

# Anomaly Detection of 2.4L Diesel Engine Using One-Class SVM with Variational Autoencoder

Gye-bong Jang<sup>1</sup>, Sung-Bae Cho<sup>2</sup>

<sup>1,2</sup> *Soft Computing Laboratory, Department of Computer Science, Yonsei University, Seoul 03722, Korea*

*gyebong.jang@yonsei.ac.kr*  
*sbcho@yonsei.ac.kr*

## ABSTRACT

Despite intensive research, it is still at its early stage to prevent breakdown of construction machine, leading to a need to develop an autonomous and robust solution that minimizes equipment downtime and ensures the rigidity of equipment through predictive diagnostics. In particular, engine failure is critical because it causes the entire system to stop, highlighting the importance of determining and predicting the symptoms before failure occurs. However, at present, specific indicators based on domain knowledge should be set in order to judge a failure. This paper proposes an anomaly detection model for a 2.4L diesel engine, and verifies the model for two main faults. The proposed method based on deep learning extracts 130 feature parameters with autoencoder and distinguishes between normal and abnormal states by one-class support vector machine (OCSVM). An autoencoder can automatically extract useful features from multiple sensors on an excavator engine, and a variational autoencoder (VAE) extracts latent values from its input variables to generate new information. In this paper, a VAE is applied to extracting feature vector from the vibration signal for robust modeling, and OCSVM detects abnormal state and distinguishes between the two different faults and unknown factors. The experimental results show the accuracy of about 73%, and the false alarm related to the reliability of the model can be minimized to about 17%. Finally, to resolve the issues of reliability and interpretability of the model based on the deep learning, the Local Interpretable Model-agnostic Explanations (LIME) analysis is applied to listing the sensor data that affect the determination of the abnormal state. We intend to improve the accuracy of the model by adding expert knowledge to the data-driven model, because experts can easily make professional judgments about abnormal conditions and build a model with a continuously increasing sets of known data about faults and symptoms.

---

Gye-bong Jang et al. This is an open-access article distributed under the terms of the Creative Commons Attribution 3.0 United States License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

## 1. INTRODUCTION

The excavator is normally used at large construction sites, such as mines, quarries and other construction sites. Performance failure of an excavator directly affects the productivity, resulting in a great loss to the customer. Identifying such failures in advance and minimizing downtime are critical for both the manufacturer and the customer. The fault diagnosis approach can be categorized into three sub-approaches: model-based, data-based, and hybrid. Recently, data-driven approach has emerged as a way to solve many of the previously unexplained issues or to allow an even more efficient implementation. For example, in hidden Markov model (HMM) (Soualhi, Razik, Clerc and Dong Doan, 2014), data-driven approach (Yin, Ding, Xie and Luo, 2014; Erfani, Baktashmotlagh, Rajasegarar, Karunasekera and Leckie, 2015; Zimek, Schubert, Kriegel, 2012) and hybrid approach combine multiple models (Gao, Cecati and Ding, 2015).

Recently, with advancement in computer science, machine learning-based methods are gaining popularity. There are several models using deep learning technology (Jia, Lei, Lin, Zhou and Lu, 2016; Kim, Bu and Cho 2018) and the hybrid method (Lei, Jia, Lin, Xing and Ding, 2016). There have been several attempts to apply these methods in many fields: Using the correlation between sensors of small power generator to improve the performance compared to a single sensor feature (Zhao, et al., 2017); using data such as PHM08 to analyze with LSTM in order to prove the suitability of LSTM in predicting RUL (Remaining Useful Life) (Zheng, Ristovski, Farahat & Gupta, 2017); and comparing decision tree (DT), support vector machine (SVM) (Pal and Foody, 2010; Ruff, et al., 2018), k-nearest neighbor (kNN), and random forest against each other to prove the superiority of SVM and kNN to analyze data for prognostics (Shafi, Safi, Shahid, Ziauddin and Saleem, 2018).

The performance of many of these data-driven proactive diagnostics relies heavily on the degraded data to be applied. However, an engine has many sensors, which have different sensitivity to degraded engine performance. As some sensors

are more sensitive, it is necessary to choose appropriate sensor parameters with data more sensitive to engine failure tendencies, when selecting a set of training data for diagnostic prediction. In addition, it is not easy to obtain data on failure, and most of the failure data show only the information of the state after a failure has occurred, with no information about symptoms. Therefore, we come up with anomaly detection as critical in failure diagnosis.

If a training data set contains unusual data, detection of abnormal data may become difficult. In this paper, autoencoder is used to extract proper features and diagnose abnormality from the data measured in a real equipment. As one of the deep learning methods, autoencoder restores the most similar data to the input data based on encoder and decoder. We use a feature of the intermediate node, named latent space, as the feature for the whole data. In addition, OCSVM is used as an anomaly detection model. Since the degree of imbalance between the abnormal and normal data is significant, the normal data are grouped into one class while the effectiveness of anomaly detection is verified through outlier verification of two failure cases.

