Application of Auto-Encoder for Time Series Classification with Class Imbalance

Yi-Ching Yeh and Chia-Yu Hsu

September 5, 2019

Application of Auto-Encoder for Time Series Classification with Class Imbalance

Yi-Ching Yeh
Department of Industrial Engineering and Management,
National Taipei University of Technology,
Taipei, Taiwan
t107378049@ntut.org.tw

Chia-Yu Hsu*
Department of Industrial Engineering and Management,
National Taipei University of Technology,
Taipei, Taiwan
chiayuhsu@mail.ntut.edu.tw

Abstract*—Prognostics and health management (PHM) is important to increase the reliability of production equipment and to detect failure events of equipment in advance. In order to model the equipment status, data-driven approaches have been used to extract key feature from various sensors installed on the equipment and build a model to diagnosis the occurrence and end time of the plant failure and the sort of the fault. In particular, the data from each sensor were recorded by specific time period and are also called time series data. To consider the condition of monitor equipment, it is defined as time series classification problem. In recent years, machine learning methods have been widely used to detect plant failure events. Each sensor data typically are separated into several indicators based on their process steps. However, the variation need to be incorporated into time series classification model. In addition, although auto-encoder are widely used to image classification, it still has a challenge to use to time series data. This paper presents an auto-encoder (AE) method of time series classification to distinguish different time series pattern for failure diagnosis. Auto-encoder is a fault detection way of identifying the normal or abnormal of each sensor data. In order to get better results, deep learning model for multivariate time series classification is used to extract the time sequence characteristics. To evaluate the performance of proposed model, the data collected from PHM 2015 were used to compare with the Random Forest, Xgboost and LSTM-based model for performance evaluation. In particular, a plant with minimum proportion fault type was used to examine the effect of imbalanced class. According to the experimental results, the proposed AE outperforms better than other machine learning classification models.

Keywords*—data-driven, prognostics and health management (PHM), long-short term memory (LSTM), classification, anomaly detection, auto-encoder

I. INTRODUCTION

Predictive maintenance is play an important role in the manufacturing fabrication. Effective and accurate fault prediction system helps to prevent serious accidents, saves downtime costs and increases productivity. The methods to fault detection can be generally classified into two type: model-based methods and data-driven methods. The traditional model-based methods are based on the physical models of the systems and the related experience [1]. Data-driven method have been widely used on data analysis. Thus, we need to use the data to train a model, which can be used to predict the failure of the machine in the future.

Sensor data is a time series data, which is characterized by a series of data points indexed in chronological order. It exists in various application fields, such as stock price forecasting, weather forecasting and equipment fault detection and monitoring. Predictive maintenance problem could be divided into two types. One is to predict the remaining useful life (RUL) of the machine, and the other is to diagnose whether the machine has a fault. In this paper, the problem is the diagnosis whether the machine has fault or not.

Most of existing studies were subject of diagnosing faults, and most of them focus on using machine learning methods to solve related problems. However, the key features from these time series data are not easily identified. This paper aims to propose a deep learning method for sensors using condition monitoring and failure detection based on their own collected measurements. Since the problem of data imbalance is encountered in the past to explore this type of fault diagnosis problem, the normal data set is fewer than the abnormal dataset. Therefore, this paper solve the data in the proposed method. In the case of data imbalance, the time of occurrence of the abnormal can still be correctly diagnosed, and this paper focuses on maintaining high accuracy when the proportion of failure is the lowest. Thus, we mainly focus on the lowest proportion of failures in each factory.

The rest of this paper is organized as follows. Section II Problem Description. Section III presents related surveys and method. Section IV introduces the proposed algorithm. Section V present the experiment. Section VI presents detailed discussion. The conclusions and future work are presented in Section V.

II. LITERATURE REVIEW

A. PHM 2015 Data Challenge paper

Refer to the author's method of winning the first place in the PHM 2015 competition, Xiao [2] proposed the method based on ensemble machine learning model to predict industrial plant faults based on classification methods such as penalized logistic regression, random forest and gradient boosted tree. Although this paper performance is well, the author does not consider each plant the least proportion of fault type. In addition, the second contestant Kim [3] in the competition also proposed a fault log recovery method, which is a based a machine learning-based fault classification approach FDA classifier for fault classification problem. The third contestant Cong Xie[1] proposed ensemble decision tree methods, including RF (Random Forest) and GBDT (Gradient Boosting Decision Tree) as the classifiers.

