

Sound Based Machine Fault Diagnosis System Using Pattern Recognition Techniques

Caleb Vununu[†], Kwang-Seok Moon^{**}, Suk-Hwan Lee^{***}, Ki-Ryong Kwon^{****}

ABSTRACT

Machine fault diagnosis recovers all the studies that aim to detect automatically faults or damages on machines. Generally, it is very difficult to diagnose a machine fault by conventional methods based on mathematical models because of the complexity of the real world systems and the obvious existence of nonlinear factors. This study develops an automatic machine fault diagnosis system that uses pattern recognition techniques such as principal component analysis (PCA) and artificial neural networks (ANN). The sounds emitted by the operating machine, a drill in this case, are obtained and analyzed for the different operating conditions. The specific machine conditions considered in this research are the undamaged drill and the defected drill with wear. Principal component analysis is first used to reduce the dimensionality of the original sound data. The first principal components are then used as the inputs of a neural network based classifier to separate normal and defected drill sound data. The results show that the proposed PCA-ANN method can be used for the sounds based automated diagnosis system.

Key words: Pattern Recognition, Machine Learning, Machine Fault Diagnosis, Sound Processing, Principal Component Analysis, Artificial Neural Network

1. INTRODUCTION

Among the many problems in industries, the most important factor is for each machine systems to work in a normal state. In order to maintain a normal condition of a machine system, fault prediction and diagnosis systems are necessary. When failures occur, the failures should be detected as soon as possible, because if these machines run continuously under abnormal conditions, it may result in great damage and even loss of human lives. In the past, many studies have been based on the traditional methods of establishing a mathematical model, analyzing a variety of parameters and then

judging the operating conditions of the machine [1]. However, the complexity of the real world machine system and the obvious existence of nonlinear factors such as unwanted noises that can corrupt the used signal make the mathematical models based approaches very difficult to handle and not efficient in terms of accuracy. Since three decades now, machine learning techniques have been widely used in machine fault diagnosis [2]. Sounds and vibrations data have shown effectiveness in damage detection systems, especially because of their capability of carrying machine operating conditions characteristics. Many studies have demonstrated that fault diagnosis can be effectively made by an-

* Corresponding Author : Ki-Ryong Kwon, Address: (608-737) 599-1, 45 Yongso-ro, Namgu, Busan, Korea, TEL : +82-51-629-6257, FAX : +82-51-629-6230, E-mail : krkwon@pknu.ac.kr

Receipt date : Jan. 3, 2017, Approval date : Jan. 17, 2017

[†] Dept. of IT Convergence and Application Engineering, Pukyong National University
(E-mail : exen.xmen@gmail.com)

^{**} Dept. of Electronics Engineering, Pukyong National University (E-mail : ksmoon@pknu.ac.kr)

^{***} Dept. of Information Security, Tongmyong University (E-mail : skylee@tu.ac.kr)

^{****} Dept. of IT Convergence and Application Engineering, Pukyong National University

* This work was supported by a Research Grant of Pukyong National University(2016 year)

alyzing the sounds or/and the vibrations emitted by the machines. Vibrations and sounds data are in most of studies processed with the same methods for detecting faults.

Tandon and Nakra [3] reported a detailed review of the different vibration and acoustic methods, such as sound measurements, vibration measurements, the shock pulse method and the acoustic emission technique, for fault diagnosis in rolling bearings. By just hearing the sound of a machine during its running, an experienced operator can even identify and locate some defined faults in the machine. This shows that the sound signals are strong indicators of the condition of the machine. Compared with vibrations, sounds can be collected easily by any operator who wants to build a diagnosis system, while the sensors that can capture vibrations of the machines are, in practice, difficult to find. That makes the sound-based analysis cheaper and simpler to set up while the vibration-based analysis can be expensive and a more complicated task. Processing the sounds can help to reliably identify the machine faults [4].