Finally, analysis is required to prove the reliability of the results obtained from the machine learning method. In machine learning, this is called explainable AI. We have applied LIME (Local Interpretable Mode-Agnostic Explanations) to provide the information of key sensors that have the biggest impact on the judgment, in order to facilitate the analysis of experts.

## 2. BACKGROUNDS

Recently, there are many detection methods including OCSVM-based, clustering-based and reconfigurable error-based methods. OCSVM uses kernel tricks to project input functions into high dimensional geometry space to separate normal and new data from each other. In other words, OCSVM is used to find the farthest origin, data points, and decision boundaries that regard the closest point to the origin as a novelty. However, SVM is actually memory-intensive and time-consuming, and the complexity increases seriously with the number of data.

### 2.1. Static Data Feature Selection

This section explains the feature extraction. It is difficult to extract features from raw data in the case of high-dimensional, time-series data with various environmental variables, such as engine operating data (Niennattrakul, Keogh and Ratanamahatana, 2010). To reduce the dimensionality of data, a high-level representation is built where a set of significant features are calculated. These features provide an approximation of the original raw data. Many statistical features are calculated for each time series variable  $X = \{x_1, x_2, \dots, x_n\}$ ,  $i = 1, 2, \dots, N$ . Figure 1 shows the main groups

of feature selection methodologies for static data, and explains the details of these measures.

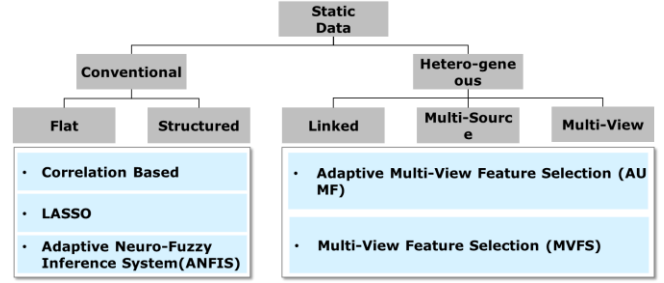


Figure 1. Approaches of the static data conversion.

Statistic conversion is used as a preprocessing method for feature extraction. Median, standard deviation, skewness, and kurtosis for each window size are found. This also improves the performance by learning secondary and tertiary conversion values, not the primary conversion. Each statistical value is defined as follows:

1. Mean is the estimated central value, and best known as the *mean* of the values  $x_1, \dots, x_N$ ,

$$\bar{x} = \frac{1}{N} \sum_{j=1}^N x_j. \quad (1)$$

2. With a central value of a distribution being characterized, “width” and “variability” around the central value can be conveniently characterized. Here, more than one measurement is possible: Most commonly the variance,

$$VAR(x_1 \dots x_N) = \frac{1}{N-1} \sum_{j=1}^N (x_j - \bar{x})^2 \quad (2)$$

and its square root, the standard deviation,

$$\sigma(x_1 \dots x_N) = \sqrt{VAR(x_1 \dots x_N)}. \quad (3)$$

3. Skewness is a measure of symmetry, or more precisely, the lack of symmetry. A distribution, or a data set, is symmetric if it looks the same on the left and the right of the center point. The usual definition is

$$Skew(x_1 \dots x_N) = \frac{1}{N} \sum_{j=1}^N \left[ \frac{x_j - \bar{x}}{\sigma} \right]^3 \quad (4)$$

where  $\sigma$  is the standard deviation in Eq. (3).

4. Kurtosis is a measure of whether the data are heavy-tailed or light-tailed when compared to the normal distribution. That is, data sets with high kurtosis tend to have heavy tails, or outliers. Data sets with low kurtosis tend to have light tails, or less outliers. A uniform distribution would be an extremely rare case. The conventional definition of kurtosis is

$$Kurt(x_1 \dots x_N) = \left\{ \frac{1}{N} \sum_{j=1}^N \left[ \frac{x_j - \bar{x}}{\sigma} \right]^4 \right\} - 3 \quad (5)$$

where the  $-3$  term makes the value zero for a normal distribution.

These values represent statistical values of raw data and are used to extract the features not recognized in raw data.

## 2.2. Autoencoder

Autoencoder has been introduced as one kind of artificial neural network (Liou, Cheng, Liou and Liou, 2014), unsupervised algorithm for learning how to decode its own input  $(x_1, x_2, \dots, x_n)$  to its output  $(y_1, y_2, \dots, y_n)$  based on  $(x_i = y_i)$ , which becomes possible by reducing the gap between the inputs and the outputs (Guo, et al., 2016).

