The Classification Model for Hard Disk Drive Functional Tests under Sparse Data Conditions

S. Pattanapairoj and D. Chetchotsak

***Abstract*—** This paper proposed classification models that would be used as a proxy for hard disk drive (HDD) functional test equitant which required approximately more than two weeks to perform the HDD status classification in either "Pass" or "Fail". These models were constructed by using committee network which consisted of a number of single neural networks. This paper also included the method to solve the problem of sparseness data in failed part, which was called "enforce learning method". Our results reveal that the constructed classification models with the proposed method could perform well in the sparse data conditions and thus the models, which used a few seconds for HDD classification, could be used to substitute the HDD functional tests.

***Keywords*—**Sparse data, Classifications, Committee network

I. Introduction

HDD manufacturing processes in general consist of more than 5,000 processes. They can be grouped into merely six main processes as shown in Fig. 1, starting from the wafer processes (not fabricated in Thailand), following by bar processes, the slider processes, head gimbals assembly (HGA), head stack assembly (HSA) and hard disk drive assembly (HDA). Before reaching the hands of the customers, all HDD products must go through to the functional test process which normally takes more than two weeks for each particular drive. The functional test is done so as to classify the products into either "Pass" or "Fail". If the test time of such a test procedure can be shorten, production cycle time can be significantly reduced.

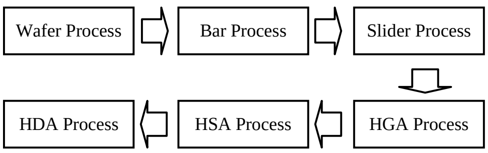

Fig. 1 Hard disk drive manufacturing processes

This can be accomplished by means of an intelligent system like artificial neural networks which can be trained to replicate the functional test system, where the test procedure can be done within less than a second. Hence, if the neural network classification model is used, substantial productivity could be potentially improved.

The HDD manufacturing processes typically have a number of process attributes or variables, which affect the quality of HDDs. Such a large number of process variables will cause the classification models to consume a considerable processing time. As a result, the feature selection algorithm is used to selected subsets of the process variables and attributes which have a significant impact on the HDD product quality.

Due to the fact that, the HDD manufacturing yields normally are above 80-90%, the number of sample data of defective product is essentially small. Therefore the sparseness of data with respect on the defective HDDs commonly occurs. In this regard, constructing the classification model to achieve good performance on defective HDDs' prediction is very difficult. This is because the amount of sample data on failed parts is not enough for models construction. Hence, the challenge of this study is how to construct the classification model under sparse data conditions.

There have been a number of published papers attempting to apply the committee network in many fields of classification problems. For example, [13] used committee network trained on different data set in medical field, [10] and [15] proposed the committee network in the image classification field while [1] used the neural network and committee network to classify type of yogurts. In addition, [4], [8] and [9] applied the neural network in the HDD manufacturing which has a much larger number of input variables than previous proposed fields. However, they were not able to solve the problem of sparseness on the defective samples. In this regard, [3] and [14] were reviewed. They proposed the method to improve classification performance under sparse data condition. These papers can convince that the committee network trained on difference input data sets can help to solve the classification problem under sparse data.

The focus of this paper therefore is to propose a classification model to be used as a proxy for the HDD functional test equipment. The classification model is based on the committee network and the proposed enforced learning method. The following section describes background and these models construction including the committee network and neural network.

S. Pattanapairoj is with the Department of Industrial Engineering, Faculty of Engineering, Khon Kaen University, 40002, Thailand; e-mail: may_ie32@ hotmail.com.

D. Chetchotsak is with the Department of Industrial Engineering, Faculty of Engineering, Khon Kaen University, 40002, Thailand; e-mail: cdanai@ kku.ac.th

II. Background

Here, the classification models were constructed using a number of single neural networks as a classifier. The HDD manufacturing process variables were used as the input for neural as show in diagram below.