Principal component analysis (PCA) [5] is widely used for its capacity of de-correlating data in signal processing. It is also used as a feature extractor or for the process of dimensionality reduction which is very practical for pattern recognition problems [6]. Artificial Neural Networks (ANN) have shown an impressive learning and memory capability. They have been widely used for automatic detection of faults in different ways [7] and in other pattern recognition problems [8]. ANN have the ability to reproduce arbitrary nonlinear functions and are very suitable for complex pattern recognition tasks [9]. Which means that it is proven that ANN can separate even extremely complex data. That is the main reason why we have chosen to apply them as a classifier for our fault diagnosis system. In this research, sound data are collected from undamaged and defected drills. We use PCA to time series data as our dimensionality

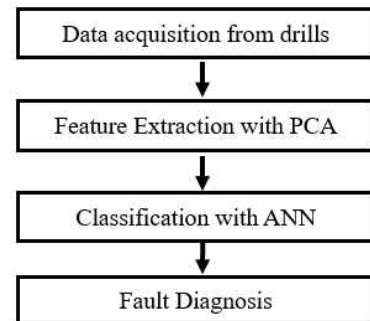


Fig. 1. Summary of the proposed PCA-ANN method.

reduction method, then the first five principal components are fed to the neural network in order to perform the classification task. The PCA-ANN proposed method in this paper is summarized in Fig. 1.

2. PATTERN RECOGNITION

Machine learning based approaches mainly consist of two steps: the feature extraction process or and the classification task (Fig. 2). Collected data from real world always contain unwanted elements that can corrupt or disturb the analysis. Furthermore, collected data are usually large vector that contain many sample points. Fig. 3 shows an example of sound data in time domain collected during the experiments. Even with a short time signal (0.05s), the number of sample is big. The larger is the size of the data to be processed, the more complicated can be the processing analysis. These two difficulties, combined together, increase the complexity of the classification task for any kind of machine learning algorithms. That is why it is usually admitted to avoid using the raw collected data as the inputs of the classifier. Instead, we can choose a set of features that strongly char-

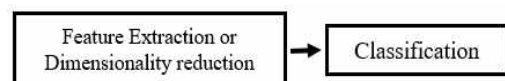


Fig. 2. The 2 steps of the machine learning based pattern recognition.

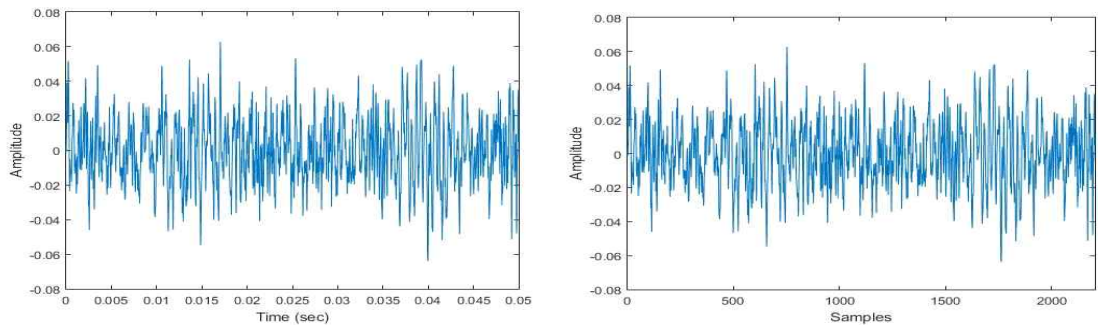


Fig. 3. An example of sound data collected during the experiments. A 50 ms signal (left) with a sampling frequency of 44K Hz gives a vector of size 2205 (right).

acterize the data and use them as the inputs of the classifier. Selecting a defined number of features has the advantage of maximizing the classification task by reducing the complexity and the size of the original data. Next, we will discuss about some of the widely used feature extraction methods.

2.1 Feature Extraction

Many methods exist for selecting good features according to a specific problem. Choi, for example, presented a feature extraction method based on discriminant analysis for face recognition [10]. We will list some that have been widely used in machine fault diagnosis literature. Many studies use statistical features for the dimensionality reduction purpose. For each data sound or vibration, some statistical numbers are computed. And then those numbers will be used as the inputs of the machine learning classifier, neural network or support vector machines. In the reference [11], the authors compute 11 statistical features: the mean, the standard error, the median, the mode, the standard deviation, the sample variance, the kurtosis, the skewness, the range, the minimum and the maximum. They have used support vector machine (SVM) as the classifier. Which means, for each data sound and vibrations, instead of processing all the time series data, the authors just select 11 numbers that will represent the data. They have gone even further in dimensionality reduction be-

cause they reduced the eleven statistical features to seven values by applying independent component analysis (ICA).

