Design and implementation of a CPS-based predictive maintenance and automated management platform

Huei-Yung Lin1, Fang-Ning Yang2

1Department of Electrical Engineering, Advanced Institute of Manufacturing with High-Tech Innovation, National Chung Cheng University, 168 University Road, Chiaiyi 621, Taiwan
2Department of Electrical Engineering, National Chung Cheng University, 168 University Road, Chiaiyi 621, Taiwan
E-mail: lin@ee.ccu.edu.tw

Abstract: The concept of cyber-physical systems (CPSs) is a core technology indispensable to Industry 4.0. The development of an intelligent automatic operational management platform is presented in this study. The platform is used to collect on-site data, provide predictive maintenance, and monitor and control the mobile robots. For predictive maintenance, it is targeted at products from manufacturing big data to cloud computing, and predictive maintenance for all factories around the world. The efficiency of the management system is improved by the performance evaluation strategy. In the experiments, the authors implement and discuss the predictive maintenance platform with generated synthetic data and real-world data collected from a gearbox plant. The results show that the important variables possess large drops according to mean decrease Gini. A warning generation for battery simulation and manufacturing execution is also presented to demonstrate the proposed system design. The results show that the discharge curve drops sharply at the last stage, and the warning value can be set accordingly.

1 Introduction

In Hannover Industrial Fair in 2011, Germany first proposed the concept of ‘Industry 4.0,’ the United States, Japan, China, South Korea, and many other countries also introduced their advanced manufacturing strategies [1, 2]. Its main features were intelligence and objectivity, and the major purpose was to combine traditional industrial production with modern information technologies. Under the original automation technology and architecture, it realised the transformation from centralised control to the basic mode of decentralised enhanced control, and established a highly flexible, personalised production model that integrated products and services. In this mode, production automation technologies can make the equipment smarter through self-diagnosis and self-correction to better assist workers in production [3]. Thus, the communication function and integration ability of the automation equipment are required to be stronger, and the automation requires more powerful analysis, processing and data sharing capability for the enterprise.

The cyber-physical system (CPS) [4–6] is a core technology indispensable to Industry 4.0. It is an integrated control system combining various sensors, actuator devices, and computing power. It can also be thought of as an intelligent system which is capable of autonomous thinking, acquiring data from the environment, searching and mining useful information, and controlling devices in accordance with the system. These behaviours in term have a controlled impact on the environment. Cyber is the discrete and logical management of computing, communication, and control systems. It integrates the internet of things (IoT), cloud computing, and big data analysis to realise smart factories with adaptability and resource efficiency in Industry 4.0.

Since the emerge of the Industry 4.0 concept, the investigation of CPS has been an important research and development objective in both the engineering and applied science in the world and a priority in the industries. The CPS-based automated management platform in the substantive process can be embedded in the main functions of computing and communications such that the substantive process can interact with the information systems. This concept is widely applied to many different areas, for instance, the national smart grids, single devices such as pacemakers, and agricultural irrigation networks. Generally speaking, there are several sub-systems contained in a CPS which is also commonly a hetero-system. The sub-systems might be constructed and deployed differently. It is also possible to be scattered in many locations and their tasks are coordinated by a smart network [7].

In the era of big data, it is a trend for people to use database materials to create competitiveness. People begin to accept the hidden meanings from the data, hoping to bring more value, and often can analyse the possible reasons from the results of the data. Thus, this study aims to apply the manufacturing execution system (MES) to the daily process for abnormality collection. We use big data analysis and cloud computing to investigate whether the abnormalities between the processes are related and the key impact factors of product anomalies. We extract valuable potential knowledge and provide engineers with the means to follow and solve problems in the next encounter with the same or similar process anomalies. In addition, there is an opportunity to improve production yields by using timely detection, prevention and prediction of process anomalies that have occurred.

The digital plant is a concept deeply rooted in many aspects of core business philosophy. Its idea is to employ the information technology to integrate the production digitally. In the existing literature, production planning control systems with the computer-aided design are usually built around a relatively fixed central control system. On the other hand, the basic concept of this work is that the concentration reduction of the product is important and should be decentralised. To make the production process more open and more flexible, one way is to link the manufacturing equipment and the Internet by increasing the self-control of production equipment through the CPS. As a result, more optimisation possibilities can be provided. Furthermore, the efficiency of production is also able to be effectively improved. Therefore, one of our main objectives is to construct a virtual platform which is able to connect and communicate with the real-world objects. Yet another important goal is to construct an independent system through the CPS-based automated management platform. It will be served as a standard test platform for many other research purposes.

