IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received November 12, 2019, accepted November 27, 2019, date of publication December 10, 2019,
date of current version December 23, 2019.

Digital Object Identifier 10.1109/ACCESS.2019.2958570

Optimal Burn-In Strategy for High Reliable Products Using Convolutional Neural Network

YI LYU[1], JUNYAN GAO[2], CI CHEN[3], YIJIE JIANG[4], HUACHUAN LI[2], KAIRUI CHEN[4], AND YUN ZHANG[4]

[1]School of Computer, University of Electronic Science and Technology of China, Zhongshan Institute, Zhongshan 528400, China
[2]School of Computer Science and Engineering, University of Electronic Science and Technology of China, Chengdu 611731, China
[3]School of Electrical and Electronic Engineering, Nanyang Technological University, Singapore 639798
[4]School of Automation, Guangdong University of Technology, Guangzhou 510006, China

Corresponding author: Yi Lyu (yi.lyu@foxmail.com)

This work was supported by the Science and Technology Program of Guangdong: 2017A010101004, and the Major Special Projects of Zhongshan: 2017B1023.

ABSTRACT Burn-in test is widely used to improve the product reliability from the customer's perspective by identifying and screening out defective individuals before they are marketed. For those high reliable products whose failures are caused by gradual degradation, burn-in test not only could pick out weak units, but also increases the degradation of normal units, and hence the test duration is regarded as one key factor in the test policy optimization. In this paper, a new burn-in framework is proposed, which combines a sliding window strategy with one-dimensional convolutional neural network, completes the off-line training for classification model, and then obtains the optimal burn-in time with a group-accuracy strategy. And an online optimization algorithm is constructed to reduce the burn-in time as much as possible without deteriorating the screening effect, thereby to reduce the unnecessary lifetime loss of normal units involved in the test. The effectiveness of the presented framework is validated by the experiment. Compared to conventional strategies based on degradation models, the proposed method has better performance and robustness.

INDEX TERMS Burn-in, deep learning, degradation, sliding window, online optimization.

I. INTRODUCTION

Due to the improvement of manufacturing technology, the reliability of various products tends to be higher and their service life is thus longer, which means that even in the burn-in environment with special settings, the failure time of products is longer than that of general products. Burn-in test aims to eliminate defective products before they are put into service. Meanwhile, the unnecessary life loss of normal products should be reduced as much as possible, which requires that burn-in time should be taken into account when formulating burn-in strategies. Therefore effective burn-in test for highly reliable products is becoming a big challenge.

Burn-in test can be classified into two types according to the failure mechanism. Traditional burn-in, which is based on catastrophic failure, is conducted by subjecting all units to normal or accelerated working load for a suitable duration. Then, the failed subjects are screened to prevent these weak subjects from being shipped to customers. Detailed discussions on the optimization of this type of burn-in policy can be found in Sheu and Chien [1], Cha and Finkelstein [2], Cha [3], and Ye *et al.* [4]. However, for highly reliable products, this approach is likely to be ineffective, because short burn-in time is usually not enough to make defective products fail, while longer time will lead to more performance loss of normal products. In reality, there exists a broad category of products or components whose reliability is directly related to the degradation of some quality characteristics (QC) [5]. With the aid of modern measurement techniques and Internet of Things, degradation data have become increasingly accessible at a relative low cost, and there will be a growing number of products whose failure can be defined by their degradation level [6]. For these degradation-failure products, although weak units have higher degradation rates than normal ones, they can still survive for a relatively long duration in traditional burn-in testing [7]. Therefore, manufactures have to resort to the degradation measurement, that is, degradation-based burn-in, to perform screening.

A lot of studies have contributed to the development and extension of degradation models, which can be referred to the recent review in [8]. The stochastic process is generally used in the research on degradation modeling [9], [10], of which,

The associate editor coordinating the review of this manuscript and approving it for publication was Baoping Cai.

the three most common classes are the inverse gaussian (IG) process [11]–[14], the wiener process [15]–[18], and the gamma process [19]–[21]. Ye and Chen [12] systematically investigated the IG process and showed that the IG process is an important family of degradation analysis because of its superb properties in dealing with covariates and random effects. The wiener process has been widely employed to construct degradation-based burn-in models [22]. For example, Tseng and Peng [23] constructed a burn-in model using the integrated Wiener process and discussed burn-in scheme optimization. Ye *et al.* [24] studied the burn-in planning of products with two competing risks. Ye *et al.* [25] applied the Wiener process to model the measured degradation and discussed the optimal burn-in plan by jointly considering the burn-in test cost and maintenance expense. In the work of Peng [26], a degradation model considering random effects and measurement errors was proposed and the burn-in test was extended into a classification problem with several subpopulations. Zhai *et al.* [27] recently studied the optimization of degradation-based burn-in plan with considering the measurement errors. Moreover, The gamma process has also been used in degradation-based burn-in optimization. For application examples of the gamma process in degradation-based burn-in optimization, see Singpurwalla [28], Lawless and Crowder [29], and Park and Padgett [28], [30].

However, there are some problems with the methods of degradation-based burn-in. Firstly, establishing the physical model to describe the degradation trend depends on fully understanding the failure mechanism of products. This kind of information is hard to obtain, especially when the degradation is complicated [8], [22], [26]. Secondly, burn-in strategies are formulated offline, and fixed strategies are difficult to cope with complex and changeable degradation. Influenced by raw materials, manufacturing process, and different initial degradation state, the degradation rate of products in different batches may vary, and the corresponding optimal time will also change. The strategy developed offline is likely to cause low completion of screening tasks or unnecessary high performance loss of normal products. In addition, Zhai *et al.* [27] discuss the effects of measurement errors on decision parameters and propose that measurement errors always exist and can not be negligible, otherwise the burn-in strategies at the inferior level will be formulated. Finally, the effectiveness of conventional burn-in strategies is challenged by the correctness of some prior information, such as the unit cost of misclassification and the class distribution [31]–[34]. Inaccurate information and unreasonable settings will always lead to burn-in strategies with low application value.

Recently, deep learning (DL) has emerged and achieved much success in computer vision, machine translation, and social network filtering. DL attempts to learn high-level representations from data through multiple layers, and obtain a hierarchical feature representation automatically instead of designing hand-crafted features. Among various DL models, convolutional neural network (CNN) has been applied extensively in various classification tasks [35], [36]. In particular,

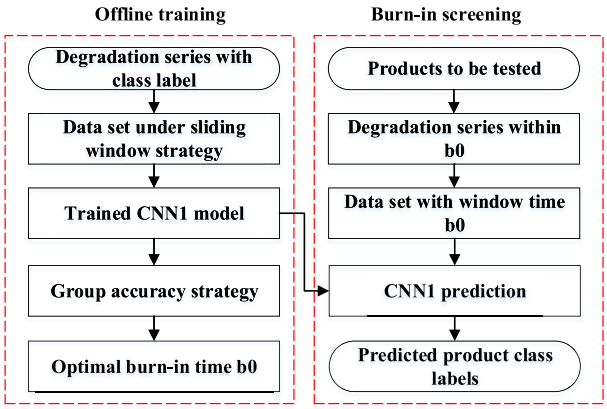


FIGURE 1. Flowchart of the proposed burn-in method.

the application of CNN in field of fault diagnosis can be found in [37]–[42]. Essentially, burn-in test which aims to screen out defective products, could be regarded as a classification problem. Therefore, the relevant methods of CNN can be applied to burn-in test in theory. However, burn-in test needs to process time-series data, and differing from simple classification in fault diagnosis, its decision-making involves burn-in time. The difficulty lies in how to obtain reasonable burn-in time based on given degradation information and apply it to online screening tasks, and this is why DL has a lot of research and application in the field of fault diagnosis, but little development in the burn-in test.

