Jionghua Jin
Assistant Professor
Department of Systems & Industrial Engineering,
The University of Arizona,
Tucson, AZ 85721-0020

Jianjun Shi[1]
Assistant Professor
Department of Industrial and Operations Engineering,
The University of Michigan,
Ann Arbor, MI 48109-2117
e-mail: shihang@engin.umich.edu

Diagnostic Feature Extraction From Stamping Tonnage Signals Based on Design of Experiments

Diagnostic feature extraction with consideration of interactions between variables is very important, but has been neglected in most diagnostic research. In this paper, a new feature extraction methodology is developed to consider variable interactions by using a fractional factorial design of experiments (DOE). In this methodology, features are extracted by using principal component analysis (PCA) to represent variation patterns of tonnage signals. Regression analyses are performed to model the relationship between features and process variables. Hierarchical classifiers and the cross-validation method are used for root-cause determination and diagnostic performance evaluation. A real-world example is used to illustrate the new methodology. [S1087-1357(00)00302-6]

1 Introduction

A tonnage signal representing a stamping force is the most effective monitoring signal for stamping process control. Tonnage signal analysis provides rich information about stamping process variable changes and has the potential to predict part quality. In general, if a tonnage signal change is detected, it normally reflects that the conditions of process variables have changed, and the part quality may also be affected.

Due to the complexity of a stamping process, many process variables can influence tonnage signals. Thus, it is very complex and challenging to conduct root-cause determination through tonnage signal analysis. Most past research focused on process monitoring, i.e., finding the changes of tonnage signals (or extracted features) to detect the process faults. For example, a Shewhart control chart for single measurement monitoring was often used as the basis in tonnage monitoring systems. In those systems, a peak tonnage feature or other features extracted from a tonnage signal are checked with the respective control limits individually [1–3]. In this approach, the correlation of those features is not considered and the control limits are set for each feature independently. Thus, a high false alarm rate (Type I error) is often encountered. Recently, an effort has been made, by using a multivariate T^2 control chart [4], to monitor the whole waveform of a tonnage signal [5]. In this study, the detection of a waveform change is considered to be a multivariate detection problem. Further research has also been made to achieve feature-preserving data compression of tonnage signals to effectively extract features for process monitoring and diagnosis [6]. By using those approaches, a tonnage monitoring system can identify whether the process is normal or abnormal, based on whether the tonnage signal features are within the control limits. However, those approaches could not identify the root causes if an abnormal process condition is detected.

In root-cause diagnosis, very little research has been done in stamping process control. Among the few publications in this field, most research focuses only on the single-fault situation without considering the interactions among the process variables. A few examples in this category include the detection of material thickness and hardness change [2,5], nitrogen cushion change, shut height change, bearing wear-out, loose tie-rod [3,7], and punch breakage and wear-out [8]. Those detection criteria are effective only when all other variables are considered to be unchanged. In practice, the interactions among the stamping process variables are very significant and complex. As a result, the applicability would be very limited if the root-cause diagnosis method is based on the assumption of a single variable change. Thus, it is desirable to develop root-cause diagnostic methodologies with consideration of multiple variable interactions. As an essential component, diagnostic feature extraction from tonnage signals with consideration of variable interactions has to be studied first.

In this paper, a new diagnostic feature-extraction methodology is proposed to consider the variable interactions by using the fractional factorial design of experiments (DOE). The general framework of the proposed methodology is shown in Fig. 1. The proposed methodology consists of three steps. In the first step, the important process variables are determined and the experimental design is made based on the complexity of the relationship between the process variables and the tonnage signals. (The term ''variable'' used in the paper is the same as the term ''factor'' used in the design of experiments [9]. We use ''variable'' rather than ''factor'' to be consistent with the terminology used in the root-cause determination, as in the process control literature. Based on the DOE, tonnage signals are collected under different setup conditions. In the second step, data reduction is conducted by using principal component analysis (PCA), where the fewer significant eigenvectors are selected to represent the major variation patterns of tonnage signals. The principal components defined as features are obtained by projecting tonnage signals to the selected eigenvectors. Furthermore, in Step 3, a regression model is used to describe the relationship between the features (principal components) and the process variables. Based on the regression analysis, the diagnostic features and diagnostic variables are identified to form the hierarchical classifiers for root-cause determination. The effectiveness of the extracted diagnostic features is finally evaluated according to their classification performances.

When the process variable interactions are considered, the classification of waveform patterns in terms of the process variable setups (conditions) is very complicated. The fault caused by one variable change may generate different patterns of waveform signals if other interaction variables are under different setup levels. Therefore, the traditional fault classification based on a single-fault assumption without considering the variable interactions cannot be applied in this case. For this purpose, a new hierarchical classification structure is proposed in the paper for solving the multivariate classification problem where the variable interactions are considered.

The outline of this paper is listed as follows. After a brief introduction, an overview of a stamping process and the design of experiments are given in Section 2. Section 3 reviews the basis of

[1]Author to whom all correspondence should be addressed.

Contributed by the Manufacturing Engineering Division for publication in the JOURNAL OF MANUFACTURING SCIENCE AND ENGINEERING. Manuscript received Oct. 1998; revised June 1999. Associate Technical Editor: E. C. DeMeter.

