Chapter 4
Big Data Analytics for Predictive Maintenance Strategies

C. K. M. Lee
The Hong Kong Polytechnic University, China

Yi Cao
The Hong Kong Polytechnic University, China

Kam Hung Ng
The Hong Kong Polytechnic University, China

ABSTRACT

Maintenance aims to reduce and eliminate the number of failures occurred during production as any breakdown of machine or equipment may lead to disruption for the supply chain. Maintenance policy is set to provide the guidance for selecting the most cost-effective maintenance approach and system to achieve operational safety. For example, predictive maintenance is most recommended for crucial components whose failure will cause severe function loss and safety risk. Recent utilization of big data and related techniques in predictive maintenance greatly improves the transparency for system health condition and boosts the speed and accuracy in the maintenance decision making. In this chapter, a Maintenance Policies Management framework under Big Data Platform is designed and the process of maintenance decision support system is simulated for a sensor-monitored semiconductor manufacturing plant. Artificial Intelligence is applied to classify the likely failure patterns and estimate the machine condition for the faulty component.

DOI: 10.4018/978-1-5225-0956-1.ch004

INTRODUCTION

Maintenance can be defined as all actions which are necessary to retain or restore a system and a unit to a state, which is necessary to fulfill its intended function. The main objective of maintenance is to preserve the capability and the functionality of the system while controlling the cost induced by maintenance activities and the potential production loss. Correspondingly, failures can be defined as any change or anomaly in the system causing an unsatisfactory level of performance. Although only certain failures will cause severe risk in productivity and safety, most failures lead to disruptive, inconvenient, and expensive breakdowns and loss of quality. Maintenance plans are designed to reduce or eliminate the number of failures and the costs related to them.

There are two broadly accepted methodologies aiming at continuously enhancing maintenance excellence, with different focuses. As a human factor management oriented policy, total productive maintenance (TPM) involves all employees, especially the operators, in the maintenance program in order to achieve optimality in overall effectiveness and zero breakdowns. Through the operators' participation in maintenance, such as through inspections, cleaning, lubricating and adjusting, early detection of hidden defects, before service breakdown. TPM aims to diminish and eliminate six significant losses of equipment effectiveness – i.e. breakdowns, setup and adjustment, idling and stoppages, reduced speed, defects in process, and reduced yield (Jardine & Tsang, 2013).

Reliability-centered maintenance (RCM) is another approach to strengthening the system's reliability, availability and efficiency which focuses on design and technology. RCM program is based on systematic assessment of maintenance needs after a complete understanding of the system function and the types of failure causing function losses.

Types of Maintenance

Maintenance activities can be categorized into three types:

1. Reactive or corrective maintenance,
2. Preventive maintenance (PvM), and
3. Predictive maintenance (PdM).

The following terms are also respectively used for the above three categories interchangeably as:

1. Breakdown maintenance or unplanned maintenance,

2. Planned maintenance, and
3. Condition based maintenance (CBM) or prognostic and health management (PHM).

Reactive or corrective maintenance follows the run-to-failure methodology, which is the repair and/or replacement work after an equipment outage has occurred. This primitive maintenance approach, which has been applied in industry for decades, and is still considered the best maintenance policy for non-critical components with short repairing time in the system. However, in most cases, an equipment failure can lead to unexpected production delay and lower the production efficacy rate, or more seriously, cause severe damage to other components and/or injury to people. One goal of a proactive maintenance plan is to reduce the overall requirement for reactive maintenance and to apply PvM and/or PdM strategies on any feasible occasion.

Preventive maintenance is performed based on a certain periodic interval to prevent and correct problems before breakdown without considering the actual health condition of a system. Basic preventive maintenance, including inspections, lubrication, cleaning and adjustment is the first step to be undertaken. After that, rectification or replacement can be undertaken only for components identified with defects and/or considerable risk of failure. Generally, most PvM actions can be implemented by operators with basic training.

Predictive maintenance is a trend-oriented policy that begins with identifying the states of each component within the equipment. PdM greatly relies on engineering techniques and statistical tools to process the data and analyze the health condition in order to predict possible equipment failure (Lee, Ardakani, Yang, & Bagheri, 2015). The prediction of the equipment condition is based on the finding that most types of failures, which occur after a certain degradation process from a normal state to abnormalities, do not happen instantaneously (Fu et al., 2004). Through degradation monitoring and failure prediction, PdM reduces the uncertainty of maintenance activities and enables identifying and solving problems before potential damage. Condition-based maintenance, the alternative term for PdM, imposes more emphasis on real-time inspections using RFID devices and wireless sense networks (WSNs). The three key steps of CBM are monitoring and processing, diagnosis and prognosis, and maintenance decision making.

BIG DATA ACHITECTURE

Big Data is not only a matter of volume increased of the collected data, but also includes the evolution of data peculiarities, namely data variety and data velocity. The intrinsic pattern of comprehensive data becomes the major driver for companies to investigate the use of big data analytics. The features of big data are broadly recognized as "4Vs" – i.e. volume, velocity, variety and value (Xin & Ling, 2013).

- **Data Volume:** Data volume is the primary attribute of data. Due to the dramatically boost in data volume and the rising demand of data storage, database management has been modified as petabyte-scale data management. The storage needs can be easily satisfied by increasing data warehouse capacities but at the same time, the data transfer and processing will become too slow to be operated and handled by a single computer. Therefore, a super-computer or high capacity server is required for operation. On the other hand, the cost is also a major concern for implementing with advanced computing equipment. The data growth in the supply chain and logistics system becomes too large and too complex in the traditional data warehouse and system architecture. In addition, big data analytics makes good use of multiple modes during the computation. This Big Data Analytics method can handle large volume data.
- **Data Velocity:** With the use of automatic data retrieval systems (e.g. sensor network, wireless network and electronic data interchange through the Intranet and Internet), the speed of data transaction is rapidly increased. Moreover, this allows companies to collect instant information and accelerate the speed of data production and streaming. The data transfer and transformation are more intensive nowadays. If the company chooses to extract real time information from several automatic data retrieval systems, the speed of data processing from big data must be as expeditious as possible to meet the requirement of quick response.
- **Data Variety:** The analytics process of data mining has been expanded from structured data to unstructured data, typically the images, videos, audio and text, etc. The systematic relationship is longer limited by numerical results, but also prohibits pattern finding in unstructured data. This motivates companies to investigate semi-structured and unstructured data for their decision making process. The major purpose of big data analytics is to resolve the problem of incompatible data formats and non-aligned data structures of unstructured data with data mining techniques.
- **Data Value:** Big data analytics creates value for data mining in order to find the intrinsic and multidimensional attributes from the enormous amounts of data. To a certain extent, a big data-driving model can perform as a support vector machine to establish supervised learning, which allows the model to adapt and evolve from time to time. Value creation is a significant process contributing to organizations' continuous improvement and demand prediction. We need to understand that big data analytics is not only in statistical analytics, but also in more complicated and tailor-made analytics. In general,

big data analytics come into existence for resolving data storage problems as well as providing valuable insights for organizations.

The acquisition and processing of big data largely improves the transparency along the supply chains providing accurate and timely information for managerial decision making. Companies and organizations operate on the huge amounts of data by classifying trends and identifying patterns to produce invaluable knowledge.

Meanwhile the flood of big data with high speed and many variations has challenged the limited storage and conventional data mining methods. Challenges are also from processing and analyzing the large amount of unstructured data which are the major components in the big data acquired. Technologies have been advancing towards better performance in the big data context regarding integrated platforms, predictive analytics, and visualization (Lee, Kao, & Yang, 2014). Big data and predictive analysis are strongly interconnected. Without proper analytics, big data is just a deluge of data, while without big data, predictive analytics, the strength of statistics, modeling, and data mining tools for analyzing current and historical conditions will be undermined.