There are several kinds of statistical features. M. Subrahmanyam and C. Sujatha [7] for example, used different kinds of statistical features of acoustic emission signals for feeding the neural network.

In our case, we have chosen principal component analysis (PCA) [5] as the feature extraction process. PCA is a statistical procedure that uses an orthogonal transformation in order to convert a set of correlated data into a set of values that are linearly uncorrelated. Those values are called principal components. The number of principal components is less than or equal to the number of the original variables of the observations. PCA is defined such a way that the first principal component has the largest possible variance, which makes it to be sensitive to the variability of the data. Each succeeding component in turn must also have the highest variance possible under the constraint that it must be orthogonal to the preceding components. Figure 4 shows an example of PCA of a given 2-dimensional dataset. The obtained vectors (see Fig. 4 for illustration) will be used as a new basis on which we will project the original data to obtain the new data on the PCA space (see Fig. 5). For example, in figure 4, all the data points from the original space must be projected onto the

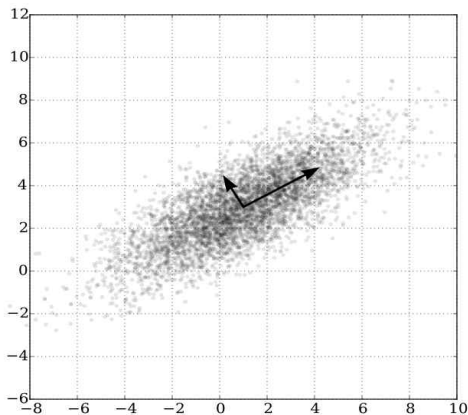


Fig. 4. PCA of a given 2-dimensional dataset. The first principal component follows the direction of the highest variance. The second, must also find the highest variance while being perpendicular to the first principal component.

new orthogonal basis set. That is the part of dimensionality reduction capability of the PCA.

Principal Component Analysis is mostly used as a tool in explanatory data analysis. The very first principal components contain almost all the significant characteristics of the original dataset. Let's say we have a dataset containing vectors of size m . As said before, the total number of the found principal components is less than or equal to the vector's size (m). Before projecting the data onto the new basis set, we may choose, say, the n first principal components. The number n will always

be less than m . And because the very first principal components contain almost all the characteristics of the original data, we may just have to choose a few number of them in order to perform the projection. After the projection, all the original vector data that we have will be of size n . Which means that we have just reduced the dimension of our original data vector from m to n .

Principal component analysis de-correlate the data. That means, it forces the uncorrelated data in the dataset to be separated. Because of that, after the projection, the set of data that are highly correlated between them will be put together in one cluster. Data that are not correlated at all between them will be separated. In Fig. 5, after the projection of all the original data onto the component space, the data points represented by an "x-sign" are clustered together because they are highly correlated between them, and in the same manner, the circle data and the square data points are separated because they are not correlated. That is, the PCA is a very good tool to not only separate data between them, but also to investigate if the data that we have (in our case the normal and abnormal sounds) are different between them. We have performed PCA on our data and chosen the first 5 principal components for the projection. The results are discussed in section 4.