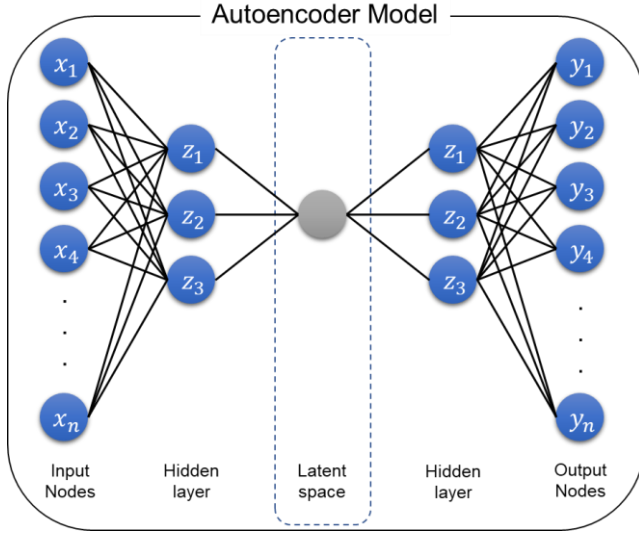


Figure 2. Autoencoder model architecture.

As shown in Figure 2, the input and output neurons for autoencoder are the same, and the hidden layer is a compressed or learned latent value. In general, autoencoder has a structure of a neural network with at least one hidden layer and learned the hidden layer (latent space) to minimize the reconstruction error.

Reconfigurable error in autoencoder is defined as follows to learn to minimize the reconstruction error through model learning.

$$E(X, Y) = \|X - \varphi(\varphi(X))\|_2^2 = \sum_{i=1}^m (x_i - y_i)^2. \quad (6)$$

## 2.3. One-Class SVM

OCSVM for unsupervised anomaly detection has expanded the idea of SVM largely applied in classification (Schoelkopf and Smola, 2002). The classic SVM aims to maximize the margins between the data points by looking for hyperplanes, whereas in OCSVM, hyperplanes learn to best distinguish data points from the origin. SVMs can usually capture nonlinearities using the kernel. The kernel method maps the data points from the input feature space of  $\mathbb{R}^d$  in the higher dimensional space of  $\mathbb{R}^D$  (where  $D$  is potentially infinite), where the data become linearly separable by transformation  $\mathbb{R}^d \rightarrow \mathbb{R}^D$ .

The most commonly used kernel is radial basis function (RBF), defined by the similarity mapping between any two points  $x$  and  $x_0$  of the input feature space and  $K(x, x') = \exp(-\frac{\|x-x'\|^2}{2\sigma^2})$ , where  $\sigma$  is the kernel bandwidth. Here,  $w$  and  $\rho$  are the vectors representing all the weights of all dimensions in the kernel space, and the offset parameter determines the distance from the origin to the hyperplane. The objective of the OCSVM is to isolate all data points from the origin with the maximum margins in relation to some constraint mitigation, and this can be written in a quadratic program as follows:

$$\begin{cases} \min \frac{1}{2} w^2 + \frac{1}{vN} \sum_{i=1}^N \xi_i - \rho \\ \text{s. t. } (w \times \psi(z)) \gg \rho - \xi_i, \xi_i \geq 0 \end{cases} \quad (7)$$

where  $\xi_i$  is a slack variable and  $v$  is the regularization parameter. Theoretically,  $v$  is the upper bound of the fraction of anomalies in the data, and the main tuning parameter for OCSVM. Additionally, by replacing  $\xi_i$  with the hinge loss, unconstrained objective function becomes

$$\min_{w, \rho} \frac{1}{2} \|w\|^2 - \rho + \frac{1}{vN} \sum_{i=1}^N \max(0, \rho - w^T \phi(x_i)). \quad (8)$$

Let  $g(x) = w\phi(x_i) - \rho$ , and the decision function of OCSVM is

$$f(x) = \text{sign}(g(x)) = \begin{cases} 1 & \text{if } g(x) \geq 0 \\ -1 & \text{if } g(x) < 0 \end{cases} \quad (9)$$

The optimization problem of SVM in Eq. (8) uses Lagrangian multiplier to solve convex optimization problem in double space, thereby reducing complexity and enhancing feasibility. However, solving SVM in dual space can affect the size of the data due to  $O(n^2)$  complexity, and the function  $K$  between each pair of data sets must be computed and stored in a matrix. Here,  $n$  is the size of the data set.

## 3. METHODOLOGY

This section illustrates the use of deep learning in autoencoding reconstruction error to describe novel detection of unlabeled data. The proposed method has two basic functions: dimension reduction and novelty identification. General architecture of the proposed method is shown in Figure 3.

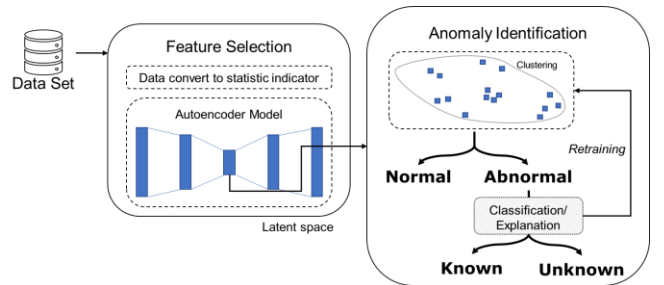


Figure 3. Architecture of the proposed model.

In this paper, the OCSVM is used as an anomaly detection model. In addition, the model is designed to make judgment of abnormal data more conveniently when experts use known information and explanation from LIME. The ultimate objective of the learning model is to constantly update the expansion of the known area, allowing for more precise judgment of failures.