BIG DATA ANALYTICS

The descriptive tasks of big data analytics identify the common characteristics of data with the purpose of deriving patterns and relationships existed in the data. The descriptive functions of big data mining include classification analysis, clustering analysis, association analysis, and logistic regression.

- **Classification Analysis:** Classification is a typical learning models used in big data analytics, which aims to build a model for making prediction on data feature from the predefined set of classes according to certain criteria. A rule-base classification is used to extract IF-THEN rules to classify as different categories. The examples include neural network, decision trees and support vector machine.
- **Clustering Analysis:** Clustering analysis is defined as the process of grouping data into separate cluster of similar objects, which helps to segment and acquire the data features. Data can be divided into different subgroups according to the characteristics. The practitioners may formulate appropriate strategies for different clusters. The common example of clustering technique are K-means algorithm, self-organizing map, hill climbing algorithm and density-based spatial clustering.

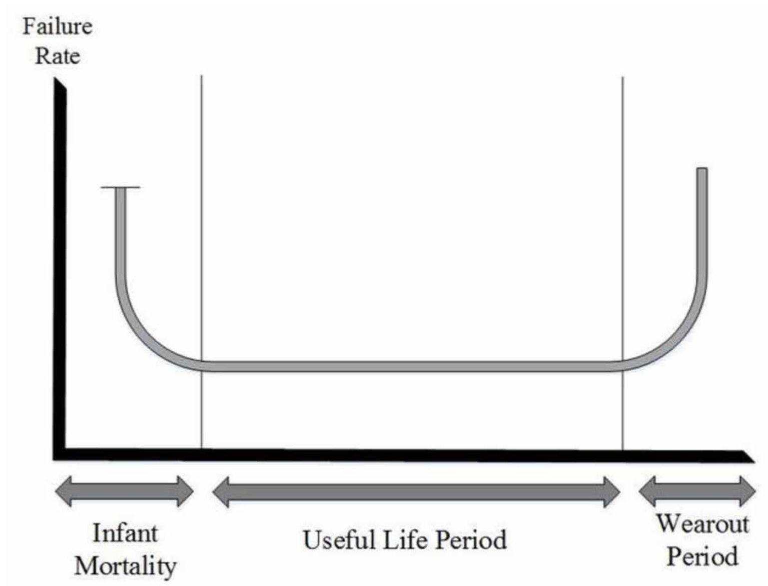
- **Association analysis:** Association model helps the practitioners to recognize groups of items that occur synchronously. Association algorithm is developed for searching frequent sets of items with minimum specified confidence level. The criteria support and confidence level helps to identify the most important relationships among the related items.
- **Regression Analysis:** Regression represents the logical relationship of the historical data. The focus in regression analysis is to measure the dependent variable given one or several independent variables, which support the conditional estimation of expected outcome using the regression function. Linear regression, non-linear regression and exponential regression are the common statistical method to measure the best fit for a set of data.

MAINTENANCE STRATEGIES

The Big Data platform has the ability to handle huge amounts of data in manufacturing or production logistics databases along with the development of computerized maintenance management systems (CMMS), which assist decisions making so as to formulate maintenance strategies.

Maintenance procedures will be undertaken when a machine failure has occurred in the CrM strategy. Manufacturers are required to keep components inventories for maintenance, repair and operations (MRO) in order to prevent disruption of the overall production by failure of machine parts or equipment. Compared to the CrM strategy, maintenance performance in the PvM strategy follows a fixed time, interval basis or condition based schedules to avoid fatal machine failure. The design of PvM is a protective, process-oriented approach in which machine failure and downtime cost could be reduced by taking proper prevention and prevention to smoothen the production. Decisions on maintenance schedules is based on a machine's physical properties or asset condition. Extra resources are spent on non-value-added activities to estimate and measure the condition rules for PvM (Exton & Labib, 2002). However, PvM attempts to provide an empirical basis for the development of a framework design of manufacturing flexibility at machine idle periods and during maintenance activities. The assumption behind a PvM policy is that the machine failure follows the bathtub curve in Figure 1. Scheduled maintenance happen in the wear-out phase in order to reduce the failure rate (Sikorska, Hodkiewicz, & Ma, 2011). However, the most conspicuous deficiency in PvM is still the apparent random failure within the useful life period. The impact of failure in a critical machine is a tremendous risk to the downtime costs and, it in turn becomes bottleneck in production logistics operations.

Figure 1. Classical Bathtub Curve
(Klutke, Kiessler, & Wortman, 2003)



To remedy random machine failure in maintenance management, PdM has been well developed carrying out observation of the machine degradation process and symptoms from normal to flawed situations (Wu, Gebraeel, Lawley, & Yih, 2007). PdM is a sensor-based content-awareness philosophy based on the foundation of "Internet of Things". The intelligent maintenance prediction support system monitors the machine status by utilizing real time sensory data (Kaiser & Gebraeel, 2009). Advance maintenance in PdM policy is able to provide insights for maintenance scheduling in advance in order to eliminate unanticipated machine breakdowns, and minimize maintenance costs as well as downtime, before the occurrence of random machine failure (Garcia, Sanz-Bobi, & del Pico, 2006).

The important factors associated with PdM in Maintenance Policy Management (MPM) emphases criticality, availability of sensory data, reliability, timeliness, relevance and knowledge-oriented strategy.

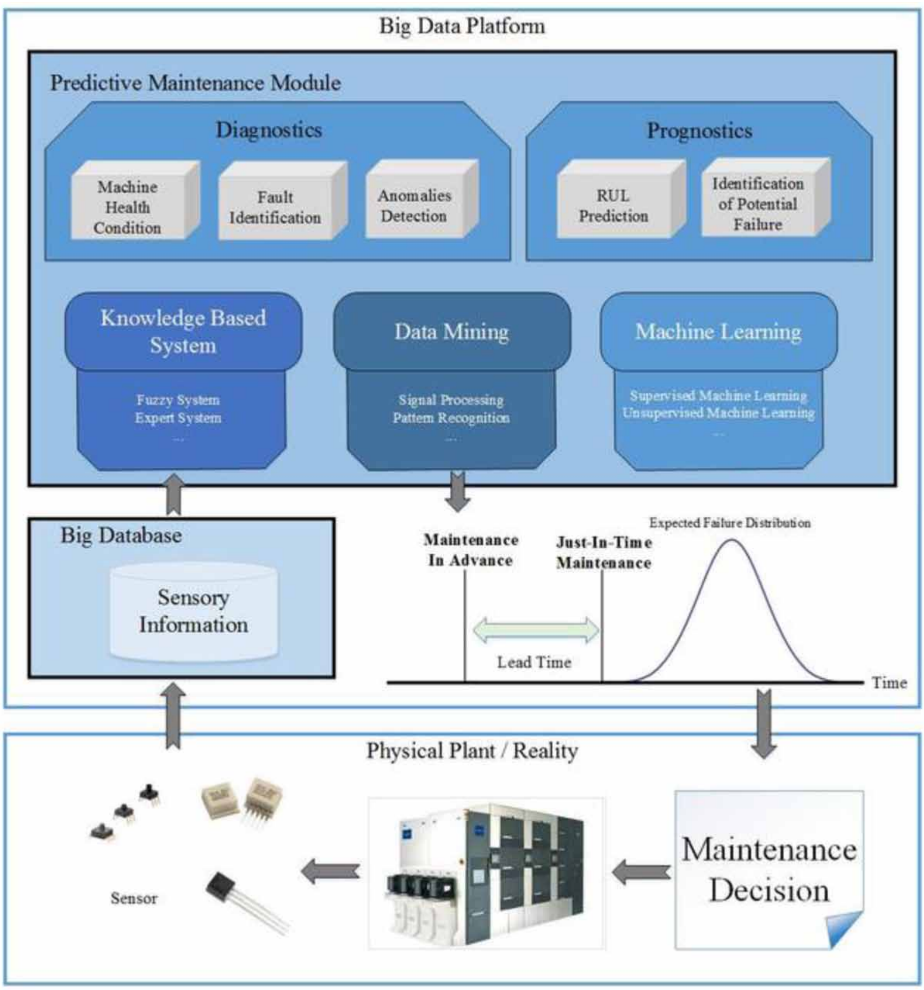
- **Criticality in Failures:** PdM strategy has been heavily concentrated in real time machine condition monitoring through diagnostics and prognostics for reimbursement of foreseeable machine downtime cost. In reliability study, the critical assets must have a higher rank in priority formulate PdM strategy to predict the most likely time for the next machine breakdown and random error, as this will have the greatest impact on the production operations This changes the maintenance objective from avoiding breakdown to accepting downtime and taking maintenance action ahead of the schedule.