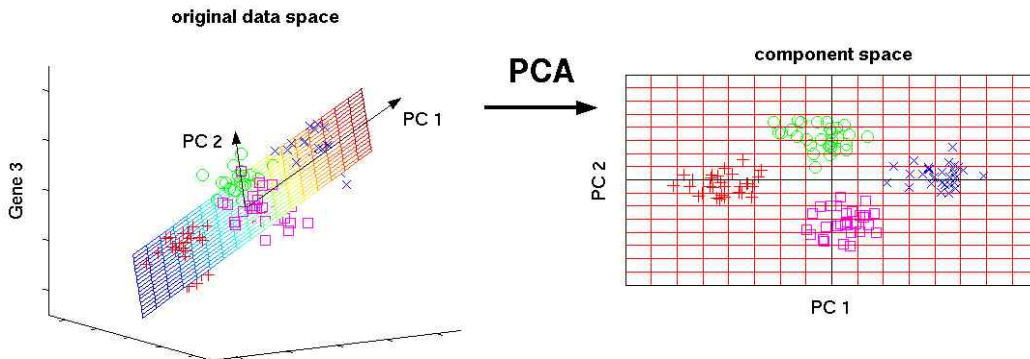


Fig. 5. Projection of the original data onto the principal component basis, the component space. The original data have 3 dimensions, Just the first 2 principal components (PC 1 and PC 2) were chosen in this example. After the projection, all the data change their size. The dimensionality reduction of the data is from 3 to 2.

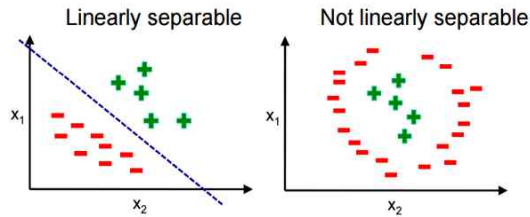


Fig. 6. Linearly separable and not linearly separable dataset examples.

2.2 Classification

After the feature extraction process, the final step of the machine learning based pattern recognition method is the classification. Classification is the process of separating the dataset into different classes, or different instances. If the data are linearly separable (see Fig. 6), a simple line that can separate the instances can be easily found. In this case, we call the separator line a linear classifier. But, when the dataset is not linearly separable, the linear classifier is obviously useless. In this case, we have to find a nonlinear classifier that can separate the dataset. Artificial Neural Networks have the ability to reproduce arbitrary nonlinear functions, which means that they can separate any nonlinear dataset [9]. That is the main reason why we have chosen to apply them as a classifier for our fault diagnosis system.

2.3 Artificial Neural Network

Artificial neural network (ANN) is an interconnected group of artificial neurons. These neurons use a mathematical or computational model for information processing. ANN is an adaptive system that changes its structure based on information that flows through the network [12]. The neural network acquires the ability to generalize based on the training data and, if the training data contains all the characteristics, all the meaningful information possible of the processed sounds or vibrations, the network can predict outcomes for new, previously unseen data sound or vibrations. One of the commonly used structures of ANN is

the multi-layer perceptron. Using supervised learning, the sample $\{X_k\}$ is fed to the network and produces an output $\{y\}$.

The input pattern $\{X_k\}$ is then propagated through the network in the following way:

$$y_i = f \left[\sum_{j=1}^M w_{ij}^{(2)} \times f \left(\sum_{k=1}^N w_{jk}^{(1)} x_k \right) \right] \quad (1)$$

where y_i denotes the output of a given neuron i , and N the number of input neurons, while M denotes the number of hidden neurons. $w_{ij}^{(n)}$ is the weighted sum in this form: j represents the input neuron that comes to feed the neuron i , and n denotes the layer where we are ($n=1$ represents the first layer). To implement this procedure, one needs to calculate the error derivative with respect to weight in order to change the weight by an amount that is proportional to the rate at which the error changes as the weight is changed. The backpropagation algorithm [13] is used in this research paper. The activation function f is a sigmoid function (see Fig. 7) and is defined as:

$$f(x) = \tanh(x) \quad (2)$$

3. EXPERIMENTAL SETUP AND DATA ACQUISITION

We want to detect defected drill by analyzing the sounds emitted by it during its operating time. We have collected data sounds from an undamaged and a damaged drill using a microphone. Fig. 8 shows the tools used for the experiment. We put

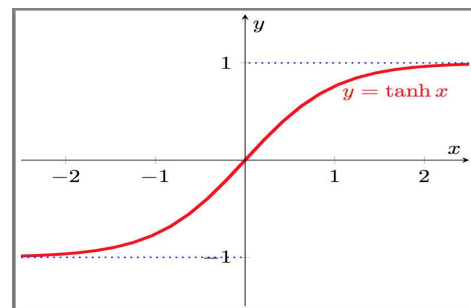


Fig. 7. Hyperbolic tangent used as the activation function for the neural network.

