

Apply Machine Learning to Improve Fault Detection and Classification in Cyber Physical System

Chien-De TSAI and Ming-Chuan CHIU¹

Department of Industrial Engineering and Engineering Management, National Tsing-Hua University, Hsinchu, Taiwan

Abstract. Cyber Physical System (CPS) is one of the key technologies of industry 4.0. CPS is an integrated system that combines computing, sensors and actuators. CPS is controlled by computer-based algorithms that integrates people and internet. However, the performance of CPS is limited by its computational complexity. How to implement CPS with less computational complexity in a real time performance remains an important issue. Therefore, the study attempts to establish a CPS based on machine learning to reduce its complexity and improve applicability. This study applies a Fault Detection and Classification system (FDC) with machine learning method in the virtual subsystem of CPS. This study consists of two steps. Based on the data collected from sensors of CPS, Random Forest (RF) is used to figure out key factors that reduces the data dimension for less computational complexity. The process engineers will know which factors are causing the machine fails. The second step is to establish a time series model which bases on Long Short-Term Memory (LSTM) network to achieve real time monitoring of the machine. Before the machine fails, FDC will alarm that allow the engineer to adjust parameters of machines to avoid machine fails. The study in order to achieve the goals of predicting machines fails, figuring out key factors, predicting maintenance and increasing the productivity.

Keywords. Cyber Physical System, Fault Detection and Classification, Random Forest, Long Short-term memory, Predictive Maintenance

Introduction

Many traditional manufacturing industries use human resource to collect data. First, it probably causes lots of mistakes because of management. Second, it is impossible to monitor the machine downtime timely. In order to reduce labor costs, improve product quality, and optimize processes, the manufacturing industry began to develop into Industry 4.0. Industry 4.0 integrates the enterprise information systems, upload production data to the cloud for storage, analysis, decision-making, big data analysis and calculation with CPS. CPS is an indispensable technology currently in the manufacturing industries. Importing CPS will improve the quality and added value of the product, and its complexity and usability will be improved relatively. CPS is limited by its complexity and usability itself [1].

¹ Corresponding Author, Email, : mcchiu@ie.nthu.edu.tw.

The motivation for this study is to try to establish a method to strengthen CPS. Reducing the complexity of CPS itself and improving usability. It includes how to monitor several parameters and issue alerts for abnormal parameter ranges.

The purpose of this study is to effectively provide decision-making to the shop floor for predicting the downtime. In order to maintain the normal operation of the machine and the smooth production line. FDC allows engineer to monitor the operation of the machine in real time which ensures product yield and determines whether the machine needs to perform maintenance operations through predictive maintenance. When the shutdown occurs in the incubation period, the appropriate production line adjustment can also be carried out first, so that the production line will not stop operating due to the shutdown of one machine and the production capacity will be increased.

At first, this study introduces what is CPS and FDC the applications of FDC, including multivariate analysis and machine learning and discuss what is the predictive maintenance. This study uses the prediction models of FDC and CPS to improve the production capacity and fault diagnosis. The case study of this study utilizes parameters from sensors of CPS to collect data, and employs FDC to analyze the collected parameters to make decisions and compares other previous studies. Conclusions will summarize the contribution and discuss the future research directions.

1. Literature review

1.1. CPS (Cyber Physical Systems)

The term CPS is defined by the 2006 National Science Foundation (NSF). CPS is a mechanism which is controlled by computer-based algorithms that covers complex systems and embedded computing. CPS links between physical devices and computer computing network. In further, CPS provides various data processing services on the Internet. Using the data collected from the physical world from the Design Process to the Manufacturing Process and then reacting to the physical production environment through software operations in order to let each entity and virtual information tightly integrate. Akkaya et al. [2] proposed that CPS can be used to control engineering, mechanics, thermodynamics, sensors, electronics, networking, and software engineering. CPS can capture the values of the variables of its process currently, including (1) The measured variable of sensors (2) The control variable of signals [3]. CPS currently offers an important opportunity for Industry 4.0 and smart manufacturing.

