Analysis and Visualization of Metabolic Syndrome

Using

Self Organizing Maps (SOM)

Peter Kamita Kihato

A thesis submitted in fulfillment for the degree of Doctor of Philosophy in Electrical and Electronic Engineering in the Jomo Kenyatta University of Agriculture and Technology

2009

DECLARATION

This thesis is my original work and has not been presented for a degree in any other University.

Signature: ..... Date: .....

Peter Kamita Kihato

This thesis has been submitted for examination with our approval as University Supervisors:

Signed: ..... Date: .....

Dr. John Ndichu Nderu,

JKUAT, Kenya.

Signed: ..... Date: .....

Prof. Masaaki Ohkita,

Tottori University, Japan.

Signed: ..... Date: .....

Prof. Yutaka Fukui,

Tottori University, Japan

DEDICATION

My loving wife and children for their understanding and continued support.

ACKNOWLEDGEMENT

I am greatly indebted to both the Kenya Government-Directorate of Personnel Management (DPM) (sponsor), and the Japanese Government for affording me this chance to study and pursue research at Tottori University Japan and Jomo Kenyatta University of Agriculture and Technology Kenya.

I am grateful to both staff members in Tottori University especially in Electrical Engineering and those of Jomo Kenyatta, Faculty of Engineering for their kind assistance throughout my split programme course. In particular, very special thanks and deep appreciation goes to the following for their invaluable guidance and advice during the three years of the course:

1. Prof. Yutaka Fukui
2. Prof. Masaaki Ohkita
3. Prof. Heizo Tokutaka
4. Dr. John N. Nderu

I would like to express my deep and sincere gratitude to my Japan advisor in areas of SOM Prof. H. Tokutaka for accepting and guiding me incessantly throughout the period I was doing this work. In addition, I wish to recognize the members of the research team whose profession was related to health namely; Y. Kurozawa, K. Kotani and Y. Maniwa all from Japanese hospitals. They provided the raw data to research on as well as giving professional advice whenever need arose. Frankly speaking I learnt a lot both in the field of SOM and general approach to research. This culminated in the presentations done in conferences and published papers in refereed international journals as shown at the end of this thesis.

It is unforgettable that the above could not happen without the full consent and encouragement from my project supervisors Prof Ohkita and his entire team, Prof. Ohki, K. Fujimura and students in his laboratories.

Special thanks go to my local supervisor Dr. John N. Nderu for continually giving me words of encouragement and focus to issues of research. His unreserved comments and guidance in research will always linger in my mind.

Prof. Fukui has been a father to many Kenyans in the academic world and is my inspiration when it comes to furthering my studies.

Finally, to my wife Njeri, sons Kihato and Githuku, and my daughter Wanjiru, for the long years they endured when I was away from home attending this course.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF APPENDIXES

LIST OF ABBREVIATIONS

AI	Artificial intelligence
ANN	Artificial neural networks
BAM	Bi-directional associative memory
BMI	Body mass index
BMU	Best matching unit
BPN	Back propagation network
Ca^{2+}	Calcium ions
CAD	Coronary artery disease
CNS	Central nervous system
DNA	Deoxyribonucleic acid
EDA	Exploratory data analysis
ES	Expert system
ESOM	Emergent SOM
FAT	Fatty acid transporter
FL	Fuzzy logic
FNN	Fuzzy neural network
HBP	High blood pressure
HDL	High density lipoproteins
GA	Genetic algorithms
GLU	Blood glucose
K^{+}	Potassium ions

KMC	*K*-means cluster
LBP	Low blood pressure
LDL	Low-density lipoproteins
LVQ	Learning vector quantization
MLP	multiple layer perceptron
MDS	Multidimensional scaling
MSE	Mean square error
MSP	Metabolic syndrome points
Na^+	Sodium ions
PDP	Parallel distributed processing
PCA	Principal Component Analysis
PNS	Peripheral nervous system
POD	Proper Orthogonal Decomposition.
SOFM	Kohonen's self organizing feature map
SOM	Self organizing map
SSOM	Spherical SOM
TG	Triglycerides
TSP	Travelling salesman problem
VLDL	Very low density lipoprotein

LIST OF SYMBOLS

Θ	Predetermined threshold value
$\varphi(\Theta)$	Activation function
$\alpha(t)$	learning rate during SOM training
σ	Width of the neighborhood function
S_j	Output of the Summed input layers data
X_i	Input data
W_{ij}	Weight between input layer and output layer
X_j	Output data
E	Error Function
D_n	n^{th} desired output
O_n	n^{th} output
$\|.\|$	Euclidean norm

ABSTRACT

In recent times, a lot of research has been going on in the field of nervous systems with a view of grasping and utilizing the acquired knowledge in the area of artificial intelligence. One of the branches of science inspired by the functioning of the brain is artificial neural networks. Self Organizing Maps (SOM) falls under artificial neural networks, and can be viewed as a visualization tool that projects high-dimensional dataset onto a two-dimensional plane thereby simplifying the complexity of the monitored data. The simplification in effect discloses much of the hidden details for easy analysis, clustering and visualization, but still preserving the details of original data. The pioneer of SOM algorithms, T. Kohonen, developed plane or flat SOM data mining tool. The tool has drawbacks in that it does not consider the neighborliness or relationship between the nodes appearing at the corners and edges of the lattice. The clusters formed at these regions have no similarity. In this research improved SOM tools Torus and Spherical that overcame the flat SOM drawbacks were developed.

One of the threatening trends of human health in recent years has been metabolic syndrome. Metabolic syndrome is a cluster of conditions that occur together resulting in simultaneous health disorders related to ones metabolism. Such disorders as obesity, particularly around the waist, elevated blood pressure, elevated level of the blood fat (triglycerides) (TG), low level of high-density lipoprotein cholesterol((HDL)) and resistance to insulin (a hormone that helps to regulate the amount of sugar in the body). The disorders are taken as parameters (variables) affecting a healthy system. Having one

component of metabolic syndrome means one is more likely to have others. The more components you have, the greater the risks to ones health.

The developed data mining tools were therefore, subsequently used to analyze and visualize metabolic syndrome as a risk to human health. The dataset parameters were Body Mass Index (BMI), High Blood Pressure (HBP), Blood Glucose (GLU), TG, Low Blood Pressure (LBP) and HDL. Using the developed Torus and Spherical SOM, real health data (4007 females and 2450 male test data) was used in the simulation. The contribution each risk (parameter) had on the syndrome was analyzed. Combination of parameters and their priority to induce the syndrome risk were also investigated. Also investigated using the same tools were the probable causes of the syndrome to both male and female examinees. The results obtained from the analysis compared very well with those diagnosed by the physicians thereby validating the Torus and spherical SOM. Referring to the sampling done on examinees, the ones diagnosed to be metabolic by the physicians were also found to be metabolic using the developed software. The developed simulators even reviewed trends the physicians could not have obtained at a glance. Moreover, after identifying the dominant parameters that contribute to the syndrome risk, specific (software) tools were developed to evaluate metabolic syndrome more accurately, particularly focusing on these main contributors. The developed tools were further used to formulate future trends the parameters may follow for a particular examinee. The metabolic SOM tools developed, display the metabolic syndrome risks in percentages with the most risk given 100 mark-points (MK). Examinees can equally observe the risk status they may be in. The developed tools can even predict the risk status the examinee may be in if the observed trend is not corrected through medical,

psychological or physical means in good time. The risk factors were further analyzed based on age. Using spherical SOM with the simulated data moderated to read metabolic points, the age cluster trends were formulated from which it was observed that each cluster responded differently towards the syndrome risk factors.

Analysis and visualization approach to metabolic syndrome developed here has initiated a different concept of understanding and appreciating the sources of the syndrome. The visualization tools are very handy for development of the trends the risk parameters may be taking. While the actual definition of metabolic syndrome may vary for the physicians, the clustering that occurs after training SOM becomes a useful map to aid the diagnosis of the examinees. Component maps generated from the trained SOMs showing how each risk parameter affects the overall metabolic map, become very helpful to the physician since dominant risk parameters become known. Furthermore, using the resulting trend maps, physicians are able to monitor the trends the risk parameters are taking.

The developed tools become an added opinion to the physician diagnosis. The examinees are themselves advantaged by the fact that SOM tools are self-explanatory maps and therefore they can observe the risk levels and the parameters causing them to be in the position their measured parameters have mapped them. It should however be noted that physician's comment need to be considered as the professional opinion to this form of SOM application. With this in mind, the tools were deliberately developed under constant consultations with physicians. The obtained results were precise and in agreement with the interpretations from the physicians, who are the experts. From the analyses TG, HBP and BMI were found to be the highest risk factors to metabolic

syndrome. Age clustering analysis isolated HBP as the most dominant risk factor. Finally, it is notable that the physician expert advice coupled with knowledge gained from examinee's interpretation of the maps, become an enhanced healing process and hence a quickened recovery period.

CHAPTER 1

INTRODUCTION

In a 1977 article [1] on Artificial intelligence (AI), the late pioneer of AI, Allen Newell, foresaw a time when the entire man-made world would be permeated by systems that cushioned us from dangers and increased our abilities: smart vehicles, roads, bridges, homes, offices, appliances, even clothes. According to Newell, systems built around AI components would increasingly monitor financial transactions, predict physical phenomena and economic trends, control regional transportation systems, and plan military and industrial operations amongst others. Artificial intelligence for the most part does not produce stand-alone systems, but instead adds knowledge and reasoning to existing applications, databases, and environments, to make them friendlier, smarter, and more sensitive to user behavior and changes in their environments. It has become one of the key technologies in many novel applications, ranging from banking systems that detect attempted credit card fraud, to telephone systems that understand speech, to software systems that notice when you're having problems and offer appropriate advice.