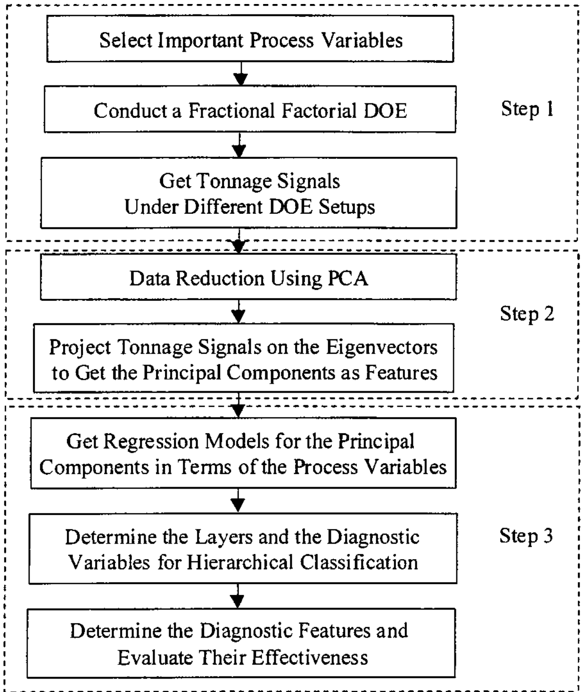

Fig. 1 The framework for diagnostic feature extraction based on DOE

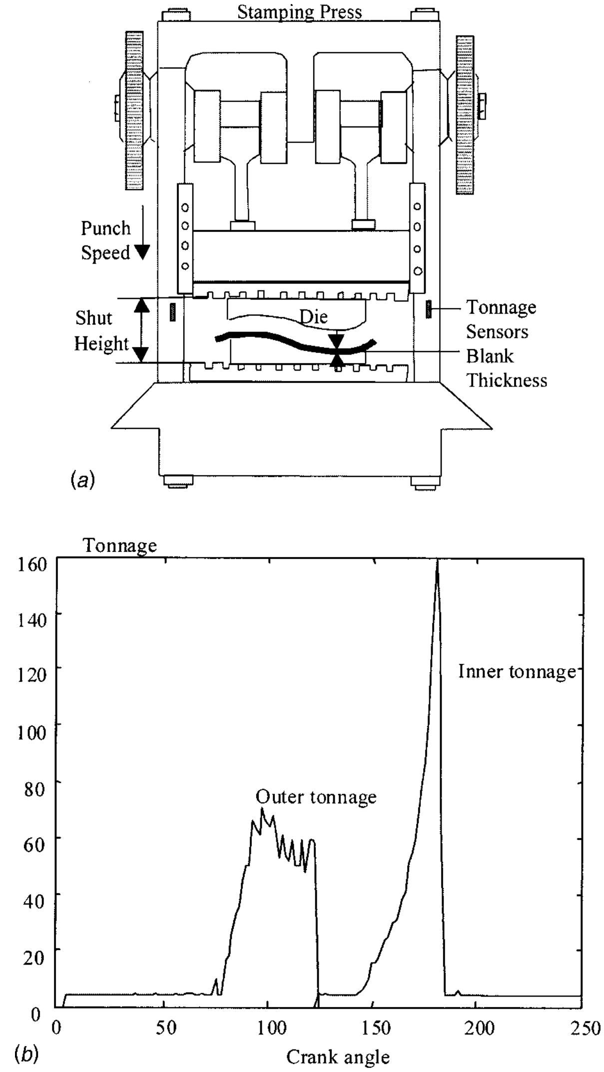

Fig. 2 (*a*) An example of a stamping press. (*b*) One cycle of a tonnage signal

PCA and presents a PCA decomposition model to describe tonnage waveform signals. Feature extraction is further conducted using the principal components. In Section 4, a DOE regression analysis is used to analyze the effect of each process variable on the features. The diagnostic process variables, which are considered as the major root cause of the tonnage-signal change, are selected based on the defined contribution index. Based on the regression analysis, a hierarchical classification structure for fault diagnosis is presented in Section 5. A real-word example is presented in Section 6 to demonstrate the analysis procedures and the effectiveness of the proposed method in stamping process control. Finally, a summary and future work are provided in the last section.

2 Overview of a Stamping Process and Experimental Design

2.1 Overview of a Stamping Process. Sheet metal stamping is a very complex manufacturing process. Stamping process faults are generally caused by the process variable changes, most of which cannot be measured directly or are very expensive to be measured on-line. In recent years, stamping tonnage sensors have been widely used to measure stamping forces. This tonnage signal contains rich information about stamping process changes, which are closely related to process faults. An example of a stamping press with some process variables is shown in Fig. 2(*a*). In order to measure the stamping tonnage force, four tonnage sensors (strain gauge sensors) are mounted on the four press uprights (or on the two linkages). The total stamping force is obtained by the summation of tonnage forces on the four uprights (or on the two linkages). Figure 2(*b*) shows one cycle of a total tonnage signal measured from a double-action forming process. The press crank angle in the *X*-axis is generally used as the reference to represent the rotation position of the main shaft of a press, which can also be translated into the press slide position based on the press motion curve. The crank angle from 0° to 360° corresponds to a complete cycle of a stamping operation. Figure 2(*b*) shows only the effective portion of a stamping force, where the outer tonnage corresponds to the stamping force on the outer binder, and the inner tonnage corresponds to the stamping force on the inner punch.

A stamping process has more than forty process variables. Those variables can be classified into four broad categories: blank material characteristics, die condition and setup parameters, press performance and working parameters, and the interaction variables [7,10]. With so many variables, it is very difficult to conduct a test that studies all those process variables. In the paper, six important process variables are selected based on the process engineering understanding of the characteristics of stamping operations, the performance of the tooling and press machine, the frequency of variable changes, and the resultant severe effect on part quality due to the process variable changes, etc. [11]. Those selected process variables are lubrication, material thickness, outer shut height, inner shut height, punch speed, and blank wash pressure, which are denoted as *A*, *B*, . . . ,*F* respectively in the following sections. If there is no sufficient process engineering knowledge, a screening experiment is suggested as an initial test to determine the effective test variables [9].

2.2 Description of the Design of Experiments. A fractional factorial design of experiments can be conducted to study the effect of the selected six important process variables. In order to obtain the fault patterns of tonnage signals, the training samples

Table 1 Tested process variables and setup values

Variable	A	B	C	D	E	F
Name	Lubri-cation	Material Thickness	Outer Shut-Height (inch)	Inner Shut-Height (inch)	Punch Speed (rpm)	Blank Washer Pressure (psi)
Normal (-)	10%	normal	83.6553	95.9435	7	8
Abnormal (+)	20%	thick	83.6725	95.9649	14	16

Table 2 (*a*) Test matrix for the first test day; (*b*) Test matrix for the second test day

	Test Order															
Variable	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
A	-	-	-	-	-	-	-	-	+	+	+	+	+	+	+	+
B	+	+	+	+	-	-	-	-	-	-	-	-	+	+	+	+
C	+	+	-	-	-	+	+	-	-	+	-	+	+	+	-	-
D	-	+	+	-	+	+	-	-	-	-	+	+	-	+	-	+
E	-	+	-	+	-	+	-	+	+	-	-	+	-	+	+	-
F	-	-	-	-	+	+	+	+	-	-	-	-	+	+	+	+

(a)

	Test Order															
Variable	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32
A	+	+	+	+	+	+	+	+	-	-	-	-	-	-	-	-
B	+	+	+	+	-	-	-	-	-	-	-	-	+	+	+	+
C	-	-	+	+	-	+	+	-	+	-	-	+	-	-	+	+
D	-	+	-	+	-	-	+	+	+	-	+	-	+	-	+	-
E	-	+	+	-	-	+	-	+	-	-	+	+	+	-	-	+
F	-	-	-	-	+	+	+	+	-	-	-	-	+	+	+	+

(b)

considering the effect of variable interactions will be generated from this DOE. In addition to the main effects of the selected variables, the interactions between two variables are also a major concern in the study. The interactions among three or more variables are neglected due to the difficulty in their interpretation. Thus, a standard two-level fractional factorial design with resolution five was used in this DOE plan, i.e., 2_V^{6-1} with the generator F=ABCDE [9]. In this DOE, each main factor aliases with a five-factor interaction, and each two-factor interaction aliases with a four-factor interaction. Therefore, if three or more factor interactions are negligible, this design can provide clear estimates of the effects of the main factors and two-factor interactions. Considering the restriction of testing time, two test days were required, corresponding to two blocks generated by ABF=CDE in the DOE plan. In each block, sixteen different setups were tested on each test day, and nine replicates were used under each setup. Thus, the total number of observations is equal to 288 $(2\times16\times9=288)$. Table 1 gives the real test values for each variable corresponding to the normal setup level and the typical abnormal setup level respectively.

Tables 2(a) and 2(b) show the test matrix in the DOE plan, where "−" and "+" correspond to the normal and abnormal levels of the test variables. Tests 1–16, shown in Table 2(a), were conducted on the first test day, and tests 17–32, shown in Table 2(b), were conducted on the second test day. Due to the complexity of adjusting a stamping process, it is very difficult to conduct this DOE in a completely random sequence. Some variables, such as lubrication and blank washer pressure, are very difficult to adjust accurately. Therefore, the tests for these variables should be conducted in a sequence so that the test requires the minimum number of setup changes of these variables. However, for other easily adjustable variables, such as shut height and punch speed, the tests should be conducted in such a sequence that the test levels of these variables are changed frequently to avoid systematic errors due to different test times.

3 Feature Extraction Based on a PCA Decomposition Model

PCA is used as a multivariate statistical analysis method to handle large numbers of highly correlated data via orthogonal projection. It has the ability to reduce the dimensionality of a monitoring space by projecting original data into a lower-dimensional orthogonal space defined by a few significant eigenvectors [12,13]. Therefore, PCA can be used to simplify the construction and interpretation of multivariate control charts for process monitoring [14–16]. In addition to these process-monitoring applications, recent efforts have been made to integrate PCA with engineering knowledge for root-cause determination in the autobody assembly process with great success [17,18]. In this paper, a new analysis method is developed to consider the variable interactions by integrating the PCA method with the DOE.