The main objective of this work is to find out the information of the hidden abnormalities in the process of defective products, and to improve the yield, and finally to predict the key impact factors of defective products. We develop a predictive maintenance big
data platform from the perspective of Industry 4.0 smart factory. Therefore, the following three points are addressed in this study:

i. Based on the IoT, CPS, big data analysis, and cloud computing, we propose and implement a predictive maintenance big data platform for Industry 4.0.

ii. Analysis module in the cloud: Using massive data technology, we adopt the offline association rule algorithm and offline predictive maintenance to quickly identify the relevance of parameters and the key parameters in the process. Finally, we summarise the valuable rules or knowledge to provide a reference for decision makers.

iii. Compared to the previous methods, we consider not only finding a single problem factor, but also the correlation between multiple factors.

2 Related work

By the general definition, CPS is an integrated system constructed with the integration of computer, communication, and control technologies [8–11]. It should be able to instantly sense and control any engineering and information interactive systems, and provide the related and necessary information services. From the perspective of industrialisation and informatisation, the significance of the CPS is its capability to deal with different physical devices and integrate them into one network system. Moreover, the physical equipment are able to access the functions on instant messaging, data computing, precise control, self-management, and remote coordination through the transformation of the information and communication systems. Therefore, a CPS can be thought as a network system which also possesses the control system capabilities. However, they have more functions than the common control systems such as sharing, openness, and information exchange [12, 13].

As shown in Fig. 1, a CPS in the entire factory is a monitoring station. In the case of non-interference with the internal information of the factory, the data is handed over to the console via wireless communication (e.g. Wi-Fi). It is then transmitted from the main control station to the monitoring station, and receives data from the AGV. In this way, the production history will be generated, and finally analysed by big data, predictive maintenance, and early warning monitoring.

An MES [14] provided the instant information that was transmitted from the production line to the completion of the product, and the product was delivered on the production line. This included product scheduling, production dispatching, equipment monitoring/detection repair, direct and indirect production staff coordination management, manufacturing rule management, detailed material control, and actual cost tracking. In addition, every important processing event on the production line was traced. On one hand, it could be used as the production traceability of the product, and the electronic data could also be used to trace the product of the production process. On the other hand, when the number of production experiences increases, the data could be isolated later.

Predictive maintenance utilises the actual operating conditions of equipment, materials, and systems to optimise manufacturing operations. A standard predictive maintenance management program utilises a combination of most cost-effective tools such as vibration monitoring, process parameter monitoring, temperature recording, tribology, and visual inspection to obtain the actual operating conditions of critical plant systems. These actual manufacturing data are used to schedule all maintenance activities [15].

One of the main requirements for effective predictive maintenance is to obtain a sufficient amount of data from various parts of the manufacturing process. This is also the main drawback of predictive maintenance implementation in manufacturing. The greater the amount of data, the higher the accuracy of the maintenance interval prediction for machines, materials, tools, and any other critical components in the manufacturing process. The increase in the amount of manufacturing data not only brings about an increase in prediction accuracy, but also leads to increasing demand for data processing. This work implements statistically based predictive maintenance and improvements. Our message from all anomalous products helps to develop statistical models for predicting anomalies and preventive maintenance strategies.

There are quite a few mature tools in the market available for cloud computing [16], and a variety of management and data analysis tools such as Microsoft Azure [17], Google Cloud Platform [18, 19] etc. However, if the system has to maintain a flexible architecture on the platform, it is required to build the system through a customised interface. The devices and interfaces use Wi-Fi as a bridge for communication, and the collected data are built on the database and manipulated using SQL after embedded with equipment independent codes. In general, the simulation platform is used to establish the mirror module of each subsystem [20], and develop a set of management algorithms to optimise the management unit configuration. These simulation platform and algorithms can be flexible in different system architectures. Based on the Bayesian prediction model framework, the early warning system uses the historical data and the statistical probability of accidents to calculate the possible instants of the events.

In general, manufacturing big data [21] can be divided into three categories: device data, product data, and command data. Using a cloud platform, the collected manufacturing big data can be further analysed and used to create innovative applications such as active preventive maintenance, line optimisation, and energy optimisation. However, these applications do not take into account the active preventive maintenance in intelligent manufacturing for product data in manufacturing big data. In this work, we focus on finding the hidden information from the defective products to predict the key impact factors. A predictive maintenance big data
platform is developed and illustrated from the perspective of smart factory.

3 System architecture

As described in the previous section, there are proposals for the analysis of different types of manufacturing big data, and there are also many new methods for predictive maintenance of manufacturing. However, we are unable to discover from the latest technical literature that it is possible to implement a complete big data solution for predictive maintenance of product data. Therefore, the motivation of this work is to help implement the big data environment through the existing reliable predictive maintenance methods. Moreover, we can execute in the cloud and expand the system framework according to the requirements.

3.1 Problem statement

Before presenting a big data solution, it is important to clarify and define the problem, confirm the final goal and direction. We deal with this problem with three aspects:

i. For manufacturing plants, the goal is to find the hidden correlation information about the problem-related relevance of defective products in the process, and identify key factors to improve the yield.

ii. As the automation of manufacturing plants is introduced and the number of electronic products increases, more and more data are generated. Many enterprises need to add more resources for maintenance, and thus the overhead is increased.

iii. The predictive maintenance based on traditional technologies requires each plant to have its own individual system. Since each factory operates independently, the enterprise is not easy to maintain.

