PREDICTION OF RECORDING HEAD RELIABILITY DEGRADATION USING NEURAL NETWORK FUSION WITH NEGATIVE CORRELATION

CHALERMCHAI JORJONG

A THESIS SUBMITTED IN PARTIAL FULLFILLMENT

OF THE REQUIREMENT FOR THE DEGREE OF

MASTER OF ENGINEERING IN DATA STORAGE TECHNOLOGY

INTERNATIONAL COLLEGE

KING MONGKUT'S INSTITUTE OF TECHNOLOGY LADKRABANG

2013

KMITL-2013-IC-M-005-013

Thesis: Prediction of Recording Head Reliability Degradation using Neural Network Fusion with Negative Correlation.

Student: Mr. Chalermchai Jorjong

Student ID: 52600627

Degree : Master of Engineering (M.Eng)

Program: Data Storage Technology

Year: 2013

Thesis Advisor Asst. Prof. Dr. Yuttana Kitjaidure

ABSTRACT

Hard disk drive industry is growing years over years with many new technologies. New technology always has a challenge in reliability. The determination of the reliability failures is another challenge. As in any industry, the drive industry reliability test takes long test time and almost needs to wait till the end of the test to determine the results. This delay has an impact in cost and the product launching time. In order to avoid the time delay, the decision making in current practice is normally done based on the early results during the reliability testing. This paper focuses on predicting the reliability failure based on the parametric data using combination of Homogeneous and Non Homogeneous Classifier with Negative Correlation (NC) on Neural Network (NN) fusion techniques. The results show that the combination models which have the most robust performance with higher level of accuracy. By this way the test time could be very well reduced.

Keywords— Classification, Reliability, Negative correlation, Neural Network fusion.

Acknowledgements

I would like to gratefully acknowledge an Assist Professor Dr. Yuttana Kitjaidure my advisor for his helpful and support during research at KMITL. I would also like to thank CW Mark, Kittiphan, Pricha L and Selvan HOD of HDD Engineering for critically reviewing of material, and particularly for pointing out the head reliability degradation performance with parametric data.

Thanks are also due to my managements Vankat, Somboon and Weerapat HOD of Quality Assurance for encouraging the work and un-restrictively allowing me to use as much time as I needed , Kritsada, Wiwat and Waraporn in the conduct of the various research studies reported and the preparation for suggestion, advice and consultant to success this thesis.

I wish to thank Western Digital for supporting on the Master scholarship and for providing data that is necessarily for the paper and hope the contents of this research can contribute on product reliability.

The research was supported in part by a grant from cooperation among the Western Digital Company, Department of Reliability Engineering, NECTEC Institute of Thailand's Goverment and King Monkut Institute of Technology Lard Krabang.

I would also like to thank my parents and family for the support all these years.

Chalermchai Jorjong

Contents

Page

Contents (Cont.)

List of Figures

Fig Page

List of Figures (Cont.)

Fig Page

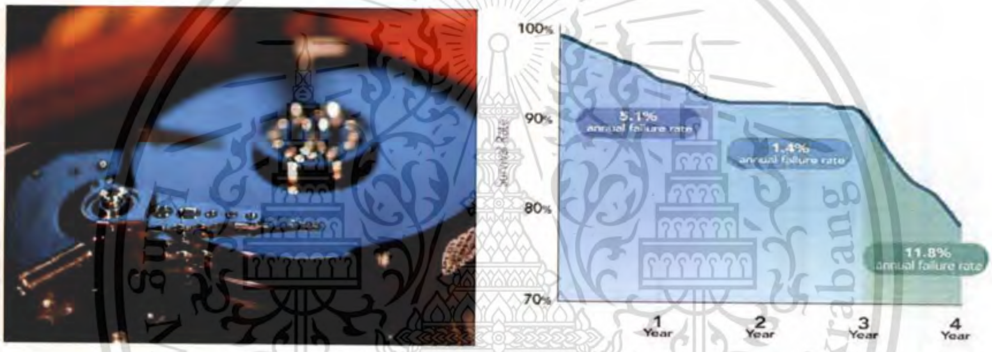
List of Tables

Chapter 1

Introduction

1.1 Statement and Significance of the Problem

A hard disk drive is very complex which consists of electrical, mechanical and PCBA (Print circuit board assembly) part as shown in Fig.1.1 (a). This is an example of report published regard hard disk drives reliability. How long do hard drives actually live for ?.



(a) (b)

Fig. 1.1 Backblaze hard drive failure rate [1]

In [1], Back blaze is online backup company that keeps 25,000 hard drives failure record in the past 4 years. They summarized the results and published graphs as shown in Fig. 1.1 (b). At 92% of drives survives the first 18 months, and almost all of those (90%) then go on reaching three years and under 80% of all hard drives survive to their fourth anniversary. Figure 1.2 shows the bathtub curve that reliability engineers use to explain the three distinct phases of a product's lifecycle. It can segregate in 3 periods as Infant mortality, stead state and wear out periods. The first year are primarily caused by manufacturing defects. Between 18 and 36 months, drive deaths are induced by random failures. Afterwards, as the drive moves into its fourth year, failure rates increase as drives start to wear out.

According to the Fig. 1.2, HDD manufacture uses reliability test to capture the field failure in early infant mortality and steady state periods.

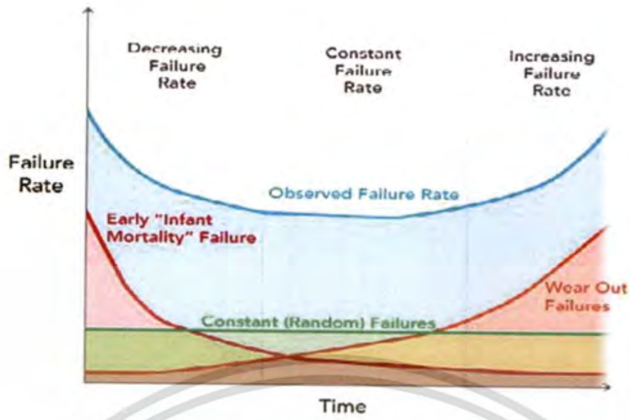


Fig. 1.2 The bathtub curve of hard disk drives. [1]

In [2], The reliability program plan is utilized to verify the product. It covers the chain of supply from product definition to volume ramp and from suppliers to clients. The purpose is to emulate the field failure. Every product will possess its own reliability program covering a design for reliability grow, reliability check list, the and reliability demonstration testing and field reliability tracking. In HDD, we use reliability demonstration test (RDT) to demonstrate the product reliability and predict the annual failure rate (AFR) and mean time to failure (MTTF).

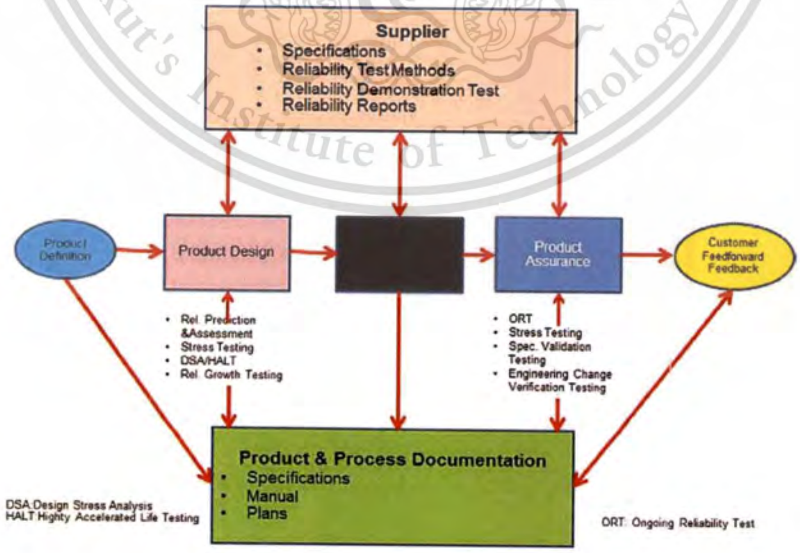


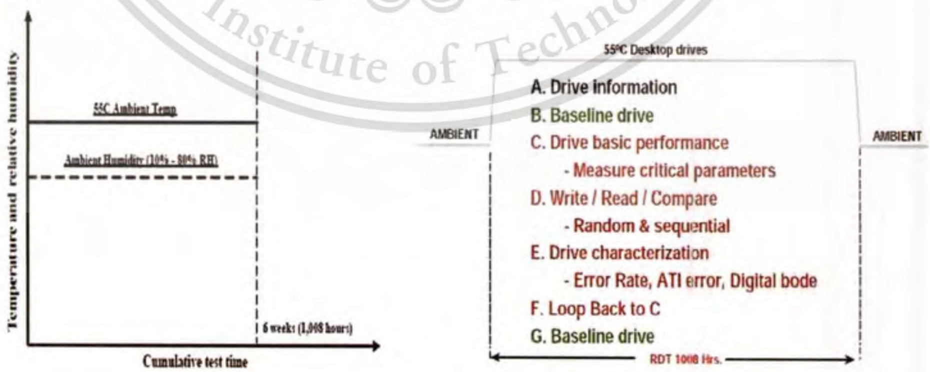
Fig. 1.3 HDD factory reliability program plan [1].

The purpose of RDT is to show that the product in volume production or its equivalent meets a minimum specified MTTF in order that the steady state MTTF goal can be executed in the field. In the meantime, RDT give management with decision/action criteria to assure timely consistent, proportionate, problem-solving responses with closed-loop feedback. Currently, we use Advance Single Plug Tester (ASPT) to conduct the reliability test as shown in Fig. 1.4.

Test Attributed	Product A	Product B
Duyt cycle	100%	100%
Rad Vs writes	64:36	50:50
Random vs sequential	40:60	70:30
Temperature	Base Casting @60C	Base Casting @50C
Power Cycles	50/day	10/day
Voltage	5v/12v nominal	5v/12v nominal
MTTF Calculation POH	3120	8760
Drive Quantity	1000	1200
Duration hours	1008	1008

Fig. 1.4 The ASPT and test condition

All drives will be operated by using RDT test script. The drive will run read/write by random and sequential, comparing with spindle idle and standby functions to emulate specific operations. The power will be cycled once per day as shown in Fig. 1.5.



(a) (b)

Fig. 1.5 RDT test profile and test operation

All drives were kept at constant high temperature/humidity for 1-1.5 months duration time. The temperature/humidity were set at 55°C+/-3°C and 50% RH at nominal input voltage based on specification. Figure 1.5 (b) shows that the test operation starting from drive information collected subsequently by using baseline drives and drive basic performance measurement after performing write/read comparison and Hard Disk Drive (HDD) characterization and then repeating until the test is completed.

The general RDT procedure can be summarized as follows:

1. Install hard disk drive samples into the advance single plug tester (ASPT).
2. Keep drives operate functional at environment specification.
3. Monitor test result with automate tool in daily basis.
4. Take drives out when finding error message and the time to failure (TTF) recorded.
5. The defect was send to Failures Analysis laboratory to identify the possibly root cause.
6. Reliability engineers perform reliability assessment (AFR and MTTF) and work with various engineering teams for corrective/preventive action.

The key parameters of RDT can be described as follows:

1. Duty cycle is account the failure rates with how much the device is used.
2. Write and Read is the process to writing the signal into the media and reading back the data.
3. Random vs sequential is the sequential and random for read/write signal.
4. Transfer length is a transfer rate data while read and head function.
5. Temperature is temperature accelerate stress condition.
6. Power cycles is the loop number of power turn on/off supply to disk drives.
7. Voltage is the supply voltage to hard disk drives.
8. MTTF calculation (POH) is the average time to failure of a population of units under specified conditions.
9. Drive quantity is the number of test sample.
10. Duration hours is how long drives perform in environmental specification.

AFR is a commonly utilized term in hard disk drive and system industry for reliability performance in the field. AFR can be reckoned on the basic of the equation 1.1. MTTF is mean time to failure and can be calculated via power on hours divided by AFR number based on two-phase bathtub curve as shown in equation 1.2.

Temperature and Work load are important factors which affect AFR and MTTF number for modern Hard disk drives.

$$AFR = fx\frac{POH}{t_T}x(\mathrm{UT})^{0.6}x\exp\left|-\frac{0.4eV}{k}(TM)\right| \tag{1.1}$$

Where

f is raw test failure percentage

POH is Power on Hours

AFR is Annualized failure rate

UT is User Term

TM is Thermal Term

WL is Work Load

Tr is test duration

T is Absolute temperature

K is Boltzmann't constant (8.2e-5eV/K)

MTTF is mean time to failure

$$\mathrm{MTTF} = \frac{\mathbf{POH}}{\mathbf{AFR}} \tag{1.2}$$

Figure 1.6 shows that at highest total host transfer rate, the MTTF number is getting lower while AFR % is getting higher. A higher MTTF is to ensure the product survival at the design life.

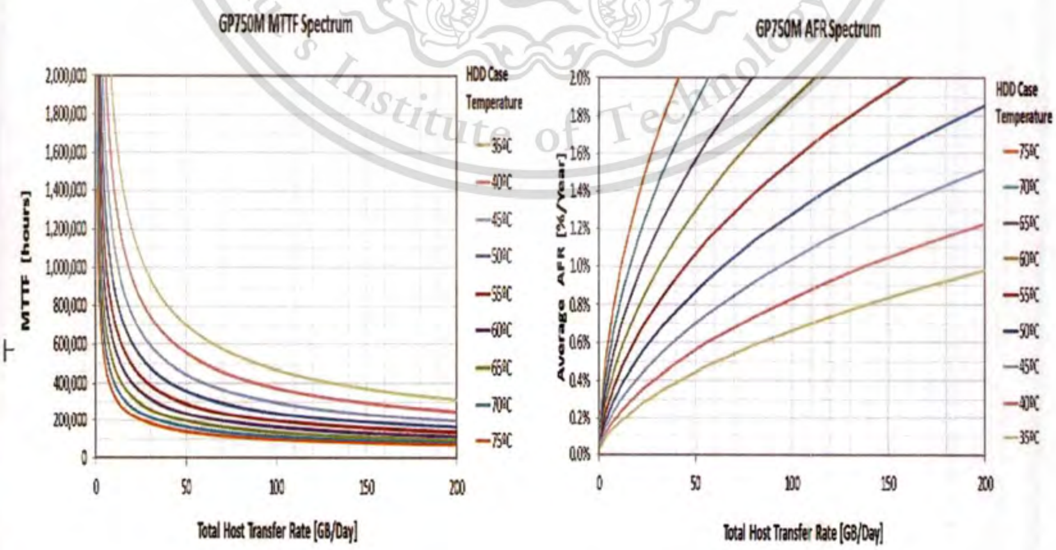


Fig. 1.6 MTTF/AFR Vs Total Host Transfer Rate

The sample size is derived from this and explains 20% allowable sampling error at a 70% one-sided confidence. The sample size *n* and number of failures *np* are discrete quantities and are described by the binomial distribution. We need to define the acceptable confidence level and acceptable error margin is given by.

$$n = \frac{1}{p}\left(\frac{z}{e}\right)^2 \tag{1.3}$$

Where

n is required sample size

p is test failure rate (percentage)

z is expresses confidence level

e is acceptable error

A total of 300pcs disk drives were RDT tested for 600 hours during which a total of 32 failures were detected. The failure top pareto is BER/EM degrades, BLN and BLP failures and Scratch failure in turn. Thus, head related failure is a major failure bucket relating to BER degradation over time. In reliability test, there are 6 critical parametric uses to monitor as shown in Fig. 1.7.

.

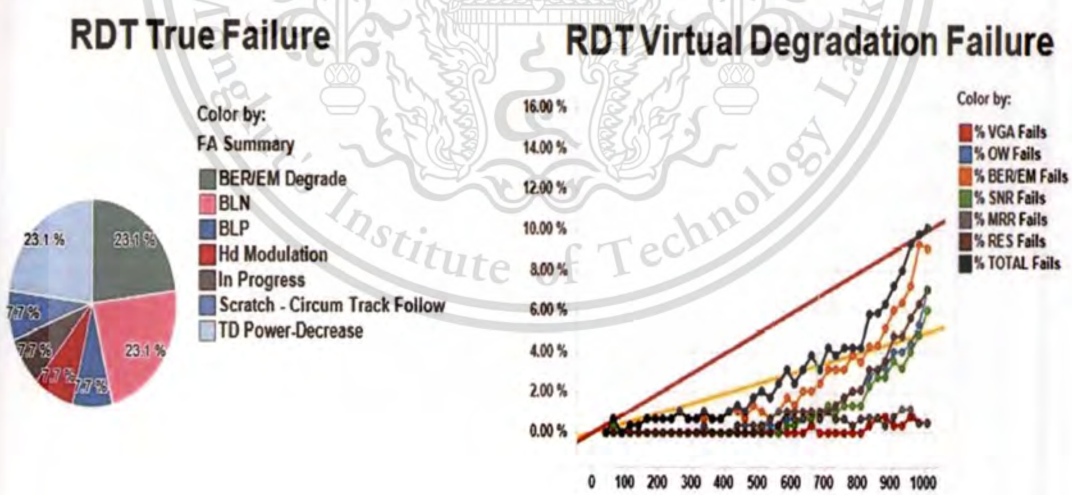


Fig. 1.7 The RDT test Result

Figure 1.8 shows the example of key parametric used to monitor in reliability demonstration test (RDT), namely Bit Error Rate (BER), Over Write (OW), Resolution (RES) and Mag Spacing parameter. Each color line represents head ID and X-axis represents duration time. Y-axis is bit error rate delta percent change. The bit error rate (BER) will degrade when test time is longer. This is to prove that the head can perform at high environmental stress and functional condition over the time.