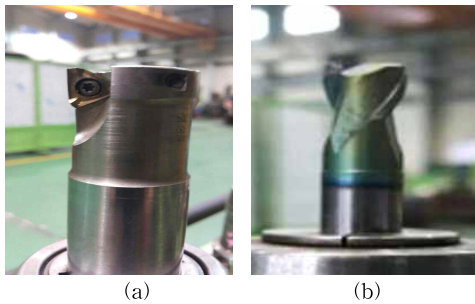


Fig. 8. (a) The tip of a “good” drill, (b) The tip of a damaged drill.

the undamaged drill in the cutting machine then we’ve recorded the sounds for one minute and half. Then, we take out the undamaged drill, replace it by the damaged one. We’ve recorded the sounds from both the undamaged drill and the damaged one. Sound being a nonstationary data, we’ve segmented the recorded sound into many parts to construct our dataset for the pattern recognition. The sampling frequency used was 44100 Hz. The signal length 50 ms (0.05 s). We’ve created a dataset of 1038 data sounds, 496 from the “good” drill and 542 data from the damaged drill. Some samples are shown in Fig. 9.

4. RESULTS AND DISCUSSION

Applying pattern recognition method on data whose size exceeds 2000 without dimensionality reduction is a bad idea. We’ve started by applying

PCA on our dataset. Analyzing the results, we found out that the data from the good drill are highly correlated between them. That is why the PCA outputs all of the normal data extremely close between them. In Fig. 10, we just see the abnormal data (data from the damaged drill). This is because there are strong variations between them. And those variations come from the fact that the sounds emitted by the damaged drill are very noisy and unstable over time. Because the variance of the abnormal dataset is very significant compared to the one of the normal dataset, the normal data are confined around the origin of the component space. The reason of this confinement around the origin is that the variance between the sounds emitted by an undamaged drill is close to zero. These sounds are not very disturbed by noises and they are found to be stable over time, which makes them highly correlated between them.

If we zoom around the origin, we obtain this Figure below (see Fig. 11). We can clearly distinguish the normal data around the origin.

We have selected the first five principal components for feeding our neural network (see Fig. 12). Which means that our network will have 5 inputs neurons. The network has 3 layers. The input layer has 5 neurons, the hidden layer has 10 neurons and the last layer has 2 neurons for our 2 instances (normal and abnormal).

The results are really convincing. The perform-

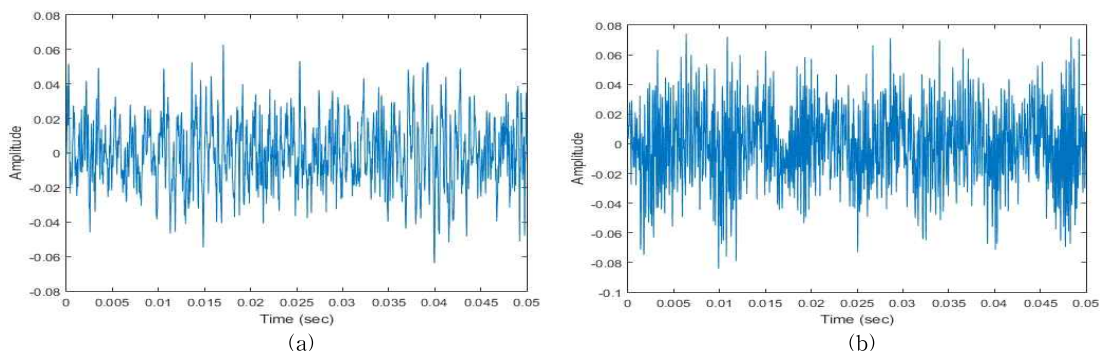


Fig. 9. Two randomly selected samples from our dataset, (a) Sound data from undamaged drill, (b) Sound data from damaged drill.