### 3.1. Analysis of the Input Data

Data was collected from 123 different sensor values measured in 0.1 sec of actual equipment for 12 days, and two different failure cases such as PCR (Pressure control regulator) and SmkLim (Smoke Limitation) failure, were labelled in the working log. Table 1 shows the summarized list of sensors.

Table 1. Measured Sensor List.

Variable	Unit
Engine speed	rpm
Sensed fresh air mass flow	Kg/h
Temperature down stream of charged air cooler	°C
Desired EGR ratio	Kg/h
Charge state of the DPF	-
Desired air mass	Kg/h
Gas mass flow at throttle valve	Voltage
Standardized accelerator pedal position	%
Battery voltage	Voltage

Characteristics of these measured data are summarized as follows:

- 12 non-consecutive days of operation including one PCR and one SmkLim day.
- Each day consisted of 130 attributes and 80,000+ tuples where data were gathered in every 100 milliseconds.
- Depending on the operation condition, different sensors were activated; some sensors were binary, but others were not.
- Some were seemingly useful when detecting a fault and its previous alarms.

The result of correlation analysis between 123 sensors with each day as class confirmed that there were positive and negative correlation intermingled as Figure 4. However, a correlation specifically with the fault data was not found because most types of faults or symptoms occurred in the form of point failures. Therefore, the anomaly detection model conducted data preprocessing and model learning in order to combine several sensors to perform fault diagnosis and condition diagnosis.

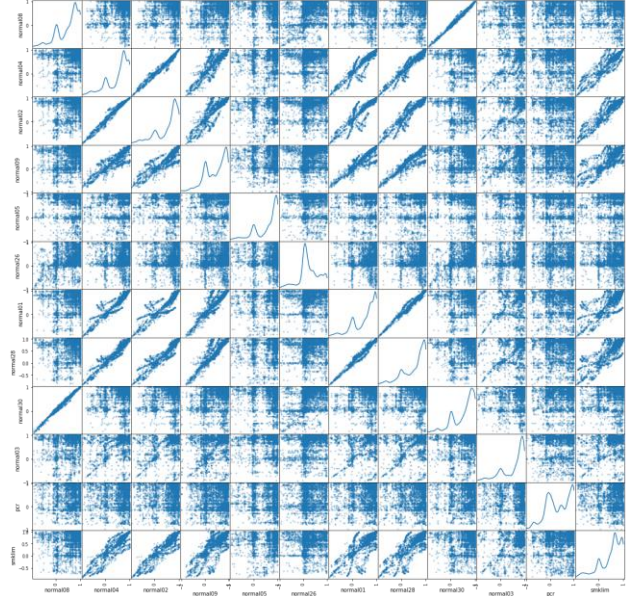


Figure 4. Correlation between each day with 123 sensor.

### 3.2. Feature Selection with Autoencoder

The principal component analysis (PCA) or autoencoder (AE) is applied to reduce the dimension of multi variable dimensional sensor data for performance improvement of the model. However, raw value cannot be applied to autoencoder because of the large learning dimension of 369K. Therefore, statistical values is used instead, and they include median, variance, skewness, quantiles, and deciles, which are mainly used in statistical view of probability for autoencoder.

Conversion of raw data into statistical values during the preprocessing stage makes it possible to reduce the dimensionality and express the characteristics of the data more prominently. As we transform the data into six statistic value types and compare the results of each conversion, we find that effective results have been obtained when the deciles are used. This leads us to reduce the learning dimension to 3.3K, and the representation of the two failure cases becomes more abundant. The summarized result of conversion is shown in Figure 5.

(Originally 123 Features)	FE Result	# of Rows to Reach 5min Span	Total Features	AE Performance	Result
W/O Stats.	123	0.1 (sec.)	369K	NA	
With Stats.	Median	300 (sec.)	1.2K	No Go	Rejected
	Variance				
	Skewness				
	Kurtosis				
Quantiles	5	1 (sec.)	1.5K		
Deciles	11		3.3K	Go	Selected

Figure 5. Analysis results according to data type for application to autoencoder model.

We also compare the feature extraction performance for a fault using two methods. Compared to PCA, autoencoder is not only better at distinguishing both faults, but also requires less features to be extracted. PCA can only distinguish PCR, and requires 10 features to be extracted. In Figure 6, the blue shaded area indicates the area of real failure detection. However, the PCA-applied data do not find any particular differences in the area.

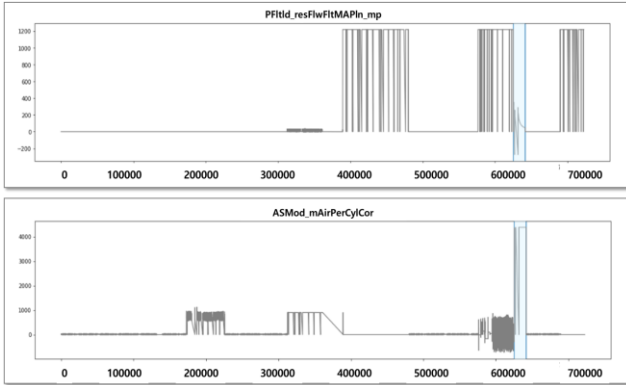


Figure 6. Multi-dimensional PCA result.