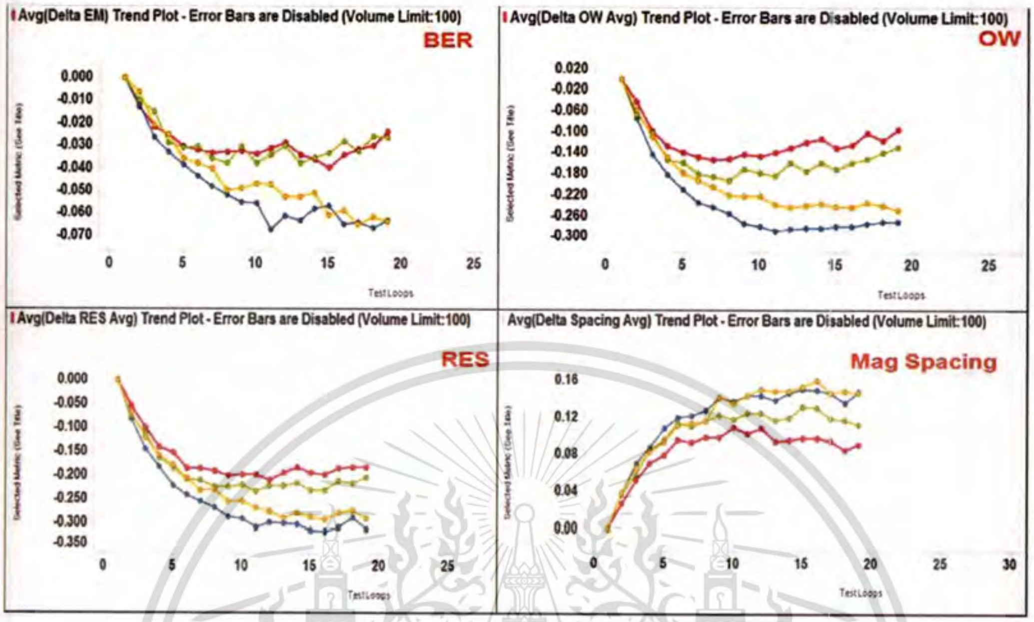


Fig. 1.8 The parametric monitoring in RDT.

All of the key parameters will be gathered during the tests for each test profile including head parametric for head degradation monitoring (Head Asymmetry, Head Amplitude, Bit Error Rate (BER), all errors both recoverable and unrecoverable, throughput performance, total bytes read and written, and SMART attributes. In this paper, we select the bit error rate (BER) to study because it leads to a head related failure as indicated in virtual failure rate (VFR) result.

The key parametric monitoring in RDT can be described as follows:

1. BER is gain of the amplifier to read the signal back.
2. The overwrite (OW) capability is to measure the writer performance during write process. If the poor over OW occurs, it means that the writer does not work properly as expected.
3. RES stand for resolution
4. The non-repeatable run out (NRRO) is the index to reflect the mechanical performance during Hard Disk Drives operation.

The challenge of HDD manufacture is to introduce the innovated product and launch to the market with short period of time with high product reliability. Nowadays, the HDD manufacturer uses statistical technique to predict the reliability failure of their products, which depend on the test result to determine the field failure. A modern HDD generation has higher level of complexity, which is too complicated to determine the failures. Consequently, we introduce a new technique by using the parametric data namely bit error rate (BER) to predict the product reliability performance. The bit error rate is one of critical parameters which is collected during reliability testing. This parameter is the most important indicator to determine the head degradation and clearly represents the passed and failed criteria for the drive failure mode in HDD reliability test. The HDD reliability test takes longer test time for more complexity and almost needs to wait until the end of the test to determine the results. The long test time impacts HDD and product launch schedule. In order to avoid the time delay, the reliability failure prediction model is aim to study in this thesis in order to find the new model that can provide early decision making.

1.2 Objective

1.2.1 To study the head reliability degradation pattern based on critical head parametric data input namely bit error rate.

1.2.2 To study the prediction model of head reliability failure by using the combination of Homogeneous and Non Homogeneous Classifier with Negative Correlation (NC) on Neural Network (NN) Fusion techniques.

1.2.3 To provide the high accuracy reliability prediction result.

1.2.4 To study the reliability test time using the failure prediction pattern.

1.3 Expectation

1.3.1 This study expect to create the new feature extraction and prediction model to determine the reliability of recording head in HDD.

1.3.2 The prediction model is able to provide higher accuracy of reliability prediction.

1.3.3 The prediction model is able to reduce the reliability test time.

1.4 Conceptual Framework

In order to predict the bit error rate degradation pattern, we use neural network methodology, which is widely used in intelligent system to discover the relationship between time periods. The learning and prediction of the neural networks are based on time series data input, which is collected from the 1^{st} day of reliability test to the last day of reliability test. A neural-network approach is proposed for recording head reliability prediction of HDD by using the combination of Homogeneous and Non

Homogeneous Classifier. This algorithm provides the best combination of classifier and is the approach of a reliability monitoring tool as a pre-alert mechanism to screen/detect the potential of head related failure in HDD final test process. To enhance the prediction model capability, negative correlation is introduced. The advantage of this algorithm is to eliminate the insignificant classifier output in order to reduce the complexity of the system. This research topic is challenging because less research paper is published results. Therefore, research opportunity is in this topic.

1.5 Hyphothesis

In HDD reliability top failure mode analysis reveals the highest head related failure which leads to BER degradation over the test time. RDT is required longer time affecting product launch schedule. Nowadays, it takes 1-1.5 months to be completed. Shorter test time is beneficial to the testing cost and the product launch lead time. The motivation of this research is it will be better if we are able to predict the head reliability failure, based on bit error rate (BER) and do not wait until the test is completed.

1.6 Thesis Outline

The remainder of the thesis is organized as follows:

Chapter 2 presents the background information and reviews of theories, methods and parameter in the literature.

Chapter 3 explained the feature extraction, an introduce the new reliability prediction modeling algorithms namely Homogenous and Non- Homogeneous with negative correlation via neural network fusion techniques.

Chapter 4 described the experimental result and discussion.

Chapter 5 will conclude the final model and the suggestion for further work.

Chapter 2

Literature Review

In this chapter, we reviewed the principle of Hard disk drive reliability; the concepts of neural network along with related previous work are discussed as follows:

2.1 Hard Disk Drive
2.2 Hard Disk Drive Reliability
2.3 Bit Error Rate
2.4 Related previous work

2.1 Hard Disk Drives

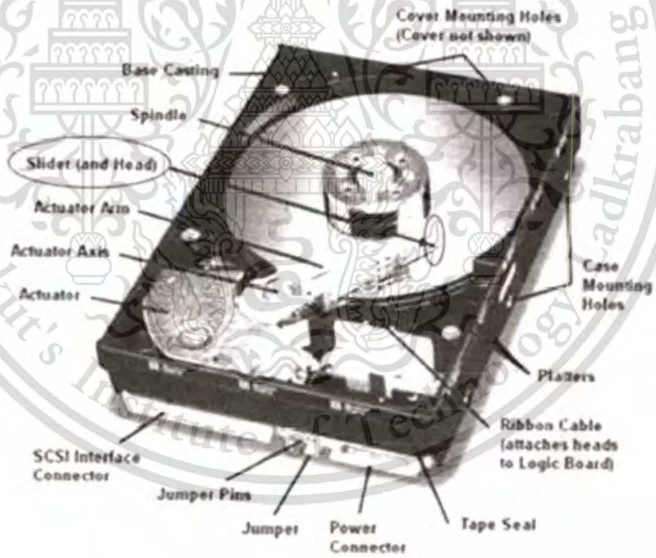


Fig. 2.1 A basis Hard Disk Drive [4]

Hard Disk Drives (HDD) is data storage device consisting of electrical and mechanical component such as magnetic head, magnetic recording media, spindle motor and Print circuit board assembly (PCBA) as shown in Fig. 2.1. The Head and Media are most critical components where the head is transducer to convert electrical pulses to magnetic transitions and the media is used to store magnetic transitions as magnetic bits. Modern HDD generation over generation has increasing area density to extend the capacity support customer demand with an increase of digital multi-media information. In the market, HDD competitors have launched the new product with short period of time. Lots of models were introduced to the market from time to time so that the challenge of HDD manufacture could renovate their products with lower cost, smaller size, higher capacity and faster execution. In parallel, HDD makers also concentrate and improve on product reliability performance to make their customers fully satisfied.

In general, for most products and applications, the top failure pareto majority of Hard Disk Drive failures are driven with a variety of failure modes as Head related failures, Hard Defect and High Fly light and the other failures are no fault found (NFF) and PCBA (Print circuit board assembly) dominant as shown in Fig. 2.2.

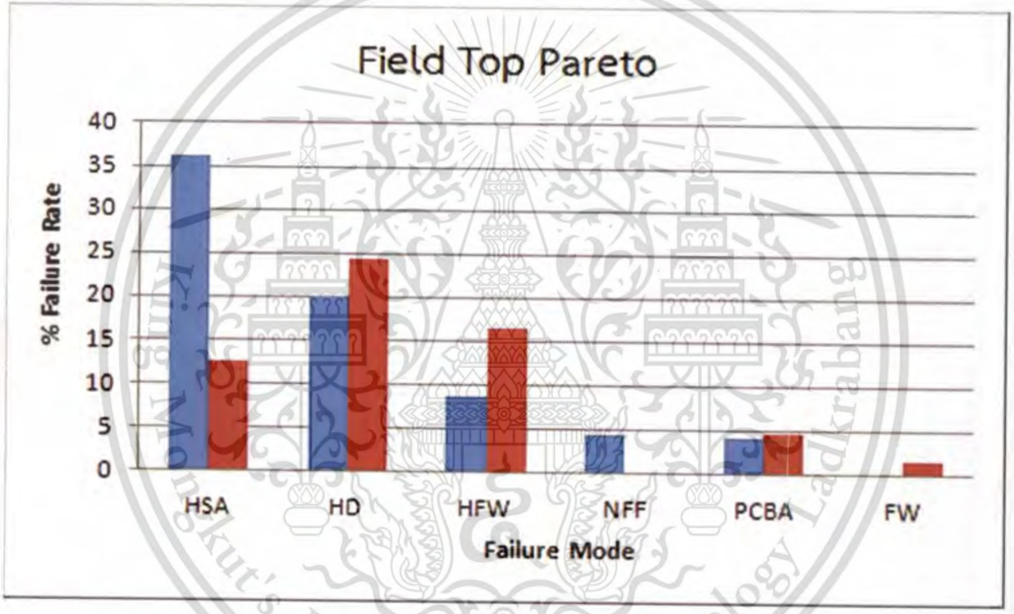
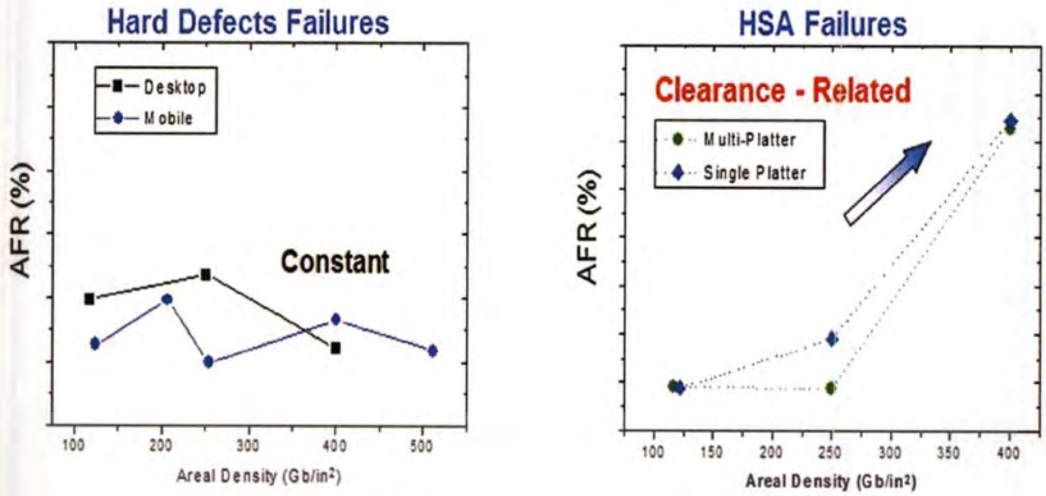


Fig. 2.2 The top failure mode pareto.

Figure 2.2 shows that the hard defect failure is the top failure pareto of field failure in the past HDD generation. Hard Defect is largely driven by contamination / particulates and by head-disk contacts. Constant hard defect failure rate is likely to be the result of extensive particulate reduction activities in the factory and supplier. Nowadays, for modern HDD generation, the failure pareto changes from Hard Defect to Head related failure (HSA). The increase in the total failure rate product over product is dominated by an increase rate of Head related failure such as BLP, OW degradation or etc. As a result, HDD manufactures concentrate on the head degradation over time with reliability failures.



(a) (b)

Fig 2.3: Hard Defect/HSA Versus AFR

2.2 Hard Disk Drives Reliability

Traditionally, the reliability of individual hard disk drives have been quantified by time to failure (TTF) at a specified operating temperature, and specified functional duty cycle. Basically, the reliability test can be defined into two types, destructive test, and parametric test as shown in Fig. 2.4.

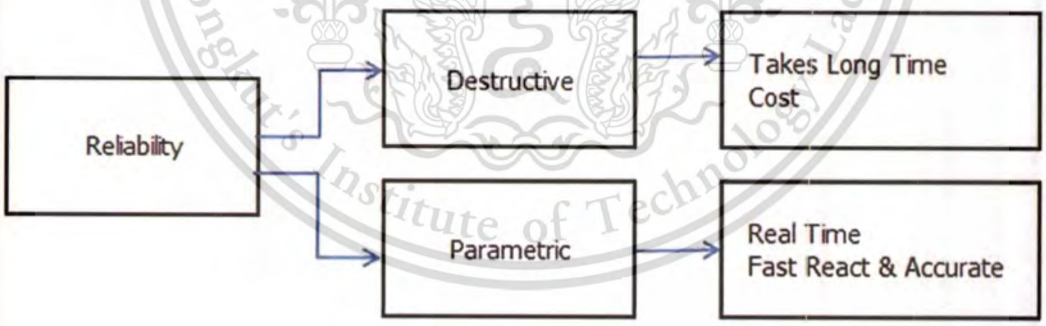


Fig. 2.4 Two types of reliability test.

2.2.1 Destructive test is reliability test in HDD industry that brings the product to be tested until it fails.

2.2.2 Parametric test is reliability demonstration test which is currently used in HDD industry. This test performs with shorter test time by accelerating temperature, the humidity, voltage and the functional test.

The reliability demonstration test is accelerated stress test for the head-disk interface that critically evaluates the robustness of any tribological design before shipment. It was designed for head/media and interface improvement because current product has more complexity and the customer field failure changes. The RDT provides visibility to cover expected drive life. The failure occurs when qualification can be segregated in two groups namely reliability real failure and reliability virtual failures. Reliability real failure is a failure reported which contains a variety of failure modes such as hard defect, particulate failure, head related failure. Nowadays, we collect the parametric data to monitor product reliability performance since the beginning of the test until completion. Any drive that exceeds the critical parameter limit on any head. It is rejected by drives having head exceeding the degradation triggers. As the result, HDD factory uses reliability real failure to predict the AFR and MTTF number.

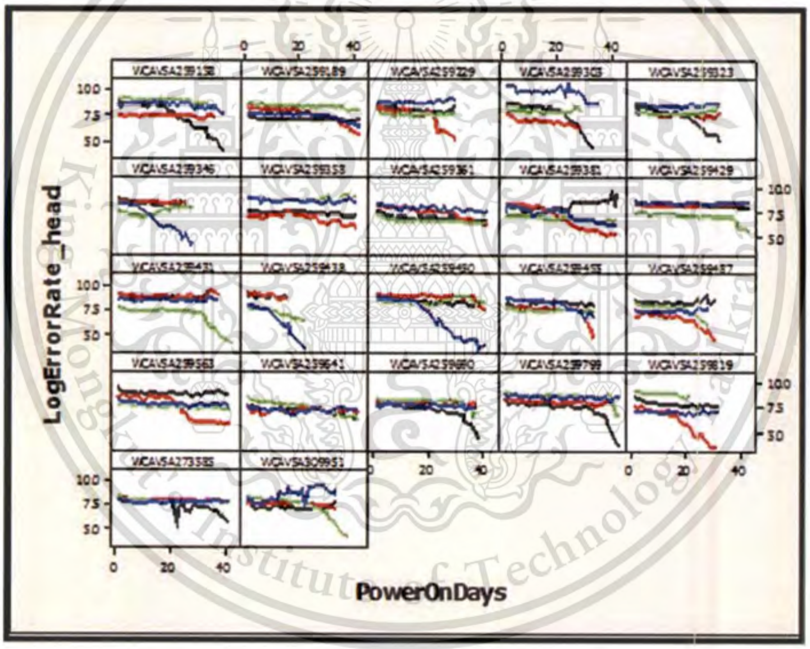


Fig. 2.6 The BER versus power on days

Figure 2.6 shows the relationship of bit error rate degradation over time. This parameter data was collected during the RDT and lead indicator of drive failures. Each single line represents of each head ID and the bit error rate performance will degrade when time is longer.

