# A New Neuron Model Based on Dendritic Mechanism and Its Applications
# (樹状突起のメカニズムに基づく新型ニューロンモデルとその応用)

*A dissertation submitted to the*

*Graduate School of Science and Engineering for Education,*

*University of Toyama*

*Toyama, Japan*

*in partial fulfillment of the*

*requirements for the degree of*

***DOCTORATE OF PHILOSOPHY (Ph.D.)***

*in Advanced Mathematics and Human Mechanisms Science*

***March 2014***

*By*

***Zijun Sha***

CONTENTS

ACKNOWLEDGEMENTS

I would like to express my special appreciation and thanks to my advisor Professor Tang, you have been tremendous mentor for me. I would like to thank you for encouraging my research and for allowing me to grow as a research scientist. Your advices on both research as well as on my career have been invaluable.

I would also like to thank my committee members, Professor KIKUSHIMA, Professor YAMAZAKI and Professor Gao for serving as my committee members even at hardship. I also want to thank you for letting my defense be an enjoyable moment, and for your brilliant comments and suggestions, thanks to you.

I would also like to pay my appreciation to my senior: Ken Takeuti and the members in the laboratory. They are so kind and optimistic to make the lab a lovely place. So I can be in a good mood every day and do my study horse and foot.

A special thanks to my family. Words cannot express how grateful I am to my mother and my father for all of the sacrifices that you've made on my behalf. Your prayer for me was what sustained me thus far. I would also like to thank to my beloved wife. Thank you for supporting me for everything, and especially I can't thank you enough for encouraging me throughout this experience.

# Abstract

All along, many scientists, one after another, in order to clarify the functions of the human brain to make a lot of effort. From the appearance of neural networks to the study of brain waves, the exploration of the brain has not stopped. With the development of science and technology, more and more new technologies are being used on the research of brain. Research in this area has attracted more and more attention. In order to understand the process of brain, the principle of the complex neural systems, many researchers will shift attention to the neurons, the basic building blocks of the nervous system. Breast cancer is a preponderant disease in the world, and it is one of the most major death causes of women. In order to predict the cancer in women, recent years, artificial intelligent (AI) has been widely used in the scopes. This thesis deals with the application of a novel a neuron model based on dendritic mechanism for classifying breast cancer on Wisconsin breast cancer database (WBCD). As its name suggests, the dendrites mechanism is the main computation of the neuron model. The model neuron is composed of a set of independent branches and a soma. Instead of being weighted simply, the inputs of the neuron model are processed nonlinearly rather than being weighted simply to realize excitatory synapse, inhibitory synapse, constant-1 synapse or constant-0 synapse. The signal of each branch is weighted and performed due to the input. The Soma receives the signals transmitted from the branches to produce the output. The performance of the neuron model based on dendritic mechanism is compared with the classic back propagation neural networks (BPNNs). Simulation results indicate that the neuron model based on dendritic mechanism holding superior capability at the accuracy, convergence speed, stability and AUC. In addition, it is worth of note, through learning, an arbitrarily dendrite of neurons with different initial synapses can develop an internal structure which depends on the location of synapses in the branch, and the type of synapse. Furthermore, in this simulation, the developed structure may suggest some inspirations to the detection of breast cancer. In addition, owing to non-decrease on classification accuracy after

eliminating the useless branches of the neuron model, the computation load can be released. In this thesis, the trading data from January 2004 to October 2014 in the Shanghai stock market is selected to verify the overreaction on Shanghai stock market. And, the overreaction from 2007 is found to turn to weaken with the time going by and the influence of the overreaction turn to disappear from 2011. Moreover, the neuron model based on dendritic mechanism, for the first time, is also proposed to fit and predict the changes about abnormal returns of ill-performed and well-performed stocks in test period. The result shows that the neuron model possesses high computational ability and successes to predict the tendency of overreaction.

***key words:*** *Neuron Model with Dendritic Nonlinearity（NMDN）, Wisconsin breast cancer database (WBCD), back propagation neural networks(BPNNs), dendrites mechanisms, overreaction, stock market*

# Chapter 1

# Introduction

What is life? What is the mean of “Man Alive”? How brain thinks? Countless questions emerge in the human mind. The Chinese view of man also arrived at the idea that man is the “Lord of the Creation” unlike other species, because the brain of human is extremely complex, since it is the crystallization of evolution by millions of years. In the past 600 million years, the organisms have evaluated into a large number of neurons linked to each other by the formation of the neural network to solve the problems that how the human brains handle a variety of complex information in a complex and changing environment. Especially, highly developed human cognitive capacity results that human become “Lord of the Creation” to dominate the world. Recent years, all the countries have invested a lot of resources for specialized research. For example, the United States has sponsored the plan of “ten years of the brains”; European has determined the “two decades of brain research program”, Japan has also named the 21st century as the “brain science century”. The brain science research booms worldwide. Scientists have proposed the three goals, “Understanding the brain, protecting the brain and create brain”. It is believed that the brain science research will help human beings to understand themselves, protect themselves, prevent from brain diseases and treat them, and even advance the development of the brain and other aspects of the potential to make a significant contribution. Facing the 21st century science, “Understanding the brain, recognize their own” is the biggest challenge.

## 1.1 Brain

Brain, with the aim of perceiving outside environment by advanced information processing ability of itself, can abstract the characteristics from the information of outside and judge the result which is generated by the abstraction through being accepted and inputted the sound, lights, etc. from outside. So, it's very important for the application of the engineering and the comprehension of the brain that the research of information processing in brain. However, it's very difficult to realize all the functions of the brain by the artificial neural network. As a result, most functions of the brain are unknown to us [1].

Research also found that the weighs of an adult brain is about 3.3 pounds, the volume is 1.5 liters, and there is one hundred billion nerve cells, as well as more than $10^{14}$ synapses in it [2]. The structure and functions of brain in vivo are the most complex organization, can accept external signals resulting in the feeling, the formation of consciousness, logical thinking and can issue a directive to produce behavior headquarters, which is in charge of the daily human language, thought, feeling, emotion, movement and other senior activities. The human brain is extremely delicate and a perfect information processing system. Because of the structure and functions of the human brain are extremely complex, it will be possible to reveal its mysteries by researching and integrating molecular, cellular, systems and behavior.

## 1.2 Neural network and Neurons

An ANN is an information processing system that roughly replicates the behavior of a human brain by emulating the operations and connectivity of biological neurons [3]. Most of the artificial neural network can change the internal structure on the basis of outside information. Modern neural network is a nonlinear statistical data modeling tool used to model complex relationships between inputs and outputs, or to explore the data model. ANNs possess a variety of alternative features such as massive parallelism, distributed representation and computation, generalization ability, adaptability and inherent contextual information processing [4-11]. In the fields of clinical medicine and biomedical engineering, ANNs have been used to solve complex and chaotic problems without the need of mathematical models and a precise understanding of the mechanisms involved [12-18].

However, the backbone of many artificial neuronal networks is the starting from the original work of McCullough & Pitts (1943) to the present day. The node of neural network, also called neuron, is the tradition neuron model, McCulloch-Pitts Model, which is widely used in the traditional neural networks to solve many complicated problems by incorporation into the multilayer networks and whose structure is too simple to solve the Exclusive OR problem alone. Hence, McCulloch-Pitts Model is unable to reflect the real importance of neurons in the nervous system.

Recently, more detailed characteristics appear in the detailed analysis of the nerve cells. A certain ions exist in the receptor of dendrites, and when the ions enter the receptors whose potential changes. The responses of receptors determine are based entirely on the synapses which are excitation or inhibition. Furthermore, it has also been shown that inhibition of nonlinear interactions is located among synapses [19, 20].

In 2013, Spencer L. Smith et al. succeeded in making incredibly challenging electrical and optical recordings directly from the tiny dendrites of neurons in the intact brain while the brain was processing visual information by examining neurons in areas of the mouse brain which are responsible for processing visual input from the eyes.

The experiment result shows that the local information processing is realized in dendrites [21]. The result challenges the widely held view that the computation is achieved only by large numbers of neurons working together, and also demonstrates that the basic components of the brain possess an exceptionally powerful computation. In other words, dendrites act as small-scale computing devices for detecting and amplifying specific types of input. This new property of dendrites adds an important new explanation to the computation in the brain.

In 2000, Tang et al. proposed a neuron model based on dendritic mechanisms [22]. In this model, in order to reflect the physiological morphology and the functions of nerve cells, the logical AND (owing to the soft-minimum function) are used to realize the switch function and the logical OR (owing to the Soft-maximum function) are used to deal with when there are two or more inputs and there are switch in parallel. Furthermore, a logical NOT (owing to the sigmoid function) is also required when a signal is transmitted in the dendrites. Thus, a nonlinear interaction belonging to the dendrites can be expressed by logic operation AND, OR and NOT. Moreover, the neuron model can heighten and fix the practical dendrites and synapses, filter out the worthless ones by training to form a mature dendrites shape and preserve the worth synapse [23]. However, owing to the logic operation, there is a key limitation of this neuron model is that it can only solve the problem of the binary value smoothly. In order to generalize the computation of neuron model into a continuum of values between 1 and 0, Tang et al. have proposed to use the multiplication and addition to replace the logical AND and OR, respectively.

## 1.3 Outline of this thesis

In this thesis, a neuron model based on dendritic mechanisms, for the first time, is proposed for solving two real world problems, classifying the breast cancer lesions as benign or malignant on the WBCD database and predicting the trends of overreaction in Shanghai stock market.

In chapter1, the functions, features and the relationship of brain, neural networks and neurons are introduced. In chapter 2, two types of neuron model are described in detail. They are McCulloch-Pitts Model and Koch-Poggio-Torre model. In chapter 3, a very detailed description of the new neuron model which is called Neuron Model with Dendritic Mechanism and can be also called Neuron Model with Dendritic Non-linearity (NMDN) is carried out. In chapter 4, NMDN is used for the detection of breast cancer and compared to traditional back-propogation neural networks (BPNNs) for the first time. In chapter 5, NMDN is also for the first time to be proposed to predict the change trends of the overreaction on Shanghai stock market on the basis of the trading data from January 2004 to October 2014. In chapter 6, it is the conclusion of the results of two applications on classification and prediction to prove that a single neuron has a very superior computing power.

# Chapter 2

# The review of neuron model

The brain is a large-scale information processing system that consists of $10^{11}$ neurons with perhaps $10^{15}$ interconnections between them [24]. Moreover, there are no new neurons being formed after one-year old human, but neurons will be lost at a rate of roughly 200,000 per day (a net loss of 2 to 5% by age 50). The brain weight is achieved maximally at about age 21. A neuron can connect with more than 1,000 other neurons. Aside from neurons, the other important brain cells are the glia, which are more numerous than neurons (a human brain may contain a trillion glial cells). The portion of brain cells which are glia are 25% in the fruit fly, 65% in the mouse, 90% in the human, and 97% in the elephant. The four types of brain glia cells are: (1) astrocytes, (2) oligodendrocytes, (3) microglia, and (4) ependymal cells [24, 25]. Furthermore, although the shape and size of neurons are different, the structure of the neurons is unique, which consists of three portions, cell body, dendrite and axon. The incoming signal from other neurons or sensors received by the dendrite is computed at the synapse and transmitted to the cell boy. When the input into the cell body exceeds the holding threshold, the neuron will fire, and the signal be transmitted to other neurons through axon.

## 2.1 Neuron

A neuron, also known as nerve cell, is the functioning unit of the nervous system, which includes the brain, spinal cord, which together comprise the central nervous system (CNS), and the ganglia of the peripheral nervous system (PNS); specialized to receive, integrate, and transmit information (Fig.2-1) [24].

A typical neuron possesses a cell body (soma), dendrites, and an axon (Fig.2-2). Dendrites arise from the cell body, often extending for hundreds of micrometres and branching multiple times, giving rise to a complex “dendritic tree”. “An axon is a special cellular extension arising from the cell body at a site called the axon hillock and travels for a distance, as far as 1 meter in humans or even more in other species [26].” The cell body of a neuron frequently generates multiple dendrites, but never to more than one axon, although the axon may branch hundreds of times before terminating. At the majority of synapses, signals are sent from the axon of one neuron to a dendrite of another.

Neuron is also a kind of electrically excitable cell processing and transmitting information through electrical and chemical signals (Fig.2-3). The signals among neurons occur through synapses, specialized connections with other cells. “Specialized types of neurons include: sensory neurons which respond to touch, sound, light and all other stimuli affecting the cells of the sensory organs that then send signals to the spinal cord and brain, motor neurons that receive signals from the brain and spinal cord to cause muscle contractions and affect glandular outputs, and interneurons which connect neurons to other neurons within the same region of the brain, or spinal cord in neural networks [26].”

“All neurons are electrically excitable, maintaining voltage gradients across their membranes by means of metabolically driven ion pumps, which combine with ion channels embedded in the membrane to generate intracellular-versus-extracellular concentration differences of ions such as sodium, potassium, chloride, and calcium [26].” The potential changes in the cross-membrane can change the function of voltage-dependent ion channels. If the potential changes by a large enough amount, an action potential is generated. The action potential travels rapidly through the cell's axon, and can activate synaptic connections with other cells on its arrival.

In addition, the cell division does not appear among neurons. In most cases, mere some special types of stem cells can turn into neurons. However, a type of glial cell, is called astrocytes, has also been observed to turn into neurons by virtue of the stem cell. In humans, the neurogenesis almost ceases during adulthood, however, in two brain areas, the hippocampus and olfactory bulb, there is strong evidence for generation of substantial numbers of new neurons [27].

Moreover, during the past few years, there has been an explosion of interest in dendrites, driven by the progressed powerful new imaging and recording techniques. There is more and more evidence disclosing that dendrites substantially enhance the neuron's computational power by introducing nonlinear interactions between synapses and subcompartments of the cell [28]. In other words, "additional linear and nonlinear mechanisms in the dendritic tree are likely to serve as computational building blocks, which combined together playing a key role in the overall computation performed by the neuron. [29]"

## 2.2 McCulloch-Pitts Model

An artificial neuron is a mathematical function conceived as a model of biological neurons. Artificial neurons are the constitutive units in an artificial neural network.

In 1943, McCulloch and Pitts first proposed a simple neuron model (Fig. 2-4), also called McCulloch-Pitts Model, where the dendrites and synapses are independents and there are no effects on them of each other, besides their functions are as mere weight and transmission, and the cell body is treated as the main calculation unit. The McCulloch-Pitts Model receives one or more inputs (representing dendrites) and sums them to produce an output (representing a neuron’s axon). The sums of each node are weighted, and the sum is passed via a nonlinear function known as an activation function or transfer function [30]. The transfer functions usually have a sigmoid shape (Fig. 2-5). In addition, the standard transfer functions is given by

$$OutPut = \frac{1}{1 + e^{\sum w_i x_i - \theta}} \qquad (1)$$

where $w_i$ and $\theta$ are the weights and threshold respectively, and $x_i$ is the input.

Moreover, Minsky and Papert have shown that some rather elementary computations could not be done by one-layer of McCulloch-Pitts cells [31]. Since the computation of dendrites is not considered into this model, the morphology of neurons is the same.

## 2. 3 The Koch-Poggio-Torre model

As mentioned before, the dendrites play a key role in the overall computation of neurons. Synaptic inhibition may veto an excitatory signal which depends on the location and the strength which is from the inhibitory conductance, hence the passive integration of incoming signals the branching pattern and morphology of the dendritic tree are very important (i.e. no local response or dendritic spike occurs) [29]. Koch, Poggio and Torre studied the interpretation of dendritic architecture which is about the processing function of cells and analyzed the interaction of excitatory synaptic input with steady-state shunting inhibitory input in α, β, γ and δ retinal ganglion cells through the cable theory [20]. They have found that nonlinear synaptic interactions are maximal for γ and β cells and relatively weaker for α and β cells, besides, the results instruct δ-like cells strongly as the morphological substratum for directional selectivity in the retina. They also have shown that the logic operations can be connected with the less formal notions of computation used by physiologists to design a model of a retinal ganglion neuron owing a directional selectivity to moving visual inputs.

In Fig. 2-6, it shows δ-cell dendrite with excitatory inputs (●) and inhibitory inputs of the shunting type (■), and its highly branched pattern in terms of logical operations. Because any inhibitory cannot reject only more distal excitations and influence other inputs more proximal to the soma, the activity would be reasonably of the type. The logical relations can be given by

$$(e_1 \text{AND NOT } i_1)\text{OR}(e_2 \text{AND NOT } i_2)\text{OR}$$
$$\{[(e_3 \text{AND NOT } i_3)\text{OR}(e_4 \text{AND NOT } i_4)$$
$$\text{OR}(e_5 \text{AND NOT } i_5)\text{OR}(e_6 \text{AND NOT } i_6)] \text{ AND NOT } i_7\}$$

Moreover, Koch et al. have considered the main problem is the addressing of the correct synapses to the correct synapses to the correct dendrite. And recent research has identified hundreds of neurons, each of which holds a unique shape of dendritic tree. A slight morphological difference would result in great functional variation. Type-specific dendrite morphology has important functional implications in determining which signals a neuron receives and how these signals are integrated [32].

However, there is no clear definition that the dendritic computation can itself provide a constraint for targeting of synaptic inputs at the appropriate locations [29, 33, 34]. And, in the early stages, redundant synapses and dendrites are found in the nervous system, and the unnecessary ones are soon filtered out and the necessary ones are strengthened and fixed, then form the ripened neural network function [35]. The morphologies of dendrites in these neurons remain unclear and there is no effective method to discover them. Manual analysis of neuronal morphology is time-consuming, labor-intensive, and subject to human error and bias [36].