3.2 Cloud-assisted architecture

Under the cloud architecture, we break the geographical restrictions and remotely monitor the sensor data around the world, as shown in Fig. 2. Designing a predictive maintenance big data platform requires the establishment of a smart factory, and need to consider three aspects: (i) data collection, (ii) data persistence, and (iii) data processing. This is to enable predictive maintenance systems to be distributed remotely around the world, as well as to provide high performance, scalability, and fault tolerance features.

Data collection: This work uses traditional data preprocessing to collect data, which mainly includes extract, transform, and load (ETL). The ETL representative data is extracted from the data source, transformed, and then loaded into the destination data transmission process. The destination is usually a data repository for business intelligence analysis. We use Apache Hive [22], a data warehousing tool built on top of the Hadoop architecture.

Data persistence: This work uses the Hadoop [23] distributed file system (HDFS) for data persistence. HDFS is fault tolerant and scales horizontally. So even if anything goes wrong, HDFS will continue to load large amounts of data from the manufacturing plants in a distributed manner.

Data processing: This work uses Apache Spark [24] for data processing. Spark is a computational framework based on memory. When it is operating, the data generated in the middle is temporarily stored in the memory, so the execution speed can be accelerated. According to the official documents, it can process data 100 times faster than traditional approaches. In addition, Spark is an extensible system that can scale to hundreds or thousands of machines in a fault-tolerant manner. These features make it ideal for processing and analysing large amounts of data collected from manufacturing plants.

The design core of the system: First, the data preprocessing includes the ETL steps. The Hive decentralised database is used in conjunction with HDFS [25] to implement data storage and retrieval. Then Spark is used to build a huge data processing mechanism. Finally, a predictive maintenance system is achieved based on the process parameter correlation analysis and the process parameter critical analysis through the huge data exploration technology. The architecture is shown in Fig. 3.

4 Data processing in the cloud

As described in Section 3, this work proposes a cloud-assisted architecture, and now we present the product data processing in the cloud under this architecture. There are two types of active maintenance based on manufacturing big data: offline association rule algorithm and offline predictive maintenance. The former method can be used to find the implied association information in the process based on the product data from the process. The latter method is suitable for finding potential key factors for anomalous problems and focuses on predicting anomalies in the product process. The goal of both the methods is to find out the hidden reasons which cause defective products. The overall system for data analysis and process is shown in Fig. 4.

4.1 Offline association rule algorithm

When a product is abnormal during the process, it is judged to be defective after tested. The product is then often eliminated or repaired from scratch. Usually, we do not know what the problem is, and use general classification methods to find out a possible factor. However, in terms of actual production lines, many times the occurrence of anomalies is not only a problem in one aspect, but many of them are related. There are often many hidden factors or combinations leading to abnormalities. In this work, we use the association rule algorithm to analyse and explore the correlation among the data fields from a huge dataset, and derive the association rules among them. This can be used to more extensively analyse the correlation between the information in the
Agrawal et al. [26] proposed the Apriori algorithm for association rule analysis, which contains three important values: (i) support, (ii) confidence, and (iii) lift. They are used to measure and judge whether the associated association rules are meaningful. The three metrics for the association rules are detailed below:

**Support**: The probability of the intersection of premise event $A$ and result event $B$, as defined by (1). It is used to show the breadth of this rule.

$$\text{Support}(A \rightarrow B) = P(A \cap B) \quad (1)$$

**Confidence**: The conditional probability that event $B$ occurs when the premise event $A$ occurs, as defined by (2). It is used to show the correctness of this rule.

$$\text{Confidence}(A \rightarrow B) = P(B|A) = \frac{P(A \cap B)}{P(A)} \quad (2)$$

**Lift**: The ratio of the probability of the rule of confidence to the outcome of event $B$ alone, as defined in (3). It is used to show the necessity of the existence of premise events in the rules. $\text{Lift} > 1$ means that the premise event $A$ has a significant effect on the occurrence of the resulting event $B$.

$$\text{Lift}(A \rightarrow B) = \frac{P(A \cap B)}{P(A)P(B)} = \frac{P(B|A)}{P(B)} \quad (3)$$

The main process includes two steps:

* Find all the frequent itemsets: Candidate itemsets are generated in a joint manner, and frequent itemsets are generated in a prune manner. Search the database repeatedly until all the frequent itemsets are found.

* Derive association rules from the frequent itemsets: The rules must be frequent itemsets, and meet the minimum support and minimum confidence given by the decision maker.