The ability to create intelligent machines has intrigued humans since ancient times [2]. Today, with the advent of the computer and engagement in research over the years into AI programming techniques, the dream of smart machines is becoming a reality. Within the area of AI various branches have emerged that focus on specialized sections depending on the concepts adopted by the researchers. This approach has in effect made

it possible to have new areas of research and development that can broadly be categorized as Expert Systems (ES) and Artificial Neural Networks (ANN).

1.1 Expert Systems

Expert systems is regarded as hard computing or precise computing where the procedures developed for computing the data are fixed. It is basically an intelligent computer program that is designed to implant the expertise of the human being in a certain domain based on the expert knowledge of the designer. It is concerned with both the concepts, methods of symbolic inference, or reasoning, and how the knowledge used to make those inferences will be represented inside the computer. The expert knowledge is represented as data or rules within the computer and often embedded as part of the programming code, so that as the knowledge changes, the program has to be changed and then rebuilt. These rules and data can be called upon when needed to solve problems. Thus, a different problem within the domain of the knowledge-base can be solved using the same program without reprogramming. The ability of these systems to explain the reasoning process through back-traces and to handle levels of confidence provides an additional feature that conventional programming does not handle. Building an expert system is known as knowledge engineering and its practitioners are called knowledge engineers. The knowledge engineer must make sure that the computer has all the knowledge needed to solve a problem [3].

1.2 Artificial Neural Networks (ANN)

Artificial neural networks deal with soft or approximate computing where a degree of uncertainty is accommodated. It encompasses Fuzzy Logics (FL), Genetic Algorithms (GA) and Self Organizing Maps (SOM).

1.21 Fuzzy Systems

Fuzzy systems are static or dynamic systems that make use of fuzzy sets or fuzzy logics to formulate rule-based systems that can be used to describe system behaviors in linguistic and qualitative terms. They can also be seen as nonlinear function approximation schemes. Fuzzy systems have close relationship with neural networks. In many cases there is a one-to-one mapping between fuzzy systems and neural networks and hence the term 'neuro-fuzzy' systems. A Fuzzy system can be viewed as a special parameterization form of nonlinear mapping. Fuzzy modeling and control, therefore, from many respects can be seen as a special class of nonlinear modeling and control. Fuzzy control is a 'soft computing' technique, which mimics the ability of the human mind to learn and make rational decisions in an uncertain imprecise environment [4]. Fuzzy system contains three main components, fuzzification, rule base and defuzzification. Fuzzification is used to transform the so-called crisp values of the input variables into fuzzy membership values. Afterwards, these membership values are processed within the rule base using conditional 'IF-THEN' statements. The outputs of the rules are summed and defuzzified into a crisp output value. Fuzzy logic controllers have become a technique of choice for many researchers in robotics and have allowed for

the integration of high-level, human designed plans to operate along side immediate, reactive actions in a successful manner.

1.22 Genetic Algorithms (GA)

A Genetic algorithm is an iterative procedure that consists of a constant-size population of individuals, each one represented by a finite string of symbols, known as the genome, encoding a possible solution in a given problem space. This space, referred to as the search space, comprises all possible solutions to the problem at hand. Generally speaking, the GA is applied to spaces which are too large to be exhaustively searched. The symbol alphabet used is often binary, though other representations have also been used, including character-based encodings, real-valued encodings, and most notably tree representations. The standard GA proceeds as follows: An initial population of individuals is generated at random or heuristically. At every evolutionary step, known as a generation, the individuals in the current population are decoded and evaluated according to some predefined quality criterion, referred to as the fitness, or fitness function. To form a new population (the next generation), individuals are selected according to their fitness. Many selection procedures are currently in use, one of the simplest being Holland's original fitness-proportionate selection, [5], where individuals are selected with a probability proportional to their relative fitness. This ensures that the expected number of times an individual is chosen is approximately proportional to its relative performance in the population. Thus, high-fitness (``good") individuals stand a better chance of ``reproducing", while low-fitness ones are more likely to disappear. Selection alone cannot introduce any new individuals into the population, that is, it

cannot find new points in the search space. These are generated by genetically-inspired operators, of which the most well known are crossover and mutation [6].

1.23 Self Organizing Maps (SOM)

SOM as described by Kohonen [7] are "visualization and analysis tools for high dimensional data" but can also be used for clustering, dimensionality reduction, classification, sampling, vector quantization and data mining. In this research SOM is taken as a mapping routine where multi-dimensional data is mapped onto two-dimensional surface for easy visualization, clustering and hence analysis and interpretation of the original complex data.

Kohonen's plane or flat SOM data mining tool does not consider the neighborliness or relationship between the nodes appearing at the corners and edges of the lattice. The clusters formed at these regions have no similarity. It is due to this drawback that necessitated the development of Torus and Spherical SOM tools. Torus SOM projects the trained data on a two dimensional surface that can be folded in toroidal form. Edge nodes become neighbors of other edge nodes. The corners of the trained lattice are also made to be adjacent. The clusters formed in these regions become neighbors. Spherical SOM has the trained data distributed around the sphere in relation to similarity and neighborliness of the trained data.

The developed tools can be applied to many fields requiring data mining. Some engineering applications would be in robotics, power system analysis and protection, telecommunications, image processing and motor vehicle industry. Having noted with a lot of concern the serious danger metabolic syndrome has posed to Kenya citizens due to

their lifestyle [Appendix A-2]; the developed tools were tested with the analysis and visualization of the syndrome.

1.3 Metabolic syndrome

Metabolic syndrome is a complex symptom of human body disorder that can be decoded using SOM tools with an intention of visualizing and analyzing health behavior patterns of an individual. Metabolic syndrome, also called insulin resistance syndrome or syndrome X can be defined as a cluster of metabolic risk factors that come together in a single individual increasing the risk of heart disease, stroke and diabetes [8]. Having metabolic syndrome means one has several disorders related to one's metabolism at the same time. These disorders include obesity, elevated blood pressure, elevated level triglycerides, low level of high-density lipoprotein cholesterol and resistance to insulin (a hormone that helps to regulate the amount of sugar in one's body). Having one component of metabolic syndrome means one is more likely to have others. Furthermore, the more components one has, the greater are the risks to one's health. Other conditions associated with the syndrome include physical inactivity, aging, hormonal imbalance and genetic predisposition.

In this work data mining tools (Torus and Spherical) were used in the testing of the syndrome using real dataset of patients' health parameters.

1.31 Causes of Metabolic syndrome

As is true with many medical conditions, genetics and the environment both play important roles in the development of the metabolic syndrome. Genetic factors influence

each individual component of the syndrome, and the syndrome itself. Research into the complexity underlying process linking the group of conditions involved in metabolic syndrome is ongoing [9]. As the name suggests, metabolic syndrome is tied to the body's metabolism, possibly to a condition metabolism, possibly to a condition called insulin resistance “medical experts’ dilemma”. Insulin is a hormone made by the pancreas to control the amount of sugar in the bloodstream. Normally, the digestive system breaks down some of the foods we eat into sugar (glucose). The blood carries the glucose to the body tissues, where the cells use it as energy sources. Glucose enters the cells with the help of insulin. In people with insulin resistance, cells then don't respond normally to insulin. As a result, glucose can't enter the cells easily. The body then reacts by churning out more and more insulin to help glucose get into the cells. The result is higher than normal levels of both insulin and glucose in the blood. Although perhaps not high enough to qualify as diabetes, an elevated glucose level still interferes with the body processes. Increased insulin raises triglyceride level as well as blood fat levels. It also interferes with how the kidneys work, leading to higher blood pressure. These combined effects of insulin resistance put one at risk of heart disease, stroke, diabetes and other conditions.

1.32 Diagnosing metabolic syndrome

There is no well-accepted criterion for diagnosing the metabolic syndrome. The criteria proposed by the National Cholesterol Education Program (NCEP) Adult Treatment Panel III (ATP III), with minor modifications, are currently recommended and widely used [10]. The American Heart Association and the National Heart, Lung, and

Blood Institute [11] recommend that the metabolic syndrome be identified as the presence of three or more of the components below:

1. **Abdominal obesity:** a waist circumference over 102 cm (40 in) in men and over 88 cm (35 inches) in women.
2. **Triglycerides:** 150 mg/dl or above.
3. **HDL cholesterol:** 40mg/dl or lower in men and 50mg/dl or lower in women.
4. **Blood pressure** of 130/85 or more.
5. **Blood glucose:** 110 mg/dl or above (some groups say 100mg/dl)

The World Health Organization (WHO) has slightly different criteria for the metabolic syndrome:

1. **High insulin levels**, elevated blood glucose with at least 2 of the following criteria:
2. **Abdominal obesity** as defined by a waist to hip ratio of greater than 0.9, a body mass index of at least 30 kg/m2 or a waist measurement over 94 cm (37 in).
3. **Cholesterol:** panel showing a triglyceride level of at least 150 mg/dl or HDL cholesterol lower than 35 mg/dl.
4. **Blood pressure** of 140/90 or above (or on treatment for high blood pressure).

In this work all parameters (components) considered in the medical field to have any contribution to metabolic syndrome were investigated.