In this paper, we attempt to solve problems stated above with a new method which mainly contains two phases. During the offline training phase, one-dimensional convolutional neural network(CNN1) is combined with the sliding window strategy to build the relationship between degradation trend and measured data, and here, a well-trained CNN1 will be obtained to conduct the screening task. Then, combine the well-trained CNN1 with our group-accuracy strategy, and the optimal burn-in time will be obtained. During Online testing phase, through the online optimization algorithm, the actual burn-in time of products to be screened will be adjusted and the task of product screening then will be completed. The remainder of this paper is organized as follows. Section II elaborates the important theories and methods. Section III introduces the proposed online optimization algorithm. Section IV is the experimental design and method verification. The conclusion is made in Section V.

II. THEORIES AND METHODS

To address the limitations of conventional methods, CNN1 is used to learn data sets under a sliding window strategy, which is the first and key step. On this basis, the optimal burn-in time b0 can then be obtained under a group- accuracy strategy. After completing the training of the model and obtaining b0, new samples to be screened are tested within b0, and finally automatically predicted by CNN1 to complete the screening task. The flowchart of the proposed method is shown in Fig. 1. The main theories and techniques are described as follows.

A. SLIDING WINDOW STRATEGY

The original run-to-failure data set consists of degradation series of QC, each of them has a unique class label representing the true class of the corresponding product, i.e. weak class or normal class. However, due to the principle of burn-in test, it is useless to directly training a model with the original data set. For this reason, a sliding window strategy is applied to process the original data set. The new data set obtained is called a sliding window set, and the sample in which is called a window sample. The sliding window strategy is to give a fixed size window and slide the window along a specified direction, which is effective only when the window is filled with degradation data. When the window slides effectively, the data in the window and the last measurement time will form a window sample. Particularly, the measurement time will be regarded as the actual burn-in time of the window sample, referred to as window time (WT). Denote the window size by ws, and ms denotes the number of steps for each movement. Under the sliding window strategy, $\left\lfloor \frac{s-ws}{ms} \right\rfloor + 1$ samples then can be obtained from the degradation series with length s, each of them contains $ws + 1$ features. It is worth noting that at any WT, the product has only one sliding window sample, then the predicted label and the true label will be consistent with the product.

B. CONVOLUTIONAL NEURAL NETWORK

In this paper, one-dimensional convolution neural network is used to perform the product screening task. In our design, CNN1 model consists of one data input layer, four convolution calculation layers, three fully connection layers and one classification layer. Besides, the pooling layer is not used in the convolution network design, and the same mode of a convolution operation is proposed. The designed structure is shown in Fig.2. Since the same mode is used in convolution layers, and the down-sampling operation is not conducted, the size of the feature maps will remain unchanged after convolution. The output of the convolution layer is calculated as

$$m_i^r = \phi\left(\sum_j m_j^{r-1} \otimes k_{ji}^r + b_i^r\right) \tag{1}$$

where m_i^r, m_j^{r-1} mean the ith feature map of the convolution layer r and the jth feature map of the convolution layer $r-1$, respectively. k_{ji}^r and b_i^r are the corresponding filters and biases. $\phi(\cdot)$ is the activation function used in the convolution layer. Rectified Linear Unit (ReLU) is used in our convolution layers and is defined as

$$\phi(x) = \begin{cases} x, & if\ x > 0 \\ 0, & if\ x \le 0 \end{cases} \tag{2}$$

Each fully connection hidden layer contains a certain number of neurons. The relationship between input and output of neurons in the hidden layer is

$$a_i^l = f\left(\sum_k w_{ik}^l \cdot a_k^{l-1} + b_i^l\right) \tag{3}$$

where w_{ik}^l denotes the weight connecting Neuron i in Layer l and Neuron k in Layer $l-1$, a_i^l denotes the activation value of Neuron i in Layer l, b_i^l is the bias set on Neuron i in Layer l, and $f(\cdot)$ denotes the activation function. The exponential linear unit function (ELU) is used in fully connection hidden layers, and ELU with $\alpha > 0$ is defined as

$$f(x) = \begin{cases} x, & x > 0 \\ \alpha \cdot \left[\exp(x) - 1\right], & x \le 0 \end{cases} \tag{4}$$

In the output layer of a classification model, probability scores of the input on each class are calculated based on the soft-max function, and the predicted label is the one with the highest score. Assuming that there are C classes and the true class label of sample x_i is c_i, the formula for calculating the probability score of x_i on class c_i is

$$p(y = c_i|x_i) = \frac{\exp\left(a_{c_i}^L\right)}{\sum_{j=1}^{C} \exp\left(a_j^L\right)} \tag{5}$$

Note that the summation of denominator is performed on all output neurons.

In terms of the working principle, CNN has two characteristics: local perception and parameter sharing. CNN perceives the local information, and then merges the local information at a higher level to get all the characterization information. The neural units at different layers of CNN adopt the local connection mode, which ensures that the learned convolution kernel has the strongest response to the input spatial local mode. CNN's weight sharing network structure makes it more similar to the biological neural network, reduces the complexity of the network model, and reduces the number of weights. Due to the small number of parameters, the training samples required are relatively small, so to some extent, it is not easy to have a fitting phenomenon.

C. GROUP-ACCURACY STRATEGY

The simulated degradation paths of high-power semiconductor lasers (HPSL), light-emitting diodes (LED) and micro-engine inside the microelectro mechanical systems (MEMS) are shown in Fig. 3, from which, it is readily found that the differentiation between two-class products is positively correlated with the burn-in time and tends to stabilize after reaching the peak value, and this phenomenon is common in various types of products. Assume that the prediction accuracy of products by CNN1 is positively correlated with the differentiation degree of two-class products, then the prediction accuracy of window samples by CNN1 is also positively correlated with WT. Based on this assumption, a group-accuracy strategy is adopted to obtain the optimal burn-in time.