2.3 Bit Error Rate

In [2], the BER is high-density disk recording which is a subject of intense investigation and identifies the various sources of error. Yeh and Wachenschwanz (1997) studied this problem in systems using sampling detector known as PRML (partial-response maximum-likelihood) detection. They enlarge the concept of signal to noise ratio (SNR) to embrace signal variation arising from linear and nonlinear intersymbol interference (ISI) in pseudorandom sequences, the variance in read back amplitude becomes the sum of the noise variance (power) and the ISI variance. They called this mean-square error (MSE) variance shown in equation 2.1.

$$\sigma_{MSE}^2 = \sigma_{noise}^2 + \sigma_{ISI}^2 \tag{2.1}$$

So with proper treatment, interference is "noise like" [see Williams, p.253, in Arnoldson and Nunnelley (1992). Yeh and Wachenschwanz (1997) discussed estimation of the BER using a complementary error function of the signal to MSE ratio, in which the "signal" is normalized to unity shown in Equation 2.2.

$$BER \sim 0.75 erfc\left(\frac{1}{2\sqrt{2}\sigma_{MSE}^2}\right) \tag{2.2}$$

The analysis of amplitude variation in Viterbi detectors provided a more accurate estimate of BER. The plotted on a log linear basis shown in Fig. 2.8

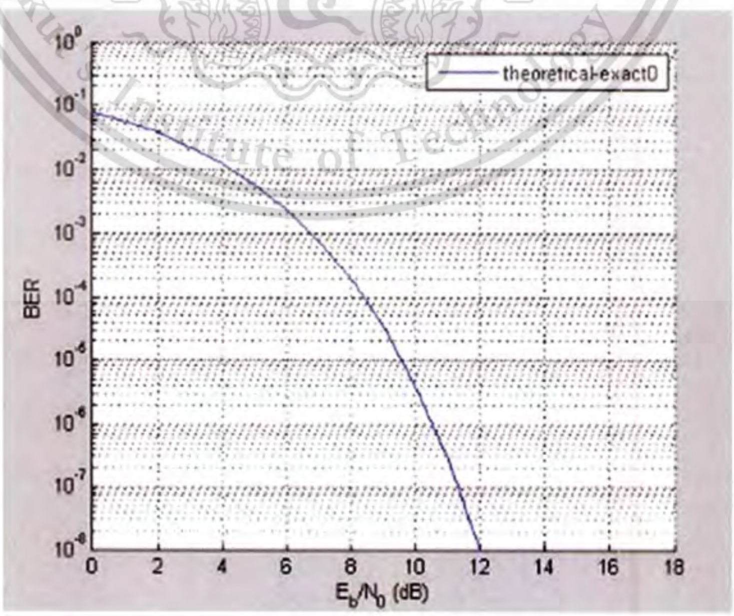


Figure 2.8 Log to base 10 of the BER [3]

The bit error rate (BER) is the number of bit errors divided by the total number of transferred bits during a studied time interval. BER is a unit less performance measure, often represented as a percentage. The BER is considered as an approximate estimate of the bit error probability. The BER as function of the Eb/N0 shown in Equation 2.3.

$$BER = \frac{1}{2} erfc\,(\sqrt{Eb/No}) \tag{2.3}$$

Bit error rate is the test verification parameter that performs at final test process. This station is measured the bit error rate of all zones and collects the results by Pbert per head, per zone and per drive. It was continuous numbers with quantified degradation by the maximum percent change of the worst head. It is critical parameter that indicating to head reliability health in long life service and the parametric degradation reliably proven to be correlated to head failure probability in the field application. The head degradation noticed not only the field but also in reliability testing is driven by contacting with the disk. DFH actuation simply during the transfer of data to/from the disk (reading/writing).

As the failure mechanism changes from the particulate to the head disk interface failure, it is affected to reliability test which has to focus on reliability of head disk interface. The head parameter degradation is clearly sign indicator of failed drives. The drive which has parameter over degradation limits will become real failure subsequently. Based on the extensive studies, one hundred critical parameter both drive and HSA level are collected during hard disk drives reliability demonstration test. There are six critical parameters representing head performance such as Bit error rate (BER), EM (Error Margin), Voltage Gain Amplitude (VGA), Magnetic MRR and Non-repeatable run out (NRRO). In this paper, we choose the bit error rate (BER) parameter for the experiment. Head degradation is recently a key important role used to monitor and predict the future trend. To identify the passer and failed drives, we can consider significant change with regard to increase, decrease and fluctuation trend as shown in Fig. 2.9.

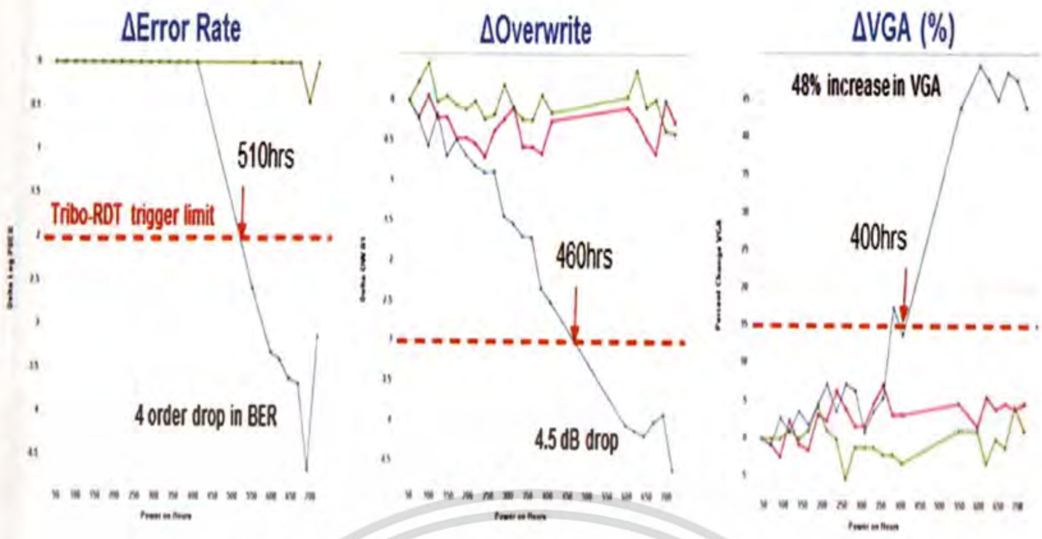


Fig. 2.9 Example of critical head parameter BER, Overwrite and VGA.

A brief of head related failure.

1. BER/EM degradation is error rate/error margin degraded from factory value.
2. VGA degradation is a malfunction/change of the variable gain ampliter in the end unable an accurate gain to the signal read. VGA stands for Variable Gain Ampliter. It is a part of read/write channel optimization process. The core function is to apply a gin to any signal coming out from preamp's low pass filter after reading from magnetic media.
3. Baseline Popping (BLP) is a signal typical of an unstable head which periodically leads to a signal spike in the read-back signal or means characterization by the spurious popping of baseline between readback pulses from magnetoresisitve/TuMR heads.
4. Overwrite (OW) is the write ability of the head to re-write on the exiting signal with no old signal left or it is characteristic of the ability of a writer to write or erase over "old" data.

2.4 Reliability Prediction

To comprehend the concept of this research, there are several papers reviewed that relate to this topic. The paper is organized as follows;

In [2], Jimmy Yang and Feng-Bin Sun presented how Quantum's effort to meet customer requirement. According to a shorter development cycle, a shorter product life cycle in the field and high MTTF goal of 400,000 Hours challenged to HDD industry to achieve on this expectation. This paper is explained from early reliability prediction (ERP), design for reliability, reliability demonstration (RDT) test through mass production including field operations (field performance tracking and evaluation). The field failure rate pattern of Hard disk drives and Quality/Reliability program plan includes AFR and MTTF discussed.

In [4], Eduardo Pinheiro, Wolf-Dietrich Weber and Luiz Andre Barroso presented and analyzed the correlation between failures and several parameters that affected hard disk drive life time based on data gathered from a production internet services deployment. The parameters in SMART are Seek Error, CRC error, Power cycles, Calibration retries, Spin Retries; Power-on hours, Vibration and Temperature studied.

In [5], Bing Ping and team introduced the new algorithm to predict the drive failure based on SMART attributes by using Back propagation neural network. They compared the ANN prediction model between Back propagation (BP) algorithm and Support Vector Machine (SVM) model. A real-word dataset 23,395 drives is used to train and test the predictive model. The experiment result shows that BP neural network is higher prediction accuracy as compared to SVM technique and gives better result than the previous works.

In [6], Zhi-Sheng Ye, Yu Wang investigated Wiener processes with measurement errors. It was compared with the traditional Wiener process with positive drifts compounded with i.i.d. Gaussian noises, and improved its estimation efficiency in comparison with the existing inference procedure.

In [7], Loon Ching Tang and Dong Shang Chang presented a frame work for reliability evaluation from nondestructive accelerated degradation data (NADD) by using power supply units to perform the experiment. The relationship between the parameters and the associated stresses is explored using regression. The failure-time of power-supply units is modeled by the Birnbaum-Saunders distribution, for which the confidence bounds and tolerance limits can be easily obtained. The result shows that it can give better results.

2.4 Intelligent Classification in HDD quality and Reliability

To understand the concept of this research, there are several papers reviewed that relate to this topic. The paper is organized as follows;

In [8], Hirokazu Takahashi and Narade Pecharain presented the reliability of neural network with 3 types of patterns whether the judgment of NN is right or wrong. Researches on pattern recognition using NN have mainly been studied about the speed of training or the recognition rate. This paper pays attention to the hidden layer and extracted output of the units on hidden layer. The results show that the reliability rate of NN is calculated by using 3 types of training data to achieve a high reliability rate.

In [9], Joanne Beechta Dugan and James Watterson presented a simple method to improve reliability by using neural network. In this paper, they addressed the problem of estimating the reliability of neural network. The effectiveness is based on inherent reliability of the component duplicated. This methodology can apply to a variety of neural networks and utilize any several models for component failure mode.

In [10], Meyer, Sebastian; Wohlrabe, Heinz and Wolter, Klasu-Jurgen presented the proposed method to predict the quality and reliability for BGA solder joints by comparing with the conventional statistical technique. The back propagation networks (BPNN) and radial basis function networks (RBFNN) were studied. The results demonstrate that a comparison of two neural networks is better than the BPNN approach.

In [11], Yong Liu, Xin Yao and Testusy Higuchi studied evolutionary ensembles with negative correlation learning (EENCL) to address the issues of automatic determination of individual neural networks (NNs) in an ensemble and the exploitation of the interaction between individual NN design and combination. This paper presented the evolutionary with EENCL for designing NN ensembles automatically based on negative correlation learning and evolutionary learning on the basis of evolutionary programming.

In [12], Huanhuan Chen and Xin Yao analyzed the NCL and pointed out that NCL was prone to over fitting the noise. This paper proposes regularized negative correlation learning (RNCL) algorithm which incorporates an additional regularization term for the ensemble and demonstrates how to decompose the training of ensemble into a set of sub tasks. The result shows that RNCL is as ensemble learning algorithm that addresses drawbacks of NCL and the noise-robustness characteristic of RNCL is important when contaminated with noise.

In [13], Shuo Wang and Xin Yao proposed a new negative correlation learning (NCL) algorithm for classification ensembles which had better performance than the standard AdaBoost and other NCL algorithms. This paper gives insight into AdaBoost. NC algorithm can produce less correlated ensembles and improve classification performance.

In [14], Y. Liu and X.Yao presented learning approach negative correlation learning for neural network ensembles. They used time series for 3 years to study and show why negative correlation learning worked. It was trained simultaneously and interactively through the correlation penalty learns in their error functions. Then, it uses a totally different error function. The experiment result shows that negative correlation learning can produce neural network ensembles with good generation ability.

In [15], Kittiphan Pomdoung and Yuttana Kidjaidure presented the classification techniques to recognize and reduce defects on the media surface in Hard Disk Drive (HDD) manufacturing. This paper presents a statistical feature extraction technique with rank-level fusion based on neural networks to improve recognition and classification of media defects, discriminant function, regression trees and neural network to improve recognition and classification of media defects. The experiment results show that the purpose method is capable of improving the classification performance significantly, which is better than both conventional fusion and individual classifier. Based on the finding, they finally proposed a rank-level Fusion with neural network as a based automatic recognition system to classify the media defect patterns.

In [16], Waraporn Taepin and Yuttana Kidjaidure presented how to predict the customer failure by using neural network rank-level fusion. The goal is to classify the empirical data in two classes namely "Pass"/"Fail" based on customer failure data set that focused on Head Disk I interface group failure mode. Principle component analysis (PCA) is used to eliminate the complexity of input before put through the classifier model. There are four classifier methods namely neural network, discriminant analysis, bayes networks and support vector machine studied. Neural network gives the result better than other classifiers. Therefore, neural network lank-level fusion technique is used for consolidating rank output of each classifier and compared with Borda count and Logistic Regression method. The result shows that neural network rank level fusion gives higher prediction accuracy than other methods

In [17], Minlong Lin, Ke Tang and Xin Yao presented a selective NCL new approach algorithm for incremental learning to address the exiting learning problem. It was fixed by the size of the ensemble to a predefined number. Then, the ensemble was cloned and trained using negative correlation learning when a new data set presents. After that, the new trained is combined with the previous to a larger ensemble. The advantage of proposed method dose not eliminate the previous trained neural networks directly. However, It outperforms the GNCL on accuracy because multiple neural networks are introduced to learn the new data sets and its size does not increase with the new coming data. They compared the exiting two methods namely the Fixed size SNCL (FNCL) and the growing NCL (GNCL) for evaluating the proposed method. The results show that SNCL gives better results than GNCL and FNCL techniques.

In [18], Trupti Pathrabe and Dr.N.G. Bawane presented a paper currency recognition system using ensemble neural network (ENN) technique. It used three characteristics of paper currency as size, color and template. The paper focuses recognition system used to negatively correlate ensemble neural network. When noise pattern is applied, the network will be able to recognize.

In [20], Ahmed Z and team were studied the failure rate prediction for aircraft tires of three De Havilland Dash-8 airplanes using neural network and weibull regression model. Neural network use the feed-forward back-propagation algorithm and the result shows that the failure rate by approach ANN is better than conventional weibull model.

In [21], Di Xu Ph.D, Wenbiao Zhao Ph.D presented a degradation modeling and analysis approach involving multiple degradation measures. Reliability prediction is based on degradation modeling which can be efficient. In this paper, there are two techniques to model and analyze the system. The first proposed method is the correlation between the degradation measure and the failure which use a logistic function to define the probabilistic measure. The second method is a state-space model to describe the evolution of the degradation process by degradation dynamics and random stress effects. Finally, the degradation dynamics are used to predict the reliability functions.

Based on several topics of literature review which contains of reliability, neural network, negative correlation and neural network fusion techniques are discussed, we found a relationship of among technique and can then be applied the develop feature extraction technique to predict the reliability of Hard Disk Drive. We greatly study and adapt those very useful features to our research.

Chapter 3

Feature Extraction

In this chapter, we explained the prediction model with two approaches namely non homogenous and homogenous classifier with negative correlation and neural network fusion. Our two proposed methods are focused on a static model and the prediction model is based on Neural Network (NN) feed forward back propagation algorithm, the various classifications namely Linear Discriminant Analysis (LDA), and K-mean clustering are studied. The experimental procedure can be summarized as follows:

1. Select HDD product for experiment.
2. The quantity of HDD sample has 300 drives built from same the magnetic recording head and the magnetic recording media to reduce the variance of the data sets.
3. Conduct performance of reliability demonstration test and collect the bit error rate data since the beginning of the test to complete.
4. Focus only the head related failures that correlate to bit error rate and exclude those non-head related failures.
5. Use the clementine software to construct the predictive model.

3.1 Experiment Setup

The HDD samples are installed in ASPT chamber. Each drive is connected to a central controller providing both power and data communications; it has applied with accelerating test in temperature / humidity and functional stress level (read/write sequence) as shown in Fig. 3.1.

Fig. 3.1 The Advance Single Plug Tester (ASPT)

The advance single plug tester is used to conduct reliability demonstration test of this research. A total of 300 drives were put in this chamber and kept at temperature ~55 degree C for 30 days to demonstrate the reliability of Hard Disk Drives. The test condition is shown in Table 3.1.

Table 3.1 The test condition

Temperature/Humidity	55°C / 50% RH
Voltage	5 Volts
Period	30 Days
Test Scripts	RDT
Test Sample	300 drives

3.2 Data Collection

The bit error rate data is collected from reliability demonstration test which is a current method used to measure on product reliability; the data set is kept in central database and it is refreshing until the test is completed in daily basis.

3.3 Empiral Data / Pre-screening

A total of 300 samples with 7,800 records of bit error rate are used in this paper. This parametric data is collected by drives, heads and test loops. The samples are selected from reliability demonstration test. Before putting the empirical data to system, outlier noise is eliminated to reduce the complexity and variance of the input. The input data is 1x26 matrixes as shown in Fig. 3.2.

Fig. 3.2 The diagram of pre-processing.