The autoencoder extracts more abundant features than PCA in Figure 7. In addition, we confirm that the characteristics of the two faults, PCR and SmkLim, are expressed differently from each other. Here, the overlap between the steady state and the fault state occurs because the measured data is labeled on a daily basis, not failure basis. Therefore, we classify the anomalies distinguished from the normal using OCSVM.

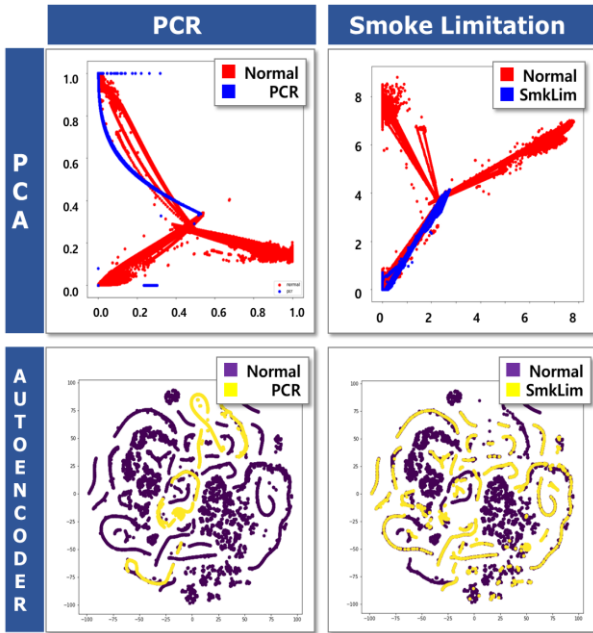


Figure 7. PCA and autoencoder results for failures.

### 3.3. OCSVM for Anomaly Detection

OCSVM is best suitable for detecting “abnormal” patterns when the target event is rare. Training requires a normal dataset only, in order to create a boundary that can optimally cover the observations as normal. However, the model training takes a lot of time (8hrs+α). The parameters of the model are Nu = 0.1 and Kernel = rbf. Figure 8 shows the concept of OCSVM.

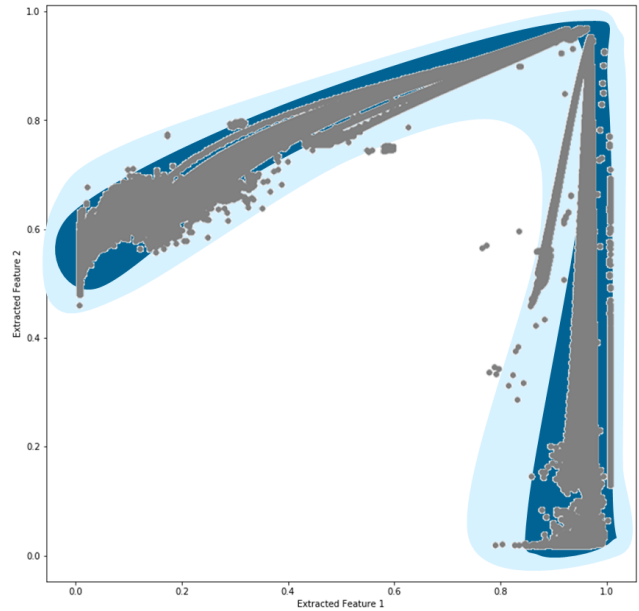


Figure 8. Conceptual results of OCSVM.

For construction of the abnormal detection model, the information of a known failure case can be provided to the model in advance in order to improve the judgment performance of the model. The hyperparameter is tuned by applying the information of the two failure cases together in order to improve the performance of the model. This enables the model to distinguish between known and unknown failures.

### 3.4. Explainable AI

Local Interpretable Model-agnostic Explanations (LIME) is an algorithm that can faithfully describe the evaluation of all classifiers or regressors (Ribeiro, Singh and Guestrin, 2016). We use this method to explain the failure classification. The model also shows which part of the sensor is used for each evaluation. Classes are predefined, but the importance of each class can vary for each failure. LIME extracts the main element to find the distance between each class and feature. The evaluation of the classifier or regressor in a faithful way is explained as follows:

$$\varepsilon(x) = \operatorname{argmax}_{g \in G} L(f, g, \pi_x) + \vartheta(g) \quad (10)$$

$$L(f, g, \pi_x) = \sum_{z, z' \in Z} (\pi_x(z)(f(z) - g(z'))^2) \quad (11)$$

Table 2 shows the algorithm for LIME.

