sampling algorithm (DyS) and improved ensemble neural network (IENN). The DyS algorithm dynamically determines the training set after oversampling the minority class, while the IENN can effectively suppress the oscillation in the iterative process of the DyS algorithm and improve the overall classification accuracy by removing the redundant and ineffective networks from the ensemble neural network. The experiment results indicate that the proposed method outperforms other methods in terms of accuracy and stability for early warning of tardiness in RFAP.

Keywords Early warning of tardiness · Rocket final assembly · Imbalanced learning · Ensemble neural network · Dynamic sampling

Introduction

The rocket final assembly process (RFAP) is the process of assembling the mechanical system, electrical system, control system, etc., which accounts for more than 50% of the manufacturing cycle of the whole rocket (Rui et al. 2016). However, the cycle time of RFAP fluctuates intensively with a high risk of delay due to the dynamic events encountered during the process such as random outage, temporary worker deployment, unfixed assembly time, etc. To avoid delay, the method based on manual experience is often utilized to evaluate and adjust the assembly progress in actual production. However, the factors considered are limited, subjective, and hysteresis, which leads to frequent overtime work near the delivery date and high additional production costs. Therefore, it is crucial to ensure the timely delivery of final assembly orders and reduce additional production costs by establishing a scientific early warning mechanism for RFAP (Sheng et al. 2019).

✉ Wei Qin
wqin@sjtu.edu.cn

[1] School of Mechanical Engineering, Shanghai Jiao Tong University, Shanghai, China

The research on early warning problem originated from the demand of human beings to resist natural disasters (Joseph 2011; Wilhite 2012). With the convenience of data collection in the fields of construction, manufacturing, etc., the early warning of large-scale engineering projects has aroused extensive interest (Xiong et al. 2014; Zhang et al. 2017; Jain and Lad 2019). The early warning methods can mainly be divided into two categories. The first method equates early warning with the prediction (Wang et al. 2017), that is, to directly predict the remaining completion time under the current system state by using probability model (Wang et al. 2016a; Yuan et al. 2017), artificial neural network model (Zheng et al. 2017; Gürbüz et al. 2019) and simulation model (Ramezani and Rothe 2017; Cao and Lam 2018), and then identify the warning level according to the overdue degree of the predicted manufacturing cycle. The other is to conduct early warning based on prediction. Specifically, the prediction result is taken as one of the indicators to describe the future development trend of the system and combined with other indicators to further identify the early warning level. The former is easy to implement but has low accuracy of early warning, and unable to trace the source of warning information when the warning occurs. The latter separates the prediction process from the early warning process and pays more attention to the macroscopicity and generality of the input parameters of the early warning model, which can better reflect the source of warning information and help exclude warning information pertinently.

Compared with conventional completion time prediction problem, the early warning of tardiness in RFAP remains a tough task due to the following two challenges.

One challenge comes from the unsystematic design of warning indicators and warning levels, as well as the complex nonlinear relationship between them. Many factors affect the completion time of RFAP, but few studies have investigated the definition of warning indicators and warning levels, which cannot effectively support the early warning of tardiness in RFAP (Sheng et al. 2019). Besides, when establishing the model between the input warning indicators and the output warning levels, the commonly used method is the linear weighting method, which is not suitable for the early warning of tardiness in RFAP with the complex nonlinear relationship between input and output.

The other challenge stems from the imbalanced distribution of early warning data in RFAP. Due to the strict time control implemented in RFAP, the sample with delay risk accounts for a low proportion of the whole sample in actual historical data, and the higher the risk, the lower the proportion. The imbalanced data distribution of RFAP will easily lead to prejudice in the identification of early warning level. Specifically, a small number of serious warning samples will reduce the identification ability of the early warning model to the serious warning, resulting in serious losses. The existing early warning literature has not yet studied this important characteristic of early warning data.

To address these two issues, this paper redesigns the warning indicators to reflect the impact of various delay risks and the warning levels to reflect the magnitude of delay risks after analyzing the variability of the rocket final assembly system (RFAS). Then, this paper develops a novel multiclass imbalanced learning method based on the dynamic sampling algorithm and improved ensemble neural network (DyS-IENN). In the DyS-IENN method, the dynamic sampling (DyS) algorithm is utilized to balance the dataset by dynamically determining the training set after oversampling the minority class, and the ensemble neural network (ENN) is introduced to suppress the oscillation in the iterative process of the DyS algorithm which is generated by selecting different hyper-parameters in the multi-layer perception (MLP) and improved by removing redundant and ineffective networks in the ENN.

The rest of this paper is organized as follows. "Literature review" provides a brief literature review on warning level identification methods and strategies for dealing with multiclass imbalanced classification problems. "Problem formulation" formulates the problem of early warning of tardiness in RFAP. "DyS-IENN method" presents the multiclass imbalanced learning method based on DyS-IENN. "Experimental results" reports numerical experiments of the proposed method for early warning of tardiness in RFAP. Finally, conclusions and future work are summarized in "Conclusions and future work".

Literature review

Over the years, several methods have been developed for warning level identification, such as statistical method (Yuan et al. 2017), analytic hierarchy process (Zheng et al. 2012), and machine learning method (Wang et al. 2016b). Yuan et al. (2017) weighted the current progress deviation rate and predicted completion time deviation rate, and divided the warning interval according to the delay risk of the monitoring points to realize the warning level identification. Since the early warning indicator may be a semantic variable from subjective evaluation, Zheng et al. (2012) utilized the fuzzy analytic hierarchy process to realize the warning level identification. Wang et al. (2016b) employed support vector machine (SVM) to predict apt-galloping weather conditions, and AdaBoost classifier to realize early warning of galloping. Among these methods, the machine learning method has the advantage of balancing model accuracy and modeling cost, but it depends on a large amount of high-quality data.

With the wide application of Internet of things technology and sensing technology in manufacturing workshop, the collection, processing, and storage of data have been realized, which provides a solid data foundation for machine learning methods in high-level applications, such as fault diagnosis (Wen et al. 2017; Zhang et al. 2018; Zhou et al. 2019), cycle time forecasting (Wang and Zhang 2016; Wang et al. 2018; Tan et al. 2020) and quality control (Qin et al. 2018; Wang et al. 2019; Gonzalez-Val et al. 2020), etc. Therefore, for the manufacturing system with the complex nonlinear relationship between warning indicators and warning levels, such as the RFAS studied in this paper, it will be a promising way to establish a machine learning classification model based on a large amount of historical data to realize high-accuracy warning level identification.