Figure 3.3 shows the bit error rate patterns of all drives tested which contain pass and fail drives. The x-axis represents of test loop per day (time), the Y-axis is the bit error rate degradation with unit in decibel (dB). The multiple drives are tested and shown in different colours where a single colour represents a head data point. We can clearly see that as the test duration increases, the level of the BER gets lower. The BER degradation of observation samples defined in vectors Y is BER delta (Min -5.5 and Max 1.5, resolution 0.5) and X represent of test loop per day (measure BER degrade in daily)

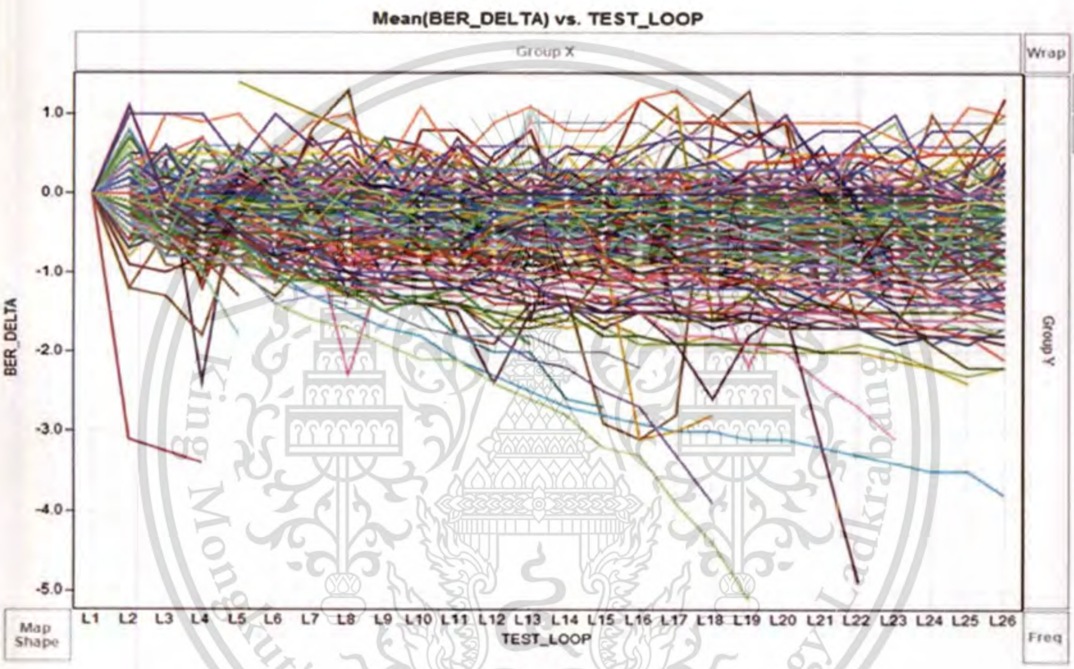


Fig. 3.3 The BER degradation pattern.

The bit error rate pattern usually provides an evidence to classify between pass and fail drives as shown in Fig. 3.4. In this study, a total of 268 drives representing of passer drives are discussed here. Figure 3.4 shows that the bit error rate trend is getting lower when time increases and the bit error rate is within threshold limit number (>3.5dB) that HDD factory is used to justify whether this drive is considered pass or failed drive.

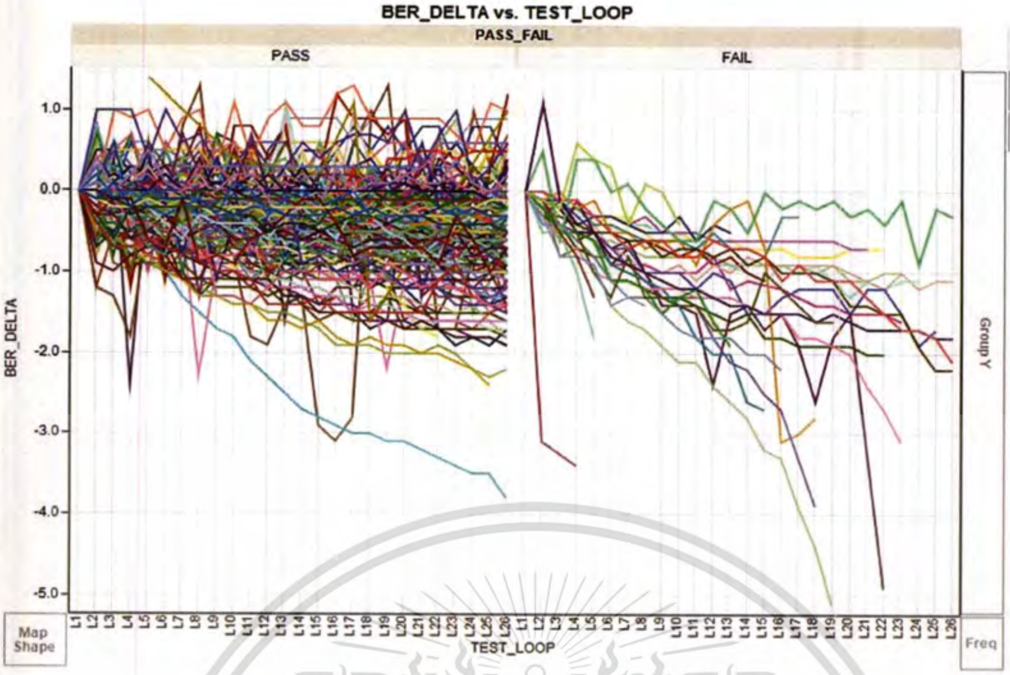


Fig. 3.4 The BER degradation pattern of Pass/Fail sample

The predictive model is learnt from this bit error rate pattern historical and adaptive weight when the data set input changes automatically. To better understand the nature of empirical data. Figure 3.5 shows the bit error rate pattern comparison between Pass / Fail group in average. The average BER degradation for Pass and Fail shows clearly separation. Then, the classification method namely Neural Network, K-Mean and LDA studied.

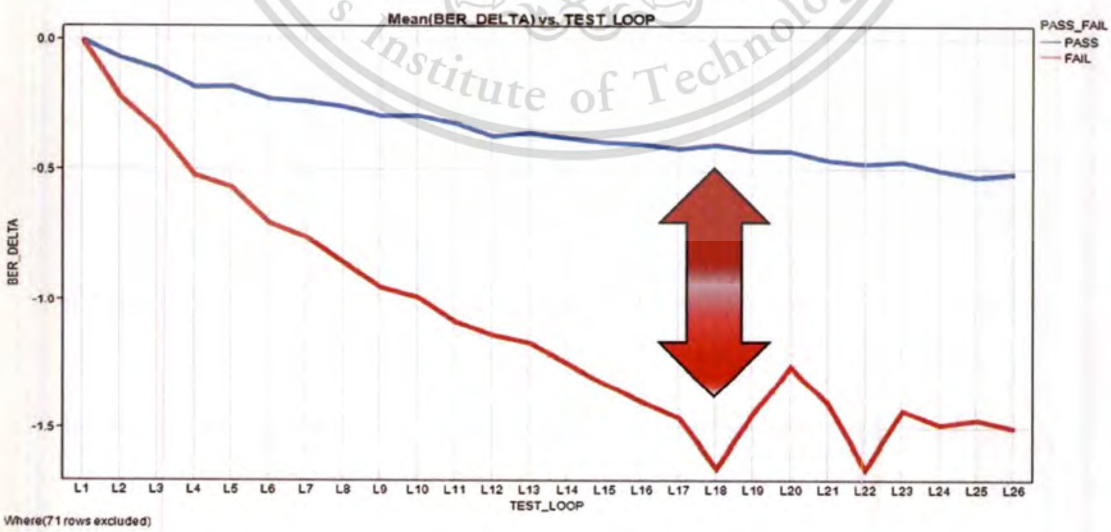


Fig 3.5 The average BER pattern of Pass/Fail group

The gap between pass and fail groups can be clearly separated and the classification method is used to classify this empirical data. Apart from this, a total of 32 drives with head related failures are described. All of them can be segregated by the time to failure (TTF) in three groups as shown in Fig. 3.6. The time to failure is used as the life measure for unrepairable items. We can see that some drives do not have the bit error rate full window because drives are taken out of the test due to several reasons.

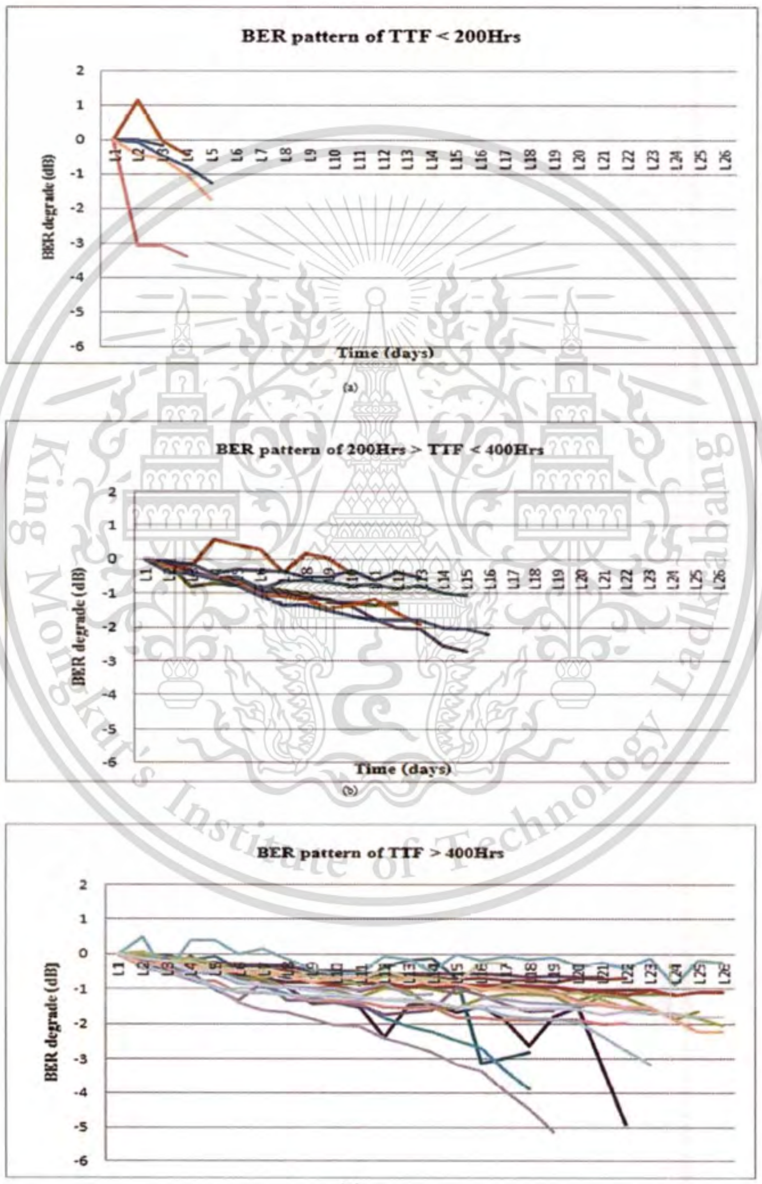


Fig. 3.6. Three types of failure pattern (a) TTF < 200 Hrs (b) TTF < 400Hrs, (c) TTF < 678Hrs.

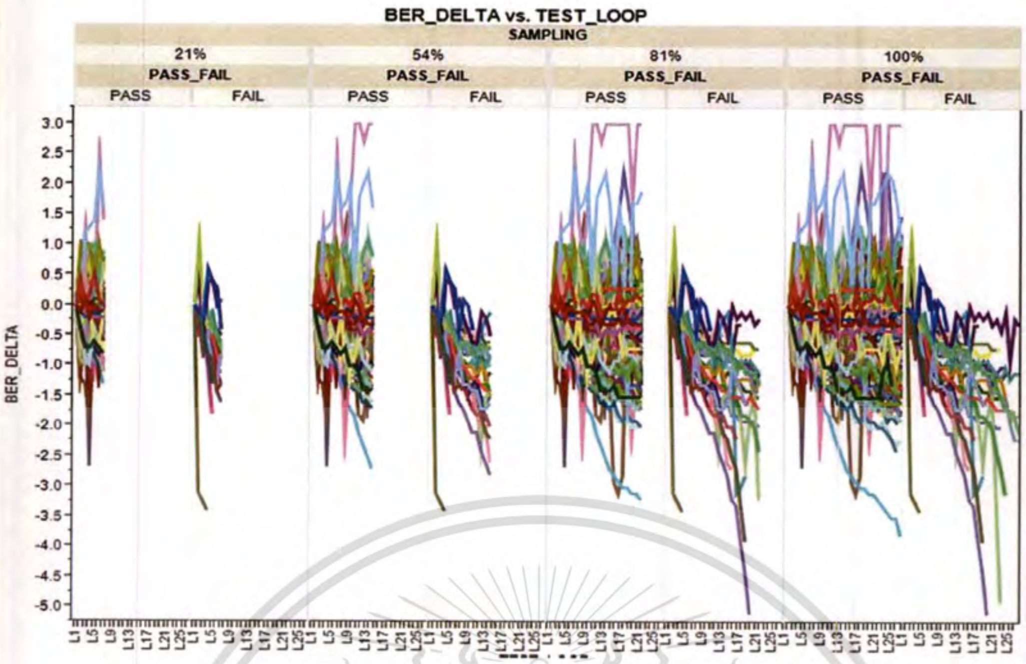


Fig. 3.7 The BER pattern of various input samples.

Figure 3.7 shows that the bit error rate degradation pattern of various input range is divided in four periods as 21%, 54%, 84% and 100% respectively. Each input contains pass and fail results in each input range. To reduce the dimension and observe the separation clearly between Pass and Fail, the compare the classification accuracy subsequently versus full range of window as shown in Fig. 3.8

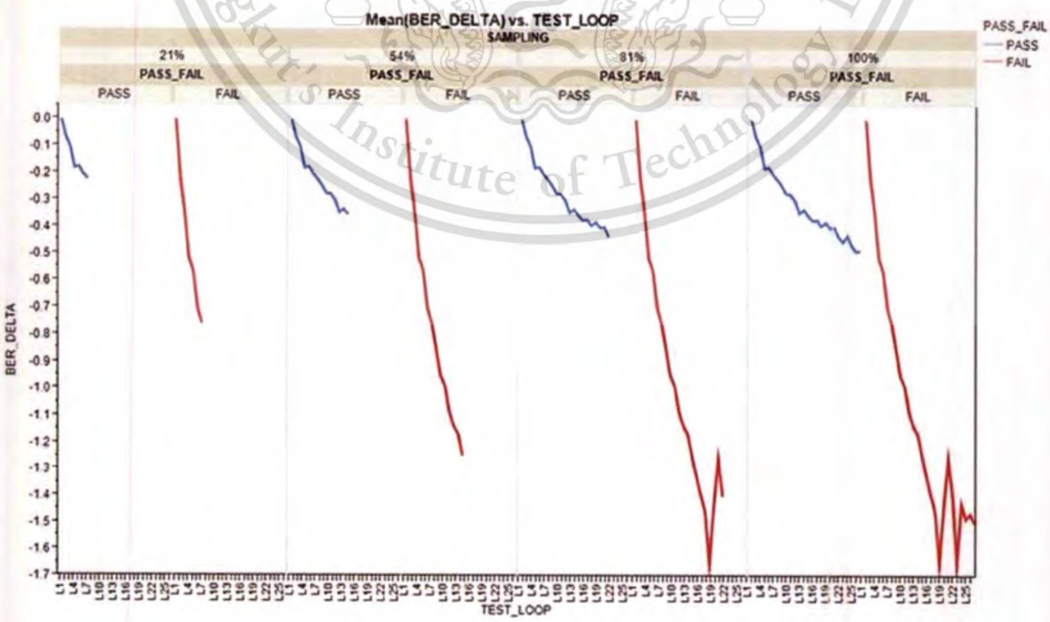


Fig. 3.8 The average of BER pattern of various input samples.

3.4 Intelligent Classification

Empirical data is then put through the two different classifier models, namely Homogeneous and Non Homogeneous Classifier models. The Non homogeneous classification model contains three different classification methods namely Neural Network, Linear Discriminant Analysis (DA) and K means clusturing. NN and LDA are supervised learning algorithms whereas K-mean clustering is unsupervised learning algorithm. For Homogenous model, it contains with 10 pull neural networks with supervised learning algorithms.

3.4.1 Neural Network (NN)

Neural Networks (NN) [8] is an intelligent system that is can recognize the pattern of data sets based on historical inputs. It is widely utilized and well-received in intelligent system to discover the relationship between the numerical. Neural network is mathematical algorithm used to reckon the complexity information by connection list technique. It is simulated like human brain comprising of input and output required to adjust the weight of each input so as to enhance the pattern recognized. The advantage of neural network can discover the knowledge and relationship among information obtained without using mathematical techniques. The concept of NN is adjusting the weight of each input in hidden layers by giving the highest score to that important. Then the output is compared with threshold limits. If the summation is more than threshold level, neuron will send the output. Conversely, if the summation is less than threshold, neuron will not trigger the output. To improve the prediction model, we utilize back propagation algorithm (BP) to better weight after training them. The weight of each node will be obtained from input multiplying weight and then passing through each hidden layer. If the summation of the result in each node exceeds the threshold level, then neural network will send out to output. If the number is less than threshold level, then neural network does not send out the output. So as to let neural network memorize the pattern, we have to supervise them to learn and to segregate the pattern. This is the advantage of neural network that can be adapted to learn new things and resolve problems which humans are not able to address. In this paper, we focus on feed forward back propagation algorithm and set up the NN by using pruning method starting with a large network and removing the weakest units in the hidden and input layers as training proceeds.

3.4.1.1 A Feed-Forward Neural Network

Artificial Neural Networks are mathematical algorithms that can learn mappings between input and output states through supervised learning, or cluster incoming information in an unsupervised manner. A feed-forward neural network composes of input, middle and output layers. Each layer will be connected to each other and send out the output if the summation is more than threshold limit as shown in Fig. 3.9.