## Figure caption

**Fig. 2-1** The location of neuron [37].
**Fig. 2-2** The Structure of the Neuron [38].
**Fig. 2-3** The transmission of signals among neurons [39].
**Fig. 2-4** McCulloch-Pitts Model
**Fig. 2-5** Sigmoid shape.
**Fig. 2-6** Example of δ-cell dendrite with its highly branched pattern in terms of logical operations form

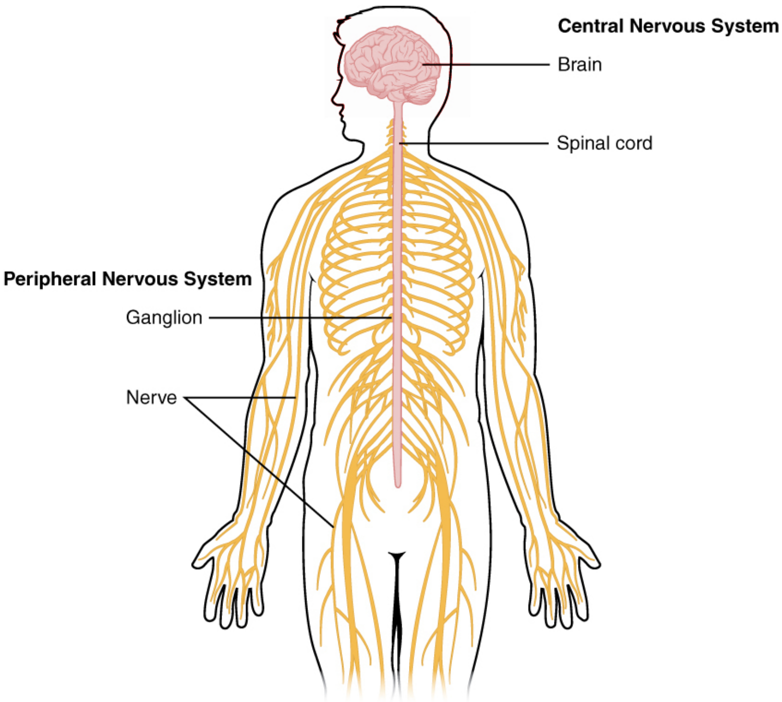



**Fig. 2-1** The location of neuron [37].

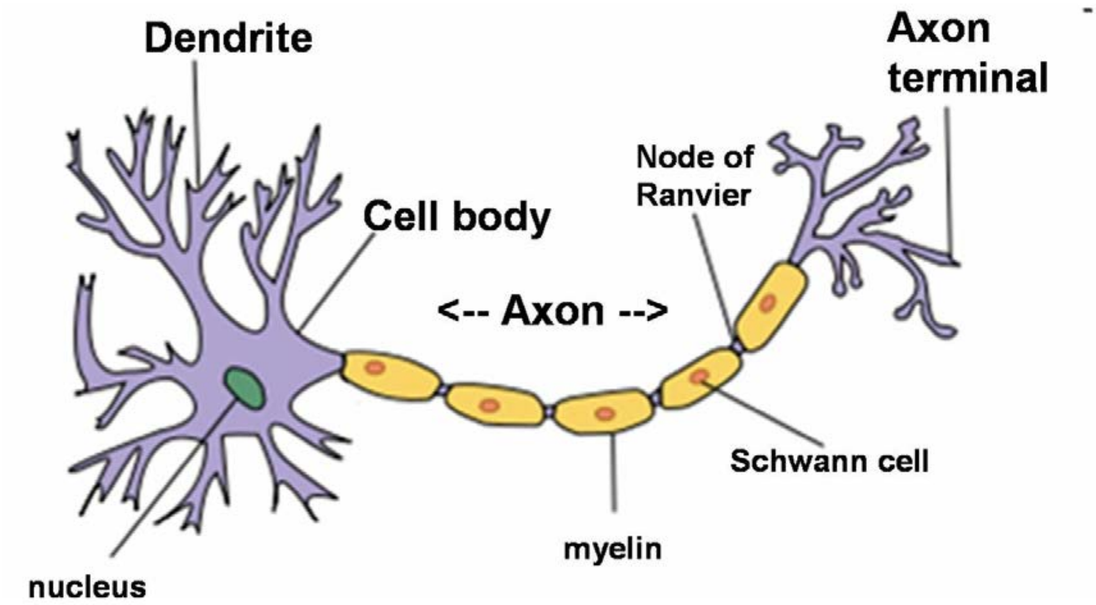



**Fig. 2-2** The Structure of the Neuron [38].

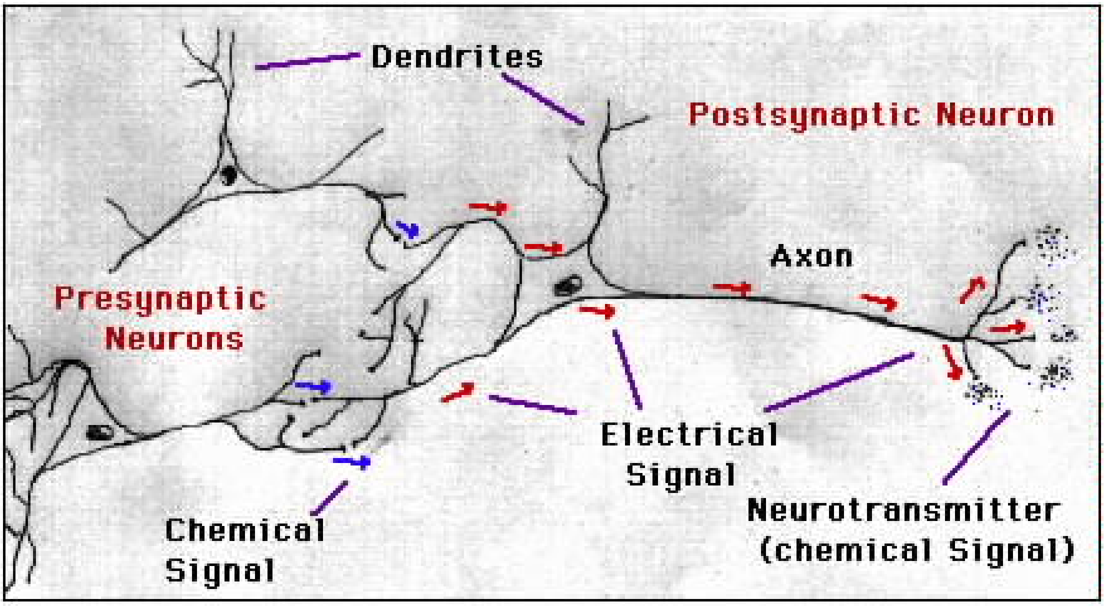



**Fig. 2-3** The transmission of signals among neurons [39].

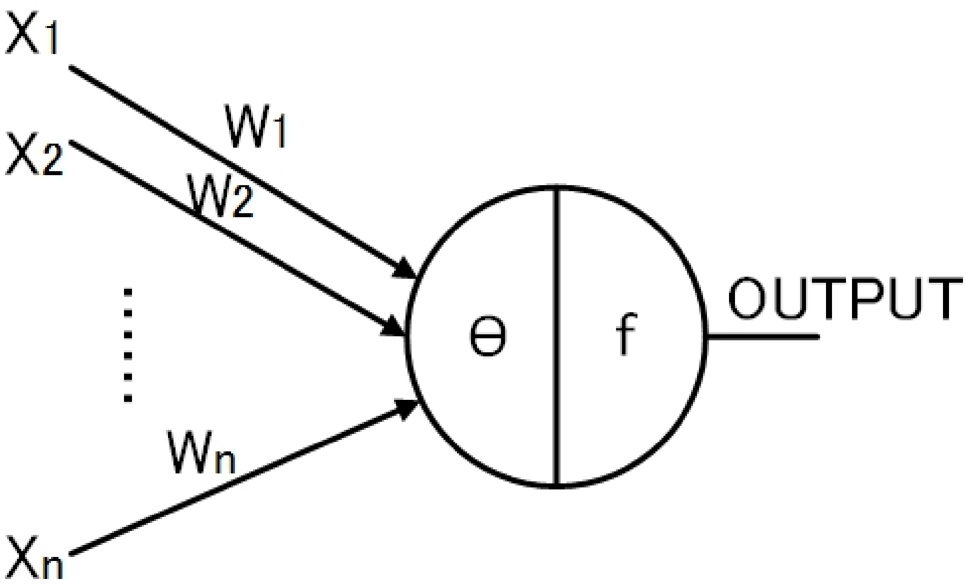



**Fig. 2-4** McCulloch-Pitts Model.

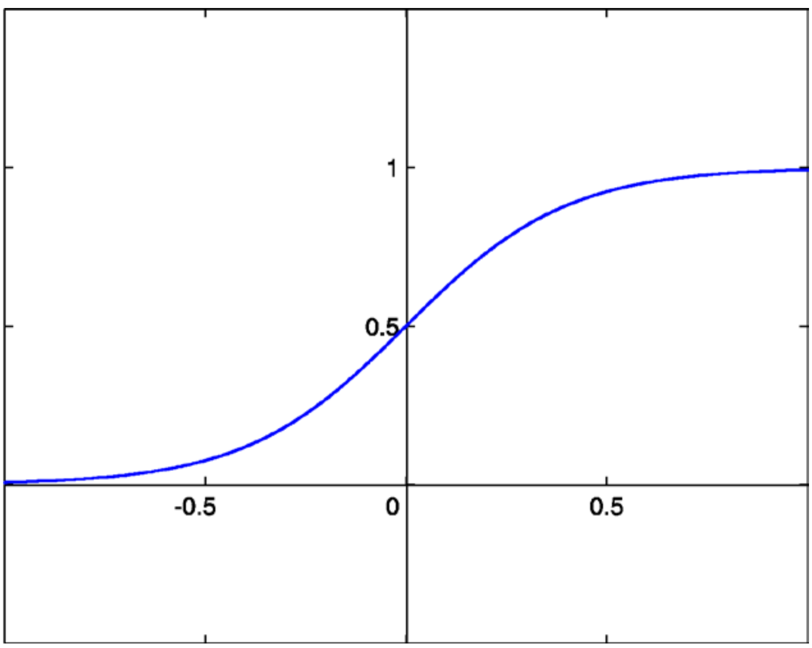



**Fig. 2-5** Sigmoid shape.

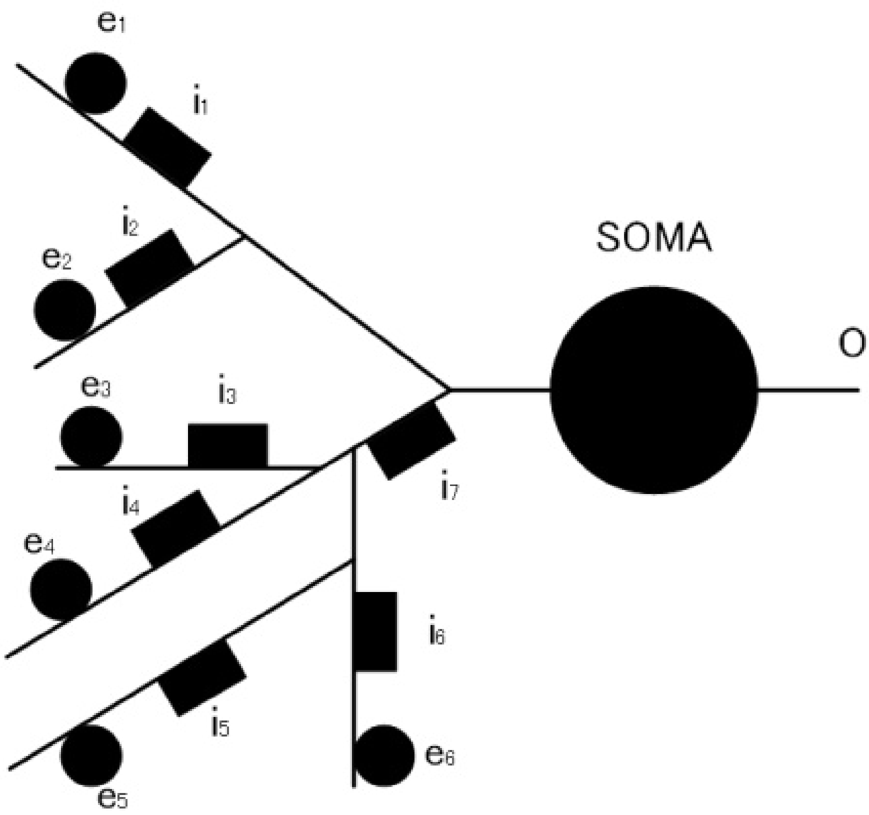



**Fig. 2-6** Example of δ-cell dendrite with its highly branched pattern in term in terms of logical operations.

# Chapter 3

# Neuron Model with Dendritic Nonlinearity

In 2000, Tang et al. proposed a neuron model based on dendritic mechanisms from the conventional Koch-Poggio-Torre model [22]. This model realizes the interaction among synapses and the dendrites and the elimination and generation of synapses to form a ripened dendrites structure holding a special function. In this model, in order to reflect the physiological morphology and the functions of nerve cells, the logical AND (owing to the soft-minimum function) is used to realize the switch function and the logical OR (owing to the Soft-maximum function) is used to deal with when there are two or more inputs and there are switch in parallel. Furthermore, a logical NOT (owing to the sigmoid function of synapse) is also required when a signal is transmitted in the dendrites. Thus, a nonlinear interaction belonging to the dendrites can be expressed by logic operation AND, OR and NOT [22]. In order not to lose generality, a assumption is used that there are nonlinear interactions among all inputs, thus all inputs will connected to all branches initially, and the ripened number of the dendritic branches, the location and the type of synapses on the dendritic branches are all unknown and will be synthesized through learning. This neuron model is successfully trained to learn the directionally selective problem and the depth rotation problem [40-45]. However, owing to the logic operation, there is a key limitation of this neuron model which it can only solve the problem of the binary value smoothly. In order to generalize the computation of neuron model into a continuum of values between 1 and 0, Tang et al. have developed a neuron model with denritic nonlinearity (NMDN) to use the simple multiplication and summation to replace the logical AND and OR function in the neuron model based on dendritic mechanisms, respectively [23]. Not only does the NMDN keep all the function of the original, but also promotes the convergence rate and accuracy.

## 3.1 Model

As shown in Fig. 3-1, NMDN possesses dendritic structure, interaction among synapses and is generalized as following:

1. The arbitrary decision can be used to initialize dendrites.

2. There is an interaction among all synapses on the same branch

3. The nonlinear interaction produced in a dendrite can be expressed by a logical network.

4. The ripened number of the dendritic branches, the location and the type of synapses on the dendritic branches will be synthesized through learning.

In Fig. 3-1, dendritic branches receive signals at the synapses (▲) and perform a simple multiplication on their own signals respectively. At the junction of branches, the outputs of the branches are summed up and conducted to the cell body (soma). When the input of the soma exceeds the threshold, the cell fires to send signal down to the other neurons through axon.

**Synaptic Function:** The synaptic Function is described by sigmoid function. And the output of the synapse whose address is from i-th (i=1, 2,…, n) input to j-th (j=1, 2,…, m) branch is given by

$$Y_{ij} = \frac{1}{1 + e^{-k(w_{ij}x_i - \theta_{ij})}} \tag{2}$$

where $w_{ij}$ and $\theta_{ij}$ are the connection parameters, and k is a positive constant. If k becomes larger, the sigmoid function will turned to be similar with step function.

Due to the values of $w_{ij}$ and $\theta_{ij}$, four kinds of synaptic connections can be defined as following (Fig.3-2):

1. 0-constant connection (⓪) (either $0 < w_{ij} < \theta_{ij}$ or $w_{ij} < 0 < \theta_{ij}$): No matters how the input changes from 0 to 1, the output is always 0.

2. 1-constant connection (①) (either $\theta_{ij} < 0 < w_{ij}$ or $\theta_{ij} < w_{ij} < 0$): No matter how the input changes from 0 to 1, the output is always 1.

3. Excitatory synapse (●) ($0 < \theta_{ij} < w_{ij}$): No matter how the input changes from 0 to 1, the output equals the input.

4. Inhibitory synapse (■) ($w_{ij} < \theta_{ij} < 0$): No matter how the input changes from 0 to 1, the output reverses the input.

The excitatory and inhibitory synapses are real connection states in neurons. However, not all inhibitory synapses and excitatory synapses will exist necessarily in the same branch of an actual dendrite. Therefore, an input is assumed to be initially connected to all branches, but some inputs may not be connected to some branches in truth. For 1-constant connection, it has no influence on the output of multiplication. For 0-constant connection, since the output of the branch is always 0, there is no influence on the summation. In other words, it is equivalent that there is no branch existing. As the values of $w_{ij}$ and $\theta_{ij}$ change, the synaptic connections will emerge the appropriate changes.

**Multiplication Function:** Multiplication Function, as is implied by the name, performs a simple multiplication on various synaptic connections of the branch. The output of j-th branch is given by

$$Z_j = \prod_{i=1}^{n} Y_{ij} \tag{3}$$

**Summation Function:** As mentioned above, the summation on signals sent from the branches is approximated by the following

$$V = \sum_{j=1}^{m} z_j \tag{4}$$

**Soma:** The function of soma can be described as sigmoid operation by following

$$O = \frac{1}{1 + e^{-ksoma(V-\gamma)}} \tag{5}$$

where the $ksoma$ is taken as a positive constant, and the $\gamma$ is taken as a threshold of 0.5.

## 3.2 Learning

Because the functions of NMDN are all differential, thus, the error back-propagation learning rule, supervised learning procedure, is used to the learning procedure. During the learning, the own output vector being produced by the input vector is compared with the target vector. The learning is aimed to reduce the difference of output vector and target vector by modifying $w_{ij}$ and $\theta_{ij}$. Finally, the synapses will converge to their own one of four synaptic connections.

The error between the target vector and the output vector can be expressed by following

$$E = \frac{1}{2}(T - O)^2 \tag{6}$$

where the T is taken as the target, and the O is taken as the output.