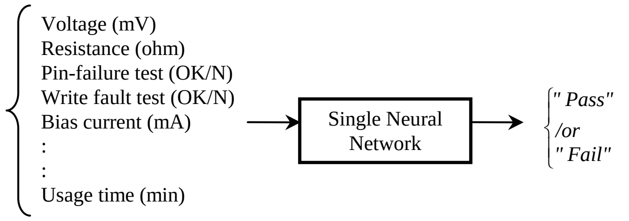

Fig. 2 Neural network diagram

A. Variables Selection Method: Feature Selection Algorithm

Feature selection in this paper was used to purify the influencing variables which affected the HDD quality by eliminating variables with unuseful information. This reduced a number of input variables for neural network and the processing time for neural network training. The algorithm involves three steps that were screening, ranking, and selecting.

In the screening step, the variables which were not useful for prediction such as all missing value, inputs that had too many values or had all constant value and so on would be detected. In the ranking step, one input at a time would be considered and evaluated how well it could predict the target in ranking using statistical test as p-values of Pearson's chi-square and p-values of likelihood ratio chi-square etc. Next, the algorithm will select the input variables based on the ranking from previous strep. For more details of the feature selection algorithm were provided in [11].

B. Classifier: The Neural Network

The neural networks were used as a tool to identify HDD functional status into either "Pass" or "Fail" since they have received good attention as a learning algorithm for several decades. In this paper a multilayer perceptron trained by the backpropagation algorithm [6] was used as a classifier. This type of neural networks has three layers: input, hidden, and output layers. The input layer received the data through input nodes and did not perform any computation. The hidden layer located between the input and output layers could have one or more intermediate layer. The nodes in this layer were fully connected to other nodes in the input and output layers. The corresponding weights connected among each node are used to determine the relationship between the input and output (HDD functional status) nodes. The training process began and carried on so as to change the weight values until the error is minimized. More details discussion was provided in [7].

C. Classification Model: Committee Network

Committee networks were constructed from many single neural networks as the concept of "many heads are better than one". The several single neural networks were used as the committee members of the committee network. Each of individual committee members should be encouraged to learn different data set so as to have different expertise. The most well known algorithm to decorrelate the training data was the bootstrap algorithm which was the general re-sampling procedure. In this regard, if the members make a mistake in training process, they would make a mistake in different ways. When the training process done, the outputs from each neural network would be combined through a fusion rule then produced the committee output. The simple methods of combining the individual neural network was the majority voting. As this method, the mistake might cancel out.

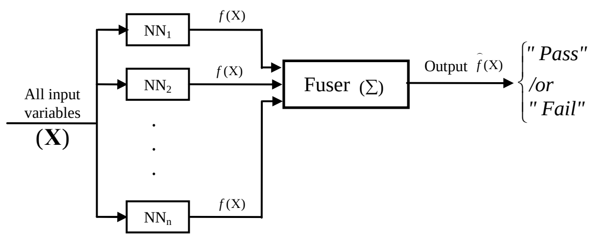

Fig. 3 Committee network diagram

Fig. 3 shows how the diagram of committee network where: $\mathbf{X}$ was a set of input data, $f(\mathrm{X})$ was the output from each neural network, and $\hat{f}(\mathrm{X})$ was the finally output into either "Pass" or "Fail" after fusion.

D. The Enforce Learning Method

This method was inspired from the concept of learning mathematics [5]. The concept of this method was that if one can not solve a problem, he or she should pay more attention in that problem field. One could learn how to solve that problem correctly then keeps practicing in that problem several times. For this paper, the neural networks could not learn well in the defective parts because the number of data sample which were collected from the defective HDDs was very small when comparing with data were collected from the good HDDs. In this regard, the neural network would most likely misclassify the instances of defective HDDs. As this result, we decided to duplicate the data which were collected from defective HDDs then add all duplicated data into the training data set which was used as input for neural networks. This approach could help the neural networks classifies the defective better than used default input data. The figure bellow shown the diagram of the enforce learning method for single neural network.