However, the underlying pattern of data collected from RFAS is not completely consistent with the hypothesis of machine learning algorithms, especially the class imbalance, which poses new challenges for the application of machine learning algorithms. The early warning of tardiness in RFAP based on the machine learning method can be naturally transformed into the multiclass imbalanced classification problem. For the multiclass imbalanced classification problem, algorithms are usually designed from the perspective of input, output, or classification model (Guo et al. 2017). The pre-sampling method balances the class distribution by over-sampling the minority class or by under-sampling the majority class. The over-sampling method duplicates the examples of the minority class, but the redundant examples due to over-sampling will waste the training time or cause over-fitting, resulting in poor generalization performance (Abdi and Hashemi 2015). The under-sampling method randomly or pertinently deletes the examples of the majority class, which reduces the time and space complexity of the algorithm, but the information losses due to under-sampling will never be recovered and thus deteriorate the performance of classifier (He and Garcia 2009). Besides, the cost-sensitivity method is also one commonly used approach that assigns higher costs for the misclassification of minority class examples (Zhou and Liu 2005). Although the cost-sensitive method is more computationally efficient, it is difficult to set a proper cost matrix for a class imbalance problem, and an inappropriate cost matrix may mislead the training process (Krawczyk et al. 2014).

These two types of methods are performed independently before the training process and remain unchanged during the training process. Another type of method borrows the idea of boosting and dynamically changes the training set during the training process, such as SMOTEBoost (Chawla et al. 2003), DataBoost-IM (Guo and Viktor 2004), RAMOBoost (Chen et al. 2010), etc. To emphasize the minority examples and difficult-to-learn examples simultaneously, the DyS algorithm with MLP as the classifier has been proposed (Lin et al. 2013), which estimates the probability for every example to decide whether the current example will be reused to update the MLP in the next iteration. However, this method still suffers some drawbacks. First, the classifier of this method is the MLP, which is easy to fall into the local minimum with insufficient classification accuracy; secondly, this method has great fluctuation in model accuracy. Therefore, this paper combines IENN with the DyS algorithm to improve the performance of the algorithm in terms of classification accuracy and stability.

Problem formulation

The early warning of tardiness in RFAP is to evaluate the current and future state of the rocket manufacturing cycle and predict the time range and criticality of the abnormal state according to the workshop state and order characteristics. The warning model takes the system warning indicators as input and the system warning levels as output. The designs of warning indicators and warning levels are described in detail in the following sections.

Design of warning indicators

As shown in Fig. 1, the RFAP is usually organized according to the rocket segments in the form of flow shops. The RFAS is a typical manufacturing system that includes the assembly personnel, equipment, materials, process, and assembly environment related to the RFAP (Sheng et al. 2019). After analyzing the sources of interference and their functional mechanisms, the uncertainty caused by incomplete observations of the early warning system can be controlled and eliminated by quantifying the various variability into the warning indicators. To fully reflect the order and production status of the manufacturing workshop, three kinds of early warning indicators are considered, namely overschedule rate, incomplete set influence rate, and worker deployment influence rate. According to whether it reflects the current situation or the future, the overschedule rate can be divided into the current overschedule rate and predicted overschedule rate, so does the incomplete set influence rate. Moreover, the above indicators can be further subdivided according to the object (rocket segments or working group). In this paper, a three-segment rocket is taken as an example, and a total of 12 indicators are used as warning indicators. The definition and calculation of warning indicators are as follows:

1. Overschedule rate

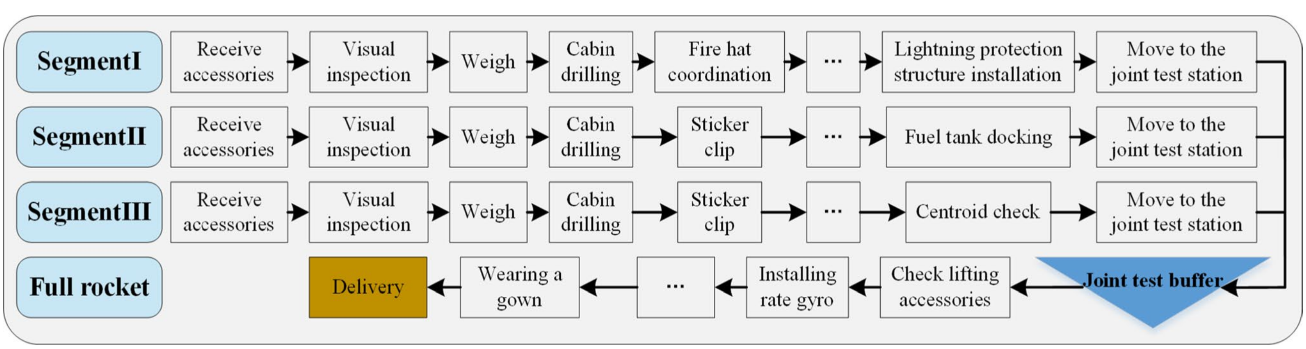

Fig. 1 Flow chart of rocket final assembly process

The overschedule rate is utilized to measure the delay in the production schedule relative to the planned working hours. The current overschedule rate and the predicted overschedule rate are defined by Eqs. (1) and (2), respectively

$$ind_1^s = \frac{\max\left\{\sum_{j=1}^{sn} LT_j^{act} - \sum_{j=1}^{sn} LT_j^{st}, 0\right\}}{\sum_{j=1}^{sn} LT_j^{st}} \times 100\% \tag{1}$$

$$ind_2 = \frac{\max\left\{\sum_{j=1}^{sn} LT_j^{act} + RCT_s^{pred} - CT^{order}\right\}}{CT^{order}} \times 100\% \tag{2}$$

where s denotes the rocket segment number, LT_j^{st} represents the standard labor time of jth process, LT_j^{act} represents the actual working hours from the end of (j-1)th process to jth process, sn is the process number finished of sth segment, RCT_s^{pred} represents the predicted remaining completion time of sth segment that can be predicted according to the current production state, and CT^{order} is the planned completion time of order.

