THE STUDY OF ANFIS-BASED AND OTHER CLASSIFIERS FOR HEAD PERFORMANCE DOWNGRADE

RANGSON HUNKON

A THESIS SUBMITTED IN PARTIAL FULFILLMENT
OF THE REQUIREMENT FOR THE DEGREE OF
MASTER OF ENGINEERING IN DATA STORAGE TECHNOLOGY
COLLEGE OF DATA STORAGE INNOVATION
KING MONGKUT'S INSTITUTE OF TECHNOLOGY LADKRABANG
2014
KMITL-2014-IC-M-005-003

THE STUDY OF ANFIS-BASED AND OTHER CLASSIFIERS FOR HEAD PERFORMANCE DOWNGRADE

RANGSON HUNKON

A THESIS SUBMITTED IN PARTIAL FULFILLMENT
OF THE REQUIREMENT FOR THE DEGREE OF
MASTER OF ENGINEERING IN DATA STORAGE TECHNOLOGY
COLLEGE OF DATA STORAGE INNOVATION
KING MONGKUT'S INSTITUTE OF TECHNOLOGY LADKRABANG
2014
KMITL-2014-IC-M-005-003

THE STUDY OF ANFIS-BASED AND OTHER CLASSIFIERS FOR HEAD PERFORMANCE DOWNGRADE

RANGSON HUNKON

A THESIS SUBMITTED IN PARTIAL FULFILLMENT
OF THE REQUIREMENT FOR THE DEGREE OF
MASTER OF ENGINEERING IN DATA STORAGE TECHNOLOGY
COLLEGE OF DATA STORAGE INNOVATION
KING MONGKUT'S INSTITUTE OF TECHNOLOGY LADKRABANG
2014
KMITL-2014-IC-M-005-003

ชื่อหัวข้อวิทยานิพนธ์:	การศึกษา ANFIS และวิธีการอื่นๆ เพื่อจำแนกการลดประสิทธิภาพของหัวอ่าน
หน่วยกิตของวิทยานิพนธ์:	12 หน่วยกิต
นักศึกษา:	นาย รังสรรค์ หุ่นโกน
อาจารย์ที่ปรึกษาวิทยานิพนธ์:	รศ.ดร.พรชัย ทรัพย์นิธิ
อาจารย์ที่ปรึกษาวิทยานิพนธ์ร่วม:	รศ.ดร.ปิยะ โควินท์ทวีวัฒน์
หลักสูตร:	วิศวกรรมศาสตรมหาบัณฑิต
สาขาวิชา:	เทคโนโลยีการบันทึกข้อมูล
คณะ:	วิทยาลัยนวัตกรรมการจัดการข้อมูล

บทคัดย่อ

อาการเสียของฮาร์ดดิสก์ไดรฟ์ (HDD) ที่เกิดขึ้นในกระบวนการทดสอบคุณภาพนั้น เป็นตัวแปรสำคัญที่จะส่งผลกระทบต่อประสิทธิภาพ การขับเคลื่อนของโรงงานและผลผลิตฮาร์ดดิสก์ไดรฟ์ โดยทั่วไปอาการเสียที่เกิดขึ้นนั้น ส่วนมากมีสาเหตุมาจาก หัวอ่าน/เขียน ที่เกิดประสิทธิภาพการทำงานลดลง ดังนั้น งานวิจัยชิ้นนี้จะนำเสนอวิธีในการจำแนกอาการเสียของฮาร์ดดิสก์ไดรฟ์ (HDD) ซึ่งจะสามารถช่วยปรับปรุงคุณภาพและความน่าเชื่อถือของฮาร์ดดิสก์ไดรฟ์ (HDD) โดยการใช้เครื่องมือวิเคราะห์ ที่เรียกว่า Adaptive neuro-fuzzy inference system (ANFIS) เพื่อที่จะจำแนกการลดประสิทธิภาพการทำงานของหัวอ่าน จากนั้นเปรียบเทียบความถูกต้อง (Accuracy) กับวิธีการจำแนกแบบอื่นๆ เช่น neural network (NN), naive bayes (NB) และ support vector machine (SVM) เพื่อหาโอกาสที่จะลดอาการเสียที่เกิดขึ้นในกระบวนการทดสอบคุณภาพ จากการทดลองครั้งนี้ เราพบว่า เครื่องมือวิเคราะห์ Adaptive neuro-fuzzy inference system (ANFIS) นี้สามารถใช้ในการทำนายอาการเสียที่เกิดขึ้นในกระบวนการทดสอบคุณภาพของฮาร์ดดิสก์ไดรฟ์ (HDD) ได้ด้วยความถูกต้องสูงสุดประมาณ 80%

Thesis	THE STUDY OF ANFIS-BASED AND OTHER CLASSIFIERS FOR HEAD PERFORMANCE DOWNGRADE
Student	Mr. Rangson Hunkon
Student ID	53600605
Degree	Master of Engineering
Program	I/U CRC in Data Storage Technology and Application
Year	2014
Thesis Advisor	Assoc. Prof. Dr. Pornchai Supnithi
Thesis Co-Advisor	Assoc. Prof. Dr. Piya Kovintavewat

ABSTRACT

The hard disk drive (HDD) quality failure is an important issue that crucially impacts factory efficiency and productivity. In general, a major failure is caused by head performance downgrade. Thus, this paper proposes a method to predict this failure, which in turn can help to improve HDD reliability. Specifically, we apply an analytic tool called an adaptive neuro-fuzzy inference system (ANFIS) to predict the head performance downgrade and then compare accuracy performance with other classifiers such as neural network (NN), naive bayes (NB) and support vector machine (SVM), to find out an opportunity to reduce the quality failure. We found from an experiment that the ANFIS model can be employed to predict a failure from head performance downgrade at quality test with the best accuracy of about 80%.

Acknowledgements

I would like to thank you Assoc.Prof.Dr.Pornchai Supnithi (King Mongkut's Institute of Technology Ladkrabang) and Assoc.Prof.Dr.Piya Kovintavewat (Data Storage Technology Research Center, Nakhon Pathom Rajabhat University), who provided me the guidance, advices and reviews my work. Also, I would like to thank WD Company and my supervisor K. Surasith Phongtharapat to support everything during my study and my thesis. Finally, I would like to specially thank my family for the motivation and encouragement during my study.

Mr. Rangson Hunkon

Contents

PAGE

Contents (Con't)

Contents (Con't)

PAGE

List of tables

PAGE

List of figures

List of figures (Con’t)

PAGE

Chapter 1

Introduction

1.1 Introduction

Hard disk drive (HDD) is a growing business due to worldwide customer demand. HDDs are important for data storage in every company and personal use (e.g., desktop, notebook, tablet, game boxes, and so on). Hence, HDD manufacturer must improve the processes continuously to achieve the highest output with the highest quality to support customer requirements. Generally, the HDD manufacturing process is separate into two processes. The first step is done in a class-100 clean room [1], where contamination is carefully controlled. Next, the second one is performed at back end area where drives are tested for performance and quality control, then ship them to the customers, as illustrated in Figure 1.1.

Practically, HDD manufacturer must continuously improve the manufacturing processes and the quality performance so as to support customer's requirements. However, we still found many HDD failures returned from customers. In general, the failure rate at the quality test stage is about 1.5%, whereby a major cause is head performance downgrade. Head downgrade is a top one of failure in hard disk drive process because drives do not meet target of the quality test, there are many root causes of head performance downgrade, for example, incoming head low performance, assembly process induced, back end test process induced, handling induced, etc. This thesis is aimed to apply existing classifiers to identify such potential failures and predict the failures at the quality test stage so as to let them fail at the back end test process and as a result prevent failed drives shipped to the customers.

To study the prediction and reduce head performance downgrade at quality test, we found that the existing key head stack parameters are baseline popping (BLP), variable gain amplifier (VGA), signal-to-noise ratio (SNR), magnetic erasure width (MEW), overwrite (OW), and error margin (EM) from the back end test process. These parameters are used as input for the proposed analytic tool called an adaptive neuro-fuzzy inference system (ANFIS) to predict the failures.

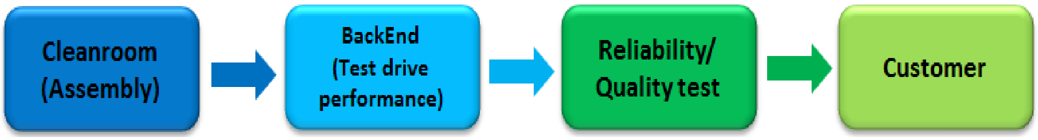


Figure 1.1 A flow of hard disk drive manufacturing process.

1.2 Problem statement

Hard disk drive manufacturing has many minor processes starting from clean room assembly to back end test. During these processes, some drives may pass the performance test limit but they fail during test at reliability due to head performance downgrade. Currently in the production test, we found that some drives must be downgraded during the test process because they do not support the target capacity from the first build in the clean room. The effects of this problem impact reliability failures and cause the customer drives to be returned. It is therefore necessary to predict and reduce reliability failures head downgrade issue.

1.3 Objectives

The goal of this study is to predict the significant parameters which affect head downgrade at the reliability test. To achieve this goal, the specific objectives are as follows:

a) Study the key head parameters concerned with head downgrade.
b) Classify the key head parameters that relate to head downgrade.
c) Predict failures by using key head parameters in back end test process.
d) Use the ANFIS model to analyze the relationship between each parameter and the head downgrade.
e) Use the ANFIS model to predict failure probability in the current production.
f) Evaluate the proposed ANFIS models on test data in the real production.
g) Compare ANFIS model prediction performance with other data mining methods such as neural network, naive bayes and support vector machine in RapidMiner Studio 6.

1.4 Outcomes and expectation

The outcomes of this thesis are the study result and understanding of the key head parameters that cause the head performance downgrade at the reliability test and improve the knowledge for classification data technique likes ANFIS model, neural network model, naive bayes model and support vector machine model to predict the reliability failures in hard disk drive manufacturing process. The challenge of this thesis is accuracy of prediction head downgrade must be above 80%. We expect that this thesis can be used in the hard disk drive production to predict and prevent drives before shipped to the customers.

1.5 Hypothesis

Since there are many head parameters which can be used to identify head performance. This thesis will study six key parameters: baseline popping (BLP), variable gain amplifier (VGA), signal-to-noise ratio (SNR), magnetic erasure width (MEW), overwrite (OW), and error margin (EM) from the back end test process to predict head downgrade at the reliability test.

1.6 Scope of the study

The scopes of this thesis are described below.

a) Use one HDD product model to study.
b) Study & use six parameters to experiment for prediction head downgrade.
c) Use ANFIS model to analyze and experiment data.
d) Use data mining models such as neural network, naive bayes and support vector machine to compare performance with ANFIS model.
e) Measure accuracy from experiment data and conclusion.
f) Apply concept head downgrade prediction into production.

Chapter 2

Theory

In this chapter, we overview hard disk drive (HDD) technology and the theories related to an adaptive neuro-fuzzy inference system (ANFIS) model, data mining techniques including neural network, naive bayes and support vector machine, advantages of fuzzy logic versus data mining.

2.1 History of hard disk drives

Recording data on hard disk drive is based on the magnetic recording principle. The magnetic recording technology has been improved so that it can increase the storage capacity. This section will provide the history of the hard disk drive recording technology and the structure of the recording system.

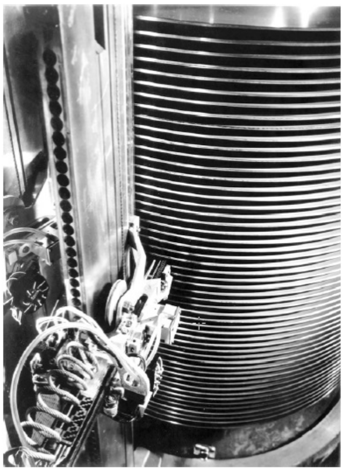


Figure 2.1 Hard disk drive IBM model 305 RAMAC. [2]

The IBM 305 RAMAC [2] was the first commercial computer that used a moving-head hard disk drive (magnetic disk storage) for secondary storage as illustrated in Figure 2.1. The system was publicly announced on September 14, 1958, with test units installed at the U.S. Navy and at private corporations. It can record the data up to 5 Megabytes on 50 recorded disks and each disk has 24 inches of diameter [2]. The capacity per area is 2 Kbits/in2 (kilobits per square inch).

2.1.1 Reading and writing of hard disk drive. The magnetic recording technology can be explained by using north and south pole direction of magnetic where both are aligned with the two states of binary data which are "0" and "1". In the past, longitudinal recording was used whereby the recording data direction was parallel with the magnetic recording medium as shown in Figure 2.2. However, there is some limitation for longitudinal magnetic recording since it has reached the saturation point for increasing recording capacity [3]. Consequently, the recording technology has changed to perpendicular magnetic recording whereby the direction of recorded data is perpendicular to magnetic recording medium as shown in Figure 2.3. Moreover, the perpendicular magnetic recording can increase the capacity up to 1 Tbit/in^2 by using the discrete track perpendicular magnetic recording. [3]

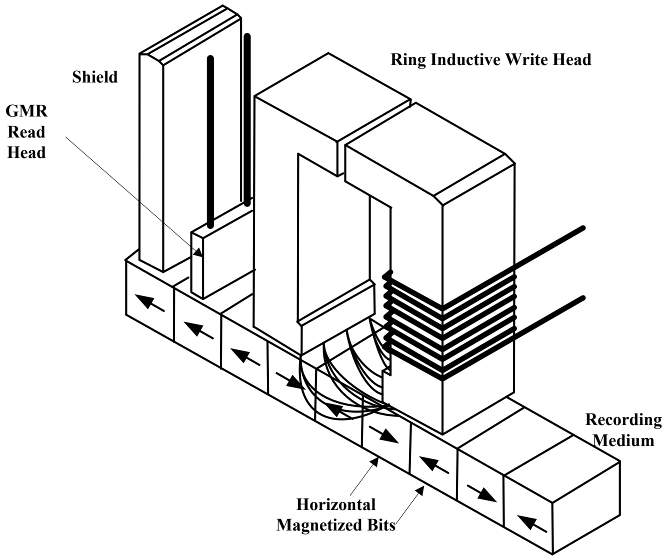


Figure 2.2 Longitudinal magnetic recording. [3]

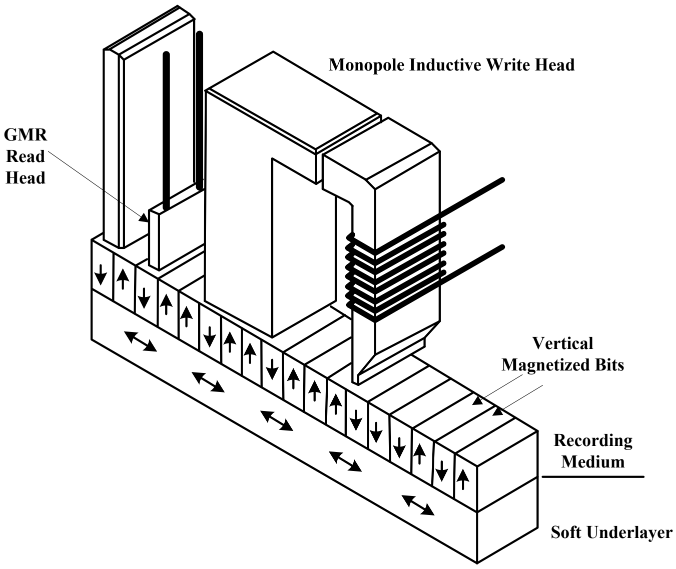


Figure 2.3 Perpendicular magnetic recording. [3]

2.2 Technology roadmap of hard disk drives

According to the latest roadmap of ATSC (Advanced Storage Technology Consortium), mechanical storage densities increase at an accelerated rate starting from a few years as illustrated in Figure 2.4, after heat-assisted and bit-patterned recording kicked into gear. It is supposed to continue at least until year 2025, when the areal densities become high enough to enable 100TB hard disk drives.

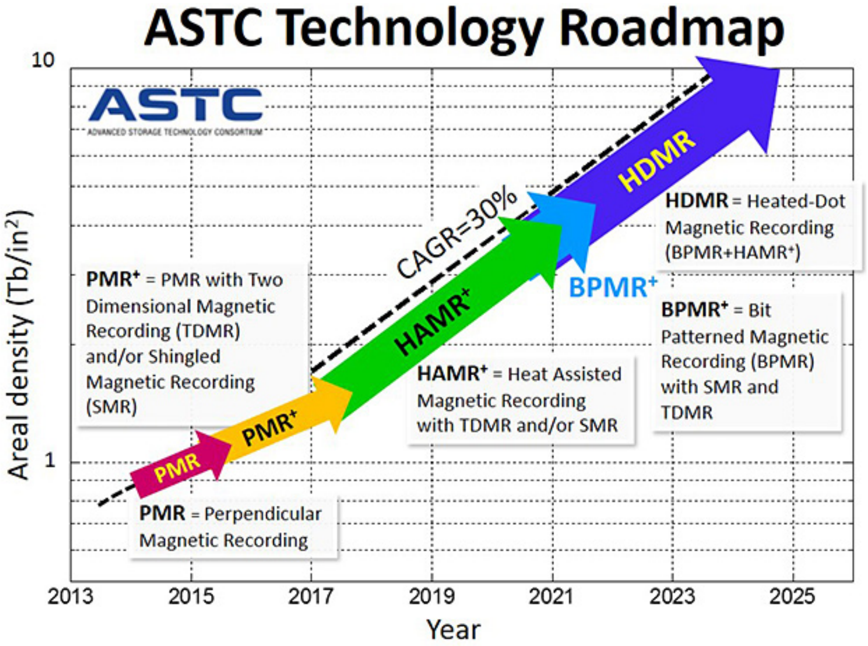


Figure 2.4 Technology roadmap of hard disk drives. (*Source: ASTC)*

At present, we are in the area of perpendicular technology, with areal densities under 1Tb/in². Shingled magnetic recording has already appeared in datacenter drives optimized for cold storage. Along with the two-dimensional recording, a similar and potentially complementary approach, shingled technique poised to drive density scaling for at least until year 2017. However, after heat-assisted and bit-patterned recording are combined which is around year 2021, it will push areal densities up to 10Tb/in² by year 2025.

There is the hard disk drive technology roadmap to increase data storage areal densities by techniques as below.

2.2.1 Shingled magnetic recording (SMR) is a magnetic storage data recording technology used in hard disk drive (HDDs) to increase storage density and overall per-drive storage capacity. Conventional hard disk drives record data by non-overlapping magnetic tracks writing in parallel to each other, while shingled recording writes new tracks on the overlap part of the previously written magnetic track, leaving the previous track thinner and allowing for higher density. Thus, the partially overlap track are similar to roof shingles. This approach was selected because its physical limitations can prevent the recording magnetic heads from having the same width with the reading heads, leaving recording heads wider.