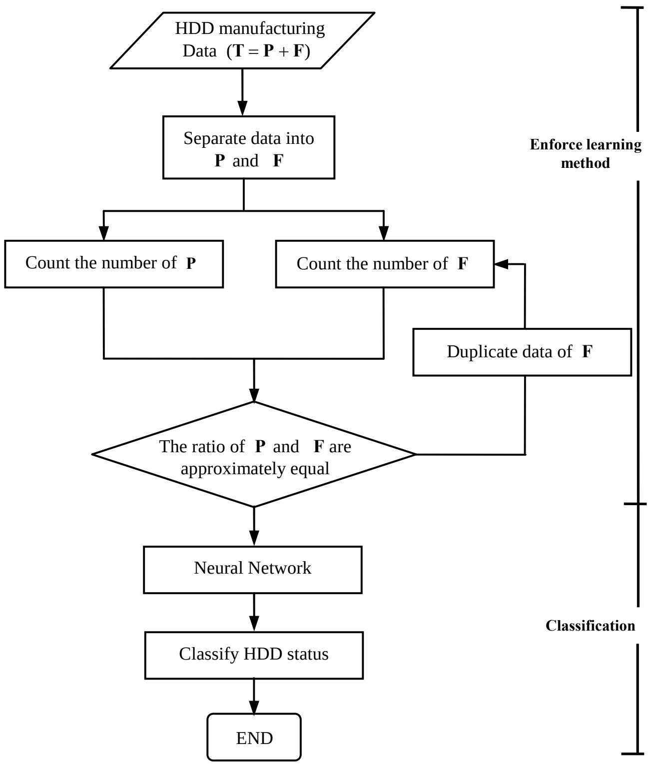

Fig. 4 Enforce learning with neural network diagram

E. Bootstrap Method

The bootstrap method was the method of re-sampling the given data for estimating the variability of statistical quantities and for setting confidence regions [12]. The name "bootstrap" referred to the analogy with pulling oneself up by one's own bootstraps. Given observations, $X_1, X_2, ..., X_n$, bootstrap samples were drawn with replacement from $X_1, X_2, ..., X_n$, putting equal probability mass $\frac{1}{n}$ at each X_i. For example, with sample size $n=6$ and distinct observations $X_1, X_2, X_3, X_4, X_5, X_6$ one might obtain $X_2, X_1, X_3, X_4, X_4, X_6$ as bootstrap re-sample. In fact, there could be 462 distinct bootstrap samples in this case. For more details of bootstrap algorithm, the readers were referred to [2]. In this matter, the bootstrap algorithm for construction of training data is given in Table I [14].

TABLE I
THE BOOTSTRAP ALGORITHM FOR CONSTRUCTION OF TRAINING DATA

i.	Let $\hat{F}$ be the empirical probability distribution where $\mathbf{T}$ with n observation is drawn.
ii.	Let $t_1, t_2, ..., t_n$ be a collection of training set $\mathbf{T}$.
iii.	Specify the number of bootstrap samples b to produce b training data set These set were referred to as $\mathbf{T}^{*1}, \mathbf{T}^{*2}, ..., \mathbf{T}^{*b}$.
iv.	Randomly choose t_i from $\mathbf{T}$ for $i = 1,2,...,n$ with replacement and equal probability mass $\frac{1}{n}$ to construct each $\mathbf{T}^{*b}$.
v.	Repeat Step iv. b times to construct $\mathbf{T}^{*1}, \mathbf{T}^{*2}, ..., \mathbf{T}^{*b}$.
vi.	Train neural network1, neural network2,…, neural network b using $\mathbf{T}^{*1}, \mathbf{T}^{*2}, ..., \mathbf{T}^{*b}$.