- **Availability of Sensory Data:** PdM policy is highly dependent on extract-transform-load (ETL) operational data in close condition-based monitoring. The current operational status and abnormal performance could be assessed by equipping sensors to identify failure modules or machines. Lack of sensory data may result in unpromising maintenance prediction.
- **Reliability:** Maintaining critical machine performance and leveraging the overall cost to sustain production are the major targets of PdM policy. The system must provide the correct measures and reliable performance in prediction to address feasible and foreseeable machine failure, and build confidence in operation.
- **Timeliness:** The prediction for maintenance modules must have a high level of confidence level before the undesired event occurs, and the data size and data transmission speed administered in a timely manner. The time series of the maintenance schedule and delivery of MRO should be taken into consideration the maintenance management in order to facilitate the production, with zero tolerance of equipment failure.
- **Relevance:** The MPM system needs to be developed based on the opinions of experts. The collected sensory information must be recorded and analyzed on a real time basis. In order to improve data quality, extraction of relevant data for maintenance decision making is crucial in regard to engineering aspects. Inappropriate integration of a sensor and machine may cause poor estimation and inaccuracy prediction of the current machine performance.
- **Knowledge-Objective Oriented Strategy:** The concept of the PdM strategy involves a belief that the implicit knowledge from collaboration of sensory information did contribute to the maintenance in advanced. The knowledge transfer system facilitates the disclosure of implicit information to maximize production efficiency and minimizes the adverse impact of idling time under maintenance and unawareness of potential failure. The decision of PdM policy could be assessed by the involvement of Big Data Mining Techniques to detect and defeat anomalies at an early stage.

PREDICTIVE MAINTENENACE IN BIG DATA FRAMEWORK

The ideology of a PdM is to create transparency of the machine condition and in the utilization of available information for maintenance decision making. In Figure 2, the framework of a big data platform in PdM is designed for closer integration of data acquisition and the maintenance decision support system (MDSS), which highlights the dataflow process in diagnostics and prognostics modeling for PdM. Case examples are provided in the following section to describe the framework of MPM for a semiconductor machine under Big Data Platform, and operations of MDSS. The operating data from vibration, heat and pressure sensors, which provide sensory information stored in the Big Database, are embedded on the semiconductor machine to evaluate the machine condition. The diagnostics and prognostics process may involve real time data and historical data for data mining procedures. The advantage of Big Data Architecture are capable to manage huge units of data and perform ETL in a timely manner by using appropriate data processing algorithms, such as Map-Reduce technique. The Big Data platform can build an intelligent agent

Figure 2. Predictive maintenance model in big data platform



to connect the PdM module and the sensor knowledge from the actual machine. The PdM module consists of two major systems: a diagnostics system and a prognostics system. PdM analysis is a powerful tool to identify machine/components failure, and provide surprisingly accurate future breakdown using time-series data by well-developed analytic processes. The predictive maintenance module closely supervises the machine condition and constantly aids the analytics process of diagnosis and prognosis. The results from Artificial Intelligence (AI) will then be further processed and evaluated by the MDSS. Certain operational guidance and estimation of failure events, such as foreseeable situations, time to breakdown and estimated downtime, are recommended for maintenance decision making. With the implementation of the big data platform, more precise sensory data acquisition and accurate maintenance decisions could be made from the suggested algorithms in order to manage critical machines, with the objectives of maintenance in advanced or just-in-time (JIT) maintenance for supply chain management.

It is practical for a decision makers to select appropriate analytic processes and recognize the functionalities of algorithms for maintenance planning. Therefore, a comprehensive discussion of algorithm design is presented herein. The common practices of AI techniques can be classified as Knowledge Based System (KBS), Data Mining (DM) and Machine Learning (ML) (Faiz & Edirisinghe, 2009).

- **Knowledge Based System:** These types of analytics process require logical deduction and cognitive reasoning to resolve complex problems and support decision making. KBS attempts to extract rules for algorithm contexts by human intelligence and expert opinion, which are of practical significance. The practical necessity of KBS is increased due to the advancement of sensor-based PdM. KBS encourages a more flexible way to increase quality in problem solving and in extracting relevant data into knowledge for decision making, i.e., machine failure identification, classification in maintenance policies. A variation of sensory information causes data-booming in the analytics process. The rule-based and inference engine expert system have the ability to simulate a human expert in reducing the complexity in MPM and in discovering hidden machine failures.
- **Data Mining:** The goal of the DM process is to create a constructive model of patterns recognition and feature analysis, which is able to classify data into groups, detect irregular features and measure the dependencies of data. Moving toward a total productive manufacturing system, DM is an instrument that is able to mine all kinds of manufacturing knowledge, such as job shop scheduling, manufacturing process, quality control, yield improvement, and even predictive maintenance strategy. In the data mining technique, the accuracy of the information discovered increases along with the increase in

gathered data from the sensors and the historical maintenance data, which is able to foresee failure from pattern behavior of the operating machine data and the increased reliability of MDSS.

- **Machine Learning:** ML is another dimension of the analytics process. The mechanism of KBS and DM are either to discover knowledge and insight beforehand for the working process of the algorithm, which is concerned information and knowledge extraction from massive data. However, ML deals with automatic reasoning and artificial cognitive resolution by an intelligence agent. ML works as an online measurement of a health detection system to reveal machine degradation and anomalies from the models. Self-learning and reinforcement in ML, together with normal degradation allow the forecasting of random machine failure effectively and efficiently in order to plan for the best before failure occurs.

THE RELATIONSHIP BETWEEN DIAGNOSTICS AND PROGNOSTICS

In most cases, there is a measureable process of degradation before a machine fails. Figure 3 illustrates the degradation process of a system, sub-system, or component into failure. Through functioning life, the system may continuously degrade to a condition with an observable drop in its performance level and initial faults may occur during the degradation progress. The incipient defects continuously proliferates and the severity gradually increases which causes the system to fail to perform its required function and fails. In order to predict and prevent failures in advanced, diagnostics and prognostics techniques are studied and employed to evaluate the current health conditions and forecast future performance.