1.4 Problem Statement

The real cause of metabolic syndrome to this date is not known and research is still going on with intensions of both understanding and developing more accurate methods of

diagnosing it. The global definition of having any three of the accepted parameters that are said to contribute to the syndrome is yet to be agreed upon by the medical experts. This is mainly due to the fact that the methods used to decode the syndrome are unable to narrow down to the root cause(s) of an individual's metabolic syndrome. The expert's argument that patients having one of the parameters affecting them may have other parameters accompanying it is not a satisfactory way of diagnosing patients' ailments. Precision on the side of the expert is not a debatable issue since we are dealing with human life. The vagueness appearing on the expert's advice to the patients has in itself psychological impact to the patients' attitude in response to medication and lifestyle changes deemed necessary by the expert.

Kenyans lifestyle has resulted in endangering their lives due to metabolic syndrome as reported in Kenya's daily [Appendix A-2]. The mentioned ambiguities in the definition and diagnosis of the syndrome indicate that the causes may be from complex combinations of input parameters. It thus became apparent that analysis and visualization of the syndrome became a better choice of testing the developed tools. The interrelations of the parameters believed to be associated with the syndrome and the trends the syndrome may follow on individuals or collectively when considering a cluster of members needed investigations. Furthermore the relation of the syndrome to age and sex required investigation.

Using the findings reported in this research, a lot of new correlations of various diseases namely blood pressures and diabetes to the syndrome has been reported [12-15]. This work was done in conjunction with physicians who are experts in this field. The findings reported were in agreement with their diagnosis. This prompted the need to

develop a metabolic evaluation tool for use in syndrome related ailments. The developed tools can also be used to predict the syndrome future trends with the help of examinees' health parameters database accumulated over the years. Individuals whose future trends fall in risky zones can be consoled and given medication depending on the correlations of the other monitored diseases. If the syndrome trend falls on healthy zone and particularly if the examinee's syndrome is from a risky zone, it becomes a sigh of relieve to the examinee.

Analysis of the syndrome through SOM also demonstrates the importance of passing more information to the examinee by including him or her in the decision making process. In effect the patient becomes a health conduit to solving any syndrome issue and thereby preventing patients being affected by the secondary disease "psychology". The research findings and developed tools therefore become necessary tools to physicians when dealing with syndrome diagnosis.

1.5 Objectives

The objectives of this work are:

1. Develop data analysis tools based on SOM.
2. To analyze the effects the variables identified by experts have on causing metabolic syndrome using SOM.
3. Develop metabolic syndrome evaluation tools derived from SOM that decode health dataset of patients and give their degree of metabolic risk in percentages.
4. Using the developed tools, identify the most probable causes of the syndrome and the trends the variables of the dataset take in promoting the metabolic syndrome risk.

5. Evaluate the results obtained from the developed tools by comparing medical reports of sampled diagnosis of patients.
6. Improve on the tools to give prediction of patients' metabolic syndrome with the help of their databases.

1.6 Organization of the thesis

The thesis is divided into six chapters:

Chapter 1, "**Literature Review,"** gives an overview of artificial intelligence and the branches that have emerged from it. Special attention is given to one data mining tool developed from the branch of artificial intelligence SOM. In this work, improved SOM tools developed from SOM-PAK (developed by Kohonen research team) idea are used in the analysis and visualization of metabolic syndrome.

Chapter 2 "Nervous System," discusses the functions of the biological nervous system with a view to understanding the links between the functions of nervous system and artificial intelligence.

Chapter 3, "Artificial Neural Networks (ANN)," focuses on artificial neural networks; the interpretation of the biological concepts of the brain, the learning phases, the propagation of the information and the tuning of the output to obtain the near exact image of the input.

Chapter 4, "Self Organizing Maps (SOM)," discusses SOM and expounds on the needs and benefits this area of science has to **the present generation**. Outline of SOM applications has also been given with a view of giving the reader an insight into its importance.

Chapter 5, "ANN Applications in Human Health," discusses SOM applications in particular to human health. Metabolic syndrome being a symptom of body disorder is investigated using SOM tools. The various parameters understood by physicians to be or suspected to be having an effect on the syndrome risk are used in the investigation. Pre-processing, simulations, mapping, visualizing, analysis and physician interpretation of the outputs are covered in this chapter.

Chapter 6, "Conclusions and Recommendations," gives an overall summary of the research findings giving recommendations and future projections of the research.

1.7 Chapter Summary

The chapter reviews the roots of ANN and its growth to various branches that are currently being applied in various fields. These branches include Fuzzy logic, Genetic algorithms and SOM to mention a few. Methods of data analysis are discussed highlighting SOM as the method used to develop metabolic syndrome software evaluation tools.

Input variables used in this work are:

1. **Abdominal obesity:** a waist circumference over 102 cm (40 in) in men and over 88 cm (35 inches) in women.
2. **Triglycerides:** 150 mg/dl or above.
3. **HDL cholesterol:** 40mg/dl or lower in men and 50mg/dl or lower in women.
4. **Blood pressure** of 130/85 or more.
5. **Blood glucose:** 110 mg/dl or above (some groups say 100mg/dl)
6. **Low blood pressure** of 90/60 or more.

CHAPTER 2

NERVOUS SYSTEM

2.1 General overview

The Nervous System is the most complicated and highly organized of the various systems which make up the human body. It is the mechanism concerned with the correlation and integration of various bodily processes and the reactions and adjustments of the organism to its environment. The system is a highly specialized network whose principal components are nerves called neurons. Neurons are interconnected to each other in complex arrangements and have the property of conducting, using electrochemical signals, and a great variety of stimuli within the nervous tissue as well as towards and from most of the other tissues [16]. Thus, neurons coordinate multiple functions in organisms. Nervous systems are also found in many multi-cellular animals but differ greatly in complexity between species. The system can be divided into two parts namely the central nervous system (CNS) and the peripheral nervous system (PNS)

The CNS of the vertebrate nervous system is enclosed in meninges. It contains the majority of the nervous system, and consists of the brain (in vertebrates which have brains), and the spinal cord. Together with the peripheral nervous system, it has a fundamental role in the control of behavior. The CNS is contained within the dorsal cavity, with the brain within the cranial cavity, and the spinal cord in the spinal cavity.

The brain is also protected by the skull, while the spinal cord is, in vertebrates, also protected by the vertebrae.

The PNS resides or extends outside the CNS, to serve the limbs and organs. It relays voluntary and involuntary information to and from the CNS: Unlike the central nervous system, however, the PNS is not protected by bone, leaving it exposed to toxins and mechanical injuries. PNS main components are its sensory and motor systems. The sensory nerves gather information from the environment; send it to the spinal cord, which then sends the message to the brain. The brain then makes sense of that message and fires off a response. Motor nerves on the other hand deliver the instructions from the brain to the rest of your body. PNS can be divided into somatic nervous system and the autonomic nervous system (ANS).

The somatic nervous system is the part of the peripheral nervous system associated with the voluntary control of body movements through the action of skeletal muscles, and with reception of external stimuli, which helps keep the body in touch with its surroundings (e.g., touch, hearing, and sight). The autonomic nervous system is the part of the peripheral nervous system that acts as a control system, maintaining homeostasis in the body. These activities are generally performed without conscious control or sensation. The ANS affects heart rate, digestion, respiration rate, salivation, perspiration, diameter of the pupils, micturition (urination), and sexual arousal to mention a few. Whereas most of its actions are involuntary, some, such as breathing, work in tandem with the conscious mind.

2.2 Brain

2.21 Introduction

The brain is the most complex and highly specialized of all mammalian organs. Understanding the complexity of its functions remains one of the greatest challenge to man. It produces our every thought, action, memory, feelings and experience of the world. This jelly-like mass of tissue, weighing approximately 1.4 kilograms, contains a staggering one hundred billion nerve cells, or neurons. The complexity of the connectivity between these cells is mind-boggling. Each neuron can make contact with thousands or even tens of thousands of others, via tiny structures called synapses. Our brains form a million new connections for every second of our lives. The pattern and strength of the connections is constantly changing and no two brains are alike. It is in these changing connections that memories are stored, habits learned and personalities shaped by reinforcing certain patterns of brain activity, and losing others [17].

The functional unit of the brain is the neuron, or excitable nerve cell which makes anatomic and chemical connections with other units in the brain that can then be termed as a system. Many of the essential biochemical connections of the nerve cell are dependent upon special morphological features: synaptic contact is mediated by chemical molecules, neurotransmitters which ensure the continued propagation of electrical impulses through sequential units of the system, chemical energy expended in maintaining distribution gradients of cations across cellular membranes. Nerve cells are unique in their ability to trigger off and maintain conduction of electrical impulses over long distances (in meters) without loss of strength of the conducted impulse. These unique features rest in the possession of semi-permeable excitable membranes which can

be caused, rapidly and transiently, to undergo changes in permeability to small chemical molecules and to cations. One aspect of the biochemical function of the brain can be seen in its efficient production of energy required to support the entire processes. The brain depends absolutely for its ability to function normally on a constant supply of glucose and oxygen from the blood stream. The importance of the constant blood supply of essential nutrients can readily be appreciated if we remember that this organ is about 3% of the total adult body weight but consumes about 20% of the glucose required by the whole body. The constant blood supply is such that one-fifth of the output from the heart passes through the brain.

2.22 Grey matter

While people often speak of their one's grey matter, the brain also contains white matter. The grey matter is the cell bodies of the neurons, while the white matter is the branching network of thread-like tendrils called dendrites and axons that spread out from the cell bodies to connect to other neurons.