III. METHOD

In this paper, the classification models constructed from committee network and the enforce learning method was proposed to classify the HDD status as the functional machine testing. The neural network was used as a classifier while the enforce learning method and bootstrap algorithm were used to modify the input data for each neural network. The Table II shows the enforce learning method with the committee network [5].

TABLE II
THE ENFORCE LEARNING METHOD FOR THE COMMITTEE NETWORK

i.	Let $\mathbf{T}$ was the training set with n records. And let $\mathbf{P}$ was the subset of $\mathbf{T}$ that contains only good HDD data while let $\mathbf{F}$ was the subset of $\mathbf{T}$ that contained only defective HDD data.
ii.	Duplicating $\mathbf{F}$ m times and added it to $\mathbf{T}$, where m was chosen such that the ratios of defective HDD samples and good samples were approximately equal.
iii.	Suppose $m_1, m_2, ..., m_p$ were chosen. Then p modified training sets can be built. These set were referred to as $\mathbf{T}^{+m_1}, \mathbf{T}^{+m_2}, ..., \mathbf{T}^{+m_p}$.
iv.	Train neural networks using $\mathbf{T}^{+m_i}$, $i = 1,2,..., p$. The output of the neural networks should be as $f_i[\mathbf{T}^{+m_i}, \mathbf{x}]$, $i = 1,2,..., p$. In this case the output was interpreted in the range of [0.000, 1.000].
v.	In the fusion rule, the output from each neural network was written as $\hat{f}_i[\mathbf{T}^{+m_i}, \mathbf{x}] = \begin{cases} \text{"Pass", if } \hat{f}_i[\mathbf{T}^{+m_i}, \mathbf{x}] \le 0.300 \\ \text{"Fail", if } \hat{f}_i[\mathbf{T}^{+m_i}, \mathbf{x}] \ge 0.700 \\ \text{"UID", otherwise.} \end{cases}$
vi.	From a committee network using a certain fusion rule. Let W and were the number of neural networks that predicted "Pass" and "Fail" respectively, for a particular HDD.
vii.	To form a committee machine using the threshold, the output should be $\hat{f}[\mathbf{T}, \mathbf{x}] = \begin{cases} \text{"Pass", if } W \ge TH \\ \text{"Fail", if } Z \ge TH \\ \text{"UID", otherwise} \end{cases}$, Where "UID" stranded for unidentified.

The status of defective HDD was coded as 0.000 while the status of good HDD was coded as 1.000.

World Academy of Science, Engineering and Technology
International Journal of Industrial and Manufacturing Engineering
Vol:4, No:10, 2010

TABLE III
THE MODELS CONSTRUCTION

Model	Input variable selection	Input data modifier	Input data set variegation	Classifier	Number of neural networks	Fusion schemes
CM_1	Feature Selection method	-	-	Single Neural network	1	-
CM_2		Enforce learning method with 7 data sets, there are $\mathbf{T}^{original}, \mathbf{T}^{+2}, \mathbf{T}^{+3}, \mathbf{T}^{+6}, \mathbf{T}^{+10}, \mathbf{T}^{+15}, \mathbf{T}^{+30}$	-	Committee network with 21 different architectures	21	Threshold $(TH) = 15$ or 2in 3
CM_3			Bootstrap $\mathbf{F}$ of each data set 3 times	Committee network with 21 different data sets	21	Threshold $(TH) = 15$ or 2in 3
CM_4			Bootstrap $\mathbf{F}$ of each data set 5 times	Committee network with 35 different data sets	35	Threshold $(TH) = 24$ or 2in 3
CM_5			Bootstrap $\mathbf{F}$ of each data set 7 times	Committee network with 49 different data sets	49	Threshold $(TH) = 34$ or 2in 3
CM_6			Bootstrap each data set 3 times	Committee network with 21 different data sets	21	Threshold $(TH) = 15$ or 2in 3
CM_7			Bootstrap each data set 5 times	Committee network with 35 different data sets	35	Threshold $(TH) = 24$ or 2in 3
CM_8			Bootstrap each data set 7 times	Committee network with 49 different data sets	49	Threshold $(TH) = 34$ or 2in 3

$\mathbf{T}^{original}$ is the raw input data set with out adjustment.