Figure 3. Fault to failure progression

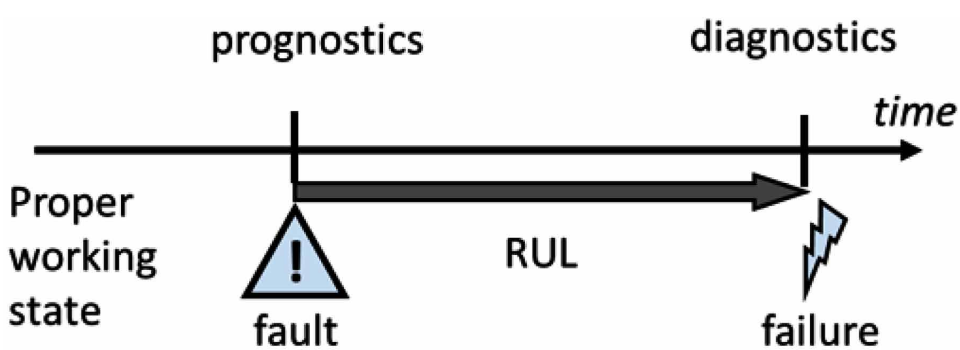


Figure 4 indicates the inputs and outputs of the machine diagnostics and prognostics process. The diagnostics process determines how the system has degraded and investigates the cause or nature of such a degraded condition. Faults are detected, isolated and identified to create a diagnostic record, and the faults possibilities are computed to find the potential failure pattern. The diagnostic operations follow an analysing approach from effect to cause. On the other hand, prognostics considers time as a vital factor and focuses on predicting the future condition of the system or component and in calculating accurately the remaining useful life before the failure. The prediction is conducted from cause to effect. The analysing and computing process is based on current system conditions and future operational requirements. Although similar information and knowledge bases are shared, the difference between the two concepts now becomes obvious. Diagnostics is to investigate and determine a failure mode within a system; while prognostics is to compute a rather accurate result of the remaining useful life before final failure.

The following sections present a diagnostics and prognostics system in Big Data Analytics with a case example. The system flowchart is shown in Figure 5. Fault identification is one of the diagnostics modules. The classification of machine failure is critical to smoothen the production, as not all the failures require emergency maintenance. The result from the diagnostic model is able to assist in the development of prognostics model. The reliability of the prediction could be increasingly improved from a known machine failure.

Figure 4. Inputs and outputs with the diagnostic–prognostic process structure

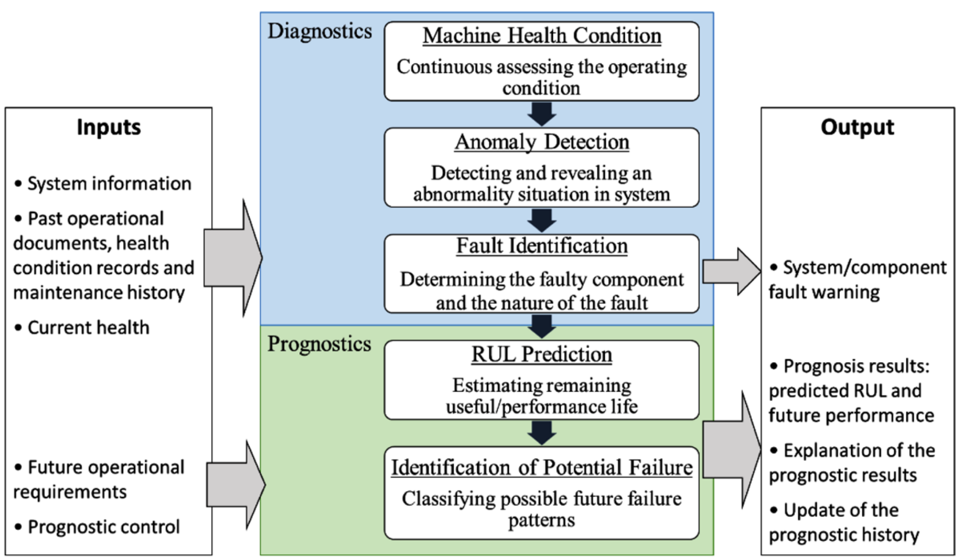
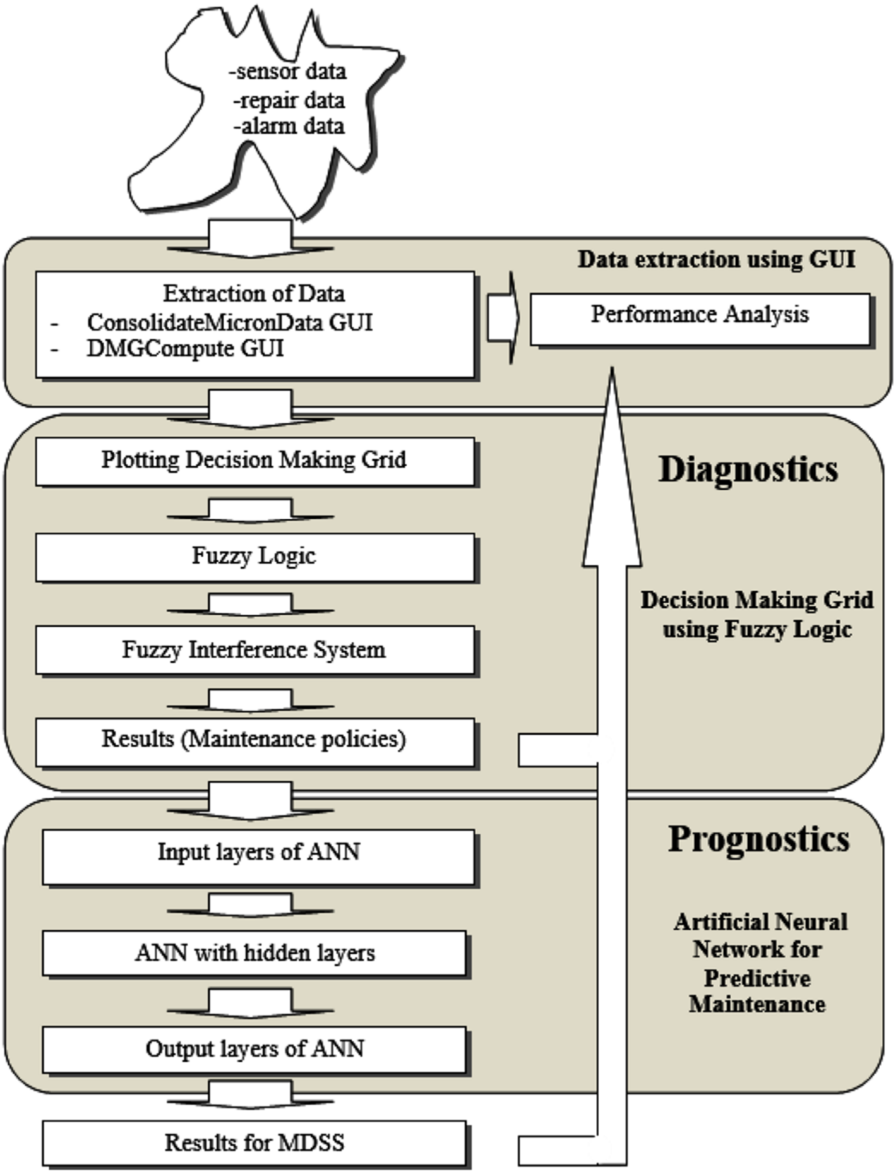


Figure 5. Flowchart of the proposed MDSS for case company



The machine failure identification has been developed and incorporated with the current decision making grid (DMG) in MDSS for the case company. The DMG model works like a map which categorizes the machines according to a set of pre-defined parameters/criterions. The criteria for this project are the downtime and the frequency of failures from the sensor data and historical record (Exton & Labib,

Figure 6. Decision Making Grid by the case company

Frequency	**Downtime**		
	Low	Medium	High
Low	OTF	**FTM1**	CBM
Medium	**FTM1**	**FTM2**	**FTM3**
High	SLU	**FTM3**	DOM

2002). Other criteria, such as cost, availability of MRO and the bottleneck problem, can also be included by expanding it from merely a 2 dimensional analysis to a 3 or more dimensional analysis model. Higher dimensional models can produce a more comprehensive and accurate analysis. The DMG then proposes the different types of maintenance policies based on the state in the grid which then determine the appropriate maintenance actions for the MDSS. These maintenance policies subsequently lead to be the formation of the following strategies; Operate to failure (OTF), fixed time maintenance (FTM), skill level upgrade (SLU), condition based monitoring (CBM) and design out machine/component (DOM), as shown in Figure 6.