The brain also has another, even more numerous type of cell, called glial cells. These outnumber neurons ten times over. Once thought to be support cells, they are now known to amplify neural signals and to be as important as neurons in mental calculations.

Brain structure is shaped partly by genes, but largely by experience. It has been observed that new brain cells are being born throughout our lives a process called neural genesis. The brain has burst of growth and then periods of consolidation, when excess connections are pruned. The most notable bursts are in the first two or three years of life, during puberty, and also a final burst in young adulthood.

How brain ages also depends on genes and lifestyle too. Exercising the brain and giving it the right diet can be just as important as it is for the rest of the body.

2.23 Microscopic appearance

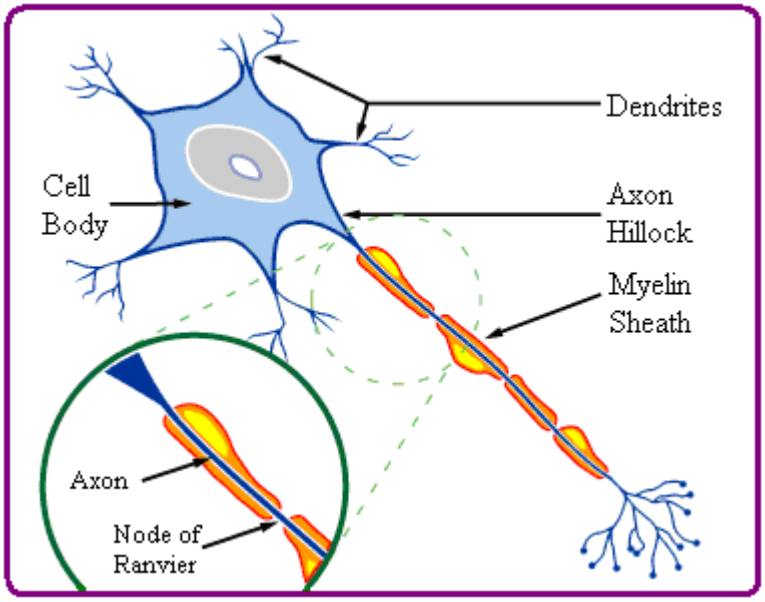
The brain is made of approximately 100 billion nerve cells, called neurons. Neurons have the amazing ability to gather and transmit electrochemical signals. They are something like the logic gates and wires in a computer. Neurons share the same characteristics and have the same parts as other cells, but the electrochemical aspect lets them transmit signals over long distances (a few meters) and pass messages to each other. Neurons have three basic parts:

Cell body: This main part has all of the necessary components of the cell, such as the nucleus (contains Deoxyribonucleic acid (DNA)), endoplasmic reticulum and ribosomes (for building proteins) and mitochondria (for making energy) [18]. It contains most of the cell's genetic material, organized as multiple long linear DNA molecules in complex with a large variety of proteins, such as histones, to form chromosomes. The genes within these chromosomes are the cell's nuclear genome. The function of the nucleus is to maintain the integrity of these genes and to control the activities of the cell through regulating gene expressions. If the cell body dies, the neuron dies.

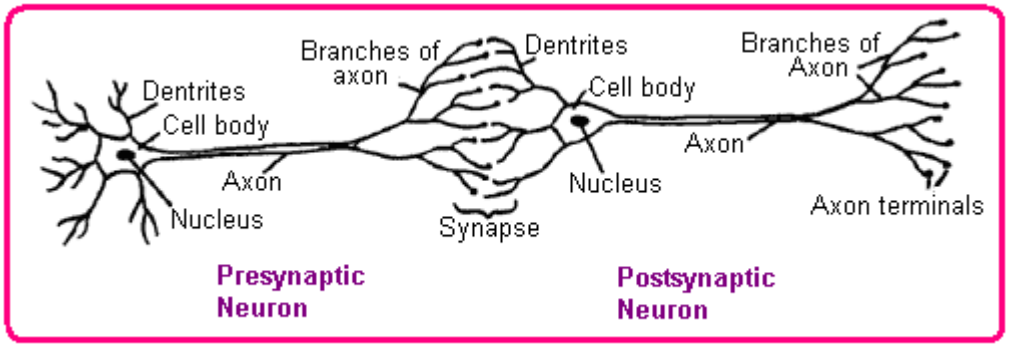
Axon: This long, cable-like projection of the cell carries the electrochemical message along the length of the cell.

Dendrites or **nerve endings:** These small, branch-like projections of the cell make connections to other cells and allow the neuron to talk with other cells or perceive the environment. Dendrites can be located on one or both ends of the cell.

Figure 2-1 (a) shows the three basic parts of the neuron. A section of the axon is magnified to indicate the ranvier node whereas Figure 2-1 (b) shows how information is propagated along the neurons. Information passes from the axon of the pre-synaptic neuron to the dendrites of the postsynaptic neuron.



(a)



(b)

Figure 2- 1: (a) Basic human neuron (b) Neuron network

2.24 Types of nerve cells

Neurons come in many sizes from sizes of millimeters to arms lengths as illustrated by figure 2-2. They have different shapes depending on what they do. Motor neurons that control muscle contractions for example have a cell body on one end, a long axon in the middle and dendrites on the other end; sensory neurons have dendrites on both ends, connected by a long axon with a cell body in the middle.

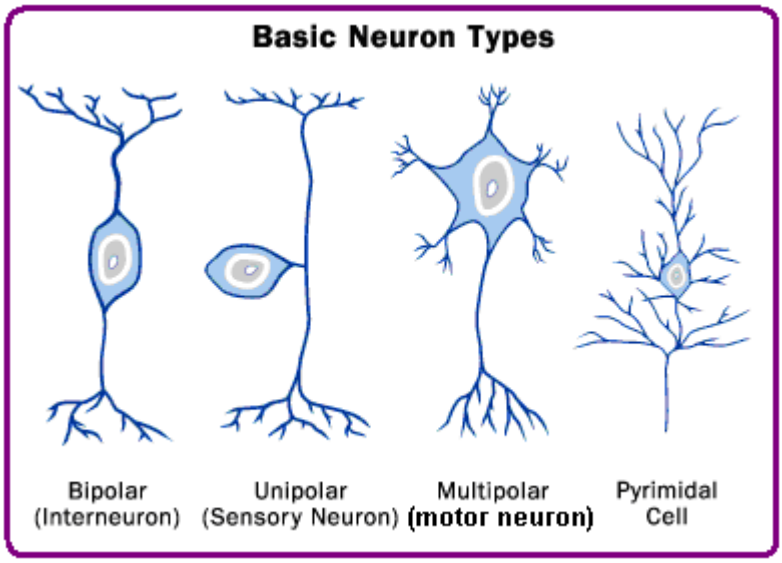


Figure 2- 2: Types of neurons

Neurons can be classified by their number of processes (or appendages), or by their functions. If the number of processes classifies them, they fall into three categories:

Unipolar neurons: Have a single process (dendrites and axon are located on the same stem), and are most common in invertebrates.

Bipolar neurons: Have the dendrite and axon as neuron's two separate processes. Bipolar neurons have a subclass called pseudo-bipolar neurons, which are used to send sensory information to the spinal cord.

Multi-polar neurons: Most common in mammals. Examples of these neurons are spinal motor neurons, pyramidal cells and Purkinje cells (in the cerebellum).

If classified by their function, neurons again fall into three separate categories. The first group is sensory or afferent neurons, which provide information for perception and motor coordination. The second group provides information (or instructions) to muscles and glands and is therefore called motor neurons. The last group, inter-neuronal, contains all other neurons and has two subclasses. One group called relay or projection inter-neurons have long axons and connect different parts of the brain. The other group called local inter-neurons are only used in local circuits [19], [20].

Glial cells provide support and protection for neurons, the other main type of cell in the brain. They are thus known as the "glue" of the brain. The four main functions of glial cells are to surround neurons and hold them in place, to supply nutrients and oxygen to neurons, to insulate one neuron from another, and to destroy pathogens and remove dead neurons.

2.25 Basis of excitability and impulse propagation

Neurons convey information using electrical and chemical signals. Understanding how neurotransmission occurs is crucial to understanding how the brain processes and integrates information. Interruption of neural communication causes changes in cognitive processes and behavior.

A neuron may have many dendrites, which branch out in a treelike structure, and receive signals from other neurons. A neuron usually only has one axon which grows out from a part of the cell body called the axon hillock. The axon conducts electric signals generated at the axon hillock down its length. These electric signals are called action

potentials. The other end of the axon may split into several branches, which end in a pre-synaptic terminal. Action potentials are the electric signals that neurons use to convey information to the brain. All these signals are identical. Therefore, the brain determines what type of information is being received based on the path that the signal took. The brain analyzes the patterns of signals being sent and from that information it can interpret the type of information being received. Myelin is a fatty tissue that surrounds and insulates the axon though along the same length of the axon, some parts are not insulated called nodes of Ranvier. At these nodes, the signal traveling down the axon is regenerated (boosted). This ensures that the signal traveling down the axon travels fast and remains constant (i.e. very short propagation delay and no weakening of the signal).