In NMDN, modifications are made only to the connection parameters $w_{ij}$ and $\theta_{ij}$ of the connection function during learning. During the learning procedure, these parameters are corrected to decrease the error. If the gradient descent learning method is used to decreasing the value of E, the connection parameters should be corrected as shown in the following formulas:

$$\Delta w_{ij}(t) = -\eta \frac{\partial E}{\partial w_{ij}} \tag{7}$$

$$\Delta \theta_{ij}(t) = -\eta \frac{\partial E}{\partial w_{ij}} \tag{8}$$

where $\eta$ called the learning constant is a positive constant. In this thesis, the $\eta$ is 0.01.

$$w_{ij}(t+1) = w_{ij}(t) + \Delta w_{ij}(t) \tag{9}$$

$$\theta_{ij}(t+1) = \theta_{ij}(t) + \Delta \theta_{ij}(t) \tag{10}$$

where $w_{ij}(t+1)$ and $\theta_{ij}(t+1)$ are the following values of $w_{ij}$ and $\theta_{ij}$ after modified, and $w_{ij}(t)$ and $\theta_{ij}(t)$ are their current value. Thus, the partial differentials of E with respect to $w_{ij}$ and $\theta_{ij}$ can be computed respectively as:

$$\frac{\partial E}{\partial w_{ij}} = \frac{\partial E}{\partial O} \cdot \frac{\partial O}{\partial V} \cdot \frac{\partial V}{\partial z_j} \cdot \frac{\partial z_j}{\partial Y_{ij}} \cdot \frac{\partial Y_{ij}}{\partial w_{ij}} \tag{11}$$

$$\frac{\partial E}{\partial w_{ij}} = \frac{\partial E}{\partial O} \cdot \frac{\partial O}{\partial V} \cdot \frac{\partial V}{\partial z_j} \cdot \frac{\partial z_j}{\partial Y_{ij}} \cdot \frac{\partial Y_{ij}}{\partial w_{ij}} \tag{12}$$

In addition, the momentum and variable learning rate are also introduced into NMDN. The method of momentum optimization involves adding a term to the weight and threshold adjustment. The term is proportional to the amount of the previous weight and threshold changes. The momentum optimization can be represented as:

$$\Delta w_{ij}(t) = -\eta \frac{\partial E}{\partial w_{ij}} + \alpha \Delta w_{ij}(t-1) \tag{13}$$

$$\Delta \theta_{ij}(t) = -\eta \frac{\partial E}{\partial w_{ij}} + \alpha \Delta \theta_{ij}(t-1) \tag{14}$$

where $\alpha$ is a positive constant. In this thesis, it is 0.1. If the weights and threshold are to be changed in the same direction as in the previous step, the rate of changes is increased. Alternatively, if the change in the current step is not in the same direction as that in the previous step, the rate of change is decreased.

In order to accelerate the convergence rate, the method of variable learning rate optimization is introduced. In the gradient ascent algorithm presented above, the steps taken in the direction of the gradient are constant. The variable learning rate can be generalized as:

- If the mean error over the entire training set has decreased, the learning rate need to be increased by multiplying a constant (In this thesis, it is 1.1.).
- If the mean error has increased more than some threshold (In this thesis, it is 1%), the learning rate need to be decreased by multiplying a constant (In this thesis, the value is 0.9).
- If the error is increased less than the threshold, the learning rate remains unchanged.

## Figure caption

**Fig. 3-1** Neuron Model with Dendritic Nonlinearity
**Fig. 3-2** Synaptic connections on different $w_{ij}$ and $\theta_{ij}$.

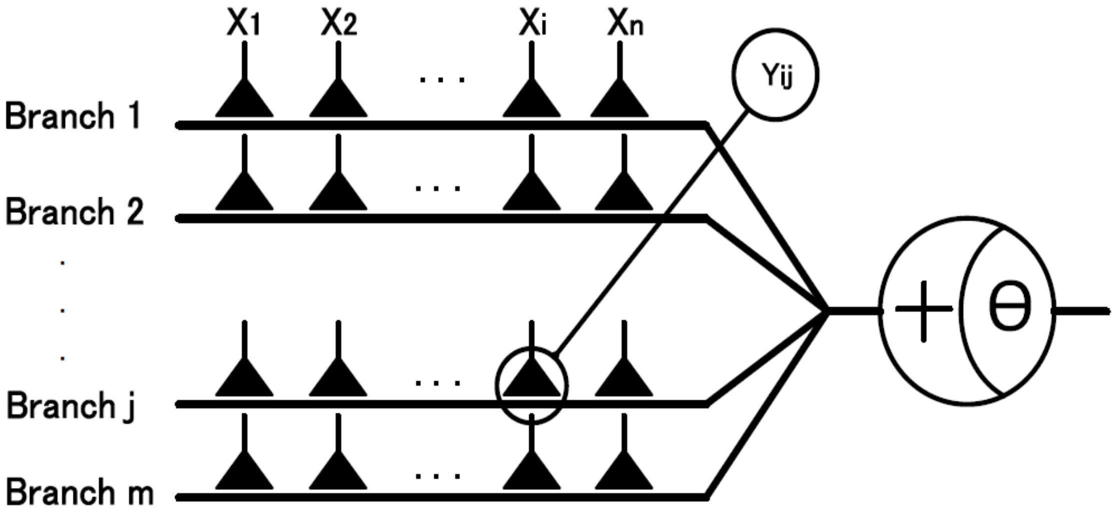
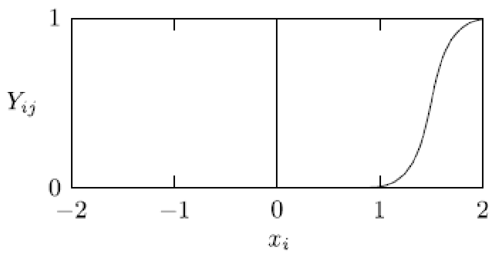
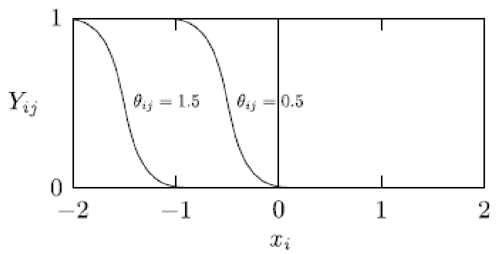


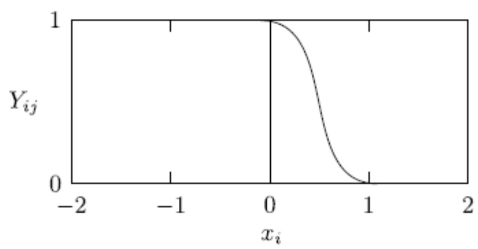

**Fig. 3-1** Neuron Model with Dendritic Nonlinearity



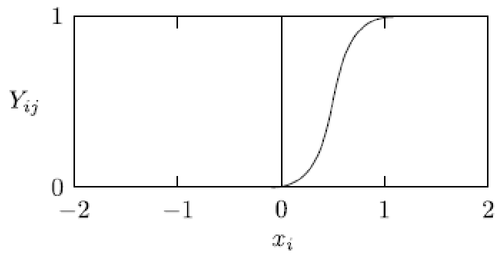

(a) 0-constant connection $0 < w_{ij} < \theta_{ij}$ (e.g. $w_{ij} = 1.0,\ \theta_{ij} = 1.5$).



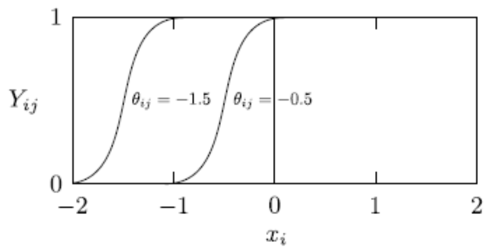

(b) 0-constant connection $w_{ij} < 0 < \theta_{ij}$ (e.g. $w_{ij} = -1.0,\ \theta_{ij} = 0.5$ or $1.5$).



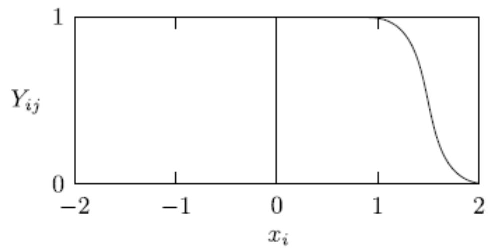

(c) Inversed connection $w_{ij} < \theta_{ij} < 0$ (e.g. $w_{ij} = -1.0,\ \theta_{ij} = -0.5$).




(d) Direct connection $0 < \theta_{ij} < w_{ij}$ (e.g. $w_{ij} = 1.0,\ \theta_{ij} = 0.5$).




(e) 1-constant connection $\theta_{ij} < 0 < w_{ij}$ (e.g. $w_{ij} = 1.0,\ \theta_{ij} = -0.5$ or $-1.5$).




(f) 1-constant connection $\theta_{ij} < w_{ij} < 0$ (e.g. $w_{ij} = -1.0,\ \theta_{ij} = -1.5$).

**Fig. 3-2** Synaptic connections on different $w_{ij}$ and $\theta_{ij}$.

# Chapter 4

# Application on Classification

In the world, breast cancer is ascendant disease, and a lot of women died because of it. At the same time, the women often have a relapse in a high rate [46]. For the reason, the exactly procedure allows doctors to distinguish benign breast tumors from malignant ones, and this is very important to find a right diagnosis ways, but it has high costs of money and society about the forecast of constant breast cancer. Therefore, this problem attracts a lot of researchers working in the Artificial Intelligent (AI) topic.

## 4.1 Related work on Breast Cancer

When ANNs are used in medical diagnosis, they are not affected by factors such as human fatigue, emotional states and habituation. They are capable of rapid identification, analyses of conditions and diagnosis in real time. For the breast cancer, the researchers have used a lots of AI devices, especially the Artificial Neural Network (ANN). There have been researches on medical diagnosis of breast cancer with WBCD using Artificial Neural Networks (ANNs) in literature in recent years.

Local linear wavelet neural networks (LLWNN) have been introduced as a very effective scheme for statistical pattern recognition problem and non-linear complex predictions. MR Senapati et al. have proposed to use recursive least square and firefly algorithm to optimize the training of the LLWNN, the result shows optimization very robust, effective and gives better correct classification [47-49]. In 2012, Chris Ninness et al. have proposed to use Self-Organizing Map (SOM) neural network to apply to diversified nonlinear data distributions in the areas of behavioral and physiological research. The result implicates that the future investigations of the SOM within the behavioral/physiological [50]. In 2011, Marco Vannucci and Valentina Colla describes a novel binary classification method named LASCUS. The method is proposed to train the datasets by SOM and uses the fuzzy inference system to calculate the threshold to determine the data belonging to which cluster. The result of the thesis suggests a good correctly detects in 3 out of 5 tests [51]. Fuzzy and Neural Approaches presents a detailed examination of the fundamentals of fuzzy systems and neural networks and then joins them synergistically, combining the feature extraction and modeling capabilities of the neural network with the representation capabilities of fuzzy systems [3]. In 2013, Manjeevan Seera and Chee Peng Lim have proposed a hybrid intelligent system that consists of the Fuzzy Min-Max neural network to exploit the advantages of the constituent models, the classification and regression tree to explain predicted outputs and the Random forest to achieve high classification performances. They aim to exploit the advantages of the constituent models and at the same time alleviate their limitations. In the experimental, not only can the system produces good results but elucidates its knowledge base with a decision tree [52]. Fuzzy Neural Networks (FNN) comprises an integration of the merits of neural and fuzzy approaches, enabling one to build more intelligent decision-making systems. But increasing the number of inputs causes exponential growth in the number of parameters in Fuzzy Neural Networks (FNN) and computational complexity increases accordingly. So hierarchical fuzzy neural network (HFNN) and Fuzzy Gaussian Potential Neural Network (FGPNN) were proposed for breast cancer detection problem by Somayeh Naghibi et al. in 2011 [53]. And Cheng-Jian Lin and Chi-Feng Wu proposed to use functional neural fuzzy network (FNFN) consisting of Functional link neural networks (FLNNs) and Neural fuzzy networks (NFNs) to deal with the classification [54, 55]. Radial basis function neural network, as an aspect of neural networks, (RBFNN) uses radial basis function as active function. M. R. Senapati et al. proposed a RBFNN for breast cancer detection by extending the application of a variation of particle swarm optimization called K-particle swarm optimization (KPSO) and the technique provides more accurate result and better classification [56]. Recently, to mimic biological neural networks further, one type of artificial neural network which comprises spiking neurons and use action potential (spike) as a computing interface is called a spiking neural network (SNN) and has drawn increasingly more attention from the researchers. Hung-Yi Hsieh proposed a probabilistic spiking neural network (PSNN) with unimodal weight distribution, possessing long- and short-term plasticity. In the experiment, the PSNN is proved to be hardware friendly, the convergence speed fast [57]. Cornelius Glackin et al. presented a supervised training algorithm that implements fuzzy reasoning on a SNN. The experiment provides a rationale for the assembly of biological components such as excitatory and inhibitory neurons, facilitating and depressing synapses, and RFs. In particular, the major contribution is how RFs may be configured in terms of excitation and inhibition to implement the conjunctive AND of the antecedent part of a fuzzy rule [58]. Because search space in artificial neural networks (ANNs) is high dimensional and multimodal which is usually polluted by noises and missing data, the process of weight training is a complex continuous optimization problem. Alireza Askarzadeh and Alireza Rezazadeh aim to deal with the application of a recently invented metaheuristic optimization algorithm, bird mating optimizer (BMO), for training feed-forward ANNs. The simulation results indicate the superior capability of BMO to tackle the problem of ANN weight training [59]. L.M. Sasu et al. proposed to Bayesian ARTMAP (BA) to analyze the efficiency of regression problems. And (i) they generalize the BA algorithm using the clustering functionality of both ART modules, and name it BA for Regression (BAR); (ii) they prove that BAR is a universal approximator with the best approximation property [60]. Ahmad Taher Azar et al. used probabilistic neural networks (PNN) to apply for the purpose of detection and classification of breast cancer. The results reveal that PNN is a quite good classifier by achieving accuracy rates of 100 and 97.66 % in both training and testing phases, respectively [61]. A. A. Kalteh et al. presented a novel hybrid intelligent method for detection of the breast cancer patterns. The proposed method includes two main modules: clustering module and the classifier module. In the clustering module, the input data will be clustered by a technique being a suitable combination of the modified imperialist competitive algorithm (MICA) and K-means algorithm (K-MICA algorithm). And in the classifier module, several neural networks, such as the multilayer perceptron, probabilistic neural networks and the radial basis function neural networks are investigated. The result reveals that K-MICA algorithm and RBF neural networks hold the highest performance [62]. Mustafa Serter Uzer et al. suggested to using SFSP or SBSP combining with SCG NN to find a solution for the problem of breast cancer. By the two hybrid methods (SFSP and SBSP), the dimension of the feature space for input has been decreased from 9 to 5 thanks to the selection of these two hybrid features, and the performance of SBSP + NN is much better and the five types of inputs are the Clump thickness, the Uniformity of cell size, the single epithelial cell size, the bare nuclei and the normal nucleoli, respectively [63]. Brijesh Verma et al. proposed two types of hybrid combination Parallel Neural-based Clusters Fusion (PNCF) and Parallel Neural-based Strong Clusters Fusion

(PNSCF) which are a novel hybrid ensemble approach for classification which can be defined as a process of combining various algorithms and techniques in such a way that it can utilize the strengths of each individual technique and compensate for each other's weaknesses [64]. A. Marcano et al. presented to use the biological metaplasticity property of neurons and Shannon's information theory to the multilayer perceptron (MLP). This model defines artificial metaplasticity as a learning procedure that produces greater modifications in the synaptic weights with less frequent patterns than frequent patterns, as a way of extracting more information from the former than from the latter [65].

As showed of the earlier works of breast cancer using ANN above, much effort has been devoted over the past several decades to the development and improvement of pattern classification models for breast cancer detection. Three trends of researches of classification models using ANN can be extracted: (i) propose the new type neural network model; (ii) using new training algorithms; (iii) propose hybrid system which using the ANN as the classifiers. Owing to the further application of cancer detection, not only do the patients or doctors need to know the classification result, but to know the symptoms that derive this result. The traditional neural networks and most of hybrid system have almost obtained high classification. However, there is a black box on their decision process, with no explanation as to how the decisions are attained [52, 66]. Hybrid heuristic methods like GA or neural networks combining with fuzzy rules will handle this problem caused by black box approaches, but it is also a key limitation of identifying which input factors are more significant than the others [52, 67].

## 4.2 Wisconsin breast cancer database overview

**Table 4-1** WBCD description of attributes

| Attribute Numeber | Attribute Description | Value of attributes |
|---|---|---|
| 1 | Clump thickness | 0.1-1 |
| 2 | Uniformity of cell size | 0.1-1 |
| 3 | Uniformity of cell shape | 0.1-1 |
| 4 | Marginal adhesion | 0.1-1 |
| 5 | Single epithelial cell size | 0.1-1 |
| 6 | Bare nuciei | 0.1-1 |
| 7 | Bland chromatin | 0.1-1 |
| 8 | Normal nucieoll | 0.1-1 |
| 9 | Mitoses | 0.1-1 |

In this study, the Wisconsin Breast Cancer Database (WBCD) (UCI Machine Learning Repository) which is from Fine Needle Aspirates (FNA) is used for datasets of classifiers. The datasets comprise 699 samples and each record has nine attributes which are showed in Table 4-1. We designate the mensuration as value between 0.1 and 1, and 0.1 is closed to benign and 1 is the most anaplastic. Either benign or malignant, this dataset has 16 samples with missing attribute values. Because these data instances are rejected by some classification algorithms and we use the same method which uses 683instances in order to compare. So there are 444 (65.0%) benign instances and 239 (35.0%) malignant instances.

## 4.3 Simulation parameters

Table 4-2 shows the Mean squared error (MSE), epoch of learning (Epochs), learning constant, parameters of NMDN and numbers of patterns used in the training and testing phases. And in the test samples, the number of benign samples is 133, the malignant is 72. In order to compare the performance between NMDN and BPNNs, we carry out an experiment with Wisconsin breast cancer database. They are both implemented in MATLAB 2013b (MATLAB Neural Network Toolbox, version 8.1 is for BPNNs) and on an intel Core-i5 computer of 3.4 GHz with 8 GB of RAM.

**Table 4-2** Simulation parameters applying to WBCD.

| Types of classifiers | MSE | Epochs | Learning constant | NMDN parameters | | Number of patterns | |
|---|---|---|---|---|---|---|---|
| | | | | k | theta | Train | Test |
| BPNNs | 0.01 | 1000 | 0.01 | NA* | NA* | 478 | 205 |
| NMDN | 0.01 | 1000 | 0.01 | 3 | 0.5 | 478 | 205 |

*NA: not applicable

In the simulation, for the purpose of analysis, different quantities of branches are chosen. If the number of inputs is N and the number of branch is M, the quantity of modifying $w_{ij}$ and $\theta_{ij}$ of the NMDN is $2M \times N.$ Moreover, if the number of inputs is N and the number of node of hidden layer is M, the quantity of modifying weights and threshold is $M \times N + M + 1.$ Table 4-3 is the structure of the BPNNs and NMDN.