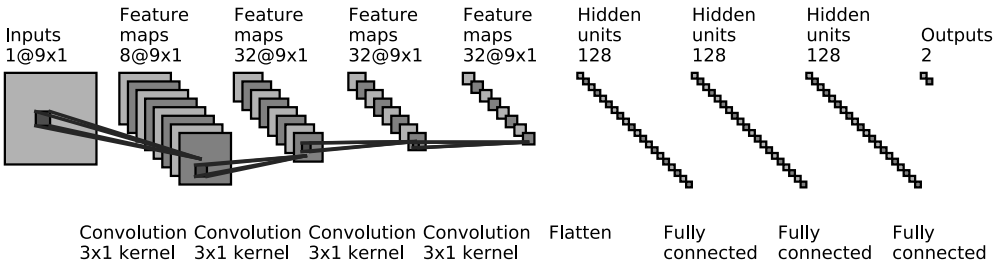
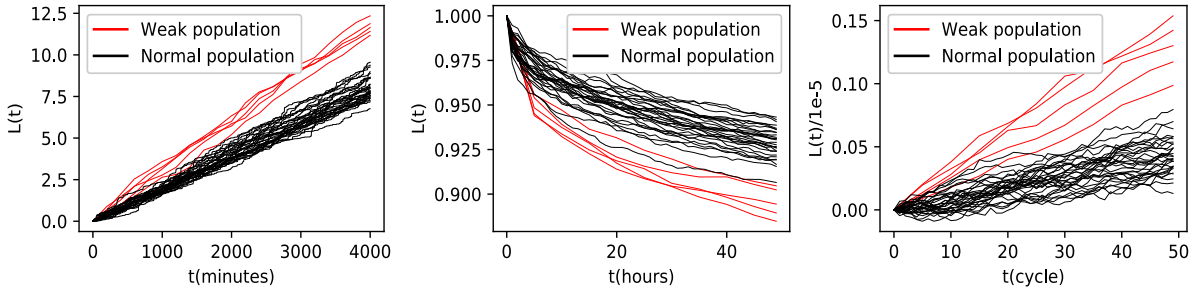


FIGURE 2. Network structure.



(a) HPSL current (b) LED output optical power (c) MEMS errors

FIGURE 3. Quality characteristic degradation of HPSL, LED and MEMS.

When using the group-accuracy strategy, we regroup the sliding window set first, and make sure samples of the same group has the same WT. Then, group tests are conducted by well-trained CNN1 and the accuracy is obtained for each group. Finally, the WT of the group with the highest accuracy and the minimum WT is the optimal burn-in time b0. For a group, denote the number of window samples by n, WT by t, the jth sample by $v_j(t)$, the predicted label and true label respectively by $j_{\text{pred}}(t)$, $j_{\text{true}}(t)$. Then the group accuracy is defined as

$$group\_acc\,(t) = \frac{1}{n}\sum_{j=1}^{n} v_j\,(t) \tag{6}$$

where

$$v_j\,(t) = \begin{cases} 0, & if\ j_{\text{pred}}(t) \neq j_{\text{true}}(t) \\ 1, & if\ j_{\text{pred}}(t) = j_{\text{true}}(t) \end{cases}$$

And then the optimal burn-in time can be formulated as

$$b0 = \min\{t_{imax},\ i = 1, 2, \cdots, m\} \tag{7}$$

where $t_{imax} = \arg\max\{group\_acc(t_i)\}$.

Actually, there is more reasonable approach to calculate b0. First, preset an accuracy threshold $gate\_acc$. When the group accuracy is not less than this value, the corresponding WT becomes a candidate of the optimal burn-in time b0. Then select the minimum window time from all the candidate values as b0. That is

$$b0 = \min\{ti|group\_acc(ti) > gate\_acc,\ i = 1, 2, \cdots, m\} \tag{8}$$

Furthermore, more burn-in time leads to more burn-in cost and more loss of normal unit life involved in the test. At the same time, more burn-in time will also promote the recognition rate of defective products, so as to reduce the risk cost of defective products sold. Therefore, in practical application, the above two situations should be considered when determining the accuracy threshold gate_acc.

III. ONLINE OPTIMIZATION

Online optimization refers to the process of collecting degradation data online and synthesizing the information of current classification results to decide whether to continue a burn-in test. As a valuable research subject, online optimization can be used to reduce the cost of burn-in test, such as time cost, measurement cost, and the life loss of products. Based on the method introduced in Section II, an online optimization algorithm is proposed. The improved online testing process is shown in Fig.4. The main techniques are described briefly as follows.

A. DEFINITION OF EFFECTIVE INFORMATION

Influenced by raw materials, manufacturing process, and different initial degradation state, the degradation distribution characteristics of products in different batches may vary, and the corresponding optimal burn-in time will also change. For example, the degradation rate of weak products in different batches may be quite different. The actual time needed for burn-in test is shorter when QC degrade faster, but increased when QC degrade slower. When a large number of weak products are not identified, the increase of burn-in time will inevitably bring more effective information (EI). With the process of burn-in test, EI will become less and less, and eventually tends to zero. Denote the class prediction results of products within burn-in time t by $RS(t)$, and the prediction

of the *i*th product is denoted by $RS_i(t)$, and $RS_i(t) \in \{0, 1\}$. When $RS_i(t) \neq RS_i(t-1)$, an EI will be produced. Then, $EI(t)$ with n products to be tested is defined as

$$EI(t) = \frac{1}{n}\sum_{i=1}^{n}\tau_i(t),$$
$$s.t.\ \tau_i(t) = \begin{cases} 0, & RS_i(t) = RS_i(t-1) \\ 1, & RS_i(t) \neq RS_i(t-1) \end{cases} \quad (9)$$

In the later stage of burn-in test, EI decreases. Although the increased time t will help to further screen out weak products, it will also lead to more cost and life loss of normal products, resulting in zero or negative actual gains. Hence, a gate of effective information (GEI) can be set artificially based on prior knowledge. When the current EI is below the GEI, then stop the test; otherwise, continue to conduct the test.

B. ONLINE OPTIMIZATION ALGORITHM

Algorithm 1 Online Optimization Algorithm

Input:

$CNN1$, $b0$, $window\_size$, $move\_step$, $numl$, $numr$, and GEI.

Output:

t and $RS(t)$.

1: Initialization:
Denote the burn-in time by t, $t \in [lb, rb]$, and the actual burn-in time is denoted by t_{actual} which will be updated automatically throughout the burn-in test process. Then the parameters are initialized as follows
$lb = max(window\_size, b0 - numl \cdot move\_step)$.
$rb = b0 + numr \cdot move\_step$.
$t = lb$, $t_{actual} = 0$,
$RS(t - move\_step) = \{0, 0, \cdots, 0\}$.

2: Conducting the burn-in test of products:
When $t_{actual} = t$, interrupt the test.

3: Obtaining $RS(t)$:
Input the window samples with $WT = t$ into $CNN1$ and get predicted labels.

4: Calculating $EI(t)$:
Substitute $RS(t)$, $RS(t - move\_step)$ into formula 8.