2. Incomplete set influence rate

The influence rate of the incomplete set is measured by the proportion of delay caused by the corresponding key parts. The current and future of incomplete set influence rates are defined by Eqs. (3) and (4), respectively

$$ind_3^s = \frac{\max\limits_{1\le j\le sn}\{ktd_j\}}{\sum_{j=1}^{sn} LT_j^{st}} \times 100\% \tag{3}$$

$$ind_4 = \frac{\max\{\alpha_{sn} E_{sn}(T)\}}{CT^{order}} \times 100\% \tag{4}$$

where $\max\limits_{1\le j\le sn}\{ktd_j\}$ represents the maximum time delay of all assembly stations finished in sth segment rocket caused by the incomplete set, $E_{sn}(T)$ represents the total working delay estimate caused by the incomplete set that is estimated according to the distribution law of arrival time of key parts, and α_{sn} is the correction factor of the estimator.

3. Worker deployment influence rate

The deployment of working groups will not only lead to the stagnation of current assembly tasks, but also affect other rocket assembly tasks that require the same working group, thus forming bottlenecks and waiting queues. The rocket final assembly system involves three assembly groups, namely the mechanical group, electric group, and control group, and these three groups complete independently their respective assembly tasks. The time delay caused by worker deployment can be defined by Eq. (5)

$$ind_5^g = \frac{DT_{sn}^g}{CT^{order}} \times 100\% \tag{5}$$

where g $\in \{1, 2, 3\}$ denotes three assembly groups respectively, DT_{sn}^g represents the amount of time delay caused by worker deployment during the current process, which is estimated by using the deployment duration when it is known, otherwise using the average of historical data.

Design of warning levels

The warning level in the early warning of tardiness in RFAP is the evaluation result of delay risk, which is determined by the severity of predicted delay and the difficulty of adjustment. Accordingly, the warning level is defined by Eq. (6), where $label_{basic}$ represents the basic warning level.

$$label = \frac{\sum_{j=1}^{sn} LT_j^{st}}{CT^{order}} \times label_{basic} \tag{6}$$

According to the actual production, the warning levels are divided into four categories, namely no warning, mild warning, medium warning, and serious warning, which indicate the delay of 0 days, 0–3 days, 3–7 days, and more than 7 days, respectively. Figure 2 shows the data distribution of warning levels after calibrating the warning samples, which indicates the characteristics of data imbalance with a ratio of no warning samples to serious warning samples up to 23:1. It is not difficult to infer that if the general classification algorithms such as SVM, decision tree (DT), MLP, etc., are employed to establish the relationship model between input warning indicators and output warning levels, the model may have a serious bias with extremely low warning accuracy for serious warnings. In Section IV, a novel multiclass imbalanced learning method based on DyS-IENN is described.

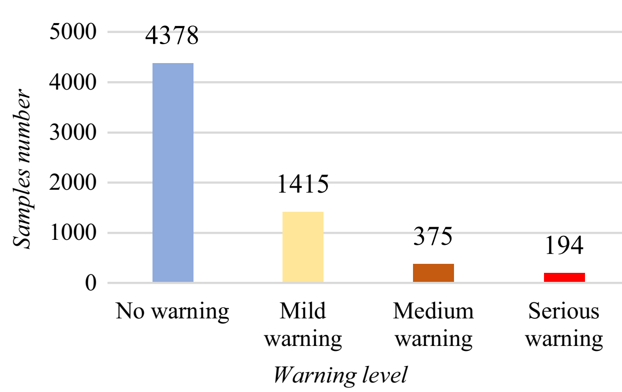

Fig. 2 Distribution of the warning level of samples

DyS-IENN method

The overall framework of the proposed multiclass imbalanced learning method for early warning of tardiness in RFAP is shown in Fig. 3. The method takes 12 warning indicators of five classes proposed in the previous section as input and warning levels defined in Eq. (6) as output. Then it constructs the non-linear mapping relationship between input and output with the DyS-IENN algorithm. Two main modules of the DyS-IENN algorithm are described in detail in the following sections.

Dynamic sampling algorithm

The DyS algorithm is a hybrid sample balancing method (Lin et al. 2013). First, it oversamples the minority class to make the dataset balanced. Then, it feeds the training set into a MLP model to obtain the probability of each training sample belonging to different classes. Finally, it calculates the probability of each training sample being reused as the training sample in the next iteration of MLP according to the difference between the predicted result and real result. Except for the first iteration, the training samples in each iteration are selected dynamically.

The core of the DyS algorithm is to calculate the probability of the ith sample being reused, which is defined by Eq. (7)

$$p^{(i)} = \begin{cases} 1 & if\ \delta^{(i)} \leq 0 \\ \exp(-\delta^{(i)} \cdot \alpha \cdot r_c^{(i)} / \min_k\{r_k\}) & if\ \delta^{(i)} > 0 \end{cases} \tag{7}$$

where c denotes the correct category of the ith sample, $r_c^{(i)}$ represents the number of samples in the category to which ith sample belongs before balancing, $\min_k\{r_k\}$ is the number of samples in the minority class, α is a hyper-parameter which adjusts the basic range of possibility of each category. $\delta^{(i)}$ is the predicted error of ith sample, which can be calculated by Eq. (8):

$$\delta^{(i)} = y_c^{(i)} - \max_{k \neq c}\{y_k^{(i)}\} \tag{8}$$

where y_c denotes the possibility of ith sample being classified into the correct category, $\max_{k\neq c}\{y_k^{(i)}\}$ denotes the maximum possibility of the category to which ith sample belongs (except for the correct category). When the ith sample is misclassified, $\delta^{(i)} < 0$. When the ith sample is correctly classified, $\delta^{(i)}$ indicates the confidence that the ith sample is correctly classified. The larger the $\delta^{(i)}$, the higher the confidence of being correctly classified, and the lower the necessity of being reused as a training sample in the next iteration.

It is easy to conclude from Eqs. (7) and (8) that in the next iteration, the misclassified sample will be directly selected as the training sample, and the correctly classified sample will be repeatedly selected with probability $p^{(i)}$ which is determined by previous classification results and the number of samples in the category to which the ith sample belongs.