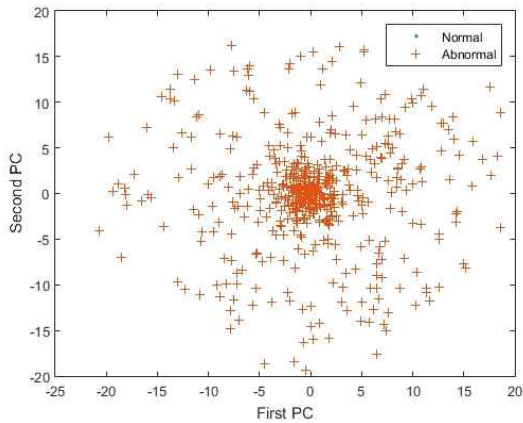


Fig. 10. Projection of the original data onto the principal component space. We can see that only the abnormal data are visible.

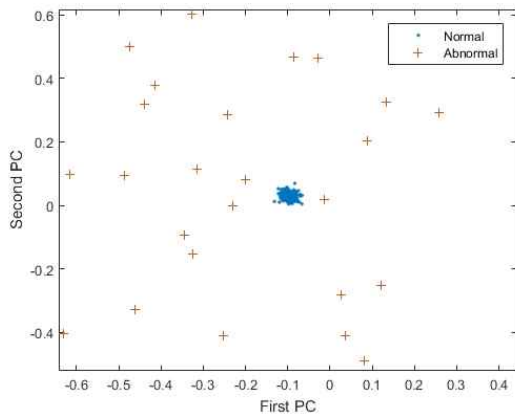


Fig. 11. By zooming Figure 10 around the origin, we see the normal data right in the middle.

ance, computed by using the mean square error is really low. In Fig. 13, we show the performance scores over the different dataset. The best performance of the validation dataset is about and was reached at the 42th iteration. The convergence has been completed after 48 iterations.

As said before, we have collected 1038 data during the data acquisition process. 496 normal and 542 abnormal data. 70% of each dataset was used for training and the remaining 30% was used for testing.

Normal dataset: 496 data, 348 were used for training and 148 for testing. All the 148 data were

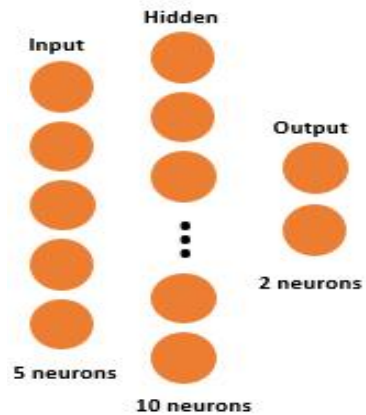


Fig. 12. The architecture of the fully connected neural network used in this paper.

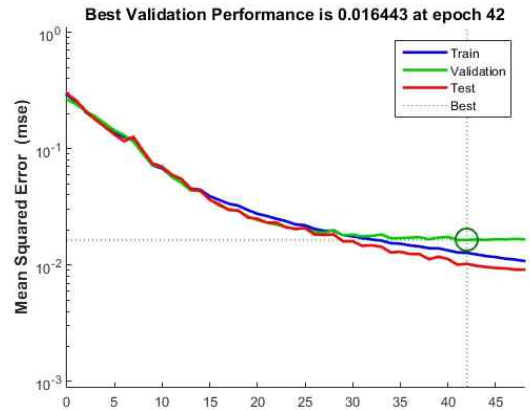


Fig. 13. Performance table.

classified as normal, as it is shown in Fig. 14. That is a 100% accuracy. As discussed before, the normal data are highly correlated between them. The total variance of the normal dataset is close to zero, as it has been demonstrated by the PCA analysis in Fig. 10 and 11. Because of a really small variance, all the normal data are confined like a point around the origin of the component space. That is the reason why the classifier does not have any problem to classify any data from this dataset.

Abnormal dataset: 542 data, 378 were used for training and 164 for testing. 160 data were classified as abnormal and 4 were classified as normal, as shown in Figure 14. That means that we have 4 cases of misclassification and an accuracy of

Output Class	Normal	148	4	Total Accuracy 98.75%
	Abnormal	0	160	
	Target Class	Normal	Abnormal	

Fig. 14. Confusion matrix of the testing data.

97.5% as it is shown in Fig. 14. As discussed before, the abnormal data are noisy, unstable over time and have strong variations between them. The total variance of the abnormal dataset is really high, as it has been demonstrated in Fig. 10. Though, by analyzing Figure 11, we see that there are some abnormal data points that are significantly close to the normal dataset. Those data can be misclassified. That is the reason why we don't have a 100% accuracy for this dataset.