Table 2. Sparse linear explanations using LIME.

```

Input: Classifier  $f$ , Number of samples  $N$ 
Input: Instance  $x$ , and its interpretable version  $x'$ 
Input: Similarity kernel  $\pi_x$ , Length of explanation  $K$ 
1: begin
2:    $Z \leftarrow \{\}$ 
3:   for  $i \in \{1, 2, 3, \dots, N\}$ 
4:      $z'_i \leftarrow \text{sample around}(x')$ 
5:      $Z \leftarrow Z \cup \langle z_i, f(z_i), \pi_x(z_x) \rangle$ 
6:   end for
7:    $\omega \leftarrow \text{K-Lasso}(Z, K)$ 
8: return  $\omega$ 
    
```

Thus, using the results of the expert analysis from LIME, it is possible to define the symptom and continuously update the relationship between the faults and the symptoms. Figure 9 illustrates this process.

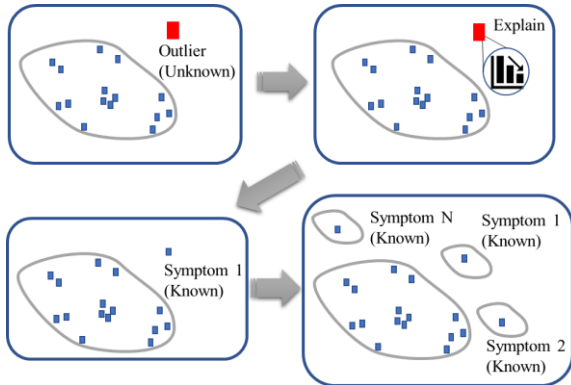


Figure 9. Defining the correlation between errors and symptoms using LIME.

## 4. EVALUATION

### 4.1. Implementation

As seen in Figure 10, we have developed a program to judge the data received from real machine based on the learned model. The program detects the current status as either abnormal or not in real-time. If there is abnormal, the program shows the explainable sensor list to help the investigation of experts.

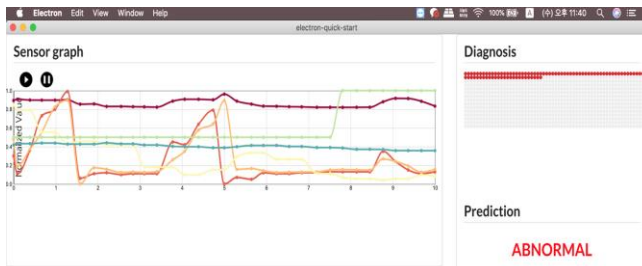


Figure 10. Representation of abnormal data in OCSVM.

## 4.2. Experiments

Experimental data were collected from 123 sensors for 12 days. The target engine is a 2.4L diesel engine, which is the data collected from the engine controller as shown in Figure 11. Each sensor data was measured in 100ms unit such as pressure, torque and temperature, etc. The total data capacity was approximately 346MB. There were two failure cases, PCR and SmkLim, on different dates. Each failure occurred in the following cases:

1. SmkLim: When the required fuel amount exceeds the limit fuel amount, a lot of smoke is prevented (fuel amount limitation).
2. PCR: When the actual sensing pressure is lower than the threshold pressure.



Figure 11. 2.4L diesel engine as target component.

Both are critical failures frequently found in engines and lead to emissions beyond the standard or equipment degradation. The 10 days data without a failure case were used for training, and the data of the other 2 days with the two fault cases were used to verify the outlier detection ability.

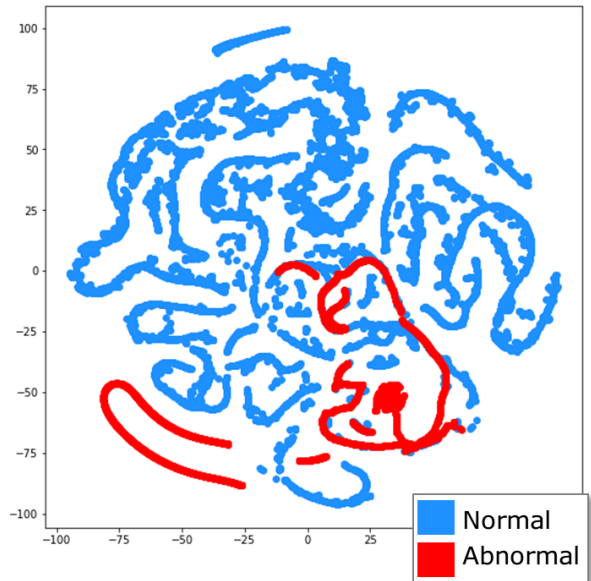


Figure 12. Representation of abnormal data in OCSVM.

Red area in Figure 12 shows the data for the fault detection, and it confirms that some data are detected as abnormal in a form similar to normal.

The constructed model is validated with the preliminary data, and accuracy is calculated as 73%, and recall as 83%. (1-Precision is the value for false). In addition to the accuracy, one of the important factors in evaluating the reliability of the model is the increased number of recalls because the model is able to reduce false alarms.