Although the DyS algorithm is superior to the general method in terms of sample balancing, it still suffers from instability. When there are multiple dominant categories with similar sample sizes, the classification accuracy of dominant categories may oscillate repeatedly during the experiment, resulting in great fluctuation in the classification accuracy of dominant categories. As shown in Fig. 4, during the iterative process of the DyS algorithm, it may oscillate repeatedly between these categories with similar numbers of samples and fail to converge. To avoid the oscillation phenomenon and increase the stability of the algorithm, this paper will replace the MLP with IENN to take full advantage of ensemble learning with low volatility.

Improved ensemble neural network

Ensemble learning is a machine learning method that completes learning tasks by building and combining multiple learners (Zhang and Ma 2012). It first produces a group of base classifiers and then combines the classification results produced by these base classifiers through some

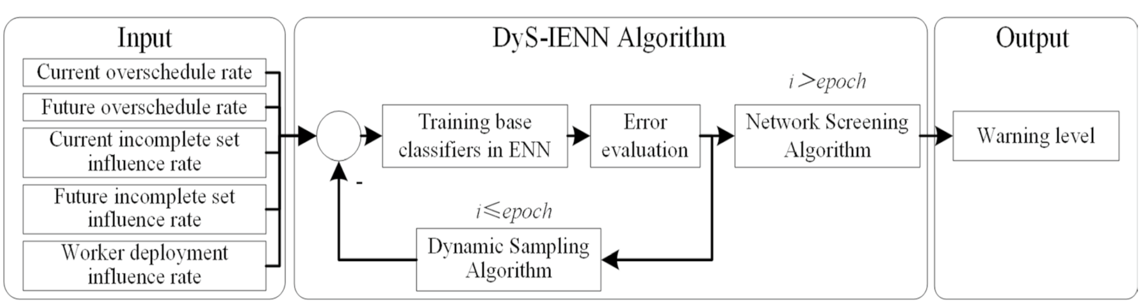

Fig. 3 The overall framework of the multiclass imbalanced learning method based on DyS-IENN

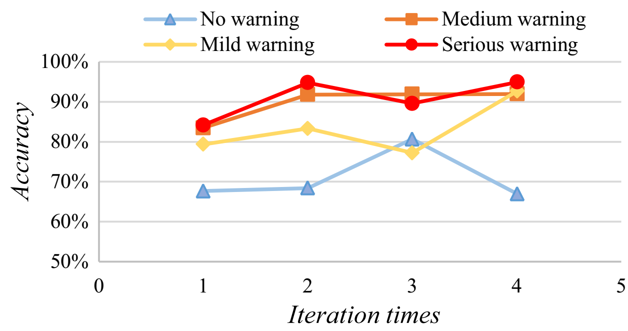

Fig. 4 Trend chart of the classification accuracy during the iteration of DyS

combination strategy to obtain the final classification results. Experimental results show that the combination of base classifiers can achieve significantly better generalization performance than the single learner (Zhou et al. 2002).

The combination strategy of ensemble learning includes serial integration and parallel integration. To solve the instability problem of the DyS algorithm and improve the model accuracy, this paper takes the parallel integration strategy which combines multiple homogeneous base classifiers. Compared with other base classifiers, neural networks usually have the strongest generalization in solving nonlinear classification problems (Rojas 2013), and thus the ensemble learning model using neural networks as the base classifiers will theoretically have the highest classification accuracy. However, due to a large amount of computation, few studies have investigated the ensemble neural networks. To pursue higher classification accuracy, this study develops an improved ensemble neural network, which takes the parallel integration strategy and network screening strategy to balance model accuracy and algorithm efficiency.

The error-ambiguity decomposition model has well stated that the ensemble learning model with higher-accuracy base learners and greater diversity between base learners will have better performance (Krogh and Vedelsby 1995). The randomization process enables different neural networks to evolve under different optimization directions and fall into different local optimums, resulting in greater diversity between multiple neural networks. The randomization strategies involved in this paper include:

a. *Randomized training samples.* The random sampling method generally adopts the Bootstrap sampling method (Kotsiantis et al. 2006). To avoid over-fitting caused by repeated selection of the minority class samples, and better integrate with the DyS algorithm, the completely random sampling method is utilized to make a large difference in the sample space of base classifier.
b. *Randomized activation function.* The activation function of the hidden layer adopts the randomized leaky rectified linear unit (RReLU), which is an asymmetric function and can effectively reduce the probability of generating symmetric local minimum value in the network optimization process. Besides, different alpha values can be taken up by different neural networks which makes the network different as well.
c. *Randomized initial learning rate.* It will also make the learning process of base classifier different by selecting different initial learning rate to train the base classifier.

The base classifier of ENN is the MLP with three hidden layers, where the activation function of hidden layers adopts RReLU and the output layer takes the SoftMax function. All hyper-parameters are listed in Table 1.

The base classifiers of ENN are generated by inheriting most basic parameters and randomly generating some parameters from MLP parameters, and trained independently by using the training samples obtained by the completely random sampling. When classifying new samples, multiple base classifiers are combined organically. Specifically, the average of all base classifier classification labels is viewed as the probability that the sample belongs to each class, and the class with the highest probability is viewed as the class of the sample. The final output class L is defined by Eq. (9):

$$\hat{L} = \arg\max_{i} \left\{ \frac{1}{k} \sum_{j=1}^{k} 1(\hat{y}_j = i) \right\} \quad i = 1, 2, 3, 4 \tag{9}$$

where k is the number of base classifiers and $1(\hat{y}_j = i) = 1$ when the classification result of the jth base classifier is the ith class.

The ENN generates a large number of base classifiers, which can effectively reduce the fluctuation of the DyS algorithm, but does not necessarily improve the classification accuracy. Due to poor structural parameters and insufficient training, some base classifiers do not reach the ideal local

Table 1 Hyper-parameter settings of multi-layer perception

Hidden layer number	Neurons of 3rd hidden layer	Neurons of other hidden layers	Regularization coefficient	Initial learning rate
3	[25,36]	[50,72](2), [100,144](1)	0.0001	[0.01,0.05]
Iteration times	Learning rate decay	Optimized algorithm	Terminate in advance	Activation function
500	0.95	RMSProp	50	RRelu[0,0.25]

optimization and cannot contribute to improving the overall accuracy of the model. Besides, the hyper-parameters of some base classifiers are too close to each other, which may easily cause the overall model effect to degenerate to a base classifier. Therefore, to further improve the overall accuracy of the model, it is necessary to screen base classifiers within the base classifier set.