Action potential voltage as shown in figure 2-3 (a), results from the flow of ions across the neuronal cell membrane. Neurons, like all cells, maintain a balance of ions inside the cell that differs from the balance outside the cell. This uneven distribution of ions creates an electrical potential across the cell membrane. This is called the resting membrane potential. In humans, the resting membrane potential ranges from −40 mV to −80 mV, with −65 mV as an average resting membrane potential. The resting membrane potential is, by convention, assigned a negative number because the inside of the neuron is more negatively charged than the outside of the neuron. This negative charge results from the unequal distribution of sodium ions (Na^+), potassium ions (K^+), chloride ions (Cl^-), and other organic ions. The resting membrane potential is maintained by an energy-dependent Na^+-K^+ pump that keeps Na^+ levels low inside the neuron and K^+ levels high inside the neuron. In addition, the neuronal membrane is more permeable to K^+ than it is to Na^+, so K^+ tends to leak out of the cell more readily than Na^+ diffuses into

the cell. If for some reason the Na^+ channels are opened so that the permeability is increased, Na^+ will flow into the cell thereby reducing the membrane potential to a less negative potential hence depolarization. If on the other hand the membrane permeability for K^+ is increased, the membrane potential will be more negative than the resting potential (hyper polarization) [21]

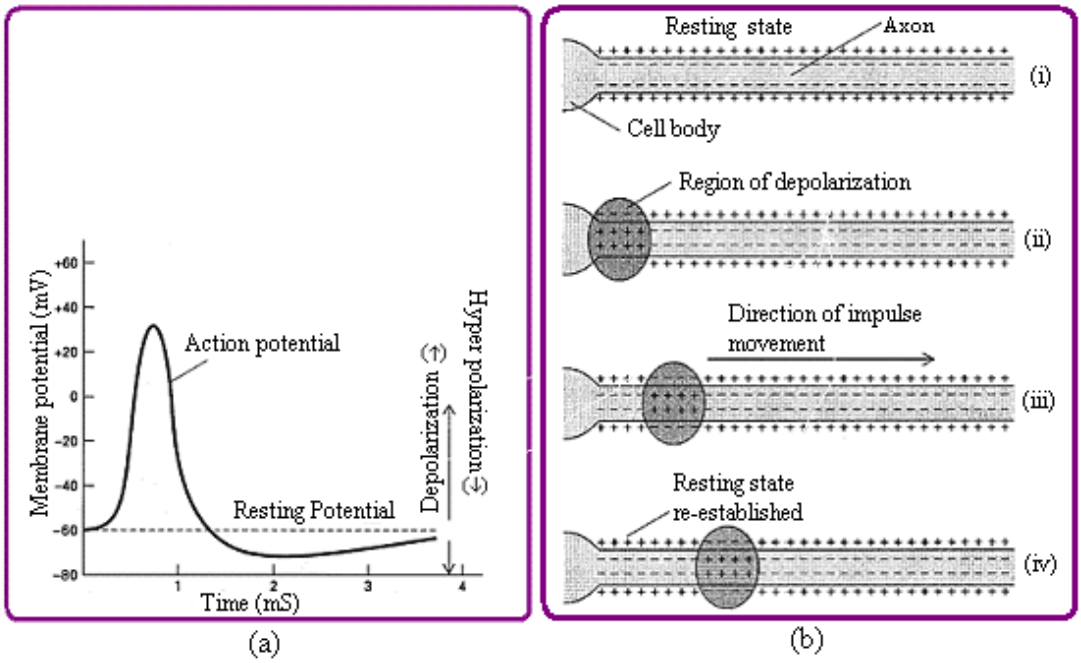


Figure 2- 3: Neuron: (a) Action potential voltage (b) Impulse propagation.

In the event of a stimulus occurring at the end of a nerve fiber, a change in the permeability of the neuronal membrane occurs. An action potential develops with Sodium ions rushing into the neuron, making the inside of the cell more positive. The Na^+-K^+ pump then restores the balance of sodium and potassium to resting levels. However, the influx of Na^+ ions in one area of the neuron fiber starts a similar change in the adjoining segment as shown in figure 2-3 (b) (ii), and the impulse moves from one end of the neuronal fiber to the other.

Action potentials are an all-or-none phenomenon. Regardless of the stimuli, the amplitude and duration of an action potential are the same. The action potential either occurs or it doesn't. The response of the neuron to an action potential depends on how many action potentials it transmits and the time interval between them. This leads us to a frequency coding phenomenon. Many neurons produce trains of action potentials in rapid successions and then pause for a while before a new train of impulses is produced. The frequency of action potential in some neurons may be more than 100Hz whereas others may be much lower. Each neuron type has its characteristic frequency pattern caused by differences in membrane properties and synaptic inputs. The code for the information carried by an axon is the frequency and pattern of action potential, since the action potential always is the same.

2.26 Synaptic junction

The synapse is the area of contact between two neurons. The neurons do not physically touch but are separated by the synaptic cleft. The synaptic junction filled with neurotransmitter fluid accelerates or retards the flow of electrical charges. The adjustment of the fluids controls the impedance or conductance of the synaptic gap and in effect contributes to the memory or learning process of the brain. This leads us to believe that the brain has a distributed memory or intelligence characteristics giving it the property of associative memory, but not like a digital computer's central storage memory

Figure 2-4 is a model of a synapse junction. When the electrical signal reaches the end of the axon, it triggers a series of chemical changes in the neuron. Calcium ions (Ca^{2+}) flow into the neuron. The increased Ca^{2+} in the axon terminal then initiates the release of neurotransmitters (stored in membranous sacs called vesicles**)** from the pre-

synaptic neuron terminals into the synaptic cleft. Each vesicle contains thousands of molecules of a neurotransmitter. For neurons to release their neurotransmitter, the vesicles fuse with the neuronal membrane and then release their contents [22].

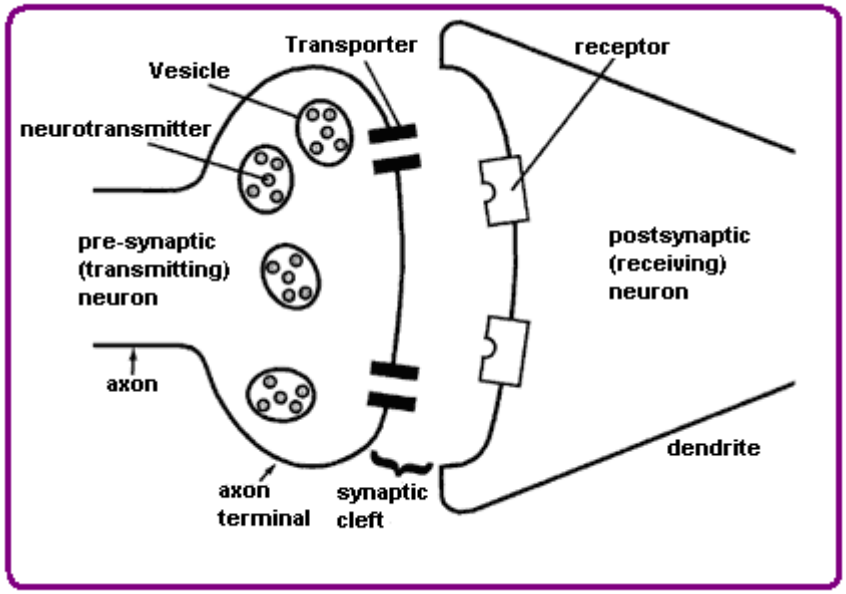


Figure 2- 4: Synaptic Junction

The released neurotransmitter molecules can then bind to specific receptors on the postsynaptic neuron membrane to elicit a response as shown in figure 2-5.

The released neurotransmitter as well as any neurotransmitter that did not bind to a receptor, is either degraded by enzymes in the synaptic cleft or taken back up into the pre-synaptic axon terminal by active transport through a transporter or reuptake pump. Once the neurotransmitter is back inside the axon terminal, it is either destroyed or repackaged into new vesicles that may be released the next time the neuron is stimulated. Different neurotransmitters are inactivated in different ways.

It is worth pointing out that an average neuron forms approximately 1,000 synapses with other neurons. It has been estimated that there are more synapses in the human brain than

there are stars in our galaxy; furthermore, synaptic connections are not static. Neurons form new synapses or strengthen synaptic connections in response to life experiences. This dynamic change in neuronal connections is the basis of learning. Once the target neuron receives the chemical message, it then reconverts it to an electrical impulse to continue the process of propagation.

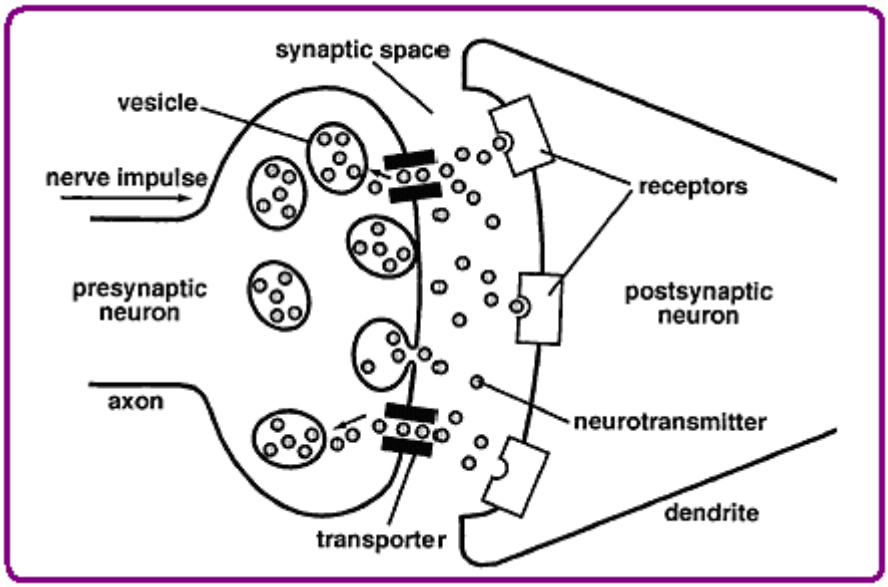


Figure 2- 5: Synaptic junction response to an electrical impulse

The postsynaptic neuron often receives both excitatory and inhibitory messages. The response of the postsynaptic cell depends on which message is stronger. Keep in mind that a single neurotransmitter molecule cannot cause an action potential in the responding neuron. An action potential occurs when many neurotransmitter molecules bind to and activate their receptors. Each interaction contributes to the membrane permeability changes that generate the resultant action potential.