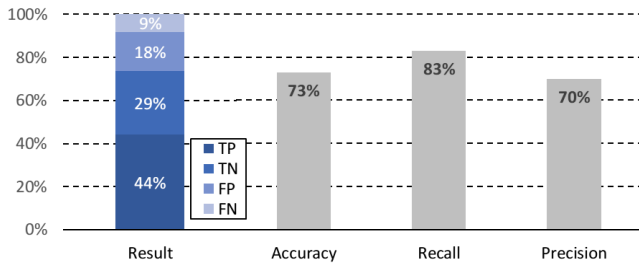


Figure 13. Result of accuracy, recall and precision.

Another factor for model reliability is expandability. This is because customer / user needs to determine the credibility of the model in finding abnormalities, if the model regards anything as abnormal. Therefore, a plausible explanation is needed to understand the cause of a failure. XAI (eXplainable AI) techniques are implemented in the model to provide transparency and interpretability of the resulting diagnosis. Explanation given by LIME can help an engineer to make a judgment by listing by prioritizing a set of sensor data which affect the decision of the model in detecting abnormality (Zeldam, Jong, Loendersloot and Tinga, 2018). Figure 14 shows the LIME result for one case, and the result from LIME confirms the fault CASE by using NOT Normal, PCR, and NOT SMK. For example, DFES\_stChk\_01 and 8 sensors are the biggest factors that distinguishes PCR.

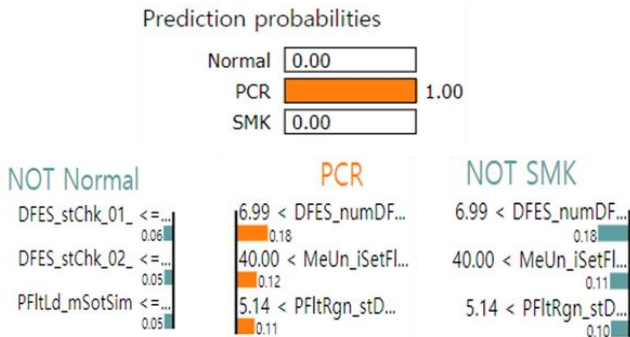


Figure 14. Representation of abnormal data in OCSVM.

5. CONCLUDING REMARKS

In this paper, we have proposed an anomaly detection model for two fault cases based on the engine state information with 123 parameters. The model is appropriate for the multi-dimensional sensor data which shows dynamic changes in

values. By using the latent space of autoencoder, we could reduce multi-dimensional data and propose a method to extract appropriate features. In addition to this, one-class SVM is applied to merge the methods to distinguish the normal and the abnormal, in order to resolve the unbalancing problem between the normal and the abnormal data. Verification of the two fault cases show the performance of 73%, and the reliability of the model is secured by lowering the false alarm rate, an important factor in anomaly detection, to 17%. In addition, we have proposed a real-time program to apply and operate the proposed model in the real environment on the actual equipment. In terms of abnormal data detection, the reliability of the model can further be improved by outputting the correlated sensor information through LIME, which allows analysis of the abnormal cases facilitated by expert opinions. Finally, diagnostic performance of the model can be improved by reducing the number of unknown abnormal cases, which is possible by updating the model with knowledge of experts on the symptoms as collected from the expert opinions. This machine learning method allows more accurate data collection and status diagnosis for anomalies.

ACKNOWLEDGEMENT

This work has been supported in part by the Doosan Infracore Co., Ltd.

REFERENCES

Erfani, S., Baktashmotlagh, M., Rajasegarar, S., Karunasekera, S., Leckie, C., (2015). R1SVM: A Randomised Nonlinear Approach to Large-Scale Anomaly Detection, *Proceedings of the Twenty-Ninth AAAI Conference on Artificial Intelligence*, (pp.432-438).

Gao, Z., Cecati, C., Ding, S., (2015). A survey of fault diagnosis and fault-tolerant techniques-Part II: Fault diagnosis with knowledge-based and hybrid/active approaches, *IEEE Trans. Ind. Electron*, vol. 62(6), pp. 3768–3774, doi: 10.1109/TIE.2015.2419013.

Guo, Y., Liu, Y., Oerlemans, A., Lao, S., Wu, S., Lew, M.S., (2016). Deep Learning for Visual Understanding: A review, *Neurocomputing*, vol. 187, pp. 27–48.

Jia, F., Lei, Y., Lin, J., Zhou, X., Lu, N., (2016). Deep neural networks: A promising tool for fault characteristic mining and intelligent diagnosis of rotating machinery with massive data, *Mech. Syst. Signal Process.* vol. 72–73, pp. 303–315.

Kim, J.-Y., Bu, S.-J., Cho, S.-B., (2018). Zero-day malware detection using transferred generative adversarial networks based on deep autoencoders, *Information Sciences*, vol. 460-461, pp. 83-102.

Lei, Y., Jia, F., Lin, J., Xing, S., Ding, S. (2016). An intelligent fault diagnosis method using unsupervised feature learning towards mechanical big data, *IEEE Trans. Ind. Electron*, vol. 63(5), pp. 3137–3147.