To ensure that the remaining neural networks are "good but different", this paper designs the Eqs. (10)–(12) to evaluate each neural network and filter out neural networks with low accuracy or highly similarity to other neural networks

$$Score_{cl} = Gmean_{cl} - Diversity_{cl} \tag{10}$$

$$Gmean_{cl} = \left(\prod_{i=1}^{K} \frac{tr_i^{cl}}{r_i} \right)^{\frac{1}{K}} \tag{11}$$

$$Diversity_{cl} = \frac{1}{cluster} \frac{1}{|S_{valid}|} \sum_{i=1}^{cluster} \sum_{j=1}^{|S_{valid}|} \left| \hat{y}_j^{cl} - \hat{y}_j^{i} \right| \tag{12}$$

where $Gmean_{cl}$ represents the geometric mean classification accuracy of the model cl on each category, tr_i^{cl} represents the number of correctly classified by the model cl for samples belonging to the ith class, r_i denotes the total number of samples belonging to the ith class, and K is the number of categories. $Diversity_{cl}$ measures the difference between one base classifier and other base classifiers by using the difference of their prediction results $\sum_{j=1}^{|S_{valid}|} \left| \hat{y}_j^{cl} - \hat{y}_j^{i} \right|$. Both Eqs. (11) and (12) are calculated on the validation set S_{valid}.

The network screening algorithm is shown in algorithm 1, where *Sort*(M, *Score*) represents the ranking of original neural network M, *ValidateSamples*(M_{opt}, S_{valid}) represents the classification accuracy of neural network subset M_{opt} on the verification set S_{valid}, M_{opt}[1:argmax(*GmeanScore*)] represents the optimal subset with highest Gmean, which will have the highest average classification accuracy. The schematic diagram of IENN is shown in Fig. 5. After training all base classifiers, the network screening algorithm is utilized to make the base learners of the obtained ensemble learning model have higher accuracy and greater diversity, which is essential for improving the performance of the ensemble learning model.

Algorithm 1: Network screening algorithm

Input: Neural network set M, Validation set S_{valid}
Output: Optimized network subset M_{opt}
for *cl* = 1 to *cluster*
  $Score_{cl} = Gmean_{cl} - Diversity_{cl}$;
end for
M' = *Sort*(M, *Score*);
$M_{opt} = \varnothing$;
GmeanScore = 0;
for *cl* = 1 to *cluster*
  $M_{opt} = \{M_{opt}, P'_{cl}\}$;
  $GmeanScore_{cl}$ = *ValidateSamples*(M_{opt}, S_{valid});
end for
$M_{opt} = M_{opt}$[1:argmax(*GmeanScore*)];
return M_{opt}

The proposed DyS-IENN method is shown in Algorithm 2, where ρ_i denotes the copy rate of ith class sample, *Dupicate*(S, ρ) duplicates the original training set S with copy rate ρ, *PickSample*(S', p) randomly selects training samples according to sample possibility p, *PickParameter*(P) extracts the hyper-parameters of MLP according to Table 1, *TrainMLP*() and *UpdateMLP*() represent training neural network and updating neural network respectively, and VotePool is the voting pool of ensemble neural network.

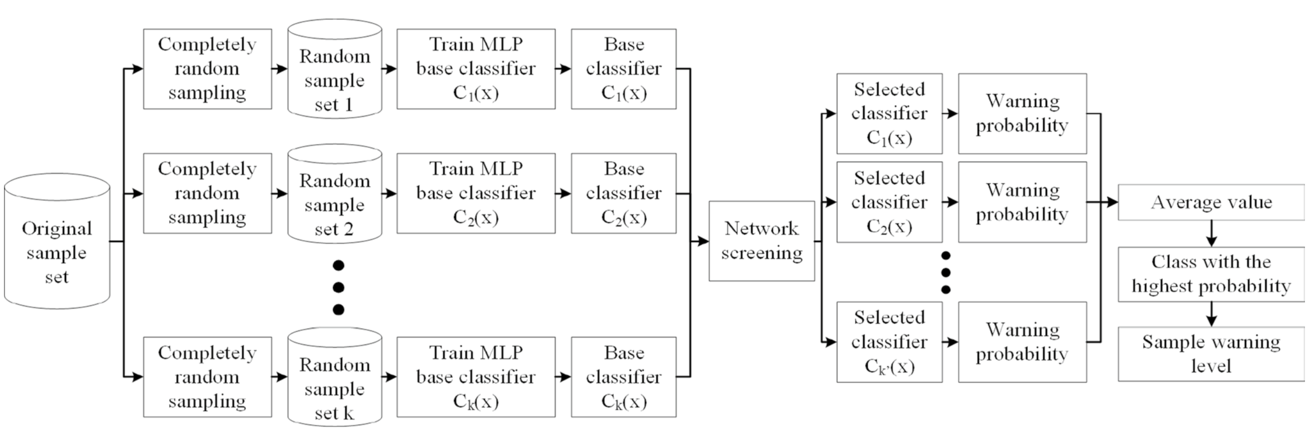

Fig. 5 Schematic diagram of IENN

Algorithm 2: DyS-IENN

Input: Training set **S,** Iteration times **epoch**, Number of neural network clusters **cluster**, hyper-parameter set **P**, categories K

Output: Neural network set **M**

```
r_max = max{r_i};
r_min = min{r_i};
for i = 1 to K
    ρ_i = r_max / r_i − 1;
end for
for ep = 1 to epoch
    S' = Dupicate(S, ρ)
    VotePool = 0;
    if ep = 1 then
        p = 1;
        for cl = 1 to cluster
            S'_cl = PickSample(S', p);
            P'_cl = PickParameter(P);
            M_cl = TrainMLP(S'_cl, P'_cl);
        end for
    else
        for cl = 1 to cluster
            S'_cl = PickSample(S',p);
            M_cl = UpdateMLP(S'_cl, P'_cl);
        end for
        V_cl = Validatesamples(M_cl, S);
        VotePool = VotePool + V_cl;
    end if
    VotePool = VotePool/cluster;
    y = argmax(VotePool);
    δ = y_c - max_{i≠c}{y_i};
    for s = 1 to |S|
        if δ^s <0 then
            p^s = 1;
            ps = exp(-δ^s · α · r_c^s / min_k{ r_k });
        end if
    end for
    if ep > 1 then
        for i = 1 to K
            ρ_i =ρ_i/ln(i·e);
        end for
    end if
end for
return M
```

Experimental results

To investigate the effectiveness of the proposed multiclass imbalanced learning method based on DyS-IENN for early warning of tardiness in RFAP, the actual production data of the enterprise is utilized for testing and analysis. Two groups of experiments are considered. The first group of experiments is designed to verify the superiority of the basic DyS method (Lin et al. 2013) in solving the warning level identification problem in RFAP. The second group of experiments is conducted to validate the suppression effect of the DyS-ENN method on the oscillation of the basic DyS method and the effectiveness of the network screening algorithm on improving the classification accuracy of DyS-ENN method. All the algorithms are coded in Python and implemented in a PC with Intel Core i7 CPU (3.4 GHz × 4), GTX 1080Ti GPU and 16 GB RAM.