The connection between neurons is through the synapse, it is reasonable to guess that whatever changes occur during learning, takes place there. If either the pre-or

postsynaptic cell were altered as a whole, other responses could be reinforced that are unrelated to the conditioning information.

2.3 Chapter Summary

This chapter focuses on how the brain trains in response to various input stimulus and some of the activities involved in the process of training. Signal propagation and its effect to a neuron and thereafter influencing neighboring neurons is addressed. The brain concepts are borrowed in the research and growth of artificial intelligence.

CHAPTER 3

ARTIFICIAL NEURAL NETWORKS (ANN)

3.1 Introduction

3.11 ANN background

An ANN is a system based on the operation of biological neural networks, or is an emulation of biological neural system. The first ANN was invented in 1958 by psychologist Frank Rosenblatt and was called Perceptron. It was intended to model how the human brain processed visual data and learned to recognize objects. Other researchers have since used similar ANNs to study human cognition [23]. Eventually, someone realized that in addition to providing insights into the functionality of the human brain, ANNs could be useful tools in their own right. Their pattern-matching and learning capabilities allowed them to address many problems that were difficult or impossible to solve by standard computational and statistical methods. By the late 1980s, many real-world institutes were using ANNs for a variety of purposes. ANNs are among the newest signal-processing technologies in the engineer's toolbox and serve two important functions: as pattern classifiers and as nonlinear adaptive filters.

An ANN operates by creating connections between many different processing elements, each analogous to a single neuron in a biological brain. These neurons may be physically constructed or simulated by a digital computer. Referring to figure 3-1, each neuron takes many input signals, then, based on an internal weighting system, produces a

single output signal that is typically sent as input to another neuron. The neurons are tightly interconnected and organized into different layers. The input layer receives the input; the output layer produces the final output. In addition, one or more hidden layers are sandwiched between the two layers. This structure makes it impossible to predict or know the exact flow of data.

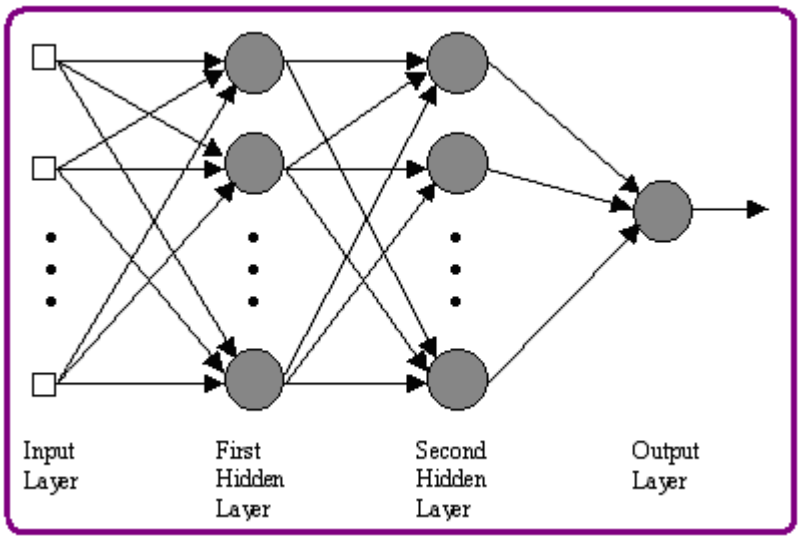


Figure 3- 1: Common multilayer perceptron

3.12 Reasons for using neural networks

Implemented on a single computer, an ANN is typically slower than a more traditional algorithmic solution. The ANN's parallel nature, however, allows it to be built using multiple processors, giving it a great speed advantage at very little development cost. The parallel architecture also allows ANNs to process very large amounts of data very efficiently. When dealing with large, continuous streams of information, such as speech recognition or machine sensor data, ANNs can operate considerably faster than their linear counterparts.

3.13 Economic Uses

The economic uses of ANNs may be the most exciting. Large financial institutions have used ANNs to improve performance in such areas as bond rating, credit scoring, target marketing, credit card transactions to detect likely instances of fraud and evaluating loan applications. ANNs are used to discover other kinds of crime, too. Bomb detectors and rooting out of corrupt officers in government offices. ANNs that have been used to adjust temperature settings and diagnose malfunctions in machine controls are now monitoring automated and robotic factories. These ANNs can augment or replace skilled labor, making it possible for fewer people to do more work.

3.2 Modes of Training ANNs

3.21 Structure of an ANN

When creating a functional model of the biological neuron, there are three basic components of importance. First, the synapses of the neuron are modeled as weights. The strength of the connection between an input and a neuron is noted by the value of the weight. Negative weight values reflect inhibitory connections, while positive values designate excitatory connections [24]. The next two components model the actual activity within the neuron cell. An adder sums up all the inputs modified by their respective weights. This activity is referred to as linear combination. Finally, an activation function controls the amplitude of the output of the neuron. An acceptable range of output is usually between 0 and 1, or -1 and 1 as detailed in 3.22. A neural network unit thus performs a relatively simple job: receive input from neighbors or external sources, processes it to compute an output signal which is propagated to other units.

ANNs typically start out with randomized weights for all their neurons. This means that they don't "know" anything and must be trained to solve the particular problem for which they are intended. Broadly speaking, there are two methods for training an ANN, depending on the problem it must solve. A self-organizing ANN (often called a Kohonen after its inventor) is exposed to large amounts of data and tends to discover patterns and relationships in that data. Researchers often use this type to analyze experimental data. A back-propagation ANN, conversely, is trained by humans to perform specific tasks. During the training period, the teacher evaluates whether the ANN's output is correct. If it's correct, the neural weightings that produced that output are reinforced; if the output is incorrect, those weightings responsible are diminished. This type is most often used for cognitive research and for problem-solving applications.

All ANNs are variations on the parallel distributed processing (PDP) idea. The system is inherently parallel in the sense that many units can carry out their computations at the same time. The architecture of each neural network is based on very similar building blocks which perform the processing.

The interconnection of artificial neurons results in an ANN and its objective is to emulate the functions of the human brain to solve scientific, engineering and many other real-life problems. These networks may be feedforward and feedback (recurrent) types. Feedforward networks have signals flowing from neuron to neuron in only the forward direction. Some of the feedforward models available are:

- Perceptron
- Adaline and Madaline
- Back propagation network

- Fuzzy neural network (FNN)

Examples of recurrent types are:

- Hopfield network
- Kohonen's self organizing feature map (SOFM)
- Bi-directional associative memory (BAM)

The above types of models are not exhausted, but it is worth noting that any problem that can be solved by recurrent network can also be solved by a feed-forward network with the proper external connections.

3.22 Perceptron

Using a simple two-layer perceptron as an example of an ANN, there is one layer of input nodes (layer 1) and the other of output nodes (layer 2). Figure 3-2 illustrates the model representation of the neuron (perceptron). Each layer is fully connected between the other, but no connections exist between nodes in the same layer. When layer 1 sends a signal to layer 2, the associated weights on the connections are applied and each receiving node on layer 2 sums up the incoming values. If the sum exceeds a given threshold, that node in turn fires an output signal.

The input nodes are assumed to be of unity values such that the outputs from them are joined directly to output layer (j) as shown in figure 3-3.

The outputs are summed across all the inputs (X_i) received by a node (j) in the output layer. From the model shown in figure 3-3, the interval activity of the neuron can be shown to be:

$$S_j = \sum_{i=0}^{n} X_i W_{ij} \quad \text{If } S_j > \Theta \text{ THEN } X_j = 1. \text{ If } S_j \leq \Theta \text{ THEN } X_j = 0 \qquad (3.1)$$

where Θ is the predetermined threshold value.