3.1 Waveform Signal Representation Using PCA. In this paper, tonnage waveform signals are denoted as a matrix $\mathbf{X}\in\mathbf{R}^{m\times p}$, in which each row vector $\mathbf{x}_i^T$ is a complete cycle of a tonnage signal with p measurement points, and m is the total number of observations. The mean vector of tonnage waveform signals μ^T can be estimated by $\overline{\mathbf{x}}^T=1/m\Sigma_{i=1}^{m}\mathbf{x}_i^T$. After subtracting the vector of $\overline{\mathbf{x}}^T$ from $\mathbf{x}_i^T$, i.e., $\mathbf{y}_i^T=\mathbf{x}_i^T-\overline{\mathbf{x}}^T$, the data matrix $\mathbf{Y}\in\mathbf{R}^{m\times p}$ can be decomposed using PCA [13]:

$$\mathbf{Y}=a_1\mathbf{v}_1^T+a_2\mathbf{v}_2^T+\cdots+a_p\mathbf{v}_p^T \quad (1)$$

where the matrix $\mathbf{Y}$ contains vector $\mathbf{y}_i^T=[y_{i1},y_{i2},\ldots,y_{ip}]$, $(i=1,\ldots,m)$, which has a zero mean. The vector $\mathbf{v}_i\in\mathbf{R}^{p\times1}$ $(i=1,\ldots,p)$ is the normalized eigenvector of the sample covariance matrix $\mathbf{S}$ of $\mathbf{Y}$; that is:

$$\mathbf{S}\mathbf{v}_i=\lambda_i\mathbf{v}_i \quad (2)$$

$$\mathbf{S}=\frac{1}{m-1}\sum_{i=1}^{m}\mathbf{y}_i\mathbf{y}_i^T \quad (3)$$

where λ_i is the eigenvalue corresponding to the eigenvector $\mathbf{v}_i$. The matrix $\mathbf{V}$ consists of eigenvectors $\mathbf{v}_i$ $(i=1,\ldots,p)$ and forms an orthogonal basis for the space spanned by $\mathbf{Y}$. The principal component $a_j\in\mathbf{R}^{m\times1}$ is obtained by projecting $\mathbf{Y}$ onto the vector $\mathbf{v}_j$ $(j=1,\ldots,p)$:

$$a_j=\mathbf{Y}\mathbf{v}_j, \quad (4)$$

3.2 Selection of Significant Variation Patterns. When a data set is projected to eigenvectors, it is often found that only the first few eigenvectors, corresponding to larger eigenvalues, are associated with the systematic process variations, while the remaining eigenvectors reflect the variations of the process noise [13]. The noise in this case relates to the uncontrollable process variations and instrumental variations arising from random disturbances. Therefore, dimension p of the original PCA model shown in Eq. (1) can be reduced into a smaller dimension r_0 ($r_0<p$) in a simplified PCA model for the data reduction. The eigenvectors corresponding to these r_0 largest eigenvalues are used as the decomposition basis in the simplified PCA model. So, the residual variance resulting from this dimensional reduction can be obtained by:

$$S_{res}^2=\text{trace}(\mathbf{S})-\sum_{i=1}^{r_0}\lambda_i=\sum_{i=r_0+1}^{p}\lambda_i, \quad r_0<p \tag{5}$$

At the first stage of PCA modeling, the value r_0 needs to be determined. A two-step analysis is provided for the selection of the r_0 value. In the first step, the initial value r_0 is determined so that the residual variance, S_{res}^2, is commensurate with the process noise. For this purpose, an order selection index η_r is defined as:

$$\eta_r=\frac{\sum_{i=1}^{r_0}\lambda_i}{\sum_{i=1}^{p}\lambda_i} \tag{6}$$

An appropriate r_0 is selected so that η_r is close to one. In practice, η_r is determined based on the signal-to-noise ratio requirement. Based on this initial r_0, a regression analysis is conducted for these candidate variation patterns. Then, the value r_0 is further determined in the second step based on the modeling capability of principal components in terms of the process variables. A low R-square value in the regression analysis [9] indicates that this linear model is not sufficient to describe the change of the respective principal component. Therefore, r_0 is finally determined by the total number of the principal components that can be modeled in the DOE regression analysis. The detailed rules are discussed in Section 6.

It can be verified that the PCA model is an optimal linear model with a model dimension equal to r_0, which has a minimized S_{res}^2 over all possible choices of bases [19]. Because the effective dimension of the PCA model is reduced from p to r_0, the effective dimension of feature (principal component) space is r_0. Consequently, the original large dimension of a tonnage signal $\mathbf{x}_i^T=[x_{i1},x_{i2},\ldots,x_{ip}]$ can be expressed by a smaller dimension of features, i.e., $[a_{i1},a_{i2},\ldots,a_{ir_0}]$. The effect of the process variable changes on the tonnage waveform signals will be represented as the changes of these features. Thus, these principal components will be potentially used as diagnostic features in the subsequent analysis. Furthermore, the diagnostic variables, which are considered as the root causes of the variation pattern, can be identified in terms of their significant effects on the changes of the respective features; this will be discussed via the DOE regression analysis in the following section.

4 Diagnostic Variable Identification Based on DOE Regression Analysis

4.1 DOE Regression Analysis. The principal components calculated from Eq. (4) are considered as the potential diagnostic features and will be used as the response variables in the DOE regression analysis. The analysis of variance (ANOVA) is used to identify whether the process variable has a significant effect on the selected features due to different variable setups. The basis for this analysis is based on the following hypothesis testing:

$$H_0: \mu(a_{ik})=\mu(a_{jk})$$

$$H_1: \mu(a_{ik})\neq\mu(a_{jk}) \quad (i\neq j; \quad i,j=1,\ldots,N; \quad k=1,\ldots,r_0) \tag{7}$$

where k is an index of the kth principal component corresponding to the kth eigenvector; i and j are indices of the test setups. $\mu(a_{ik})$ represents the mean value of the principal component a_{ik} under setup i. N is the total number of different setups conducted in the DOE. Because the eigenvectors are sorted based on their eigenvalues, the first eigenvector and the first principal component ($k=1$) represent the most significant variation pattern of tonnage signal changes under the different variable setups in the DOE. The significance of other variation patterns will be determined by their respective eigenvalues.

For a given eigenvector $\mathbf{v}_k$, the change of $\mu(a_{ik})$ ($i=1,\ldots,N$) due to different variable setups represents the effect of the variable changes on feature a_k. The significance of each variable is reflected by its contribution to the variability of feature $\mathbf{a}_k$. The significant variables having the major contributions to the feature's variability are called diagnostic variables. Because each variable shows different significances on different variation patterns $\mathbf{v}_j$, ($j=1,\ldots,r_0$), the diagnostic variables should be identified corresponding to each variation pattern. The detailed analysis procedure for diagnostic variable identification is summarized in Fig. 3.

In Fig. 3, PCA is used to obtain all eigenvectors and the associated principal component vectors. Starting from the first principal component vector of the first variation pattern, DOE regression analysis is used to determine the diagnostic variables, which are the major contributors to the variability of this principal component. This iteration analysis is repeated until all significant variation patterns (r_0) have been analyzed. The identified diagnostic variables are used to form the clusters from the DOE observations, and a hierarchical classification structure is used to simplify the interactions between the diagnostic variables. The clustered observations can further serve as training samples in the development of the hierarchical classifier for the root-cause determination in terms of the diagnostic variables.