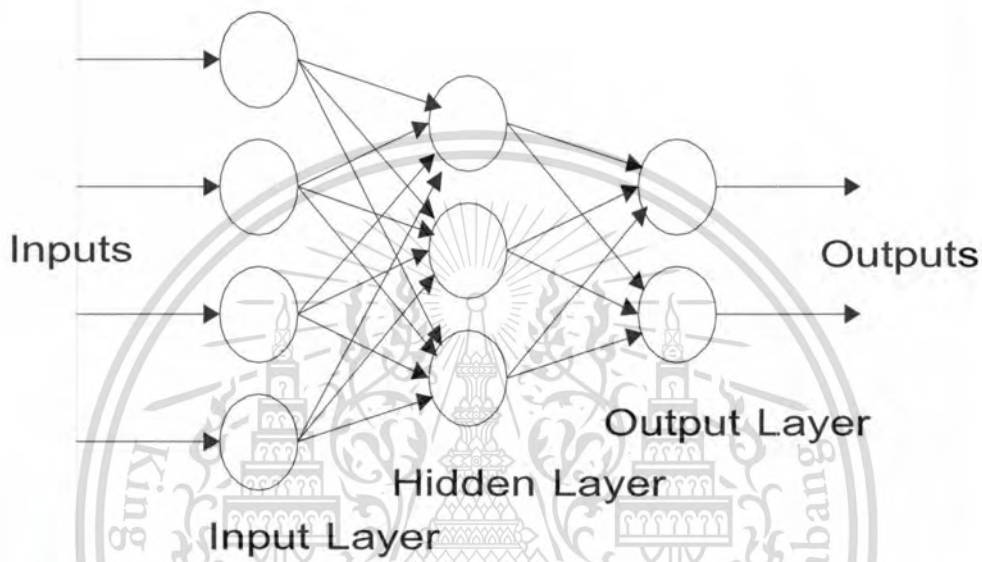


Fig. 3.9 The basis architecture of feed forward neural network.

3.4.1.2 Back-propagation algorithms

Multilayer perceptron has been efficiently applied to solve some difficult various problems by training them in a supervised manner with a very popular algorithm known as the error back-propagation algorithm. This algorithm is on the basis of the error-correction learning rule. Fundamentally, the error back-propagation process comprises of two passes through the different layers of the network:a forward pass and a backward pass. In the forward pass, activity pattern (input vector) is applied to the sensory nodes of the network and propagates have an effect on the network layer by layer. Eventually, a set of outputs is produced as the actual response of the network. During the forward pass the synaptic weights of network are all fixed. During the backward pass, in contrast, the synaptic weights are all adjusted in accordance with the error-correction rule. Particularly, the actual response of the network is subtracted from a desired (target) response to yield an error signal. This error signal is then propagated backward through the network, against direction of synaptic connections hence the name "error back-propagation". The synaptic weights are adjusted to render the actual response of the network move closer the desired response.

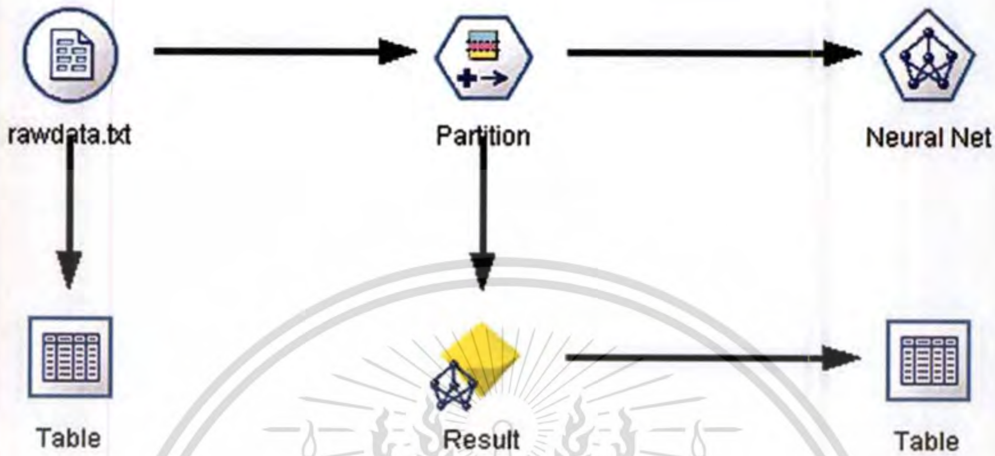


Fig. 3.10 A stream designed of neural networks.

Figure 3.10 shows the classifier model with neural network; staring with the source node (raw data) consists of bit error rate data of 300 samples which contains 7,800 records. A partition node is used to divide the data sets into two parts as training and testing. We have randomly divided 60% allocation for training data sets and 40% allocation for testing sets. A neural networks using prune method is used to produce the prediction model and then we can obtain the prediction results shown in Chapter 4. The outputs of this model will act as the input of negative correlation (NC) and neural network fusions next.

3.4.2 Linear Discriminant Analysis

Linear Discriminant Analysis (LDA) is classified by constructing a predictive model based on linear combination of predictive factor with priority assigned and it segregates the datasets similar to each other. It can be used to grouping nearly the similar data and distance which are used to judge whether it is the same group or not as shown in Fig. 3.11.

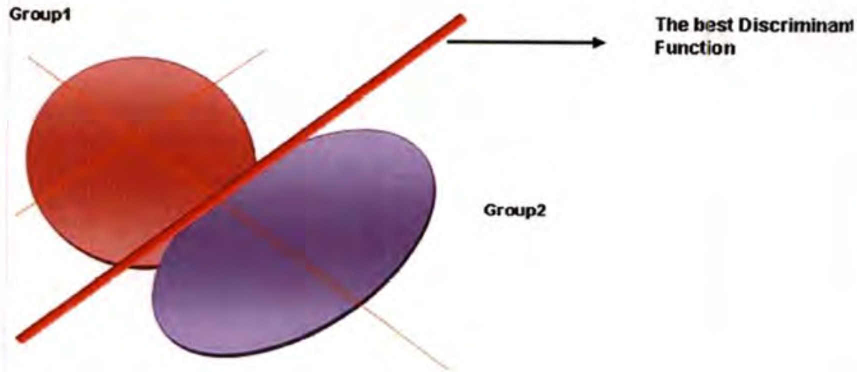


Fig. 3.11 Linear Discriminant Analysis.

Linear Discriminant Analysis (LDA) is used to find a linear combination of features which characterizes or separates two or more classes of objects or events. The resulting combination may be used as a linear classifier, or more commonly, for dimensionality reduction before later classification.LDA is used in many fields such as face recognition, marketing or product management.

In face recognition, each face is represented by the large number of pixel values. Linear discriminant analysis is primarily used to decrease a number of features to a more manageable number before classification.

In marketing, discriminant analysis was often used to determine the variables distinguishing different types of customers and/or products based on surveys or other forms of collected data.

LDA involves the determination of a linear equation like regression that will predict which groups of the case belong to. The formula is shown in equation 3.1.

$$D = v_1X_1 + v_2X_2 + v_3X_3 \ldots\ldots\ldots\ldots\ldots\ldots\ldots v_iX_i + a \qquad (3.1)$$

Where D = Discriminate function

v = The discriminant coefficient or weight for that variable

X = Respondent's score for that variable

a = A constant

i = The number of predictor variables

The function is like a regression equation. There are unstandardized discriminant coefficients analogous to the regression equation. These are maximizing the distance between the means of the criterion (dependent) factors. Standardized discriminant coefficients can also be used in common with beta weight in regression. Good predictors are likely to have large weights. What you want this function to do is to maximize the distance between the categories and come up with an equation that has strong discriminatory power between groups. After using an existing set of data to reckon the discriminant function and classify cases, the new cases can then be classified. The number of discriminant functions is one less than the number of groups. There is merely one function for the basic two group discriminant analysis. There are various purposes of DA.

1. To investigate differences between groups based on the attributes of the cases, indicating which attributes contribute to group separation. The descriptive technique successively identifies the linear combination of attributes known as canonical discriminant functions (equations) contributing maximally to group separation.
2. Predictive DA addresses the question of how to assign new cases to groups. The DA function uses a person's scores on the predictor factor to forecast the category to which the individual belongs.
3. To determine the most parsimonious way to differentiate between groups.
4. To classify cases into groups. Statistical significance tests using chi square help you observe how well the function separates the groups.
5. To test the theory whether cases are classified as forecasted or not.

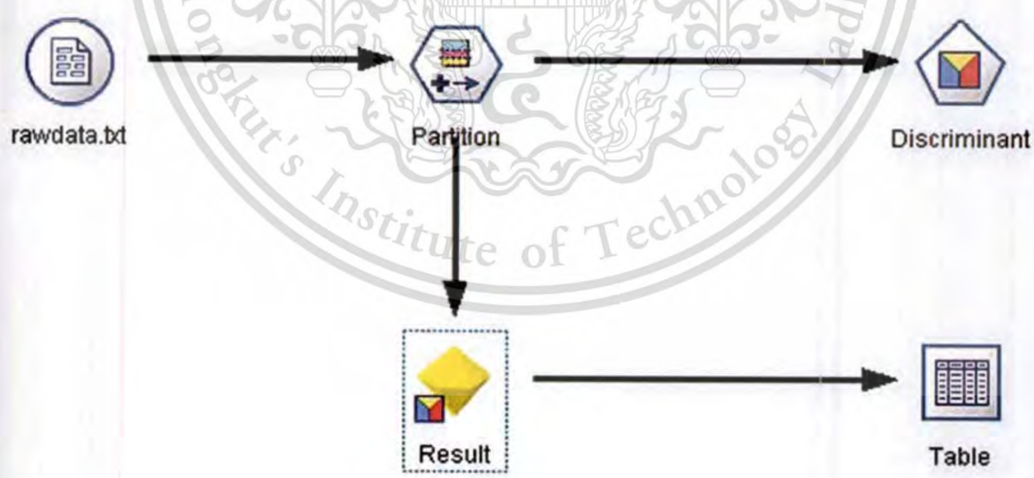


Fig 3.12. A stream designed of Linear discriminant analysis

Figure 3.12, represents a stream designed for classifier with discriminant analysis. The source node comprises of bit error rate degradation data of 300 samples which contains test loops, serial number and the test result. A partition node is classified into two parts as training and testing. 60% allocation is given to training data sets and 40% allocation is for testing data sets. A discriminant analysis utilizes stepwise model to generate the classifier model which we can get the prediction test result as shown in Chapter 4. The output of this section will be the input of negative correlation used to get rid of insignificant factors.

3.4.3 K-Mean Clustering

The k-means is a clustering method, used to group records on the basis of similarity of values for a set of input fields. The primary idea is to try to discover k clusters, such that the records within each cluster is similar to each other and distinct from records in other clusters as shown in Fig. 3.13. K-means is an iterative algorithm. An original set of clusters is defined and the clusters are repeatedly updated until more improvement is impossible (or a number of iterations exceed a specified limit).

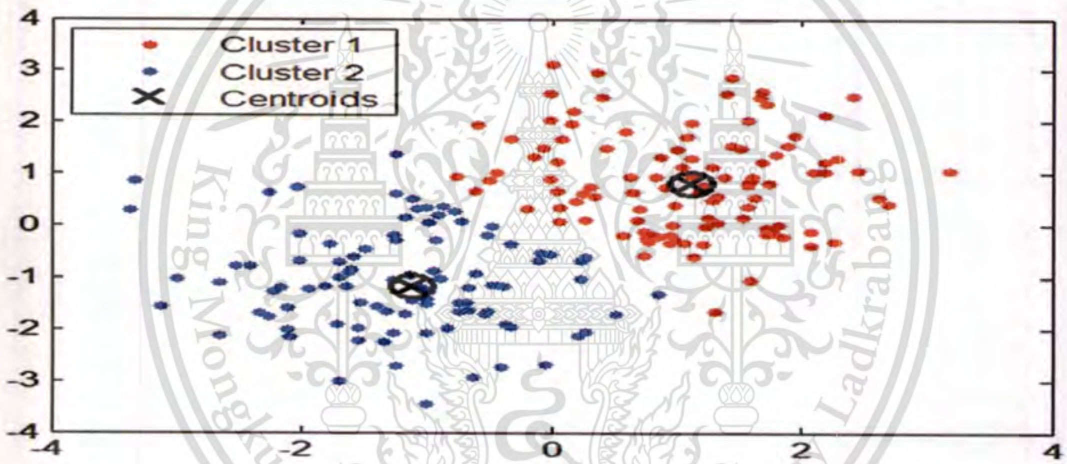


Fig. 3.13 K-mean clustering.

In building the k-means model, input fields are encoded to gather differences in measurement scale and type, and the clusters are defined and updated to yield the final model. In most datasets, a great deal of variability occurs in the scale of range fields. Euclidian distance is technique to check the distant between groups from centroid points. For example, consider age and number of cars per household. Depending on the population of interest, age may take values up to 80 or even higher. Values for number of cars per household, however, are unlikely to exceed three or four in the vast majority of cases. If you use both of these fields in their natural scale as inputs for a model, the age field is likely to be given much more weight in the model than number of cars per household, simply because the values (and therefore the differences between records) for the former are so much larger than for the latter.

To compensate for this effect of scale, range fields are transformed so that they all have the same scale. In Clementine, range fields are rescaled to have values between 0 and 1. The transformation used is where x'i is the rescaled value of input field x for record i, xi is the original value of x for record i, xmin is the minimum value of x for all records, and xmax is the maximum value of x for all records.

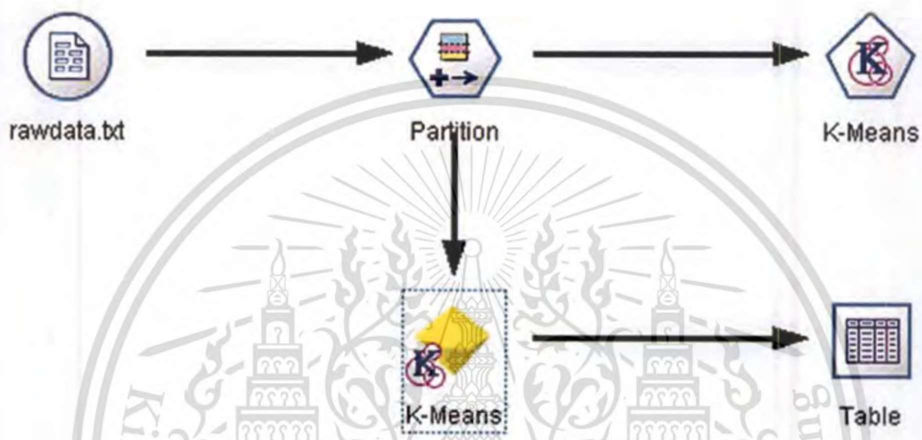


Fig 3.14. A stream designed of K-mean

Clementine software is used to construct the classifier model as shown in Fig. 3.14, the source node makes up of bit error rate degradation data of 300 samples, containing test loops, serial number and the test results. A partition node is used to divide the data sets into two parts as training and testing. Training data gets 60% allocation while training set gets 40% allocation. K-mean clustering is used to classify the datasets to Pass and Fail class. However, it is needed to determine the number of the K-mean clustering at initial stage. The output of this classifier model will be the input of negative correlation further.

3.5 Negative Correlation (NC)

Negative correlation learning is a successful approach to designing neural network (NN) ensembles. NCL has shown a number of successes on various applications, including classification problems, regression problems and time-series prediction. It has consistently demonstrated very competitive results with other ensemble learning techniques such as mixtures of experts, bagging, and boosting. There a sonis, when using NCL to train an NN ensemble, the individual NN softhe ensemble are highly different among one. Another and at the same time have high accuracy. So, when new Training data are presented, the diversity among the individuals makes the madapt to new data in different ways. Such difference may help NN better adapt to the new data, and there by enhance the overall performance. We suppose a training set and then set the classifier method whose output is simple averaging of outputs of the set of neural networks from equation 3.2.

$$\mathrm{F}(n) = \frac{1}{\mathrm{M}}\sum_{\mathrm{i}=1}^{\mathrm{M}} \mathrm{F}i(n) \tag{3.2}$$

Where

$Fi(n)$ is the output of classifier

i is on the nth class

$\mathrm{F}(n)$ is the output on the n class.

M is the number of the individual classifier

NC employs the standard back-propagation (BP) algorithm to train the individual NNs in parallel. The key to the success of NCL is the usage of a different error function. It is well known that for an ensemble to generalize well, the individual NNs of it should be both accurate and diverse. To encourage these, NCL uses the sum of the mean squared error (MSE) and a penalty term as the error function during the learning procedure. When the nth training pattern is presented, the ith NN is trained to minimize the error function

The penalty function pi may have the form:

$$Pi(n) = (Fi(n) - F(n))\frac{1}{N}\sum_{j\neq 1}(Fj(n) - F(n)) \tag{3.3}$$

We adjust the number of neurons in the hidden layer and the output layer of neural networks and use the diverse network structures and algorithms to analyze and modify the results. Accordingly, an optimum prediction system is achieved.

3.6 Neural Network Fusion

Anil K. Jain [16] gave a detailed discussion of classifier combination in their review. There were many reasons as follows: (a) the use of different classifiers with a different context of the same problem; (b) different classifiers trained on the same data may produce strong local differences besides their global performance and (c) Neural Networks might show different results with different initializations and therefore one could compound the results to exploit the attempts to learn from the data.

Data fusion provides a theoretical, computational, and implementation framework for combining data and knowledge from different sources with the aim of maximizing the useful information content. In this way, reliability and discrimination capability are improved while the amount of required data is minimized. Through the three overlapping stages: preprocessing, data alignment, and decision making, the performance of a system is improved. Data fusion spans disciplines such as signal detection, pattern recognition, and tracking, with applications in domains such as military, robotics, medicine, and space research.

In pattern recognition, where the use of multiple classifiers and combiners are common, the large number of experimental studies has shown that classifier combination is able to enhance the recognition accuracy. Tumer and Ghosh proposed a quantitative analysis of the improvements in classification accuracy by combining multiple neural networks. They showed that combining networks using a linear combiner or order statistics combiner significantly lowered the variance of the actual decision boundaries around the optimum boundary. Neural Network Fusion was applied in data mining [13] and it helped increase the prediction accuracy to higher level after negative correlation. The adaptive weight of each classifier input is generated by using the concept of neural network fusion for calculating the final ranking. The target of Rank-level Fusion is to consolidate the rank output by individual subsystems so as to obtain a consensus rank for each identification.