### 4.2 Offline predictive maintenance

The above-mentioned offline association rule algorithm is used to obtain the relevant parameters of the highest rule. The next is to find the key factors and obtain the importance evaluation of its parameters. We set up an offline predictive maintenance system and finally confirm whether the key factors are consistent with the associated parameters. Random forests (RFs) are used to obtain key factors because RFs are fault tolerant and can ignore missing values when estimating the missing data. Even if there is a large amount of data missing, RFs can better maintain the accuracy.
RFs are the ensemble method proposed by Breiman and Cutler [27]. This method uses a classification and regression tree, and uses random sampling to select data samples and data features when constructing each decision tree. The basic structure is shown in Fig. 5. First, the Bagging method is used to randomly sample $N_{\text{tree}}$ samples from the training dataset as a subset of the data, and the sample size of each sample is the same as the training set. Then $N_{\text{tree}}$ decision tree models are established for $N_{\text{tree}}$ samples, and $N_{\text{tree}}$ classification results are obtained for the decision tree classifier. The results of the $N_{\text{tree}}$ classification are used to vote, and the classification prediction results are determined by the majority decision.

The randomisation method is as follows:

- Each tree naturally divides and grows without pruning.
- Randomly select sample training sets: Use the Bagging method to form a training set for each tree.
- Randomly select the classification sample features: randomly select a certain number of sample features from all features, and then select the best classification method.

Random selection of sample data means that the tree in the RF is generated by Breiman’s Bagging, Bootstrap Aggregating [28]. It selects $N$ sub-trainset samples from $N$ data by means of uniform random number, and randomly selects the data used by the sub-training set from the original training dataset each time, and each selected data will be returned to the training dataset. Therefore, some data will be re-selected. The probability that each sample is not selected after $N$ sampling is about 0.368, as shown in the following equation:

$$\left(1 - \frac{1}{N}\right)^N \approx e^{-1} = 0.368 \quad (4)$$

Overall, approximately one-third of the training materials are not selected at the end of the sample. These data are called Out-Of-Bag Data (OOB Data). It can be used to assess the importance of the classification tree for the importance of the input factor, measure the training results of this RF, and assess the strength and correlation of the forest.

The randomly selected sample features refer to the sample characteristics selected by each tree in the RF, and the $m$ value is given by the user (the $m$ value must be less than the $M$ value). The $m$ sample features are selected from the $M$ features in a uniform random manner, and the selected sample features are no longer placed back into the sample, so the selection is not repeated. If there are $N$ data and $M$ sample features (variables, attributes), the whole algorithm process can be briefly described as follows:

1. Determine the parameter item: the total number of trees in the forest $N_{\text{tree}}$, the sample feature $m$ used for each tree.
2. Use Bagging to select a training subset from the training data by repeating $N$ times sampling, and randomly select $m$ sample features for each subset.
3. Construct a decision tree by taking the data selected in Step 2 as a sub-training set.
4. Repeat Steps 2 and 3 for $N_{\text{tree}}$ times to construct an RF model with $N_{\text{tree}}$ decision trees.
5. When predicting and classifying the verification data, the $N_{\text{tree}}$ trees are predicted separately, and the final prediction results are synthesised by voting.

The OOB data can be used to replace the test set error estimation method. Its error is calculated as follows: For RFs that have been generated, the performance is tested with the OOB Data. Assuming the total number of the OOB Data is $a$, they are used as the input to bring in the RF classifier that has been generated before. The classifier will give the corresponding classification of $a$ data because the type of this $a$ data is known. The correct classification is compared to the results of the RF classifier. The number of classification errors is set to $b$, then the OOB Data error size is equal to $b/a$.

Next, the way we estimate the importance of variables in a RF model is to calculate the node purity. The node purity is based on the Gini value, which measures the total reduction in node impurities in each split of the measurement tree and averages the results of all trees. The total reduction of the node is the mean decrease Gini (MDG). The MDG calculates the Gini impurity of each tree according to the formula, and averages the results of all trees. Thus, we need to calculate the Gini impurity. At the classification node $t$, the Gini impurity is given by

$$G(t) = 1 - \sum_{k=1}^{c} p^2(k|t) \quad (5)$$
The Bagging method is used to randomly from the training dataset as a subset of the data. We find the relevant parameter rules that are useful when the exception code is generated, as shown in Table 2.

Table 2 Association rule parameter selection. We find the rules related to the occurrence of each anomaly code through association rules

| Abnormal code | Minimum support | Minimum confidence | Minimum parameter | Number of rules |
|---------------|-----------------|--------------------|-------------------|-----------------|
| error 1       | 0.00155         | 0.01               | 8                 | 8               |
| error 2       | 0.001355        | 0.01               | 8                 | 8               |
| error 4       | 0.00125         | 0.01               | 8                 | 8               |
| error 4       | 0.001137        | 0.01               | 8                 | 9               |

where \( e \) represents the total number of categories for the target variable, \( p(k|\pi) \) represents the conditional probability where the target variable is the \( k \)th class in node \( t \).