- Liou, C., Cheng, W., Liou, J., Liou, D., (2014). Autoencoder for Words. *Neurocomputing*, vol. 139, pp. 84–96.
- Niennattrakul, V., Keogh, E. J., Ratanamahatana, C. A., (2010). Data editing techniques to allow the application of distance-based outlier detection to streams, *Proc. IEEE ICDM*, pp. 947–952, doi: 10.1109/ICDM.2010.56.
- Pal, M. and Foody, G. M. (2010). Feature selection for classification of hyperspectral data by SVM. *IEEE Transactions on Geoscience and Remote Sensing*, vol. 48(5), pp. 2297–2307.
- Ribeiro, M., Singh, S., Guestrin, C. (2016). why should i trust you? explaining the predictions of any classifier. *Proceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining*, August 13-17, San Francisco, CA, pp. 1135-1144.
- Ruff, L., Robert, V., Nico, G., Lucas, D., Shoaib, S., Alexander, B., Emmanuel, M., Marius, K., (2018). Deep One-Class Classification. *Proceedings of the 35th International Conference on Machine Learning*, July 10-15, Stockholm, Sweden.
- Shafi, U., Safi, A., Shahid, A.R., Ziauddin, S., & Saleem, M.Q. (2018). Vehicle Remote Health Monitoring and Prognostic Maintenance System, *Journal of Advanced Transportation*, vol. 2018, DOI:10.1155/2018/8061514
- Schoelkopf, B., Smola, A., (2002). Learning with Kernels. *MIT Press*, Cambridge MA, ISBN: 9780262256933.
- Soualhi, A., Razik, H., Clerc, G., Dong Doan, D. (2014). Prognosis of Bearing Failures Using Hidden Markov Models and the Adaptive Neuro-Fuzzy Inference System. *IEEE Trans. Industrial Electronics*, vol. 61(6). pp. 2864-2874, doi: 10.1109/TIE.2013.2274415.
- Yin, S., Ding, S., Xie, X., Luo, H. (2014). A review on basic data-driven approaches for industrial process monitoring *IEEE Trans. Ind. Electron.* vol. 61(11), pp. 6418–6428.
- Zeldam, S., de Jong, A., Loendersloot, R., Tinga, T. (2018). Automated failure diagnosis in aviation maintenance using explainable artificial intelligence (XAI). *Proceedings of the European Conference of the PHM Society*, vol. 4.
- Zhao, P., Kurihara, M., Tanaka, J., Noda, T., Chikuma, S., Suzuki, T. (2017). Advanced correlation-based anomaly detection method for predictive maintenance, *IEEE International Conference on Prognostics and Health Management (ICPHM)*, Dallas, TX, pp. 78-83, doi: 10.1109/ICPHM.2017.7998309.
- Zheng, S., Ristovski, K., Farahat, A., Gupta, C., (2017). Long Short-Term Memory Network for Remaining Useful Life estimation, *IEEE International Conference on Prognostics and Health Management (ICPHM)*, Dallas, TX, pp. 88-95, doi: 10.1109/ICPHM.2017.799831.
- Zimek, A., Schubert, E., Kriegel, H.P. (2012). A survey on unsupervised outlier detection in high-dimensional numerical data. *Statistical Analysis and Data Mining*, vol. 5, pp. 363–387. doi: 10.1002/sam.11161

## BIOGRAPHIES



**Gye-bong Jang** is a Ph.D. candidate in computer science at Yonsei University, Seoul, South Korea. He received M.S. degree in signal processing from Korea University, Seoul, South Korea.

His research interest focuses on prognostics and health management for construction equipment system using deep-learning or machine learning.

**Sung-Bae Cho** received the B.S. degree in computer science from Yonsei University, Seoul, Korea and the M.S. and Ph.D. degrees in computer science from KAIST (Korea Advanced Institute of Science and Technology), Daejeon, Korea. He was an Invited Researcher of Human Information Processing Research Laboratories at ATR (Advanced Telecommunications Research) Institute, Kyoto, Japan from 1993 to 1995, and a Visiting Scholar at University of New South Wales, Canberra, Australia in 1998. He was also a Visiting Professor at University of British Columbia, Vancouver, Canada from 2005 to 2006. Since 1995, he has been a Professor in the Department of Computer Science, Yonsei University. His research interests include neural networks, pattern recognition, intelligent man-machine interfaces, evolutionary computation, and artificial life. Dr. Cho was awarded outstanding paper prizes from the IEEE Korea Section in 1989 and 1992, and another one from the Korea Information Science Society in 1990. He was also the recipient of the Richard E. Merwin prize from the IEEE Computer Society in 1993. He was listed in Who's Who in Pattern Recognition from the International Association for Pattern Recognition in 1994, and received the best paper awards at International Conference on Soft Computing in 1996 and 1998. Also, he received the best paper award at World Automation Congress in 1998, and listed in Marquis Who's Who in Science and Engineering in 2000 and in Marquis Who's Who in the World in 2001. He is a Senior Member of IEEE and a Member of the Korea Information Science Society, INNS, the IEEE Computational Intelligence Society, and the IEEE Systems, Man, and Cybernetics Society.