According to these papers, we can find that this study focused on using machine learning to solve the problem of time classification problem. However, in recent year, deep learning model have been rapidly developed, and it have proven to be good methods for automatic feature extraction.

So, many deep learning methods are applied to time series classification.

B. Long Short-term Memory for time series data

LSTM model can be traced back to the algorithm [4], which is an advanced model based on RNN. LSTM architecture has the input gate, output gate and forget gate, which can avoid to vanishing gradient and exploding gradient. GRU [5] is an advanced model based on LSTM model. Because this dataset is a time series type of data, the first model, which we directly think, is LSTM.

One of the paper also uses the same dataset, he proposed weighted deep representation learning model (wLRCL-D) for imbalanced fault diagnosis. The deep learning model contains 2-layer CNN and 2-layer inner LSTM and the loss function is a class-imbalance weighted loss function that takes the weight of minority and majority classes into consideration. CNN and LSTM method is the most straightforward method to deal with time series classification problem. Thus, the other author proposed a MVCNN model, whose model contains many one-dimensional CNN layers. According to the above, it is known that many papers have used the CNN and LSTM methods in this dataset. No one has tried to use auto-encoder in this dataset.

C. Auto-encoder Method

Auto-encoders are simple learning circuits which aim to transform inputs into outputs with the least possible amount of distortion [6]. This deep learning method have been shown on a number of challenging classification and fault detection problem. Many studies use the deep auto-encoder model based on normal data and the mean reconstruction error as the threshold to how far it is from the nominal condition and is used as an anomaly indicator [6]. The method proposed in this paper is to use the feature of compressing and restoring the feature. The training model only learns the normal data. The loss value indicates the degree of data reduction, and the value is used to distinguish between normal and abnormal. In addition, we will verify that the proposed model is suitable for different plants.

III. METHOD

A. Pipeline Overview

The research flow of preprocessing the imbalanced fault detection is composed of three main step as Figure 1.

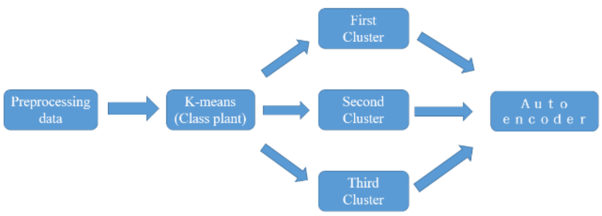

Figure 1. research flow

The research flow includes (1) Data Preprocessing: Organize all the folders containing the features into a folder, and merge the folders containing the labels into a folder. (2) Cluster: Considering the difference in the number of parts and the number of zones and the type of faults in each plant, k-mean is used as a method of grouping. (3) Model Training: the deep learning model auto-encoder is proposed, which consists of two layers of encoder and two layers of decoder in fault detection.

B. Data Preprocessing

In this paper, we take 31 plants from 2015 PHM train data as all data, but plant 2 and plant 27 are eliminated, which are outlier. For each plant is regard as independent. So, train and test data are split from each plant. The preprocessing data can be following step, which some step follow competition rule.

(1) Convert all timestamp into one unit every fifteen minutes, and drop duplicate record at the same time.

(2) Merge the features of the same time but different components and zones into the same row.

(3) Fill in the feature values of the previous time point into the missing features.

(4) Feature engineering, we added the month, date and hour of the time as a new feature.

(5) Combine the A and B files of the specified plant to produce a feature file.

(6) Combining the start and end time label, we do not consider the start and end time problems. We only diagnose the current time whether there is a fault.

C. K-means algorithm

However, due to the differences in the characteristics of the factories and the distribution of faults in the data types, the factories are characterized by the number of different components, the number of zone zones, the number of fault classes, and the proportion of each of the six categories of faults. We use K-means algorithm [7], which is an unsupervised model. The objective of traditional K-means can be expressed as

$$\frac{1}{n}\sum_{i=1}^{n}\left[\min_{j} d^2(x_i, y_j)\right] \quad (1)$$

Given a dataset of n data points $x_1, x_2, \cdots, x_n$ such that each data point is in R^d, the problem of finding the minimum variance clustering of the dataset into k point $\{y_j\}(j = 1,2,\cdots,k)$ in R^d such that is minimized, where $d(x_i, y_j)$denotes the Euclidean distance between $x_i$and y_j. The points $\{y_j\}(j = 1,2,\cdots,k)$ are known as cluster centroids. The problem in Eq. (1) is to find k cluster centroids, such that the average squared Euclidean distance (mean squared error, MSE) between a data point and its nearest cluster centroid is minimized. [7]

It is mainly divided into three groups, and then data analysis for different groups of factories has found a model suitable for each cluster. Since the method proposed in this paper is to solve the problem of serious imbalance of data, we only discuss the least proportion of fault type in each factory. As shown in Table 1, taking factory one as an example, F4 has the least proportion, and so on.