We can obtain an estimate of the importance of the calculated variables for each parameter. The lower the Gini impurity of the parameter, the smaller the uncertainty of the parameter. That is to say, the higher the MDG value, the higher the node purity, and the higher importance of the parameters, so that we can obtain the specific indicators of the key factors.

5 Case study analysis I

The Cloud Layer is based on Microsoft Azure HDInsight's Microsoft R server combined with enterprise R-analysis software, the Apache Hadoop ecosystem framework and the Apache Spark ecosystem framework. The MES collects a lot of information, and we are interested in the gearboxes containing relevant data collected in the manufacturing process of good and bad products, sharing a total of 388,744 data for case study analysis.

5.1 Design and implementation of data generation module

As shown in Table 1, we set the data format in accordance with the manufacturing process of the gearbox. The simulation data are generated randomly using the Gaussian distribution. When the simulation data are generated, we detect the current consumption, positive and negative rotational speeds, and product weight. In addition, in order to increase the authenticity of the work, we assume that the yield is 90% according to the manufacturing statistics.

If the random variable \( x \) obeys a probability distribution with a positional parameter \( \mu \) and scale parameter \( \sigma \), then the probability density function is given by

\[
f(x) = \frac{1}{\sigma \sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}
\]

The mathematical expectation value \( \mu \) of the Gaussian distribution is equal to the positional parameter, which determines the position of the distribution; the standard deviation \( \sigma \) is equal to the scale parameter, which determines the magnitude of the distribution.

5.2 Results of the offline association rule algorithm

If the support level and the confidence level are set too low during the association rule analysis, too many association rules will be generated, and it takes more time to find useful rules. On the other hand, if the setting is too high, the rules found are biased toward common sense. In this work, we find the rules related to the occurrence of each anomaly code through association rules, and find the relevant parameter rules that are useful when the exception code is generated, as shown in Table 2.

On the whole, we can obtain the relevant parameters of the highest rule from the association rules very clearly. We can see from the results that the employee class staffGroup and the week weekend are strongly related. It means that the employees are different in efficiency on different working days, especially the weekly products are most likely to cause abnormalities. In addition, we can also know from the results that the part batch number supplierLog is also the main cause of product abnormality. The batch number of the same batch is easy to cause abnormality in the products shipped by the factory (F) and the product model (Model).

5.3 Results of the offline predictive maintenance

We calculate the importance of the variables for the stochastic model. As shown in Table 3, we mainly observe according to MDG, and more the drop, more important the variable. The same is also true for the node purity. Looking at the top five parameters of high importance, we can see that the important indices of the two parameters staff and productLog are much higher than other parameters. The three parameters, staffGroup, Date, and supplierLog, are parameters with relatively high importance indices.

5.4 Discussion of integration results

When integrating the results of association rules and RFs, we can find that the associated parameters of the highest rule obtained by
Table 3  Importance of calculating variables. We calculate the importance of the variables for the stochastic model

| Variable     | Error 1 | Error 2 | Error 3 | Error 4 |
|--------------|---------|---------|---------|---------|
| staff        | 81.0107 | 55.60058| 125.61148| 88.44953|
| productLog   | 73.52063| 100.04113| 201.23617| 98.06811|
| staffGroup   | 48.33080| 29.45665| 61.81928| 54.19893|
| date         | 43.12649| 53.53120| 74.62358| 64.20019|
| supplierLog  | 38.29331| 31.67822| 61.07547| 43.07972|
| weekend      | 18.29483| 25.48112| 36.34361| 21.07829|
| supplier     | 16.73032| 7.76779 | 16.68718| 13.03682|
| station      | 12.87603| 8.02184 | 16.07365| 17.27101|
| material     | 7.62854 | 3.09614 | 6.07693 | 6.05201 |
| model        | 5.19538 | 9.46672 | 7.54784 | 6.18649 |
| operation    | 4.72184 | 2.30622 | 4.42951 | 6.21596 |
| factory      | 1.97559 | 1.98220 | 2.66628 | 5.85533 |

Fig. 6  IoT platform of the proposed system. It consists of three parts: the evaluation, early warning, and mirror modules

Fig. 7  Yurdakul's multi-criteria of manufacturing. The manufacturing performance is classified by five criteria: reliability, quality, flexibility, time, and cost

the association rules and the key factors obtained by RFs complement each other. We integrate the information to find hidden correlations by finding potential key factors. It shows that the employee class staffGroup, the employee code staff, and the part number supplierLog are relatively important indicators for the entire process. The parameters of the associated relationship are week weekend and date date. In this work, through the association rules and the predictive maintenance system constructed by the RF algorithm, we obtain the correlation and key parameters to solve the important problem: finding the information of the abnormal association in the process of observing the defective products. The major limitation is that the data driven approach highly depends on the information accessible to the operators. It is important to reduce the uncertainty via collecting different aspects of processing data, which is the main focus of the improvement in future work.