2.2.2 Two dimensional magnetic recording (TDMR) is a novel recording architecture intended to support densities beyond those of conventional recording systems. The gains from TDMR primarily come from the more powerful coding and signal algorithms processing that allow the bits to be more tightly packed on the disk, and yet be retrieved with acceptable error rates. The main challenges involved in TDMR arise from the small bit-area, along with an aggressive write/read process, which lead to a large amount of noise, and the two-dimensional nature of the recording process, so far not encountered in today's systems. Thus, a gamut of 2D signals processing algorithms need to be developed for the compensation of errors occurring due to the aggressive write/read processes.

2.2.3 Heat assisted magnetic recording (HAMR) is a potential technology which can enable the increasing of density in the next magnetic hard disk drive generation. Current technologies utilize a magnetic storage medium consisting of small magnetic grains, which are bistable and able to retain their magnetic state for at least 10 years. The data is written on the storage medium by using an electromagnet, which generates a magnetic field as it passes over the spinning medium and changes the magnetic state of the grains. When we try to increase the amount of data in the same physical area a problem arises. To retain sharp bit transitions, where the boundary between the “0” and “1” orientated areas of the media are well defined, the grain size of the medium must be reduced. This then requires an increase in the magnetic anisotropy of the grains to ensure that the magnetic state is stable for greater than 10 years. However, the magnetic field generated by the electromagnet (or write head) is limited by the material and theoretically it is around 2 Tesla. Now, the magnetic field has required to reverse the grains depends on the magnetic anisotropy of the material, and so there are the conflict in “thermal stability” and “writability” requirements. This problem is known as the magnetic recording trilemma and this encapsulates the key principles of magnetic media design.

2.2.4 Bit patterned magnetic recording (BPMR) is a potential future hard disk drive technology to record data in magnetic islands (one bit per island), as opposed to the current hard disk drive technology where each bit is stored in 20-30 magnetic grains within a continuous magnetic film. The islands would be patterned from a precursor magnetic film using nanolithography. It is one of the proposed technologies for the success of perpendicular recording since it would enable for the greater storage densities.

2.2.5 Heated-dot magnetic recording (HDMR) is a combination between heat assisted magnetic recording (HAMR) and bit patterned magnetic recording (BPMR).

2.3 Components of hard disk drive

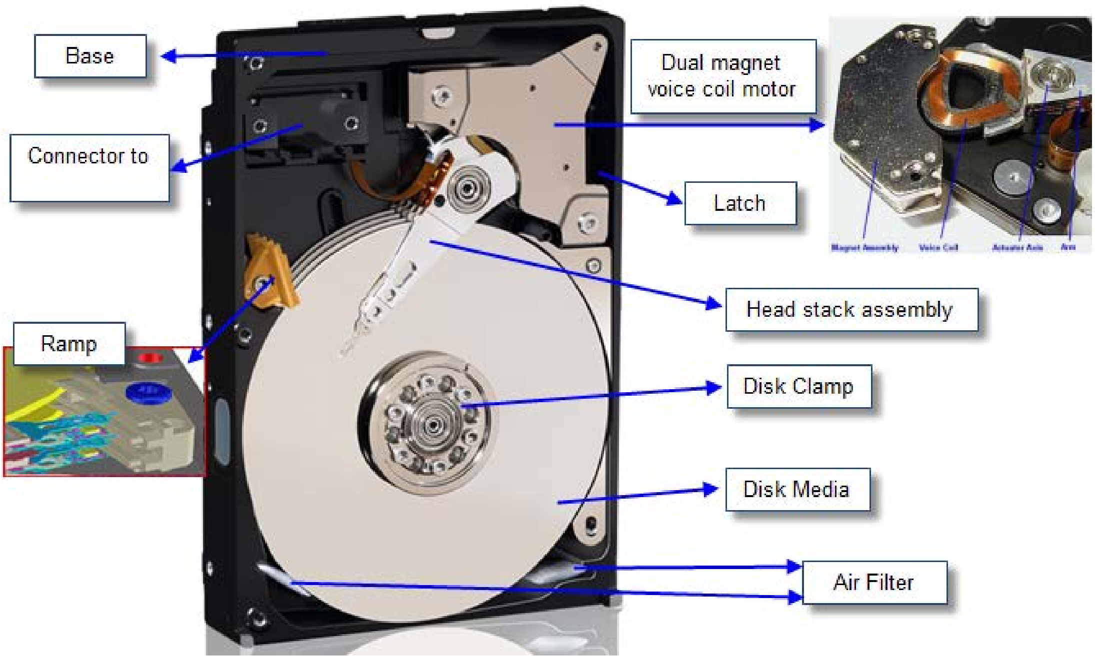


Figure 2.5 Components of hard disk drive. (*Source: CW Mak 2009*)

Hard disk drive consists of many mechanical components inside, all parts of the hard disk drive are important for the recording data system and its performance, Figure 2.5 illustrates the main hard disk drive components.

2.3.1 Disk media composes of magnetic bits used to record and retrieve data in binary format. The Capacity of a hard disk drive depends on the total amount to disk media available and areal density. It is attached to the motors by a disk clamp.

2.3.2 Head stack assembly (HSA) is a combination of slider, head arm, pivot, and voice coil motor. It reads and writes while flying 9-12 nm above the disk media around, when converting electrical pulses into magnetic transitions.

2.3.3 Dual magnetic voice coil motors are installed on the top and bottom of the voice coil and used to move the head stack assembly to the proper place. It works by Faraday's law. The current is induced when the voice coil motor moving between top and bottom magnet.

2.3.4 Filters are placed in the hard disk drive in two places. The first filter is called an air recirculation filter that used to entrap contamination. A second filter which is a breather filter allows the air to pass or leave as air pressure.

2.3.5 Disk clamp is installed on the top of the disk media to hold the media during the work of hard disk drive to help balancing the disk media.

2.3.6 Base and top cover are the components for particular external contamination protection. Some types of base can have the built in base motor.

2.3.7 Ramp is the place for the head stack assembly to stop, when turning off hard disk drive. It prevents the head landing on the media when the hard disk drive is shutting down.

2.3.8 Latch prevents the head stack from crashing with motor and controls its movement during head stack stops working.

2.4 Test process manufacturing

Hard disk drive assembly starts in a clean room and then performs test drives performance at back end area as illustrated in Figure 2.6.

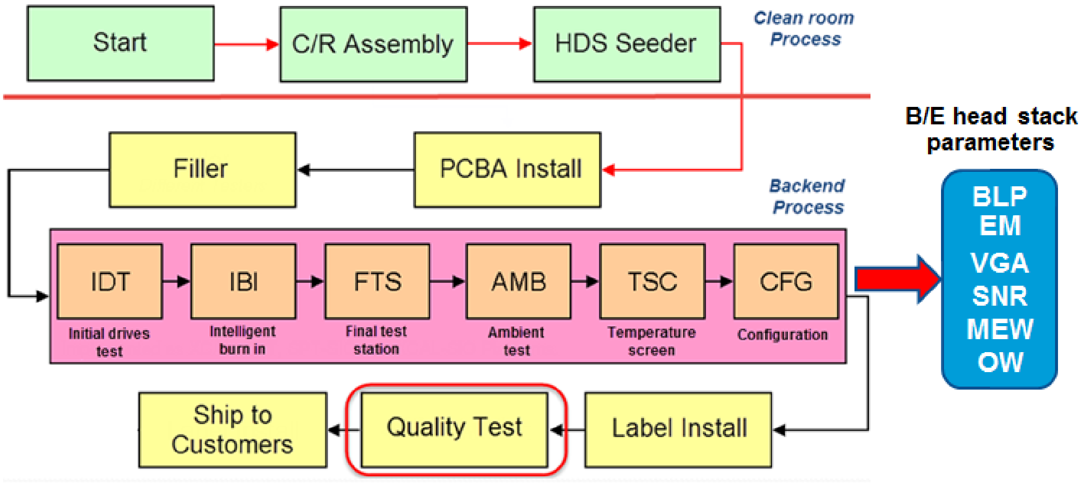


Figure 2.6 Generic HDD Process Flow.

2.4.1 Clean room process. The first process is to clean the parts with chemical liquid. The second process is to assemble the bottom voice coil magnet to the base then perform ramp installation. The ramp is the landing zone for the actuator when the hard disk drive is powered off. The next process is media installation followed by clamp bias installation for media holding while the hard disk drive is turning on. After this process, the operator must test the balancing and acceleration for the balance and imbalance of the media during spin up. Ramp installation is in the next process, and it requires the operator to inspect the gap between disk and ramp. The head stack assembly is installed before the top voice coil motor magnet is installed to complete the circuit. Latch is the last part which is installed in the hard disk drive, after the assembly process HDS seeder has performed written servo signal onto magnetic recording disk.

2.4.2 Back end test. After assembly, we send out the hard disk drive to the back end area for printed circuit board assembly (PCBA) installation, and then go to Filler to write data track onto magnetic recording disk. After Filler process, we send hard disk drive to the back end test process. This station is very important for testing the hard disk drive performance. After completing the back end process, we send them to Label installation, then Quality test, before shipping them to customers. By the way, the back end test processes can be described as below.

a) **Printed circuit board assembly (PCBA) installation,** this station is to connect the hard disk drive with the circuit board for drives controlling.

b) **Filler station**, this step is to write data track onto magnetic recording disk.

c) **Initial drive test (IDT)**, this test checks mechanical/ electrical / firmware/ drive configuration and starts component optimization. This step determines whether the drive is working properly.

d) **Intelligent burn in (IBI)**, this step starts with firmware customizing for each head and media to perform a defect scans and prepare for the user mode.

e) **Final test station (FTS)**, this test formats the drive into customer mode and test parameters related to customer performance requirements. This is a drive check after all defects have been found and mapped out.

f) **Ambient (AMB)**, this test focuses on an error rate behavior at the ambient temperature.

g) Temperature screen (TSC), this test focuses on servo behavior at hot temperature.

h) Configuration (CFG), this is the final step to check for customer requirements. All limitations and configurations must pass total of specific customer requirements.

i) Label install, this station is to attach drives label onto hard disk drive**.**

j) Quality test station, this test is performed on a sample basis to assess / monitor the mass production product quality.

2.5 Test process parameters

During the test process, all parameters are generated and uploaded to the database. Head downgrade performance concerns on the following parameters.

2.5.1 Baseline popping (BLP) is one of the crucial problems caused by the head instability, whose effect can distort the readback signal to the extent that causing a sector read failure. BLP error count will be reported during the back end test process by head & zone. The signal amplitude comes from an instable BLP head can be fluctuated and exceeded the threshold. This indicates the presence of head instability and reflects its severity.

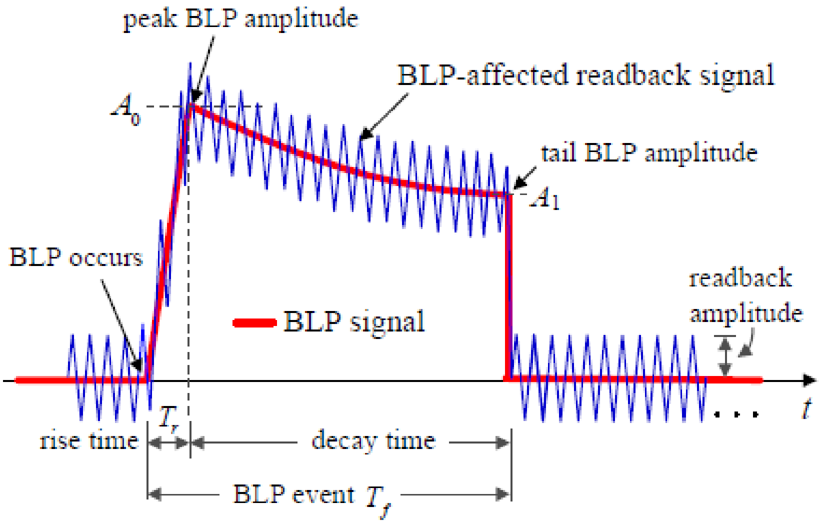


Figure 2.7 A signal of baseline popping (BLP). [4]

Usually, the BLP has a low-frequency signal with a short risen time and a moderate exponentially decay time with the peak amplitude that less than 1.5 times of the peak of normal readback signal. The BLP effect can cause an error burst in data detection, which could exceed the correction capability of the error correction codes. Thus, this results in a read failure. It is noted that the severity of the BLP effect depends on the BLP duration and its peak amplitude. The BLP signal in Figure 2.7 is modeled as.

$$u(t) = \begin{cases} A_0 t/T_r, & 0 \le t \le T_r \\ A_0 \exp(-(t-T_r)/T_d), & T_r < t < T_f \end{cases} \qquad (2.1)$$

where $u(t)$ = BLP signal, A_0 = is a peak BLP amplitude, T_f is a BLP duration. T_r is a rise time of 5T, and T_d is a decay constant depending on a tail BLP amplitude A_1.

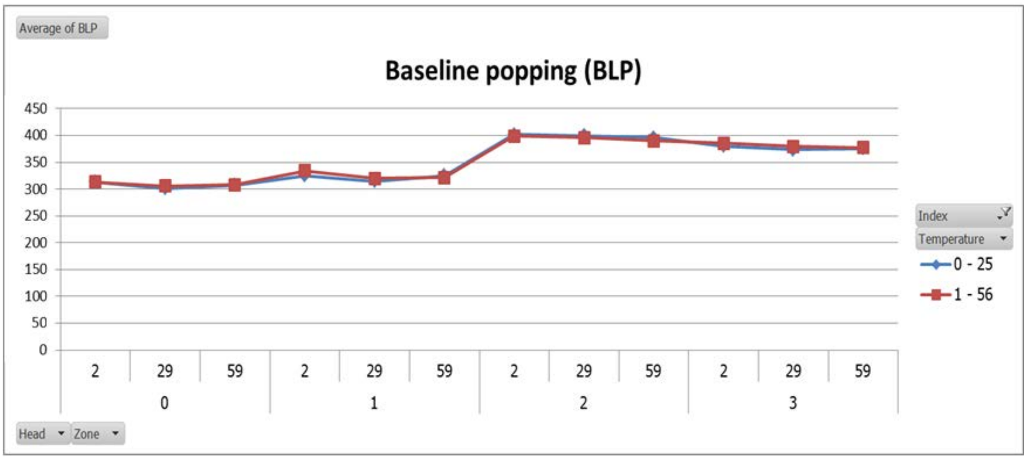


Figure 2.8 Example of BLP data analysis.

Figure 2.8 illustrates an example of BLP data analysis in hard disk drive. We get the actual data from the back end test data base. To perform failure analysis, we use pivot table to plot BLP data by head (4 headers) and by zone (60 zones) of hard disk drive with difference temperatures between 25° Celsius at ambient and 56° Celsius at hot. A higher BLP is counted more than a setting limit it represents for a severe head, which is a potential root cause of failures in the hard disk drive process.

2.5.2 Error margin (EM) is measured as the SNR distance from the cliff after completed the drive calibration and this parameter represents a direct measure of drive margin as opposed to BER (bit error rate) indirect measurement. EM aims to characterize zone performance by presenting the EM statistics of a continuous band of tracks in the zone. Figure 2.9 illustrates an example for EM measurement.

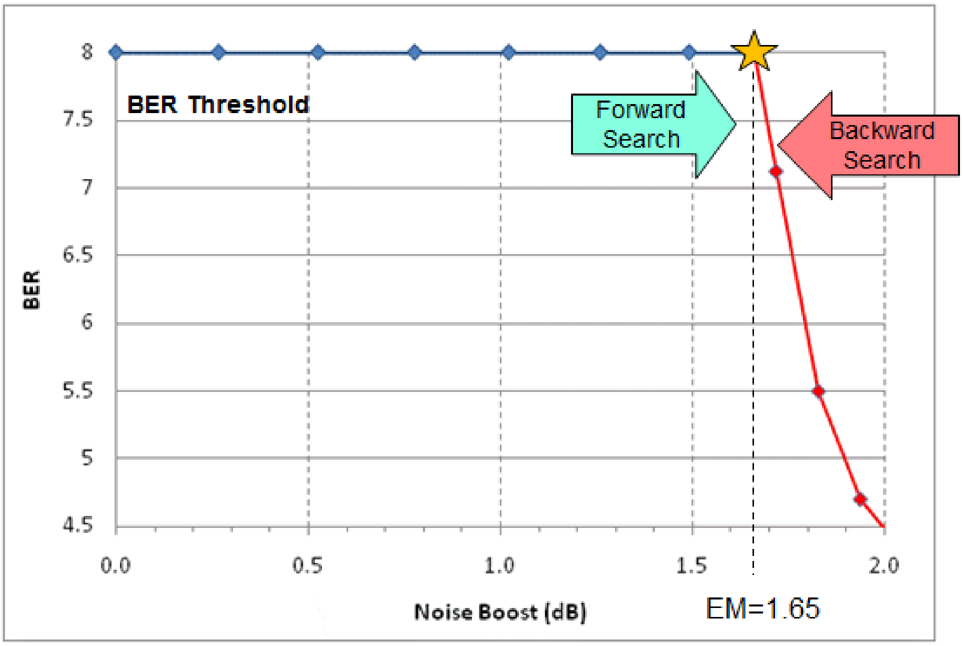


Figure 2.9 Example for EM measurement.

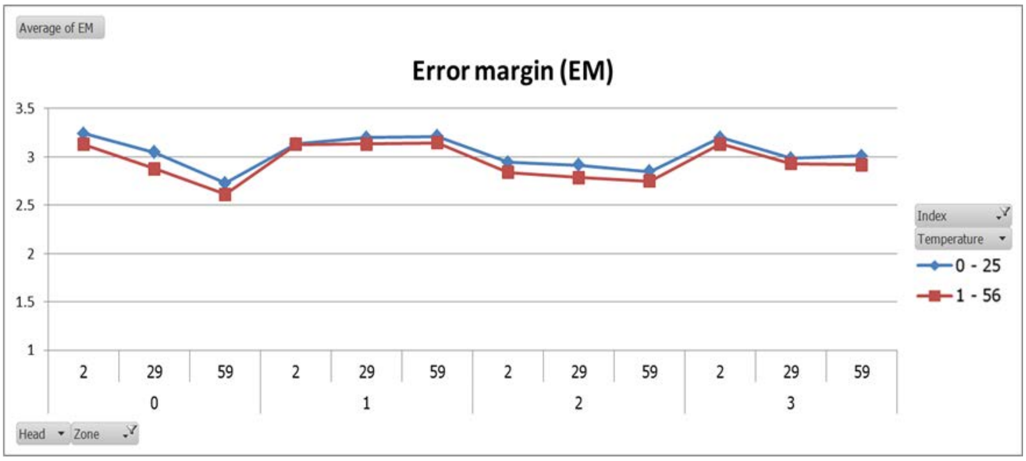


Figure 2.10 Example of EM data analysis.