The integration of Industry 4.0 and logistics causes lots of unnecessary waste due to the long waiting time. Zhang et al. [4] used the system in combination with CPS and industrial Internet of Things to establish a model of key factors, to achieve a self-distribution mechanism due to solve the unnecessary waste caused by excessive waiting time. The realization of CPS is called the smart factory, which is the application of manufacturing in IT background. Nikolakis et al. [5] simulated the production process by digital twins in CPS, and the performance would be better than the production process planned by human. While CPS collects different sensor and alarm data from sensors in the factory, fault and error diagnosis is often one of the most

complex problems because most of the problems are very fuzzy and need to be solved according to expert experience [6].

Because of the fact that the problem of failure is fuzzy usually and needs to be solved according to the experience of experts. Expert solution will be very time-consuming. Using CPS to establish a smart model of key factors will effectively solve the long waiting time, but it can not achieve immediate response failure.

1.2. FDC (Fault Detection and Classification)

FDC is a processing system that could detect machine anomalies and production process anomalies. Collect various parameters of the machine to analyze and monitor whether there is any abnormality in the production process. In the event of an abnormality, the machine or production process can be controlled immediately as well as helping the engineer to handle the abnormal machine immediately. With the evolution of process technology, the critical range of the process is relatively stable. Consequently, the importance of FDC and the application of FDC has increased so that the method of multivariate analysis and machine learning become a trend.

Multivariate analysis is an important method to solve pattern recognition problems in machine learning in recent years. Multivariate analysis mainly used to analyze data sets of multiple variables and analyze the relationship between variables of data sets. Multivariate analysis processes the input data set to form a new set which is retaining useful information of data sets and filtering out the noise [7]. Multivariate analysis is an effective method for dealing with multivariate and complex data. Commonly employed multivariate analysis methods include regression, Principal Component Analysis (PCA), Random Forest, Factor Analysis (FA), classification and regression tree, Logistic regression analysis and so on.

Factor Analysis (FA) is a multivariate statistical analysis method. It refers to the dependence of the internal relations of the correlation matrix of the research indicators, and the variables of the complex relationship are reduced to a few unrelated composite factors.

Classification and Regression Tree (CART) is a machine learning method that builds a predictive model from data. The recursive method cuts the data and builds a predictive model. Decision tree analysis uses innovative and powerful statistical techniques to determine important factors [8].

Random Forest (RF) is an integrated algorithm that includes multiple decision tree classifiers. The output categories of RF determined by the number of votes output by the decision tree. Díaz -Uriarte & De Andrés [9] utilized RF algorithm to select features. The feature selection of post-dimensionality reduction data to evaluate its accuracy is better than the accuracy of data before dimensionality reduction. The popularity of random forest algorithms has contributed greatly to a wide range of predictive problems, including nonlinear problems and complex interactions. RF can measure their importance for each variable [10]. Based on random forests, Genuer et al. [11] employed random forest algorithms for variable selection to find important key variables and provide some insight into the importance of variables in random forests.

CPS often has a large number of missing values in the database, however, RF can process a large number of input variables, detect variable interactions and the

importance of variables, and perform accurate classification. Compared with other multivariate methods, RF algorithms estimate missing values is excellent.

Machine learning is one of the information technologies of computer science that used to solve problems in artificial intelligence. The theory of machine learning is mainly to design and analyze algorithms that allow computers to achieve automatically learning.

Abdallah et al. [12] proposed a fault diagnosis system by machine learning method that based on a decision tree algorithm to detect wind turbine faults, damage and abnormal operations. Using a decision tree is easier than other methods to manipulate and interpret, but it requires the ability to capture the features of the input data. One of the branches of machine learning, deep learning, which has the high dimensional adaptability. Majaj & Pelli [13] mentioned that deep learning has attracted the attention of many researchers, but it is incomprehensible due to its complex architecture and data processing in the black-box operation. Deep learning is not entirely suitable for the application of FDC because of unknowing the reason which causes the equipment shutdown.