In this study we use Neural Network Fusion applied in our model, this method helped increase the prediction accuracy to higher level after negative correlation (NC). The adaptive weight of each classifier input is generated by using the concept of neural network fusion (NN Fusion) for calculating the final ranking. The target of Rank-level Fusion is to consolidate the rank output by individual subsystems so as to obtain a consensus rank for each identification. We considered the classifier combination schemes that dealt with rank and/or confidence. At the confidence level, a classifier outputs a numerical value for each class indicating the probability that the particular input pattern depends on that class. At the rank level, a classifier assigns a rank to each class with the highest rank which is the first choice. Rank value is not allowed to be used in isolation as the highest rank does not necessarily mean a high confidence in the classification.

3.7 Proposed Methodology

As empirical data is complicate and very noisy. It very difficult and takes longer time to classify the data via traditional method. Based on several literature and related works, we found that the intelligent classification could classify this data and able predict the recording head reliability degradation. Neural network and linear discriminant analysis are supervised learning and K-mean clustering is un-supervised learning algorithms used to compare the prediction the results. To reduce the complexity of the classifier model, we apply the neural network fusion and negative correlation to improve and eliminate the insignificant classifier input. Next section will describe the more details.

3.7.1 Non Homogeneous Classifier with Negative Correlation on NN Fusion.

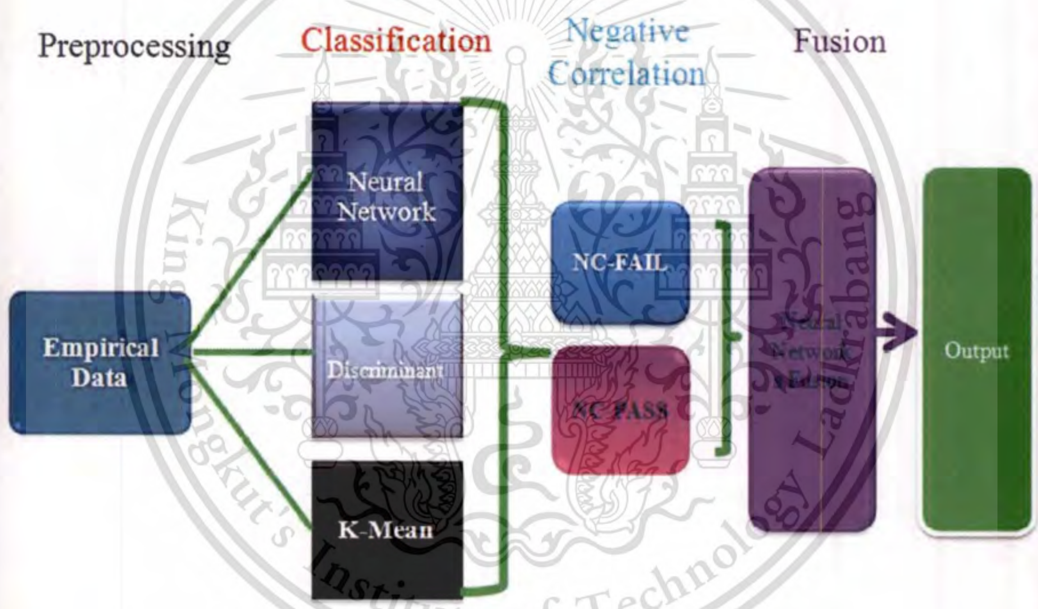


Fig. 3.15 Non Homogeneous classifier model flow incorporating NC and NN Fusion.

Figure 3.15 shows the NN fusion model for Non Homogeneous classifier containing 26 inputs, 1 hidden layer with 18 neurons and has 2 outputs which are passing and failing at the output layer. In this arrangement, empirical data is presented to three classifiers, namely NN, LDA, and K-Mean. Each classifier independently achieves a model for a two-class output, namely "PASS" and "FAIL". The best classifier for the "PASS" and the "FAIL" classes is separately selected via NC. In this way, the most optimum classifier required for each of the two output classes ("PASS" and "FAIL") derives from the set of empirical data presented. The results are then aggregated via NN fusion to yield the predictive model for the empirical data presented.

3.7.2 Homogenous NN Classifier with Negative Correlation on NN Fusion

The Homogeneous classification model contains the pool of ten Neural Networks. The purpose of this classification model is to classify the empirical data into two classes namely, "PASS" and "FAIL", according to the results derived from reliability testing of the Hard disk drives (HDD) as shown in Fig. 3.16.

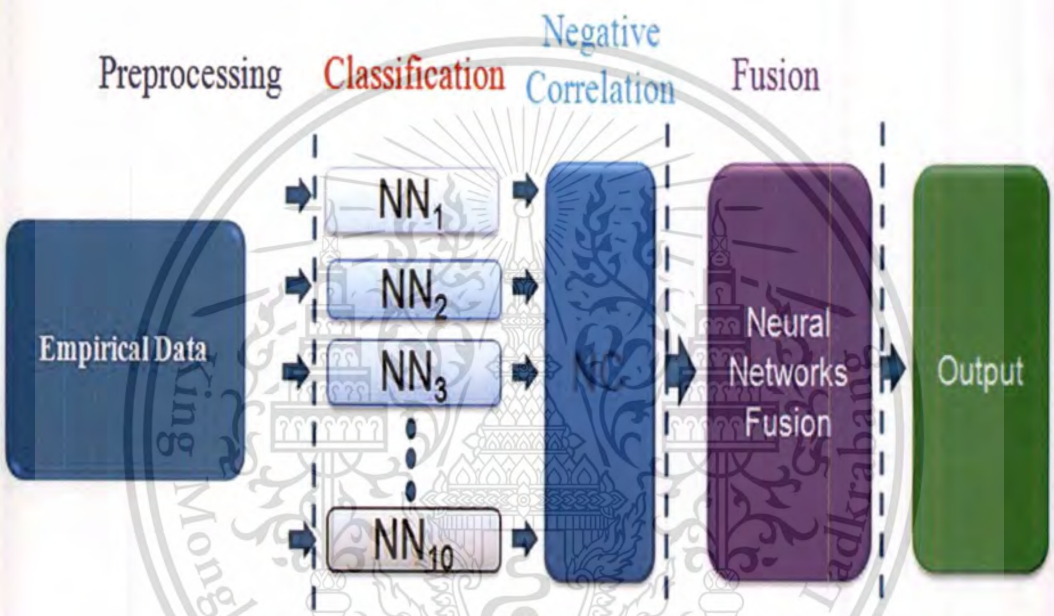


Fig. 3.16 Homogenous classifier model flow Incorporating NC and NN Fusion.

First, the NN fusion model for Homogeneous classifier contains 26 inputs, 1 hidden layer with 18 neurons and has 2 outputs which are passing and failing at the output layer.In this arrangement, empirical data is presented to ten pull of Neural Networks (NN). There is only one time series input parameter applying to the NN classifiers which is the bit error rate degradation data. In each network, the training / testing data sets are forced to be re-established but with the same 60- 40 ratio between training and testing. Through the ten iterations, ten classification models will be achieved. The outputs of each NN are two-class models, namely "PASS" and "FAIL". The best classifiers for "PASS" and "FAIL" are then obtained separately via NC. The final result is then aggregated via NN fusion to derive the optimum predictive model for the empirical data presented.

Table 3.2 Neural Network parameters

Network parameters	Details
Training algorithm	Back Propagation
Input data	26 inputs, 1 hidden layer and 2 output
Weights	[-1,1]
Activation function	Sigmoid function
Output data	2=Failed, 1=Passed
Prediction	Percentage of predicted accuracy of pass/fail

3.8 Various Input Range Sampling

Since we are able to classify the bit error rate degradation pattern to "Pass "and "fail" class by non-homogenous and homogenous classifier with neural network fusion with negative correlation techniques, we will conduct the experiment in the next step by applying with different various input ranges. The input range is dived into 4 periods as 27%, 54%, 84% and 100%. The objective of this work is to see how far to forecast the result. The benefit we gain is the reduction in test time in reliability demonstration test.

Chapter 4
Experimental result

In this chapter, we described the experiments results of Non-homogenous and Homogenous classifier with negative correlation on neural network fusions and the various input sampling test results.

4.1 Empirical data / Pre–Processing

After studying the critical parameter that related to head reliability performance of Hard disk drives (HDD), the empirical data bit error rate are gathered in time series since starting until completing the full test. Due to large variation of the failure mode, we picked up simply the head failure that correlates to bit error rate degradation performance and excludes those non-head related. Prior to inputting the system, we need to normalize the data sets to eliminate the null and data missing. In this study, a total of 300 samples (90% of samples were the "PASS" class and 10% of samples were the "FAIL" class) were picked from reliability demonstration test with a homogenous set. The total samples were divided as by random. The 60% and 40% of samples were the training and testing & validation sets respectively.

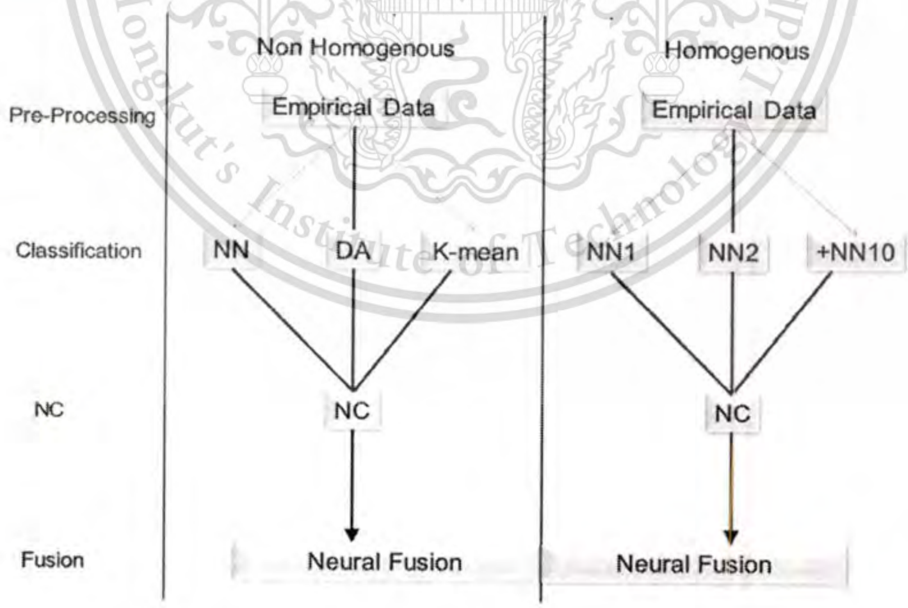


Fig. 4.1 An overview of proposed methods for Non-Homogenous and Homogenous classifier using neural network fusion with negative correlation.

Figure 4.1 shows the architecture of Non-homogenous and Homogenous classifier with Negative correlation on neural networks fusion. It consists of four process steps namely preprocessing, classification, negative correlation and neural network fusion. We compared the prediction results of both proposed methods and validation with various input ranges to optimize the best prediction accuracy of test time reduction purpose.

4.2 Non Homogenous NN Classifier and Negative Correlation Result

The first proposed method is non-homogenous classifier with negative correlation and neural network fusion. The empirical data will be used in the same datasets as the input of three classifier categories namely Neural Network, Discriminant analysis and K-mean clustering method as shown in Figure 4.2. The model contains 26 inputs, 1 hidden layer with 18 neurons and has 2 outputs which are passing and failing at the output layer.

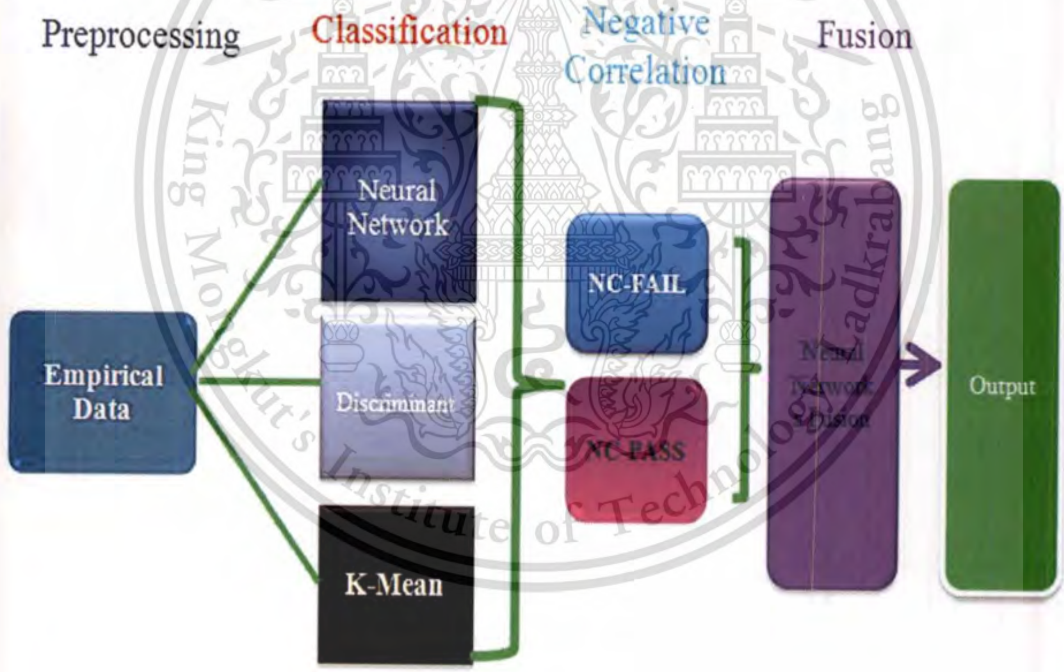


Fig. 4.2 Non-Homogenous with Negative Correlations on NN fusion diagram

In this paper, we use clementine software to construct the prediction model. It provides users with powerful predictive model that can be quickly developed. The predictive model is constructed as shown in Figure 3.9, 3.11, and 3.13. Therefore, we can obtain the prediction results of each individual classifier for non-homogenous classifier as shown in Table 4.1.

Table 4.1 Results from Non Homogenous Classifier

Classifier	Actual	Predict	Testing Accuracy
NN	PASS	TRUE	99.60%
		False (Type I error)	0.400%
	FAIL	False (Type II error)	53.100%
		TRUE	46.90%
	CUM	***Pass/Fail***	***94.00%***
DA	PASS	TRUE	88.10%
		False (Type I error)	11.90%
	FAIL	False (Type II error)	81.20%
		TRUE	18.80%
	CUM	***Pass/Fail***	***80.70%***
K-Mean	PASS	TRUE	70.90%
		False (Type I error)	29.10%
	FAIL	False (Type II error)	25.00%
		TRUE	75.00%
	CUM	***Pass/Fail***	***80.70%***

From Table 4.1, the neural network is given prediction result "Fail" 46.90%" and "Pass" 99.60% which gives the highest prediction accuracy in "Pass" portion. K-mean is given prediction result "Pass" 70.90% and "Fail" 75.00% which is the highest prediction accuracy in "Failed" group. Linear Discriminant Analysis (LDA) is given prediction result "Pass" 88.10% and "Fail" 18.80%" which is lower result as compared to previous two classifier techniques. Afterwards, Negative correlation (NC) is introduced to eliminate the complexity of data sets. In this study, we reduced the complexity of system from three to two methods based on negative correlation technique as shown in Table 4.2.

Table 4.2 Results from Non Homogenous Classifier and Negative Correlation

Model	NC Value of Pass Group	NC Value of Fail Group
Neural Network	-3878.013	-1493.9142
Discriminate Analysis	-13231.42	-1002.8491
K-Means	-2783.006	-4944.759

The table 4.2 shows the comparison of NC value results of Pass/Fail between supervised and unsupervised learning groups with three models namely neural network, discriminant analysis and K-means. The result from Negative Correlation (NC) shows that neural network is given the better result than the LDA's and K-mean's for pass group of the former column, and the K-means is given better result than the LDA's and Neural Network's for fail group.

4.3 Homogenous NN Classifier and Negative Correlation Result

The second proposed method is Homogenous NN classifier and negative correlation as shown in Figure 4.3 and described in Chapter 3, the model is constructed with ten pull neural networks using the same data set in order to select whether NN has given the high prediction accuracy and the prediction result is provided in table 4.3

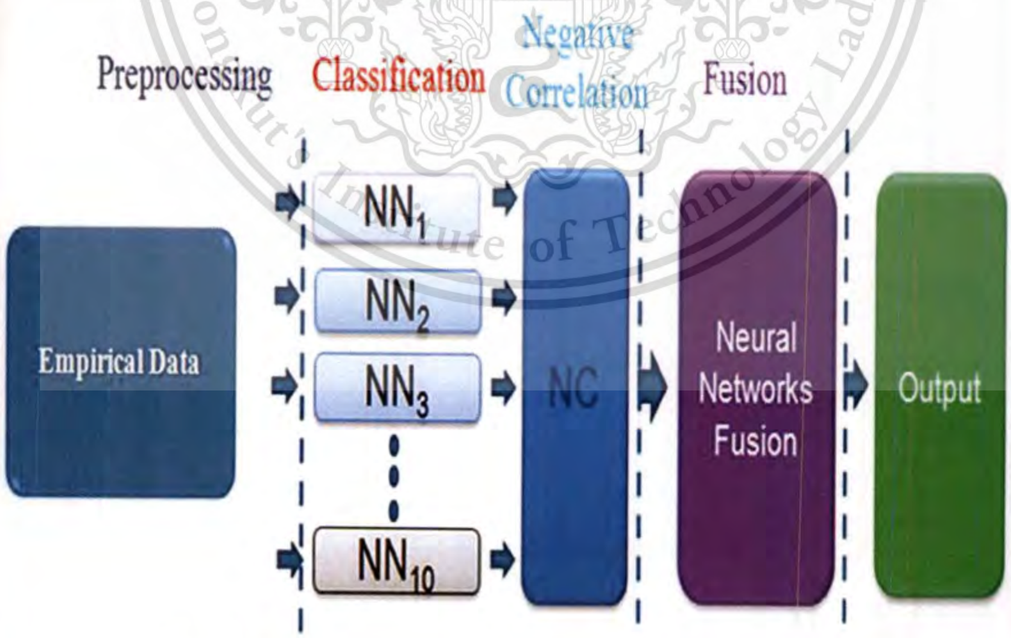


Fig. 4.3. Homogenous NN classifier and negative correlation.