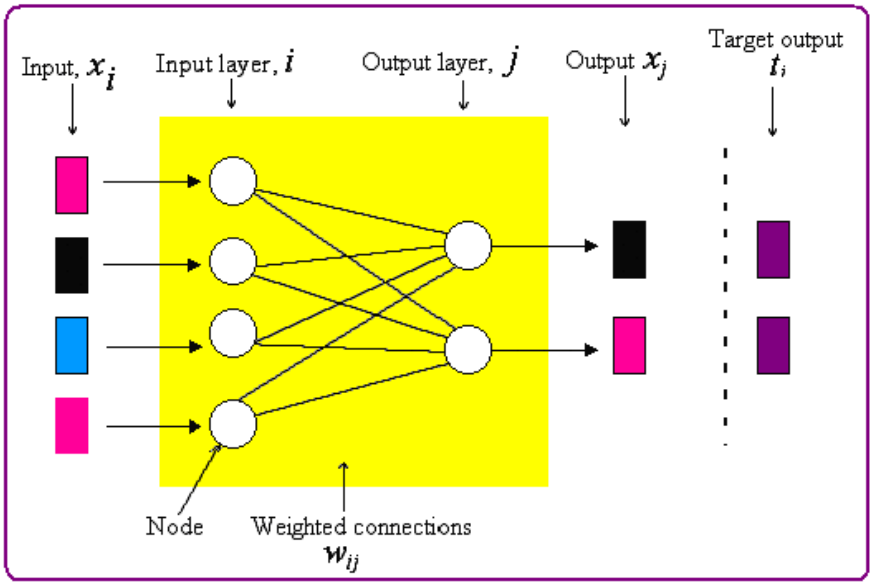


Figure 3- 2: Simple perceptron

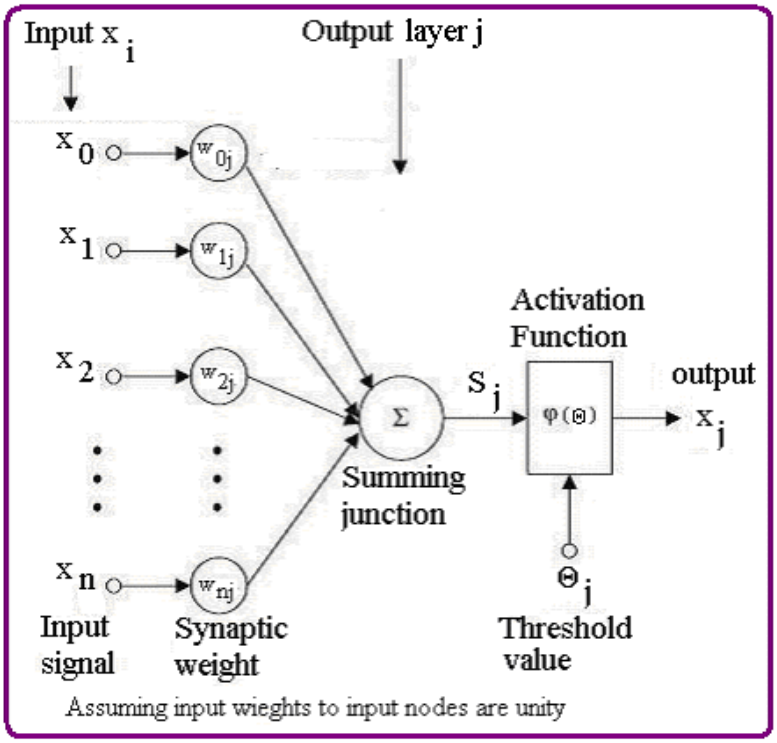


Figure 3- 3: Model representation of a neuron

3.3 Training ANNs

The learning process can be categorized into two distinct parts namely supervised and unsupervised:

Supervised learning or **Associative learning**: The learning process proceeds by way of presenting the network with a training set composed of input patterns together with the required response pattern. The net will then produce some firing activity on its output layer which can be compared with a `target' output. By comparing the output of the network with the target output for that pattern we can measure the error the network is making. These input-output pairs can be provided by an external teacher, or by the system which contains the neural network (self-supervised). A back-propagation ANN, conversely, is trained by humans to perform specific tasks. During the training period, the teacher evaluates whether the ANN's output is correct. If it's correct, the neural weightings that produced that output are reinforced. If the output is incorrect, those weightings responsible are diminished. This type is most often used for cognitive research and for problem-solving applications.

Unsupervised learning or Self-organization: An output unit is trained to respond to clusters of pattern within the input. In this paradigm the system is supposed to discover statistically salient features of the input population. Unlike the supervised learning paradigm, there is no apriori set of categories into which the patterns are to be classified; rather the system must develop its own representation of the input stimuli. A self-organizing ANN (often called a Kohonen after its inventor) is exposed to large amounts of data and tends to discover patterns and relationships in that data. Researchers often use this type to analyze experimental data.

Typical examples are the Hebbian learning rule, and the competitive learning rule:

A simple version of Hebbian learning rule is that when unit *i* and unit *j* are simultaneously excited, the strength of the connection between them increases in proportion to the product of their activations.

Competitive learning is a rule based on the idea that only one neuron from a given iteration in a given layer will fire at a time. Weights are adjusted such that only one neuron in a layer, for instance the output layer, fire. Competitive learning is useful for classification of input patterns into a discrete set of output classes. The "winner" of each iteration, element i^*, is the element whose total weighted input is the largest

Example Hebbian Learning:

> When an axon of cell ***A*** is near enough to excite a cell ***B*** and repeatedly or persistently takes part in the firing it, some growth process or metabolic change takes place in one of the cells such that ***A's*** efficiency as one of the cells firing ***B*** is increased.

Suppose: Excitation of ***C*** caused by sight of food is sufficient to excite ***B***, causing salivation. Suppose without any additional stimulus, the excitation of ***A***, resulting from the hearing of a bell is not sufficient to cause firing of ***B***. Let us now consider excitation ***C*** causes ***B*** to fire and during this time excitation ***A*** is added. Hebb's assumption dictates that, some changes occur between ***A*** and ***B*** so that ***A's*** influence on ***B*** increases. If the experiment is repeated often enough, ***A*** will eventually be able to cause ***B*** to fire even with the absence of visual stimulus ***C***.

An example of a competitive learning:

In pattern recognition, if a new pattern is determined to belong to a previously recognized cluster, then the inclusion of the new pattern into that cluster will affect the representation (e.g., the centroid) of the cluster. This will in turn change the weights characterizing the classification network. If the new pattern is determined to belong to none of the previously recognized cluster, then (the structure and the weights of) the neural network will be adjusted to accommodate a new class (cluster).

Reinforcement Learning: This type of learning may be considered as an intermediate form of the above two types of learning. Here the learning machine does some action on the environment and gets a feedback response from the environment. The learning system grades its action good (rewarding) or bad (punishable) based on the environmental response and accordingly adjusts its parameters. Generally, parameter adjustment is continued until an equilibrium state occurs, following which there will be no more changes in its parameters. The self organizing neural learning may be categorized under this type of learning.

Training of the ANN shown in figure 3-2 is accomplished by adjusting the weights on the connections between layers to match a desired output. If there is a difference between the actual and the target outputs, the weights are adjusted on the adaptive layer to produce a set of outputs closer to the target values. New weights are determined by adding an error correction value to the old weight. The amount of the correction is determined by applying Equation 3.2. If the input node output (X_j) is a 1, that connection weight is not adjusted, and if it sends 0, its connection weight is subsequently adjusted.

Thus, the process can be summed as follows:

$$w_{ij}(new) = w_{ij}(old) - C(t_j - x_j)x_i \text{; where C is the learning rate constant} \qquad (3.2)$$

This training procedure is repeated until the network's performance no longer improves. The network is then said to have "converged". At this point, it has either successfully learned the training set or it has failed to learn all of the answers correctly. If it is successful, it can be given new sets of input and generally produce correct results on its own.

The functional limitation of a two-layer perceptron, however, is that it can only recognize linearly separable patterns due to only having one adaptive layer. A linearly separable pattern is one that can be separated into two distinct classes by drawing a single line. However, this limitation fell to the wayside after the introduction of the back error propagation paradigm, that extends the perceptron by implementing a multiple, hidden layer network which is also referred to as a multiple layer perceptron (MLP) [25], [26].

When two or more layers of weights are adjusted, the network has middle or hidden layers of processing units. Each hidden layer responds to specific features in the input pattern. These feature detectors (hidden layers) organize as learning takes place, and are developed in such a way that they accomplish the specific learning task presented in the network.

Activation functions

In general, there are three types of activation functions, denoted by $\varphi(\Theta)$. First, there is the threshold function that takes on a value of 0 if the summed input is less than or equal to a certain threshold value (Θ), and the value 1 if the summed input is greater than the threshold value as shown in Equation (3.1). Equation 3.3 refers to a Piecewise-Linear

function. This function again can take on the values of 0 or 1, but can also take on values between depending on the amplification factor in a certain region of linear operation.

$$\varphi(\Theta) = \begin{cases} 1 & \Theta \geq \frac{1}{2} \\ \Theta & -\frac{1}{2} < \Theta < \frac{1}{2} \\ 0 & \Theta \leq -\frac{1}{2} \end{cases} \qquad (3.3)$$

Thirdly, there is the sigmoid function (Equation 3.4). This function can range between 0 and 1 as shown in figure 3-4, but it is also sometimes useful to use the -1 to 1 range.

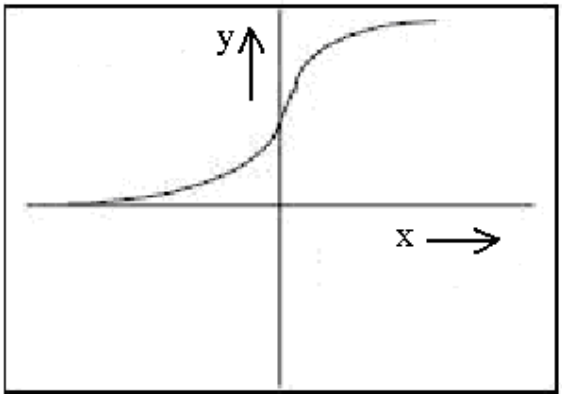


Figure 3- 4: Sigmoid function

$$y = \frac{1}{1 + e^{-x}} \qquad (3.4)$$

3.4 Back propagation network (BPN)

3.41 Background

When an ANN is presented with data, the output will not be the desired output since the network has not undergone training. Since the network weights are initially random, it is likely that the initial output value will be very far from the desired output. To

improve the behavior of the network and to know which connection weights need modification and by how much so as to achieve the objective, an algorithm commonly used is the back-propagation algorithm. This is simply a gradient descent method of minimizing the total squared error of the output computed by the network. The principal advantages of back-propagation are simplicity and reasonable speed. Back-propagation is well suited to pattern recognition problems.