Predictive maintenance is a information technology based on FDC which maintain the statuses of equipment. Perform monitoring and fault diagnosis of components during the operation of the equipment, confirm the status of the equipment and predict the future development trend of the equipment status. The purpose of Predictive Maintenance is to (1) Evaluate the future state of the equipment. (2) Develop a maintenance plan. (3) Provide information based on dynamic sampling.

Li et al. [14] examine the potential trends of predictive maintenance in big data and CPS environments. Chiu et al. [15] perform predictive maintenance in the factory based on predictive maintenance systems and propose health indicator hierarchy. Liu et al. [16] proposed that predictive maintenance use lots of big data technology. Using the predictive maintenance method of the recurrent neural network, it is possible to find out the data relationships in a specific time series. As a result, life cycle and maintenance time of equipment will be known.

1.3. Summary

This study integrated CPS and big data analysis in Industry 4.0. RF algorithm can process large number of input variables, detect variable interactions and the importance of variables, accurate classification and estimate missing values. Moreover, RF can explain the key parameters of data certainly. Furthermore, using the deep learning to build the prediction model, compare to the traditional statistical methods. First, the required time that build the machine learning model will be reduced generally. Second, the model will achieve accurate predictive maintenance. This is first time that integrate FDC with multivariate analysis, deep neural networks, and CPS to increase productivity.

2. Method

The purpose of this study utilizes the machine learning method to identify the key variables of the process. The predictive model is established to process these key variables and these key variables will be adjusted by CPS in order to increase production, as shown in Figure 1. The method of this paper uses FDC to achieve the

above objectives. The establishment of FDC is mainly divided into two phases. In the first step, random forests are applied to find the key variables of the process. In the second step, the time series model is established by using the key variables that obtained in the first step to perform machine monitoring, predict the time of machines abnormality. At the last, CPS will send the abnormality information from back-end to front-end immediately so that engineers on the shop floor can perform parameter improvement for predictive maintenance.

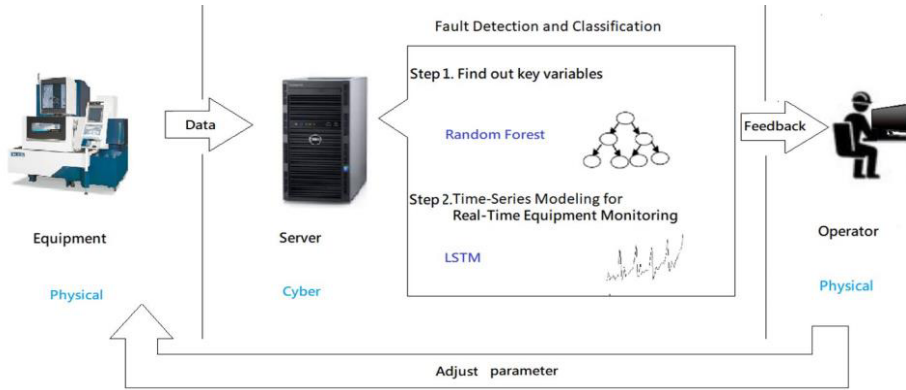


Figure 1. The structure of CPS.

2.1. Random Forest (RF)

Random Forest is an algorithm proposed by Breiman [17]. It uses the binary data to classify, regress, and combines the classification trees into random forest data for generating many classification trees randomly. The results of the summary classification tree, as shown in Figure 2. RF has a significant reduction in the amount of computation, which perform good prediction up to thousands of variables. The classification tree utilizes the largest information gain in the variable to divide each node. RF employs the optimal threshold as splitting each node with randomly generating subsets. Compared with other classifiers, such as support vector machines, RF have strong predictive power for data overfitting. Pal [18] uses RF and support vector machines for land classification. Consequently, the result shows that RF are classified accurately and raining time is better.