In this arrangement, empirical data is presented to ten neural networks (NN). In each network, the training/testing data sets are forced to be re-established but with the same 60-40 ratio between training and testing. Through the ten interactions, ten classification models will be achieved. The outputs of each NN are a two-class model, namely "Pass" and "Fail". The best classifier for "Pass" and "Fail" is then obtained separately via NC (Negative Correlation). The final result is the aggregated via NN fusion to derive the optimum predictive model for the empirical data presented. There is only one time series input parameter applying to the NN classifiers which is the BER. The output of the classifiers determines the pass and fail. the model contains 26 inputs, 1 hidden layer with 18 neurons and has 2 outputs which are passing and failing at the output layer.

Table 4.3. Results from Homogenous Classifier

Group/Predict	Testing Accuracy				
	PASS		FAIL		CUM
	TRUE	FALSE	TRUE	FALSE	*Pass/Fail*
NN1	100.0%	0.00%	46.9%	53.13%	***94.3%***
NN2	98.9%	1.12%	53.1%	46.88%	***94.0%***
NN3	96.3%	3.73%	62.5%	37.50%	***92.7%***
NN4	98.9%	1.12%	65.6%	34.38%	***95.3%***
NN5	99.6%	0.37%	46.9%	53.13%	***94.0%***
NN6	99.3%	0.75%	56.3%	43.75%	***94.7%***
NN7	99.6%	0.37%	59.4%	40.63%	***95.3%***
NN8	99.3%	0.75%	62.5%	37.50%	***95.3%***
NN9	98.9%	1.12%	59.4%	40.63%	***94.7%***
NN10	99.3%	0.75%	53.1%	46.88%	***94.3%***

From Table 4.3 shows the compararision of testing accuracy result for pass/fail of each neural network prediction for Homogenous NN classifier. The result shows that NN4th, NN^{7th}, and NN^{8th} are given the highest prediction accuracy. In order to reduce the complexity of the system, the negative correlation technique is used to eliminate the insignificant classifier output before fusing. Therefore, we can get the NC results from Homogenous NN Classifier model as shown in Table 4.4.

Table 4.4. The comparison of NC value results of PASS/FAIL each neural network

Number of NN	NC Value of Pass Group	NC Value of Fail Group
NN1	-127.267	-15.5476
NN2	-192.601	-1.96255
NN3	-17.5755	-8.61703
NN4	-64.8204	-21.1564
NN5	-83.8564	-7.33383
NN6	-301.589	-12.649
NN7	-222.896	-1.11895
NN8	-44.5341	-3.57071
NN9	-48.7347	7.675001
NN10	-12.7083	1.063913

Table 4.4 shows the comparison of NC value results of pass/fail of each neural network prediction for Homogeneous NN classifier. The result from NC shows that NN 6th has better prediction for the pass group and NN 4th has better Prediction for the fail group.

4.4 Final Homogenous and Non Homogenous NN Classifier with Negative Correlation and NN Fusion Result

The NN fusion model for Homogenous and Non Homogenous classifier contains 26 inputs, 1 hidden layer with 18 neurons and has 2 outputs which are passing and failing at the output layer. To measure the advantage of the proposed techniques, the classification accuracy is compared between Homogenous and Non homogeneous classifier with negative correlation on neural network fusion as shown in Table 4.5.

Table 4.5 summary table of Non-Homogenous versus Homogeneous

Classifier	Actual	Predict	Testing Accuracy
NON HOMOGENOUS NN + DA + K-Mean & NC & NN Fusion	PASS	TRUE	99.30%
		False (Type I error)	0.70%
	FAIL	False (Type II error)	56.20%
		TRUE	43.80%
	CUM	Pass/Fail	93.30%
HOMOGENOUS NN Classifiers & NC & NN FUSION	PASS	TRUE	98.90%
		False (Type I error)	1.100%
	FAIL	False (Type II error)	34.40%
		TRUE	65.60%
	CUM	Pass/Fail	95.30%

From Table 4.5, shows the comparison of Homogeneous and Non Homogeneous Classifier with Negative correlation on NN fusion after NC selection of the best pass and fail predictive model. In overall results, the both models are comparable. The Homogenous NN classifier with NC performs slightly better than Non Homogeneous classifier after fused by NN at 95.30%.

4.5 Various input sampling results

To reduce the test time, we sample the various inputs for both proposed methods to see the prediction accuracy. The results are able to be summarized as shown in Table 4.6.

Table 4.6 shows the prediction result of pass/failed for neural network modelling constructed.

Total Input Range (% of full test time)	Non Homogenous Classifier with NC+NN Fusion	Homogenous Classifier with NC+NN Fusion
27%	91.78%	92.81%
54%	92.96%	93.64%
81%	93.47%	94.12%
100%	95.33%	95.56%

From table 4.6 shows the prediction results of pass/failed for neural network modelling constructed. The experimental result shows that at the full test 100% provides the highest accuracy at 95.56%. However at 27% of full test time, there is the accuracy more than 90% and we can conclude that the prediction can be compromised. As a result, the reliability test process can be saved by 70%.

Chapter 5

Conclusion and Future Work

5.1 Final Model

In this paper, we have studied and proposed the new algorithm by predicting the recording head reliability degradation based on parametric data, namely the bit error rate which is the most significant factor that relates to head failure. The combination of Homogeneous and Non Homogeneous classifier with Negative Correlation (NC) on Neural Network (NN) fusion techniques is used to obtain highest recognition accuracy.

The first approach is multi non homogenous classifiers which consist of neural networks, discriminant analysis and K-mean cluster techniques to classify the empirical data for predicting the pass/fail of bit error rate degradation pattern and then combine with negative correlation that work as a selective best model prior NN Fusion. The experimental results show that the prediction accuracy gives more than 93%; however, the disadvantage of the predictive model is a complex system and obtains lesser prediction accuracy. Therefore, we have proposed the second approach which uses pool of ten neural networks or homogeneous classifier. The predictive model is used to feed forward back propagation algorithm and log-sigmoid transfer function. The experimental result shows that the prediction accuracy is given more than 95% after combined with negative correlation and NN Fusion which is significantly improved in comparison with the previous approach method.

The both types need a NN fusion to combine their outputs to give the final results, the experiment gives the results that the two proposed methods can be achieved by competitive generalization performance and they have the most robust performance with accuracy level more than 90% and Homogenous NN Classifier is highly recommended due to lesser complexity in implementations and better accuracy.

In the final step of study, we apply the various ranges of input data (25%, 51%, 84% and 100%) for both approach methods, the result shows that the predictive model able to obtain a robust performance even smaller of input range with an acceptable accuracy level not less than 90%. By this way, the test time could be very well reduced.

5.1 Future work

The predictive model is not flexible to apply for all programs due to different drive configurations and technology. We need to study the other critical parametrics to feasibility with various failure modes and enhance the predictive model ability prior to implementation in mass volume. In the future, to implement this predictive model for real time monitoring, it requires more experimental results with a variety of HDD product segments such as Datacenter, Desktop and Mobile. Since each product has built with different magnetic recording heads, recording media type and different technologies are constructed.

References

[1] **Sebastian Anthony on November 12, 2013 at 6:00 am [Online]** Available: http:// www.extremetech.com/computing/170748-how-long-do-Hard-drives-actually-live-for.

[2] Frank San a, Jimmy Yang, **"A Comprehensive Review of Hard-Disk Drive Reliability",** PROCEEDINGS Annual RELIABILITY and MAINTAINABILITY Symposium, IEEE 0-7803-5143-6/9, 1999

[3] Edgar M.Williams, **"Design and analysis of magneto resistive recording heads"** A Wiley-Interscience Publication New York, 2001.

[4] Eduardo Pinheiro, Wolf-Dietrich Weber and Luiz Andre Barroso, "Failure **Trends in a Large Disk Drive Population**", the 5th USENIX Conference on File and Storage Technology (FAST'07), February 2007.

[5] Bingpeng Zhu, Gang Wang, Xiaoguang Liu, Dianming Hu, Sheng Lin, and Jinwei Ma, " **Proactive Drive Failure Prediction for Large Scale Storage Systems**", IEEE 978-1-4799-0218-7/13, 2013.

[6] Zhi-sheng Ye, Yu Wang, Kwok-Leung Tsui, and Michael Pecht, " **Degradation Data Analysis Using Wiener Process With Measurement Errors**", IEEE Transactions on reliability, Vol. 62, No. 4, December 2013.

[7] Loon Ching Tang and Dong Shang Chang, **"Reliability Prediction Using Nondestructive Accelerated-Degradation Data: Case Study on Power Supplies"**, IEEE Transactions on reliability, Vol. 44, No. 4, December 1995.

[8] Hirokazu Takahashi, Narade Pecharanin, Yoshinao Akima and Motota sone, "The **Reliability of Neural Network on Pattern Recognition**", IEEE 0-7803-1901-X/94 @ 1994.

[9] Joanne Bechta Dugan and James W.Watterson, **"Reliability Analysis of Artificial Neural Networks",** IEEE 0149-144X/91/0000, 1991.

[10] Meyer, Sebastian; Wohlrabe, Heinz and Wolter, Klaus-Jurgen, **"Neural Network Modeling to Predict Quality and Reliability for BGA Solder Joints"**, IEEE 978-1-4244-6412-8/10 @ 2010. 2010 Electronic Component and Technology Conference.

[11] Yong Liu, Xin Yao, and Testsuya Higuchi, **"Evolutionary Ensembles with Negative Correlation Learning"**, IEEE Transactions on evolutionary, Vol. 4, No. 4, November 2000.

[12] Huanhuan Chen and Xin Yao, **"Regularized Negative Correlation Learning for Neural Network Ensembles"**, IEEE Transactions on neural network, Vol. 20, No. 12, December 2000.

[13] Shuo Wang and Xin Yao, "**The Effectiveness of A New Negative Correlation Learning Algorithm for Classification Ensembles**", IEEE International Conference on Data Mining Workshops 2010.

[14] Y.Liu and X.Yao "**Ensemble learning via negative correlation**", Neural networks 12 (1999) 1399-1404.

[15] K. Pomoung. and Y. Kidjaidure, "**Rank-Level Fusion with Neural Network in Classification of Media Defect Patterns.**", The 3rd International Data Storage Technology Conference. Thailand – Conference, Jan 9-12, 2012.

[16] Waraporn Tepin and Yuttana Kidjaidure, "**Customer Failure Modes Prediction for Hard Disk Drive using Neural Networks Rank-Level Fusion**", the 8^{th} Electrical Engineering/Electronics, Computer, Telecommunications and Information Technology (ECTI) Association of Thailand- Conference 2011.

[17] Minlong Lin, Ke Tang and Xin Yao, "**Selective Negative Correlation Learning Algorithm for incremental Learning**" IEEE 978-1-4244-1821-3/08, 2008.

[18] MS.Trupti Pathrabe, Dr.N.G.Bawane, "**Feature Extraction Parameters for Genuine Paper Currency Recognition & Verification**" International Journal of Advanced Engineering Sciences and Technologies, Vol No. 2, Issue No. 1, 085-089.

[19] Anil K.Jain, Robert P.W.Duin, and Jaianchang Mao. "**Statistical Pattern Recognition: A review**", IEEE Transactions on Pattern Analysis and Mechine Intelligence, Vol. 22, Jan 2000, pp-4-37.

[20] Ahmed Z and Ahmad Jamal, "**Neural Network-based failure rate prediction for De Havilland Dash-8 tires**", Engineering Application of Artificial Intelligence 19 (2006) 681-691.

[21] Di Xu Ph.D, Wenbiao Zhao Ph.D, "**Reliability Prediction using Multivariate Degradation Data**", IEEE 0-7803-8824-0/05, 2005.

[22] Martin T. Hagan, Howard B.Demuth and Mark Beale 1996, "**Neural Network Design**" PWS Publishing Company.

[23] Patrick D.T.O'connor, David Newton, and Richard Bromley 2002, **Practical Reliability Engineering**. John Wiley & Sons.

[24] C. Moura, Eduardo 1991, **How to determine sample size and estimated failure rate in life testing**. ASQC Quality Press.

[25] Wayne Nelson 1983, "**How to analyze reliability data**", ASQC Quality Press.

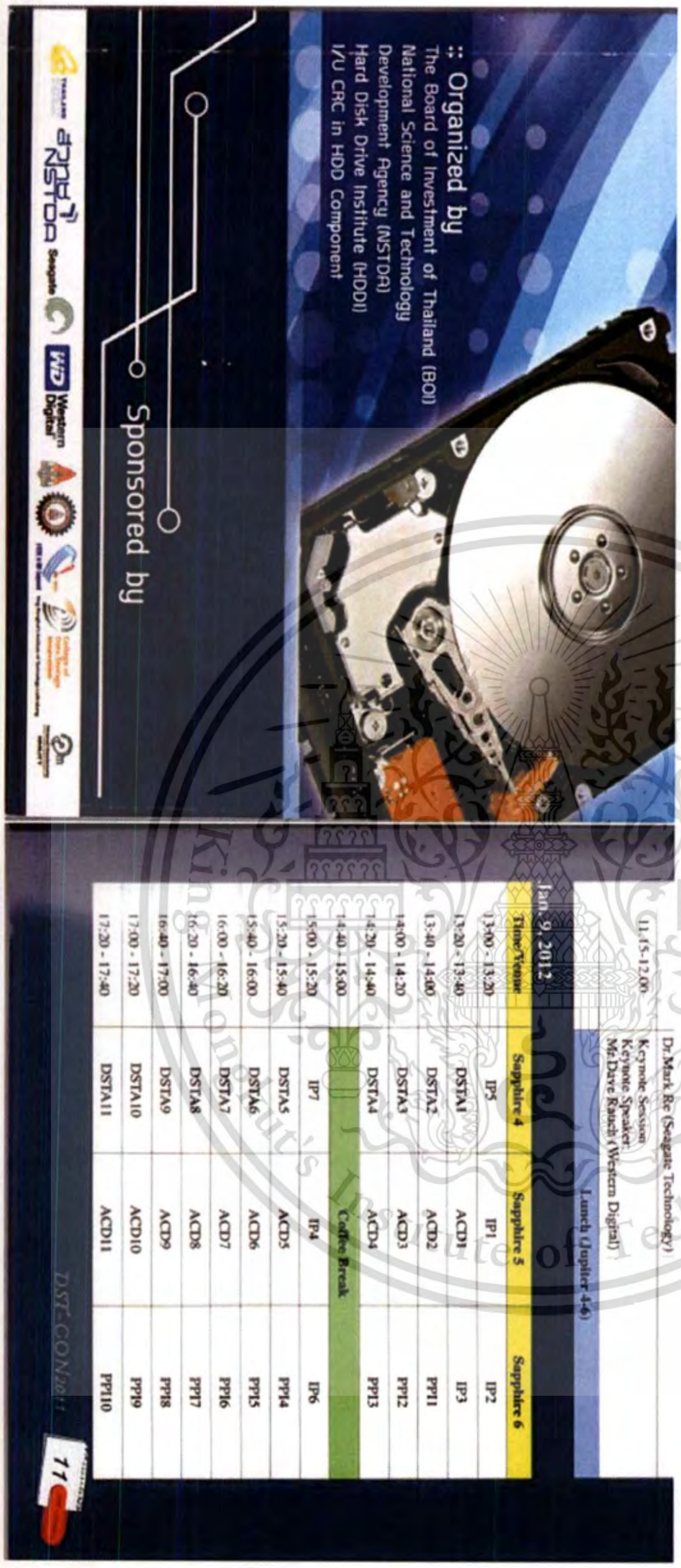
[26] Kanu G. Ashar 1997, "**Magnetic Disk Drive Technology**", Institute of Electrical and Electronics Engineering, New York.