In addition, the sigmoid function is chosen as the node function of the BPNNs. For either NMDN or BPNNs, in order to analyze accurately and reflect their performance completely, the simulation will run 100 times with different random initial values on each condition.

**Table 4-3** The structures of NMDN and BPNNs.

| | NMDN | BPNNs | |
|---|---|---|---|
| Input | Branch | HL* | Output |
| 9 | 5 | 8 | 1 |
| | | 9 | |
| | 10 | 16 | |
| | | 17 | |
| | 15 | 24 | |
| | | 25 | |
| | 20 | 32 | |
| | | 33 | |
| | 25 | 40 | |
| | | 41 | |

*HL: the node of Hidden layer

## 4.4 Performance on the accuracy

Classification accuracy: In this simulation, classification accuracy for the data sets is measured using the equation:

$$accuracy = \frac{TP + TN}{TP + TN + FP + FN}\% \tag{12}$$

where TP, TN, FP, and FN denote true positives, true negatives, false positives, and false negatives, respectively.

True positive (TP): An input is detected as malignant and the teacher targets also label so.

True negative (TN): An input is detected as benign as diagnosed by the teacher targets.

False positive (FP): An input is detected as malignant, although the teacher targets show the opposite.

False negative (FN): An input is detected as benign breast cancer, although the teacher targets label is malignant.

We use the following expressions for sensitivity and specificity analysis,

$$Sensitivity = \frac{TP}{TP + FN}\% \tag{13}$$

$$Specificity = \frac{TN}{TN + FP}\% \tag{14}$$

The average sensitivity and average specificity listed in the Table 4-4 belong to average performance of the BPNNs and NMDN, and the average accuracy is shown in Table 4-5.

As mentioned above, in order to compare the performance, in Table 4-4 and Table 4-5, the NMDN and BPNNs are chosen the similar quantity of modification, and generated 100 runs separately. The following analysis of the other sections is also under the same condition. As shown in the Table 4-4 and Table 4-5, NMDN is superior to classical BPNNs training in most of the cases especially on the average accuracy. And, the sensitivity is lower than specificity and it shows that the detection of malignant is more difficult than benign, indeed. Additional, with the growth of quantity of branch in NMDN and node of the hidden layer in BPNNs, the accuracy raises higher, too. In Table 4-4, the average sensitivity of NMDN raised from 50.7778% to 95.3611%, and the BPNNs' raised from 61.1389% to 81.1389%. The increment of NMDN is 44.5833%, and the BPNNs' is 20%. However, there is basically no growth on the average specificity of NMDN and BPNNs. That means growth of sensitivity of NMDN raises faster than BPNNs', and the growth of accuracy depends on the growth of sensitivity. However, when the branch number of NMDN is 5, the average accuracy of NMDN is lower than the BPNNs. And from the branch is 10, the accuracy of NMDN is better than BPNNs. In [35], redundant synapses and dendrites are found in the nervous system at the early stages, and the unnecessary ones are soon filtered out and the necessary ones are strengthened and fixed, then form the ripened neural network function. So, an assumption can be proposed that there is some connection with the quantity of redundant synapses and branch of initial states.

**Table 4-4** The average sensitivity and specificity of NMDN and BPNNs on WBCD

| Branch/ Hidden Layer node | | The average Sensitivity | | The average Specificity | |
|---|---|---|---|---|---|
| NMDN | BPNNs | NMDN | BPNNs | NMDN | BPNNs |
| 5 | 8 | 50.7778 | 61.1389 | 99.0301 | 98.5789 |
| | 9 | | 63.0139 | | 93.8271 |
| 10 | 16 | 71.1687 | 65.2361 | 98.7594 | 97.5263 |
| | 17 | | 69.1528 | | 97.4436 |
| 15 | 24 | 87.7917 | 71.4861 | 98.8195 | 97.5789 |
| | 25 | | 77.6528 | | 94.6316 |
| 20 | 32 | 91.2778 | 77.7500 | 98.7218 | 97.5639 |
| | 33 | | 79.6250 | | 96.4337 |
| 25 | 40 | 95.3611 | 81.1389 | 99.1203 | 98.4962 |
| | 41 | | 80.4444 | | 97.3233 |

**Table 4-5** Classification accuracies of classifiers used for detection of breast cancer.

| Branch/ Hidden Layer node | | The average accuracy | | The best accuracy | |
|---|---|---|---|---|---|
| NMDN | BPNNs | NMDN | BPNNs | NMDN | BPNNs |
| 5 | 8 | 82.0829 | 85.4293 | 98.5366 | 99.0244 |
| | 9 | | 83.0049 | | 99.0244 |
| 10 | 16 | 89.0683 | 86.1854 | 99.5122 | 98.5366 |
| | 17 | | 87.5037 | | 98.5366 |
| 15 | 24 | 94.9463 | 88.4164 | 99.5122 | 98.0448 |
| | 25 | | 88.6683 | | 99.0244 |
| 20 | 32 | 96.1073 | 90.6049 | 99.5122 | 98.5366 |
| | 33 | | 90.5561 | | 98.5366 |
| 25 | 40 | 97.8000 | 92.4000 | 100 | 98.5366 |
| | 41 | | 91.3951 | | 98.5366 |

## 4.5 Performance on convergence rate

The convergence rate is one of the most prominent evaluation strategies of the classifiers' performance. In this simulation, we suppose the accuracy whether reach 90% as the foundation of the convergence. Furthermore, the convergence rate is shown with the epochs. The test condition is:

1, if accuracy≥ 90%, stop training and return the epochs;

2, if epochs = 1000, stop training.

**Table 4-6** The average epochs of NMDN and BPNNs.

| NMDN | | BPNNs | |
|---|---|---|---|
| Branch | Average epochs | HL node | Average epochs |
| 5 | 123.09 | 8 | 207.80 |
| | | 9 | 203.44 |
| 10 | 112.92 | 16 | 226.66 |
| | | 17 | 275.86 |
| 15 | 133.10 | 24 | 310.21 |
| | | 25 | 303.28 |
| 20 | 123.45 | 32 | 343.11 |
| | | 33 | 384.59 |
| 25 | 63 | 40 | 337.5 |
| | | 41 | 395.87 |

As shown in Table 4-6, in each condition, the NMDN's average epoch of convergence is fewer than the corresponding BPNNs'. And, with the growth of calculation, on the whole, the average epoch of NMDN turns to decrease. In the opposite, the BPNNs' turns to increase. The reasons may be include as:

1: The structure of the NMDN. Because, there are 4 types of synaptic connection, with the growth of the branch, the patterns types of the connections will increase in a geometric ratio on the whole. In the other word, after training, the ideal shape of dendrites may be formed more easily.

2: To the BPNNs, there are no conceptions about the structure after training, so the more Hidden Layer nodes, the more hard to form a stable structure.

Especially, when the branch number of NMDN is 25, the average epoch is 63. However, the corresponding BPNNs' average epoch is 337.5 and 395.87. In other words, the average convergence epoch of BPNNs is 5.35 and 6.28 times that of

NMDN's, respectively. As shown above, when branch number of NMDN is 25, the accuracy and convergence performance better than others. So the following comparison with the NMDN and BPNNs will be focused in this condition.

In addition, Fig. 4-1 shows the performance of the average mean squared error convergence curve of 100 experimental repetitions respectively on different branches of NMDN. As shown in Fig. 6, the convergence rate performs better than others on the 25 branches. So, the following comparison will be focused between the 25 branches of NMDN and its corresponding BPNNs.

$$Mean\ Error = \frac{1}{R}\sum_{a=1}^{R}[\frac{1}{2}\sum_{b=1}^{S}(E_b - O_b)] \tag{15}$$

Moreover, as shown in Fig. 4-2, it is the average mean squared error convergence curve of 100 experimental repetitions respectively when the NMDN's dendritic branch is 25 and the BPNNs' hidden layer nodes are 40 and 41. And, the mean error is defined as Eq. 15. R and S are defined as the number of experimental repetitions and the training data. It is clearly that the convergence rate of NMDN is higher than BPNNs.

## 4.6 Performance on stability

The stability is another prominent evaluation strategies of the classifiers' performance. In this simulation, we use the initial and final errors to judge the stability.

As shown in Fig. 4-3, the initial error of NMDN shows that there are just a few of light fluctuations and the initial error is approximately from 0.13 to 0.1. However, the BPNNs' range of variation is much huger than NMDN, and frequency is higher. In Fig. 4-4, the finial errors of NMDN converge to less than 0.05 and more stable than BPNNs. And, the BPNNs can fall into the local minimum more easily than NMDN by the final error. It indicates that the NMDN is more stable than BPNNs.

Furthermore, the data set of final errors is analyzed statistically by generating a box and whiskers plot [68] in Fig. 4-5. Integrating with the Fig. 4-5 and Table 4-7 and Table 4-8, once again explains that the NMDN is more stable than BPNNs. And owe that all the outliers of NMDN are mild outliers and the NMDN also is trapped into local minimum, combing with Table 4-5, the classification accuracy is high. That is to say, NMDN is surmised to possess the inhibition of the local minimum problem.

**Table 4-7** The important values of box and whisker for final errors in NMDN and BPNNs.

| | NMDN-25 | BPNNs-40 | BPNNs-41 |
|---|---|---|---|
| Upper whisker | 0.037768 | 0.12785 | 0.12962 |
| Upper Quartile (Q3) | 0.031862 | 0.087018 | 0.088252 |
| Median | 0.029425 | 0.067329 | 0.068585 |
| Lower Quartile (Q1) | 0.027771 | 0.057607 | 0.058141 |
| Lower whisker | 0.025298 | 0.038776 | 0.040378 |
| Upper outer fence | 0.044135 | 0.175251 | 0.178585 |

**Table 4-8** The outliers of box and whisker for final errors in NMDN and BPNNs.

| NMDN-25 | BPNNs-40 | BPNNs-41 |
|---|---|---|
| 0.038082 | 0.131893 | 0.144542 |
| 0.038268 | 0.132584 | 0.174069 |
| 0.038609 | 0.136827 | 0.24886 |
| 0.03867 | 0.142562 | 0.315718 |
| 0.038904 | 0.171905 | 0.331993 |
| 0.039414 | 0.185008 | 0.652302 |
| 0.042752 | 0.314692 | 0.668047 |
| 0.042648 | 0.648997 | |

## 4.7 Performance on AUC

One of the method of evaluating classifier performance was used in calculating the Receiver Operating Characteristic (ROC) curve which is a two dimensional measure of classification performance, widely used in biomedical research to assess the performance of diagnostic tests [69, 70]. In order to compare the performance of classifiers and reduce ROC performance to a single scalar value representing expected performance, a common method is to calculate the area under the ROC curve (AUC) [71]. When an AUC is close to 1, it indicates that the classifier is a very reliable diagnostic test [72]. The AUC can be computed by integrating the area under the ROC curve (summing the areas of trapezoids) or by the Mann-Whitney-Wilcoxon test statistic [73, 74]. In this simulation, the average AUC values of 100 runs of NMDN and BPNNs are listed in Table 4-9. Moreover, the ROC curve is also shown in Fig. 4-6

**Table 4-9** The AUC of NMDN and BPNNs.

| NMDN | | BPNNs | |
|---|---|---|---|
| Branch | AUC | HL nodes | AUC |
| 25 | 0.9741 | 40 | 0.9314 |
| | | 41 | 0.9151 |

As showed in Table 4-9, the average of AUC of NMDN is 0.9741, and the BPNNs' are 0.9314 and 0.9151. It indicates that the NMDN classifier is more efficient than BPNNs. This indicates once again the superiority of NMDN over BPNNs, in this particular case.

## 4.8 Performance on the Dendrite Morphology

As mentioned above, the dendrite morphology of NMDN is one of the most the characteristic in the neuron model's study. Following, the accurate prediction of NMDN's dendrite morphology is presented.

In NMDN, due to the values of $w_{ij}$ and $\theta_{ij}$ of sigmoid in synaptic function, four types of synaptic connections can be defined. The excitatory and inhibitory synapses are real connection states in neurons. After learning, if 0-constant connection exists in some branch, the output of the branch will be 0 (owing to the multiplication among synapses in one branch). In other words, the branch equals to be filtered out through learning. As Fig. 4-7 shows that it is one of the dendrite morphology of NMDN holding best accuracy after eliminating the 0-constant existing branches.

Furthermore, after learning, if 1-constant connection exists in some branch, the output of the corresponding synapse will be 1. However, in a branch, the interaction among synapses is given by a simple multiplication. Thus, for 1-constant connection, there is no influence on the output of the branch. So, the 1-constant connection can be eliminated in branches. As shown in Fig. 4-8, it is the ripened dendrite morphology of NMDN after learning. And, there are only the excitatory synapses existing as the effective ones in the ripened dendrite morphology. In addition, the performance of classification on AUC doesn't have changed after eliminating branches and synapses in Fig. 4-9.

As mentioned above, the NMDN is developed for continuum of values between 0 and 1. Thus, the ripened dendrite morphology can also be given by

$$V = x_1 \cdot x_3 \cdot x_8 + x_3 \cdot x_6 \cdot x_7 = x_3(x_1 \cdot x_8 + x_6 \cdot x_7) \tag{16}$$

Integrated Table 4-1 and Eq. 15, the Clump thickness, Uniformity of cell shape, Normal nucieoll, Bare nuclei, and Bland chromatin can be extracted as critical factors of breast cancer detection.

As the related work showed above, in [52], owing to the capability of decision tree, the uniform of cell shape, Bare Nuclei, Bland Chromatin and Clump Thickness can be collected as the critical attributes of detecting breast cancer. However, in [63], the Clump thickness, the Uniformity of cell size, the single epithelial cell size, the bare nuclei and the normal nucleoli are selected as more probability to promote the accuracy of detection. In other words, these critical factors in [63] possess the first probability to determine whether the breast cancer is benign or malignant. To compare our simulation result with the ones gotten from [52] and [63], we can't define which one is more reliable. However, the Clump thickness and Bare Nuclei may be surmised as important factors to detect the breast cancer, because, in our simulation, [52] and [63], these two attributes exist. Furthermore, another two attributes, marginal adhesion and mitoses, can be proposed to be removed from the detection items in the future. Because, these two attributes are not involved in the results of our simulation, [52] and [63].

Furthermore, the NMDN is for the first time to be used into real world problem and continuum values problem. Although, as shown above, there is no influence on eliminating 0-existing branch and 1-existing synapses at the best performance of NMDN, the further confirmation of the influence on eliminating branch and synapse has been introduced. Fig. 4-10 shows that after eliminating the 1-existing synapses, the average ROC of 100 repeat experiments decreases. That is to say, it can be surmised as a criteria which if a branch was eliminated, whether there is some influence on the classification that the 0-connection existing in the branch or not in the continuum values.