Figure 2.10 illustrates an example of EM data analysis in hard disk drive. We get the actual data from the back end test database to perform failure analysis as well as BLP's practice. If EM values are lower than the setting limit and it will effect on the head performance downgrade, which is a potential root cause of failures in hard disk drive process.

2.5.3 Over write (OW) this test verifies the head and media performance during the final test station at the back end. The test writes data with the 1st frequency and 2nd frequency, and then measures the residual amplitude resulting from the 1st frequency in dB. In an idle test, we should not find any residual amplitude from the 1st frequency because the writer should overwrite the previous data for 100 percent. A bad overwrite may result from a problem during the optimization process (optimization between head and media during the initial test to find optimum points). The overwrite function is given by

$$OW_{dB} = 20\log_{10}\left(\frac{Vf_1}{Vf_2}\right) \tag{2.2}$$

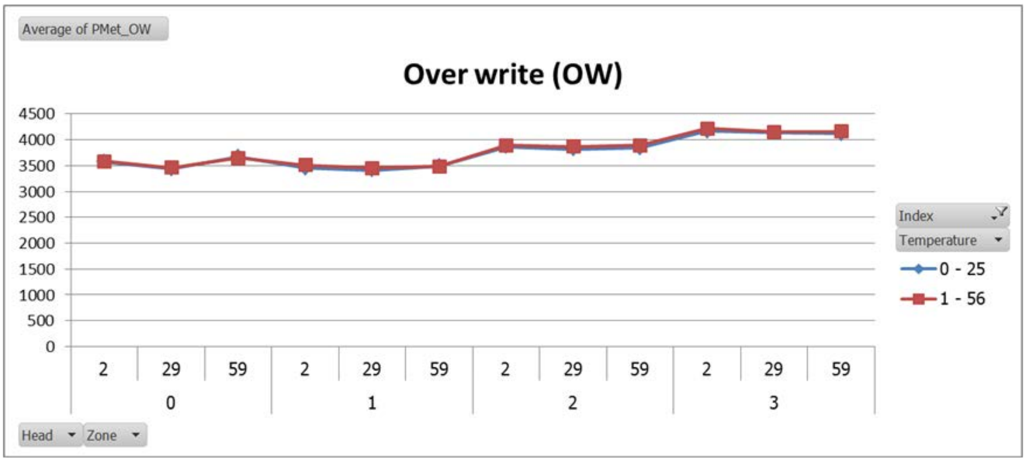


Figure 2.11 Example of OW data analysis.

Figure 2.11 illustrates an example of OW data analysis in hard disk drive. We use the failure analysis as well as BLP's practice. If OW values are lower than the setting limit, it will represent a severe head, which is a potential root cause of the failures in hard disk drive process.

2.5.4 Variable gain amplifier (VGA) is a signal-conditioning amplifier with an electronically settable voltage gain. There are the analog variable gain amplifiers and digital variable gain amplifiers. An analog voltage controls the gain and either a functional source, a digital to analog converter or a dc source in which can provide the control. In analog variable gain amplifiers, gain in dB is a linear function of the input voltage. It performs for read operation testing at the initial test. This test can screen changeable head. VGA test starts to let the VGA adapt into the signal for each drive and read the newly adapted VGA setting, compares to a setting that being captured earlier in the process.

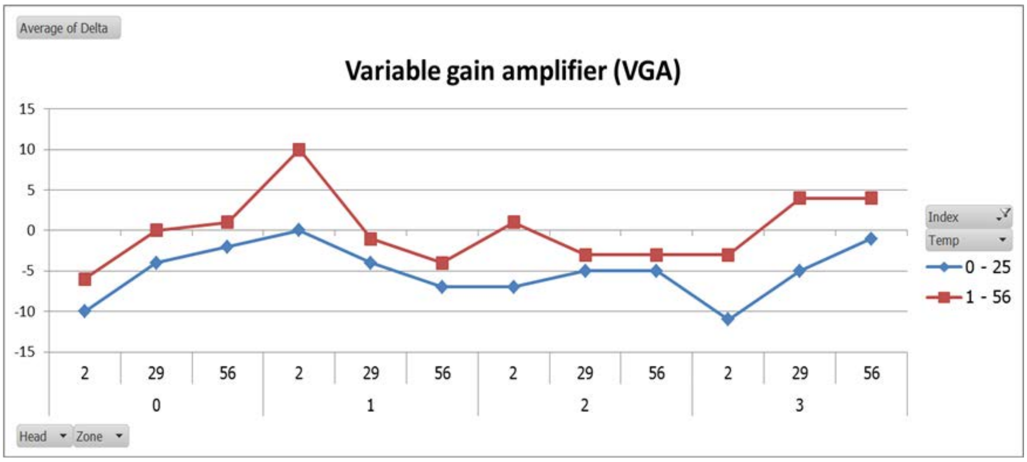


Figure 2.12 Example of VGA data analysis.

Figure 2.12 illustrates an example of VGA data analysis in hard disk drive. We use the failure analysis as well as BLP's practice. If VGA values are worse than the setting limit, it will represent a severe head, which is a potential root cause of failures in hard disk drive process.

2.5.5 Signal to noise ratio (SNR) is a measure used in science and engineering to compare between the level of a desired signal and the level of the background noise. SNR is an indicator that can be used to determine the system performance. There are many SNR definitions used in the literature, it depends on how and where they are used. While, SNR definition used in the magnetic recording systems analysis normally takes into account the additive noise and media noise. It is defined as the ratio of signal power to the noise power. The SNR function is given by

$$SNR = \frac{E_t}{N_0} \tag{2.3}$$

$$SNR_{dB} = 10\log_{10}\left(\frac{E_t}{N_0}\right) \tag{2.4}$$

where E_t = energy in an isolated transition pulse and N_0 is a single-sided power spectral density (psd) of the random noise.

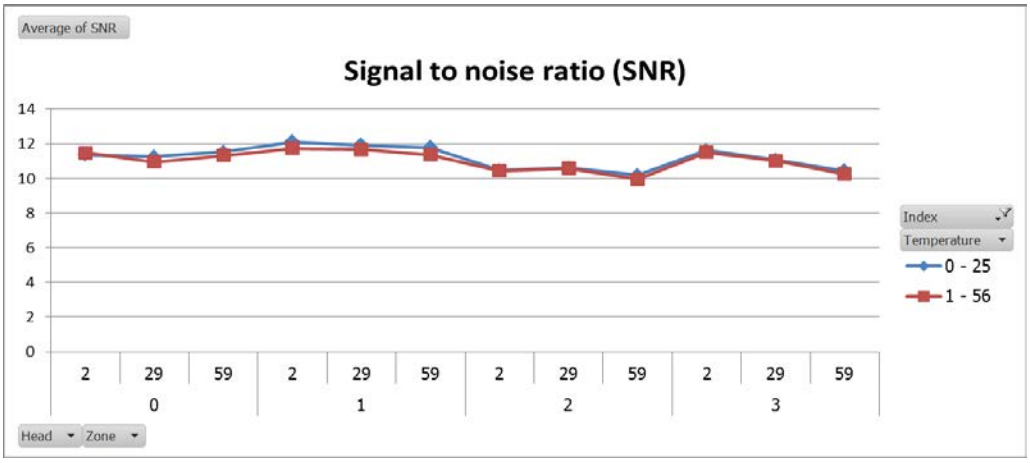


Figure 2.13 Example of SNR data analysis.

Figure 2.13 illustrates an example of SNR data analysis in hard disk drive. We use the failure analysis as well as BLP’s practice. If SNR values are lower than the setting limit, it will represent a severe head, which is a potential root cause of failures in hard disk drive process.

2.5.6 Magnetic erasure width (MEW) this measurement involves the direct magnetic erasure measurement. Particularly, the recording field bubble footprint width has a good correlation with the final TPI of the hard disk drive. The unit of magnetic erasure width (MEW) is measured as micro inch.

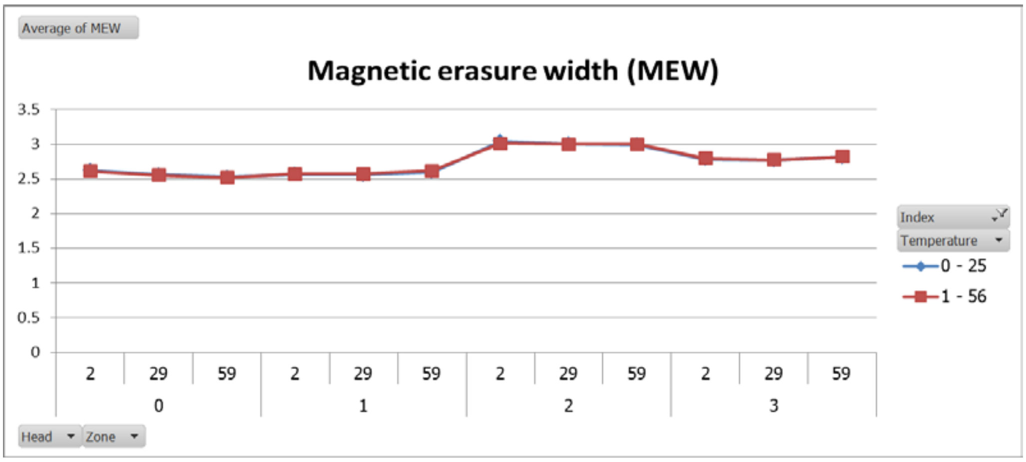


Figure 2.14 Example of MEW data analysis.

Figure 2.14 illustrates an example of MEW data analysis in hard disk drive. We use the failure analysis as well as BLP's practice. If MEW values are lower than the setting limit, it will represent a severe head, which is a potential root cause of failures in hard disk drive process.

2.6 Neural network (NN)

The term neural network (NN) [5] has been traditionally used to refer to a network or circuit of biological neurons. The modern usage of the term often refers to the artificial neural networks that composed of artificial neurons or nodes. Neural networks are the simple models on the way that the nervous system operates. The basic units are neurons, which are typically organized into layers. A neural network (or multilayer perceptron) is a simplified model for the way that human brain processes information. NN is widely applied in many areas such as character recognition, image compression and stock market prediction.

In character recognition, the ideas of character recognition become very important as handheld devices like the Palm Pilot becomes increasingly popular. Neural networks can be used to recognize handwritten characters as well.

In image compression, NN can receive and process vast amounts of information at once, making them useful in image compression. With the internet explosion and more sites using more images on their sites, using neural networks for image compression is a worth looking.

Stock market prediction, the day-to-day business of the stock market is extremely complicated. Many factors weigh in whether a given stock will go up or down on any given day. Since neural networks can quickly examine on a lot of information and sort it all out, they can be used to predict stock prices.

In this work, we use neural networks (NN) which is a function within a Rapid Miner Studio 6 to simulate the data from input parameter. Thus, we introduce only the used model. NN works by simulating a large number of interconnected simple processing units that resemble abstract versions of neurons. The processing units are arranged in layers. There are typically three parts in a neural network, an input layer with the units that represent for the input fields, one or more hidden layers, and an output layer with a unit or units represent for the output field. The units are connected with varying connection strengths (or weights). Input data are presented to the first layer, and values are propagated from each neuron to every neuron in the next layer. Eventually, a result is delivered from the output layer.

The network is learnt by individual records examining, generating a prediction for each record, and making adjustments to the weights whenever it makes an incorrect prediction.

This process is repeated many times, and the network continues to improve its predictions until one or more of the stopping criteria have been met.

Initially, all weights are random, and the answers that come out of the net are probably nonsensical. The network learns through training. Examples for which the output is known are repeatedly presented to the network, and the given answers are compared to the known outcomes. Information from this comparison is passed back through the network and gradually changed the weights. As training progresses, the network becomes increasingly accurate in replicating the known outcomes. Once trained, the network can be applied for the future cases where the outcome is unknown.

We can define the error information and sensitivity terms which calculated by multiplying the difference between these two terms with the derivative of the activation function.

$$e = t - a \tag{2.5}$$

where e is an error from the target, t is the target values, and a is the output

$$S^M = -2F^M(n^M)(t-a) \tag{2.6}$$

$$S^m = F^M(n^M)(\mathrm{W}^{m+1})^T S^{m+1} \tag{2.7}$$

where S^M is sensitivity of output layer, and S^m is sensitivity of the hidden layer

This error and sensitivity values are propagated to compute a weight (2.8) adjustment and biases (2.9) so that over time the neural network will be accurately predictable based on the input variables that it is presented.

$$W^m(k+1) = W^m(k) - \alpha S^m(a^{m-1})^T \tag{2.8}$$

$$b^m(k+1) = b^m(k) - \alpha S^m \tag{2.9}$$

where W is weight and b is bias

Figure 2.15 illustrates a general of neural network model. The input nodes are taken to be an information. The information is presented as activation values. Each node has given a number, the higher the number, the greater the activation. This information is then passed throughout the network. Based on the connection strengths (weights), inhibition or excitation, and transfer functions, the activation value is passed from node to node. Each of the nodes sums the activation values it received, then modifies the value based on its transfer function. The activation flows through the network and through hidden layers until it reaches the output nodes. The output nodes then reflect the input in a meaningful way to the outside world.

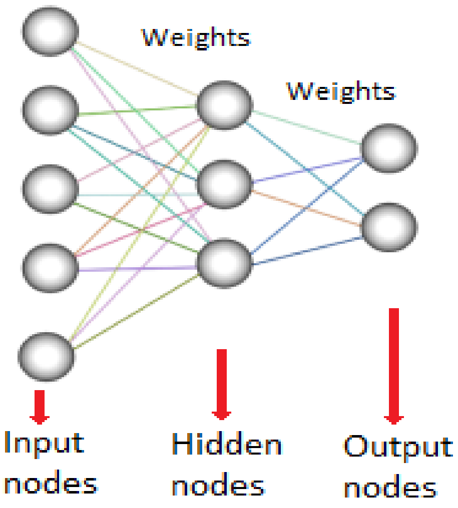


Figure 2.15 A general of neural network model. [6]

Generally, neural networks (NN) are computational models inspired by biological neural networks and used to approximate functions that are generally unknown. Particularly, they are inspired by the behavior of neurons and the electrical signals they convey between input (such as from the eyes or nerve endings in the hand), processing, and output from the brain (such as reacting to light, touch, or heat). Some NNs are the adaptive systems and used for example to model populations and environments, which constantly change. Neural networks can be hardware or software-based (computer models) and can apply with variety of topologies and learning algorithms.

2.6.1 Feed forward neural network. The feed forward neural network was the first and arguably most simple type of artificial neural network devised. In this network, the information moves in only one direction forwards by the input nodes, the data goes through the hidden nodes (if any) and to the output nodes. There is no cycle or loop in the network. Feed forward networks can be constructed from different types of units while the simplest example is the perceptron. Continuous neurons, frequently with sigmoidal activation are used in the context of back propagation of error.

2.6.2 Radial basis function (RBF) network. Radial basis functions are the powerful techniques for interpolation in multidimensional space. A RBF is a function which has built into a distance criterion with respect to a center. Radial basis functions have been applied in the area of neural networks where they may be used as a replacement for the sigmoidal hidden layer transfer characteristic in multi-layer perceptions. RBF networks have two layers of processing. Firstly, input is mapped onto each RBF in the “hidden” layer. The RBF chosen is usually a Gaussian. While secondly, in regression problems; the output layer is then a linear combination of hidden layer values representing mean predicted output. The interpretation of this output layer value is the same as a regression model in statistics. In classification problems the output layer is typically a sigmoid function of a linear combination of hidden layer values which represented for a posterior probability. Performance in both cases is often improved by shrinkage techniques, known as ridge regression in classical statistics.

2.6.3 Kohonen self-organizing network. The self-organizing map (SOM) invented by Teuvo Kohonen performs a form of unsupervised learning. A set of artificial neurons are studied to map points in an input space to coordinates in an output space. The input space can have different dimensions and topology from the output space while SOM attempts to preserve these.

2.6.4 Recurrent neural network. Recurrent neural networks (RNNs) are modeled with bi-directional data flow. While a feed forward network linearly propagates data from input to output. RNNs also propagate data from later processing stages to the earlier stages. RNNs can be used as general sequence processors.

2.6.5 Modular neural network. Biological studies have shown that the human brain function is not similar to a single massive network, but as a collection of small networks. This realization gave birth to the concept of modular neural networks, in which several small networks cooperate or compete to solve problems.

2.6.6 Physical neural network. A physical neural network includes with the electrically adjustable resistance material to simulate artificial synapses. Examples include the ADALINE neural network developed by Bernard Widrow in the 1960s and the memristor based neural network developed by Greg Snider of HP Labs in 2008.

2.6.7 Hopfield neural network. This is a single layer recurrent neural network. The Hopfield neural network is trained via an algorithm that teaches it to learn to recognize the patterns. The Hopfield network will indicate that the pattern is recognized by echoing it back. Hopfield neural networks are typically used for the pattern recognition.

2.6.8 Advantages and disadvantages of neural network

Advantages:

a) A neural network can be used to solve linear as well as non-linear programming tasks.

b) As a component of an NN fails, the net continues to operate.

c) A neural network learns and does not have to be re-programmed.

d) An NN can be used to solve classification, clustering and regression related problems.

Disadvantages:

a) Most NNs require a training phase to operate function.

b) As an NN architecture differs from microprocessors. NN has to be emulated

c) Large NN requires rather powerful hardware to run (to accomplish reasonable execution times).

2.7 Adaptive neuro-fuzzy inference systems (ANFIS)

Fuzzy inference systems [7] are efficient techniques for studying the behavior of nonlinear systems by using fuzzy logic rules. ANFIS is an adaptive neuro-fuzzy inference system that uses the learning techniques of neural networks, which is usually employed in many applications in control and prediction. Practically, ANFIS utilizes a hybrid learning algorithm to specify parameters. Specifically, it uses the least-square method with the back propagation gradient descent method to train the ANFIS membership function parameters based on a given training data set.

In general, the ANFIS structure is similar to a neural network structure. Figure 2.16 illustrates an example of the ANFIS model based on Takagi and Sugeno model [8]. Clearly, it consists of 5 layers connected through direction links, where the 1st is a fuzzy layer, the 2nd is a product layer, the 3rd is a normalized layer, the 4th is a de-fuzzy layer, the 5th is a total output layer, x1 and x2 are the inputs, and y is the output. Note that each layer is characterized by a node function with fixed adjustable parameters.