6  Case study analysis II

In this case study, the data are collected from the AGVs moving in the environment. The data consist of the displacement, current position, angle of rotation, battery level of an AGV. The items on the AGVs' shelves for warehousing management, logistics management, and the documentation of the processed result of each item are also recorded. All of the information is organised and delivered to the main console of the AGV via a Wi-Fi network, and then routed to a remote monitor console through another Wi-Fi channel. The CPS plays a key role as the monitor console in this architecture. The AGV data are all received within and then used to generate the traceability of the production. At the final stage, the production history is analysed, followed by performing the pre-alarm monitoring on the record. In addition, the acquired numerical data are transmitted to the IoT platform via a Wi-Fi network in the CPS. As depicted in Fig. 6, the proposed IoT platform consists of three parts: the evaluation module, the early warning module, and the mirror module.

6.1  Evaluation module

In this paper, we adopt an evaluation module, which is to verify the performance of the given tasks or events. Our evaluation will be established based on the analysed quantitative data. An effective evaluation module can help our system to detect errors and omissions as soon as possible. It also plays an important role in improving the countermeasures to maximise the efficiency of the management system. In the previous work, Yurdakul [29] constructs a multi-criteria manufacturing performance system (as shown in Fig. 7). The evaluation of manufacturing performance is divided based on five criteria: reliability, quality, flexibility, time, and cost. In the approach, the performance index is constructed for each part in a hierarchical manner as shown in the figure. The weight for each performance index is computed so as to promote the manufacturing competitiveness.

6.2  Early warning module

This research lies in the establishment of automated factory monitoring and warning systems [30, 31]. After a large amount of data collected by the IoT platform, a database is built and used to construct a set of real-time warning modules. It is used to monitor logistics management on the shelves of the AGVs and warehousing management. In addition, the long-term record data can be used to analyse the relationship between the inbound and outbound logistics of the shelves, as well as their relations with the workers. With this information, the system can continuously monitor and provide warning. When the observed values reach pre-set thresholds, the system will issue a warning signal and notify the administrator to perform the pre-disposition so as to achieve the functionality of early warning. There are two early simulated warning systems set up in this work: battery simulation warning system and MES.

6.2.1  Battery simulation warning system: Compared with other rechargeable batteries, lithium batteries have the advantages of high energy density, high operating voltage, large operating temperature range, long service life, and better discharge capacity. With the same size of batteries, a lithium battery can store more energy. Typically the voltage of a single lithium battery is about 3.8 V, while other batteries are with 1.2 V. Thus, the use of lithium batteries can easily reach the required voltage. Compared to other batteries which reduce 1–5% of the daily electricity, lithium batteries are not discharged for several months and maintain the status of almost full charge. A simple mathematical model is established to describe the discharge behaviour of lithium batteries

This is an open access article published by the IET under the Creative Commons Attribution -NonCommercial License
(http://creativecommons.org/licenses/by-nc/3.0/)
Table 4 Parameters \(m, n, p, q\) of the function at 0.4 A current

|   | \(m\)       | \(n\)       | \(p\)       | \(q\)       |
|---|-------------|-------------|-------------|-------------|
| \(A(t)\) | -0.001624  | 2.22 \times 10^{-19} | -0.01149  | -0.0002799 |
| \(B(t)\) | -4.7114   | 7.20 \times 10^{-17} | 11.96     | 0.02127    |
| \(C(t)\) | 49.48     | -2.28 \times 10^{-10} | 7.374     | -1.907     |
| \(D(t)\) | 270.7     | -4.095 \times 10^{-15} | 105.2     | 0.4304     |
| \(E(t)\) | 185.1     | -3.127 \times 10^{-15} | 4.466     | 0.8868     |

According to the formula, the parameters \(m, n, p,\) and \(q\) of the function at 0.4 A current are shown in Table 4.

6.2.2 Manufacturing execution system: The establishment of a smart factory in the import of information and communications technology to strengthen the quality management, the establishment of the production traceability system, and the formation of upstream and downstream manufacturers linked to the production and marketing supply chain, these will surely be able to enhance the competitiveness of enterprises. Smart factories integrate various automation technologies. An MES has a large number of sub-systems, including in-process management, statistical process control, equipment management, anomaly alarm etc. It possesses such features as the post-operation barcode, materialised electronic management, and production traceability transparency. This is a transition from the past with the handwriting error-prone traditional way to the fool-proof precision digital operation. The electronic data are conducive to product traceability of the production process. Data analysis for processes that needs to be improved can also save some time on data collection. It allows managers to spend more time to set the best course of action.