- **Operate to Failure:** There are too many low downtime and low downtime frequency components which made it impossible or too expensive for applying scheduled PdM maintenance. These components/machines are allowed to operate to failure as they are deemed as having minimal impact on the system. Ideally, if operators of the machine are also maintenance technicians, which agrees with the TPM policy, OTF 'faults' can be further reduced without the need for reporting and waiting for the technician to bring the machine back to normal operating state. A CrM approach is suggested for OTF situation.
- **Fixed Time Maintenance:** PvM is also called Fixed Time Maintenance. A more flexible PvM can be chosen which is determined either by availability, machine not in use, or by the severity of the faults. Faults with higher downtime and frequency faults are allowed to have shorter fixed time maintenance,

whereas faults with lower downtime and frequency can extend the duration of their fixed time maintenance.

- **Skill Level Upgrade:** This strategy falls on the high frequency and low downtime grid. Faults that fall under this strategy are faults caused by human errors, such as the accidental pressing of the wrong switch/button. For this grid, the human factor oriented policy will be emphasized, either through the human/operator or through the machine. From the human perspective, operators can undergo skill level upgrading so that they can ratify the problem personally without any assistance from a maintenance technician. Whereas from the machine side, human errors can be reduced by designing more human oriented machine systems.
- **Condition Based Monitoring:** This strategy falls in the low frequency and high downtime grid. This matches RCM policy where studies and measurement need to be done to determine the underlying reliability condition of the machines. Sensors are usually be applied to feed monitoring data for machine learning in a PdM approach.
- **Design Out Machine/Component:** This strategy falls on the high frequency and high downtime grid. Machines/components that are prescribed in this grid region are usually unable to manage the current production level or are subjected to substantial wear and tear after prolonged usage. Either the need to upgrade to a better machine/components or a replacement of a new machine has to be undertaken.

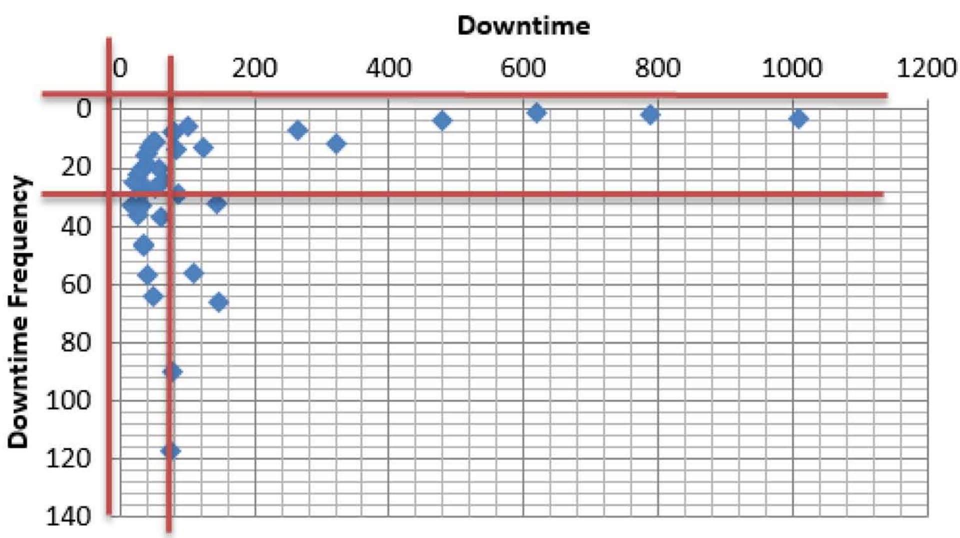
FUZZY LOGIC FOR DIAGNOSTICS MACHINE FAILURE

Fuzzy logic is an approximate reasoning model to estimate the possible outcome based on a set of rules. This method has been proven to be a prominent control system that have been implemented in different engineering application, hardware monitoring and conditional-based assessment. The process of Fuzzy Logic is a simple, rule-based by using IF-THEN statements that imitate the decision making of humans to classify the type of responsive performance by defining the intermediate possibilities. Uncertainty in maintenance management, such as non-linear, imprecise and incomplete knowledge representation can be resolved by the adoption of fuzzy logic. As a consequence, conclusions are derived with certainty factor from a predefined fuzzy sets.

The research focus of predictive maintenance is to discover and recognize critical random failure, involving low frequency and high downtime of maintenance. The guidance from diagnostics in MDSS presents satisfactory knowledge rules to suggest maintenance action afterwards. Fuzzy Logic in KBS is merely a suitable algorithm for machine faliure identification in an imprecise enironment. The linkage between machine failure and classification of maintenance policies are frequently vague and subject to various sensory information. In practice, however, a need to refine the DGM model is required due to two scenarios. The first scenario is when data is located close to each another but at different sides of the policy boundary, leading to applying different strategies, despite being closely similar with one another. The next scenario is when two data points are located at both ends of the boundary within the same grid, leading to applying the same strategies, despite being far apart, as shown in Figure 7. Therefore, the concept of Fuzzy Logic can be applied to reshape the rigid boundaries,and make it more logical for the system.

A Fuzzy Logic design for semiconductor equipment is provided with two factors: downtime records and frequency records as in Figure 8. All membership functions (MF) are selected as trapezium shape, with two numerical inputs and one numerical output involved, as shown in Figure 9. Defuzzification is a method of extracting a crisp value from a fuzzy set as a representative value. In general there are five methods, for defuzzifying a fuzzy set A in a universe of discourse Z. These methods are the Centroid of Area (COA) or Center of Gravity, Mean of Maximum (MOM), Bisector of Area (BOA), Smallest of Maximum (SOM) and Largest of Maximum (LOM). COA is selected as the defuzzification strategy for the model. A fuzzy if-then rule is applied on the interface, as shown in Figure 10. When all rules and membership functions are settled, the Fuzzy Logic Rule Viewer and Fuzzy Logic Surface Viewer can be examined in Figure 11. With the help of fuzzy logic, the

Figure 7. Decision making grid formaintenance policies



known machine failure problem associated with breakdown time and frequency can be reviewed for the current machine. Figure 12 shows the result of machine fault identification from semiconductor A and B