The total accuracy of the classifier is 98.75% (see Fig. 14).

We have conducted a comparative study with another method developed in here [14]. As we said before, pattern recognition problems solving has two steps: feature extraction and classification. In this paper, we demonstrate the feasibility of a PCA-ANN method, having the principal component analysis as feature extractor and artificial neural network as classifier. In [14] instead, they have used statistical features (for feature extraction) and neural network (for classification). Using the vibration signals of ball bearing components, they have computed for each signal 6 statistical features that have been fed as inputs of the neural network. These statistical features are:

- a) **Range**: refers to the difference between maximum and minimum value of a signal.
- b) **Mean value**: Average value of a signal.
- c) **Standard deviation**: the measure of energy

content of the signal.

$$Std = \sqrt{\frac{1}{N-1} \sum_{i=1}^N |x_i - \mu|^2} \tag{3}$$

with x being the signal, N the total number of observations in x and μ the mean of x

- d) **Skewness**: the measure of symmetry of the signal.

$$Skewness = \frac{E(x - \mu)^3}{\sigma^3} \tag{4}$$

where μ is the mean of x , σ , the standard deviation of x , and $E(t)$ represents the expected value of the quantity t .

- e) **Kurtosis**: the measure of whether the data are peaked or flat relative to a normal distribution.

$$Kurtosis = \frac{E(x - \mu)^4}{\sigma^4} \tag{5}$$

using the mean and the standard deviation as in the previous equation.

- f) **Crest factor**: equals to the peak amplitude of a waveform divided by the root mean square.

$$Crest\ factor = \frac{\max(x)}{\sqrt{\frac{1}{N} \sum_{i=1}^N x_i^2}} \tag{6}$$

where $\max(x)$ represents the maximum value of x

For each data sound in the dataset, we have computed these six statistical features and used them as the inputs of a three layers neural network. The inputs layer has 6 neurons, the hidden layer 10 and the output layer 2. The results are shown

Output Class	Normal	148	16	Total Accuracy 95.1%
	Abnormal	0	148	
	Target Class	Normal	Abnormal	

Fig. 15. Confusion matrix of the testing data using statistical features as feature extraction,

Table 1. Results summary

	P.K. Kankar <i>et al.</i> [21]	Present work
Feature Extraction Method	6 statistical features (range, mean value, standard deviation, skewness, kurtosis and crest factor)	Principal Components Analysis (5 first principal components)
Classifier	Neural Network	Neural Network
Accuracy	95.1	98.75

in Fig. 15 and we can see that using the statistical features [14], the normal dataset completes also a 100% accuracy. Though, the accuracy decreases at 90% for the abnormal dataset because of their instability and strong variations, making the total accuracy to be about 95%. In Table 1, we made a summary of the results of the 2 methods.

6. CONCLUSION

The PCA-ANN method proposed in this study proves that it can reach outstanding performance in pattern recognition. We have used PCA on raw sound data, then we have selected the first five principal components as the inputs of our neural network based classifier. And the method shows that it can have a very good performance. However, our processed data were highly separable. We should now improve our method and try to apply it to non-separable dataset.

The PCA-ANN method can be used in any other fault diagnosis system. If it can separate normal and abnormal sounds of the drill, it can be also used for vibrations and other acoustic emission data.