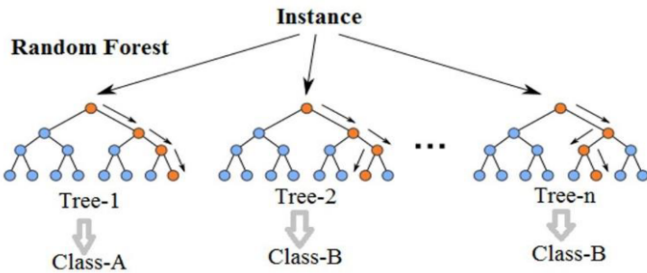


Figure 2. The structure of Random Forest.

2.2. Time Series Model for Process Monitoring

Time series is a sequence that is sorted by time and has the interaction of time with each other. Time series is a method that can reflect the development process, the phenomena of social economic and its developing trends. The time series is a set of observations X_t , and each observation is recorded at a specific time t . In the case of observation at fixed time intervals, the set T_0 of time of discrete time series observations is a discrete set. When recording some observations continuously, use a continuous time series as the time interval, for example $T_0 = [0, 1]$.

Long Short Term Memory Network is one of Recurrent Neural Network (RNN), which proposed by Hochreiter & Schmidhuber (1997). LSTM has high learning performance, long-term preservation of input information, repeated learning, and dealing with nonlinear functional problems. Therefore, LSTM Network is one of the most common deep learning models in sequence data. In addition, because the error gradient function of the neural network propagates through the neural network every time in the recurrent neural network model. However, it will be scaled and the gradient of the error will exponentially explode or disappear with recursion. That is to say nodes at later times reduce the perception of the previous nodes. Thus, LSTM Network models are processed. When predicting time-series data, the predicted performance will be much better than recursive neural network model generally.

Root-Mean-Square Error (RMSE) indicates the standard deviation of the difference between the predicted value and the observed value. The individual differences are called residuals when calculating the sample of the estimated data. It is called the prediction error when using the calculation of the data samples. Namely, RMSE is the difference between the prediction result and the real result. Assume we have n model sample errors ε calculated as $(e_i, i = 1, 2, \dots, n)$. Additionally, uncertainties of observation errors that used to compare models and observations are not considered here [20].

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n e_i^2} \quad (1)$$

Mean Absolute Deviation (MAD) is the average distance between each data and the mean, that is, the absolute value of the deviation of all single observations from the arithmetic mean. It allows us to better understand the actual situation of the error, and the average absolute error measures the accuracy of the prediction by averaging the absolute value of each error. The value of the mean absolute error can be calculated using the following formula [21].

$$MAD = \frac{\sum_{i=1}^n |e_i|}{n} \quad (2)$$

3. Case study

This case study is a semiconductor manufacturing company G company, which main business is research, development, production, manufacturing of silicon wafer semiconductors. Including crystal growth, cutting, polishing and epitaxy. Adopting

advanced process technology, adhering to the quality of its products and paying attention to customer service is company G's manufacturing strategy. In addition, company G has strict quality control in every step, enabling company G to provide stable silicon wafers for a long time. Simultaneously, it can be adjusted according to the specific requirements of customers and delivered to customers with the highest output value. This study analyzed the sensor data of the wire saw machine, including normal data and breakage data of wire. As a result, confirm its key variables according to the following method.

The data collected by the G company shows that the output of slurry temperature detection is taken as an example. The front sensor collects data and then the original signal is converted and passed to the Programmable Logic Controller (PLC) through the transceiver. Moreover, data is transmitted directly to the back-end database with the server by PLC. It will cut from the highest point to the lowest point when

cutting the wafer. After cutting, return to the highest point again. The sensor collects data every minute and has 10172 data approximately in three months. Finally, FDC is used to make prediction on the back end.