Data description

A rocket final assembly workshop in Shanghai is taken as the experimental object, which is mainly responsible for the final assembly of three types of rockets. The assembly process of each type of rocket can be regarded as an assembly process composed of three sub-segment assembly and final rocket assembly. In terms of the input and output indicators proposed in Section III, the relevant data has been collected from rocket assembly information systems of the enterprise such as ERP, MES, etc. After the extraction, conversion, and loading of historical production data, a data warehouse has been established with the theme of early warning of tardiness in RFAP. After further data cleaning, a total of 6362 pieces of examples from 31 rockets in the data warehouse can be used. An example is shown in Table 2, where COR denotes the current overschedule rate, FOR denotes future overschedule rate, CIS denotes the current incomplete set influence rate, FIS denotes future incomplete set influence rate, WD denotes worker deployment influence rate, and WL denotes the warning level. The complete details of the experimental datasets are described in Table 3.

Table 3 Description of experimental datasets

Dataset	Training set	Test set	Validation set	Total
No warning	3504	437	437	4378
Mild warning	1133	141	141	1415
Medium warning	301	37	37	375
Serious warning	156	19	19	194
Total	5094	634	634	6362

Table 2 An example of the data for warning level identification

COR1	COR2	COR3	COR4	FOR	CIS1	CIS2
0.38089	0	0	0.369518	0.10481	0	0
CIS3	FIS	WD1	WD2	WD3	WL	
0	0	0	0	0	1	

Comparison of imbalanced learning methods

Since the imbalanced learning method is usually composed of sampling strategy and classification method, to verify the superiority of the basic DyS algorithm (DyS-MLP) over other imbalanced learning methods, twenty-five imbalanced learning methods are obtained by combining five typical sampling strategies and five commonly used classification algorithms. The sampling strategies include random over-sampling (ROS), random under-sampling (RUS), cluster-based under-sampling (CUS), adaptive synthetic sampling (AdaSyn) and DyS (He and Garcia 2009; Lin et al. 2013), while classification algorithms include SVM, DT, random forest (RF), gradient boosting decision tree (GBDT) and MLP (Kotsiantis et al. 2006). The five-fold cross-validation is executed ten times for all the methods. Besides the classification accuracy of every class, the geometric means of the classification accuracy of every class (G-mean) is employed as another performance metric in our experimental study to evaluate the overall classification performance, which is defined by formula (11). The results of imbalanced learning methods for early warning of tardiness in RFAP are shown in Table 4.

From Table 4, it can be seen that SVM has the lowest classification accuracy compared with other classification algorithms, which may be due to that SVM is not suitable for multi-classification problems. It is also found that the parallel ensemble learning method can greatly improve the accuracy of warning level identification by comparing DT with RF, and the parallel ensemble learning method is superior to the serial ensemble learning method by comparing RF with GBDT. Through the comparison between DT and MLP, it is known that MLP is slightly better than DT as the base classifier when the sampling strategy adopts over-sampling methods (AdaSyn, ROS, and DyS). Besides, the performance of over-sampling methods exceeds that of under-sampling methods in most cases. More importantly, the DyS-MLP method achieves the highest classification accuracy of 83.71%. Therefore, the method combining the over-sampling method and parallel ensemble learning method may achieve higher early warning accuracy, and improving the DyS-MLP by introducing the parallel ensemble learning method with MLP as the base classifier may be a promising way.

The classification accuracy of every class achieved by the five highest accuracy algorithms in Table 4 is shown in Fig. 6. It is observed that the basic DyS method has higher classification accuracy than most algorithms on the category with warning (mild warning, medium warning, and serious warning), but the deviation between the accuracy of different warning levels is significantly larger than RF. Therefore, the classification accuracy and stability of the algorithm need to be further improved.

Comparison of algorithm improvements

To validate the suppression effect of DyS-ENN method on the oscillation of DyS-MLP method and the effectiveness of network screening algorithm on improving the classification accuracy of DyS-ENN method for early warning of tardiness in RFAP, the experimental results of these three methods (DyS-MLP, DyS-ENN, and DyS-IENN) are statistically analyzed to obtain the mean and standard deviation of the G-mean index. The increase of mean indicates that the warning accuracy of the algorithm is improved, while the decrease of standard deviation implies that the oscillation of the model is suppressed.

From the Fig. 7, it may safely conclude that replacing the MLP with ENN in the DyS-MLP method improves the overall accuracy of warning level identification by about 5.2%, and reduces the fluctuation of warning accuracy of no warning class and mild warning class by 82%. Besides, it is also found that introducing the network screening algorithm further improves the overall warning accuracy of DyS-ENN algorithm by about 4.6%, and increases the identification accuracy of serious warning to 99.74%.