TABLE I. Each fault type of plant 1

Fault type	F1	F2	F3	F4	F5	F6
ratio	25%	18%	13%	4%	10%	30%

TABLE II. The result of the fault type of each plant

Class1 plants	Fault type	Class2 plants	Fault type	Class3 plants	Fault type
1	4	5	3	3	5
4	3	10	4	6	3
7	5	24	1	18	4
8	3	31	5	21	3
9	5			23	1
11	5			25	5
12	4			28	5
13	3			29	4
14	4			30	5
15	4			32	3
17	4				

Then, we hope to find the model, which are suitable for different groups of factories.

D. *Auto-encoder(DAE)*

Auto-encoder is unsupervised learning that applies backpropagation, setting the target values to be equal to the inputs. In other word, model can be trained without label. Deep auto-encoder is constructed by a multi-layer neural network, where there is an input layer, single or multiple hidden layers and an output layer.

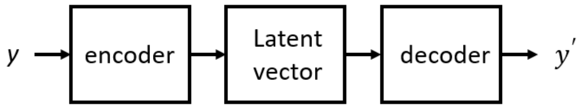

Figure 2. auto-encoder model

Let $Y = (y_1, y_2, \cdots, y_n)$ be the input vector of an auto-encoder network and encoder transforms the input Y into a low-dimensional latent vector. Since the latent vector is of low dimension, the encoder is forced to learn only the most important features of the input data. Decoder is tries to recover the input from the latent vector. $Y' = (y_1', y_2', \cdots, y_n')$ is the output data. [8] The reconstruction error is

$$MSE = \frac{1}{n}\sum_{i=1}^{n}(y_i - y_i')^2$$

If the construction error is higher than a certain threshold τ the current condition of a sensor is marked as fault. The threshold τ is a random number, which can be replaced as all row mean square error.

IV. EXPERIMENT

This paper deals with the discovery of faults in sensor data. The dataset is provided by PHM 2015 Challenge. This data is based on the factory. Each factory has three files: A, B and C. The A file has the time series about sensors signals (S1~S4) and control reference signals (R1~R4), which is depend on the component. In addition, the number of components in each factory is different. The B file has the time series about cumulative energy consumption, which is depend on the plant of zone. The C file has fault start time, fault end time and fault type.

The goal of the competition is to predict the start time and end time of each fault type. The fault type is six. But in the paper, we didn't consider the start time and end time. We tried to transform the task issue into a classification problem. [3] We only predict whether there is an abnormality at this point in time. Therefore, this paper focuses on the accurate prediction of when a fault will occur in the case of data imbalance.

Therefore, this paper focuses on the accurate prediction of when a fault will occur in the case of data imbalance. The dataset can be downloaded from NASA Ames Prognostics Data Repository.

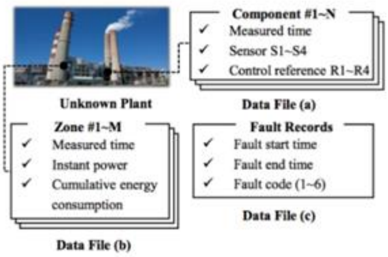

Figure 3. Datasets

The experimental environment and the settings are using GPU is Nvidia GeForce RTX-2080TI 11GB, RAM is 32GB SSD, and SATA3 480GB.

A. *Machine Learning Experiment*

- Random Forest

Data-driven or statistical approaches based on historical data are seen as the most cost-effective approach for fault detection in complex systems [10]. Machine learning is one of the fault detection way to data-driven algorithm.

Random forest is an ensemble learning method which combine many decision trees. The method has been extensively applied to the dataset. Therefore, we use the way as the benchmark.

According to the dataset that has been preprocessed, the normal data and the abnormal data are divided into two categories. Then 80% of the normal and abnormal categories split as training data sets and 20% of the normal and abnormal categories as test datasets. Based on this segmentation dataset, we want to put this kind of information into machine learning as a benchmark for comparison. Therefore, we choose the Random Forest method. The Random Forest model parameter set that n_estimators is 600.