Based on the DOE plan of 2_V^{6-1} given in Section 1, the main effects and two-factor interaction effects can be clearly estimated if the higher-order interactions are neglected. Thus, the corresponding regression model can be expressed as:

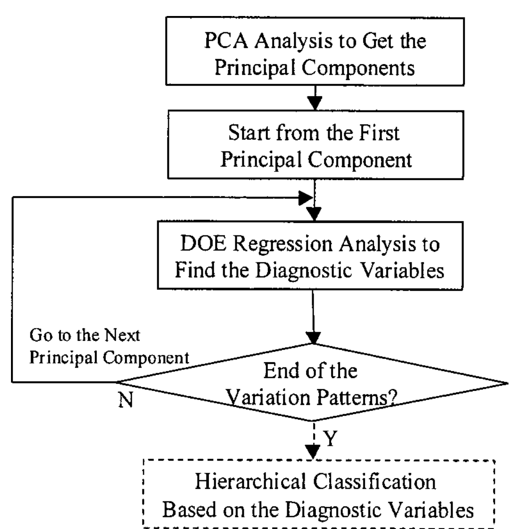

Fig. 3 Diagnostic variable identification based on DOE

$$a_{ij}=\beta_{j0}+\sum_{k=1}^{6}\beta_{jk}z_{ik}+\sum_{k=1}^{5}\sum_{l=k+1}^{6}\beta_{jkl}z_{ik}z_{il}+\beta_{j,\text{block}}z_{i,\text{block}}+\varepsilon_{ij} \tag{8}$$

where j ($j=1,\ldots,r_0$) is the index of principal components corresponding to variation pattern j in the PCA model, and i ($i=1,\ldots,m$) is the index of the observations. $z_{i1},\ldots,z_{i6}$ represent the coded variable setup values (normal condition $=-1$ and abnormal condition $=1$) in the ith observation corresponding to the six process variables $A,\ldots,F$ respectively. $z_{i,\text{block}}$ represents the test block value, i.e., $z_{i,\text{block}}=-1$ for the first test day and $z_{i,\text{block}}=1$ for the second test day. β_{j0}, β_{jk}, β_{jkl} and $\beta_{j,\text{block}}$ are the regression coefficients, which correspond to the overall mean of the observations, the estimate of one-half of the effect of variable k, the estimate of one-half of the interaction effect between two variables k and l, and the estimate of one-half of the effect of the test block respectively. ε_{ij} corresponds to the model error, which is assumed to be normally and independently distributed random variables with the mean of zero and the variance of σ^2.

4.2 Diagnostic Variable Identification. Based on the DOE regression analysis, the diagnostic variables can be selected from those variables that have the major contributions to the total model variability [9]. For this purpose, a new term, "contribution index," is defined as the percentage of the variability of variable $k(SS_k)$ to the total model variability (SS_{model}) of all variables:

$$\rho_k=\frac{SS_k}{SS_{\text{model}}}\times 100\text{ percent} \tag{9}$$

The subscript k in ρ_k and SS_k can be replaced by "ij" to represent the interaction between variables i and j when $SS_{ij}>(SS_i$ or $SS_j)$. Based on the contribution index, the diagnostic variables are selected from those which have a larger value of ρ_k. Considering the classification difficulty under a large number of clusters, the maximum number of diagnostic variables in this analysis is limited to two for each variation pattern. The decision of using one or two diagnostic variables is made according to the following rules:

$$\begin{array}{ll} \text{If } \rho_k>\xi, & \text{one diagnostic variable } k \text{ is selected;} \\ \text{else if } \rho_k+\rho_j>\xi, & \text{two diagnostic variables } k \text{ and } j \\ & \text{are selected;} \\ \text{else,} & \text{no diagnostic variables are selected;} \end{array} \tag{10}$$

where ξ is a predefined contribution limit, which is selected according to process variability and process noise in the application. The detailed discussion on the usage of the contribution limit ξ will be given in the example in Section 6.

5 Hierarchical Classification Based on Diagnostic Features

5.1 Hierarchical Classification Structure. If variable interactions exist, the effect of one variable on the tonnage signals depends on the levels chosen for the other interaction variables. As a result, a large quantity of fault clusters will be generated by considering different variables and their interactions. Thus, it is very difficult to design a classifier for simultaneous classification of all fault clusters. In order to solve this problem, a new hierarchical classifier is proposed to simplify the fault clustering and classification. Figure 4 shows an example of the hierarchical classifier. In the example, W_i represents the ith classified diagnostic variable, and "$-$" or "$+$" represents the normal or abnormal setup level of the corresponding variable.

In the hierarchical clustering structure, the classification is conducted in a sequence from the top layer to the bottom layer. Each layer may consist of one or more classifiers depending on the number of nodes at the upper layer. Each classifier is designed to identify the conditions of the nodes which are connected with the same node at the upper layer. For example, there is one classifier in Layer 1 in Fig. 4. The nodes $W_1(+)$ and $W_1(-)$ are used to design the classifier for W_1's condition identification. Similarly, there are two classifiers in Layer 2 in Fig. 4. The nodes $W_2W_3(--)$, $W_2W_3(+-)$, $W_2W_3(-+)$, and $W_2W_3(++)$ under the condition of $W_1(+)$ are used to design one classifier for identification of those conditions. The other classifier is designed for those nodes under the condition of $W_1(-)$. By using the hierarchical clustering structure, the fault diagnosis is performed to sequentially identify the diagnostic variable conditions from the upper layer to the lower layers according to the significance of the variation patterns sorted by eigenvalues. Moreover, this strategy can simplify the variable-interaction effect on the classification problem. For example, when we classify W_2W_3 conditions at the second layer, the W_1 condition is known (fixed) at the first layer. Thus, the interaction between variables W_2 (or W_3) and W_1 can be eliminated from the regression model of the second layer.

In order to develop a hierarchical classifier, two issues will be addressed in the following sections: (1) how to select the diagnostic variables to form a layer; and (2) how to select diagnostic features for classifier design in each layer.

For the first issue, the diagnostic variables are selected from each variation pattern sorted by the eigenvalues. The detailed analysis is summarized as the following two steps: (1) regression analysis of the principal components, which are selected sequentially according to the variation patterns sorted by the eigenvalues; and (2) based on the regression analysis, identification of the diagnostic variables by comparing the contribution index defined in Eq. (9) with the contribution limit ξ. The number of diagnostic variables in each layer is determined by Eq. (10). The identified diagnostic variables will then be used to form a layer in the hierarchical classification structure.

Once the number of the diagnostic variables at each layer is determined, the number of nodes N_i at the ith layer can be deter-

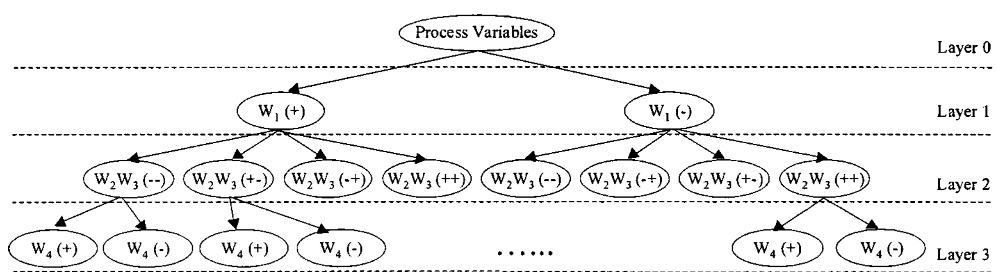

Fig. 4 An example of classification based on a hierarchical structure

mined by $N_i = d^q N_{i-1}$, where d is the number of the setup levels of test variables, q is the number of diagnostic variables to be classified in this layer, and N_{i-1} is the number of nodes in the upper layer. In this paper, d equals two because a two-level fractional factorial design is conducted in the experiment and q is equal to either one or two, which is determined by Eq. (10). Therefore, the maximum number of clusters included in each classifier is four.

The major concern in the second step is how to select the diagnostic features to classify the diagnostic variable conditions. If one diagnostic variable is identified in the first step, two clusters will be formed for this variable corresponding to its two setup levels. In this case, one feature is sufficient for the classification of these two clusters. Therefore, the principal component of the respective variation pattern can be used directly as the diagnostic feature for classification of this variable condition. If two diagnostic variables are identified, one feature may not be sufficient to classify four clusters formed by two setup levels of these two diagnostic variables. Thus, additional features should be selected from other principal components for the classification purposes. The availability of other principal components depends on whether they have the same selected diagnostic variables for their variation patterns. According to [19], the required maximum number of features is less than the number of clusters to be classified. Thus, the maximum number of features equals three in this paper.