Several neural networks were encouraged to learn different modified data sets. As a result, they would have different expertise so they could help one another to classify the HDD functional status. In this study, the data used to construct the classification models was obtained from a local HDD manufacturer in Thailand. All available data to construct the models consisted of 2,193 records with 162 input variables and one target which is HDD functional status. They were collected from the operation test during April 30 and May 8, 2010. From 2,193 data records, the ratio of defective HDD samples was only 15.37% or 337 records while the good HDD samples were 84.63% or 1,856 records. All data set was randomly divided into two parts. The first 70% was used for model training and another 30%, called the test set, was used for the first model evaluating. In addition, another unseen data set, 1,318 records, collected from May 9, 2010 was prepared to use to confirm the model performance in the final evaluating. In this experiment, to avoid dependency from sampling, the trial was replicated 20 times.

The Table III described eight classification models constructed. They were referred to as CM_1-CM_8. The first model (CM_1) was constructed from a single neural network which trained by the backpropagation algorithm. The architecture of this neural network in this model was 5,000 learning cycles and 40 hidden units. The model CM_2 was constructed using committee networks which was built from seven different data sets which generated by using enforce learning method. Each data set was trained in three levels of hidden units that are 20, 30, and 40 hidden units. Thus there would be 21 neural networks for committee formation. Model CM_3- CM_8 were trained in only one level of architecture, 5,000 learning cycles and 40 hidden units with the bootstrap algorithm. In model CM_3, CM_4 and CM_5, only the defective HDD data set ($\mathbf{F}$) of each input data set ($\mathbf{T}^{+m_i}$) were bootstrapped 3 times, 5 times and 7 times while all input data set ($\mathbf{T}^{+m_i}$) were bootstrapped 3 times, 5 times and 7 in the model CM_6, CM_7 and CM_8, they were 21, 35 and 49 different data sets. All classification models were evaluated using two measures as equations (1) and (2).

$$\text{percentage of positive accuracy} = \frac{\text{number of correct predictions} \mid \text{HDDs are "Pass"}}{\text{number of total samples} \mid \text{HDDs are "Pass"}} \times 100\% \qquad (1)$$

$$\text{percentage of negative accuracy} = \frac{\text{number of correct predictions} \mid \text{HDDs are "Fail"}}{\text{number of total samples} \mid \text{HDDs are "Fail"}} \times 100\% \qquad (2)$$

IV. THE RESULTS

The classification performances of the models were shown in Fig. 5. The results revealed that model CM_1 constructed from using the original data without adjustment and the single neural network gave the best performance with respect to the positive accuracy while yielded the worst performance with respect to negative accuracy. This means that the models predicted the HDDs' statuses as “Pass” only, no matter what their actual statuses were. Therefore the number of sample size in defective samples was very small when comparing with the number of good HDDs data samples, so overfitting occurred. In this regard, the model CM_1 was not suitable to be used as the function HDD tester.

For the other classification models built from the enforce learning method and committee network, CM_2-CM_8, even though they lost some performance for positive accuracy but they presented the increasing of negative accuracy as Fig. 5.