## Figure caption

**Fig. 4-1** The convergence rate on NMDN.

**Fig. 4-2** The convergence rate on NMDN and corresponding BPNNs

**Fig. 4-3** The initial Error of NMDN and BPNNs

**Fig. 4-4** The final Error of NMDN and BPNNs

**Fig. 4-5** The box and whisker for final errors in NMDN and BPNNs

**Fig. 4-6** The average ROC curves and AUC of NMDN and BPNNs

**Fig. 4-7** The Dendrite Morphology of NMDN without the eliminated branches after learning

**Fig. 4-8** The Dendrite Morphology of NMDN after learning.

**Fig. 4-9** The ROC and AUC of NMDN holding best performance on eliminating branches.

**Fig. 4-10** The average ROC and AUC of NMDN on eliminating branches and synapses.

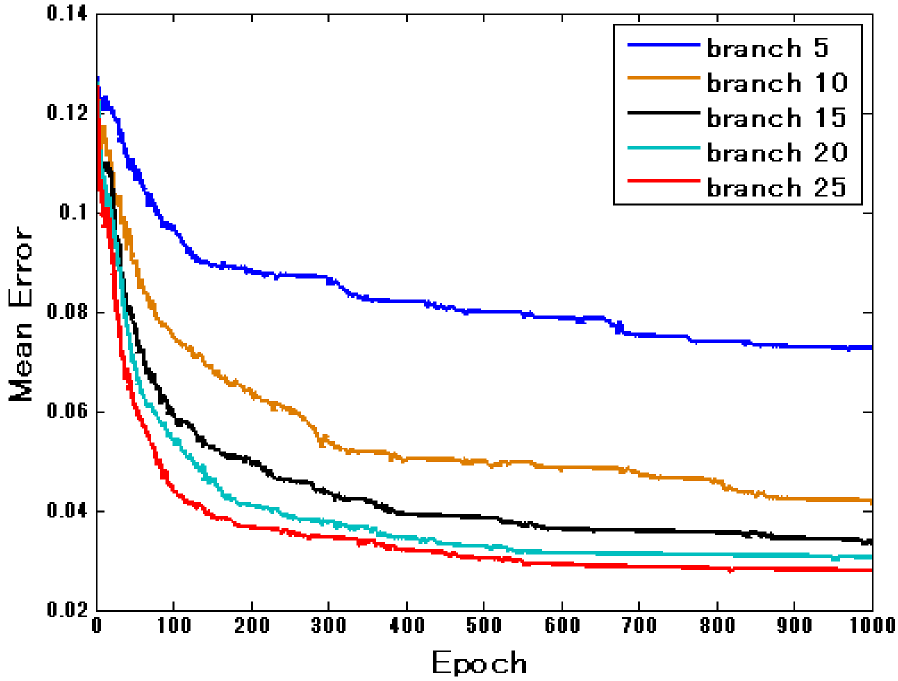



**Fig. 4-1** The convergence rate on NMDN.

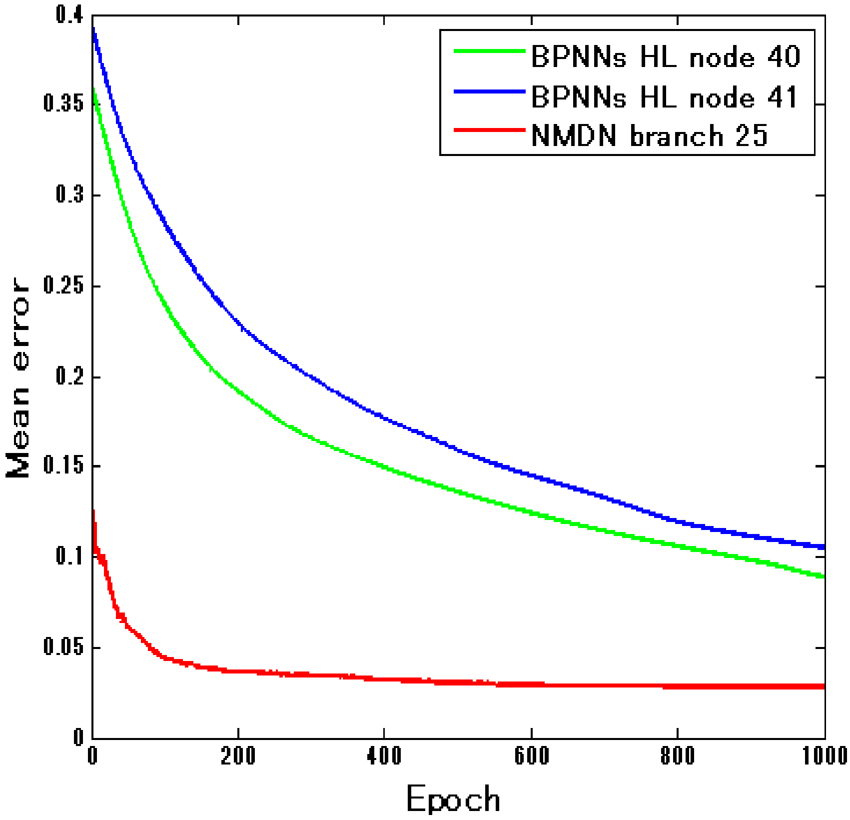



**Fig. 4-2 The** convergence rate on NMDN and corresponding BPNNs

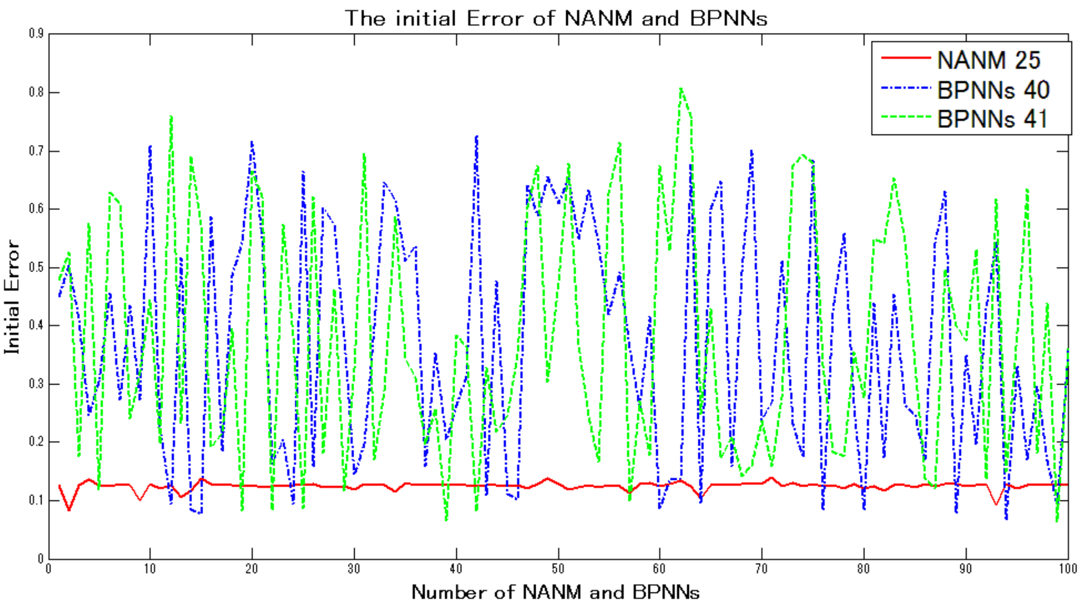



**Fig. 4-3** The initial Error of NMDN and BPNNs

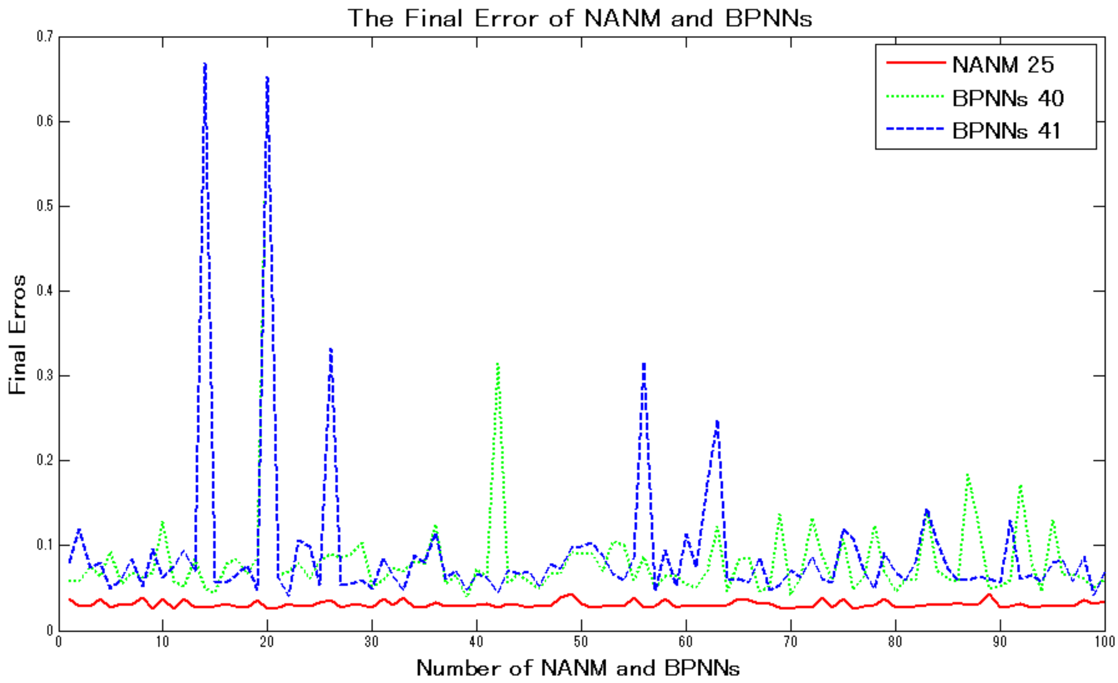



**Fig. 4-4** The final Error of NMDN and BPNNs

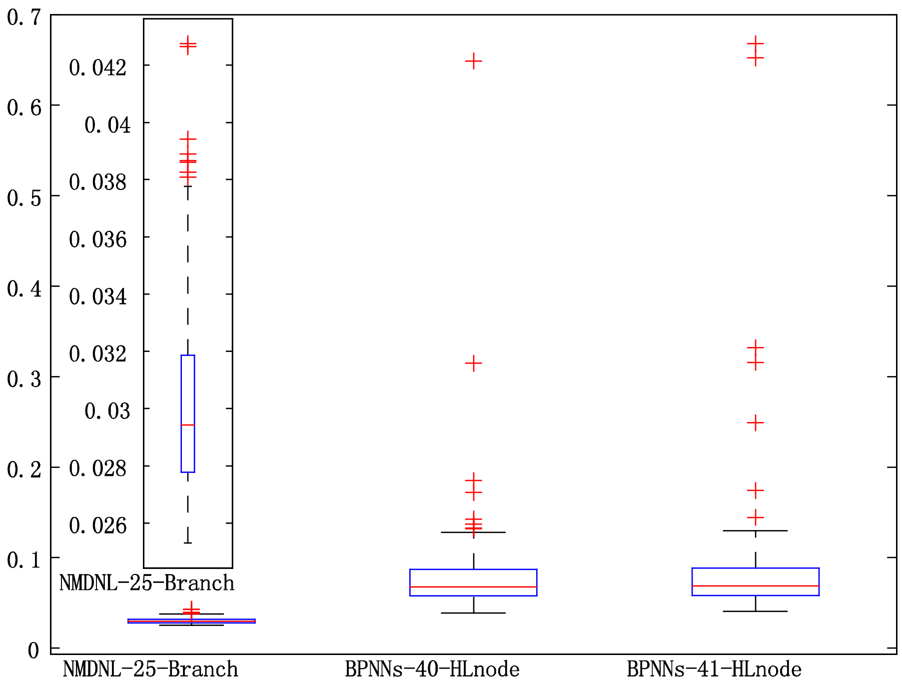



**Fig. 4-5** The box and whisker for final errors in NMDN and BPNNs

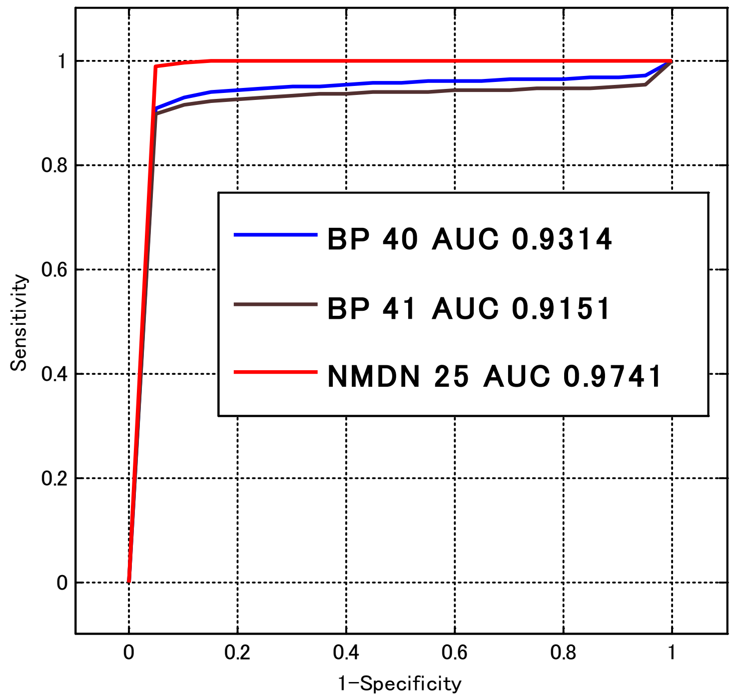



**Fig. 4-6** The average ROC curves and AUC of NMDN and BPNNs

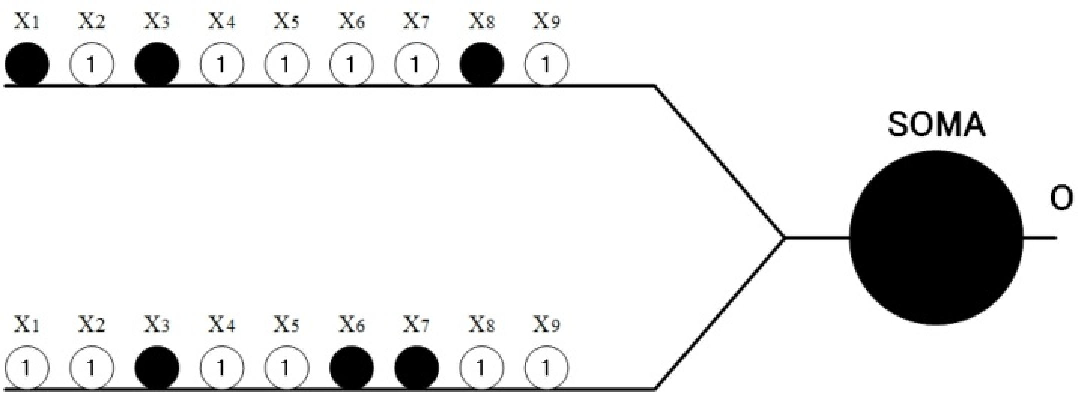



**Fig. 4-7** The Dendrite Morphology of NMDN without the eliminated branches after learning

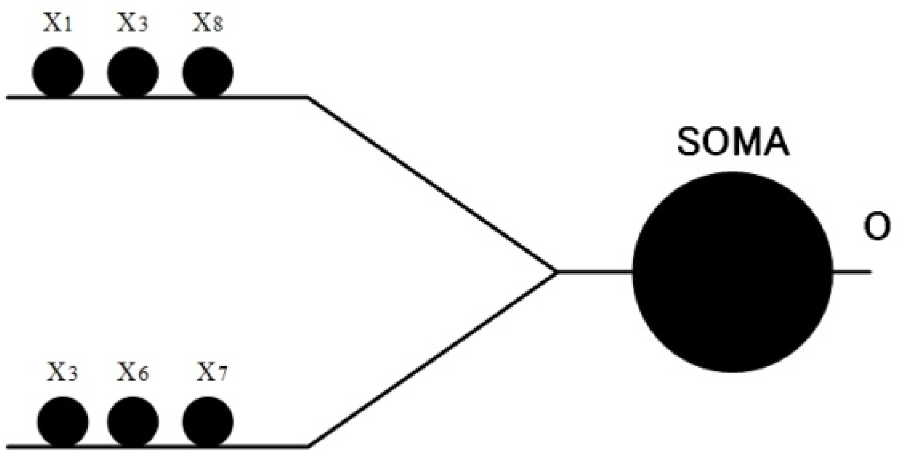



**Fig. 4-8** The Dendrite Morphology of NMDN after learning.

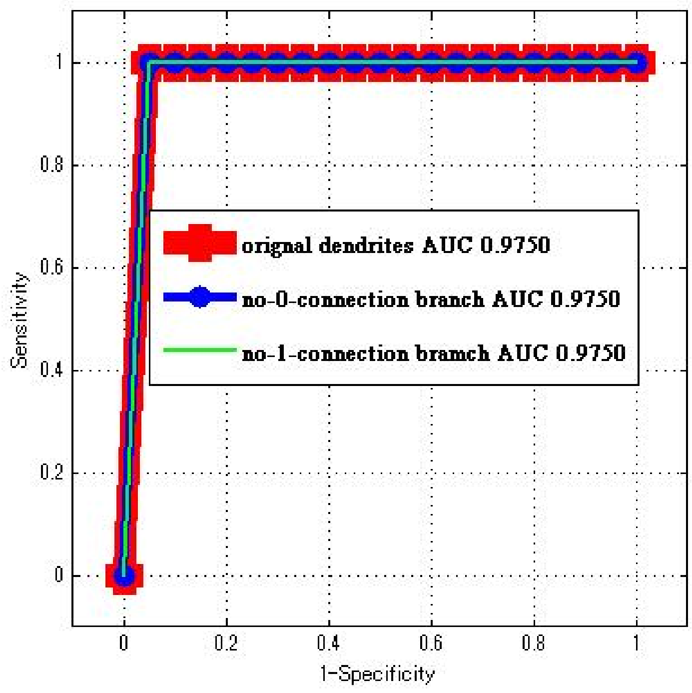



**Fig. 4-9** The ROC and AUC of NMDN holding best performance on eliminating branches.

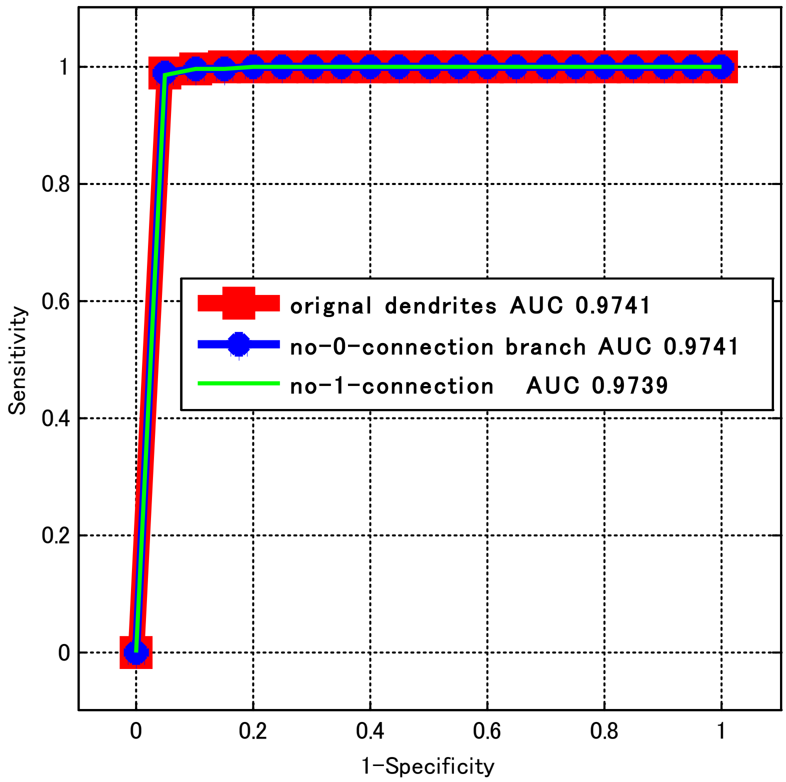



**Fig. 4-10** The average ROC and AUC of NMDN on eliminating branches and synapses.

# Chapter 5

# Application on Prediction

## 5.1 Related work on overreaction in stock market

### 5.1.1 Market Efficiency Theory

The market efficiency theory states that, “one cannot consistently achieve returns in excess of average market returns on a risk-adjusted basis, given the information available at the time the investment is made” [75].

Historically, there was a very close link between the market efficiency theory and the random walk hypothesis and then the Martingale model. The random character of stock market prices was first modeled by Jules Regnault, a French broker, in 1863 and then by Louis Bachelier, a French mathematician, in his Ph.D. thesis in 1900 [76]. His work is almost ignored until 1950s, however the independent work has proved it from 1930s. In his thesis, a small number of studies indicated that US stock prices and related financial series followed a random walk model.