Figure 8. Fuzzy logic: fuzzy inference system editor

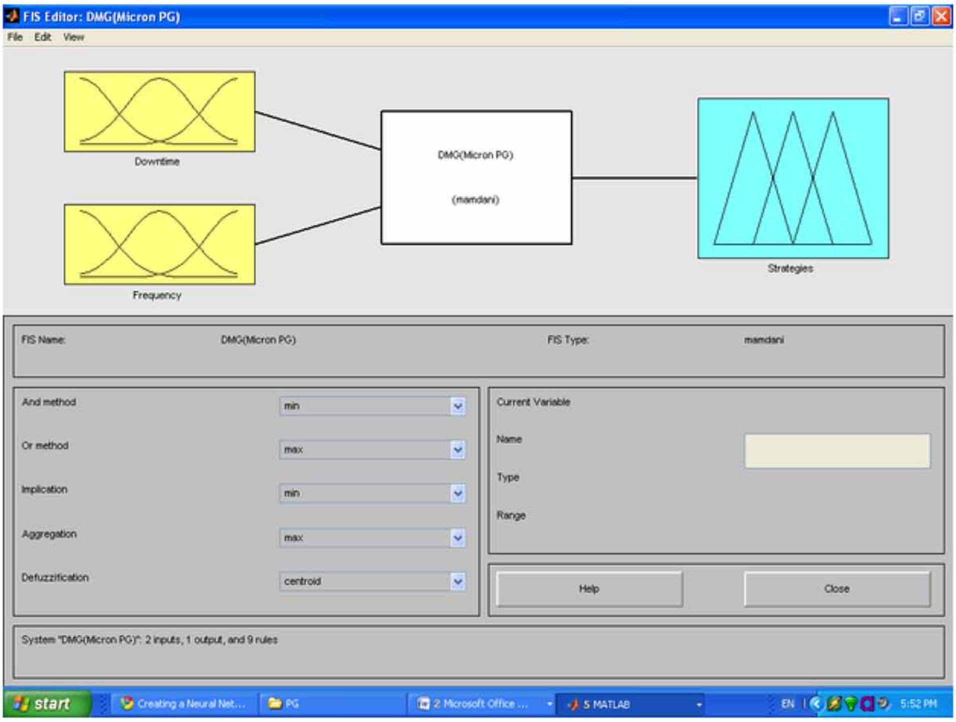


Figure 9. Fuzzy logic membership input function editor: downtime and frequency

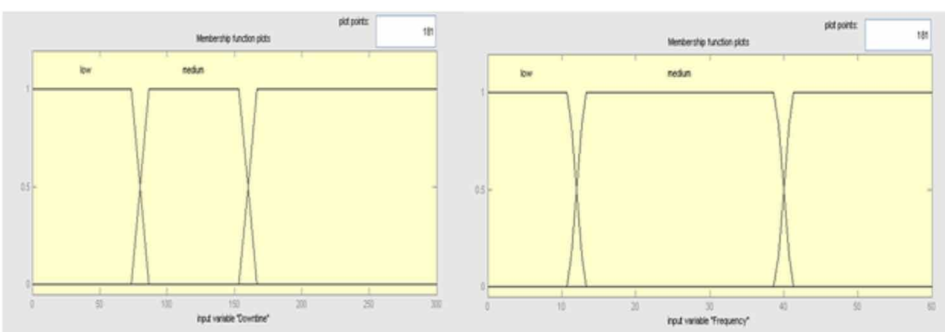


Figure 10. Fuzzy logic MF rules and output: maintenance policies

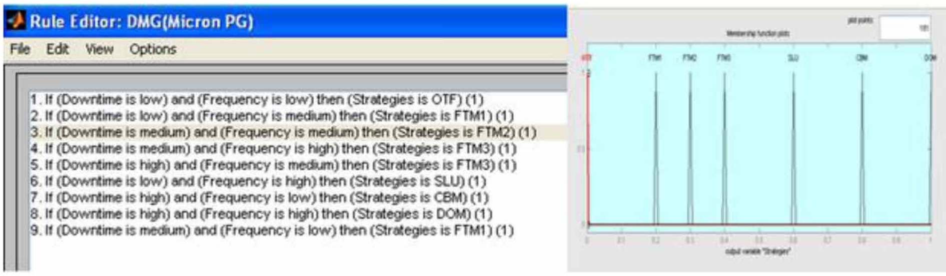
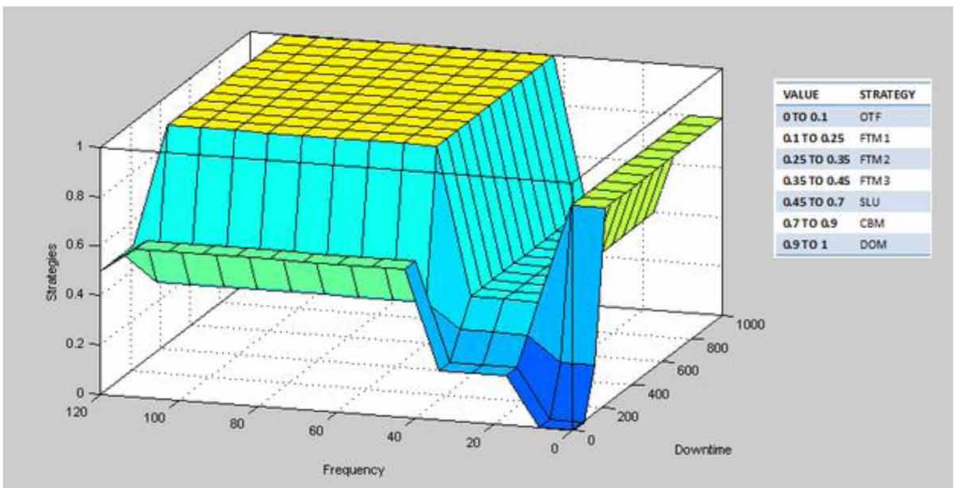


Figure 11. Fuzzy logic surface viewer



ARTIFICIAL NEURAL NETWORK FOR PROGNOSTICS MACHINE FAILURE

Prognostics is a model that is able to predict machine failure compared with normal behavior on a real time basis. Maintenance in advanced is better adopted to the necessity of repairs for an in-service machine in order to mitigate the disruption of the overall production. A health condition assessment is required to check the machine status frequently. Machine failure can be caused by normal degradation and random failure. Basic KBS and DM are incapable of distinguishing random failure and normal degradation, since machine degradation follows time progression rather than by rules. Therefore, an adoptive approach of machine learning to prognosticate the next random failure is required. The Artificial Neural Network

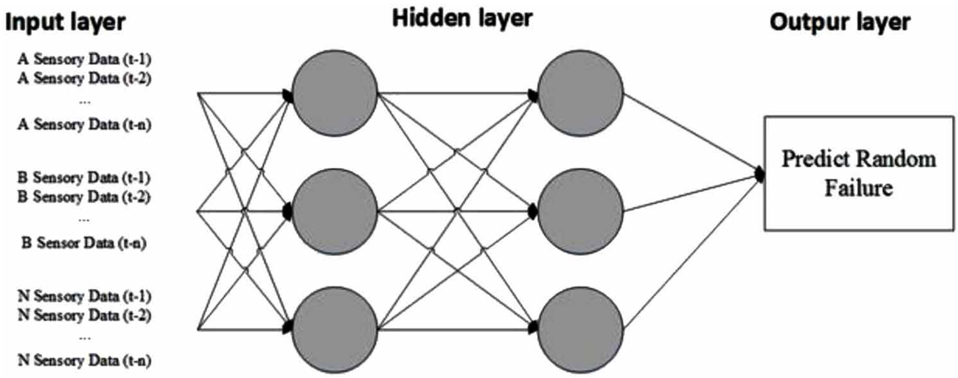
Figure 12. Maintenance policies for seminconductor machine A and B