The advantages of the presented hierarchical classifier can be summarized as follows: (1) The complicated mapping relationship due to the interactions among process variables can be significantly simplified by using a hierarchical classification structure. This simplification is achieved by fixing the pre-identified variable conditions at the upper layer to simplify the variable interactions and reduce the number of clusters in the classifier design. (2) The diagnostic variables in the higher layers are more important because they correspond to the more significant variation patterns with the larger eigenvalues. Thus, the diagnostic variable classification, which is performed from the top layer to the lower layers, can potentially lead to an effective process control strategy by sequentially adjusting the variable conditions. (3) The hierarchical clusters are formed in terms of different variable conditions; thus the classification results can provide explicit information for fault diagnosis and process improvement.

5.2 Fast Updating the Contribution Index in a Hierarchical Classifier. In DOE regression analysis, the sum of the squares for each variable variability (SS_k) is calculated simultaneously based on all test observations [9]. However, under the hierarchical structure, SS_k in Eq. (9) (except for the first layer) should be modified to consider the effect of the fixed variables at the upper layers. In order to avoid the tedious recalculation of SS_k from the samples, a ''fast-update'' algorithm is developed to modify the contribution indices in the hierarchical structure by directly using the regression coefficients, which have already been estimated based on all test observations.

When a two-level fractional factorial DOE is used, the regression coefficient β_k in Eq. (8) (for simplicity, the subscript j of the variation pattern index is ignored in this section) and SS_k can be calculated based on the defined contrast relationship [9]:

$$\beta_k = (\text{Contrast}_k)/(n \cdot 2^b) \tag{11}$$

$$SS_k = (\text{Contrast}_k)^2/(n \cdot 2^b) \tag{12}$$

where the subscript k in β_k and SS_k can be replaced by ''kl'' to represent the interaction term between variables k and l. n is the number of the test replicates, and b is the number of variables. From Eqs. (11) and (12), it can be seen that for a given n and b, SS_k is proportional to the square of the regression coefficients. Therefore, the contribution index defined in Eq. (9) can be calculated by:

$$\rho_k = \frac{\beta_k^2}{\sum_{i=1}^{b} \beta_i^2 + \sum_{i=1}^{b-1} \sum_{j=i+1}^{b} \beta_{ij}^2} \times 100 \text{ percent} \tag{13}$$

Because a hierarchical structure is used for classification, when we classify the diagnostic variables at the current layer, the levels of the diagnostic variables at the upper layer are already known (fixed). Therefore, the regression model at the current layer should not contain the diagnostic variables of the upper layer. The regression coefficients, which represent the interaction effect of the diagnostic variables at the upper layer, should be modified. A simple example is given here to illustrate this idea.

Assume two variables, A and B, are studied in the test. Variable A is the diagnostic variable at the top layer, and variable B is the diagnostic variable at the second layer. For $A=$''+'' fixed at the upper layer, the original interaction coefficient β_{AB} should no longer exist in the regression model at the second layer, but should be included in the modified coefficient β_B^* by:

$$\beta_B^* = \beta_B + \beta_{AB}|_{A=\text{''+''}} \tag{14}$$

Proof: $\beta_B + \beta_{AB}|_{A=\text{''+''}}$

$$= (\text{Contrast}_B + \text{Contrast}_{AB|A=\text{''+''}})/(n2^b)$$

$$= 2(\text{Contrast}_{B|A=\text{''+''}})/(n2^b)$$

$$= (\text{Contrast}_{B|A=\text{''+''}})/(n2^{b-1}) = \beta_B^* \qquad \#$$

In this proof, the number of the variables included in the second regression model is reduced from $b=2$ to $b-1=1$ because variable A is eliminated from the second regression model. Similarly, if three variables, A, B, and C, are studied and two diagnostic variables, A and B, are fixed at the upper layer, such as $A=$'' $-$'' and $B=$''+,'' then the modified coefficient β_C^* of variable C in the current regression model can be calculated by:

$$\beta_c^* = \beta_C - \beta_{AC} + \beta_{BC} \tag{15}$$

Generally, under the hierarchical classification structure, the modified regression coefficients can be simply calculated by a summation or a subtraction of the original regression coefficients, which depends on the fixed conditions of the diagnostic variables at the upper layers.

Based on Eq. (13), at different layers except for the top layer, the modified contribution index of a variable can be quickly updated with the modified regression coefficients by:

$$\rho_k^* = \frac{\beta_k^{*2}}{\sum_{i=1}^{b^*} \beta_i^{*2} + \sum_{i=1}^{b^*-1} \sum_{j=i+1}^{b^*} \beta_{ij}^*} \tag{16}$$

where * is used to represent the modified parameters after fixing the levels of the diagnostic variables at the upper layer. b^* represents the number of the variables excluding the diagnostic variables fixed at the upper layers. β_i^* and β_{ij}^* correspond to the modified coefficients for variable i and the interaction between variables i and j, where variables i and j do not include the diagnostic variables fixed at the upper layers.

6 An Example: Tonnage Signal Classification Based on the DOE

6.1 DOE-Based PCA Decomposition and Regression Analysis. Based on the DOE discussed in Section 2.2, 32 different setups with 9 test replicates in each setup have been conducted, as shown in Tables 2(a) and 2(b). The total of 288 tonnage signals are obtained and analyzed using PCA. The first ten largest eigenvalues (initial $r_0=10$) and the order selection index of η_k ($k=1,\ldots,10$) are obtained by using Eqs. (2), (3), and (6), and shown in Table 3. The other eigenvalues are ignored because they

Table 3 Eigenvalues of PCA and *R*-squares of regression models

Variation Pattern (k)	1	2	3	4	5	6	7	8	9	10
Eigenvalue	7352.0	545.4	83.6	42.7	36.6	31.6	20.2	16.4	11.0	10.4
$\eta_k = \sum_{i=1}^{k} \lambda_i / \sum_{i=1}^{p} \lambda_i \, (\%)$	89.31	95.94	95.96	97.48	97.92	98.30	98.55	98.75	98.88	99.01
R-Square	0.99	0.95	0.86	0.66	0.72	0.41	0.35	0.46	0.46	0.27

Table 4 Regression model parameters

Model Parameters	Variation Patterns				
	j = 1	j = 2	j = 3	j = 4	j = 5
β_{j0}	380.72	-160.96	24.02	-39.10	7.52
β_{j1}	7.49	-2.22	0.02	-1.60	0.57
β_{j2}	12.10	-17.09	3.04	1.25	-1.06
β_{j12}	-10.84	5.28	-0.49	0.42	-0.04
β_{j3}	7.09	2.49	-1.47	0.67	-2.48
β_{j13}	3.72	-0.70	0.18	-0.64	-0.21
β_{j23}	3.44	-0.51	0.46	0.68	0.50
β_{j4}	-77.65	-7.01	-1.85	-1.31	-0.05
β_{j14}	0.98	-1.08	-0.23	-0.41	0.03
β_{j24}	2.17	-1.56	0.71	-1.66	-0.15
β_{j34}	1.06	0.60	0.24	0.42	3.86
β_{j5}	13.77	-9.69	-7.12	1.34	0.97
β_{j15}	3.99	-1.03	-0.24	-0.78	0.075
β_{j25}	3.05	-1.13	0.85	0.92	1.54
β_{j35}	1.38	0.07	0.36	-1.52	-0.73
β_{j45}	3.66	-0.35	0.60	1.17	-0.45
β_{j6}	9.92	-4.24	0.61	-1.33	0.03
β_{j16}	-8.97	3.20	0.006	1.58	0.39
β_{j26}	-9.37	4.09	-0.28	1.58	-0.68
β_{j36}	3.53	-0.81	0.036	-0.27	-0.01
β_{j46}	1.08	-1.74	-0.096	-0.49	-0.10
β_{j56}	3.81	-0.60	-0.002	-0.97	0.16
$\beta_{j,block}$	-7.36	6.04	-1.77	0.23	-0.65

are so small that the residual variance can be ignored according to η_k. The eigenvectors corresponding to these ten eigenvalues are then used as the decomposition bases to obtain ten vectors of the principal components based on Eq. (4). Based on Eq. (8), each principal component vector is analyzed with a regression model, and the modeling adequacy is described by the *R*-square value, as shown in Table 3. It can be seen that when the variation pattern index $k>5$, *R*-square values are very small. So the final effective dimension of the PCA model is equal to five, i.e., $r_0=5$. The regression coefficients in Eq. (8) of these five regression models are listed in Table 4.