There are 20 sensors in the wafer dicing machine, as shown in Figure 3, "9. Wire guide roll bearing temp at fix end 1", "5. Wire guide roll bearing temp at fix end 2" and "13. Wire guide roll bearing temp at fix end 3 " indicates the temperature at the fixed end of the wire that is in different positions. In addition, "12. Wire guide roll bearing temp mov 1", "16. Wire guide roll bearing temp mov 2", and "8. Wire guide roll bearing temp mov 3" indicate that the moving end of the wire is in different positions. Furthermore, other numbers, "6" means Slurry temp wire web in front, "7" means slurry out let temp, "4" means Chilled slurry outlet temp, "1" means Chilled water valve position, "20" means Chilled water Inlet temp, "10" means Slurry mass flow at the front side, "11" means Slurry mass flow at the rear side, "14" means Slurry density, "3" means Wire speed, "2" means left wire tension and "18" means right wire tension.

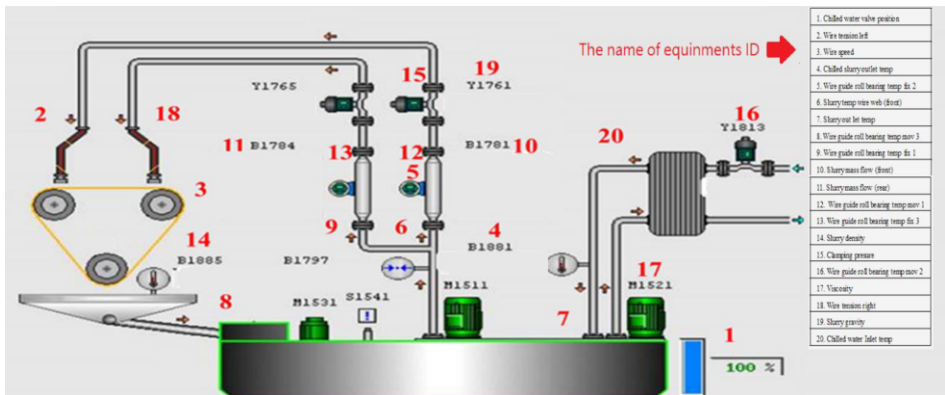


Figure 3. The wire saw machine.

In this study, after using the RF algorithm for importance testing of key variables, the first 7 variables were selected from the 20 variables according to the value of the Increased node purity (IncNodePurity). IncNodePurity is related to the loss function. The best choice for dividing tree based on the loss function is an indicator that is evaluating the importance of parameters after obtaining a model from a random forest. In this study, the first seven key variables were selected from 20 variables according to IncNodePurity greater than 0.05.

Table 1. The importance of variable.

Variable Name	IncNodePurity
1. Clamping pressure	0.3316
2. Slurry density	0.1012
3. Slurry gravity	0.1011
4. Wire speed	0.0946
5. Position	0.0906
6. Chilled slurry outlet temp	0.0609
7. Slurry temp wire web (front)	0.0510

The results of this study are shown in Table 1. It is found that the clamping pressure of the wafer cutting machine is the biggest factor causing the wire breakage. In addition, the influence of the slurry is also very large for the wire breakage, including the slurry density and the slurry gravity. The slurry temperature includes the temperature of the cooling slurry outlet and the temperature of the slurry in front of the wire web. The speed of the wire and the position of the ingot are also one of the factors that affect the machine shutdown.

This study utilizes the Keras suite with the Python language to build a LSTM network with 50 neurons in the hidden layer and 1 neuron in the input layer for predicting downtime. The batch size is 72

and training for 100 rounds, the learning rate is 0.01, and the adaptive moment estimation algorithm- Adamive Moment Estimation (Adam) is used to evaluate the model with loss function - Mean Absolute Error (MAE) and Mean Squared Error (MSE) as shown in Figure 4 and Figure 5 respectively

The error analysis employs the RMSE of the data set in the training model, and R^2 of the model has the ability to interpret the prediction of the model. The error analysis is used to observe RMSE of the data in the training model, that is, R^2 provides a measure of how well observed outcomes are replicated by the model. The best of R^2 is 1.00 or 100%. 70% of R^2 has enough ability to represent models (Leinweber, 2007). The R^2 of the training and test data is 0.93 and 0.81, both of which are greater than 0.7 in Table 2. Briefly, it indicates that the model has excellent performance.