5: Decision-making:
If $t + move\_step > rb$ or $EI(t) < GEI$, stop the algorithm and output t, $RS(t)$. Otherwise, update t, $t = t + move\_step$, and go to step 2.

IV. EXPERIMENTS

A. DATA DESCRIPTION

High-power semiconductor lasers are core components for various applications such as medical surgery, photodynamic therapy, and space applications. The main quality characteristic of a high power laser diode is its output optimal power, which will degrade over working time. In some type of laser device, optimal power feedback function is designed, then the operating current will increase over time to remain the optimal power stable. When the operating current reaches a pre-fixed threshold level, the device is considered to have failed. As argued in [22], the gamma process (GP) exhibits quite well performance in capturing the degradation characteristic of laser devices. Denote GP by $\{L(t),\ t \geq 0\}$, then the degradation increments follow a gamma distribution,

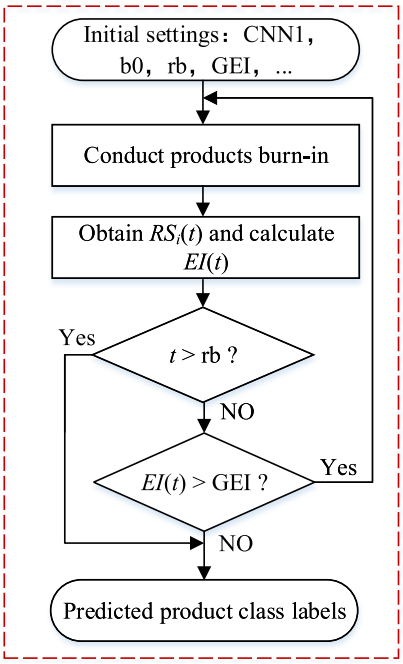


FIGURE 4. The improved online screening process.

$$L(t + \Delta t) - L(t) \sim Ga(\Delta g(t), \nu) \quad (10)$$

where $\Delta g(t) = g(t + \Delta t) - g(t)$, $Ga(\Delta g(t), \nu)$ means the gamma distribution with shape parameter $\Delta g(t)$ and scale parameter ν, $\nu > 0$. $g(\cdot)$ is a monotone increasing function.

There are also many other types of products whose QC degradation process can be described by a wiener stochastic process (WP), such as the micro-engine inside the micro-electro mechanical systems (MEMS), the light-emitting diode (LED), the contact image scanner of a copy/fax machine, plasma display panels, vacuum fluorescent displays, liquid crystal displays and digital light processing projectors and numerous other dependable systems. Among them, the degradation process of micro-engine in MEMS is a linear WP with linear drift [23], and its increment follows normal distribution. Denote the linear WP by $\{Y(t),\ t \geq 0\}$,

$$Y(t) = \beta t + \sigma B(t) \quad (11)$$
$$Y(t + \Delta t) - Y(t) \sim N(\beta \Delta t, \sigma^2 \Delta t) \quad (12)$$

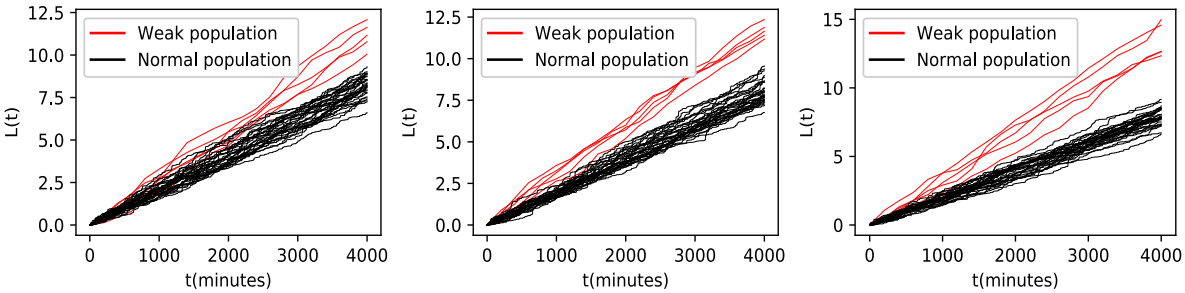
where β and σ denote the drift parameter and the variance coefficient respectively, $B(t)$ is the standard Brownian motion. Referring to the work [27], the degradation process of LED can be described as a non-linear WP which is denoted by $\{X(t), t \geq 0\}$,

$$X(t) = g(t) + \sigma B(\tau(t)) \quad (13)$$

where $g(t) = \exp\left(-\eta t^{\delta}\right)$, $\tau(t) = t^{\gamma}$.

TABLE 1. Overview of data simulation information.

Data set	Degradation model	Parameter seting	Failure threshold
HPSL	Gamma	$g_w = 0.05, v_w = 0.06$ $g_n = 0.04, v_n = 0.05$	Time:4000 min
LED	Non-linear Wiener	$\eta_w = 0.02625, \eta_n = 0.0175$ $\delta_w = \delta_n = 0.037, \gamma_w = \gamma_n = 0.6$ $\sigma_w = \sigma_n = 0.005$	Time:128 h
MEMS	Linear Wiener	$\beta_w = 2.6875 * 1e-8$ $\beta_n = 8.4832 * 1e-9$ $\sigma_w = \sigma_n = 2.2808 * 1e-8$	Degradation:1.3093*1e-5
DIG	IG_NRE	$\mu_w = 3, \delta_w = 0.15, \eta_w = 2$ $\mu_n = 5.5, \delta_n = 0.1, \eta_n = 1$ $b_w = b_n = 1.2$	Degradation:7



(a) $\mu = 2.5, \delta = 0.1$ (b) $\mu = 3, \delta = 0.15$ (c) $\mu = 3.5, \delta = 0.1$

FIGURE 5. Simulated DIG data with different parament settings.

In addition, the degradation of some products can be described by the inverse gaussian (IG) process, such as aluminum alloy specimens [43] and lithium-ion batteries [44]. The IG process has independent increments following the inverse gaussian distribution. Denoting the IG process by $\{Y_s(t), t \geq 0\}$, an IG process model with normal random effects (IG_NRE) can be defined as Eq.14, where β is the degradation rate, η is the shape parameter, $\Lambda(t)$ is the monotonously increasing drift function and $\Lambda(t) = t^b$. $N(\mu, \sigma^2)$ denotes the normal distribution with the mean μ and the standard deviation σ.