As the results, we use the Random Forest model to learn but it is not ideal. Especially the fifth and thirty-one factories are the worst because data imbalance problem of the two plants is particularly serious. The performance of precision, recall and F-score is better than XGBoost.

- XGBoost

XGBoost is a scalable machine learning system for tree boosting. [10]

For the data, we follow steps to split train and test datasets. First, the normal data and the abnormal data are divided into two categories. Then 80% of the normal and abnormal categories split as training datasets and 20% of the normal and abnormal are divided as test datasets.

Based on this segmentation dataset, the xgboost model parameter set that n_estimator is 600. As the result, the performance is worse than RF model. Especially the 5th and 31th factories performance about the precision, recall and F-score are closed to 0.

B. Deep Learning Experiment

- LSTM + attention

Because the dataset is time series data, we directly used LSTM model as a way. For the LSTM model, the data format is different, we need to do the conversion for data pre-processing. Follow step:

(1) From the data, the normal data and abnormal data are divided into two categories. Then 80% of the normal and abnormal categories split as training datasets and 20% of the normal and abnormal split as test datasets.

(2) Transfer dataset from two-dimensional (sample, feature) to three-dimensional (sample, timestamp, feature). We also called window size or timestamp, which is to split data into a time format.

(3) According to the experimental result, the window size is set to 64 as the optimal parameter value, which is given each 64 time points label. The result shape is (sample, 64, feature).

(4) Data normalization

(5) Specify a specific fault type to classify each time point to determine whether the time point fault has occurred. It belongs to the two-category problem. There is a fault occurrence label of 1 and no label is 0.

(6) The architecture of this model consists of a layer of LSTM, a layer of GRU and a layer of attention.

(7) Attention can help the model to give the different weight for features. Besides, it can depend on the important information to give the different weight in order to make the model better than before.

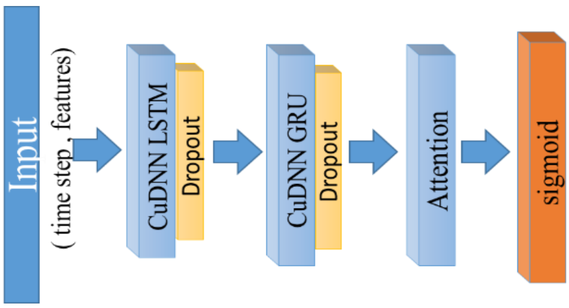

Figure 4. LSTM model

According to the results, the fault type accounts for a high proportion of the factory and the method of using LSTM accurately predict the occurrence of the fault. However, when the data is seriously balanced, the effect of using the LSTM model is not good, and it is impossible to correctly determine whether there is a fault or not. Because the model is all guessed as normal data. Although the accuracy is high, the performance of precision, recall and F-score is 0.

- Auto-encoder

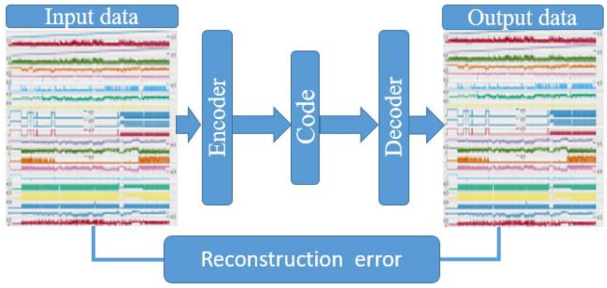

Figure 5. the architecture of this paper

In this situations, although the model LSTM predict high accuracy, there is a large difference of the performance of the precision, recall and F-score. We can see the result from the table 3. It is not good to use LSTM to predict the effect.

In general, most datasets are more normal than the abnormal data, so we proposed the method to solve this problem in this paper. These problems are defined as anomaly detection problems. By reference Scholar papers, we survey some use the auto-encoder method in the diagnosis of fault detection with data imbalance problem. [6] The paper build 11-layer deep auto-encoder model, which use the normal data, to be able to distinguish abnormal points in the case of data imbalance. But the threshold is applied on the average error such that a test run with larger error than threshold would be diagnosed as a fault. The approach is tested on data from NASA open database and demonstrates high fault detection rates (97.8%).

According to past research, many scholars have used the auto-encoder method to do fault detection. In the paper, we also use the similar method to do in the dataset. But the threshold is different with other paper setting. The detail about following the steps.