Machine A	Value	STRATEGY	Machine A	Value	STRATEGY
"'ALIGNMENT SENSOR SERCH ERRO'	0.4	FTM3	"'(4061)AUTO CAN NOT STAR'	0.2	FTM1
"'(2215)BG TAPE PEELING ERROR(PH133'	0.5183	SLU	"'H2 St T-AXIS MOTOR SERVO ALAR'	0.2	FTM1
"'VACUUM SORCE PRESSURE DRO'	0.5537	SLU	"'(1811)CUTTING ERROR(PH064'	0.2	FTM1
"'(2807)LABEL VACUUM ERROR(PSW256'	0.4	FTM3	'True Stall'	0	OTF
"'(1B8C)APPLY STAGE VACUUM ERRO'	0.3	FTM2	"'BACKSIDE OCR READ FAI'	0.2	FTM1
"'SPIN St CYCLE TIME OVE'	0.8	CBM	"'(1B6E)H4 St PAD WAFER NOT CLAMPE'	0.2	FTM1
"'FRONTSIDE OCR READ FAI'	0.6	SLU	Machine B	Value	STRATEGY
"'(2260)PEEL TABLE LOCK ERROR(PSW087'	0.8	CBM	'Storage unit front cover overload'	1	DOM
"'GR2 St WAFER THIN LIMI'	0.3	FTM2	'Disconnected communication with DGP8760'	0.8	CBM
"'(1B6B)H4 St PAD CLAMP ERRO'	0.6	SLU	'Ring frame reading error'	0.6	SLU
"'(2309)UNLOAD WORK DETECT ERR.(PH157'	0.2	FTM1	'BROKEN WAFER'	0.4	FTM3
"'(2420)OCR READ ERRO'	0.8	CBM	'Pick-up tape tip detection error'	0.6	SLU
"'GR1 St WAFER THIN LIMI'	0.8	CBM	'Wafer ID control PC communication timeou'	0.8	CBM
"'SLURRY SUPPLY UNIT ERRO'	0.6	SLU	'Peeling heater head temperature too low'	0.8	CBM
"'PORT2 BCR READ FAI'	0.3	FTM2	'Printer error'	0.6	SLU
"'(1700)W_TABLE LOW END DETECT ERROR'	0.8	CBM	'Positioning TBL: WID GRANT'	0.8	CBM
"'H1 St PAD WAFER NOT CLAMPE'	0.6	SLU	'Arm-2 Z-axis movement time up error. Pls'	0.2	FTM1
"'BROKEN WAFE'	0.2	FTM1	'Alignment unit communication timeout'	0.2	FTM1
"'(3026)FRAME CASS .NOT PLACEMENT(PH220'	0.2	FTM1	'No ring frame in frame stocker'	0.2	FTM1
"'Bubbles trapped between wafer and DC tap'	0.3	FTM2	'No workpiece on frame transfer arm'	0.6	SLU
"'(2805)LABEL IS ON PRINTER SEC.(PH255'	0.2	FTM1	'Spinner detects ionizer abnormal signa'	0.2	FTM1
"'(2822)LABEL PRINT ERRO'	0.2	FTM1	'Wafer ID reading error'	0.6	SLU
'(4100)CASS'	0.2	FTM1	'TAPE MISMOUNT OR OFFSET ON FRAME'	0.6	SLU
"'(1B84)S2F41 TIMEOU'	0.2	FTM1	'Sensor is abnormal when robot is in acti'	0	OTF
"'(1B7D)RM CYCLE TIMEOU'	0.2	FTM1	'Exhaust capacity of the machine decrease'	0	OTF
"'CHILLER WATER FLOWLES'	0.2	FTM1	'No workpiece on workpiece storage guide'	0.2	FTM1
"'(151)300mm FRAME DETECT ERROR(LS044'	0.2	FTM1	'No change in sensors when atomizing nozz'	0.8	CBM
"'POS St ALIGNMENT CYCLE TIME OVE'	0.2	FTM1	'AVG THICKNESS OUT-OF-CONTROL (OOC)'	0.2	FTM1
"'(150)200mm FRAME DETECT ERROR(LS043'	0.1002	FTM1	'True Stall'	0	OTF
"'(2803)LABEL CHUCK HEAD LOW DET.ERR(L253'	0.2	FTM1	'Host comm: Received terminal message. P'	0.2	FTM1
"'PORT-2 pdo cannot reach positon d'	0.0734	OTF	'Error recovery completed'	0.8	CBM
"'Backside stai'	0.8	CBM	'T3 Time-out in Host Communication.'	0.2	FTM1
"'GR1 St WAFER THICK LIMI'	0.2	FTM1	'Remote command timeout'	0.2	FTM1
"'(3109)FRAME STOCKER COVER OPEN(CSW001'	0.1	FTM1	'Barcode: ID of cassette B cant be read.'	0.8	CBM

(ANN) model is a supervised learning method to estimate the Remaining Useful Lifetime (RUL) of a machine from degraded failure, or to differentiate anomalies from normal machine behavior. ANN can be used as a time series or forecasting of unanticipated machine failure with online sensory data, which is able to learn patterns from the training data set to distinguish the normal machine behavior and any anomalies. The prediction in machine failure involves Big Data management with the purpose of strengthening the data quality. Continuous data collection from different sensors installed in the machine supports the quality of prediction.

Figure 13 illustrates the mechanism of the ANN model. The basic ANN algorithm estimates the feasible outputs for prediction maintenance from observation data or sensory information in order to understand the machine condition. ANN provides assurance in resolving prediction and performance assessment with suffi-

Figure 13. Artificial neural network



cient continuous input data. Vibration, heat and temperature sensors are established for the evaluation of the semiconductor machine. ANN is a simple artificial neuron processing system, which included input nodes, number of hidden layers, weighted factors of adjacent layers and output nodes. The hidden layers function as connectors between the input nodes and output nodes to transform and extract features of the input space. The prediction accuracy of ANN is dependent on the training process. The training process allows adjustment the synaptic weighting of input and develops the overall network by the construction of hidden layer. Before the actual running of the prediction, correct data and wrong data must be available in the training of supervised learning algorithm to ascertain its reliability in prediction. As a rule of thumb, 60% of the dataset is needed for training and the remaining dataset is for running test. Big Databases are able to collect huge amounts of historical and online data to enhance the accuracy of an ANN model through trial and error.

The predictive tasks in MDSS are presented in the following three main functions:

- **Health Condition Assessment:** The main goal is to monitor the machine performance and machine components degradation. In MPM, considerable maintenance actions could be resolved by component replacement. The sensor network can effectively enhance MRO order processing and improve flexibility in maintenance scheduling for the component substitution. The machine downtime caused by component damage could be reduced to meet the aim of agile production recovery.
- **Anomalies Detection Assessment:** Anomalies detection is an approach in root cause analysis. Machine anomalies may take place when the machine are not working normally. The cause of consistency anomalies is not straight-

forward to detect. ML algorithms are capable of measuring the machine performance from the current and normal conditions for identifying anomalies. The features of abnormality in machine performance are provided for further inspection by the repair technicians.

- **Remaining Useful Lifetime Assessment:** The major challenge in asset management is to optimize the machine lifetime utilization. Different machine usage may conclude variability of depreciation. In addition, longer lead time in machine replacement causes an uncertainty in the production and supply chain performance. The estimation of remaining useful lifetime by sensory information can effectively leverage the machine lifetime utilization and prediction for the machine replacement and performing a just-in-time machine replacement policy.
- **Scheduling Optimization for Maintenance in Advance:** Prediction of foreseeable machine failure is used to assess the machine expected performance in the future and optimize the maintenance schedules with less adverse effect on production disruption. Unanticipated machine failure may give a longer downtime and cease the manufacturing process. The prediction from the ANN model emphasizes proactive maintenance and provides the right time to conduct inspection. ANN captures the possible reasons, like the timeline of the machine failure event, reasons, duration and relevance information of the machines by triggered sensory information. With systematic maintenance in advanced, unplanned machine stopping can be eliminated.

MANAGERIAL IMPLICATIONS AND RECOMMENDATIONS

The entire Big Data framework requires human intelligence and expert opinion in the design stage. It is difficult for the manufacturer to manage and, control big data and select relevant information for MDSS. The availability of sensors and technological advancement enable explorative research of Big Data Analytics, and allows organizations to expand their capability to enhance the data transparency of machine status for manufacturers. The speedy data flow and collection of abundant data through WSNs enhance the potential of analytical performance. The adoption of Big Data in MDSS state a significant step in machine health condition diagnosis. The proposed approach helps to mitigate machine failure during production and uncover the hidden patterns through Big Data Analytics. However, management faces three major challenges of transformation traditional maintenance to advanced MDSS.

Various industries noticed that the size of data has been exponentially increasing and accelerating due to the comprehensive use of sensor network. The transition from conventional database to non-relational database is not only upgrading the storage capacity, but also requiring an infrastructure and expertise to process, and handle structured and unstructured data. Handling and understanding on the petabytes or even exabyte of data has become a challenge for IT teams. More advanced capability in data warehouses and network connection expedites real time data processing.