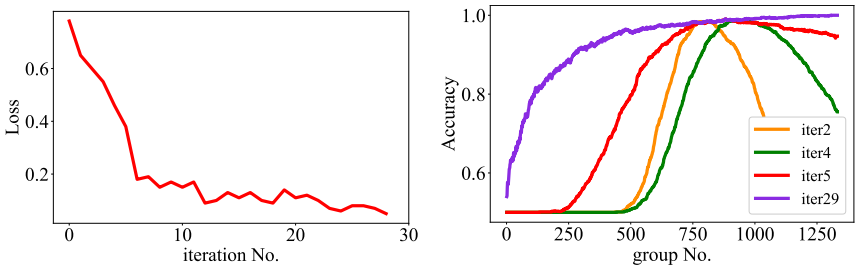
$$Y_s(t) \sim IG(\beta\Lambda(t), \eta\Lambda^2(t)), \beta^{-1} \sim N(\mu, \sigma^2) \qquad (14)$$

In our experiments, Monte Carlo method is used to simulate three kinds of data sets, namely HPSL data, LED data and MEMS data. Another kind of degradation data is generated based on IG_NRE, and here we denote this kind of data by DIG. Each data set contains one training set and 10 test subsets. Each training set has 1000 samples and each test subset has 100 samples, but the proportion of weak products to normal products is random. Denote the weak product by w and the normal by n, then an overview of data simulation information is shown in Table 1. Illustrations of degradation paths for the four data sets are shown in Fig. 3 and Fig. 5.

B. ROC EVALUATION STRATEGY

For each batch of products, the proportion of two populations is variable and unknown; the general accuracy evaluation strategy cannot reflect the performance of a classifier correctly. For example, when the proportion of the normal and weak subpopulation is 95 : 5, even if all weak products are mistaken as the normal ones, the accuracy of a classifier could still reach 95%, presenting the illusion of high performance. Hence, the performance on two subpopulations must be simultaneously taken into account in the design of an practicable classifier.

The receiver operating characteristic curve (ROC) is often used in binary classification tasks. It is a powerful tool to study the generalization performance of classifiers from the perspective of threshold selection. Inspired by the idea of ROC, a kind of ROC index is constructed to evaluate the model performance. Actually, when classifying whether a product is normal or weak, a burn-in procedure can have four possibilities which consist of two types of errors when comparing the prediction with the real situation of a product. If both the real situation and the prediction from the model are normal, they are called true normal (TN). If the product is normal, whereas the prediction is weak, it is called a false weak (FW). False normal (FN) and true weak (TW) are defined similarly. Denote the ROC index by $RocI$; Tr represents



(a) Loss (b) Group accuracy

FIGURE 6. Training and testing result on DIG data1.

the proportion of the number of normal products correctly predicted in the total number of the normal. Fr denotes the proportion of the number of normal products falsely predicted in the total number of the weak. Then, the ROC index is defined as

$$\begin{aligned} RocI &= 1-Tr+Fr \\ Tr &= \frac{TN}{TN+FW} \\ Fr &= \frac{FN}{TW+FN} \end{aligned}$$

From the definition above, it can be seen that the ROC index is non-negative, and the smaller the value is, the better the performance of a model on the two population of products will be. On the contrary, problems may exist in the model, such as the model cannot correctly classify two products, or the model has serious first or second prediction errors.

C. VALIDATION OF THE PROPOSED METHOD

In this section, we will use the DIG data to complete the validation of the proposed method. In order to evaluate the online optimization algorithm, parameters of the degradation model of weak products will be modified slightly during the Monte Carlo simulation, and three DIG data sets are finally produced, and visually shown in Fig. 5. When using the sliding window strategy to construct window samples, we set window size $= 8$, move step $= 3$, and Z standardization is applied in data preprocessing. The network design refers to the model structure in Section II and Fig. 2. The cross-entropy loss function is used in model training. Because there are many parameters to be learned, the gradient descent method is applied. The cross-entropy loss function is given as

$$J(\Theta) = -\frac{1}{B}\sum_{i=1}^{B}\log\left(p\left(y=c_i|x_i\right)\right) \tag{15}$$

where Θ, B, η, c_i denote the set of all parameters to be learned, the number of input samples in batch training, the learning rate and the true label of Sample i, respectively. In our experiment, the initialization of weight parameters follows a truncated normal distribution with mean value 0, variance 0.01; bias initial values are all set to 0.1; the dropout rate in fully connected layers is 0.5; the initial value of learning rate is 0.001, and the exponential decay of the learning rate is performed every 100 steps.

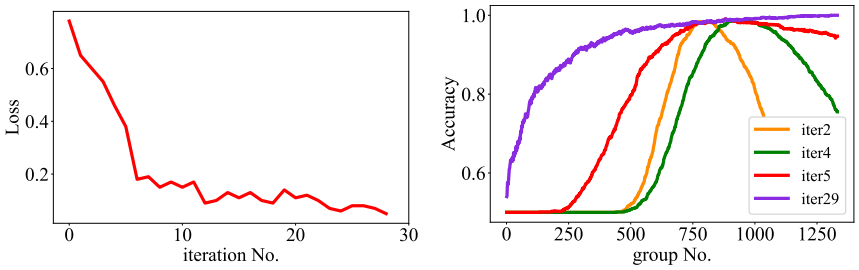
TABLE 2. Results of method validation.

DIG data	cm0			cm1		
	accuracy	RocI	actual_t	accuracy	RocI	actual_t
1	0.971	0.044	36	0.971	0.044	36
2	0.985	0.017	36	0.974	0.029	30
3	0.966	0.159	36	0.971	0.138	41

Fig. 6(a) shows the training loss on DIG data1, from which we can see that CNN1 converges rapidly. Fig. 6(b) is an illustration of testing group-accuracy over some iterations, showing that the group accuracy increases with the increase of burn-in time and tends to stabilize after reaching the peak value 1, and thus proving the corresponding hypothesis and analysis in Section II. With the setting $gate\_acc = 0.97$, the accuracy for products in the training set of DIG data1 is 0.978, the ROC index is 0.044, and the optimal burn-in time is 36.

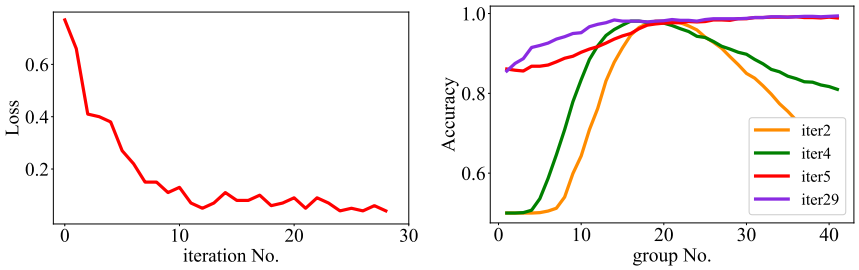
The online optimization method proposed in the paper is validated through DIG data 2, DIG data 3 and the testing set of DIG Data 1. All three sets are composed of 10 test subsets, each of which contains 100 test samples, and the proportion of weak products to normal ones is random. In the experiment, the model using the online optimization strategy is denoted by $cm1$, otherwise denoted by $cm0$. The final experimental results are shown in Table 2, where accuracy, actual-t respectively means the accuracy of product screening and the actual burn-in time. Note that the final results are averaged on 10 test subsets. When the online optimization is not performed, the actual burn-in time is b0.