Step1: Each preprocessing data of each factory is selected, and the dataset marked as normal 0 and abnormal 1 are separately selected.

Step2: Then, the normal dataset is divided into 80/20 aliquots.

Step3: 80% of the normal dataset is put into auto-encoder for embedding and reconstruction, the other leaving 20% parts and all abnormal datasets as test datasets. All faults are test sets, and the normal ratio is adjusted to 10 times the total number of faults.

Step4: The auto-encoder model consists of two layer of encoder and two layer of decoder. We also tried to add more than two layer of encoder and decoder, but the result is not better than two layer of architecture.

Step5: Put the test data set into the trained auto-encoder model to restore the data with the same dimensions but similar values.

Step6: Setting threshold

According to the model predict data compared with the original data, the mean square error is used to calculate the difference between the original data and the restored data. Therefore, each column of the data table will generate a mean square error(MSE) value. We use all the MSE values as the threshold, and the threshold value is set. From the minimum to maximum of MSE, each value is searched. As shown in figure 6. When the value is greater than threshold, it is judged as fault. Otherwise, when the value is smaller than threshold, it is determined to be normal.

t	M1_R1	M1_R2	....	MSE	threshold
2012/3/13	134.4	134.4	....	1.3	1.3
⋮	134.4	145	....	⋮	⋮
	......	......	....		
2013/2/12	167	345	....	2.7	2.7

Figure 6. The MSE of data reconstruction transfer to threshold

Use all the MSE values as the threshold, and the threshold value ranges from MSE minimum to MSE maximum. Search for each value. If the value is greater than threshold, it is judged as fault. Otherwise, when the value is less than threshold, it is judged as normal.

C. Results and Evaluations

According to the experimental results, this model has a significant effect on the data of the second group plants and the performance of 5th and 31st factories is the best in the second group. However, the performance of the first and third group plants is not better than the second group. Therefore, we can conclude that the method is not applicable to all plants.

The reason why the second set of factory results will be better than the first and third sets of factories. Because the group distinguish method is based on the number of zones and the number of components, the smaller the value of the two, the less the feature number of the conversion. Therefore, the dimensional features of the first group and the third group are larger than the second group. We can see the figure7.8.9.

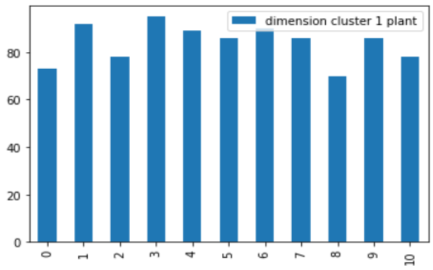

Figure 7. The dimensional feature of the first group of plants

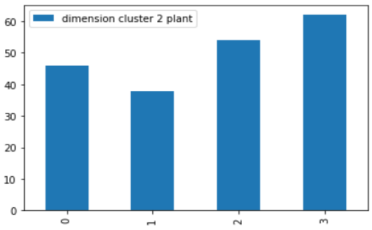

Figure 8. The dimensional feature of the second group of plants

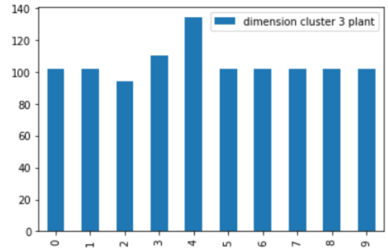

Figure 9. The dimensional feature of the third group of plants

As shown in figure7.8 and 9, it is clear that the second group of dimensional feature is almost below 60. Relatively, the first and third group of dimensional feature is higher than 60. Therefore, dimensional feature is key points that affect reconstruction data. The data dimension is so large that the reconstruction data is not good.

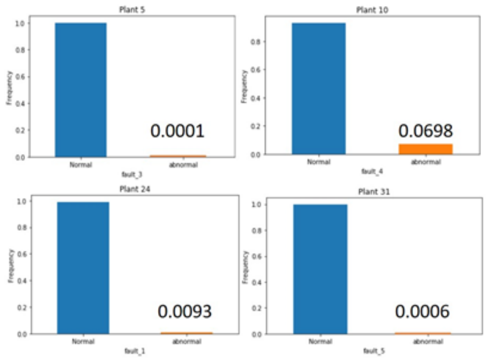

Figure 10. The second cluster plants data distribution

From the experimental results, it is clear that the method in this paper can get good results from the second group. When the data is seriously unbalanced, it also can correctly verify which is the abnormal data.