The MES starts from the receipt of orders at the beginning of the production to the completion of the product. It provides the product information produced in the production line, including the product number, shelf number, battery power, handling staff number and the products processed by the line records, and so on. The factory-generated information is presented in a real-time and accurate manner, and in a report or other forms. The results presented will change rapidly with the current situation. Its purpose is to reduce the activities without added-values, such as machine crashes etc., and to increase the effective production. In addition, every important processing event in the production line will be recorded. On one hand, it can be used as the production traceability of the product, and the electronic data can also be used to trace the product of the production process. On the other hand, when the number of production experiences increases, the data can be analysed later. When a product has a problem in the market, it takes at least a few days for the staff to find the information in the past. Now the information is electronically made available in a matter of seconds, it can make the business to respond faster to abnormal events. Furthermore, it can provide the information on site emergencies when there is an emergency in production activities, and notify managers via two or more pipelines.

This paper takes warehousing, detecting, and packaging as the three main workstations in the production line, as shown in Fig. 8. The main work of warehousing, records the entry and exit of materials, and sets a warning threshold value for each material. When the amount of material reaches the alert value, the system will issue a warning signal and carry out an early warning action to achieve status monitoring and control. Detecting is to record the tasks of actions performed on each product at the detecting station, and all detecting actions are recorded in the production traceability. Once there is a large number of production traceability, we can analyse and detect the abnormality during the production process. In addition, we also deal with the stream analysis of the staff and the number of processes. At the end of the production line there is packaging, the process is the same as warehousing. We record the total weight of the packaging materials, and set a warning value. When the total weight is no longer within the scope of the warning value, the system will issue a warning signal and perform an early warning action to achieve monitoring and warning.

6.3 Mirror module

Through the sensors on the AGVs, a variety of information is collected and used to build various types of mirror models to make the data more intuitive. In the control interface, a mirror model of each AGV is established and a map is constructed by the SLAM (Simultaneous Localisation and Mapping) technique. The mirror model is graphically displayed, and there is a corresponding function in the model for each parameter of the AGV, so as to conveniently adjust the parameters and observe the movement of the AGV. We use OpenGL to build the 2D model of the graphical
interface, and display the parameters on the console. The user or administrator can then observe the status of each AGV in real time.

6.4 Battery simulation warning system

In the battery simulation warning system, we tested a 3S1P lithium battery pack. We want to cut off the voltage from 12.5 to 9 V under a 4 W bulb plus a 5 W DC motor. The lithium battery output voltage will decrease with the discharge time, the remaining capacity will be a specific proportional relationship with the voltage \[29, 33, 34\]. If we consider the ordinate as the battery discharge voltage and the abscissa as the percentage of battery discharge, also known as capacity, the discharge behaviour is a decreasing continuous curve. According to (7) and (8), we implemented and improved the cell we tested. Fig. 9 shows the result of our experiment. The actual test and predicted curves are shown in orange and purple, respectively.

In Fig. 9, the lithium battery is characterised by the discharge curve. The early part and the middle part of the curve are gently declined. However, at the last stage of the discharge, this curve drops sharply, which is the rapid change of the remaining capacity of the battery between the battery voltage of about 9.5 – 9.0 V. This is the reason why the remaining capacity of lithium battery is difficult to control \[35\]. Thus, we give a warning value according to the declining curve of battery voltage. When the value provided by the system monitoring reaches the warning value, the system will give a warning signal and notify the system administrator to perform the pre-disposition for this function.

6.5 Manufacturing execution system

In the experiment, we generate simulated data randomly. To make the data uniform, we need to generate the Gaussian distribution. If the random variable \(x\) obeys a probability the distribution with a positional parameter \(\mu\) and a scale parameter \(\sigma\), the probability density function is given by

\[
f(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}}
\]

The normal distribution of the mathematical expectation \(\mu\) is equal to the location parameter, which determines the location of the distribution. The square root of the variance \(\sigma^2\) or standard deviation \(\sigma\) is equal to the scale parameter, which determines the distribution amplitude.

In the MES, we simulate the test data as shown in Table 5. The data are stored as a string, in the order of the date, product number, material number, the number of materials, station number, and handling staff. Table 6 shows the name and code of each field. We divide the string, and code the fields in ASCII text. It is easy to quickly segment and analyse production traceability, and also very flexible with the assignment for the current detection actions. One major difficulty of this approach is how to collect and organise different types of data. At the current stage, the simulation is performed for the concept illustration. This should be extended to the real-world scenarios and tested extensively in future work.

---

**Table 5** Simulated test data sheet used in this work

| DATE   | ID      | MAT   | STAFF        | VAL     | STA |
|--------|---------|-------|--------------|---------|-----|
| 170407 | B505-01 | MT007 | PL4A428031   | 0.132   | 1   |
| 170407 | B505-01 | MT008 | PL4A428032   | 1201    | 2   |
| 170407 | B505-01 | MT009 | PL4A428033   | 1.31    | 3   |
| 170407 | B505-02 | MT007 | PL4A428031   | 0.135   | 1   |
| 170407 | B505-02 | MT008 | PL4A428032   | 1194    | 2   |
| 170407 | B505-02 | MT009 | PL4A428033   | 1.28    | 3   |
| 170407 | B505-03 | MT007 | PL4A428031   | 0.123   | 1   |
| 170407 | B505-03 | MT008 | PL4A428032   | 1197    | 2   |
| 170407 | B505-03 | MT009 | PL4A428033   | 1.29    | 3   |
| 170408 | B505-04 | MT007 | PL4A428031   | 0.129   | 1   |
| 170408 | B505-04 | MT008 | PL4A428032   | 1204    | 2   |
| 170408 | B505-04 | MT009 | PL4A428033   | 1.31    | 3   |