Professor Eugene Fama, who is as an academic concept of study and has worked at University of Chicago Booth School of business, has developed the efficient-market hypothesis through his published Ph.D. thesis in the early 1960s at the same school. When behavioral economists became the mainstream, the hypothesis was widely accepted up until the 1990s, which had been a fringe element [77]. Empirical analyses have consistently found problems with the efficient-market hypothesis, the most consistent being that stocks with low price to earnings outperform other stocks [78-81]. Alternative theories have proposed that cognitive biases cause these inefficiencies, leading investors to purchase overpriced growth stocks rather than value stocks [77]. Although the efficient-market hypothesis has become controversial because substantial and lasting inefficiencies are observed, Beechey et al. (2000) consider that it remains a worthwhile starting point [82].

In the mid-1960s, the efficient-market hypothesis as a prominent theory emerged. Paul Samuelson had begun to circulate Bachelier's work among economists. In 1964, Bachelier's dissertation along with the empirical studies which were mentioned above were published in an anthology edited by Paul Cootner [83]. In 1965, Eugene Fama published his dissertation arguing for the random walk hypothesis [84]. Also, Samuelson published a proof showing that if the market is efficient prices will show random-walk behavior [85]. This is often cited in support of the efficient-market theory, by the method of affirming the consequent [86, 87], however in that same paper, Samuelson warns against such backward reasoning, saying “From a no-empirical base of axioms you never get empirical results” [88]. In 1970, Fama published a review of both the theory and the evidence for the hypothesis. The paper extended and refined the theory, and there are three forms of financial market efficiency which are weak, semi-strong and strong in the definitions.[89]. It has been argued that the stock market is “micro efficient” but not “macro efficient”. Samuelson, who has asserted that the EMH is more suited for individual stocks than it is for the aggregate stock market, pointed the main proponent. Samuelson's dictum has been strongly supported by research based on regression and scatter diagrams [90].

Further to this evidence that the UK stock market is weak-form efficient, other studies of capital markets have pointed toward their being semi-strong-form efficient. The grain futures market which is from Khan indicated semi-strong form efficiency following the release of large trader position information (Khan, 1986). Studies by Firth (1976, 1979, and 1980) in the United Kingdom have compared the share prices existing after a takeover announcement with the bid offer. Firth found that the share prices were fully and instantaneously adjusted to their correct levels, thus concluding that the UK stock market was semi-strong-form efficient. However, the ability of market to efficiently respond to a short term, widely publicized event such as a takeover announcement does not necessarily prove market efficiency related to other more long term, amorphous factors. David Dreman has criticized the evidence provided by this instant efficient response, pointing out that an immediate response is not necessarily efficient, and that the long-term performances of the stock in response to certain movements are better indications.

There are three major forms of market efficiency hypothesis: “Weak”, “Semi-strong” and “Strong” [75]. In weak-form efficiency, analyzing prices cannot predict future prices from the past. Excess returns using investment strategies based on historical share prices or other historical data cannot be earned in the long run. Technical analysis techniques will not be able to consistently produce excess returns, though some forms of fundamental analysis may still provide excess returns. Share prices show no serial dependencies, that is to say, there are no patterns to asset prices. This implies that future price movements are determined entirely by information not contained in the price series. Hence, prices must follow a random walk. This soft EMH does not require that prices remain at or near equilibrium, but only that market participants not be able to systematically profit from market inefficiencies. However, while EMH predicts that all price movement is random in the absence of change in fundamental information, many studies have shown a marked tendency for the stock markets to trend over time periods of weeks or longer [91] and that, moreover, there is a positive correlation between degree of trending and length of time period studied [92]. Various explanations for such large and apparently non-random price movements have been promulgated. There is a vast literature in academic finance dealing with the momentum effect identified by Jegadeesh and Titman [93, 94]. Stocks that have performed relatively well (poorly) over the past 3 to 12 months continue to do well (poorly) over the next 3 to 12 months. The momentum strategy is long recent winners and shorts recent losers, and produces positive risk-adjusted average returns. Being simply based on past stock returns, the momentum effect produces strong evidence against weak-form market efficiency, and has been observed in the stock returns of most countries, in industry returns, and in national equity market indices. Moreover, Fama has accepted that momentum is the premier anomaly [95, 96]. The problem of algorithmically constructing prices which reflect all available information has been studied extensively in the field of computer science [97, 98].

In semi-strong form of market efficiency, it is implied that once there is publicly available new information, the share prices will be rapidly adjusted to reflect the information, it also shows that excess returns will be able to be reliably produced by neither fundamental analysis nor technical analysis techniques. As a result, no arbitrage can be attained based on that information. Also, no technical analysis will be able to predict the share prices to generate excess returns. To test for it, the adjustments to previously unknown news must be of a reasonable size and must be instantaneous, and then, consistent upward or downward adjustments after the initial change must be looked for. It would suggest that investors had interpreted the information in a biased fashion and hence in an inefficient manner if there are any such adjustments.

In strong-form efficiency, share prices can reflect all public and private information, and no one can earn excess returns. If there are legal barriers to private information becoming public, as with insider trading laws, except in the case where the laws are universally ignored, strong-form efficiency is impossible. To test for this, a market needs to exist where investors cannot consistently earn excess returns over a long period of time. No refutation even of strong-form efficiency follows with lots of fund managers worldwide, even if some money managers are consistently observed to beat the market, even a normal distribution of returns should be expected to produce a few dozen star performers.

However, various phenomena suggest that the real financial market does not conform to the market efficiency hypothesis. The reasons include human irrationalities, such as overconfidence and overreaction of investors, so that they may sell winning stocks and hold on to losing stocks. There are other non-human-related errors, such as unfair distribution of information. For stock market particularly, Dreman and Berry found that, stocks with low P/E, which refers to price earnings ratio, earn greater risk-adjusted returns than high P/E stocks [99]. For the price earnings ratio anomaly, one of the explanations based on investor overreaction is that, companies with very low $P/E_s$ are thought to be undervalued, since investors are pessimistic after a series of bad news, such as poor earnings reports. The investor overreaction would further drive the stock price down. Once future earnings turn better, the price would be adjusted accordingly to more reasonably reflect the value of company. Similarly, the companies with high $P/E_s$ are overvalued, and the price would be driven down. To specifically look into the behavior of investor overreaction, Debondt and Thaler, in their study "Does the Stock Market Overreact?", suggested that people tend to overreact to unexpected and dramatic events [100].

## 5.1.2 Stock Market Overreaction

One way used to predict the extent of weak-form efficiency in stock markets is to test for the overreaction. The effect of overreaction suggests that investors overreact to good or bad news causing share prices to deviate from their equilibrium level. In particular, securities suffering abnormally low returns (losers) in the past will subsequently experience relatively higher returns while shares which have performed well in the past (winners) will do less well in the future.

The Bayes rule states that:

BAYES'S RULE

$$\mathrm{P(A|B)} = \frac{P(A_i) \cdot P(B|A_i)}{\sum[P(A_i) \cdot P(B|A_i)]}$$

where $A_1, \ldots\ldots, A_n$ is an all-inclusive set of possible outcomes given B.

One condition of rational investor behaviors is that they use Bayes Rule to form new belief as new information becomes available. Each time period new information signals are added to the information set. The investors can correctly use the new information set to update their expectations and thus determine the value of companies. Consequently, stock prices will accurately reflect fundamental values of the companies. And when there is unexpected positive or negative news, the prices will move up and down accordingly. However in real world, most investors are irrational when making decisions. Investors tend to give more weight of consideration to recent information or new data, and give less weight of consideration to historical data [101]. For instance, if a stock price drops, most irrational investors will have an incentive to buy in the stocks. Similarly, they are likely to sell the stock if stock price drops. And they will pay little attention to the long term paying power such as dividends. The price earnings ratio (P/E) anomaly, as stated earlier in this paper, describes an observation that stocks with low P/E returns earn greater risk-adjusted returns than high P/E stocks [99].

De Bondt and Thaler are the forerunners in the study of overreact on stock mark. They have used data of stock mark in the US between 1926 and 1982 to analyze whether investors overreact, and according to their research, there has been reversion on stock return in the long term because of investors' irrational behavior [100]. Based on those observations, DeBondt suggested two hypothesis if a stock experiences significant price movement, then a subsequent price movement in the opposite direction is likely to follow. Moreover, the level of extremeness is positively correlated between the initial and the following price movement [100]. To test the hypothesis, DeBondt proposed an empirical test method, using the data of monthly stock returns from 1930s to 1970s.

Lehmann (1990) has used data of stock mark in the US between 1965 and 1989 to identify the factors that have reversed the market in a short time interval, but short-term profitability can hardly be identified with overreaction. Rather, it is probably a result of pressures of price in short-term or lacking of liquidity. It thought that predictable variation in equity returns might reflect either predictable changes in expected returns or market inefficiency and stock price overreaction. These explanations can be distinguished by examining returns over short time intervals since systematic changes in fundamental valuation over intervals like a week should not occur in efficient markets. This probably reflects inefficiency in the market for liquidity around large piece changes [102].

Zarowin (1989) examines the subsequent stock return performances of firms that have experienced extreme earnings years and finds that while the poorest earners outperform the best earners by a statistically significant smaller than the best earners at the time of portfolio formation. When the poorest earners are matched with the best earners of equal size, there is virtually no evidence of differential stock return performance, indicating that the market does not overreact to extreme earnings news, and suggesting that size discrepancies between winners and losers may be responsible for the apparent overreaction phenomenon [103].

To build the stock market value, one of the best ways is to use expert systems with Artificial Neural Networks (ANN), which can easily adapt the changes of the stock market. It is observed that in most of the cases ANN models give better result than other methods. However, there have been few studies on using the ANN for stock market. Generalized version of ARCH model Generalized ARCH (GARCH) model [104], Exponential GARCH (EGARCH) model [105] and Dynamic Architecture for Artificial Neural Networks (DAN2) and so on. From the late 1990s, Chinese scholars have studied overreact on stock mark in China. Zhang Renji, Zhu Pingfang, and Wang Huai-fang (1998) have found a falling trend on winner port-folio [106]. Zhu Shaoxing (2000) has conducted no overreaction in Shenzhen stock market [107]. Song Xian-zhong and Tang Sheng (2006) have identified empirical study on overreaction and scale effect on the corporations listed in A-share market in Shanghai [108].

In recent years, there have been a growing number of studies about movements of various kinds on stock market. Both academic researchers and practitioners have made tremendous efforts to predict the future movements of stock market. Zhang Yanqing and Wan Xuhui (2007) have developed a new ANN architecture Statistical Fuzzy Interval Neural Network based on Fuzzy Interval Neural Network [109]. Zhu Xiaotian, Wang Hong, Xu Li, and Li Huaizu (2008) have found trading volume can improve the prediction performance of neural networks by using basic and augmented neural network models [110]. Liao Zhe and Wang Jun (2010) have identified some results on the global stock indices by using stochastic time effective neural network model [111].

### 5.1.3 The study of the winner and loser portfolio

In their seminal work, De Bondt and Thaler (1985) discovered patterns of return predictability in the U.S. stock market for the long-term horizon of 3 to 5 years. Stocks with poor past returns (loser stocks) outperformed those with relatively well past performance (winner stocks). In other words, winner and loser stock returns tend to reverse over time. De Bondt and Thaler (1985) suggested that investor's overreactions to good and bad news were the cause of this phenomenon. The authors postulated the overreaction hypothesis based on the findings of an experimental study in psychology conducted by Kahneman and Tversky (1982) [112], wherein individuals were found to initially overreact to the arrival of unexpected news. In a similar vein, the overreaction hypothesis states that investors tend to overweigh the significance of recent news. Investor's misjudgments cause prices to increase or decrease beyond reasonable levels. Investors then realize their error in judgment, revise their beliefs and trade in a manner that results in a return reversal. Follow-up studies have shown that the observed overreaction could not be fully attributed to seasonality (De Bondt & Thaler, 1987) [113, 114], size (Zarowin, 1990) [114] or risk (Braun, Nelson, & Sunier, 1995) [115].

In addition to the long-term overreaction documented by De Bondt and Thaler (1985), many studies have documented the existence of short-term overreaction. Among these studies, that of Jegadeesh (1990) found significant returns for contrarian portfolios that had been formed based on the previous one-month return. Additionally [116], Lehman (1990) examined whether overreaction existed in weekly returns [102]. Winner and loser stocks were selected based on the returns for the past week. Portfolio returns were then evaluated for five holding periods ranging from 1 to 52 weeks. Notable return reversals were documented for both winner and loser portfolios.

Evaluating weekly returns, Lo and MacKinlay (1990) focused on whether contrarian profits are caused by overreaction [117]. Based on their results, the authors concluded that stock market overreaction generated less than 50% of the profits. Moreover, the authors suggested that contrarian profits might not be solely driven by stock market overreaction and presented the lead lag effect as a primary contributor. However, Jegadeesh and Titman (1995) argued that contrarian profits are not generated by the lead lag effect [118]. In their study, a similar strategy to that of Lo and MacKinlay (1990) was employed where stocks were ranked using past one-week returns, and the contrarian portfolio was held for the following week. A larger sample of stocks was employed over the period from 1963 to 1990. Significant contrarian profits were reported. A decomposition of the contrarian profits revealed that a majority of the profits could indeed be attributed to the overreaction of stock prices to firm-specific information. Providing further support, Da, Liu and Schaumburg (2010) recently discovered that contrarian returns arise as a result of investor overreaction in response to the arrival of firm-specific news on discount rate as well as liquidity shocks [119].

Kang, Liu and Ni (2002) found short-term contrarian returns for the Chinese stock market [120]. Unfortunately, the loser minus winner portfolio which was formed based on the past 1-week return yielded significant returns for only the holding period of 1 week. Whilst returns were largely positive from weeks 2 to 26, none of the returns were significant. A later study by Wang, Burton and Power (2004) corroborated the evidence by documenting significant returns for only the first week after portfolio formation [121]. Returns for weeks 2 to 20 were insignificant. In contrast to situation in the Chinese market, Chou, Wei and Chung (2007) documented highly profitable contrarian returns for the Tokyo stock exchange [122]. For the one-month formation period, the returns were significant for all holding periods, from 1 to 24 months. Recently, a study by Griffin, Kelly and Nardari (2010) covered 56 stock markets with loser minus winner portfolios constructed based on 1-week holding and formation periods [123]. Argentina, Zimbabwe, Canada and Pakistan recorded some of the highest average weekly returns for the contrarian portfolio. Overall, returns were significant for 21 out of the 26 developed stock markets and 14 out of the 17 emerging markets that were examined.

One of the earliest studies of the Malaysian stock market was conducted by Mohd Arifin and Power (1996) [124]. The authors investigated overreaction using weekly data from 1990 to 1994. It should be noted that the study was severely limited in terms of sample size, as only 47 stocks were studied. Moreover, only the top and bottom 10 stocks were selected for the winner and loser portfolios. Thus, the stocks in the winner and loser portfolios were limited compared to previous studies. The KLSE composite index was used to compute market-adjusted excess returns, and the average cumulative excess return (ACER) was examined over ten weeks. The winner stocks exhibited negative returns for weeks one to three, and the loser stocks yielded positive returns throughout the ten weeks, indicating the existence of return reversals. The ACER of the loser minus winner stocks was also positive for all ten weeks. However, the statistical significance of the ACER could not be assessed as the *t*-value and/or *p*-value was not provided by the authors. Nevertheless, the CER *p*-value indicated that the returns were positively significant for one week following portfolio formation. Though accompanying data on significance was not provided, the authors concluded that overreaction is statistically significant for the first two weeks.

Ahmad and Tjan (2004) found that winner and loser stocks experience return reversals and claimed that overreaction does occur in Malaysia [125]. However, loser minus winner portfolios did not yield any significant positive returns. On the contrary, the returns were negative and insignificant. As stipulated by De Bondt and Thaler (1985), the difference between the loser and winner portfolios has to be significantly positive to justify that overreaction is indeed present. The lack of evidence found could be attributed to the sample selection method employed. Rather than screening the entire stock universe and choosing a reasonable percentage or number of stocks for portfolio formation, only the top 10 best and worst performing stocks which were as reported by the local newspaper were selected. Moreover, the sample was tested for only a one-year period. The holding period was restricted to 1, 2 and 3 weeks. The authors also investigated the effect of the 1997 Asian financial crisis by dividing the sample into pre-crisis (January to June 1997) and crisis (July to December 1997) periods. Returns for the pre-crisis period remained negative but were surprisingly significant at the 5% level for the two weeks holding period with a return of 7.88%. During the crisis, contrarian returns were positive but insignificant, and the highest return of 2.99% was obtained for the two weeks holding period. Overall, the evidence pointed towards an unprofitable contrarian strategy for 1997.

Recently, Ali, Nassir, Hassan and Abidin (2010) also studied short-term overreaction in the Malaysian stock market [126]. However, the study was limited to gauging the market reaction to specific events. In particular, 13 individual events which took place between January 1987 and December 2006 were investigated. Overall, the results were inconclusive. Overreaction was found for some events (e.g. political events), but not for others such as international events. Nevertheless, it should be noted the actual sample size used in the study was not specified. Moreover, neither the method for computing abnormal returns nor the market proxy was detailed.

In our previous study, we select the trading data from January 2007 to June 2011 in stock market in Shanghai, and the result of empirical study shows that there has been overreaction in the stock market [127]. In this simulation, the data from 2004 to 2014 is chosen to verify the overreaction in shanghai stock market once again, and it is for the first time to take advantage of a neuron model with dendritic nonlinearity to fit and predict the tendency of overreaction. This new approach is aimed to provide a novel solution for future research.

Chapter 5 is organized as follows: Section 5.2 gives the sample selection and test methods; Section 5.3 shows the experiments and results. The final section gives the conclusion and recommendations for future researches. This study will not only make contribution to the ANN research but also to the business implementations of stock market.

## 5.2 Empirical Study on the Overreaction

### 5.2.1 Sample Selection and Test Methods

This thesis extracts randomly 200 stocks from the shanghai stock market, the timeframe under examination is set from January 2004 to October 2014, and the data examined are the daily closing prices in that period. Considering stock dividends and right offerings, the returns of each stock is calculated on the price which excludes right. If a stock is in suspension, it means that the stock's closing price remains the same. Based on the sequence of the level of cumulative abnormal returns in formation period, the 20 top stocks constitute the winner portfolio, and the loser portfolio is composed of the 20 lowest stocks.