From Table 2, it can be found that (i) When online optimization is not performed, the accuracy of the three testing sets is 0.971, 0.985 and 0.966 respectively. (ii) When online optimization is conducted, the accuracy of DIG data 2 decreases to 0.974 while the actual burn-in time is cut down to 30 from 36. At the same time, the accuracy of DIG data 3 increases to 0.971 with the actual burn-in time 41. Therefore, it can be seen that through online optimization, burn-in time can be dynamically adjusted according to the actual degradation information of products, so as to better complete the screening task and reduce the unnecessary burn-in time.



(a) Loss (b) Group accuracy

FIGURE 7. Training and testing result on HPSL data.



(a) Loss (b) Group accuracy

FIGURE 8. Training and testing result on LED data.

D. COMPARISONS WITH CONVENTIONAL METHODS

In this section, we will carry out comprehensive comparisons between the method designed in this paper and the other three representative conventional methods which are respectively from the literature [22], [23], [27].

Conventional methods to obtain the burn-in strategies are mainly based on modeling degradation data and minimizing cost functions. The method in [22] uses a mixed gamma process to model HPSL data; the method in [23] uses an integrated wiener process to simulate the cumulative degradation path of LED data, and takes full account of the information contained in the whole degradation sequence when producing the burn-in strategy; the method in [27] discusses the influence of measurement errors on decision-making parameters, and particularly takes Gaussian measurement errors into account when designing the degradation model of MEMS data. The above three methods are respectively denoted by *lgm*, *lwm*, and *mwm*. In order to fully evaluate the performance of methods, Gaussian measurement errors are introduced to data. For example, the gamma process model $\{L(t),\ t \geq 0\}$ with Gaussian measurement errors is defined as follows.

$$Z(t) = L(t) + \varepsilon(t), \varepsilon(t) \sim N(\mu, \sigma^2) \tag{16}$$

In the experiment, each testing set is further divided into two groups, there are no measurement errors and there are Gaussian measurement errors. Specifically, the settings of measurement error parameters are $\mu = 0$ and $\sigma^2 = 0.01$ in HPSL data, $\mu = 0$ and $\sigma^2 = 0.001$ in LED data, and $\mu = 0$ and $\sigma^2 = 3 * 1e - 7$ in MEMS data respectively. CNN1 was respectively trained with the training set of HPSL data, LED data and MEMS data. Note that the three training sets all contain no noise. Then, each well-trained CNN1 was tested with the corresponding test sets, including the one with noise and the other without noise. Fig.7-9 respectively show the training loss and the testing group-accuracy over some iterations on the corresponding data set.

TABLE 3. Results of comparisons on HPSL data.

Test set	cm0		cm1		lgm	
	RocI	actual_t	RocI	actual_t	RocI	actual_t
HPSL	0.094	1640	0.266	1124	0.071	2000
HPSL*	0.094	1640	0.147	1529	0.078	2000

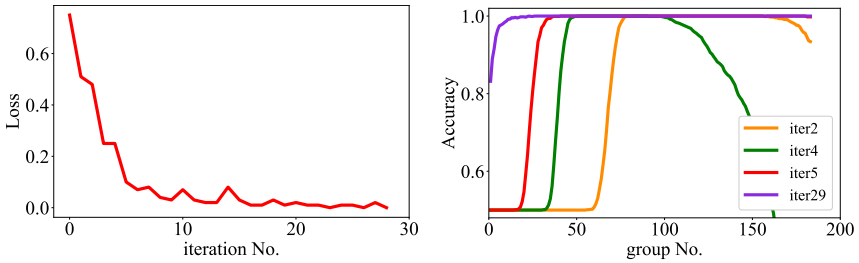
TABLE 4. Results of comparisons on LED data.

Test set	cm0		cm1		lwm	
	RocI	actual_t	RocI	actual_t	RocI	actual_t
LED	0.052	38	0.075	23	0.624	8
LED*	0.085	38	0.061	26	0.632	8

TABLE 5. Results of comparisons on MEMS data.

Test set	cm0		cm1		mwm	
	RocI	actual_t	RocI	actual_t	RocI	actual_t
MEMS	0.045	23	0.045	23	0	68
MEMS*	0.403	23	0.018	80	0.130	68

The final experimental results are shown in Tables 3-5, where data sets with measurement errors are marked with *. It should be noted that strategies produced by *lgm*, *lwm*, and *mwm* are developed offline, which means that the strategy based on the training set will not change even though



(a) Loss (b) Group accuracy

FIGURE 9. Training and testing result on MEMS data.

there are measurement errors. Similarly, the optimal time b0 produced by CNN1 using the training set will remain unchanged.

By analyzing the experimental results in Tables 3-5, conclusions are drawn as follows. (i) The performance of conventional methods is affected by measurement errors. In the experiment, when measurement errors exist in data, performance of *lgm*, *lwm*, *mwm* are all affected and the ROC index increases from 0.071, 0.624, 0 to 0.078, 0.632, 0.130, respectively. There are two main reasons behind this. On the one hand, only one degradation measurement at the end time point is used in the burn-in test. On the other hand, measurement errors are not considered in the degradation modeling. (ii) There is a big risk when offline burn-in strategies are used for online testing. In the experiment, the ROC index of *lwm* is relatively high (0.624 and 0.632), thus the corresponding burn-in strategies can be considered to have low application value. (iii) Measurement errors will also affect the performance of *cm*0, but through online optimization algorithm, the influence can be reduced or even eliminated completely. As shown in Table 5, when Gaussian errors exist, the ROC index of *cm*0 increases from 0.045 to 0.403 with the same burn-in time 23. When the online optimization is performed, the ROC index decreases from 0.403 to 0.018 and the actual burn-in time increases to 80.

V. CONCLUSION

In this paper, we employ the deep learning method to optimize the degradation-based burn-in test for highly reliable products. An burn-in optimization framework is proposed, which consists of the sliding window data-processing method, one-dimensional convolution neural network, the group-accuracy strategy and an online optimization algorithm, and its effectiveness has been validated by experiments. Moreover, compared to conventional methods, our method has the following advantages.

First, it is completely driven by degradation data and can be widely applied to the burn-in test of various products, without human exploration of degradation failure mechanism and the priori assumption of model parameter distributions. The main reason behind this is that during the training phase, CNN1 can learn a hierarchical feature representation automatically and perform the classification task. When trained properly for a particular dataset, CNN1 can optimize both feature extraction and classification tasks according to the problem at hand.

Second, it has stronger robustness in the classification task owing to the proposed online optimization algorithms. In the validation experiment, different degradation rates are adopted in three testing sets. In the comparative experiment, Gaussian measurement errors are introduced into one group of testing sets. According to the experimental results, the screening task is well done by CNN1 when the online optimization is performed.

Third, it can be easily extended to deal with multiple quality characteristics by replacing CNN1 with a two-dimensional convolutional neural network. However, traditional methods often focus on the analysis of single degradation signal, which is difficult to deal with the situation of multiple quality characteristics, otherwise data fusion technology has to be applied.