REFERENCES

- [1] M.M. Polycarpou and A.T. Vemuri, "Learning Methodology for Failure Detection and Accommodation," *IEEE Control Systems*, Vol. 15, Issue 3, pp. 16-24, 1995.
- [2] N.R. Sakhthivel, V. Sugumaran, and B.B. Nair, "Application of Support Vector Machine (SVM) and Proximal Support Vector Machine (PSVM) for Fault Classification of Mono Block Centrifugal Pump," *International Journal of Data Analysis Techniques and Strategies*, Vol. 2, No. 1, pp. 38-61, 2010.
- [3] N. Tandon and B.C. Nakra, "Vibration and Acoustic Monitoring Techniques for the Detection of Defects in Rolling Element Bearings-A Review," *The Shock and Vibration Digest*, Vol. 24, No. 3, pp. 3-11, 1992.
- [4] P.W. Tse, G.H. Xu, L. Qu, and S.R. Kumara, "An Effective and Portable Electronic Stethoscope for Fault Diagnosis by Analysing Machine Running Sound Directly," *International Journal of Acoustics and Vibration*, Vol. 6, No. 1, pp. 23-31, 2001.
- [5] H. Hotelling, "Analysis of a Complex of Statistical Variables into Principal Components," *Journal of Educational Psychology*, Vol. 24, Issue 6, pp. 417-441, 1933.
- [6] V. Emamian, M. Kaveh, and A.H. Tewfik, "Robust Clustering of Acoustic Emission Signals Using the Kohonen Network," *Proceedings of the 2000 IEEE International Conference on Acoustics, Speech, and Signal Processing*, pp. 3891-3894, 2000.
- [7] M. Subrahmanyam and C. Sujatha, "Using Neural Networks for the Diagnosis of Localized Defects in Ball Bearings," *Tribology International*, Vol. 30, No. 10, pp. 739-752, 1997.
- [8] T. Villmann, "Neural Network Approaches in Medicine-A Review of Actual Developments," *Proceedings of the European Sym-*

posium on Artificial Neural Networks, pp. 165–176, 2000.

- [9] R.O. Duda, P.E. Hart, and D.G. Stork, *Pattern Classification*, Wiley, New York, 2001.
- [10] Sang-Il Choi, “Construction of Composite Feature Vector Based on Discriminant Analysis for Face Recognition,” *Journal of Korea Multimedia Society*, Vol. 18, Issue 7, pp. 834–842, 2015.
- [11] M. Saimurugan and K.I. Ramachandran, “A Comparative Study of Sound and Vibrations Signals in Detection of Rotating Machine Faults Using Support Vector Machine and Independent Component Analysis,” *International Journal of Data Analysis Techniques and Strategies*, Vol. 6, No. 2, pp. 188–204, 2014.
- [12] M.J. Zurada, *Introduction to Artificial Neural Systems*, Jaico Publishing House, Delhi, 1999.
- [13] D.E. Rumelhart, G.E. Hinton, and R.J. Williams, “Learning Representations by Back-propagating Errors,” *Nature*, Vol. 323, No. 9, pp. 533–536, 1986.
- [14] P.K. Kankar, Satish C. Sharma, and S.P. Harsha, “Fault Diagnosis of Ball Bearings Using Machine Learning Methods,” *Expert Systems with Applications*, Vol. 38, Issue 3, pp. 1876–1886, 2011.



Caleb Vununu

He received his B.S. degree in Computer Science in Youngsan University, Republic of Korea in 2015. Since September 2015, he's a Master degree student in the department of IT Convergence and Application Engineering in

Pukyong National University. His research interests include Signal Processing and Machine Learning.



Kwang-Seok Moon

He received the B.S., or Studies M.S., and Ph.D. degrees in Electronics Engineering in Kyungpook National University, Korea in 1979, 1981, and 1989 respectively. He is currently a professor in department of

Electronic engineering at Pukyong National University. His research interests include digital image processing, video watermarking, and multimedia communication.



Suk-Hwan Lee

He received a B.S., a M.S., and a Ph. D. degrees in Electrical Engineering from Kyungpook National University, Korea in 1999, 2001, and 2004 respectively. He is currently an associate

professor in Department of Information Security at Tongmyong University. His research interests include multimedia security, digital image processing, and computer graphics.



Ki-Ryong Kwon

He received the B.S., M.S., and Ph.D. degrees in electronics engineering from Kyungpook National University in 1986, 1990, and 1994 respectively. He worked at Hyundai Motor Company from 1986–1988 and at Pusan

University of Foreign Language from 1996–2006. He is currently a professor in Department of IT Convergence and Application Engineering at the Pukyong National University. He has researched University of Minnesota in USA in 2000–2002 with Post–Doc. and Colorado State University on 2011–2012 with visiting professor. He was the President of Korea Multimedia Society in 2015–2016. His research interests are in the area of digital image processing, multimedia security and watermarking, bioinformatics, weather radar information processing, and machine learning.