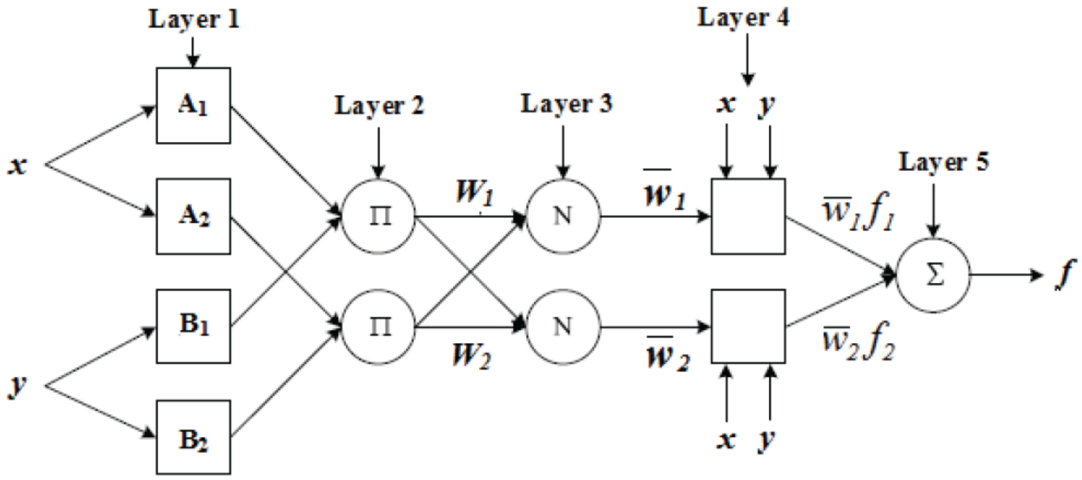


Figure 2.16 An example of adaptive neuro-fuzzy inference system (ANFIS).

According to the Takagi and Sugeno model [8], the rule sets and the function of each layer are as follows.

Rule Set:

If *x* is A_1 and *y* is B_1 then $f_1 = p_1 x + q_1 y + r_1$

If *x* is A_2 and *y* is B_2 then $f_2 = p_2 x + q_2 y + r_2$

where p_1, p_2, q_1, q_2, r_1 and r_2 are linear parameters and A_1, A_2, B_1 and B_2 are non-linear parameters.

Layer 1: It is an input fuzzy layer. Every node in this layer is an adaptive node that satisfies the following equations.

$$O_{1,i} = \mu_{A_i}(x) \ , \qquad i = 1, 2 \tag{2.10}$$

where $O_{1,i}$ denote the output functions, μ_{A_i} and μ_{B_i} denote of the membership functions.

A membership function for a fuzzy set *A* on the universe of discourse *x* is defined as $\mu_A(x) \longrightarrow$ [0, 1], where each element of *x* is mapped to a value between 0 and 1. This value quantifies the grade of membership of the element in *x* to the fuzzy set *A*. The membership function (MF) allows to graphically represent a fuzzy set, where the *x* axis represents the universe of discourse, and the *y* axis represents the degrees of membership in the [0,1] interval. For instance, if the triangular MF is employed, μ_{A_i} is given by

$$\mu_{A_i}(x) = \max\left[\min\left(\frac{x-a_i}{c_i-a_i}, \frac{c_i-x}{c_i-b_i}, 0\right)\right], \quad (2.11)$$

where a_i, b_i, and c_i are the MF parameters. On the other hand, if the generalized bell-shaped MF is used, μ_{A_i} will be given by

$$\mu_{A_i}(x) = \frac{1}{1+\left(\frac{x-c_i}{a_i}\right)^{2b_i}}, \quad (2.12)$$

where a_i, b_i, and c_i are the MF parameters

Layer 2: It is a product layer, where each node in this layer computes the impact of each rule through the multiplication by

$$O_{2,i} = w_i = \mu_{A_i}(x)\,\mu_{B_i}(y), \quad (2.13)$$

where $O_{2,i}$ is the output of layer 2.

Layer 3: It is a normalization layer, which computes the normalized effect of a given rule by

$$O_{3,i} = \overline{w}_i = \frac{w_i}{w_1 + w_2}, \quad (2.14)$$

where $i \in \{1, 2\}$, and $O_{3,i}$ is the output of layer 3, which can be called "normalized firing" strength are normalized with a maximum equal to 1 and a minimum equal 0.

Layer 4: It is a de-fuzzy layer, where the parameters in this layer are considered as consequent parameters that follow.

$$O_{4,i} = \overline{w}_i\, f_i = \overline{w}_i\, (p_i x + q_i y + r_i), \quad (2.15)$$

where $O_{4,i}$ is the output of layer 4 and p_i, q_i, r_i are called linear parameters or consequent parameters.

Layer 5: It is a total output layer to calculate the sum of the output for all incoming signals given by

$$O_{5,i} = \sum_i \overline{w}_i \ f_i = \frac{\sum_i w_i f_i}{\sum_i w_i} \ , \tag{2.16}$$

where $O_{5,i}$ is the total output layer and $\sum$ computes the sum of all incoming signals.

Generally, ANFIS has integrated the best features of fuzzy system and neural network. The best features of fuzzy system is representation of prior knowledge into a set of constraints (network topology) to reduce the optimization search space and the best feature of neural network is adaptation of back propagation to structured network to automate fuzzy control (FC) parametric tuning. Thus an ANFIS also has application to controllers system and modeling system.

In additional, fuzzy control has 3 types of it, there are type I which fuzzy control must be turned manually, type II which fuzzy control (Takagi-Sugeno type) have an automatic right hand side (RFS) tuning, and the type III which ANFIS system.

2.7.1 Data normalization. Normalization is the process of organizing the attributes and tables of a relational database to minimize data redundancy. In the simplest cases, the normalization is the rating means adjusting values measured on different scales to a notionally common scale, often prior to averaging. In more complicated cases, it refers to more sophisticated adjustments where the intention is to bring the entire probability distributions of adjusted values into alignment.

2.7.2 Advantages and disadvantage of ANFIS. There are many advantages of ANFIS, such as presents a better learning ability for a similar network complexity, achieve highly nonlinear mapping than other methods of similar complexity, requires fewer adjustable parameters than those required in other neural network and system is more accurate than the other models, anyways some of the disadvantages of ANFIS there are no variations apart from hybrid learning and that there can be only one output from ANFIS. Lastly, there is no provision for adjusting the membership functions.

2.8 Data mining

Data mining or knowledge discovery in database is a useful management system for the analysis of unknown information. It helps to pattern for the relationship or significant input from huge data which collected from database. Many techniques for data mining analysis depend on objective or result that user needs.

2.8.1 Procedure of data mining

a) Data selection before starting the experiment must indicate the objective and identify the database what user needs to do data mining.
b) Data cleaning is processed to make sure that raw data which user selected can perform the data analysis. Raw data which come from database have to screen for outlier, noisy data or missing value.
c) Data transformation after screening, data will be converted into a simply pattern for the easier algorithm analysis and data mining.
d) Data mining can be separated in two types A) Predictive data mining is a technique to estimate or expect the possible output by studying from history data, B) Descriptive data mining is a technique for data classification.
e) Result analysis and evaluation is a process to interpret and estimate the result by verifying the result whether it meets with the objective or not.

2.8.2 Classification of data has the objective to classify the empirical data into two classes namely “PASS” and “FAIL” according to the results obtained from customer testing with the device. Classification is a data mining technique used to predict group membership for data instances. The goal is to uncover the previous undetected relationships among data items, identify the sub-population that the new observations are belonging to the basis of a training set of data containing observations whose sub-population is known. Thus, the requirement is that new individual items are placed into groups based on quantitative information on one or more measurements, traits or characteristics and based on the training set in which the previously decided groupings are already established. Classification analysis works well with categorical data or a mixture of continuous numeric and categorical data. Figure 2.17 illustrates diagram of the general model procedure for pattern classification and prediction process.

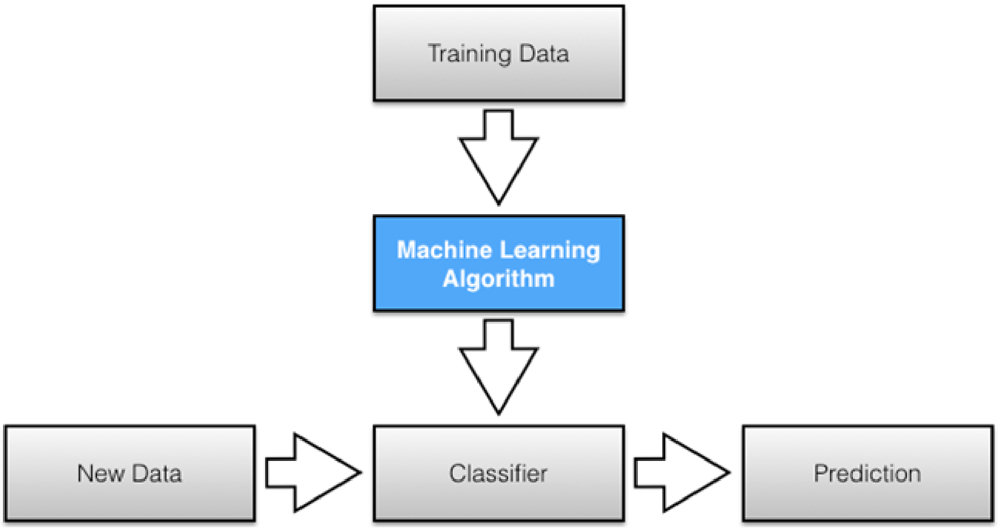


Figure 2.17 A diagram of the general model procedure for pattern classification.

2.8.3 Naive bayes (NB) [9] classifiers are linear classifiers that are known for being simple yet very efficient. The probabilistic model of naive bayes classifiers is based on Bayes’ theorem and the adjective naive comes from the assumption that the features in a dataset are mutually independent. In practice, the independence assumption is often violated, but naive bayes classifiers still tend to perform very well under this unrealistic assumption. Especially for small sample sizes, naive bayes classifiers can outperform the more powerful alternatives.

Being relatively robust, easily to implement, fast, and accurate, naive bayes classifiers are used in many different fields. Some examples include the diagnosis of diseases and make decisions about treatment processes, the classification of RNA sequences in taxonomic studies, and spam filtering in e-mail clients. However, strong violations of the independence assumptions and non-linear classification problems can lead to very poor performances of naive bayes classifiers. We have to keep in mind that the type of data and the problem to be solved dictate on the classification model we want to choose. In practice, it is always recommended to compare different classification models on the particular dataset and consider on the performances prediction as well as computational efficiency.

In the following sections, we will take a closer look at the probability model of the naive bayes classifier and apply the concept to a simple toy problem. Later, we will use a publicly available SMS (text message) collection to train a naive bayes classifier in Python that allows us to classify the unseen messages as spam or ham.

Algorithm: Bayes classifier is a simple probabilistic classifier based on bayes Theorem. Let X be the sample for unknown class, H be some hypothesis that X belongs to class C, then the probability is that H holds a given observed X, given by

$$P(H \mid X) = \frac{P(X \mid H) \cdot P(H)}{P(X)} \tag{2.17}$$

where *P*(H) is the prior probability of hypothesis H, *P*(X) is the prior probability of observing data X, *P*(X | H) (likelihood) is the probability of observing X if hypothesis H is held, and *P*(H | X)is the posterior probability of H after observing data X.

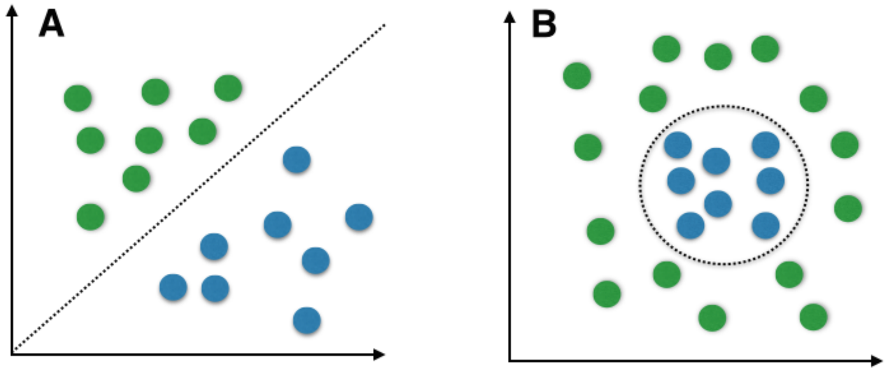


Figure 2.18 (a) Linear. (b) non-linear problems.

Figure 2.18 illustrates random samples for two different classes as shown in colored spheres, and the dotted lines indicate the class boundaries that classifiers try to approximate by decision boundaries computing. A non-linear problem (b) would be a case where linear classifiers, such as naive bayes, would not be suitable since the classes are not linearly separable.

2.8.4 Support vector machine (SVM) [10] is a set of related supervised learning methods that analyzes data and recognizes patterns, used in the classification and regression analysis. SVM has been successfully applied in many real-world problems such as text (and hypertext) categorization, hand-written character recognition, image classification and bioinformatics (Protein classification, Cancer classification). The original SVM algorithm was invented by Vladimir Vapnik and the current standard incarnation (soft margin) was proposed by Corinna Cortes and Vladimir Vapnik. The standard SVM takes a set of input data and predicts on each given input which of two possible classes; the input is a member of that makes the SVM a non-probabilistic binary linear classifier. Since a SVM is a classifier, then a given set of training examples on each mark belong to one of two categories, an SVM training algorithm builds a model that predicts whether a new example falls into one category or the other.

Intuitively, an SVM model is a representation of the examples as pointed in space so that the examples of the separate categories are divided by a gap that is as wide as possible. New examples are then mapped into that same space and predicted the belonging category based on which side of the gap they fall on.

More formally, a support vector machine constructs a hyperplane or set of hyperplanes in a high or infinite dimensional space which can be used for classification, regression or other tasks. Intuitively, a good separation is achieved by the hyperplane that has the largest distance from the nearest training data points of any class (so-called functional margin) because in common, the larger the margin the lower the generalization error of the classifier. There are many hyperplanes that can be used to classify the data. One reasonable choice for the best hyperplane is the one that represents the largest separation, or margin, between the two classes. So we choose the hyperplane so that the distance from the nearest data point on each side is maximized (strong generalization ability).

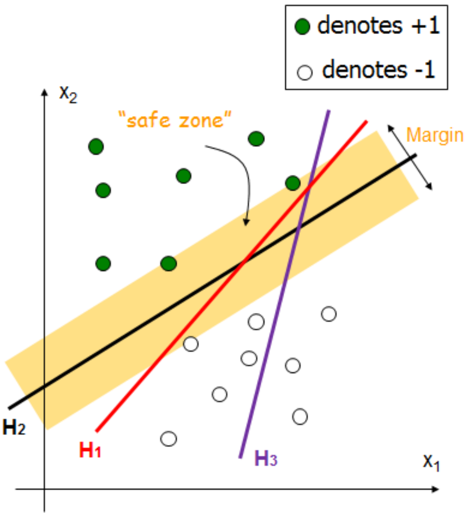


Figure 2.19 A linear SVM in 2D feature space.

In Figure 2.19, H3 cannot separate the two classes. H1 does, with a small margin while H2 with the maximum margin. The margin is defined as the width of boundary that could be increased before hitting a data point. We can find the best margin by the following.

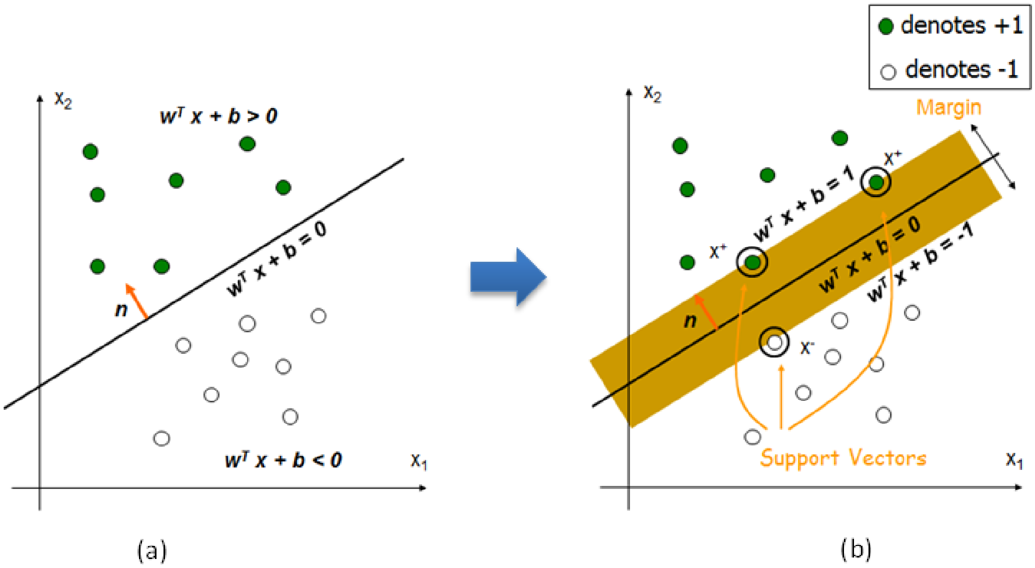


Figure 2.20 a) A simple of SVM. b) SVM with maximum margin.

From Figure 2.20 (a), *g(x)* is a linear function

$$g(x) = w^T x + b \tag{2.18}$$

where *w* is a weight vector, *x* is an input vector, *b* is bias.

(Unit-length) normal vector of the hyper-plane is

$$n = \frac{w}{\|w\|} \tag{2.19}$$

where *n* is a normal vector.

Given a set of data points $\{(x_i, y_i)\}, i = 1,2, \ldots, n$ where

$$For\ y_i = +1, \qquad w^T x_i + b > 0 \tag{2.20}$$

$$For\ y_i = -1, \qquad w^T x_i + b < 0 \tag{2.21}$$

where y_i is either 1 or -1.

With a scale transformation on both *w* and *b*, the above is equivalent to

$$For\ y_i = +1, \qquad w^T x_i + b, \tag{2.22}$$

$$For\ y_i = -1, \qquad w^T x_i + b \leq -1 \tag{2.23}$$

From Figure 2.20 (b), we know that

$$. \qquad w^T x^+ + b = 1, \qquad w^T x^- + b = -1 \tag{2.24}$$

We need to maximize the margin, that is,

$$\begin{aligned} M &= (x^+ - x^-) \cdot n \\ &= (x^+ - x^-) \cdot \frac{w}{\|w\|} = \frac{2}{\|w\|} \\ &Maximize\ \frac{2}{\|w\|} \end{aligned} \tag{2.25}$$

where *M* is a margin.