**Table 6** Correspondence of ‘Name’ and ‘Code’

| Type     | Name | Code     |
|----------|------|----------|
| date     | DATE | YYMMDD   |
| product ID | ID  | B505-01  |
| material ID | MAT | MT001    |
| staff    | STAFF | PL4A428029 |
| value    | VAL  | 0.132    |
| station  | STA  | 2        |

---

IET Cyber-Phys. Syst., Theory Appl., 2020, Vol. 5 Iss. 1, pp. 100-109
This is an open access article published by the IET under the Creative Commons Attribution -NonCommercial License (http://creativecommons.org/licenses/by-nc/3.0/)
7 Conclusions
This paper presents the design and implementation of a predictive maintenance big data platform based on the CPS under the Industry 4.0 architecture. An automated management platform is also developed for data aggregation, control, monitoring and early warning, with the performance evaluation system to maximise the efficiency of the management system. This platform is targeted at product data, from manufacturing big data to cloud computing, and provides predictive maintenance for all plants around the world. The process parameter correlation analysis and process parameter critical analysis are realised through massive data exploration technology, and the predictive maintenance system is realised based on the analysis using an offline association rule algorithm and offline predictive maintenance. The implementation is carried out with evaluation, early warning, and mirror systems. In the experiments, battery simulation warning and manufacturing execution are presented to demonstrate our system design.

8 Acknowledgments
The support of this work in part by the Ministry of Science and Technology of Taiwan under Grant MOST 106-2221-E-194-004 and the Advanced Institute of Manufacturing with High-tech Innovations (AIM-HI) from the Featured Areas Research Center Program within the framework of the Higher Education Sprout Project by the Ministry of Education (MOE) in Taiwan is gratefully acknowledged.

9 References
[1] Vogel-Heuser, B., Hess, D.: ‘Guest editorial industry 4.0: Sperequirities and visions’, IEEE Trans. Autom. Sci. Eng., 2016, 13, (2), pp. 411–413
[2] Zhou, K., Liu, T., Zhou, L.: ‘Industry 4.0: towards future industrial opportunities and challenges’. 2015 12th Int. Conf. on Fuzzy Systems and Knowledge Discovery (FSKD), Zhangjijiang, China, August 2015, pp. 2147–2152
[3] Niggemann, O., Lochow, V.: ‘On the diagnosis of cyber-physical production systems: state-of-the-art and research agenda’. Proc. of the Twenty-Ninth AAAI Conf. on Artificial Intelligence, ser. AAAI15, Texas, USA, 2015, pp. 4119–4126. Available at http://dl.acm.org/citation.cfm?id=2888116.2888294
[4] Lee, J., Bagheri, B., Kao, H.-A.: ‘A cyber-physical systems architecture for industry 4.0-based manufacturing systems’, Manuf. Lett., 2015, 3, pp. 18–23
[5] Lee, J., Kao, H.-A., Yang, S.: ‘Service innovation and smart analytics for industry 4.0 and big data environment’. Procedia CIRP, product Services Systems and Value Creation. Proc. of the 6th CIRP Conf. on Industrial Product-Service Systems, Ontario, Canada, 2014, vol. 16, pp. 3–8
[6] Letichevsky, A., Letychevskiy, O., Skobelev, V., et al.: ‘Cyber-physical systems’, Cybern. Syst. Anal., 2017, 53, (6), pp. 821–834
[7] Yang, F., Wu, C., Lin, H.: ‘Design and implementation of cps-based automated management platform’. IEEE Int. Conf. on Systems, Man, and Cybernetics, SMC 2018, Miyazaki, Japan, 7–10 October 2018, pp. 2293–2298. Available at https://doi.org/10.1109/SMC.2018.830934
[8] Navet, N., Bertholet, I.C., Hu, T.: ‘Software patterns for fault injection in cps engineering’. 2017 22nd IEEE Int. Conf. on Emerging Technologies and Factory Automation (ETFA), Limassol, Cyprus, 2017, pp. 1–6
[9] Siew, Y., Chereh, R.I.: ‘Construction objects recognition in framework of CPS’. 2017 Winter Simulation Conf. (WSC), Las Vegas, USA, 2017, pp. 2472–2483
[10] Ning, Z., Hou, W., Hu, X., et al.: ‘A cloud-supported cps approach to control decision of process manufacturing: 3D ONOC’, 2017 13th IEEE Int. Conf. on Automation Science and Engineering (CASE), Xi’an, China, 2017, pp. 458–463