Table 2. Accuracy analysis.

	Train Data	Test Data
MAE	0.2574	0.3072
MSE	0.1405	0.2059
R^2	0.93	0.81

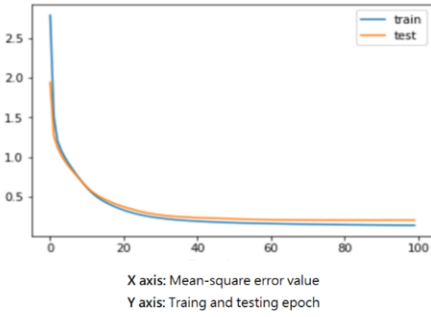


Figure 4 MSE of LSTM model.

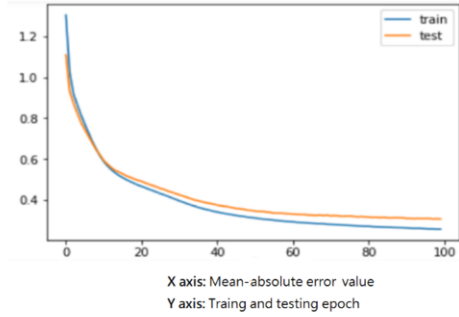
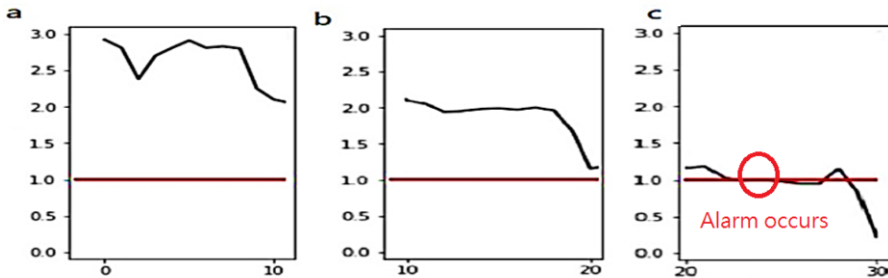


Figure 5 MAE of LSTM model.

Taking the case of once shutdown for test by using the established FDC. FDC shows the predicted results of the machine status from normal to failure. The X axis of the system is the time unit, and the Y axis is the predicted value. If the predicted value of Y is close to 0, it means that the machine is likely to failure. Therefore, it can inform the company in advance to prepare for such an emergency and to maintain the repair machine.

In order to achieve the detection error, the data collected by the machine is put into FDC every ten minutes. The X axis is the time unit, the Y axis is the predicted value. In addition, 3 is the remaining life cycle of machine is the most, the closer to 0 that means the fewer life cycles remaining. The red line in the system indicates that will issue an early warning when the predicted value is 1. The error detection every ten minutes on the machine of the length of time from start to finish, as shown in Figure 6.

In Figure 6, c is about 23 minutes and the Y value is lower than 1. At this time, FDC will issue a warning, and the actual downtime of the wire saw machine is 43 minutes after the machine startup. It will be predicted that the machine will be failure within 20 minutes. At the same time, the system uses the analysis of the important parameters of the random forest in the first stage to find out the key variable of the fault. The accuracy of the prediction is 0.81, which enables the engineer to perform predictive maintenance and evaluate the risk of machine failure.



X axis: The state of prediction. Y axis : Time (min)

Figure 6 Error detection every ten minutes on the machine.