the

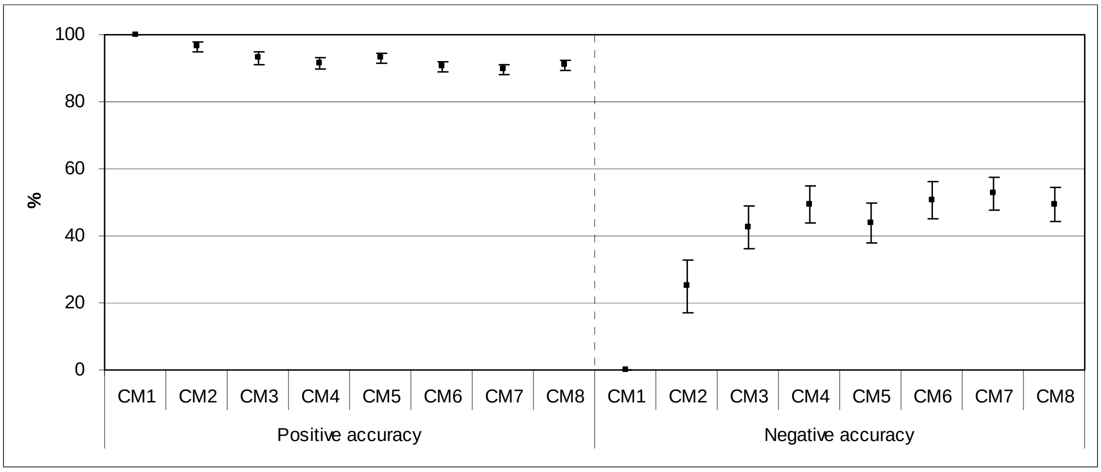

Fig. 5 The classification performance (Test set)

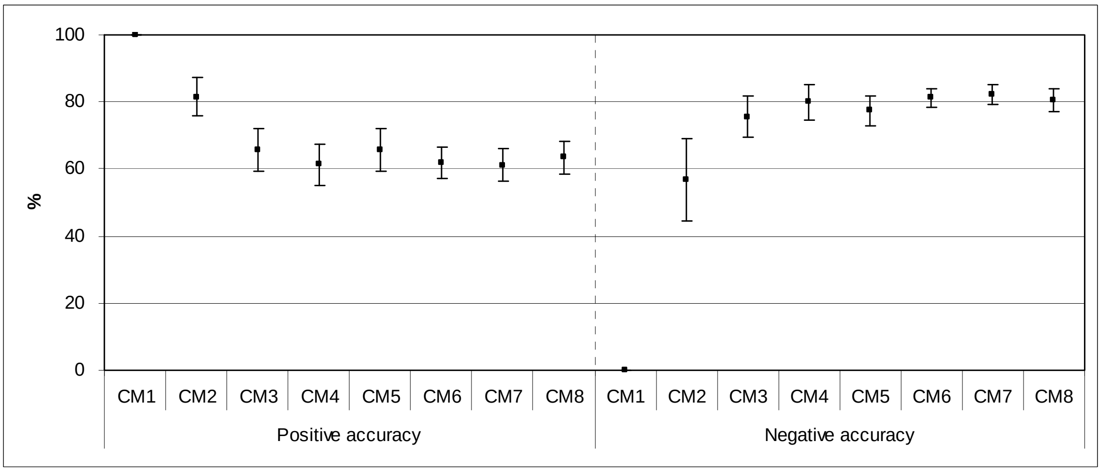

Fig. 6 The classification performance (Unseen data set)

The Fig. 6 showed the results of eight classification model using unseen data set in the final evaluating. The results from using unseen data set gave the likely performance as the result from using the test set. With out enforce learning method; the model could not classify the data in the defective HDDs products. The best negative accuracy was in the model CM_7, which constructed from using different data under bootstrapped all seven data sets 5 times. This model gave the improvement of negative accuracy from 0% - 82.07%. However, when the positive accuracy was considered, the performance to classify the good HDDs data was decreased to 61.15% using proposed method. As this result, the proposed models still need improvement before implementing on the functional operation test.

V. Conclusion

This paper proposed classification models to be used as a proxy for HDD functional test equitant. Due to the fact that most real productivity in HDD manufacturer were allowed only above 80-90%, this made the number of defective HDD data much smaller than good HDD data. Then the sparseness of data with respect to defective products commonly occurred. The enforce learning method was proposed to construct the classification models based on committee networks to deal with this sparse data problem. The single neural network in each committee network was encouraged to have different expertise through to use of different training data set in addition to the used of bootstrap algorithm. Experimental results demonstrated the ability of the constructed models to classify HDDs' status. The proposed enforce learning method with the committee network could help improve negative classification performance compare to the model constructed from using the original training data set with single neural network even though they lost some positive performance. Once this performance can be improved, the classification model will help the HDD makers to reduce the functional testing times of each drive performing from more than two weeks to only a few seconds.