6.2 Selection of Diagnostic Variables in Each Layer for Hierarchical Classification.

(1) Layer 1: In the development of a hierarchical structure, as shown in Fig. 4, the first regression model corresponding to the variation pattern with the largest eigenvalue is used to determine the diagnostic variables at the top layer. The contribution indices for all variables or interaction terms are calculated based on Eq. (9). Table 5 gives the contribution index values related to the first four most significant variables, where subscripts of β (1, 2, . . . , 6) correspond to variables $A, B, \ldots, F$ respectively.

In order to determine the number of diagnostic variables, $\xi = 70$ percent is used in this case study. Because the contribution index of variable *D* is equal to 85.46 percent in Table 5, which is larger than ξ, one diagnostic variable *D* (or variable 4) is selected at the top layer. The corresponding principal component is used as the diagnostic feature for classification. Figure 5 plots this diagnostic feature of all observations in the DOE.

(2) Layer 2: Similarly, the second regression model is used to find the diagnostic variables at the second layer. Based on Eq. (16), the contribution indices for all variables except for variable *D* are calculated using the modified regression coefficients with the consideration of the fixed variable *D* at the top layer. Table 6 gives the first two largest contribution indices of the related variables. The modified model coefficients and the corresponding formula are also shown in Table 6.

Because the maximum contribution index due to variable $2(\beta^*_{j2})$ is less than ξ, and the summation of the contribution indices of two variables 2 and 5 (*B* and *E*) are larger than ξ, these two variables are used as the diagnostic variables at Layer 2. Because these two diagnostic variables form four clusters, one feature is not sufficient for classification, and more features are needed. For this purpose, other variation patterns are investigated to select an appropriate regression model that also has *B* and *E* as its diagnostic variables.

Repeating the analysis procedures for the third regression model, the first two largest contribution indices of the related variables (or interaction terms) are given in Table 7. From this

Table 5 The contribution indices for the first variation pattern ($j=1$)

Variable	β_{j4}	β_{j5}	β_{j2}	β_{j3}
ρ_k(%)	85.46	2.69	2.07	1.7

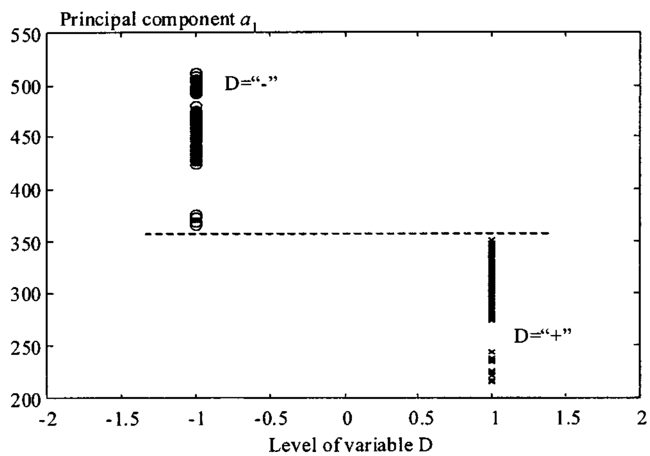

Fig. 5 Classification based on variable *D*

Table 6 The contribution indices for the second variation pattern ($j=2$)

	D= "-"		D= "+"	
	$\beta^*_{j2}=\beta_{j2}-\beta_{j24}$	$\beta^*_{j5}=\beta_{j5}-\beta_{j45}$	$\beta^*_{j2}=\beta_{j2}+\beta_{j24}$	$\beta^*_{j5}=\beta_{j5}+\beta_{j45}$
Parameter	-15.53	-9.34	-18.65	-10.04
ρ_k(%)	59.63	21.57	62.36	18.07

Table 7 The contribution indices for the third variation pattern ($j=3$)

	D= "-"		D= "+"	
	$\beta^*_{j5}=\beta_{j5}-\beta_{j45}$	$\beta^*_{j2}=\beta_{j2}-\beta_{j24}$	$\beta^*_{j5}=\beta_{j5}+\beta_{j45}$	$\beta^*_{j2}=\beta_{j2}+\beta_{j24}$
Parameter	-7.72	2.33	-6.52	3.75
ρ_k(%)	86.93	7.92	69.35	22.94

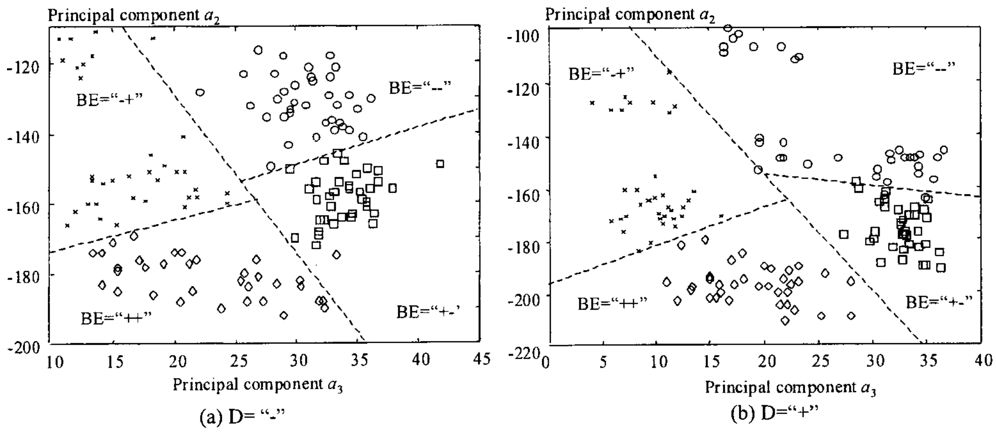

Fig. 6 Classification based on variables *B* and *E* (6(*a*) *D*=''−''; 6(*b*) *D*=''+'')

table, the same diagnostic variables 2 and 5 are identified based on the updated contribution indices when the condition of variable 4 is fixed.

Because the same diagnostic variables 2 and 5 (*B* and *E*) are selected in the second and third regression model, these two corresponding principal components are used as diagnostic features in the classification. Figures 6(*a*) and 6(*b*) plot these two features of the DOE samples under the conditions of $D=``+"$ and $D=``-"$ respectively.

(3) Layer 3: For the fourth regression model, the contribution indices under the fixed conditions of variables 2, 4, and 5 (*B*, *D*, and *E*) are shown in Table 8. Because in the fourth regression model, the contribution indices of one or two variables are less than the minimum contribution value ξ, no diagnostic variable could be found for the fourth variation pattern. Therefore, the diagnostic variables used at Layer 3 should be determined from the remaining regression models.