To analyze the cause of the machine's shutdown, many studies use key variables to identify. According to Zhang et al. [4], who established an intelligent model of key variables which realized the method of self-distribution and solved the excessive

energy consumption caused by the long waiting time. Niu et al. [8] used multivariate analysis techniques to establish a model of decision tree due to identify the most important predictor. Nevertheless, the case study presented here that the transmission of wireless signals in CPS sometimes has missing values. In this study, we overcome the drawback of the transmission of wireless signals in CPS. This study employs a RF algorithm at first that avoids the occurrence of missing values. RF algorithm is fitted to find the key variables which also can show better predictive power. Furthermore, this case study presented also is one other evidence that the data received from sensors of CPS is the sequential data which is a time series. LSTM network is very suitable for processing high-dimensional data because LSTM network have good adaptability to the sequential data. Attesting to this, Zhao et al. [22] employed LSTM networks for fault identification and analysis in chemical processes and showed that different faults could be better separated. Lin et al. [23] utilized LSTM network to accurately estimate faults and remaining life of aero engines in case of mixed faults and noise. However, in this case study, besides estimating the missing values of the transmission of signals in CPS, the cause of the faults need to be identified specifically. This study employs a RF algorithm and a neural network of LSTM to establish FDC system in order to prevent the occurrence of abnormal parameters of the machine and identify the key variables of the machine shutdown. The efficiency is faster in order to achieve timely detection and clearly distinguish the cause of the problem.

4. Conclusion

CPS and FDC are trends in Industry 4.0. This study is the first paper to integrate multivariate analysis techniques, LSTM networks, FDC and CPS to (1) Predict the downtime (2) Increase productivity (3) Key variables for fault diagnosis (4) Predictive maintenance. The research method uses the random forest to find the key variables in first step, matches LSTM network which has the high-dimensional adaptability in second step. Consequently, it shortens the model convergence time and improves the accuracy of the prediction. Furthermore, the study utilizes CPS to report the situation to the engineer on the shop floor immediately. Significantly reduce time waste, quickly and accurately provide the information to engineers in order to increase the company's production capacity. Finally, provide key variables the visualization of key variables and error detection on the results due to ensure the conclusions of this study.

Future research can be combined with statistical analysis techniques to verify key variables and analyze the type one error and type two error. Moreover, the data used in this study case is the data collected by the three-month sensor and can be collected longer in the future. Long time data will more confine the time series model, and the accuracy of prediction ability will be better. The last, an appropriate machine decay model - Degradation Mode will be established to evaluate the health indicators and degradation status of the wafer machine to achieve predictive maintenance of the machine.

References

- [1] A. De Carolis, G. Tavola and M. Taisch, *Cyber physical systems in manufacturing: Future trends and research priorities, 21st Summer School Francesco Turco*, 2016, pp. 12-17.