Also, we need to minimize the margin error, that is,

$$Minimize\ \frac{1}{2}\|w\|^2 \tag{2.26}$$

Subject to

$$y_i(w^T x_i + b) \geq 1 \qquad \forall i \tag{2.27}$$

2.8.5 Advantages and disadvantage of data mining. Data mining is useful for users to analyze from huge databases. It uses Artificial intelligence techniques (AI) that can identify the patterns or predict the possible output of a system by learning from its history. The result depends on the analysis techniques used to arid the user's objective. Data mining is a smart engine to help the users work efficiency. However, data mining can forecast only the possible outputs, thus the result still needs the verification from users and their decision before conclusion. So, the users who use data mining should fluently understand on models and algorithms.

Chapter 3

Literature review and methodology

This study presents a prediction method of head stack performance downgrade at HDD quality test. It is based on ANFIS functions in MATLAB R2008b and “RapidMiner Studio 6” by using neural networks (NN), naive bayes (NB) and support vector machines (SVM) where six input parameters (BLP, VGA, SNR, MEW, OW and EM) and one output parameter (Pass or Fail) are considered in our models. Note is made based on our observation from the historical data statistics as we found that these six input parameters in the HDD back end test process are the most relevance to the head performance at the quality test.

3.1 Literature reviews

Practically, ANFIS is a fuzzy inference system implemented in a framework of an adaptive network. By using a hybrid learning procedure, the ANFIS can construct an input-output mapping, based on human knowledge and stipulated input-output data pairs. In general, ANFIS is suitable for modeling a non-linear system and predicting a chaotic time series. For example, Roy [11] employed the ANFIS to predict the surface roughness in a turning operation for a set of given parameters with two different membership functions (MFs) and then compared the prediction accuracy. It was found that the bell-shaped MF has the prediction accuracy of 97.84%, whereas the triangular MF has the prediction accuracy of 96.13%. Altaher [12] studied a neural fuzzy classifier based on ANFIS for malware detection, which was found that this ANFIS classifier can detect the malware exe files effectively. Nazmy and Messiry [13] presented an intelligent diagnosis system using a hybrid approach of ANFIS for classification of the electrocardiogram signals, whose results indicated an accuracy level of more than 97%. Tepin [14] proposed a neural network rank level fusion applied on key parameters measured in the manufacturing process to predict customer failures resulted from head disk interaction (HDI). It was found that the result of rank level fusion classification model is able to achieve 86.61% accuracy for testing samples, and potentially to affect HDI failure. Wang et al. [15] introduced and develop a prediction model for chronic diseases to improve healthcare technology by using integrate logistic regression analysis and artificial neural networks (ANNs) model for the selection of risk factors and the prediction of chronic diseases by taking a case study of hypertension. The experimental results showed that the proposed approach achieved more than 72% prediction accuracy. Dang and Jiang [16] proposed a reliability prediction method based on degradation measure distribution and wavelet neural network to avoid the errors caused by pseudo life prediction in degradation testing. The result degradation data are utilized to verify the proposed method, reliability prediction of partial degradation data is implemented and the prediction result is acceptable. Adam et al. [17] introduced a hybrid approach, namely hybrid artificial neural network - naive bayes classifier for two-class imbalanced datasets classification in semiconductor manufacturing test process to identify good/bad products earlier and to avoid the bad units from being processed. The result of proposed hybrid approach performs better than the individual classifiers and finally overcomes the imbalanced dataset problems in semiconductor manufacturing test process. Aneja [18] proposed method to prediction of asthma in India by using the fusion approach of naive bayes and neural network, the proposed approach helps patients in their diagnosis of asthma. Then, Asawatongtip [19] introduced a method to predict the root causes of drive downgrade during the assembly and test processes by data mining using the bayesian network model, which a achieve an prediction accuracy of 80.7%.

Therefore, this paper proposes a new prediction method based on an ANFIS model by using the head stack input parameters to predict head performance downgrade at quality test and compare prediction performance with others method.

3.2 Test process study

First of all we study the process from cleanroom assembly until back end testing to understand each parameter in the test process. Then we find that the parameters concerned on the head downgrade failure and recognize each parameter meaning in depth detail, after that we collect the unseen data to test on 10000 data points in the manufacturing test process and cleaning data before starting analysis. Then, we find probabilistic model that characterizes relationship between each parameter and failures.

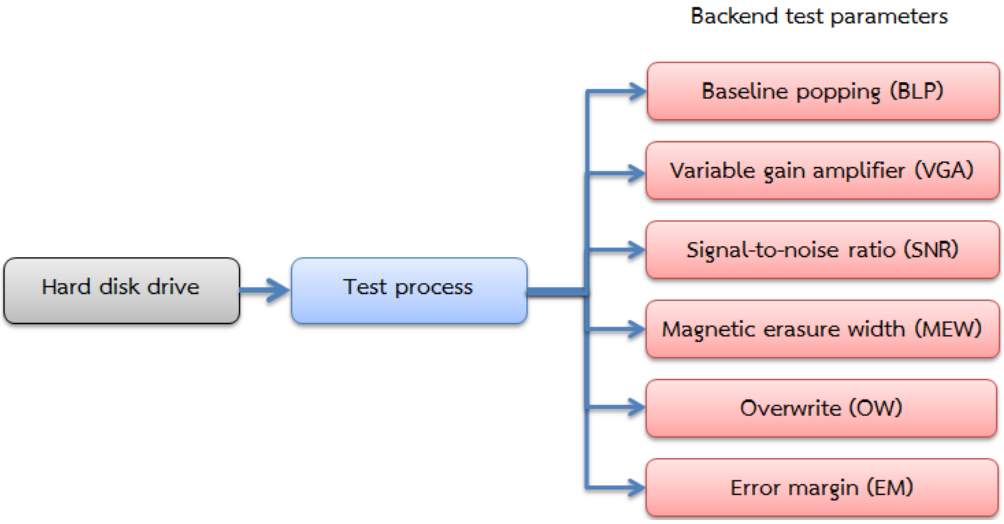


Figure 3.1 Parameters related to head performance downgrade.

Figure 3.1 illustrates six parameters during test in manufacturing process that we collected based on history of failures analysis observation, most of them related to head performance downgrade at quality test station. There are baseline popping (BLP), variable gain amplifier (VGA), signal-to-noise ratio (SNR), magnetic erasure width (MEW), overwrite (OW), and error margin (EM).

However, during the back end test process all of drives can pass the test limit, but after we sampled test at quality test, we still found fail in some drives that risk for customers return problem.

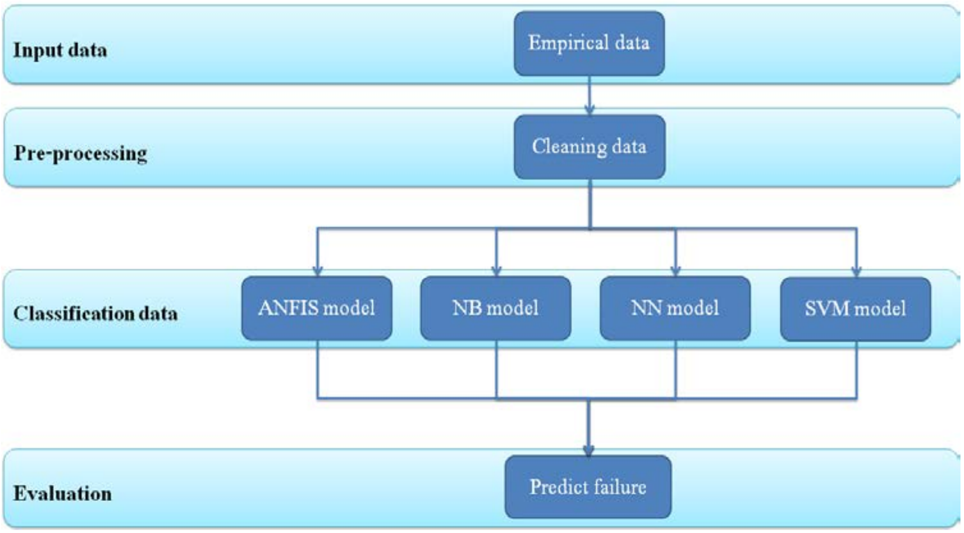


Figure 3.2 Schematic of the proposed algorithm derives from this study.

3.3 A schematic of the proposed algorithm

A schematic of the proposed algorithm derives from this study can be separated into 4 steps as illustrated in Figure 3.2.

3.3.1 Input data are collected from pass and fail samples (homogeneous) at customer integration station. The empirical parameters are grouped according to the main factors that affecting on head performance downgrade.

3.3.2 Pre-processing is to apply the Principal Component Analysis (PCA) to screen the importance of data as the number of input parameters is high (reduce the cross-correlation by selecting the main eigenvalues with optimal confidence level).

3.3.3 Classification of data with four classification methods applied to get the prediction models which are Adaptive neuro-fuzzy inference system (ANFIS), neural networks (NN), naive bayes (NB) and support vector machines (SVM).

3.3.4 Evaluation this process involves considering various models and choosing the best one based on their predictive performance.

3.4 Classification and prediction process

After the theory is reviewed, the four classification modules as described are chosen, namely, Adaptive neuro-fuzzy inference system (ANFIS), neural networks (NN), naive bayes (NB) and support vector machines (SVM).

3.4.1 Adaptive neuro-fuzzy inference system (ANFIS) is selected because of its better learning ability to generate the prediction for a complexity network. ANFIS uses the learning techniques of neural networks, which is usually employed in many applications for controlling and prediction. Practically, ANFIS utilizes a hybrid learning algorithm to specify parameters.

3.4.1.1 How to perform Adaptive neuro-fuzzy inference system (ANFIS) - First of all, we define number and types of input in which we found six input parameters (BLP, VGA, SNR, MEW, OW and EM) are correlation with the output

(Passed or Failed). Table 3.1 shows the sample of back end test with six input parameters and an output. Then we randomly divide the raw data into two groups; 80% for a training data set and 20% for testing data set. And then we normalized the input data by scaling between 0 and 1. In the training process of ANFIS, we separate it into two steps in which the first step is to find the membership function (MF) by using a “genfis2” function in MATLAB and determine the membership boundary by radius between 0.5 and 1.0. This is to determine the initial rules of model and set the condition for 100 loops training while the second step is to adjust the ANFIS model to obtain the best settings for prediction. Actually, we train and test data step by step through the varying of data from one parameter, two parameters, four parameters and six parameters, respectively, Then after the ANFIS model is obtained with proper parameters, we verify the model accuracy by using the (unseen) testing data set to verify the accuracy of ANFIS model prediction.

Table 3.1 Examples of back end test with six input parameters and one output.

BLP	VGA	SNR	MEW	OW	EM	Output
0.039	0.623	0.301	0.747	0.612	0.096	Pass
0.078	0.584	0.224	0.704	0.660	0.561	Pass
0.116	0.974	0.209	0.919	0.477	0.193	Pass
0.039	0.221	0.278	0.342	0.312	0.377	Fail
0.116	0.325	0.230	0.358	0.322	0.561	Fail
0.151	0.610	0.258	0.605	0.740	0.474	Pass
0.116	0.688	0.122	0.796	0.806	0.561	Pass
0.190	0.364	0.181	0.929	0.812	0.649	Pass
0.116	0.429	0.393	0.237	0.573	0.474	Fail
0.116	0.455	0.365	0.456	0.586	0.561	Fail
0.260	0.545	0.378	0.683	0.546	0.561	Fail
0.969	0.494	1.000	0.428	0.720	0.561	Pass
1.000	0.532	0.972	0.672	0.777	0.746	Pass
0.969	0.675	0.855	0.580	0.718	0.833	Pass
0.570	0.506	0.913	0.001	0.009	0.474	Fail
0.667	0.623	0.949	0.000	0.077	0.649	Fail
0.698	0.636	0.865	0.067	0.000	0.289	Fail
0.260	0.532	0.375	0.926	1.000	0.649	Pass

3.4.2 Neural networks (NN) is selected because of its ability to learn by examining the individual records and large data to generate a prediction. The operator learns on a model by means of a feed-forward neural network training. A feed-forward neural network is an artificial neural network where connections between the units do not form a directed cycle. In this network, the information moves in only one direction and forward from the input nodes through the hidden nodes to the output nodes. There is no cycle or loop in the network. We perform this method to compare the performance prediction with ANFIS model and others method.

3.4.2.1 How to perform neural networks (NN) - In RapidMiner Studio 6 SW, we perform NN classification by using the same inputs and output as ANFIS method; 80% for a training data set and 20% for testing data set, then cleaning and sampling data between passed and failed for testing. After that, we select operator “neural networks (NN)” for training data and then use the operator “apply model” for testing NN model. Finally using the operator “performance (classification)” to analyze and measure the accuracy of NN model prediction.

3.4.3 Naive bayes (NB) classifier is selected because the network can be used to compute the probabilities of the presence from various failures. This is a simple probabilistic classifier based on Bayes theorem applying with strong independence assumptions. We perform this method to compare performance prediction with ANFIS model and others data mining model method.

3.4.3.1 How to perform naive bayes (NB) - We start to perform NB classification by using data mining software namely “RapidMiner Studio 6”. In first step, we collect data by using the same inputs and output as in ANFIS method; 80% for a training data set and 20% for testing data set, then cleaning and sampling data between passed and failed for classification. After that, we select the operator “naive bayes (NB)” in RapidMiner Studio 6 SW for training data in which has been prepared and then used the operator “apply model” to test the data based on NB model. Finally, using the operator “performance (classification)” to analyze and measure the accuracy of NB model prediction.

3.4.4 Support vector machine (SVM) is selected as one of the classifier in our paper because it is supervised on the learning methods that analyzed the data in classification for each given input. This operator is a support vector machine (SVM) learner that support the internal multiclass learning and probability estimation based on Platt scaling for the proper confidence values after applying the learned model on a classification data set. The standard SVM takes a set of input data and predicts each given input in which the two possible classes comprises of the input, making the SVM a non-probabilistic binary linear classifier. Given a set of training examples, each marked belong to one of two categories. An SVM training algorithm builds a model that assigns new examples into one category or the other. An SVM model is a representation of the examples as pointed in space and mapped, so that the examples of the separate categories are divided by a clear gap that is as wide as possible. New examples are then mapped into that same space and predicted belong to a category based on which side of the gap they fall on.

3.4.4.1 How to perform support vector machine (SVM) - In RapidMiner Studio 6 SW, we perform SVM classification by using the same inputs and output data set as ANFIS method with 80% for a training data set and 20% for testing data set, then cleaning and sampling data between passed and failed for testing. After that we select the operator “support vector machine (SVM)” for training data, then use the operator “apply model” to test on SVM model. Finally, using the operator “performance (classification)” to analyze and measure the accuracy of SVM model prediction.

3.5 Prediction Accuracy

Accuracy [20] is used as a statistical measure of how well a binary classification test correctly identifies or excludes a condition. That is, the accuracy is the proportion of true results (both true positives and true negatives) among the total number of cases examined. The prediction accuracy evaluated by.

$$\text{Accuracy} = \frac{TP+TN}{TP+FP+FN+TN}\ , \tag{3.1}$$

where TP is True Positive (prediction of 1 when the sample test result has a 1), TN is True Negative (prediction of 0 when the sample test result has a 0), FP is False Positive (prediction of 1 when the sample test result has a 0), and FN is False Negative (prediction of 0 when the sample test result has a 1).

On the other hand, in pattern recognition and information retrieval with binary classification, precision (also called positive predictive value) is the fraction of retrieved instances that are relevant, while recall (also known as sensitivity) is the fraction of relevant instances that are retrieved. Both precision and recall are therefore based on an understanding and measure of relevance, the precision and recall given by.

$$\text{Precision} = \frac{TP}{TP+FP} \quad , \tag{3.2}$$

$$\text{Recall} \quad = \frac{TP}{TP+FN} \quad , \tag{3.3}$$

An accuracy of 100% means that the measured values are exactly the same as the given values. The accuracy for predictive analytics states that predictive models with a given level of accuracy may have greater predictive power than models with higher accuracy. It may be better to avoid the accuracy metric in favor of other metrics such as precision and recall. In situations which the minority class is more important, F-measure may be more appropriate, especially in situations with much skewed class imbalance.

Chapter 4

Experiments and results

In this section, we will verify the accuracy result and compare performance of the proposed ANFIS model versus data mining model neural networks (NN), naive bayes (NB) and support vector machines (SVM) that are used to predict the head performance downgrade, we will discuss about data distribution of six input parameters and outcome of each models and then the performance compare will discuss in the last section respectively, as depicted in Figure 4.1, all of them are the data distribution of each parameters that use for testing model.

Practically, this experiment uses western digital (WD) hard disk drive of 2.5 inch product with 500 gigabytes to study, we collected more than 10000 data points from back end test log in WD’s database, then verify and collect important parameters which are related to head performance downgrade at the quality test process, after that we use those parameters to perform prediction failures.

4.1 Data distribution of six input parameters

Figure 4.1 illustrates the data distribution of six input parameters, which is obtained by separating each input into 1 dimension and 2 dimensions for both passed (blue color) and failed (red color) drives. It is apparent that this distribution data is complicated to classify the data, thus we need the classification models with good performance to handle this kind of data for prediction.

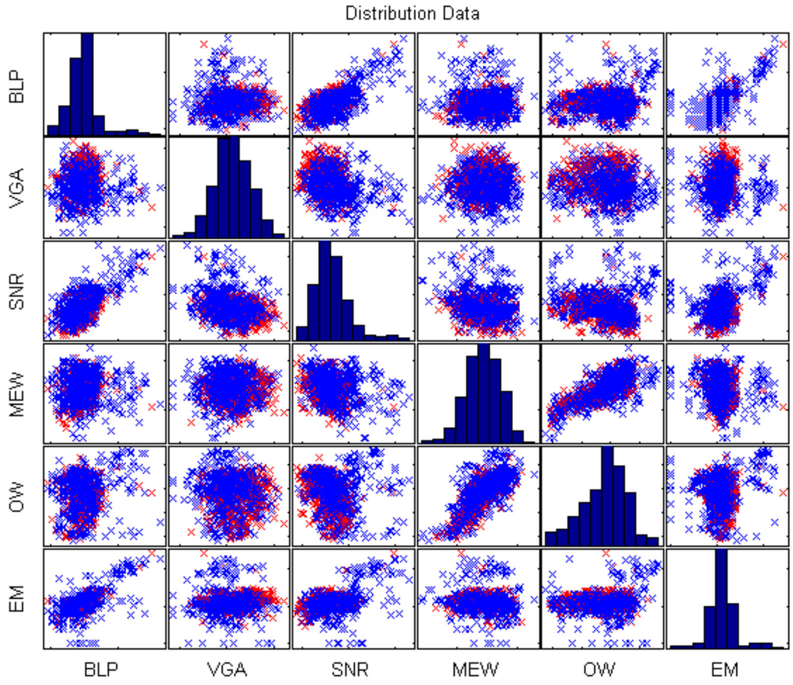


Figure 4.1 The scatter plot matrix distribution data of six input parameters.