Nevertheless, there are some aspects worth further investigation. For example, when developing the burn-in strategy for the product with multiple quality characteristics, a continuous visualization progression of degradation should be provided, which, compared with the original multi-sensor signals, is easy to understand and thus highly desired in industrial practice. Besides, the complexity of screening tasks determines the complexity of CNN, which requires the size of training set. In the case of limited original data, it should be further explored whether the amount of data can be enlarged through some existing data processing technologies to ensure that CNN can learn effectively.

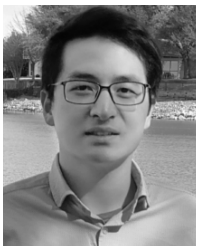
REFERENCES

[1] S.-H. Sheu and Y.-H. Chien, "Minimizing cost-functions related to both burn-in and field-operation under a generalized model," *IEEE Trans. Rel.*, vol. 53, no. 3, pp. 435–439, Sep. 2004.
[2] J. H. Cha and M. Finkelstein, "Stochastically ordered subpopulations and optimal burn-in procedure," *IEEE Trans. Rel.*, vol. 59, no. 4, pp. 635–643, Dec. 2010.
[3] J. H. Cha, "A survey of burn-in and maintenance models for repairable systems," in *Replacement Models With Minimal Repair*. London, U.K.: Springer, 2011, pp. 179–203.
[4] Z. S. Ye, L. C. Tang, and M. Xie, "A burn-in scheme based on percentiles of the residual life," *J. Qual. Technol.*, vol. 43, no. 4, pp. 334–345, 2017.
[5] W. Peng, Y.-F. Li, J. Mi, L. Yu, and H.-Z. Huang, "Reliability of complex systems under dynamic conditions: A Bayesian multivariate degradation perspective," *Rel. Eng. Syst. Saf.*, vol. 153, pp. 75–87, Sep. 2016.
[6] W. Peng, Y.-F. Li, Y.-J. Yang, J. Mi, and H.-Z. Huang, "Bayesian degradation analysis with inverse Gaussian process models under time-varying degradation rates," *IEEE Trans. Rel.*, vol. 66, no. 1, pp. 84–96, Mar. 2017.

[7] Y. Lyu, Y. Zhang, K. Chen, C. Chen, and X. Zeng, "Optimal multi-objective burn-in policy based on time-transformed Wiener degradation process," *IEEE Access*, vol. 7, pp. 73529–73539, 2019.
[8] Z.-S. Ye and M. Xie, "Stochastic modelling and analysis of degradation for highly reliable products," *Appl. Stochastic Models Bus. Ind.*, vol. 31, no. 1, pp. 16–32, 2015.
[9] M. Guida and G. Pulcini, "The inverse gamma process: A family of continuous stochastic models for describing state-dependent deterioration phenomena," *Rel. Eng. Syst. Saf.*, vol. 120, pp. 72–79, Dec. 2013.
[10] L. Wang, R. Pan, X. Li, and T. Jiang, "A Bayesian reliability evaluation method with integrated accelerated degradation testing and field information," *Rel. Eng. Syst. Saf.*, vol. 112, pp. 38–47, Apr. 2013.
[11] X. Wang and D. Xu, "An inverse Gaussian process model for degradation data," *Technometrics*, vol. 52, no. 2, pp. 188–197, May 2010.
[12] Z.-S. Ye and N. Chen, "The inverse Gaussian process as a degradation model," *Technometrics*, vol. 56, no. 3, pp. 302–311, Jul. 2014.
[13] C.-Y. Peng, "Inverse Gaussian processes with random effects and explanatory variables for degradation data," *Technometrics*, vol. 57, no. 1, pp. 100–111, 2014.
[14] X. Chen, X. Sun, X. Ding, and J. Tang, "The inverse Gaussian process with a skew-normal distribution as a degradation model," *Commun. Statist.-Simul. Comput.*, pp. 1–17, Jan. 2019.
[15] G. A. Whitmore and F. Schenkelberg, "Modelling accelerated degradation data using Wiener diffusion with a time scale transformation," *Lifetime Data Anal.*, vol. 3, no. 1, pp. 27–45, 1997.
[16] X. Wang, "Semiparametric inference on a class of Wiener processes," *J. Time Ser. Anal.*, vol. 30, no. 2, pp. 179–207, 2009.
[17] Z. Pan, N. Balakrishnan, Q. Sun, and J. Zhou, "Bivariate degradation analysis of products based on Wiener processes and copulas," *J. Stat. Comput. Simul.*, vol. 83, no. 7, pp. 1316–1329, 2013.
[18] Z.-S. Ye, N. Chen, and Y. Shen, "A new class of Wiener process models for degradation analysis," *Rel. Eng. Syst. Saf.*, vol. 139, pp. 58–67, Jul. 2015.
[19] Z. Pan and N. Balakrishnan, "Reliability modeling of degradation of products with multiple performance characteristics based on gamma processes," *Rel. Eng. Syst. Saf.*, vol. 96, no. 8, pp. 949–957, 2011.
[20] M. H. Ling, K. L. Tsui, and N. Balakrishnan, "Accelerated degradation analysis for the quality of a system based on the gamma process," *IEEE Trans. Rel.*, vol. 64, no. 1, pp. 463–472, Mar. 2015.
[21] Z.-S. Ye, M. Xie, L.-C. Tang, and N. Chen, "Semiparametric estimation of gamma processes for deteriorating products," *Technometrics*, vol. 56, no. 4, pp. 504–513, 2014.
[22] C.-C. Tsai, S.-T. Tseng, and N. Balakrishnan, "Optimal burn-in policy for highly reliable products using gamma degradation process," *IEEE Trans. Rel.*, vol. 60, no. 1, pp. 234–245, Mar. 2011.
[23] S.-T. Tseng and C.-Y. Peng, "Optimal burn-in policy by using an integrated Wiener process," *IIE Trans.*, vol. 36, no. 12, pp. 1161–1170, 2004.
[24] Z.-S. Ye, M. Xie, L.-C. Tang, and Y. Shen, "Degradation-based burn-in planning under competing risks," *Technometrics*, vol. 54, no. 2, pp. 159–168, 2012.
[25] Z.-S. Ye, Y. Shen, and M. Xie, "Degradation-based burn-in with preventive maintenance," *Eur. J. Oper. Res.*, vol. 221, no. 2, pp. 360–367, 2012.
[26] C.-Y. Peng, "Optimal classification policy and comparisons for highly reliable products," *Sankhya B*, vol. 77, no. 2, pp. 321–358, 2015.
[27] Q. Zhai, Z.-S. Ye, J. Yang, and Y. Zhao, "Measurement errors in degradation-based burn-in," *Rel. Eng. Syst. Saf.*, vol. 150, pp. 126–135, Jun. 2016.
[28] C. Park and W. J. Padgett, "Accelerated degradation models for failure based on geometric Brownian motion and gamma processes," *Lifetime Data Anal.*, vol. 11, no. 4, pp. 511–527, 2005.
[29] J. Lawless and M. Crowder, "Covariates and random effects in a Gamma process model with application to degradation and failure," *Lifetime Data Anal.*, vol. 10, no. 3, pp. 213–227, 2004.
[30] C. Park and W. J. Padgett, "Stochastic degradation models with several accelerating variables," *IEEE Trans. Rel.*, vol. 55, no. 2, pp. 379–390, Jun. 2006.
[31] D. S. Chang, "Optimal burn-in decision for products with an unimodal failure rate function," *Eur. J. Oper. Res.*, vol. 126, no. 3, pp. 534–540, 2000.
[32] K. O. Kim and W. Kuo, "A general model of heterogeneous system lifetimes and conditions for system burn-in," *Naval Res. Logistics*, vol. 50, no. 4, pp. 364–380, 2003.
[33] K. Kim and W. Kuo, "Percentile life and reliability as performance measures in optimal system design," *IIE Trans.*, vol. 35, pp. 1133–1142, Dec. 2003.
[34] J. H. Cha and J. Mi, "Optimal burn-in procedure for periodically inspected systems," *Naval Res. Logistics*, vol. 54, no. 7, pp. 720–731, 2007.
[35] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," *Nature*, vol. 521, no. 7553, p. 436, 2015.
[36] Y. Bengio, A. Courville, and P. Vincent, "Representation learning: A review and new perspectives," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 35, no. 8, pp. 1798–1828, Aug. 2013.
[37] O. Janssens, O. Janssens, V. Slavkovikj, B. Vervisch, K. Stockman, M. Loccufier, S. Verstockt, R. Van de Walle, and S. Van Hoecke, "Convolutional neural network based fault detection for rotating machinery," *J. Sound Vib.*, vol. 377, pp. 331–345, Sep. 2016.
[38] X. Guo, L. Chen, and C. Shen, "Hierarchical adaptive deep convolution neural network and its application to bearing fault diagnosis," *Measurement*, vol. 93, pp. 490–502, Nov. 2016.
[39] R. Liu, G. Meng, B. Yang, C. Sun, and X. Chen, "Dislocated time series convolutional neural architecture: An intelligent fault diagnosis approach for electric machine," *IEEE Trans. Ind. Inf.*, vol. 13, no. 3, pp. 1310–1320, Jun. 2017.
[40] L. Jing, T. Wang, M. Zhao, and P. Wang, "An adaptive multi-sensor data fusion method based on deep convolutional neural networks for fault diagnosis of planetary gearbox," *Sensors*, vol. 17, no. 2, p. 414, 2017.
[41] T. Ince, S. Kiranyaz, L. Eren, M. Askar, and M. Gabbouj, "Real-time motor fault detection by 1-D convolutional neural networks," *IEEE Trans. Ind. Electron.*, vol. 63, no. 11, pp. 7067–7075, Nov. 2016.
[42] O. Abdeljaber, O. Avci, S. Kiranyaz, M. Gabbouj, and D. J. Inman, "Real-time vibration-based structural damage detection using one-dimensional convolutional neural networks," *J. Sound Vib.*, vol. 388, pp. 154–170, Feb. 2017.
[43] S. Hao, J. Yang, and C. Berenguer, "Degradation analysis based on an extended inverse Gaussian process model with skew-normal random effects and measurement errors," *Rel. Eng. Syst. Saf.*, vol. 189, pp. 261–270, Sep. 2019.
[44] W. Peng, S.-P. Zhu, and L. Shen, "The transformed inverse Gaussian process as an age- and state-dependent degradation model," *Appl. Math. Model.*, vol. 75, pp. 837–852, Nov. 2019.