The results of the proposed multiclass imbalanced learning method based on DyS-IENN for early warning of tardiness in RFAP are shown in Fig. 8. The lower-left

Table 4 Results of warning level identification in RFAP

Sampling\ Classifier	SVM (%)	DT (%)	RF (%)	GBDT (%)	MLP (%)
AdaSyn	16.10	78.88	82.74	81.16	80.68
ROS	49.58	75.34	83.14	80.88	80.51
RUS	30.01	71.70	77.76	76.94	69.78
CUS	27.69	66.14	72.43	69.64	63.11
DyS	27.47	76.60	77.68	66.99	83.71

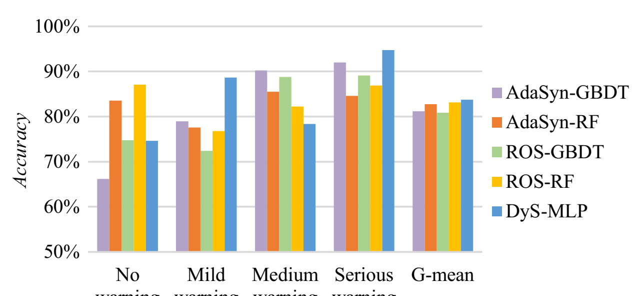

Fig. 6 Results of warning accuracies under each warning level

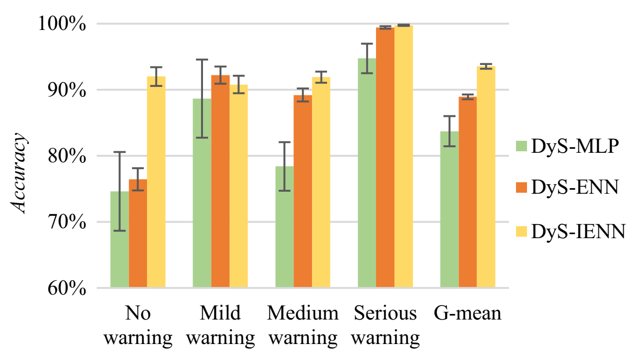

Fig. 7 Comparison of algorithm improvements

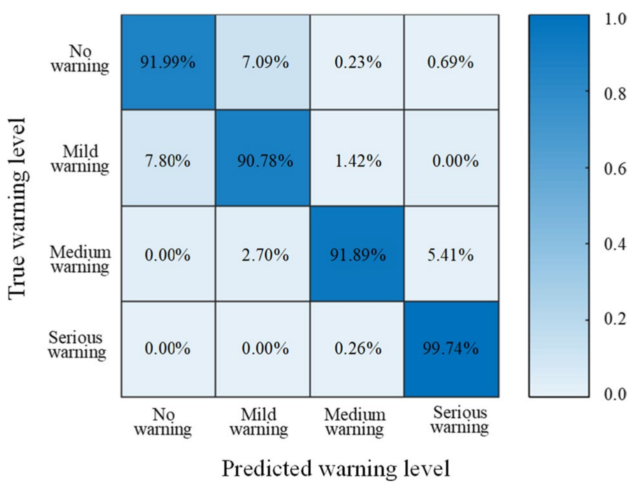

Fig. 8 Confusion matrix of identification results obtained by the DyS-IENN

corner of the confusion matrix represents the proportion in which the samples with high warning are mistakenly identified as the low warning, which is the focus of the warning level identification problem. Among them, 7.80% of the mild warning samples are misjudged as no warning, and 2.70% of the medium warning samples are misjudged as mild warning while 0.26% of the serious warning samples are misjudged as medium warning. By tracing the misidentified warning samples, it is found that the mislabeling of samples and insufficient model generalization performance are the two main reasons for misclassification, which may be solved respectively by improving the labeling accuracy of samples and increasing the sample size. In practical applications, it is an effective remedy for uncertain identification results to appropriately raise the warning level according to the actual production situation, or to directly adjust the value of disturbance factor.

Conclusions and future work

This paper deals with the problem of low warning accuracy caused by imbalanced data distribution from RFAP. After transforming this problem into a multiclass imbalanced learning problem in the field of machine learning, a novel multiclass imbalanced learning method based on DyS-IENN is proposed which integrates the DyS algorithm and IENN. The experimental results clearly show that the basic DyS algorithm is superior to other multiclass imbalanced learning methods and can be employed for early warning of tardiness in RFAP. Moreover, the proposed multiclass imbalanced learning method based on DyS-IENN can not only improve the accuracy of warning level identification, but also suppress the oscillation of the model, which is an effective improvement to the basic DyS algorithm.

The proposed DyS-IENN method can identify the warning level with high accuracy, and assist the adjustment of the RFAP by issuing appropriate warning signals. Taking all points into consideration, future work is still required to further explore the separate impact of each early warning indicator on the warning level, which can more easily guide how to take action in response to early warning. In addition, the proposed method will be investigated in more practical rocket final assembly systems and early warning issues in other similar areas.

Acknowledgements This research is supported by the National Natural Science Foundation of China under Grant No. 51775348 and No. U1637211.

References

Abdi, L., & Hashemi, S. (2015). To combat multi-class imbalanced problems by means of over-sampling techniques. *IEEE Transactions on Knowledge and Data Engineering, 28*(1), 238–251.

Cao, X., & Lam, J. S. L. (2018). Simulation-based catastrophe-induced port loss estimation. *Reliability Engineering & System Safety, 175,* 1–12.

Chawla, N. V., Lazarevic, A., Hall, L. O., & Bowyer, K. W. (2003). SMOTEBoost: Improving prediction of the minority class in boosting. In *European conference on principles of data mining and knowledge discovery* (pp. 107–119).

Chen, S., He, H., & Garcia, E. A. (2010). RAMOBoost: Ranked minority oversampling in boosting. *IEEE Transactions on Neural Networks, 21*(10), 1624–1642.

Gonzalez-Val, C., Pallas, A., Panadeiro, V., & Rodriguez, A. (2020). A convolutional approach to quality monitoring for laser manufacturing. *Journal of Intelligent Manufacturing, 31*(3), 789–795.

Guo, H., Li, Y., Shang, J., Gu, M., Huang, Y., & Gong, B. (2017). Learning from class-imbalanced data: Review of methods and applications. *Expert Systems with Applications, 73,* 220–239.

Guo, H., & Viktor, H. L. (2004). Learning from imbalanced data sets with boosting and data generation: the databoost-im approach. *ACM SIGKDD Explorations Newsletter, 6*(1), 30–39.

Gürbüz, F., Eski, İ., Denizhan, B., & Dağlı, C. (2019). Prediction of damage parameters of a 3PL company via data mining and neural networks. *Journal of Intelligent Manufacturing, 30*(3), 1437–1449.

He, H., & Garcia, E. A. (2009). Learning from imbalanced data. *IEEE Transactions on Knowledge and Data Engineering, 21*(9), 1263–1284.

Jain, A. K., & Lad, B. K. (2019). A novel integrated tool condition monitoring system. *Journal of Intelligent Manufacturing, 30*(3), 1423–1436.