As shown in figure10, the distribution information of all the second group factories and the data of the 5th plant and the 31st factories are seriously imbalanced data among the plants, but the results of the performance of the two plants are the best in this paper.

As the result, we know that can be seen from the results that the model has a significant impact on the second set of data, and the 5th and 31th sets of plants perform best in the second set. Therefore, based on the result, when the data is more seriously unbalanced, the effect of this method is more significant.

TABLE III. The second group of each plant

Plant 5	Precision	Recall	F-score
RF	1	0.33	0.5
Xgboost	0	0	0
Lstm+Attention	0	0	0
Auto-encoder	0.833	0.666	0.7407

Plant 10	Precision	Recall	F-score
RF	0.8154	0.868	0.8409
Xgboost	0.7041	0.746	0.7244
Lstm+Attention	0	0	0
Auto-encoder	0.467	0.8878	0.6121

Plant 24	Precision	Recall	F-score
RF	0.8695	0.4477	0.5911
Xgboost	0.791	0.3955	0.5273
Lstm+Attention	0	0	0
Auto-encoder	0.6268	1	0.7706

Plant 31	Precision	Recall	Fscore
RF	1	0.33	0.5
Xgboost	0	0	0
Lstm+Attention	0	0	0
Auto-encoder	0.8462	0.9166	0.8799

TABLE IV. The first and the second plants

plant	fault	Recall	Precision	F score	threshold
5	3	0.833	0.666	0.7407	0.000376
10	4	0.8878	0.467	0.6121	0.0001247
24	1	1.0	0.6268	0.7706	1.5256
31	5	0.9166	0.8462	0.8799	0.02192

V. CONCLUSION

In this paper, we proposed a deep learning based algorithm to solve the time series classification. This model has been evaluated on PHM 2015 challenge dataset and it have the good performance on the data.

According to result, we know that the method can be used on serious data imbalance, only normal data and no label situation. In this situation, we have improved the method, which is better than other ways. Using the proposed method in such cases can be more efficient than using the LSTM method, Random Forest and Xgboost.

In the future, we can try to make the auto-encoder model more complicated or change the sequence to sequence auto-encoder and Variational Auto-encoder for the sensor data. Besides, we will also try to find the model which are suitable for any group of plants.

Auto-encoder model shows the necessity of deep architecture for building a robust approach for fault detection.

REFERENCES

[1] Cong Xie, Donglin Yang, Yixiang Huang, and Donglai Sun, "Feature Extraction and Ensemble Decision Tree Classifier in Plant Failure Detection," Annual Conference of the Prognostics and Health Management Society 2015

[2] W. Xiao, "A probabilistic machine learning approach to detect industrial plant faults," 2016, arXiv:1603.05770.

[3] Hyunjae Kim, Jong Moon Ha, Jungho Park, Sunuwe Kim, Keunsu Kim, Beom Chan Jang, Hyunseok Oh and Byeng D. Youn"Fault Log Recovery Using an Incomplete-data-trained FDA Classifier for Failure Diagnosis of Engineered Systems, " Int. J. Prognostics Health Manage., vol. 7, no. 1, p. 004, 2016.

[4] Sepp Hochreiter and Jurgen Schmidhuber, "LONG SHORT-TERM MEMORY," Neural Computation 9(8):1735-1780, 1997.

[5] Rahul Dey∗, Fathi M. Salem"Gate-Variants of Gated Recurrent Unit (GRU) Neural Networks," 2017,arXiv:1701.05923.

[6] Kishore K. Reddy1,Soumalya Sarkar2, Vivek Venugopalan3, Michael Giering4, "Anomaly Detection and Fault Disambiguation in Large Flight Data:A Multi-modal Deep Auto-encoder Approach," Annual Conference of the Prognostics and Health Management Society 2016,

[7] Oyelade, O. J, Oladipupo, O. O and Obagbuwa, I. C"Application of k-Means Clustering algorithm for prediction of Students' Academic Performance," (IJCSIS) International Journal of Computer Science and Information Security,Vol. 7, No. 1, 2010

[8] Rowel Atienza, "Advanced Deep Learning with keras, "

[9] Rosca, J. PHM15 Challenge Competition and Data Set: Fault Prognostics. NASA Ames Prognostics Data Repository, 2015.

[10] Aldrich,C.,&Auret,L. (2013) "Unsupervised process monitoring and fault diagnosis with machine learning methods," Springer.