### 5.2.2 The Sorting Methods in Formation Period

The formation period in the thesis is divided into 3 months, 6 months, 12 months and 24 months and the corresponding test period is divided into 1 month, 3 months, 6 months, 12 months, 24 months and 36months. If the reference time, the length of formation period and the length of test period are set to $T_0$, $T_1$ and $T_2$, respectively, then ($T_0 - T_1$, $T_1$) is the formation period, and ($T_0$, $T_0 + T_2$) is the test period. If reference time is constant, there can be more combinations of formation period and test period. The formations period and test period of these combinations do not overlap, but the current test period and the next formation period may overlap.

The calculation of excess returns uses incorporates a marketing adjustment, as it is used by the De Bondt, Richard Thaler and Paul Zarowin. The formula is $ER_{i,k}=R_{i,k}-R_{m,k}$; i=1,2,3,…,n

$R_{i,k}$ is yield of stock I in k month.

$R_{m,k}$ is yield of market in k month.

Cumulative abnormal return of stock I in formation period is $CER_i = \sum_k ER_{i,k}$

## 5.2.3 The Test Methods in Test Period

Based on the sequence of the level of cumulative abnormal returns in formation period, the top 20 stocks make the winner portfolio and the lowest 20 stocks make the loser portfolio. Thus, the calculated average abnormal monthly returns of each combination are given by:

$$AR_{w,k} = \frac{\sum_i ER_{i,k}^{w}}{n_1} \tag{17}$$

$$AR_{l,k} = \frac{\sum_i ER_{i,k}^{l}}{n_1} \tag{18}$$

$$CAR_{w,k} = \sum_k AR_{w,k} \tag{19}$$

$$CAR_{l,k} = \sum_k AR_{l,k} \tag{20}$$

Here, $AR_{w,k}$ is the average abnormal monthly returns of winner portfolio, $AR_{l,k}$ is the average abnormal monthly returns of loser portfolio, $CAR_{w,k}$ is the average cumulative abnormal monthly returns of winner portfolio, and $CAR_{l,k}$ is the average cumulative abnormal monthly returns of loser portfolio.

## 5.3 Experiments and Results

### 5.3.1 Empirical result of overreaction

**Table 5-1** CAR of the winner and loser portfolios.

According to the method and design mentioned above, a lot of descriptive data are gotten. And, we find that the average cumulative abnormal returns (CAR) of loser portfolio has been greater than that of winner portfolio in 18 months formation period. The result is basically consistent with the former study. And that is to say, there are clearly overreaction happened on the shanghai stock market. Table 1 shows the data belonging to the 18-month formation period, CAR of winner portfolio and loser portfolio. Moreover, Figure 3 shows the alteration of CAR belonging to winner and loser portfolios. Combined Table 5-1 and Fig. 5-1, the difference of CAR between the loser and winner portfolios reaches a maximum of 17.36% at six month. Furthermore, the degree of overreaction gradually decreases with time, and finally has gone to disappear. Moreover, one may get significant arbitrage profit on the overreaction period by buying the stocks of loser portfolio at an early stage and selling the ones of winner portfolio.

However, the data listed in Table 5-1 is not enough to train the NMDN on a more refined level. In order to display the alteration of winner and loser portfolio more precisely, Fig. 5-2 shows the curve of 94-month test period with the step of one month. In

Fig. 5-2, the curve of L-W goes to under 0, meansings that the overreaction disappears. Furthermore, Fig. 5-3 shows the difference between loser portfolios' the average abnormal monthly return (AR) and winner portfolios'. The NMDN is used to predict the difference of AR and then to predict the CAR curve of L-W.

### 5.3.2 Parameters of NMDN

As mentioned in Chapter 2, there are some parameters in NMDN need to be set up. Different parameters of NMDN have distinct influence on the training and predicting, however, there has been no clear criterion to determine which values of parameters can maximize the ability of NMDN. For the reason, Table 5-2 shows the mean squared error (MSE), epoch of learning (Epochs), learning constant ($\eta$), threshold ($\gamma$), K, Ksoma and Branch in this simulation empirically.

**Table 5-2** NMDN parameters applying for training and prediction.

| K | Ksoma | Branch | $\gamma$ | $\eta$ | Epochs | MSE |
|---|---|---|---|---|---|---|
| 3 | 3 | 25 | 0.5 | 0.01 | 10000 | 0.0001 |

However, in order to train NMDN for prediction, the input data of NMDN is defined in Table 5-3. In this simulation, the average abnormal monthly returns (AR) is selected as input data for NMDN, and n-months data, indicated as the step, is used for predict the data of (n+1)-month. The NMDN has been trained 50 times on each step, respectively.

**Table 5-3** The train data of NMDN.

| | Step | | | |
|---|---|---|---|---|
| | 3 | 4 | 5 | 6 |
| Train data | 40 | | | |

### 5.3.3 The result of prediction

Fig. 5-4 shows the average curve of prediction on each step. The curve, using 3 months data as input data to predict the data of following month, has departed from the standard curve. It discloses that there may be no relationship on using 3-months data to predict the one of following month. Moreover, although it performs not so well on using the data of 4, 5 and 6 months to predict the data of the following month, NMDN has succeeded in predicting the tendency of L-W, which is decrease. In order to quantify the fitness of prediction, the goodness of fit, R, is introduced to compare the performance on different steps. Table 5-4 shows the values of R on different steps. Higher values indicate that the model fits the data better.

**Table 5-4** The goodness of fit, Rnew.

Combined Fig. 5-4 with Table 5-4, it becomes evident that the R turns larger with increasing steps. The growth of R indicates that with the increase of steps, the correlation between the data of first n-month and following month greatly improves. This result indicates a possible connection between on the first n-month and following months.

## Figure caption

**Fig.5-1** CAR curves of 18-month formation period.

**Fig.5-2** CAR curves of 18-month formation period for 94 months.

**Fig. 5-3** The difference between loser portfolio’s CAMR and winner portfolio’s.

**Fig.5-4** Average prediction result on different steps.

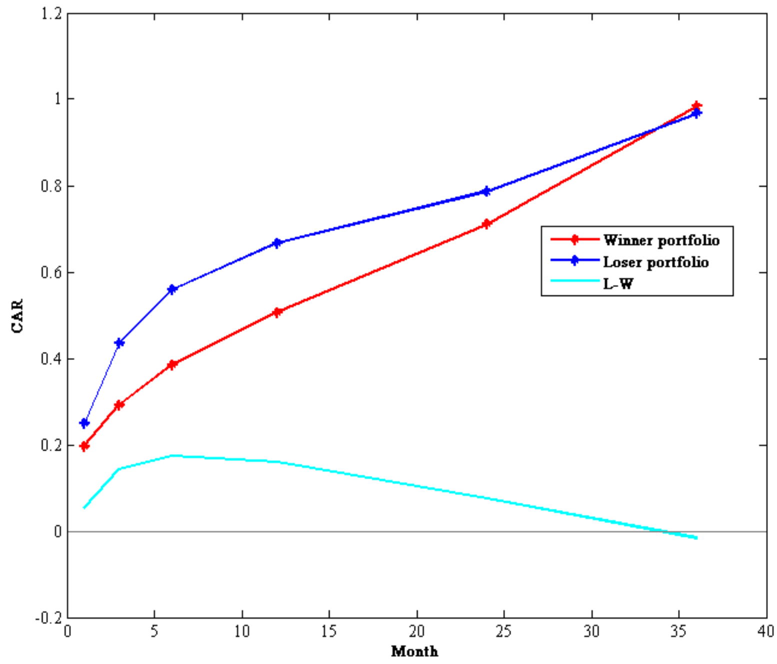



**Fig.5-1** CAR curves of 18-month formation period.

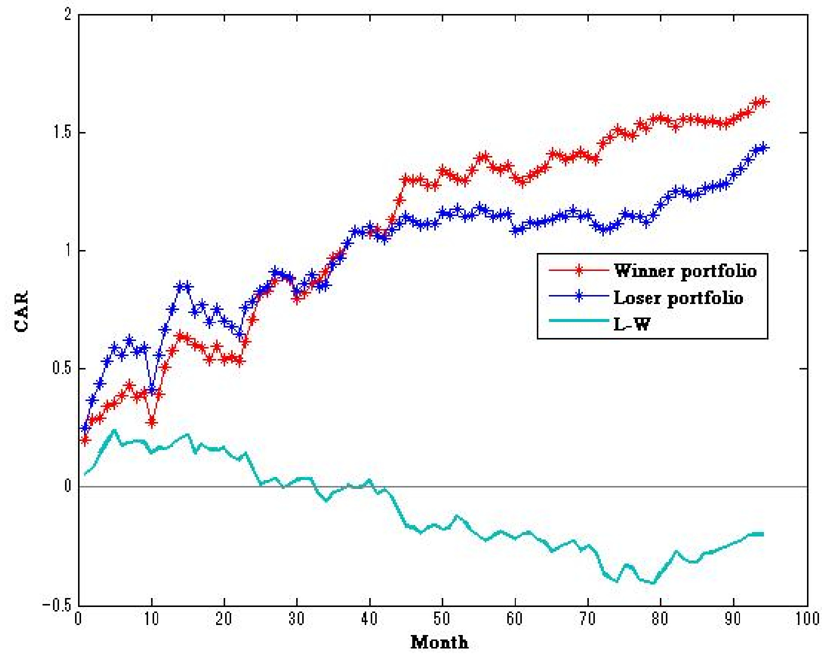



**Fig.5-2** CAR curves of 18-month formation period for 94 months.

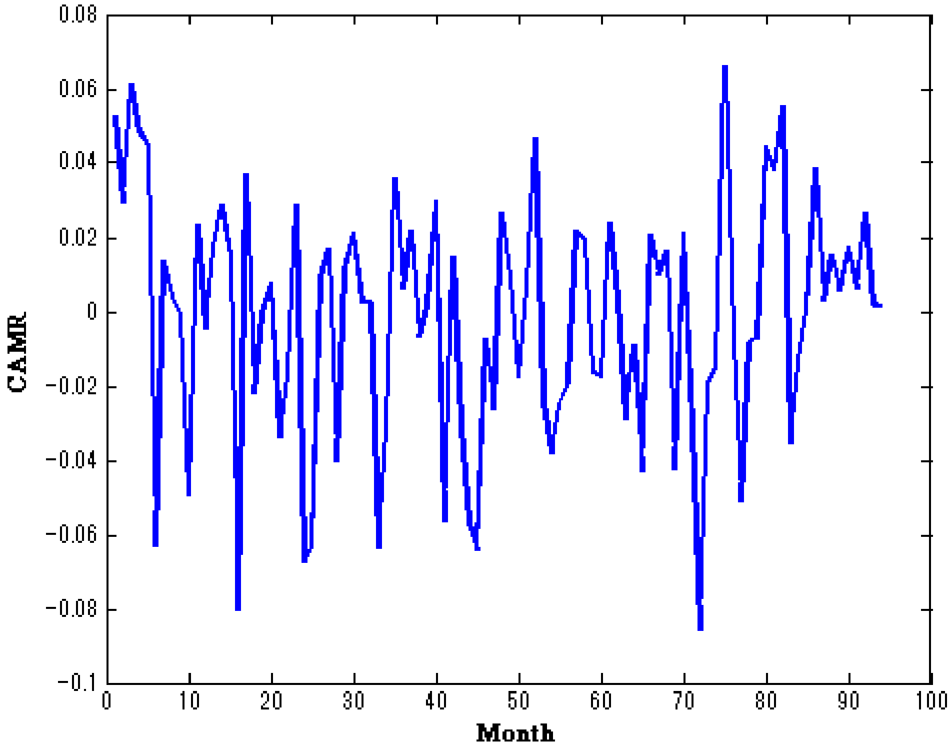



**Fig. 5-3** The difference between loser portfolio's CAMR and winner portfolio's.

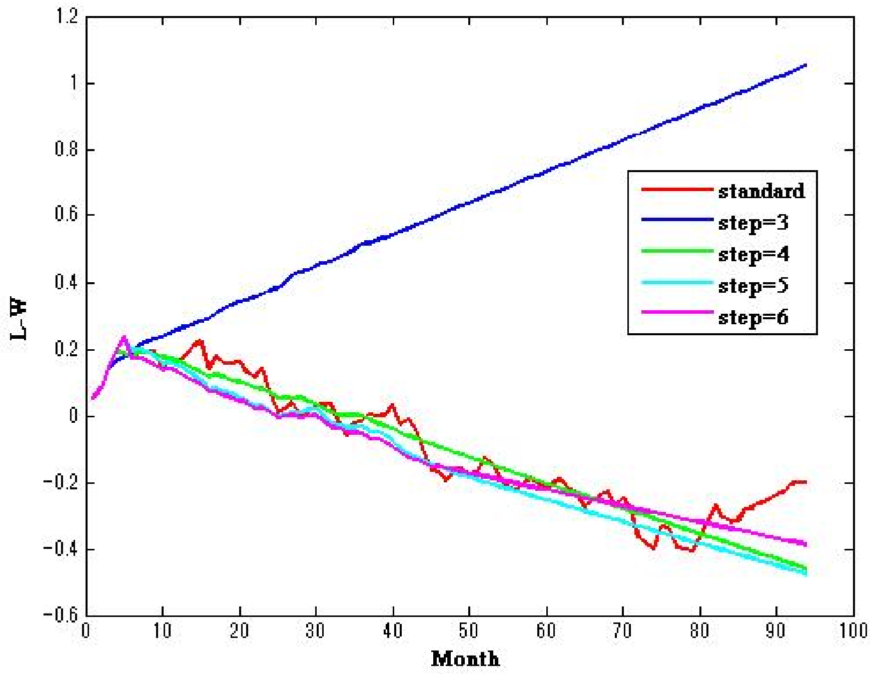



**Fig.5-4** Average prediction result on different steps.

# Chapter 6

# Conclusion

In this thesis, a neuron model with dendritic nonlinearity (NMDN) is proposed to be used for classifying the breast cancer, and is compared with the classic BPNNs on the accuracy, convergence rate, stability and AUC. The average accuracy of NMDN is 97.8%. And, with the growth of quantity of branches in NMDN and nodes of the hidden layer in BPNNs, growth of sensitivity raises. Furthermore, the growth of accuracy in NMDN is faster than BPNNs. In addition, the accuracy of NMDN having 5 branches is lower than BPNNs, thus, a conjecture is proposed that there is some connection on the quantity of redundant synapses and branches of initial state. The convergence rate of NMDN converges faster with the growth of branch; conversely, the convergence rate of BPNNs is slower with the growth of nodes in hidden layer. Through comparing the initial and final errors of NMDN and BPNNs, not only does the NMDN perform more steadily, but also releases the local minimum problem. The average AUC of NMDN is 0.9741, higher than BPNNs. It indicates that the performance of NMDN is superior and stable from another aspect. Finally, we extract a relationship on attributes to detect the breast cancer from the evolved dendrite morphology. To combine the results in this thesis with the ones of [52] and [63], The uniform shape, Bare Nuclei and the Clump thickness can be treated as the important factors to detect the breast cancer. In addition, the uniformity of cell size, marginal adhesion and mitoses may be removed from the detection of breast cancer in the future. On the basis of analysis, the results show that the NMDN processes a superior ability of classifying the possible breast cancer. Moreover, it is proved that a neuron holds a very huge computing ability and the brain of the computation is undervalued [21]. We also believe that the proposed model can be helpful to the medical researchers for them as a new choice for their final decision.

Moreover, the overreaction in Shanghai stock market is confirmed once again by using later data. With time going by, the influence of overreaction, which happened from 2007, can be found to decrease and to disappear finally.

And, for the first time, NMDN is also used to predict the tendency of overreaction. The first n-months' data is treated as input data to predict the data of following month.

The result shows that the neuron model possesses huge computational ability and successes in predicting the tendency of overreaction. Moreover, with the increase of steps on inputs, the correlation between the data of first n-month and following month turns more and more tightly together. Moreover, NMDN provides another new approach for the researchers.

# References:


[1]. Wasserman, P.D. and J.R. Meyer-Arendt, Neural computing, theory and practice. Applied Optics, 1990. 29: p. 2503.

[2]. Hajjawi, O.S., Human brain biochemistry. American Journal of BioScience, 2014. 2(4): p. 122-134.

[3]. Tsoukalas, L.H. and R.E. Uhrig, Fuzzy and neural approaches in engineering. 1996: John Wiley & Sons, Inc.

[4]. Gurney, K., An introduction to neural networks. 1997: CRC press.

[5]. Schalkoff, R.J., Artificial neural networks. 1997: McGraw-Hill New York.

[6]. Mitchell, T.M., Machine learning. WCB. 1997, McGraw-Hill Boston, MA.

[7]. Haykin, S., Neural Networks A Comprehensive Introduction. 1999, Prentice Hall, New Jersey.

[8]. Bishop, C.M., Neural networks for pattern recognition. 1995.

[9]. Cybenko, G., Neural networks in computational science and engineering. Computational Science & Engineering, IEEE, 1996. 3(1): p. 36-42.

[10]. Ripley, B.D., Pattern recognition and neural networks. 1996: Cambridge university press.

[11]. Kerekes, J., Receiver operating characteristic curve confidence intervals and regions. Geoscience and Remote Sensing Letters, IEEE, 2008. 5(2): p. 251-255.

[12]. Trujillano, J., J. March and A. Sorribas, [Methodological approach to the use of artificial neural networks for predicting results in medicine]. Medicina clínica, 2003. 122: p. 59-67.

[13]. Rodvold, D.M., et al., Introduction to artificial neural networks for physicians: taking the lid off the black box. The Prostate, 2001. 46(1): p. 39-44.

[14]. Malmgren, H., M. Borga and L. Niklasson, Artificial neural networks in medicine and biology. 2000: Springer-Verlag.

[15]. Tu, J.V., Advantages and disadvantages of using artificial neural networks versus logistic regression for predicting medical outcomes. Journal of clinical epidemiology, 1996. 49(11): p. 1225-1231.

[16]. Penny, W. and D. Frost, Neural networks in clinical medicine. Medical Decision Making, 1996. 16(4): p. 386-398.

[17]. Cross, S.S., R.F. Harrison and R.L. Kennedy, Introduction to neural networks. The Lancet, 1995. 346(8982): p. 1075-1079.