4.2 Adaptive neuro-fuzzy inference system (ANFIS) result

An adaptive neuro-fuzzy inference system (ANFIS) is one of classification techniques that are useful to the prediction of head performance downgrade at quality test. Practically, we experiment prediction head performance downgrade failure by using six input parameters and one output, and then we vary input data from one, two, four and six input parameters to verify and compare prediction results of each group. We separate 80% of data for training set from 20% of data for testing set. The prediction result for each group is shown in Table 4.1 - 4.4 respectively.

Table 4.1 Experiment results of one input parameter comparison.

Parameter	Status	Radius	Precision	Recall	Accuracy
BLP	Train	1.0	64.7%	70.5%	72.6%
VGA	Train	1.0	59.0%	64.7%	61.9%
SNR	Train	1.0	52.5%	71.9%	66.2%
MEW	Train	1.0	64.0%	66.9%	60.4%
OW	Train	1.0	68.9%	64.6%	69.7%
EM	Train	1.0	79.1%	69.8%	74.5%
BLP	Test	1.0	60.0%	74.3%	71.1%
VGA	Test	1.0	58.0%	65.7%	62.9%
SNR	Test	1.0	52.6%	71.2%	67.4%
MEW	Test	1.0	62.9%	65.3%	60.6%
OW	Test	1.0	69.5%	62.9%	68.4%
EM	Test	1.0	82.9%	68.6%	75.7%

Table 4.2 Experiment results of two input parameters (BLP and EM).

Parameter	Status	Radius	Precision	Recall	Accuracy
BLP, EM	Train	1	88.5%	63.3%	75.9%
BLP, EM	Train	0.98	88.5%	63.3%	75.9%
BLP, EM	Train	0.96	88.5%	63.3%	75.9%
BLP, EM	Train	0.94	88.5%	63.3%	75.9%
BLP, EM	Train	0.92	88.5%	63.3%	75.9%
BLP, EM	Train	0.9	89.2%	59.7%	74.5%
BLP, EM	Test	1	82.9%	65.7%	74.3%
BLP, EM	Test	0.98	82.9%	65.7%	74.3%
BLP, EM	Test	0.96	82.9%	65.7%	74.3%
BLP, EM	Test	0.94	82.9%	68.6%	75.7%
BLP, EM	Test	0.92	82.9%	68.6%	75.7%
BLP, EM	Test	0.9	82.9%	68.6%	75.7%

Table 4.3 Experiment results of four input parameters (BLP, VGA, SNR and EM).

Parameter	Status	Radius	Precision	Recall	Accuracy
BLP, VGA, SNR, EM	Train	1	79.5%	70.6%	78.0%
BLP, VGA, SNR, EM	Train	0.98	80.4%	70.7%	78.2%
BLP, VGA, SNR, EM	Train	0.96	80.4%	70.7%	78.2%
BLP, VGA, SNR, EM	Train	0.94	79.7%	69.4%	78.6%
BLP, VGA, SNR, EM	Train	0.92	79.7%	69.4%	78.6%
BLP, VGA, SNR, EM	Train	0.9	79.7%	69.4%	78.6%
BLP, VGA, SNR, EM	Test	1	80.9%	69.3%	77.7%
BLP, VGA, SNR, EM	Test	0.98	80.9%	69.3%	77.7%
BLP, VGA, SNR, EM	Test	0.96	80.9%	69.3%	77.7%
BLP, VGA, SNR, EM	Test	0.94	80.9%	69.3%	77.7%
BLP, VGA, SNR, EM	Test	0.92	80.9%	69.3%	77.7%
BLP, VGA, SNR, EM	Test	0.9	80.9%	69.3%	77.7%

Table 4.4 Experiment results of six input parameters (BLP, VGA, SNR, MEW, OW and EM).

Parameter	Status	Radius	Precision	Recall	Accuracy
BLP, VGA, SNR, MEW, OW, EM	Train	1	84.9%	75.9%	80.3%
BLP, VGA, SNR, MEW, OW, EM	Train	0.98	85.0%	76.0%	80.1%
BLP, VGA, SNR, MEW, OW, EM	Train	0.96	84.9%	75.9%	80.3%
BLP, VGA, SNR, MEW, OW, EM	Train	0.94	84.9%	75.9%	80.3%
BLP, VGA, SNR, MEW, OW, EM	Train	0.92	85.0%	76.0%	80.1%
BLP, VGA, SNR, MEW, OW, EM	Train	0.9	85.0%	76.0%	80.1%
BLP, VGA, SNR, MEW, OW, EM	Test	1	85.7%	77.1%	80.0%
BLP, VGA, SNR, MEW, OW, EM	Test	0.98	85.7%	77.1%	80.0%
BLP, VGA, SNR, MEW, OW, EM	Test	0.96	85.7%	77.1%	80.0%
BLP, VGA, SNR, MEW, OW, EM	Test	0.94	85.7%	77.1%	80.0%
BLP, VGA, SNR, MEW, OW, EM	Test	0.92	86.0%	77.4%	80.3%
BLP, VGA, SNR, MEW, OW, EM	Test	0.9	86.0%	77.4%	80.3%

Based on results from Table 4.1 – 4.4, we found that the simulation results of each parameter are correlated to output (Passed & Failed) even when we use one, two, four or six input parameters. From Table 4.1, we found the main parameter that is more related to head performance is EM parameter. Then we increase input parameters from one, two, four and six respectively, the prediction accuracy will increase also. In this study, we select model with six input parameters to perform prediction because it provides better result. Figure 4.2 illustrates the prediction result of ANFIS based on simulation result of six input parameters, we found that ANFIS model provides the prediction accuracy of about 80% with compiling time of 40 seconds.

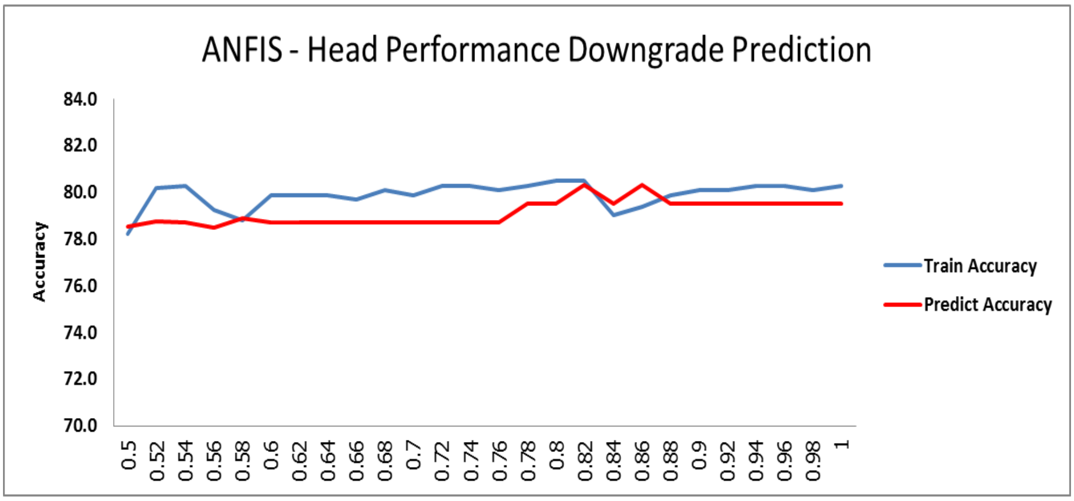


Figure 4.2 Accuracy of ANFIS prediction.

4.3 Neural network (NN) result

Neural network (NN) is one of the classification methods in data mining program Rapidminer Studio 6. We select this method since it is able to handle complexity data to prediction head performance downgrade at the quality test process. Figure 4.3 illustrates the neural network (NN) setting in Rapidminer Studio 6 program used for simulation data and then compare prediction performace with ANFIS model also, this experiment we use the same input parameters as ANFIS model with six input parameters and one output to verify prediction accuracy for each group. Then setting the NN with training loop of 200 cycles, learning rate as 0.3, and momentum as 0.2. After that we vary hidden layer from 5 layers until 17 layers.

In Table 4.5 illustrates prediction performance of neural network (NN) model by each group of hidden layers (5, 7, 9, 11, 13, 15 and 17 nodes). Based on simulation result we found that the setting hidden layers = 15 nodes (green highlight) provide better result than others. So, we select this model to represent performance of neural network (NN) model. By the way, Figure 4.4 illustrates NN simulation model of six input parameters and one output, we get 1 hidden layer with 15 nodes, all nodes are sigmoid function. Table 4.6 and 4.7 illustrate function of hidden layer there are input parameters and ouput of selected model with six input parameters in neural network (NN) model. In the experiment result, we found that neural network (NN) model provides the prediction accuracy of about 70.5% with compiling time of 31 seconds.

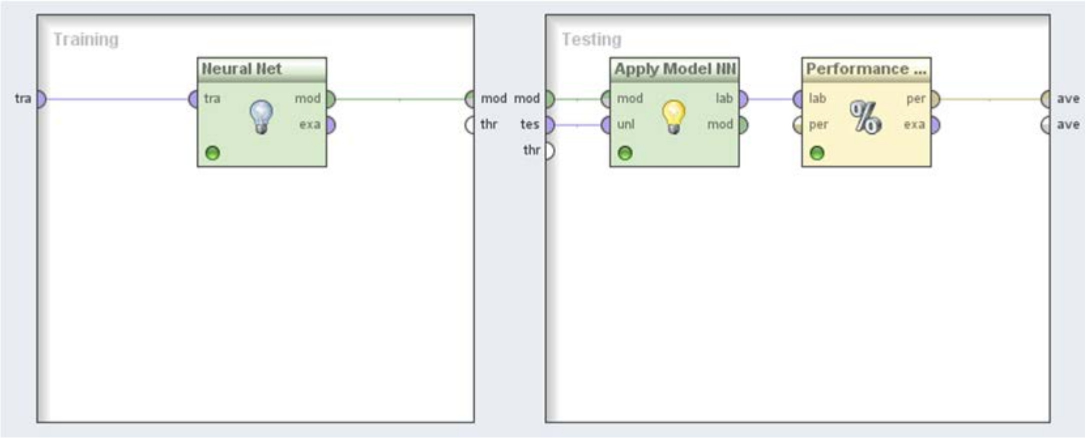


Figure 4.3 Neural network (NN) operators setting.

Table 4.5 performance of neural network (NN) model.

Hidden layer	Precision	Recall	Accuracy
5 nodes	71.6%	63.2%	68.9%
7 nodes	72.7%	63.4%	69.8%
9 nodes	71.7%	61.7%	69.2%
11 nodes	71.3%	59.5%	68.3%
13 nodes	69.9%	63.6%	68.2%
15 nodes	72.9%	65.1%	70.5%
17 nodes	70.1%	61.0%	67.1%

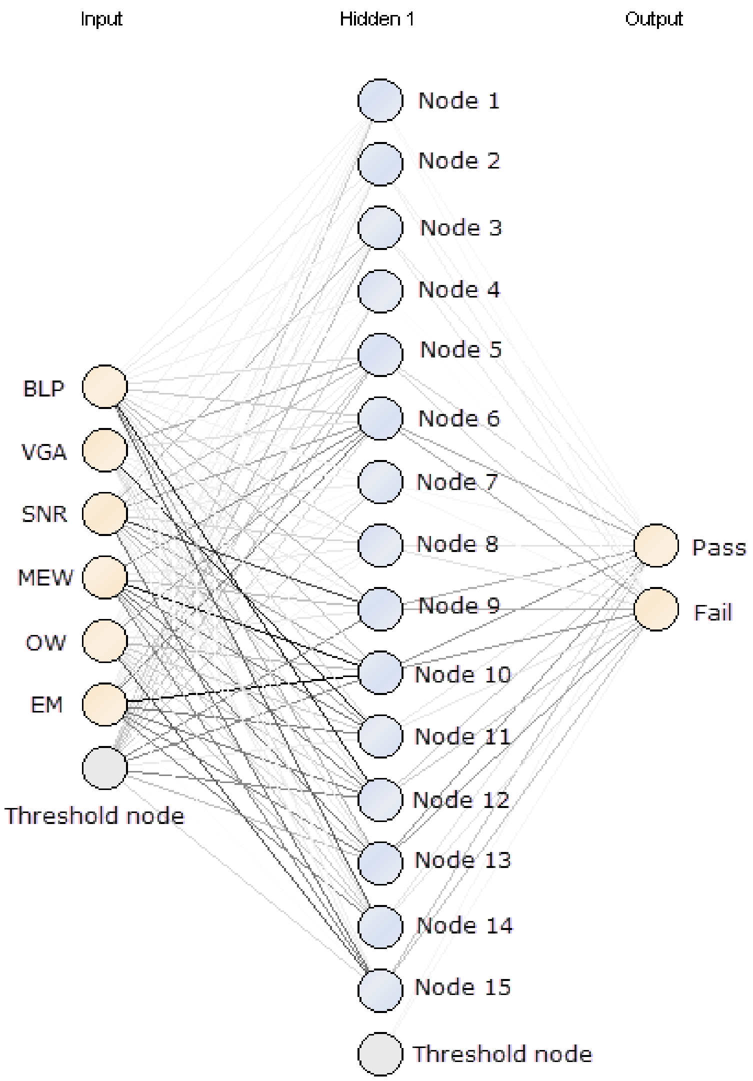


Figure 4.4 Neural network (NN) model.

Table 4.6 Hidden layer of neural network (NN) model.

Parameter	Nodes														
	N1	N2	N3	N4	N5	N6	N7	N8	N9	N10	N11	N12	N13	N14	N15
BLP	0.7	1.4	0.9	1.0	-2.3	3.5	0.8	3.1	4.8	3.9	2.3	-15.1	5.2	10.8	1.0
VGA	0.7	0.1	0.5	0.3	5.7	2.0	0.5	-1.1	-1.4	-0.6	15.0	-1.7	2.1	2.4	-1.1
SNR	-1.0	1.2	-4.9	0.1	4.3	6.0	-1.3	2.9	-12.4	-5.7	6.0	3.5	1.2	-1.2	9.4
MEW	-0.3	-0.2	-0.9	-0.4	-1.0	7.7	-0.4	1.4	-4.3	-16.0	-9.0	9.3	-9.4	-3.5	-8.2
OW	0.7	-0.5	0.8	0.1	1.6	-6.6	0.6	-0.5	1.7	4.4	-3.3	-3.5	2.4	4.0	11.6
EM	1.2	0.6	1.1	0.9	4.0	-9.1	1.2	0.9	-0.2	-17.4	8.6	8.7	8.3	-8.4	-0.2
Bias:	-3.4	-2.1	-4.5	-3.0	-4.1	-2.7	-3.4	-0.7	-7.9	-7.9	1.4	-8.2	-4.4	0.1	-2.8

Table 4.7 Output of neural network (NN) model.

Parameter	Class 'Pass' (Sigmoid)	Class 'Fail' (Sigmoid)
Node 1	0.5	-0.4
Node 2	-0.9	0.9
Node 3	1.5	-1.4
Node 4	-0.2	0.2
Node 5	2.8	-2.8
Node 6	-5.5	5.5
Node 7	0.5	-0.5
Node 8	-2.0	2.0
Node 9	5.5	-5.6
Node 10	-6.9	6.9
Node 11	1.5	-1.5
Node 12	-3.5	3.5
Node 13	-6.8	6.8
Node 14	-1.7	1.7
Node 15	-4.1	4.1
Threshold	0.6	-0.6

4.4 Naive bayes (NB) result

Naive bayes (NB) is one of the classification methods in data mining program namely “Rapidminer Studio 6 program”. A baive bayes model is easy to build without complex iterative parameter estimation which makes it particularly useful for very large datasets. Figure 4.5 illustrates the naive bayes (NB) operators setting in Rapidminer Studio 6 program, we separate training set and testing set into 2 sections. In this experiment, we use the same input parameters as ANFIS model by varying input data from one, two, four and six input parameters and one output to verify prediction accuracy for each group. Table 4.8 illustrates the mean and standard deviation values of each input parameter in naive bayes model. Table 4.9 illustrates the prediction performance of naive bayes (NB) model by each group of input parameters (one, two, four and six input parameters). Based on simulation result, we found that six input parameters provide better results than others. We select six input parameters and one output to represent performance of naive bayes (NB) model. So, we get prediction accuracy of about 68.5% with compiling time of 8 seconds.

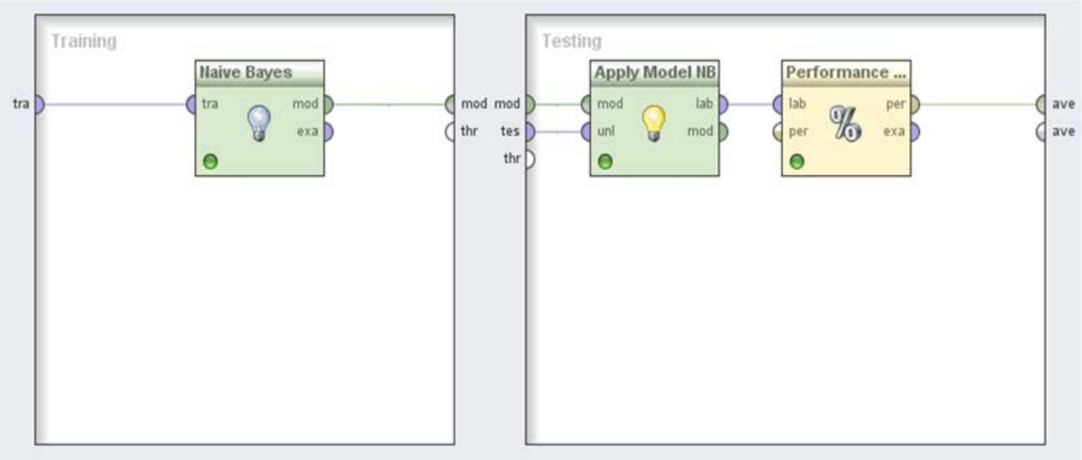


Figure 4.5 Naive bayes (NB) operator's setting.

Table 4.8 Mean and standard deviation data of naive bayes (NB) model.

Parameter	Measure	Pass	Fail
BLP	mean	0.3	0.3
BLP	standard deviation	0.1	0.1
VGA	mean	0.5	0.4
VGA	standard deviation	0.1	0.1
SNR	mean	0.3	0.4
SNR	standard deviation	0.1	0.1
MEW	mean	0.5	0.6
MEW	standard deviation	0.1	0.2
OW	mean	0.5	0.6
OW	standard deviation	0.2	0.2
EM	mean	0.5	0.4
EM	standard deviation	0.1	0.1

Table 4.9 Prediction performance of naive bayes (NB) model.