For the fifth regression model $(j=5)$, the updated contribution indices under the fixed variables *D*, *B*, and *E* are given in Table 9. It is found that under $D=``-,"$ variable *C* can be considered as the diagnostic variable at the third layer because the contribution index of variable $C(\rho_3)$ is larger than $\xi$. However, no diagnostic variable can be identified under the condition of $D=``+"$ because no consistent variables can be identified as the diagnostic variable based on Eq. (10). Figures 7(*a*), 7(*b*), 7(*c*), and 7(*d*) plot the corresponding principal component under the conditions of $DBE=``---,"$ $DBE=``--+,"$ $DBE=``-+-,"$ and $DBE=``-++."$

Based on the above analysis, the final hierarchical classifier is shown in Fig. 8. It can be seen that variable *D* is the most important variable of six variables because it associates with the most significant variation pattern. Variables *A* and *F* have the least impact on the tonnage waveform signals. This analysis result is consistent with the stamping engineering knowledge.

6.3 Classifier Design and Performance Validation

6.3.1 Piecewise Linear Classifiers. Among all given clusters, the closest cluster k^* for a given observation $\mathbf{x}_i$ can be identified by using the following piecewise linear classifier [19]:

$$\text{Cluster } k^*=\max_k\left\{\mu(k)^T\Sigma^{-1}\mathbf{x}_i-\frac{1}{2}\mu(k)^T\Sigma^{-1}\mu(k)+\ln P_0(k)\right\} \quad (17)$$

where $P_0(k)$ is the prior probability of the cluster *k*. Generally, if there is no prior knowledge available, $P_0(k)$ is assumed to be equal for all clusters. $\mu(k)$ is the mean of cluster *k* and Σ is the pooling variance of all clusters, which are estimated by:

Table 8 The contribution indices for the fourth variation pattern $(j=4)$

	D "-"				D "+"			
	BE "--"	BE "-+"	BE "+-"	BE "++"	BE "--"	BE "-+"	BE "+-"	BE "++"
ρ_1 (%)	0.00	13.27	3.96	64.86	9.47	23.51	18.61	75.31
ρ_3 (%)	42.30	6.81	62.15	0.71	17.16	15.92	41.03	2.54
ρ_{13} (%)	4.65	2.26	2.35	4.65	5.91	1.71	2.80	2.99
ρ_6 (%)	23.87	63.47	16.79	0.60	30.36	48.10	19.99	0.39
ρ_{16} (%)	28.35	13.79	14.33	28.35	36.05	10.45	17.07	18.24
ρ_{36} (%)	0.83	0.4	0.42	0.83	1.05	0.31	0.50	0.53

Table 9 The contribution indices for the fifth variation pattern $(j=5)$

	D "-"				D "+"			
	BE "--"	BE "-+"	BE "+-"	BE "++"	BE "--"	BE "-+"	BE "+-"	BE "++"
ρ_1 (%)	1.32	0.55	1.50	0.54	10.61	31.03	4.38	7.88
ρ_3 (%)	96.05	98.52	97.02	97.68	64.07	2.74	90.33	57.69
ρ_{13} (%)	0.11	0.07	0.16	0.10	1.09	5.37	0.58	1.92
ρ_6 (%)	2.13	0.60	0.75	1.34	20.47	42.35	2.69	25.86
ρ_{16} (%)	0.39	0.26	0.57	0.34	3.76	18.51	2.02	6.65
ρ_{36} (%)	0	0	0	0.00	0.00	0.00	0.00	0.00

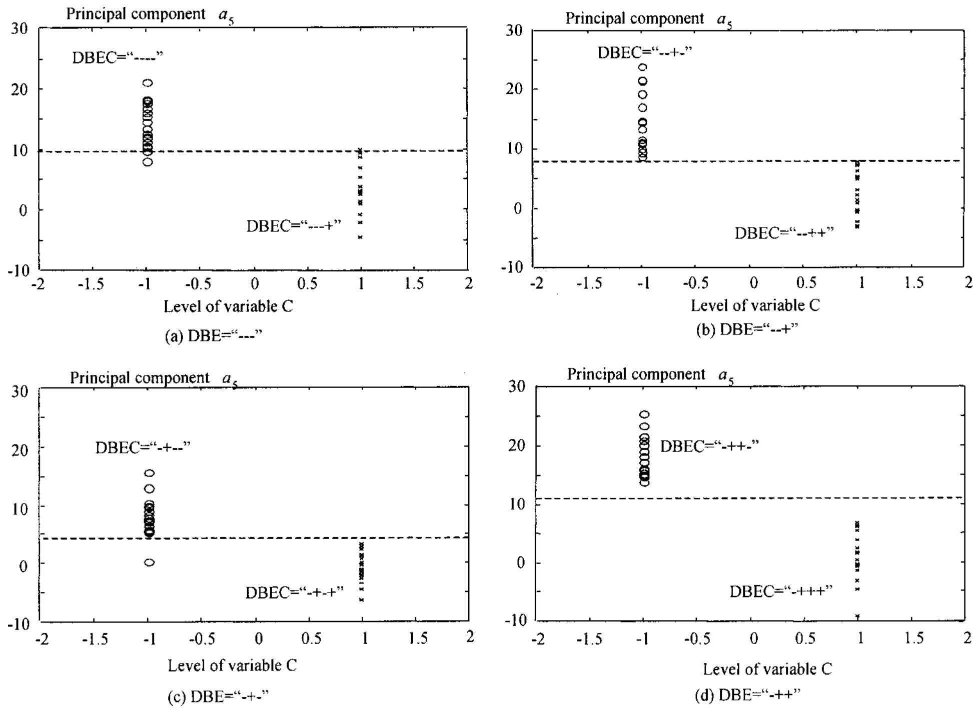

Fig. 7 Classification based on variable *C*

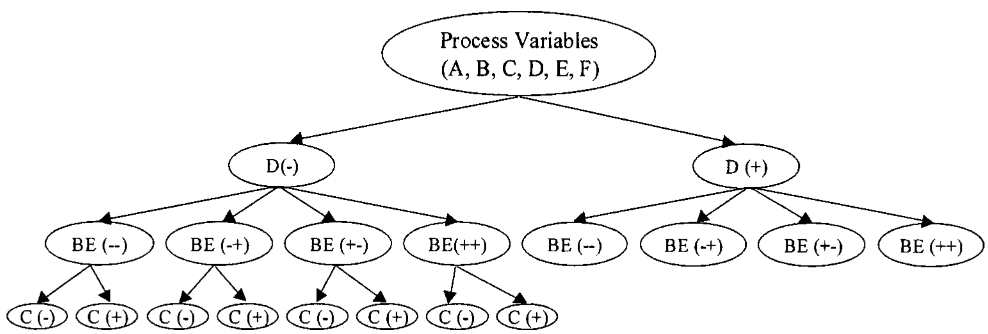

Fig. 8 Hierarchical classifier for the stamping process

$$\hat{\mu}(k)=\frac{\sum_{j=1}^{n_k} x_{j|k}}{n_k}$$

and

$$\hat{\Sigma}=\frac{1}{\left(\sum_k n_k\right)-1}\sum_k \sum_j^{n_k} (\mathbf{x}_{j|k}-\hat{\mu}(k))(\mathbf{x}_{j|k}-\hat{\mu}(k))^T \quad (18)$$

where $\mathbf{x}_{j|k}$ is sample j included in cluster k, and n_k is the number of samples in cluster k.

6.3.2 Classification Error Estimation Using the Cross-Validation Method. The limited number of samples can affect the estimation of classification errors. The cross-validation method can provide better performance than the holdout method when the number of samples is limited [19]. In this paper, the cross-validation method is used to validate the performance of the designed piecewise linear classifiers.