YI LYU received the B.S. degree in automatic engineering from the Taiyuan University of Technology, Taiyuan, China, in 2003, and the M.S. degree from the Guangdong University of Technology, Guangzhou, China, in 2006, where he is currently pursuing the Ph.D. degree with the Department of Automation. He has been an Associate Professor with the Department of Computer, University of Electronic Science and Technology of China, Zhongshan Institute, since 2014. His research interests include reliability modeling using stochastic process, burn-in, accelerated degradation test, maintenance optimization, and remaining useful life estimation for complex systems.

JUNYAN GAO received the B.S. degree in computer science and technology from Leshan Normal University, Leshan, China, in 2017. She is currently pursuing the M.S. degree in computer technology with the University of Electronic Science and Technology, Chengdu, China. Her research interests include machine learning in reliability engineering, burn-in testing, machinery prognostics, and health management.



CI CHEN received the B.E. and Ph.D. degrees from the School of Automation, Guangdong University of Technology, Guangzhou, China, in 2011 and 2016, respectively. He was a Research Assistant with the School of Electrical and Electronic Engineering, Nanyang Technological University, Singapore, from 2015 to 2016. From 2016 to 2018, he was a Research Associate with The University of Texas at Arlington and The University of Tennessee at Knoxville. He is currently with the School of Automation, Guangdong University of Technology, and the School of Electrical and Electronic Engineering, Nanyang Technological University. His research interests include reinforcement learning, nonlinear system control, and resilient control and computational intelligence. He is also an Editor of the *International Journal of Robust and Nonlinear Control* and an Associate Editor of the *Advanced Control for Applications: Engineering and Industrial Systems*.

YIJIE JIANG received the B.S. degree in electrical engineering and automation from the Hunan University of Science and Technology, Xiangtan, China, in 2016. He is currently pursuing the M.A.Eng. degree with the Guangdong University of Technology, Guangzhou, China. His research interests include machine learning in reliability engineering, burn-in testing, machinery prognostics, and health management.

HUACHUAN LI received the B.S. degree in electronic science and technology from the Chengdu University of Information Technology, Chengdu, China, in 2018. He is currently pursuing the M.S. degree in computer technology from the University of Electronic Science and Technology of China, Chengdu. His research interests include statistical machine learning, natural language processing, finger vein recognition, and degradation modeling.

KAIRUI CHEN received the Ph.D. degree in control science and engineering from the Guangdong University of Technology, Guangzhou, China, in 2017. From December 2015 to December 2016, he was a Visiting Scholar with the Automation and Robotics Research Institute, The University of Texas at Arlington. He currently holds a postdoctoral position with the School of Automation, Guangdong University of Technology. His research interests include multiagent system control, neural networks learning, adaptive control, and optimal control.

YUN ZHANG received the B.S. and M.S. degrees in automatic engineering from Hunan University, Changsha, China, in 1982 and 1986, respectively, and the Ph.D. degree in automatic engineering from the South China University of Science and Technology, Guangzhou, China, in 1998. He is currently a Full Professor with the School of Automation, Guangdong University of Technology, Guangzhou. His research interests include intelligent control systems, multiagent systems, neural networks, and signal processing.