Parameter	Precision	Recall	Accuracy
EM	62.2%	60.5%	63.0%
BLP, EM,	69.8%	54.6%	65.5%
BLP, VGA, MEW, EM	70.5%	56.1%	66.4%
BLP, VGA, SNR, MEW, OW, EM	79.3%	51.1%	68.5%

4.5 Support vector machine (SVM) result

Support vector machine (SVM) is one of the classification methods in data mining program Rapidminer Studio 6. This method is able to handle linear and non-linear. SVM performs classification by finding the hyperplane that maximizes the margin between the two classes. The vectors that define the hyperplane are the support vectors. Figure 4.6 illustrates the support vector machine (SVM) operators setting in Rapidminer Studio 6 program used for simulation data to compare performace with ANFIS model also, In this experiment we use the same input parameters as ANFIS model with six input parameters and one output, then we vary setting from 0.1 - 1.0 gamma to verify prediction accuracy for each point. Table 4.10 illustrates weight table of support vector machine for each gamma setting.

From Table 4.11, we found that using gamma = 0.7 (green highlight) provides better result than others. So, we select this model to represent performance of support vector machine (SVM) model. Based on experiment result, we found that support vector machine (SVM) model provides the prediction accuracy of about 68.3% with compiling time of 15 seconds.

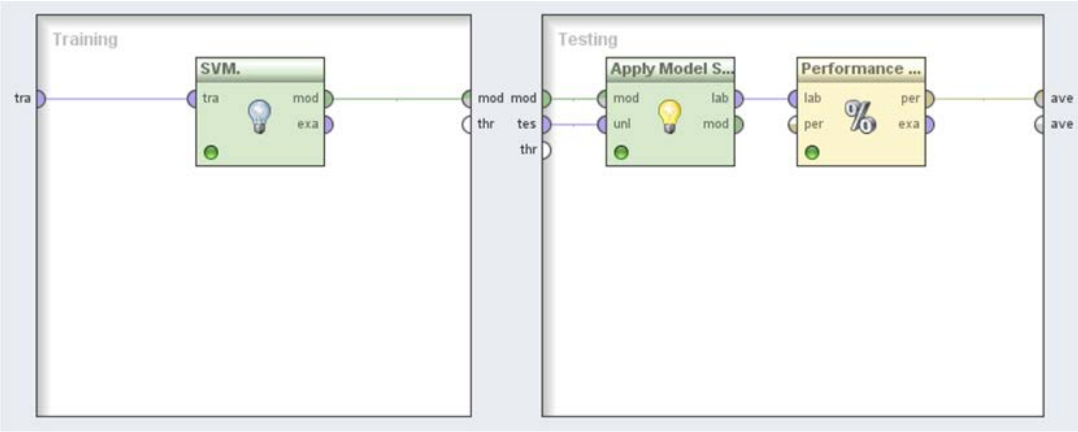


Figure 4.6 Support vector machine (SVM) operator’s setting.

Table 4.10 Weight table support vector machine (SVM) model.

Parameter	Weight									
	Gamma 0.1	Gamma 0.2	Gamma 0.3	Gamma 0.4	Gamma 0.5	Gamma 0.6	Gamma 0.7	Gamma 0.8	Gamma 0.9	Gamma 1.0
BLP	131.1	122.8	116.8	114.3	111.9	110.4	108.6	107.2	106.1	105.1
VGA	166.2	160.0	153.7	150.9	147.9	146.4	144.3	142.8	141.6	140.5
SNR	146.6	135.8	128.4	124.6	121.5	119.7	117.5	116.0	114.6	113.5
MEW	215.9	203.2	195.3	191.6	188.6	186.4	183.8	182.0	180.0	178.5
OW	213.9	201.0	193.0	189.4	186.2	183.7	181.0	179.3	177.4	175.9
EM	173.2	163.5	156.7	153.8	150.9	149.0	146.7	145.2	143.7	142.4

Table 4.11 performance support vector machine (SVM) model.

Radius	Precision	Recall	Accuracy
0.1	69.2%	55.0%	65.6%
0.2	69.1%	56.8%	66.1%
0.3	69.9%	58.5%	67.0%
0.4	70.7%	59.2%	67.7%
0.5	71.6%	59.0%	68.1%
0.6	71.8%	59.0%	68.2%
0.7	71.9%	59.3%	68.3%
0.8	71.4%	58.6%	67.8%
0.9	70.5%	58.3%	67.2%
1.0	70.5%	57.9%	67.1%

4.6 Data mining model compare performance result % accuracy

Figure 4.7 illustrates data mining operators setting to compare performance for 3 differrent methods, They consist of neural network (NN) model, naive bayes (NB) model and support vector machine (SVM) model. Firstly we prepare input data and save it at retrieve data location, then use operator multiply to separate input into each model, and then use operator X-Validation to compute performance. It has a training sub process and a testing sub-process. The training sub-process is used to train a model. Then apply in the testing sub-process. The performance of the model is measured during the testing phase.

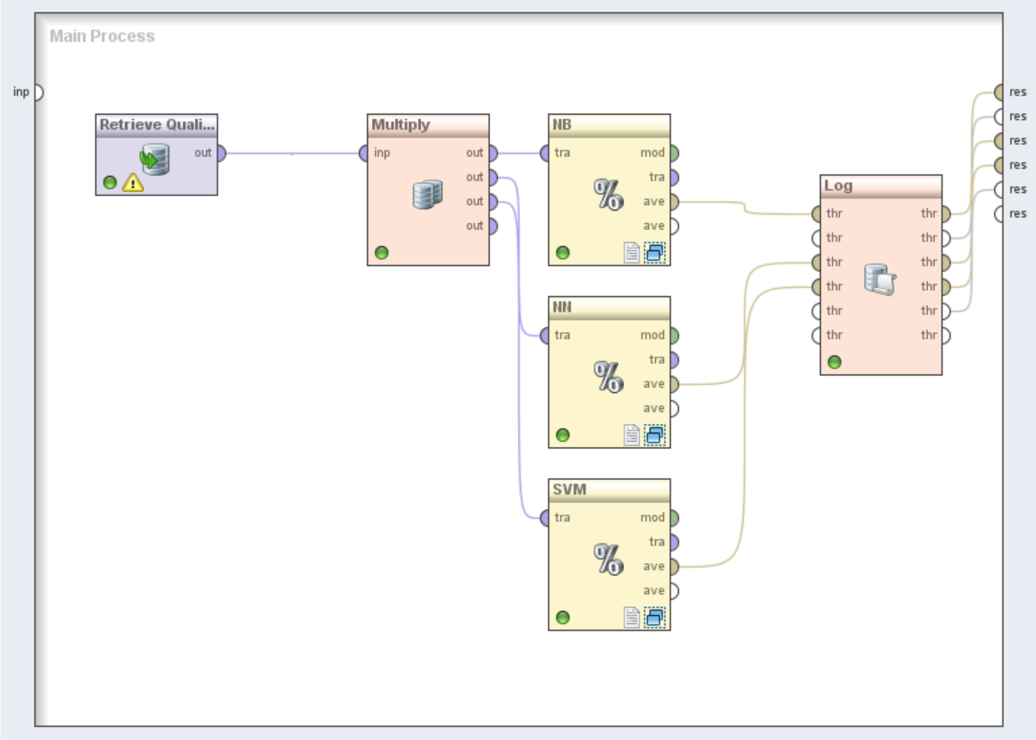


Figure 4.7 Data mining to compare performance of NB, NN and SVM.

Table 4.12 Compare prediction performance result (% Accuracy).

Method	Accuracy
Adaptive neuro-fuzzy inference system (ANFIS)	80.0%
Neural network (NN)	70.5%
Naive bayes (NB)	68.5%
Support vector machine (SVM)	68.3%

Table 4.12 illustrates the prediction result accuracy of 4 classification methods, we found that an adaptive neuro-fuzzy inference system (ANFIS) give a higher prediction performance of 80% when compared to other methods from data mining software, the result of neural network (NN) give the prediction accuracy of 70.5%, naive bayes (NB) give the prediction accuracy of 68.5% and support vector machine (SVM) give the prediction accuracy of 68.3%, respectively.

Chapter 5

Conclusions

This paper studies the prediction failure of head performance downgrade at quality test from an ANFIS method, neural networks (NN), naive bayes (NB) and support vector machines (SVM) by using the input data from the back end test process. Based on an initial study, we found that ANFIS method has a good relationship with quality failures. Specifically, we obtain the best prediction result at 80% accuracy. While neural networks (NN) model give prediction result at 70.5% accuracy, naive bayes (NB) model give prediction result at 68.5% accuracy, and support vector machines (SVM) give prediction result at 68.3% accuracy. Therefore, it can be implied that the ANFIS model can be used to classify the complicated data and to predict the failures of head performance downgrade at quality test with better results than other methods. Nonetheless, it should be pointed out that the proposed ANFIS model can still be improved to achieve higher accuracy. This can be done by using a better input parameter that is closely related to the head performance downgrade, and optimizing the ANFIS parameters. Once we obtain a better ANFIS model, it will be useful to improve the quality performance of HDD manufacturing.

Reference

[1] R. Duarte, **Environmental Monitoring of Clean Rooms**, DDK Scientific, Corp., 2008.

[2] B. Vasic and E. M. Kurtas, **Coding and signal processing for magnetic recording systems**, CRC Press, 2005.

[3] T. A. Roscamp, E. D. Boerner, and G. J. Parker, “Three-dimensional modeling of perpendicular recording with soft underlayer,” **Journal of Applied Physics,** vol. 91, no. 10, May 2002, pp. 8366.

[4] P. Kovintavewat and S. Koonkarnkhai, “Baseline Popping Detection and Correction Algorithms for Perpendicular Magnetic Recording System,” **Advanced Materials Research**, vol. 979, June 2014, pp. 50-53.

[5] X. Yao and Md. M. Islam, “Evolving Artificial Neural Network Ensembles,” **IEEE Computational Intelligence Magazine**, February 2008, pp. 31-42.

[6] Available: http://www.saedsayad.com/data_mining_map.htm.

[7] J. S. R. Jang, “ANFIS: Adaptive Neuro-Fuzzy Inference System,” **IEEE Transaction on System**, vol. 23, no. 3, June 1993, pp. 665-685.

[8] T. Takagi and M. Sugeno, “Fuzzy identification of systems and its applications to modeling and control,” **IEEE Transactions on Systems**, **Man and Cybernetics**, vol. 15, 1985, pp. 116-132.

[9] Available: http://sebastianraschka.com/Articles/2014_naive_bayes_1.html, **“Naive Bayes”**.

[10] C. Cortes and V. Vapnik, “Support-Vector Networks,” **Machine Learning**, vol. 20, September 1995, pp. 273-297.

[11] S. Roy, “Design of adaptive Neuro-Fuzzy Interface System for predicting surface roughness in turning operation,” **Journal of Scientific & Industrial research**, vol. 64, September 2005, pp. 653-659.

[12] A. Altaher, “Application of Adaptive Neuro-Fuzzy Inference System for Information Security,” **Journal of Computer Science** 8(6), 2012, pp. 983-986.

[13] T. M. Nazmy, H. EL. Messiry and B. AL. Bokhity “Adaptive Neuro-Fuzzy Inference System for Classification of ECG Signals,” **IEEE Informatics and Systems (INFOS), 2010 The 7th International Conference on**, March 2010, pp. 1-6.

[14] W. Tepin, “Customer Failure Modes Prediction for Hard Disk Drive using Neural Networks Rank-Level Fusion,” **in Proc. of ECTI-CON**, 2011, pp. 476-479.

[15] A. Wang, N. An, Y. Xia, L. Li and G. Chen, “A Logistic Regression and Artificial Neural Network-based Approach for Chronic Disease Prediction a case study of Hypertension,” **IEEE International Conference on Internet of Things (iThings 2014)**, 2014, pp. 45-52.

[16] X. Dang and T. Jiang, “Reliability prediction based on degradation measure distribution and wavelet neural network,” **IEEE Prognostics and System Health Management (PHM),** May 2012, pp. 1-5.

[17] A. Adam, L. C. Chew, M. I. Shapiai, L. W. Jau, Z. Ibrahim and M. Khalid, “A Hybrid Artificial Neural Network-Naive Bayes for Solving Imbalanced Dataset Problems in Semiconductor Manufacturing Test Process,” **IEEE Hybrid Intelligent Systems (HIS), 2011 International Conference,** December 2011, pp. 133-138.

[18] S. Aneja, “Effective Asthma Disease Prediction Using Naïve Bayes - Neural Network fusion technique,” **IEEE International Conference on Parallel, Distributed and Grid Computing,** 2014, pp. 137-140.

[19] V. Asawatongtip, “Data Mining For Predict Hard Disk Drive Downgrade,” **Asian Institute of Technology**, 2012.

[20] Available: http://en.wikipedia.org/wiki/Accuracy_and_precision.

[21] S. Wang, H. Chen and X. Yao, “Negative Correlation Learning for Classification Ensembles,” **Neural Networks (IJCNN)**, the 2010 International Joint Conference on Neural Networks, July 2010, pp. 2893-2900.

[22] Seagate Product Marketing, “Get S.M.A.R.T for Reliability,” **technology paper published**, 1999.

[23] Available: http://en.wikipedia.org/wiki/Types_of_artificial_neural_networks.

Appendix

List of international conference

[1] Rangson Hunkon[1], Pornchai Supnithi[2], Piya Kovintavewat[3] “ANFIS-Based Prediction of Head Performance Downgrade at Quality Test” The 29th International Technical Conference on Circuits/Systems, Computers and Communications (ITC-CSCC 2014)

ECTI Association

ESS

IEIE

Certificate of Participation

The 29th International Technical Conference on Circuit/Systems
Computers and Communications
(ITC-CSCC 2014)

Rangson Hunkon

is thanked for their phenomenal participation on

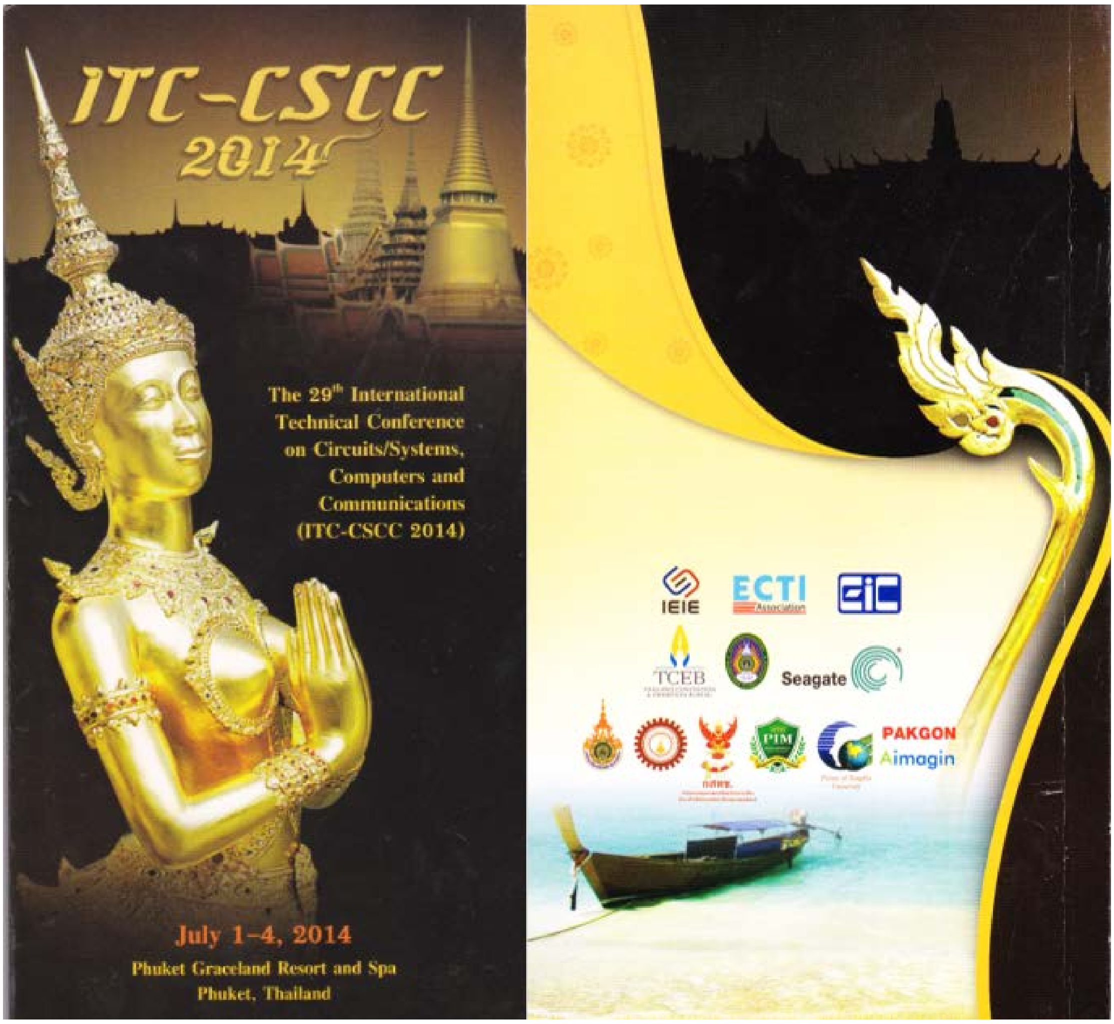
ITC-CSCC 2014, Phuket, Thailand July 1-4, 2014

LUNCHAKORN WUTTISITTIKULKIJ, PH.D.
TPC CHAIR
CHULALONGKORN UNIVERSITY

J. Heo
JUN HEO, PH.D.
TPC CO-CHAIR
KOREA UNIVERSITY

K. Hiraishi
KUNIHIKO HIRAISHI, PH.D.
TPC CO-CHAIR
JAIST

List of international conference (cont.)