The basic idea of the cross-validation method is that one sample is excluded from cluster k. Next, the parameters of the respective cluster k are re-estimated based on the remaining n_k-1 samples. Afterwards, the excluded sample will be used as the test sample to test the new classifier. This analysis procedure is repeated n_k times to test all n_k samples. The number of misclassified samples is counted to calculate the estimate of the classification error by:

$$\text{error}=\frac{N_{\text{mis}}}{N_{\text{sample}}}\times 100 \text{ percent} \quad (19)$$

where N_{mis} represents the number of misclassification and N_{sample} is the total number of the samples classified by the tested classifiers.

6.3.3 Classification Results. In the hierarchical classification structure (Fig. 8), one classifier is required to classify variable D's conditions with two clusters; two classifiers corresponding to D = ''+'' and D = ''−'' are needed to classify the conditions of variables B and E with four clusters in each classifier; and four classifiers, corresponding to the different conditions of variables B and E with the fixed D = ''−,'' are needed to classify variable C's conditions with two clusters in each classifier. From the classification plots in Section 6.2 (Figs. 5, 6, and 7), it can be seen that a piecewise linear classifier can be used for each case. It is noted

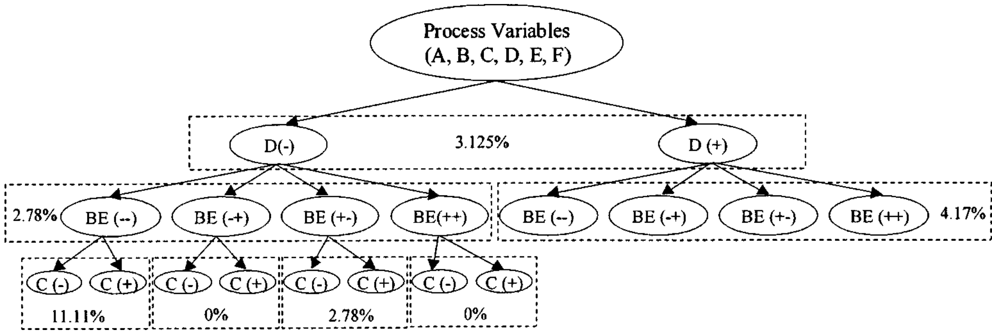

Fig. 9 Misclassification rates of the hierarchical classifiers

that in this study, the classification is conducted based on the extracted diagnostic features (principal components) instead of the observations themselves. Therefore, when we design the classifiers, $\mathbf{x}_i$ or $\mathbf{x}_{j|k}$ should be substituted by the selected features (principal components) in Eqs. (17) and (18) respectively.

The performance of the designed piecewise linear classifiers based on the selected diagnostic features can be evaluated with the classification errors defined in Eq. (19), where $N_{\text{sample}}=N_t/2^t$, $N_t=288$, and t is equal to the number of the variables fixed at the upper layers. Figure 9 shows the percentage of the misclassification errors in each classifier. It can be seen that the maximum classification error is 11.11 percent only in the classification of $C(-)$ and $C(+)$ at the lowest layer. All others, which correspond to the more important variables at the upper layers, are less than 5 percent.

7 Conclusions

A new coherent methodology for diagnostic feature extraction from the stamping tonnage signal is developed based on the design of experiments. One of the unique characteristics of the proposed methodology is that it considers variable interactions in the feature extraction for stamping process monitoring and diagnosis.

It should be pointed out that the framework of conducting diagnostic feature extraction based on DOE is innovative and generic. Even though DOE has been used often in practice to identify the significant variables and their interactions, little research has been done to use the DOE concept for feature extraction and classification in diagnostics. In the proposed methodology, the DOE plays an essential role in the analysis, which provides the basis for: 1) tonnage signal collection under different conditions; 2) identification of the diagnostic features and their relationship with process variables; and 3) determination of the hierarchical classification structure by using the contribution index based on the regression analysis. Although PCA is used in this paper for data reduction and identification of potential features (i.e., principal components), the proposed methodology of hierarchical classification based on the DOE can be applied for diagnostic feature extraction using other features, such as the segmental mean and variance, wavelet coefficients, etc. Furthermore, this methodology can also be applied to other waveform signals, such as force signals in welding, torque signals in bolting, force signals in machining, etc. Therefore, the proposed methodology has broad applications in feature extraction and diagnostic system development.

Acknowledgements

This work was partially supported by the NSF CAREER award: DMI 9624402 and NSF Grant: DMI 9713654. The authors would also like to gratefully acknowledge the referees for their many insightful comments.

References

[1] Seem, J. E., and Knussmann, K. D., 1994, ''Statistical Methods for On-Line Fault Detection in Press-Working Applications,'' Signature Technology Technical Report.
[2] Robbins, T., 1995, ''Signature-Based Process Control & SPC Trending Evaluate Press Performance,'' Metal Forming, pp. 44–50.
[3] Koh, C. K. H., Shi, J., and Williams, W., 1995, ''Tonnage Signature Analysis Using the Orthogonal (Haar) Transforms,'' NAMRI/SME Trans., **23**, pp. 229–234.
[4] Montgomery, D., 1996, *Introduction to Statistical Quality Control*, Wiley, New York.
[5] Jin, J., 1999, *Feature Extraction of Waveform Signals for Stamping Process Monitoring and Fault Diagnosis*, Ph.D. Thesis, The University of Michigan, Ann Arbor.
[6] Jin, J., and Shi, J., 1999, ''Feature-Preserving Data Compression of Stamping Tonnage Information Using Wavelets,'' Technometrics, **41**, pp. 327–339.
[7] Koh, C. K. H., Shi, J., and Black, J., 1996, ''Tonnage Signature Attribute Analysis for Stamping Process,'' NAMRI/SME Trans., **23**, pp. 193–198.
[8] Chen, P., 1997, *Punch Condition Monitoring in Sheet Metal Stamping Under Progressive Stamping Environments*, Ph.D. Dissertation, The University of Michigan, Ann Arbor.
[9] Montgomery, D., 1997, *Design and Analysis of Experiments*, Wiley, New York.
[10] Siekirk, J., 1986, ''Process Variable Effects on Sheet Metal Quality,'' J. Met., **4**, No. 3, pp. 262–269.
[11] Auto Body Consortium (ABC), 1998, ''Agile and Precision Sheet Metal Stamping-Near Zero Stamping Projects,'' Quarterly Technical Report to NIST Advanced Technology Program, Vol. NZS-301.
[12] Castro, P. E., Lawton, W. H., and Sylvestre, E. A., 1986, ''Principal Modes of Variation for Processes with Continuous Sample Curves,'' Technometrics, **28**, No. 4, pp. 329–337.
[13] Johnson, R., and Wichern, D., 1998, *Applied Multivariate Statistical Analysis*, Prentice-Hall, New Jersey.
[14] Jackson, J. E., 1980, ''Principal Components and Factor Analysis: Part I-Principal Components,'' J. Quality Technol., **12**, No. 4, pp. 201–213.
[15] MacGregor, J. F., Jaeckle, C., Kiparissides, C., and Koutoudi, M., 1994, ''Process Monitoring and Diagnosis by Multiblock PLS Methods,'' AIChE J. Proc. Syst. Eng., **40**, No. 5, pp. 826–838.
[16] Kourti, T., and MacGregor, J., 1996, ''Multivariate SPC Methods for Process and Product Monitoring,'' J. Quality Technol., **28**, pp. 409–428.
[17] Ceglarek, D., Shi, J., and Wu, S. M., 1994, ''A Knowledge-Based Diagnosis Approach for the Launch of the Auto-Body Assembly Process,'' ASME Trans. J. Eng. Ind., **116**, pp. 491–499.
[18] Ceglarek, D., and Shi, J., 1996, ''Fixture Failure Diagnosis for Auto-Body Assembly Using Pattern Recognition,'' ASME Trans. J. Eng. Ind., **118**, pp. 55–65.
[19] Fukunage, K., 1990, *Introduction to Statistical Pattern Recognition*, Academic Press, Inc.