Appendix A

Acronyms, Definitions and Symbols

This thesis document uses the following set of Acronyms, definitions and Symbols

ANN : Artificial Neural Network
LDA : Linear Discriminant Analysis
NN : Neural Network
RDT : Reliability Demonstration Test
BER : Bit Error Rate
VGA : Voltage Gain Margin
VFR : Virtual Failure Rate
TTF : Time to failure
BP : Back propagation
BLP : Base line pop
BLN : Base line noise
COW : Cold over write
EM : Error Margin
HDF : Hard Defect
HFW : Hly Fly light
SPT : Single Plug Tester
HDD : Hard Disk Drive
NC : Negative Correlations
AFR : Annual Failure Rate
SNR : Signal noise ratio
MSE : Mean Square Error
PRML : Partial-response maximum-likelihood
MRR : Magnetic Resistance Ratio



:: Organized by

The Board of Investment of Thailand (BOI)
National Science and Technology
Development Agency (NSTDA)
Hard Disk Drive Institute (HDDI)
I/U CRC in HDD Component

Sponsored by

11.15-12.00	Dr.Mark Re (Seagate Technology) Keynote Session Keynote Speaker: Mr.Dave Rauch (Western Digital)
	Lunch (Jupiter 4-6)

Jan. 9, 2012

Time/Venue	Sapphire 4	Sapphire 5	Sapphire 6
13:00 - 13:20	IP5	IP1	IP2
13:20 - 13:40	DSTA1	ACD1	IP3
13:40 - 14:00	DSTA2	ACD2	PPI1
14:00 - 14:20	DSTA3	ACD3	PPI2
14:20 - 14:40	DSTA4	ACD4	PPI3
14:40 - 15:00	Coffee Break		
15:00 - 15:20	IP7	IP4	IP6
15:20 - 15:40	DSTA5	ACD5	PPI4
15:40 - 16:00	DSTA6	ACD6	PPI5
16:00 - 16:20	DSTA7	ACD7	PPI6
16:20 - 16:40	DSTA8	ACD8	PPI7
16:40 - 17:00	DSTA9	ACD9	PPI8
17:00 - 17:20	DSTA10	ACD10	PPI9
17:20 - 17:40	DSTA11	ACD11	PPI10

DST-CON2011

11

DSTA 5

Title: Prediction of Recording Head Reliability Degradation Using Neural Network Fusion with Negative Correlation

Author: [1]Chalermchai Jorjong
[2]Yuttana Kitjaidure
[3]Kittiphan Pomoung
[4]Chee Wai Mak

Affiliation: [1]Faculty of Engineering, KMITL Bangkok, Thailand
[2]Western Digital (Thailand) Company Limited Ayutthaya, Thailand

E-mail: [1]chalermchai.jorjocg@wdc.com
[2]kkyuttan@kmitl.ac.th
[3]kittiphan.pomoung@wdc.com
[4]mak.cheewai@wdc.com

Abstract: Hard disk drive industry is growing years over years with many new technologies. New technology always has a challenge in reliability. Moreover to determine the reliability failures is another challenge. As in any industry, the drive industry reliability test takes long test time and almost needs to wait till the end of the test to determine the results. This delay has an impact in cost and the product launching time. In order to avoid the time delay, the decision making in current practice is normally done based on the early results during the reliability testing. This paper focuses on predicting the reliability failure based on the parametric data using combination of Homogeneous and Non Homogeneous Classifier with Negative Correlation (NC) on Neural Network (NN) fusion techniques. The results show that the combination models have the most robust performance with higher level of accuracy. By this way the test time could be very well reduced.

Author Biography

Name: Mr. Chalermchai Jorjong
Date of Birth: Dec 05, 1976
Place of Birth: Phrae, Thailand

Educations:

1996 - 1998 King Mongkut's Institute of Technology North Bangkok, Power Electronic Technology (Second class honor).

2009 - 2011 Scholarship for study in the M.Eng. of Data Storage Technology (English program) by NSTDA, KMITL and Western Digital (Thailand) Co., Ltd, International College, College of Data Storage and Innovations.

Publications:

Chalermchai J, Yutana, Kitiphan, CW Mark, "PREDICTION OF HEAD RECORDING RELIABILITY DEGRADATION USING NEURAL NETWORK FUSION WITH NEGATIVE CORRELATION", DST-CON 2011 The 4th International Data Storage Technology Conference.

Prediction of Recording Head Reliability Degradation Using Neural Network Fusion with Negative Correlation

Chalermchai Jorjong#1, Yuttana Kitjaidure#2, Kittiphan Pomoung*3, Chee Wai Mak*4

Faculty of Engineering, KMITL
Bangkok, Thailand

**Western Digital (Thailand) Company Limited*
Ayutthaya, Thailand

[1] chalermchai.jorjocg@wdc.com
[2] kkyuttan@kmitl.ac.th
[3] kittiphan.pomoung@wdc.com
[4] mak.cheewai@wdc.com

Abstract— Hard disk drive industry is growing years over years with many new technologies. New technology always has a challenge in reliability. Moreover to determine the reliability failures is another challenge. As in any industry, the drive industry reliability test takes long test time and almost needs to wait till the end of the test to determine the results. This delay has an impact in cost and the product launching time. In order to avoid the time delay, the decision making in current practice is normally done based on the early results during the reliability testing. This paper focuses on predicting the reliability failure based on the parametric data using combination of Homogeneous and Non Homogeneous Classifier with Negative Correlation (NC) on Neural Network (NN) fusion techniques. The results show that the combination models have the most robust performance with higher level of accuracy. By this way the test time could be very well reduced.

Keywords— Classification, Reliability, Negative correlation, Neural Network fusion.

I. Introduction

In the past decade, the weibull has been widely used to predict the reliability of products based on time to failure. Now new HDD generation has higher level of complexity, which is too complicated to determine the failures. So we introduce a new technique using the parametric data such as bit error rate (BER) to monitor the product reliability performance. The bit error rate is one of the critical parameters which are used for the disk drive testing. This parameter is a clear indicator for the head degradation performance and represents clearly the pass and fail criteria for the drive failure mode as shown in Fig 1.

To predict the bit error rate degradation pattern, NN is widely used and very famous in intelligent system to discover the relationship between the variables. The learning and predicting of the networks are based on time series data in reliability test as shown in Fig 1.

A neural-network approach in [1],[2] is proposed for Failure Mode Prediction for Hard Disk Drive using Neural Network Rank-Level Fusion. This algorithm provides the best combination of classifier and [3] is an approach of a quality monitoring tool as a pre-screening model to screen the defective work in process parts.

To enhance neural network system capability, negative correlation is introduced and the advantage of this algorithm is to eliminate the insignificant classifier output in order to reduce the complexity of the system as shown in [4],[5].

There are few review papers introduced on the degradation data [6, 7]. However, it shown a significant approach to evaluate the reliability and prediction.

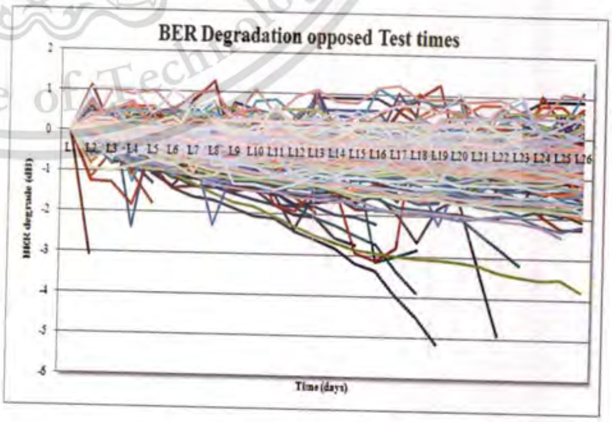


Fig 1. Bit Error Rate Degradation Pattern

Figure 1 exhibits the bit error rate degradation pattern. The x-axis indicates the time duration and the y-axis indicates the BER level. Multiple drives are tested and it is shown in different colours where a single colour represents a single head data points. We can clearly see that as the test duration increases the level of the BER gets lower which is a sign of degradation for fewer samples.

This paper proposes two different classifications to predict pass/fail of bit error rate degradation patterns. The first approach is multi classifiers which consists of neural networks, discriminate analysis and K-mean cluster techniques and the other is the pool of neural networks. So the both types need a NN fusion to combine their own outputs to give the final results.

In this study methodology is described in Section II, Experiment and Results are detailed in Section III and finally conclusion is given in Section IV.

II. METHODOLOGY

A. Pre-Processing

Empirical data, bit error rates, are gathered in time series since starting until completing the full test (26 time steps). Due to large variations of the failure mode, we picked only the head failures that correlate to bit error rate degradation performance and excluded those non-head related.

B. Classification

Empirical data are then put through the two different classifier model, namely, Homogeneous and Non Homogeneous Classifier model.

The Non Homogeneous classification model contains three different classification methods, Neural Network, Discriminant Analysis (DA) and K means.

1) *Neural Networks (NN)*: an intelligent system that is able to recognize the behaviour of data sets based on historical inputs. We use Neural Networks (NN) because of its ability to learn by examining individual records, generating a prediction. We set up the NN using pruning method, starting with a large network and remove the weakest units in the hidden and input layers as training proceeds.

2) *Discriminant Analysis (DA)*: classified by building a predictive model based on linear combination of predictive variable with priority assigned.

3) *K means (KM)*: a method of cluster analysis which aim to partition n observation into k clusters in which each observation belong to the cluster with the nearest mean.

The Homogeneous classification model contains the pool of ten Neural Networks. The objective of this classification model is to classify the empirical data into two classes namely, "PASS" and "FAIL", according to the results obtained from reliability testing of the HDD.

C. Negative Correlation

The negative correlation technique [5] is used to eliminate the insignificant classifiers outputs (as per the formula 1) before fusing. This is to reduce the complexity of the NN fusion model, which in turn improves the fusion performance and accuracy. After elimination process with NC technique, the data is fused with the final NN Fusion.

NC produces negatively correlated for the classifiers level and attempts to train individual classifier method then combine them in the same learning process. In NC, all the individual classifier method is trained simultaneously and interactively through the correlation penalty terms in their error functions.

Suppose that we have a training set. Here we form the classifier method whose output is simple averaging of outputs of the set of NNs.

$$F(n) = \frac{1}{M}\sum_{i=1}^{M} F_i(n)$$

Where M is the number of the individual classifier is, Fj(n) is the output of classifier *i* on the nth class and F (n) is the output on the n class.

The penalty function pi may have the form:

$$p_i(n) = (F_i(n) - F(n))\sum_{j=i}(F_j(n) - F(n))$$

We adjust the number of neurons in the hidden layer and the output layer of neural networks and utilize the variant network structures and algorithms to analyse and modify the results. Therefore, an optimum prediction system is acquired.

D. Aggregating Result – Combination of Non Homogenous and Homogenous with Negative Correlation on Neural Network Fusion

1) Non Homogeneous Classifier with Negative Correlation on Neural Network Fusion :

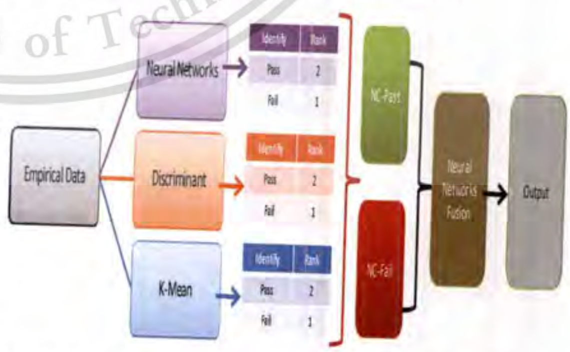


Fig 2 Diagram showing Non Homogeneous classifier model flow incorporating NC Neural Network Fusion.

In this arrangement, empirical data is presented to three classifiers, namely NN, DA and K-Mean. Each classifier independently achieves a model for a two-class output, namely "PASS" and "FAIL". The best classifier for the "PASS" and the "FAIL" classes are separately selected via NC. In this way, the most optimum classifier required for each of the two output classes ("PASS" and "FAIL") are derived for the set of empirical data presented. The results are then aggregated via NN fusion to derive the predictive model for the empirical data presented.

2) *Homogenous NN Classifier with Negative Correlation on Neural Network Fusion:*

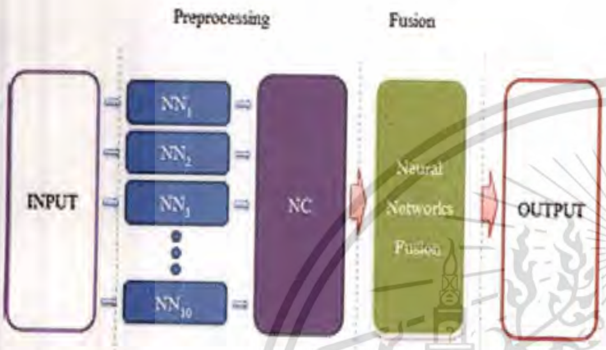


Fig3. Diagram showing Homogenous model flow incorporating NC Neural Network Fusion

In this arrangement, empirical data are presented to ten Neural Networks (NN). In each network, the training / testing data sets are forced to be re-established but with the same 60-40 ratio between training and testing. Through the ten iterations, ten classification models will be achieved. The outputs of each NN are a two-class model, namely "PASS" and "FAIL". The best classifiers for "PASS" and "FAIL" are then obtained separately via NC. The final result is then aggregated via NN fusion to derive the optimum predictive model for the empirical data presented.

There is only one time series input parameter applying to the NN classifiers which is the BER. The outputs of the classifiers determines the pass and fail.

III. EXPERIMENTAL RESULTS

A total 300 samples with 7,800 records of bit error rate were used. These samples were selected from reliability test and the data are divided into two parts with random sampling. The first part is for model training (60%) and the second part is for model testing (40%) which is unknown data sets. To ensure the data is enough for NN learning and validation.

The cross validation is used to estimating the performance of a predictive model and the practice of confirming an experimental finding by repeating the experiment.

The NN fusion model for Homogeneous and Non Homogeneous classifier contains 26 inputs, 1 hidden layer with 18 neurons and has 2 outputs which are passing and failing at the output layer. NC is subjected between Classifier and Fusion model pretend to select the best classier prior obtaining the next input to NN Fusion Model.

To measure the advantage of the proposed technique, the classification accuracy is compared between Homogeneous and Non homogeneous classifier with Negative correlation on NN Fusions.

A. *Results from Non Homogenous Classifier and Negative Correlation*

TABLE I
RESULTS FROM NEGATIVE CORRELATION ALGORITHM OF NON HOMOGENEOUS CLASSIFIER

Model	NC Value of Pass Group	NC Value of Fail Group
Neural network	-3878.013	-1493.9142
Discriminate Analysis	-13231.42	-1002.8491
K-Means	-2783.006	-4944.759

Table 1. shows the comparison of NC value results of pass/fail between supervised and unsupervised learning group with three modelling namely neural network, discriminant analysis and K-means. The result from Negative Correlation shows that neural network has better prediction for the pass group and K-means has better prediction for the fail group.

B. *Results from Homogenous NN Classifier and Negative Correlation*

TABLE III
RESULTS FROM NEGATIVE CORRELATION ALGORITHM OF HOMOGENEOUS CLASSIFIER

Number of NN	NC Value of Pass Group	NC Value of Fail Group
NN1	-127.267	-15.5476
NN2	-192.601	-1.96255
NN3	-17.5755	-8.61703
NN4	-64.8204	-21.1564
NN5	-83.8564	-7.33383
NN6	-301.589	-12.649
NN7	-222.896	-1.11895
NN8	-44.5341	-3.57071
NN9	-48.7347	7.675001
NN10	-12.7083	1.063913

Table 2. shows the comparison of NC value results of pas/fail of each neural network prediction for Homogeneous NN classifier. The result from NC shows that NN 6th has better prediction for the pass group and NN 4th has better prediction for the fail group.

C. *Results from Homogenous NN Classifier and Non Homogeneous classifier with Negative Correlation on NN fusion*

TABLE IIIII
RESULTS FROM NEURAL NETWORK FUSION

Category	Prediction Accuracy (%)
Homogneous NN Classifier + NC + NN Fusion	95.56%
Non Homogenous Classifier + NC + NN Fusion	95.33%

Table 3. shows the comparison of Homogeneous and Non Homogeneous Classifier with Negative correlation on NN fusion after NC selection of best pass and fail predictive model. In overall results, the both models are comparable. The Homogenous NN classifier with NC performs slightly better than Non Homogeneous classifier after fused by NN at 95.56%.

D. Results of Prediction at Various Input Ranges

TABLE IVV
RESULTS OF PREDICTION AT VARIOUS INPUT RANGES

Total Input Range (% of full test time)	Prediction Accuracy
27%	90.00%
54%	91.00%
81%	92.67%
100%	94.00%

Table 4. shows the prediction result of pass/failed for neural network modelling constructed. The result shows that the full test provided the highest accuracy at 94%. However at 27% of full test time, there is the accuracy > 90% and we can conclude that the prediction can be compromised. As a result, the reliability test process can be saved by 70%.

IV. CONCLUSIONS

In this paper, we have studied and proposed the new algorithm by predicting the head reliability degradation based on the parametric data using a combination of Homogeneous and Non Homogeneous classifier with Negative Correlation (NC) on Neural Network (NN) fusion techniques.

The results show that the proposed predictive model has the most robust performance with an accuracy level > 90%. By this way the test time could be very well reduced.

ACKNOWLEDGMENT

We wish to thank Western Digital for supporting on the Master scholarship and for providing data that is necessarily for the paper. We hope the contents of this paper can contribute to product reliability improvement. This work has been fully supported by Western Digital, KMITL University and NECTEC (Thailand).

REFERENCES

[1] W. Tepin and Y.Kidjaidure, "Customer Failure Mode Prediction for Hard Disk Drive using Neural Network Rank-Level Fusion", the 8th Electrical Engineering/Electronics, Computer, Telecommunications and Information Technology (ECTI) Association, Thailand - Conference, 476-479, 2011.

[2] K. Pomoung and Y. Kitjaidure Author, "Rank-Level Fusion with Neural Network in Classification of Media Defect Patterns" The 3rd International Data Storage Technology Conference. Thailand-Conference, July 30-Aug 1, 2010

[3] S. Pattanapairoj, and D. Chetchotsak, "Quality monitoring model for failed parts detection in HDD manufacturing processes," Data Storage Technology conference. Bangkok , Thailand, May. 13-15, 2009.

[4] Y. Liu and X. Yao, "Ensemble Learning via Negative Correlation", Neural Networks 12 (1999) 1399-1404, 1999.

[5] S. Wang, H.H. Chen and X. Yao, "Negative Correlation Learning for Classification Ensembles", Neural Networks (IJCNN), The 2010 International Joint Conference, 1-8, October 2010.

[6] Nima G, Lin Ma, Murthy Mittinty, Prasad Y, "A review on degradation model in reliability analysis", the 4th world congress on engineering asset management, 28-30 September 2009.

[7] Joseph Lu and William Q, "Using degradation measure to estimate a time to failure distribution" , Techno metrics May 1993, Vol 35 No2