[18]. Reggia, J.A., Neural computation in medicine. Artificial Intelligence in Medicine, 1993. 5(2): p. 143-157.

[19]. Kitajima, T. and K. Hara, A bipolar model of the nerve cell—Dynamic characteristics of synaptic film. Trans IEICE, 1987: p. 818-826.

[20]. Koch, C., T. Poggio and V. Torres, Retinal ganglion cells: a functional interpretation of dendritic morphology. Philosophical Transactions of the Royal Society of London. B, Biological Sciences, 1982. 298(1090): p. 227-263.

[21]. Smith, S.L., et al., Dendritic spikes enhance stimulus selectivity in cortical neurons in vivo. Nature, 2013. 503(7474): p. 115-120.

[22]. Tang, Z., et al., A model of the neuron based on dendrite mechanisms. Electronics and Communications in Japan (Part III: Fundamental Electronic Science), 2001. 84(8): p. 11-24.

[23]. Yuki Todo, Z.T.H.T., Neurons with multiplicative interactions of nonlinear synapses (submitted). Nature, 2014.

[24]. Kandel, E.R., J.H. Schwartz and T.M. Jessell, Principles of neural science. Vol. 4. 2000: McGraw-Hill New York.

[25]. Best, B., Basic Neuron Physiology. 1990.

[26]. From Wikipedia, T.F.E., Neuron. 2015.

[27]. Nowakowski, R.S., Stable neuron numbers from cradle to grave. Proceedings of the National Academy of Sciences, 2006. 103(33): p. 12219-12220.

[28]. Häusser, M. and B. Mel, Dendrites: bug or feature? Current opinion in neurobiology, 2003. 13(3): p. 372-383.

[29]. London, M. and M. Häusser, Dendritic computation. Annu. Rev. Neurosci., 2005. 28: p. 503-532.

[30]. From Wikipedia, T.F.E., Artificial neuron. 2014.

[31]. Minsky, M.L. and S. Papert, Perceptrons: An essay in computational geometry. 1969, Cambridge, MA: MIT Press.

[32]. Jan, Y. and L.Y. Jan, Branching out: mechanisms of dendritic arborization. Nature Reviews Neuroscience, 2010. 11(5): p. 316-328.

[33]. Chklovskii, D.B., B.W. Mel and K. Svoboda, Cortical rewiring and information storage. Nature, 2004. 431(7010): p. 782-788.

[34]. Chklovskii, D.B., Synaptic connectivity and neuronal morphology: two sides of the same coin. Neuron, 2004. 43(5): p. 609-617.

[35]. Komatsu, Y., et al., Electrophysiological and morphological correlates in the development of visual cortical circuitry in infant kittens. Developmental Brain Research, 1985. 22(2): p. 305-309.

[36]. Schmitz, S.K., et al., Automated analysis of neuronal morphology, synapse number and synaptic recruitment. Journal of neuroscience methods, 2011. 195(2): p. 185-193.

[37]. From Wikipedia, T.F.E., Central nervous system. 2015.

[38]. Hill, R.W., G.A. Wyse and M. Anderson, Animal Physiology. Sinauer. Inc., Sunderland, MA, 2004.

[39]. Humberto Gutierrez, C.O., A Brief Introduction to the Brain..

[40]. Takeuti, K., Calculate dendrites of the neuron to perceive a slope in the depth direction. 2010, The University of toyama.

[41]. Witten, I.H., et al., Weka: Practical machine learning tools and techniques with Java implementations. 1999.

[42]. 田村宏樹, 唐政与石井雅博, 入力の時間差を考慮したニューロンモデルとその運動方向選択機能. 電気学会論文誌. C, 電子・情報・システム部門誌= The transactions of the Institute of Electrical Engineers of Japan. C, A publication of Electronics, Information and System Society, 2002. 122(7): 第1094-1103页.

[43]. H. Tamura, Z.T.I.O., Directionally selective cells have a δ-like morphology, in International Symposium on Nonlinear Theory and its Applications. 1999. p. P.215-218.

[44]. SEKIYA, Y., et al., Learning-possibility for neuron model in Medical Superior Temporal area. 제어로봇시스템학회 국내학술대회 논문집, 2000: p. 516-516.

[45]. Sekiya, Y., et al. Learning-possibility that neuron model can recognize depth-rotation in three-dimension. in Proc. of 6th international symposium on artificial life and robotics. 2001.

[46]. Acevedo, M.E., M.A. Acevedo and F. Felipe, Classification of cancer recurrence with al-

pha-beta BAM. Mathematical Problems in Engineering, 2009. 2009.
[47]. Senapati, M.R. and P.K. Dash, Intelligent system based on local linear wavelet neural network and recursive least square approach for breast cancer classification. Artificial Intelligence Review, 2013. 39(2): p. 151-163.
[48]. Senapati, M.R. and P.K. Dash, Local linear wavelet neural network based breast tumor classification using firefly algorithm. Neural Computing and Applications, 2013. 22(7-8): p. 1591-1598.
[49]. Senapati, M.R., et al., Local linear wavelet neural network for breast cancer recognition. Neural Computing and Applications, 2013. 22(1): p. 125-131.
[50]. Ninness, C., et al., Behavioral and Physiological Neural Network Analyses: A Common Pathway Toward Pattern Recognition and Prediction. Psychological Record, 2012. 62(4): p. 579.
[51]. Vannucci, M. and V. Colla, Novel classification method for sensitive problems and uneven datasets based on neural networks and fuzzy logic. Applied Soft Computing, 2011. 11(2): p. 2383-2390.
[52]. Seera, M. and C.P. Lim, A hybrid intelligent system for medical data classification. Expert Systems with Applications, 2014. 41(5): p. 2239-2249.
[53]. Naghibi, S., M. Teshnehlab and M.A. Shoorehdeli, Breast cancer classification based on advanced multi dimensional fuzzy neural network. Journal of medical systems, 2012. 36(5): p. 2713-2720.
[54]. Wu, C., C. Lin and C. Lee, A functional neural fuzzy network for classification applications. Expert Systems with Applications, 2011. 38(5): p. 6202-6208.
[55]. Lin, C., S. Kuo and C. Peng, Multiple functional neural fuzzy networks fusion using fuzzy integral. International Journal of Fuzzy Systems, 2012. 14(3): p. 380-391.
[56]. Senapati, M.R., G. Panda and P.K. Dash, Hybrid approach using KPSO and RLS for RBFNN design for breast cancer detection. Neural Computing and Applications, 2014. 24(3-4): p. 745-753.
[57]. Hsieh, H. and K. Tang, Hardware friendly probabilistic spiking neural network with long-term and short-term plasticity. Neural Networks and Learning Systems, IEEE Transactions on, 2013. 24(12): p. 2063-2074.
[58]. Glackin, C., et al., Receptive field optimisation and supervision of a fuzzy spiking neural network. Neural Networks, 2011. 24(3): p. 247-256.
[59]. Askarzadeh, A. and A. Rezazadeh, Artificial neural network training using a new efficient optimization algorithm. Applied Soft Computing, 2013. 13(2): p. 1206-1213.
[60]. Sasu, L.M. and R. Andonie, Bayesian ARTMAP for regression. Neural Networks, 2013. 46: p. 23-31.
[61]. Azar, A.T. and S.A. El-Said, Probabilistic neural network for breast cancer classification. Neural Computing and Applications, 2013. 23(6): p. 1737-1751.
[62]. Kalteh, A.A., et al., A research about breast cancer detection using different neural networks and K-MICA algorithm. Journal of cancer research and therapeutics, 2013. 9(3): p. 456.
[63]. Uzer, M.S., O. Inan and N. Yılmaz, A hybrid breast cancer detection system via neural network and feature selection based on SBS, SFS and PCA. Neural Computing and Applications, 2013. 23(3-4): p. 719-728.
[64]. Verma, B. and S.Z. Hassan, Hybrid ensemble approach for classification. Applied Intelligence, 2011. 34(2): p. 258-278.
[65]. Marcano-Cedeño, A., J. Quintanilla-Domínguez and D. Andina, Breast cancer classification

applying artificial metaplasticity algorithm. Neurocomputing, 2011. 74(8): p. 1243-1250.

[66]. Markowska-Kaczmar, U. and R. Matkowski, Experimental study of evolutionary based method of rule extraction from neural networks in medical data, in Advances in Data Mining. Applications in Medicine, Web Mining, Marketing, Image and Signal Mining. 2006, Springer. p. 76-90.

[67]. Chang, P., J. Lin and C. Liu, An attribute weight assignment and particle swarm optimization algorithm for medical database classifications. Computer methods and programs in biomedicine, 2012. 107(3): p. 382-392.

[68]. Williamson, D.F., R.A. Parker and J.S. Kendrick, The box plot: a simple visual method to interpret data. Annals of internal medicine, 1989. 110(11): p. 916-921.

[69]. Zhou, X.H. and J. Harezlak, Comparison of bandwidth selection methods for kernel smoothing of ROC curves. Statistics in medicine, 2002. 21(14): p. 2045-2055.

[70]. Witten, I.H. and E. Frank, Data Mining: Practical machine learning tools and techniques. 2005: Morgan Kaufmann.

[71]. Fawcett, T., ROC graphs: Notes and practical considerations for researchers. Machine learning, 2004. 31: p. 1-38.

[72]. Bradley, A.P., The use of the area under the ROC curve in the evaluation of machine learning algorithms. Pattern recognition, 1997. 30(7): p. 1145-1159.

[73]. Bettinger, R., Cost-sensitive classifier selection using the ROC convex hull method. SAS Institute, 2003.

[74]. Purves, R.D., Optimum numerical integration methods for estimation of area-under-the-curve (AUC) and area-under-the-moment-curve (AUMC). Journal of pharmacokinetics and biopharmaceutics, 1992. 20(3): p. 211-226.

[75]. From Wikipedia, T.F.E., Efficient-Market Hypothesis.

[76]. Kirman, A., Economic theory and the crisis. VoxEU. org, 2009. 14.

[77]. Fox, J., Is the market rational?－No, say the experts. But neither are you－so don’t go thinking you can outsmart it.’. Fortune Magazine, 2002. 9: p. 59-64.

[78]. Basu, S., Investment performance of common stocks in relation to their price－earnings ratios: A test of the efficient market hypothesis. The journal of Finance, 1977. 32(3): p. 663-682.

[79]. Rosenberg, B., K. Reid and R. Lanstein, Persuasive evidence of market inefficiency. The Journal of Portfolio Management, 1985. 11(3): p. 9-16.

[80]. Nicholson, S.F., Price ratios in relation to investment results. Financial Analysts Journal, 1968. 24(1): p. 105-109.

[81]. Fama, E.F. and K.R. French, The cross－section of expected stock returns. the Journal of Finance, 1992. 47(2): p. 427-465.

[82]. Beechey, M., D.W. Gruen and J. Vickery, The efficient market hypothesis: a survey. 2000: Reserve Bank of Australia, Economic Research Department.

[83]. Cootner, P.H., The random character of stock market prices. 1964.

[84]. Fama, E.F., The behavior of stock-market prices. Journal of business, 1965: p. 34-105.

[85]. Samuelson, P.A., Proof that properly anticipated prices fluctuate randomly. Industrial management review, 1965. 6(2): p. 41-49.

[86]. Schwager, J.D., Market Sense and Nonsense: How the Markets Really Work (and how They Don't). 2012: John Wiley & Sons.

[87]. Read, C., The Efficient Market Hypothesists: Bachelier, Samuelson, Fama, Ross, Tobin and Shiller. 2012: Palgrave Macmillan.

[88]. Alajbeg, D., Z. Bubaš and V. Šonje, The efficient market hypothesis: problems with interpretations of empirical tests. Financial Theory and Practice, 2012. 36(1): p. 53-72.

[89]. Fama, E.F., Efficient capital markets: A review of theory and empirical work*. The journal of Finance, 1970. 25(2): p. 383-417.

[90]. Jung, J. and R.J. Shiller, Samuelson's dictum and the stock market. Economic Inquiry, 2005. 43(2): p. 221-228.

[91]. Saad, E.W., D.V. Prokhorov and D.C. Wunsch, Comparative study of stock trend prediction using time delay, recurrent and probabilistic neural networks. Neural Networks, IEEE Transactions on, 1998. 9(6): p. 1456-1470.

[92]. Granger, C.W. and O. Morgenstern, SPECTRAL ANALYSIS OF NEW YORK STOCK MARKET PRICES1. Kyklos, 1963. 16(1): p. 1-27.

[93]. Jegadeesh, N. and S. Titman, Profitability of momentum strategies: An evaluation of alternative explanations. The Journal of Finance, 2001. 56(2): p. 699-720.

[94]. Jegadeesh, N. and S. Titman, Returns to buying winners and selling losers: Implications for stock market efficiency. The Journal of Finance, 1993. 48(1): p. 65-91.

[95]. Fama, E.F. and K.R. French, Dissecting anomalies. The Journal of Finance, 2008. 63(4): p. 1653-1678.

[96]. Fama, E.F. and K.R. French, Multifactor explanations of asset pricing anomalies. The journal of finance, 1996. 51(1): p. 55-84.

[97]. Kleinberg, J. and É. Tardos, Algorithm design. 2006: Pearson Education India.

[98]. Nisan, N., et al., Algorithmic game theory. Vol. 1. 2007: Cambridge University Press Cambridge.

[99]. Dreman, D.N. and M.A. Berry, Overreaction, underreaction, and the low-P/E effect. Financial Analysts Journal, 1995. 51(4): p. 21-30.

[100]. Bondt, W.F. and R. Thaler, Does the stock market overreact? The Journal of finance, 1985. 40(3): p. 793-805.

[101]. Kahneman, D. and A. Tversky, Intuitive prediction: Biases and corrective procedures. 1977, DTIC Document.

[102]. Lehmann, B., Fads, martingales, and market efficiency. 1988, National Bureau of Economic Research Cambridge, Mass., USA.

[103]. Zarowin, P., Does the stock market overreact to corporate earnings information? The journal of Finance, 1989. 44(5): p. 1385-1399.

[104]. Bollerslev, T., Generalized autoregressive conditional heteroskedasticity. Journal of econometrics, 1986. 31(3): p. 307-327.

[105]. Nelson, D.B., Conditional heteroskedasticity in asset returns: A new approach. Econometrica: Journal of the Econometric Society, 1991: p. 347-370.

[106]. Renji Zhang, Pingfang Zhu and Huaifang Wang, The empirical study of overreact on stock market in shanghai, Economic Study, 1988.5 p.58-64

[107]. Shaoxing Zhu, The Study of Microstructure on Financial Market under the frame of Behavioral Finance Theory, 2000, Shanghai Jiao Tong University

[108]. Xianzhong, Song and Sheng Tang, The Empirical Study of Overreact and Scale Effect on Stock Market in China—based on the test of A shares market in shanghai, Journal of Jinan University(Philosophy and Social Sciences),2006.28(3): p. 74-78..

[109]. Zhang, Y. and X. Wan, Statistical fuzzy interval neural networks for currency exchange rate

time series prediction. Applied Soft Computing, 2007. 7(4): p. 1149-1156.

[110]. Zhu, X., et al., Predicting stock index increments by neural networks: The role of trading volume under different horizons. Expert Systems with Applications, 2008. 34(4): p. 3043-3054.

[111]. Liao, Z. and J. Wang, Forecasting model of global stock index by stochastic time effective neural network. Expert Systems with Applications, 2010. 37(1): p. 834-841.

[112]. Tversky, A. and D. Kahneman, Judgment under uncertainty: Heuristics and biases. science, 1974. 185(4157): p. 1124-1131.

[113]. BONDT, W.F. and R.H. Thaler, Further evidence on investor overreaction and stock market seasonality. The Journal of Finance, 1987. 42(3): p. 557-581.

[114]. Zarowin, P., Size, seasonality, and stock market overreaction. Journal of Financial and Quantitative analysis, 1990. 25(01): p. 113-125.

[115]. Braun, P.A., D.B. Nelson and A.M. Sunier, Good news, bad news, volatility, and betas. The Journal of Finance, 1995. 50(5): p. 1575-1603.

[116]. Jegadeesh, N., Evidence of predictable behavior of security returns. The Journal of Finance, 1990. 45(3): p. 881-898.

[117]. Lo, A.W. and A.C. MacKinlay, When are contrarian profits due to stock market overreaction? Review of Financial studies, 1990. 3(2): p. 175-205.

[118]. Jegadeesh, N. and S. Titman, Overreaction, delayed reaction, and contrarian profits. Review of Financial Studies, 1995. 8(4): p. 973-993.

[119]. Da, Z., Q. Liu and E. Schaumburg, Decomposing short-term return reversal. Available at SSRN 1551025, 2011.

[120]. Kang, J., M. Liu and S.X. Ni, Contrarian and momentum strategies in the China stock market: 1993–2000. Pacific-Basin Finance Journal, 2002. 10(3): p. 243-265.

[121]. Wang, J., B.M. Burton and D.M. Power, Analysis of the overreaction effect in the Chinese stock market. Applied Economics Letters, 2004. 11(7): p. 437-442.

[122]. Chou, P., K.J. Wei and H. Chung, Sources of contrarian profits in the Japanese stock market. Journal of Empirical Finance, 2007. 14(3): p. 261-286.

[123]. Griffin, J.M., P.J. Kelly and F. Nardari, Do market efficiency measures yield correct inferences? A comparison of developed and emerging markets. Review of Financial Studies, 2010: p. hhq044.

[124]. Mohd Arifin, N. and D.M. Power, Some evidence of short-run market overreaction for the Kuala Lumpur Stock Exchange. Capital Market Review, 1996. 4: p. 21-32.

[125]. Ahmad, Z. and S. Tjan, Short-Run Overreaction, Stock Prices and Investors' Irrationality in the Kuala Lumpur Stock Exchange. International Journal of Management Studies, 2004. 11: p. 1-19.

[126]. Ali, N., et al., Short run stock overreaction: Evidence from Bursa Malaysia. International Journal of Economics and Management, 2010. 4(2): p. 319-333.

[127]. Lin, H., et al., An Empirical Study on the Overreaction of Shanghai Stock Market. Chinese Studies, 2013. 2(01): p. 32.