- [2] I. Akkaya, P. Derler, S. Emoto and E.A. Lee, Systems engineering for industrial cyber– physical systems using aspects, *Proceedings of the IEEE*, 2016, Vol. 104, Issue 5, pp. 997-1012.
- [3] M. Krotofil and A.A. Cárdenas, Resilience of process control systems to cyber-physical attacks, In *Nordic Conference on Secure IT Systems*, Springer, Berlin, Heidelberg, 2013, pp. 166-182.
- [4] Y. Zhang, Z. Guo, J. Lv and Y. Liu, A Framework for Smart Production-Logistics Systems based on CPS and Industrial IoT, *IEEE Transactions on Industrial Informatics*. Sept. 2018, Vol. 14, Issue 9 , pp. 4019 - 4032
- [5] N. Nikolakis, K. Alexopoulos, E. Xanthakis and G. Chryssoulouris, The digital twin implementation for linking the virtual representation of human-based production tasks to their physical counterpart in the factory-floor, *International Journal of Computer Integrated Manufacturing*, 2019, Vol. 32, Issue 1, pp. 1-12.
- [6] D. Kim, S.C. Han, Y. Lin, B.H. Kang and S. Lee, RDR-based knowledge based system to the failure detection in industrial cyber physical systems, *Knowledge-Based Systems*, 2018, Vol. 150, pp. 1- 13.
- [7] S. Muñoz-Romero, V. Gómez-Verdejo and E. Parrado-Hernández, A novel framework for parsimonious multivariate analysis, *Pattern Recognition*, 2017, Vol. 71, pp. 173-186.
- [8] X. Niu, G. Liu, L. Huo, J. Zhang, M. Bai, Y. Peng and Z. Zhang, Risk stratification based on components of the complete blood count in patients with acute coronary syndrome: A classification and regression tree analysis, *Scientific reports*, 2018, Vol. 8, Issue 1, P2838.
- [9] R. Díaz -Uriarte and S.A. De Andres, Gene selection and classification of microarray data using random forest, *BMC bioinformatics*, 2006, Vol. 7, Issue 1, P3.
- [10] C. Strobl, A.L. Boulesteix, A. Zeileis and T. Hothorn, Bias in random forest variable importance measures: Illustrations, sources and a solution, *BMC bioinformatics*, 2007, Vol. 8, Issue 1, P25.
- [11] R. Genuer, J.M. Poggi and C. Tuleau-Malot, Variable selection using random forests. *Pattern Recognition Letters*, 2010, Vol. 31, Issue 14, pp. 2225-2236.
- [12] I. Abdallah, V. Dertimanis, H. Mylonas, K. Tatsis, E. Chatzi, N. Dervilis, K. Worden and E. Maguire, Fault diagnosis of wind turbine structures using decision tree learning algorithms with big data, in: S. Haugen, A. Barros, C. van Gulijk, T. Kongsvik and J.E. Vinnem (eds.) *Safety and Reliability–Safe Societies in a Changing World*, 2018, Chapter 382, pp. 3053-3061.
- [13] N.J. Majaj and D.G. Pelli, Deep learning—Using machine learning to study biological vision. *Journal of vision*, 2018, Vol. 18, Issue 13, P2-2.
- [14] Z. Li, K. Wang and Y. He, Industry 4.0: potentials for predictive maintenance, *International workshop of advanced manufacturing and automation*, Manchester, 2016,
- [15] Y.C. Chiu, F.T. Cheng and H.C. Huang, Developing a factory-wide intelligent predictive maintenance system based on Industry 4.0. *Journal of the Chinese Institute of Engineers*, 2017, Vol. 40, Issue 7, pp. 562-571.
- [16] Y. Liu, Y. Hu, R. Zhou and J. Wen, An approach based on improved grey model for predicting maintenance time of IPS2, *Procedia CIRP*, 2016, Vol. 47, pp. 204-209.
- [17] L. Breiman, Random forests, *Machine learning*, 2001, Vol. 45, Issue 1, pp. 5-32.
- [18] M. Pal, Support vector machine - based feature selection for land cover classification: a case study with DAIS hyperspectral data, *International Journal of Remote Sensing*, 2006, Vol. 27, Issue 14, pp. 2877-2894.
- [19] S. Hochreiter and J. Schmidhuber, Long short-term memory. *Neural computation*, 1997, Vol. 9, Issue 8, pp. 1735-1780.
- [20] T., Chai and R.R. Draxler, Root mean square error (RMSE) or mean absolute error (MAE)— Arguments against avoiding RMSE in the literature, *Geoscientific Model Development*, 2014, Vol. 7, Issue 3, pp. 1247-1250.
- [21] U. Khair, H. Fahmi, S. Al Hakim and R. Rahim, Forecasting Error Calculation with Mean Absolute Deviation and Mean Absolute Percentage Error, *Journal of Physics: Conference Series*, 2017, Vol. 930, Issue 1, P.012002.
- [22] H. Zhao, S. Sun and B. Jin, Sequential Fault Diagnosis Based on LSTM Neural Network. *IEEE Access*, 2018, Vol. 6, pp. 12929-12939.
- [23] J. Lin, L. Su, Y. Yan, G. Sheng, D. Xie and X. Jiang, Prediction Method for Power Transformer Running State Based on LSTM_DBN Network. *Energies*, 2018, Vol. 11, Issue 7, P1880.