ITC-CSCC 2014

The 29th International Technical Conference on Circuits/Systems, Computers and Communications (ITC-CSCC 2014)

July 1-4, 2014

Phuket Graceland Resort and Spa
Phuket, Thailand

ANFIS-Based Prediction of Head Performance Downgrade at Quality Test

Rangson Hunkon[1], Pornchai Supnithi[2]
[1]College of Data Storage Innovation,
[2]Faculty of Engineering, [1,2]King Mongkut's Institute of Technology Ladkrabang
Bangkok, Thailand
[1]rangson.hunkon@wdc.com, [2]ksupornc@kmitl.ac.th

Piya Kovintavewat[3]
[3]Data Storage Technology Research Center,
Nakhon Pathom Rajabhat University
Nakhon Pathom, Thailand
[3]piya@npru.ac.th

***Abstract*—The hard disk drive (HDD) quality failure is an important parameter that crucially impacts factory efficiency and productivity. In general, a major failure is normally caused by head performance downgrade. Thus, this paper proposes a method to predict this failure, which in turn can help improve HDD reliability. Specifically, we apply an analytic tool called an adaptive neuro-fuzzy inference system (ANFIS) to predict the head performance downgrade, and find an opportunity to reduce the quality failure. As an initial study, we found from an experiment that the ANFIS model can be employed to predict a failure from head performance downgrade at quality test with an accuracy of about 80%.**

***Keywords*—Adaptive neuro-fuzzy inference system (ANFIS), head performance downgrade, quality test.**

I. INTRODUCTION

Hard disk drive (HDD) is a growing business due to worldwide customers' demand. In practice, HDDs are important for data storage in every company and personal use (e.g., desktop, notebook, tablet, game boxes, and so on). Hence, HDD manufacturer must improve the processes continuously to get the highest output with the highest quality to support customer requirements. Generally, the HDD manufacturing process can be divided into two processes. The first build is done in a class 100 clean room [1], where contamination is carefully controlled. Next, the second build is performed in a backend area, where drives are tested for performance and quality control before shipping them to customers, as illustrated in Fig. 1.

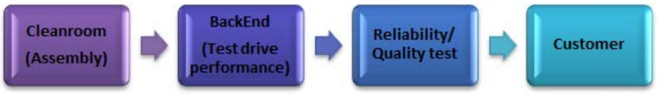


Fig. 1. A flow of hard disk drive manufacturing process.

Practically, HDD manufacturer must continuously improve the manufacturing processes and the quality performance so as to support customer's requirements. However, we still found many HDD failures returned from customers. In general, the failure at quality test is about 1.5%, whose major cause is head performance downgrade. This paper is aimed at investigating if there is a possible way to identify such potential failures and predict the failures at quality test so as to let them fail at the backend test process and also prevent failed drives shipping to the customers.

To study the prediction and reduce head performance downgrade at quality test, we found that the existing key head stack parameters are BLP (baseline popping), VGA (variable gain amplifier), SNR (signal-to-noise ratio), MEW (magnetic erasure width), OW (overwrite), and EM (error margin) from the backend test process. These parameters are required for an analytic tool called an adaptive neuro-fuzzy inference system (ANFIS) [2] to predict the failures.

Practically, ANFIS is a fuzzy inference system implemented in a framework of an adaptive network. By using a hybrid learning procedure, the ANFIS can construct an input-output mapping, based on human knowledge and stipulated input-output data pairs. In general, ANFIS is suitable for modeling a non-linear system and predicting a chaotic time series. For example, Roy [3] employed the ANFIS to predict the surface roughness in a turning operation for a set of given parameters with two different membership functions (MFs) and then compared the prediction accuracy. It was found that the bell-shaped MF has the prediction accuracy of 97.84%, whereas the triangular MF has the prediction accuracy of 96.13%. Altaher [4] studied a neural fuzzy classifier based on ANFIS for malware detection, which was found that this ANFIS classifier can detect the malware exe files effectively. Nazmy and Messiry [5] presented an intelligent diagnosis system using a hybrid approach of ANFIS for classification of the electrocardiogram signals, whose results indicated an accuracy level of more than 97%. Tepin [6] proposed a neural network rank level fusion applied on key parameters measured in the manufacturing process to predict customer failures resulted from head disk interaction (HDI). It was found that the result of rank level fusion classification model is able to achieve 86.61% accuracy for testing samples, and potentially to affect HDI failure. Then, Asawatongtip [7] introduced a method to predict the root causes of drive downgrade during the assembly and test processes by data mining using the Bayesian network, which a achieve an accuracy of 80.7%. Therefore, this paper proposes a new prediction method based on an ANFIS model by using the head stack input parameters to predict head performance downgrade at quality test.

The rest of this paper is organized as follows. Section II briefly summarizes an ANFIS model. Section III explains the experimental method. Simulation results are given in Section IV. Finally, Section V concludes this paper.

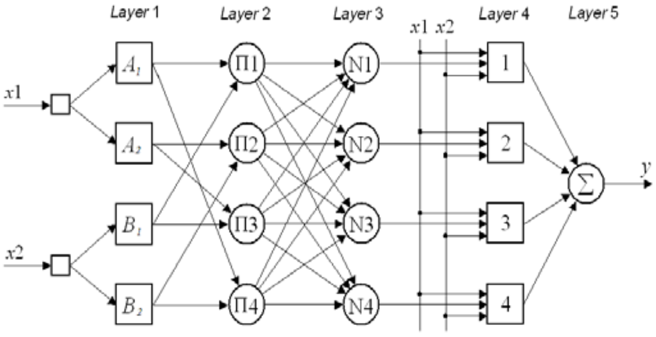


Fig. 2. Adaptive neuro-fuzzy inference system (ANFIS).

II. ADAPTIVE NEURO-FUZZY INFERENCE SYSTEM

Fuzzy inference systems are efficient techniques for studying the behavior of nonlinear systems by using fuzzy logic rules. ANFIS is an adaptive neuro-fuzzy inference system that uses the learning techniques of neural networks, which is usually employed in many applications in control and prediction. Practically, ANFIS utilizes a hybrid learning algorithm to specify parameters. Specifically, it uses the least-squares method with the back propagation gradient descent method to train the ANFIS membership function parameters based on a given training data set.

In general, the ANFIS structure is similar to a neural network structure. Fig. 2 illustrates the ANFIS model based on Takagi and Sugeno model [8]. Clearly, it consists of 5 layers connected through direction links, where the 1st is a fuzzy layer, the 2nd is a product layer, the 3rd is a normalized layer, the 4th is a de-fuzzy layer, the 5th is a total output layer, x_1 and x_2 are the inputs, and y is the output. Note that each layer is characterized by a node function with fixed adjustable parameters.

According to the Takagi and Sugeno model [8], the rule sets and the function of each layer are as follows

Rule Set:

If (x_1 is A_1) and (x_2 is B_1) then $f_1 = p_i x_1 + q_i x_2 + r_i$

If(x_1 is A_2) and (x_2 is B_2) then $f_2 = p_2 x_1 + q_2 x_2 + r_2$

where p_1, p_2, q_1, q_2, r_1 and r_2 are linear parameters and A_1, A_2, B_1 and B_2 are non-linear parameters.

Layer 1: It is an input fuzzy layer. Every node in this layer is an adaptive node that satisfies the following equations

$$O_{1,i} = \mu_{A_i}\left(x_1\right), \qquad i = 1, 2 \tag{1}$$

$$O_{1,i} = \mu_{B_i-2}\left(x_2\right), \qquad i = 3, 4 \tag{2}$$

where $O_{1,i}$ denote the output functions, μ_{A_i} and μ_{B_i} denote of the membership functions.

A membership function for a fuzzy set A on the universe of discourse x is defined as $\mu_A(x) \rightarrow [0, 1]$, where each element of x is mapped to a value between 0 and 1. This value quantifies the grade of membership of the element in x to the fuzzy set A. The membership function (MF) allows to graphically represent a fuzzy set, where the x axis represents the universe of discourse, and the y axis represents the degrees of membership in the [0,1] interval. For instance, if the triangular MF is employed, μ_{A_i} is given by

$$\mu_{A_i}\left(x\right) = \max\left[\min\left(\frac{x-a_i}{c_i-a_i}, \frac{c_i-x}{c_i-b_i}, 0\right)\right], \tag{3}$$

where a_i, b_i, and c_i are the MF parameters. On the other hand, if the generalized bell-shaped MF is used, μ_{A_i} will be given by

$$\mu_{A_i}\left(x\right) = \frac{1}{1+\left(\frac{x-c_i}{a_i}\right)^{2b_i}}, \tag{4}$$

where a_i, b_i, and c_i are the MF parameters.

Layer 2: It is a product layer, where each node in this layer computes the impact of each rule through the multiplication by

$$O_{2,i} = w_i = \mu_{A_i}\left(x_1\right)\mu_{B_i}\left(x_2\right), \tag{5}$$

where $O_{2,i}$ is the output of layer 2.

Layer 3: It is a normalization layer, which computes the normalized effect of a given rule by

$$O_{3,i} = \overline{w}_i = \frac{w_i}{w_1 + w_2}, \tag{6}$$

where $i \in \{1, 2\}$, and $O_{3,i}$ is the output of layer 3, which can be called "normalized firing" strength are normalized with a maximum equal to 1 and a minimum equal to 0.

Layer 4: It is a de-fuzzy layer, where the parameters in this layer are considered as consequent parameters that follow

$$O_{4,i} = \overline{w}_i f_i = \overline{w}_i \left(p_i x_1 + q_i x_2 + r_i\right), \tag{7}$$

where $O_{4,i}$ is the output of layer 4 and $\{p_i, q_i, r_i\}$ are called linear parameters or consequent parameters.

Layer 5: It is a total output layer to calculate the sum of the output for all incoming signals given by

$$O_{5,i} = \sum_i \overline{w}_i \ f_1 = \frac{\sum_i w_i f_i}{\sum_i w_i} \ , \tag{8}$$

where $O_{5,i}$ is the total output layer and $\sum$ computes the sum of all incoming signals.

III. EXPERIMENTAL METHOD

This study presents a prediction method of head stack performance downgrade at HDD quality test, which is based on ANFIS functions in MATLAB R2008b, where 6 input parameters (BLP, VGA, SNR, MEW, OW and EM) and 1 output parameter (Pass or Fail) are considered in our ANFIS model. Note that based on our observation from historical data statistics, we found that these 6 input parameters in the HDD backend test process are the most relevance to the head performance at the quality test.

First, we collect the unseen data used for testing about 10000 data points in the manufacturing test process. Then, we divide the data (passed and failed) into two groups randomly, with 80% for a training data set and 20% for a verifying data set. In the training process, we separate it into 2 steps, where the first step is to find the MF by using a “genfis2” function in MATLAB and to determine the membership boundary by radius between 0.4 and 1.0, and the second step is to adjust the ANFIS model to obtain the best settings for prediction. After the ANFIS model with proper parameters is obtained, we then verify the accuracy of the model by using the (unseen) verifying data set to verify the accuracy of the model.

Fig. 3 shows the data distribution of 6 input parameters, which is obtained by separating each input into 1 dimension and 2 dimensions for both passed (blue color) and failed (red color) drives. It is apparent that this distribution data is complicated to classify the data. For example, let us consider the two inputs BLP and EM (i.e., a lower-left corner figure). The scatter plot cannot tell us that the passed and the failed drives are not significant different. However, we can utilize the ANFIS model to classify and predict the failures as shown in simulation.

IV. SIMULATION RESULTS

In this section, we will verify the sensitivity and the accuracy of the proposed ANFIS model that is used to predict the head performance downgrade, as depicted in Fig. 4 and Fig. 5, respectively, where the x-axis is a radius between 0.4 and 1.0, and the y-axis is the percentage of accuracy. In addition, the blue line represents 6 inputs (all features), the orange line represents 3 inputs (BLP, SNR, EM), the red line represents 2 inputs (BLP, EM) and the yellow line represents 2 inputs (SNR, EM). These inputs are the data used as the inputs to the ANFIS model. In general, we need to consider the results of both sensitivity and accuracy so as to explain the prediction performance.

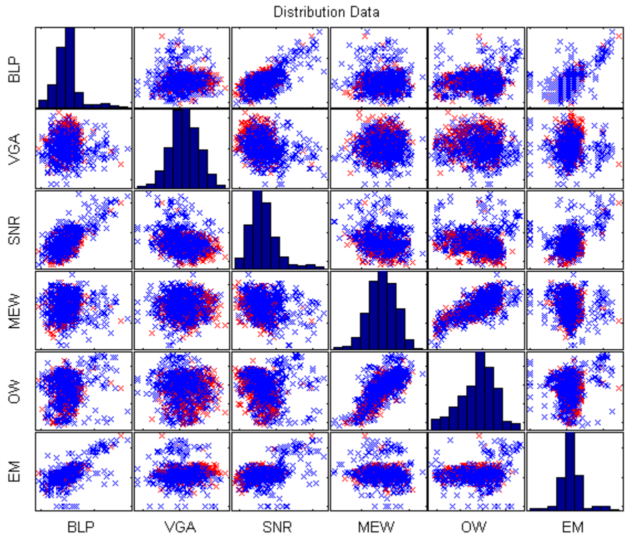


Fig. 3. The distribution data of 6 inputs parameter.

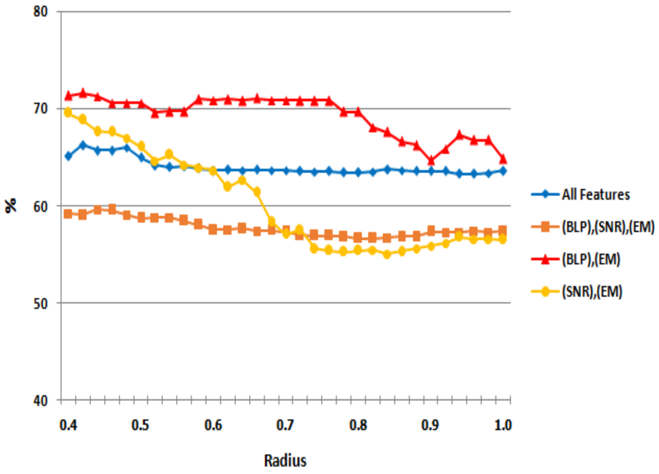


Fig. 4. Sensitivity of ANFIS prediction.

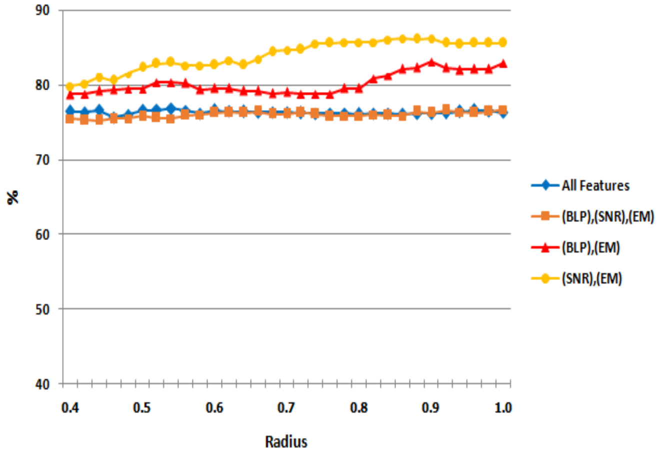


Fig. 5. Accuracy of ANFIS prediction.

In this work, the prediction accuracy and the sensitivity are evaluated by

$$\text{Accuracy} = \frac{TP+TN}{TP+FP+FN+TN}, \qquad (9)$$

$$\text{Sensitivity} = \frac{TP}{TP+FN}, \qquad (10)$$

where *TP* is True Positive (prediction of 1 when the sample test result has a 1), *TN* is True Negative (prediction of 0 when the sample test result has a 0), *FP* is False Positive (prediction of 1 when the sample test result has a 0), and *FN* is False Negative (prediction of 0 when the sample test result has a 1).

As shown in Fig. 4 and Fig. 5, if we use 6 inputs (all features) in the ANFIS model, it will approximately give the prediction error with sensitivity of 65% and accuracy of 77%. However, we found that using only 2 inputs (i.e., BLP and EM) as the inputs to the ANFIS model, we can obtain the best prediction result, i.e., the prediction error with sensitivity of 70% and accuracy of 80%.

Based on the result, we found that using two inputs will yield a better prediction result than using six inputs. This might be because six parameters may have more complexity and overfitting than two parameters. It should be noted that overfitting generally occurs when a model is excessively complex, such as having too many parameters relative to the number of observations. A model which has been overfitting will generally have poor predictive performance, as it can exaggerate minor fluctuations in the data.

V. CONCLUSIONS

This paper studies the prediction failure of head performance downgrade at quality test by using an ANFIS method and the data from the backend test process. As an initial study, we found that only two input parameters, namely BLP and EM, have a good relationship with quality failure. Specifically, we obtain the best prediction result at 80% accuracy. Therefore, it can be implied that the ANFIS model can be used to classify the complicated data and to predict the failures of head performance downgrade at quality test with good result. Nonetheless, it should be pointed out that the proposed ANFIS model can still be improved to achieve higher accuracy. This can be done by using a better input parameter that is closely related to the head performance downgrade, and optimizing the ANFIS parameters. Once we obtain a better ANFIS model, it will be useful to improve the quality performance of HDD manufacturing.

ACKNOWLEDGMENT

I would like to thank you Assoc.Prof.Dr.Pornchai Supnithi (King Mongkut's Institute of Technology Ladkrabang) and Assoc.Prof.Dr.Piya Kovintavewat (Data Storage Technology Research Center, Nakhon Pathom Rajabhat University), who provided me the guidance, advices and reviews my work. Also, I would like to thank you my WD Company to support everything during I study and do my thesis.

REFERENCES

[1] R. Duarte, "Environmental Monitoring of Clean Rooms", DDK Scientific, Corp., 2008.

[2] J. S. R. Jang, "ANFIS: Adaptive Neuro-Fuzzy Inference System", IEEE Transaction on System, vol. 23, no. 3, 1993.

[3] S. Roy, "Design of adaptive Neuro-Fuzzy Interface System for predicting surface roughness in turning operation", Journal of Scientific & Industrial research, vol. 64, 2005.

[4] A. Altaher, "Application of Adaptive Neuro-Fuzzy Inference System for Information Security", Journal of Computer Science 8(6): 983-986, 2012.

[5] T. M. Nazmy, H. EL-Messiry, "Adaptive Neuro-Fuzzy Inference System for Classification of ECG Signals", Journal of Theoretical and Applied Information Technology, 2009.

[6] W. Tepin, "Customer Failure Modes Prediction for Hard Disk Drive using Neural Networks Rank-Level Fusion", in Proc. of ECTI-CON, pp. 476 – 479, 2011.

[7] V. Asawatongtip, "Data Mining For Predict Hard Disk Drive Downgrade," Asian Institute of Technology, 2012.

[8] T. Takagi and M. Sugeno, "Fuzzy identification of systems and its applications to modeling and control", IEEE Transactions on Systems, Man and Cybernetics, vol. 15, pp. 116-132, 1985.

Author Biography

Personal Information

Name & Surname: Mr. Rangson Hunkon
Nationality: Thai
Birth of date: June 3rd, 1981
Place of birth: Ratchaburi, Thailand

Education

Bachelor degree

Field: Power Electronics Engineering
Duration: 2003 - 2005
Institute: Power Electronics, Faculty of Engineering King Mongkut’s University of Technology North Bangkok, Thailand

Master degree

Field: Data Storage Technology
Duration: 20010 - 2014
Institute: College of Data Storage Innovation, Faculty of Engineering, King Mongkut’s Institute of Technology, Ladkrabang, Thailand

Research Interests

Data Storage Innovation, Hard disk drive technology, Read channel in hard disk drive.