ACKNOWLEDGMENT

The authors would like to express their gratitude for God's grace and His everlasting love. This work has been fully supported by National Electronics and Computer Technology Center, National Science and Technology Development Agency and Industry/University Cooperative Research Center (I/U CRC) in HDD Component, the Faculty of Engineering, Khon Kaen University, Thailand.

REFERENCES

[1] A.G. da Cruz, E.H.M. Walter, R.S. Cadena, J.A.F. Faria, H.M.A. Bolini and A.M. Frattini Fileti, "Monitoring the authenticity of low-fat yogurts by an artificial neural network," *Journal of Dairy Science*., Vol. 92, pp. 4797-4804, 10 Oct. 2009.

[2] B. Efron, "Bootstrap methods: another look at the jackknife," *Ann. Stat.*, vol. 7, No.1, pp. 1–26, 1979.

[3] D. Chetchotsak, and J. M. Twomey, "Combining neural networks for function approximation under conditions of sparse data: The biased regression approach," *International Journal of general Systems*., Vol. 36, No.4, pp.479-499, Aug. 2007.

[4] D. Chetchotsak, and W. Kanarkard, "Classification model to detect failed HDD components," *Intelligent Engineering System Thorough Artificail Neural Networks.*, Vol. 18, pp. 697-703, St. Louis, MO, USA, Nov. 9-12, 2008.

[5] D. Chetchotsak, and S. Pattanapairoj, "Committee network model for HDD functional test," *ANNIE-2010.*, St. Louis, MO, USA, Nov. 1-3, 2010.

[6] D. E. Rumelhart, G.E. Hilton, and R.J. Williums, "Learning representations of back-propagation errors," Nature, London, 1986.

[7] S. Haykin, "Neural networks: A comprehensive foundation," Mc Milan College Publishing, New York, 1994.

[8] S. Pattanapairoj, and D. Chetchotsak, "Quality monitoring model for failed parts detection in HDD manufacturing processes," *Data Storage Technology conference.*, Bangkok , Thailand, May. 13-15, 2009.

[9] S. Pattanapairoj, and D. Chetchotsak, "Data mining framework for HDD yield improvement: The neural networks and association analysis approach," *TISD-2010.*, Nongkhai, Thailand, Mar. 4-6, 2010.

[10] M. Bacauskiene, A. Verikas, A.Gelzinis, and D. Valincius, "A feature selection technigue for generation of classification committee and its application to categorization of laryngeal images," *Pattern Recognition.*, vol. 42, pp. 645-654, Aug. 2008.

[11] M. Dash, and H. Liu, "Feature selection for classification," *Intelligent Data Analysis: An Int' 1J.*, Vol. 1, No. 3 pp. 131-156, 1997.

[12] M. Zribi, "Non-parametric and unsupervised Bayesian classification with bootstrap sampling," *Image and Vision Computing*, Vol. 22, pp. 1-8, June. 2003.

[13] R. E. Abdel-Aal, "Improved classification of medical data using abductive network committees trained on different feature subsets," *Computer Methods and Programs in Biomedicine,* Vol. 80, pp. 141-153, Aug. 2005.

[14] R. Nanthapodej, and D. Chetchotsak, "Classification performance of committee networks improvement under sparse data condition," *KKU Res J (GS)*, Vol. 9, No. 2, pp. 65-76, Apr.-June. 2009

[15] R. P. Pein, and J. Lu, "Multi-feature query language for image classification," *Procedia Computer Science,* Vol. 1, pp. 2533-2541, 2010.