Due to the complexity of data type, the current processing techniques could not meet the demand of Big Data Infrastructure. Due to the enormous data booming via WSNs, a cohesive platform for processing structure and unstructured data becomes an essential element for any enterprises. The major purpose of Big Data Infrastructure is to resolve the problem of incompatible data formats and non-aligned data structures. The impact on inconsistency of unstructured data requires pre-processing of data input to enhance the performance of Big Data Mining.

Investigating the unstructured data in manufacturing not only create the value for the production engineer but also support MDSS with more vigorous and sophisticated Big Data Analytics. Several research paper mentioned that analyzing the unstructured data is the first priority in decision-making and prediction (Li, Bagheri, Goote, Hasan, & Hazard, 2013; Muhtaroglu, Demir, Obali, & Girgin, 2013; Wielki, 2013). However, not all the unstructured data can be beneficial to knowledge development and decision-making process. The relevant machine data must be fit for purpose of maintenance policy selection. Proper domain experts in place are critical to interpret the sensory information for predictive maintenance during the design stage of Big Data Analytics. Furthermore, enhancing data quality by the adoption of suitable sensors in the machine is also an importance for the company. Big Data offers tremendous insight to the diagnostics and prognostics of the machine status. Nonetheless, the information reliability from predictive maintenance is only available with appropriate sensors selection and adoption. In today's Big Data Analytics, research focus has been shifted from Volume of data to quality data.

Regarding the complexity of Big Data Analytics in MDSS, collaboration of industrial expertise and scholars must be involved to have sufficient breadth and depth of domain knowledge to design an appropriate Big Data Analytics for maintenance strategies. The proactive approach to predict machine failure provides a high level reliability for excellence in maintenance management. Further benefit can be summarized as reducing the frequency of corrective maintenance, increasing machine performance and enhancing overall production reliability.

FUTURE RESEARCH DIRECTIONS

Future research is oriented to the utilization of manufacturing information in the Cyber Physical System (CPS). Big Data Analytics are able to achieve better transparency of production, which provides knowledge insight to practitioners. With the technological advancement in Internet of Things and the utilization of sophisticated prediction tools, specialized automotive networks in a manufacturing company could be developed for real time monitoring and control at the strategical level. Incorporating the manufacturing computational intelligence into the machine health monitoring allow the manufacturers to enhance the overall system reliability and production efficiency, especially in reducing the machine downtime. Predictive maintenance is not only about health assessment but also detect the abnormality of the machine before it breaks down. To have a step forward from predictive maintenance, production plant should realize the importance of just-in-time maintenance strategy for the whole production process. Implementation of CPS synthesizes data from WSNs to enable the remaining life prediction so as to improve the asset utilization. Risk assessment and impact evaluation of machine failure will also allow production engineers to estimate the production system reliability. This motive turns the predictive manufacturing system into a "self-aware-and-self-adjustment" system, with intelligent machines and sensors in Big Data era.

CONCLUSION

In this book chapter, the predictive maintenance model and Big Data Analytics in managerial aspects are presented. The feature extraction through Big Data Analytics can be beneficial to managing the machine condition and in predicting machine failure. The MDSS in Big Data is able to suggest maintenance strategies and provide insight for management to tackle maintenance issues. The proactive strategies in maintenance can be achieved by embedded sensors and real time based machine monitoring systems. Besides, the prediction from Big Data Analytics and suggested analytics processes are well-designed to reduce the maintenance turnaround time and substantially enhance the production system availability. MRO and maintenance resources can be planned in advanced to facilitate the process during the machine downtime. Moreover, it provides flexibility to design maintenance schedules to mitigate the risk of unplanned stoppings. The overall maintenance efficiency can be much improved by the implementation of predictive maintenance under a Big Data platform.

ACKNOWLEDGMENT

The research is supported by The Hong Kong Polytechnic University. The authors would like to thank the case company for providing the data. Our gratitude is also extended to the Research Committee and the Department of Industrial and Systems

Engineering of the Hong Kong Polytechnic University for support in this project (RUE9) and (RU8H).

REFERENCES

Exton, T., & Labib, A. (2002). Spare parts decision analysis–The missing link in CMMSs (Part II). *Journal of Maintenance & Asset Management*, *17*(1), 14–21.

Faiz, R., & Edirisinghe, E. A. (2009). Decision making for predictive maintenance in asset information management. *Interdisciplinary Journal of Information, Knowledge, and Management*, *4*, 23–36.

Fu, C., Ye, L., Liu, Y., Yu, R., Iung, B., Cheng, Y., & Zeng, Y. (2004). Predictive maintenance in intelligent-control-maintenance-management system for hydroelectric generating unit. *Energy Conversion*. *IEEE Transactions on*, *19*(1), 179–186.

Garcia, M. C., Sanz-Bobi, M. A., & del Pico, J. (2006). SIMAP: Intelligent System for Predictive Maintenance: Application to the health condition monitoring of a windturbine gearbox. *Computers in Industry*, *57*(6), 552–568. doi:10.1016/j.compind.2006.02.011

Jardine, A. K., & Tsang, A. H. (2013). *Maintenance, replacement, and reliability: theory and applications*. CRC Press.

Kaiser, K. A., & Gebraeel, N. Z. (2009). Predictive maintenance management using sensor-based degradation models. *Systems, Man and Cybernetics, Part A: Systems and Humans*. *IEEE Transactions on*, *39*(4), 840–849.

Klutke, G.-A., Kiessler, P. C., & Wortman, M. (2003). A critical look at the bathtub curve. *IEEE Transactions on Reliability*, *52*(1), 125–129. doi:10.1109/TR.2002.804492

Lee, J., Ardakani, H. D., Yang, S., & Bagheri, B. (2015). Industrial big data analytics and cyber-physical systems for future maintenance & service innovation. *Procedia CIRP*, *38*, 3–7. doi:10.1016/j.procir.2015.08.026

Lee, J., Kao, H.-A., & Yang, S. (2014). Service innovation and smart analytics for industry 4.0 and big data environment. *Procedia CIRP*, *16*, 3–8. doi:10.1016/j.procir.2014.02.001

Li, L., Bagheri, S., Goote, H., Hasan, A., & Hazard, G. (2013). *Risk adjustment of patient expenditures: A big data analytics approach.* Paper presented at the Big Data, 2013 IEEE International Conference on. doi:10.1109/BigData.2013.6691790

Muhtaroglu, F., Demir, S., Obali, M., & Girgin, C. (2013). *Business model canvas perspective on big data applications.* Paper presented at the Big Data, 2013 IEEE International Conference on. doi:10.1109/BigData.2013.6691684

Sikorska, J., Hodkiewicz, M., & Ma, L. (2011). Prognostic modelling options for remaining useful life estimation by industry. *Mechanical Systems and Signal Processing*, *25*(5), 1803–1836. doi:10.1016/j.ymssp.2010.11.018

Wielki, J. (2013). *Implementation of the big data concept in organizations-possibilities, impediments and challenges.* Paper presented at the Computer Science and Information Systems (FedCSIS), 2013 Federated Conference on.

Wu, S.-j., Gebraeel, N., Lawley, M. A., & Yih, Y. (2007). A neural network integrated decision support system for condition-based optimal predictive maintenance policy. *Systems, Man and Cybernetics, Part A: Systems and Humans*. *IEEE Transactions on*, *37*(2), 226–236.

Xin, N. Y., & Ling, L. Y. (2013). *How we could realize big data value.* Paper presented at the Instrumentation and Measurement, Sensor Network and Automation (IMSNA), 2013 2nd International Symposium on.