Joseph, A. (2011). *Tsunamis: Detection, monitoring, and early-warning technologies*. Cambrige: Academic Press.

Kotsiantis, S. B., Zaharakis, I. D., & Pintelas, P. E. (2006). Machine learning: a review of classification and combining techniques. *Artificial Intelligence Review, 26*(3), 159–190.

Krawczyk, B., Woźniak, M., & Schaefer, G. (2014). Cost-sensitive decision tree ensembles for effective imbalanced classification. *Applied Soft Computing, 14,* 554–562.

Krogh, A., & Vedelsby, J. (1995). Neural network ensembles, cross validation, and active learning. In *Advances in neural information processing systems* (pp. 231–238).

Lin, M., Tang, K., & Yao, X. (2013). Dynamic sampling approach to training neural networks for multiclass imbalance classification. *IEEE Transactions on Neural Networks and Learning Systems, 24*(4), 647–660.

Qin, W., Zha, D., & Zhang, J. (2018). An effective approach for causal variables analysis in diesel engine production by using mutual information and network deconvolution. *Journal of Intelligent Manufacturing*, *9*, 1–11.

Ramezani, A., & Rothe, H. (2017). Simulation-based early prediction of rocket, artillery, and mortar trajectories and real-time optimization for counter-RAM systems. *Mathematical Problems in Engineering*, *12*, 1–8.

Rojas, R. (2013). *Neural networks: a systematic introduction*. Berlin: Springer.

Rui, Z., Qi, L., Yue, Z., Jin-Cun, C., & Li-Hui, A. N. (2016). Study on lean management of carrier rocket final assembly based on the process. *Industrial Engineering and Management, 21*(2), 108–111.

Sheng, X., Zhao, X., Zhang, P., Zhang, J., Cheng, H., Liu, S., et al. (2019). A stacked auto-encoder based approach for rocket final assembly cycle time prediction. *Computer Integrated Manufacturing Systems, 11,* 2720–2730.

Tan, X., Xing, L., Cai, Z., & Wang, G. (2020). Analysis of production cycle-time distribution with a big-data approach. *Journal of Intelligent Manufacturing*. https://doi.org/10.1007/s10845-020-01544-7.

Wang, J., Xiong, X., Li, Z., Wang, W., & Zhu, J. (2016a). Wind forecast-based probabilistic early warning method of wind swing discharge for OHTLs. *IEEE Transactions on Power Delivery, 31*(5), 2169–2178.

Wang, J., Xiong, X., Zhou, N., Li, Z., & Wang, W. (2016b). Early warning method for transmission line galloping based on SVM and AdaBoost bi-level classifiers. *IET Generation, Transmission and Distribution, 10*(14), 3499–3507.

Wang, J., Yang, Z., Zhang, J., Zhang, Q., & Chien, W. T. K. (2019). AdaBalGAN: An improved generative adversarial network with imbalanced learning for wafer defective pattern recognition. *IEEE Transactions on Semiconductor Manufacturing, 32*(3), 310–319.

Wang, J., & Zhang, J. (2016). Big data analytics for forecasting cycle time in semiconductor wafer fabrication system. *International Journal of Production Research, 54*(23), 7231–7244.

Wang, J., Zhang, J., & Wang, X. (2017). Bilateral LSTM: A two-dimensional long short-term memory model with multiply memory units for short-term cycle time forecasting in re-entrant manufacturing systems. *IEEE Transactions on Industrial Informatics, 14*(2), 748–758.

Wang, J., Zhang, J., & Wang, X. (2018). A data driven cycle time prediction with feature selection in a semiconductor wafer fabrication system. *IEEE Transactions on Semiconductor Manufacturing, 31*(1), 173–182.

Wen, L., Li, X., Gao, L., & Zhang, Y. (2017). A new convolutional neural network-based data-driven fault diagnosis method. *IEEE Transactions on Industrial Electronics, 65*(7), 5990–5998.

Wilhite, D. A. (2012). *Drought assessment, management, and planning: theory and case studies: Theory and case studies*. Berlin: Springer.

Xiong, X., Weng, S., & Wang, J. (2014). An online early-warning method for wind swing discharge of the conductor toward the tangent tower and jumper toward the strain tower. *IEEE Transactions on Power Delivery, 30*(1), 114–121.

Yuan, Z., Wang, Y., & Sun, C. (2017). Construction schedule early warning from the perspective of probability and visualization. *Journal of Intelligent & Fuzzy Systems, 32*(1), 877–888.

Zhang, Y., Li, X., Gao, L., Wang, L., & Wen, L. (2018). Imbalanced data fault diagnosis of rotating machinery using synthetic oversampling and feature learning. *Journal of Manufacturing Systems, 48,* 34–50.

Zhang, C., & Ma, Y. (2012). *Ensemble machine learning: Methods and applications*. Berlin: Springer.

Zhang, Y., Xu, Y., Dong, Z. Y., Xu, Z., & Wong, K. P. (2017). Intelligent early warning of power system dynamic insecurity risk: Toward optimal accuracy-earliness tradeoff. *IEEE Transactions on Industrial Informatics, 13*(5), 2544–2554.

Zheng, Y. J., Chen, S. Y., Xue, Y., & Xue, J. Y. (2017). A Pythagorean-type fuzzy deep denoising autoencoder for industrial accident early warning. *IEEE Transactions on Fuzzy Systems, 25*(6), 1561–1575.

Zheng, G., Zhu, N., Tian, Z., Chen, Y., & Sun, B. (2012). Application of a trapezoidal fuzzy AHP method for work safety evaluation and early warning rating of hot and humid environments. *Safety Science, 50*(2), 228–239.

Zhou, Z. H., & Liu, X. Y. (2005). Training cost-sensitive neural networks with methods addressing the class imbalance problem. *IEEE Transactions on Knowledge and Data Engineering, 18*(1), 63–77.

Zhou, Z. H., Wu, J., & Tang, W. (2002). Ensembling neural networks: Many could be better than all. *Artificial Intelligence, 137*(1–2), 239–263.

Zhou, Q., Yan, P., Liu, H., & Xin, Y. (2019). A hybrid fault diagnosis method for mechanical components based on ontology and signal analysis. *Journal of Intelligent Manufacturing, 30*(4), 1693–1715.