The network learns a predefined set of input-output example pairs by using a two-phase propagate-adapt cycle. An input pattern is applied as a stimulus to the first layer of network units. It is propagated through each upper layer until an output is generated. The output is then compared to desired output. The error is transmitted backward from output later to each node in the intermediate layer that contributes directly to the output. Each unit in the intermediate layer receives only a portion of the error signal. This process repeats itself layer by layer until each node in the network has received an error signal that describes its relative contribution to the total error. One of the significant features of back propagation is that the intermediate layers organize themselves such that different nodes learn to recognize different features of the total input space. After training, when presented with arbitrary input patterns that are noisy or incomplete, the units in the hidden layers of the network will respond with an active output if the new input contains a pattern that resembles the feature the individual units learned to recognize during training.

Hidden layers inhibit their outputs if the input pattern does not contain the features they were trained to recognize. Upper layer patterns can be thought of as a pattern with features that can be recognized by units in the subsequent layer. The output pattern

generated can be thought of as a feature map that provides an indication of the presence or absence of many different feature combinations at the input.

BPN provides an effective means of allowing a computer system to examine data patterns that may be incomplete or noisy, and to recognize subtle patterns from the partial input. BPN will classify these previously unseen inputs according to the features they share with the training examples.

3.42 Selection and Preparation of Training Data

A neural network is useless if it only sees one example of a matching input/output pair. It cannot infer the characteristics of the input data for which you are looking for from only one example; rather, many examples are required. This is analogous to a child learning the difference between (say) different types of animals. The child will need to see several examples of each to be able to classify an arbitrary animal. It is the same with neural networks. The best training procedure is to compile a wide range of examples (for more complex problems, more examples are required) which exhibit all the different characteristics you are interested in. It is important to select examples which do not have major dominant features which are of no interest to you, but are common to your input data anyway. If possible, prior to training, add some noise or other randomness to your example (such as a random scaling factor). This helps to account for noise and natural variability in real data, and tends to produce a more reliable network [27].

If you are using a standard sigmoid node transfer function (not scaled), please note that the desired output must never be set to exactly 0 or 1. This is because the asymptotes values of the hidden layers are set between 0 and 1. This would then mean more data and weights are required to reach the desired output. Again the limits cannot be exceeded.

To avoid this limitation, the desired output can be lowered to (say) 0.9 for the network to ultimately reach and even overshoot. The network will converge relatively quickly. It cannot be overemphasized that a neural network is only as good as the training data. Poor training data inevitably leads to an unreliable and unpredictable network.

3.43 Modification of the neuron connection weights

Consider the example in Figure 3-5:

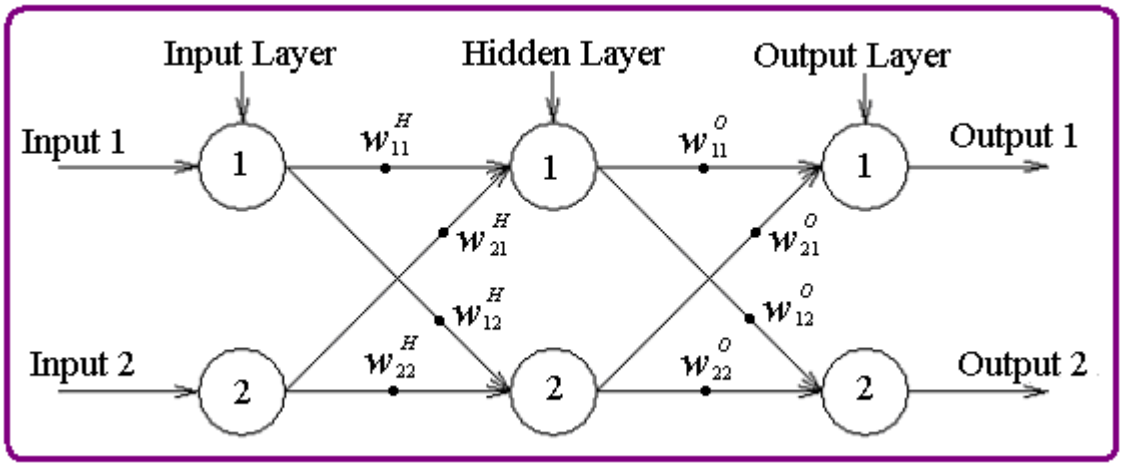


Figure 3- 5: 2-Input, 2-Output ANN.

If I_1, I_2 are the Inputs, H_1 H_2 the hidden layers outputs and O_1, O_2 are the output layer outputs respectively then:

Outputs of Hidden Node 1 and 2 are given by [28], [29]

$$H_1 = sgm\left(\sum_{l=1}^{2} I_l w_{l1}^{H}\right) \tag{3.5}$$

and

$$H_2 = sgm\left(\sum_{l=1}^{2} I_l w_{l2}^{H}\right) \tag{3.6}$$

where

$$sgm(x) = \frac{1}{1+e^{-x}} \tag{3.7}$$

Output-layer outputs are given by:

$$O_1 = sgm\left(\sum_{m=1}^{2} H_m w_{m1}^{O}\right) \tag{3.8}$$

and

$$O_2 = sgm\left(\sum_{m=1}^{2} H_m w_{m2}^{O}\right) \tag{3.9}$$

From equations (3.8) and (3.9), we can calculate the output given a particular set of inputs. This allows us to calculate the Mean Squared Error (MSE) between the actual output and the desired output for the given input in this training example. This gives us the average of what we want and what we got.

The error function can be written as:

$$E = \sum_{n=1}^{2} (D_n - O_n)^2 \tag{3.10}$$

where D_n is the n^{th} desired output.

or, using (3.5) and (3.8),

$$E = \sum_{n=1}^{2}\left(D_n - sgm\left(\sum_{m=1}^{2} sgm\left(\sum_{l=1}^{2} I_l w_{lm}^{H}\right) w_{mn}^{O}\right)\right)^2 \tag{3.11}$$

For a given training the minimum error would be the best target to our training. To find this point, the gradient of the error function with respect to each network weight would be calculated. Adjustments of the weights would then follow the opposite to the gradient.

The gradient is fairly straightforward to calculate, due to the convenient fact that the derivative of the sigmoid function can be expressed in terms of the function itself:

$$\frac{d}{dx}\left(\frac{1}{1+e^{-x}}\right)=\frac{e^{-x}}{(1+e^{-x})^2}=(1-sgm(x))sgm(x) \tag{3.12}$$

The gradient is defined as the vector of partial derivatives of the multivariate function with respect to each of variable. The error function for each network output is calculated as a set of partial derivatives with respect to each associated connection weight. All other variables except one are held constant when we calculate the partial derivative.

$$\frac{\partial O_n}{\partial W_{mn}^o}=\frac{\partial}{\partial W_{mn}^o}\sum_{k=1}^{2}W_{kn}^o H_k=H_m \tag{3.13}$$

The gradient of the error function can be calculated as:

$$\frac{\partial E}{\partial W_{mn}^o}=\frac{\partial}{\partial W_{mn}^o}\sum_{n=1}^{2}(D_n-O_n)^2$$

$$=-2(D_n-O_n)\frac{\partial}{\partial S^o}sgm(S^o)\frac{\partial S^o}{\partial W_{mn}^o}$$

$$=-2(D_n-O_n)((1-sgm(S^o))sgm(S^o))H_m \tag{3.14}$$

where $S^o=\sum_{k=1}^{2}W_{mn}^o$.

The expression '$-2(D_n-O_n)((1-sgm(S^o))sgm(S^o))H_m$' is denoted as δ_n^o

The new values for the network weights are calculated by multiplying the negative gradient with a step size parameter called the **learning rate** and adding the resultant vector to the vector of network weights attached to the current layer. This change does not take place, however, until after the middle-layer weights are updated as well, since this would corrupt the weight-update procedure for the middle layer [30].

For the middle layer, a new gradient is derived, but this time the output weights are treated as constants rather than the hidden-layer weights.

$$\frac{\partial E}{\partial W_{lm}^{H}} = ((1 - sgm(S^{H}))sgm(S^{H}))\sum_{n=1}^{2}\delta_{n}^{o}W_{mn}^{o}I_{l} \tag{3.15}$$

The middle weights are updated using the same procedure as for the output layer, and the output layer weights are updated as well. This is a complete training cycle for one piece of training data. The input layer is treated as a buffer for holding the input vector hence has no weights that need modifications.

The sample considered is a (2, 2, 2) network. Lager networks would have longer summations though the learning principle is the same.

3.44 Repetition

The above procedure causes the output move a small step towards the desired state of a minimized error. The procedure must be repeated many times until the MSE drops below a specified value. When this happens, the network is performing satisfactorily, and this training session for this particular example has been completed.

Training will be said to be successful when random data is applied to the input terminals repeatedly for 'many' times (twenty or less or ten thousand or more) depending on the application and complexity of the data and other parameters. After training the network, real data new to the network is presented to the input for classification, compression or processing.

A consequence of BPN algorithm is that there are situations where it can get stuck to a 'local minima' that traps the algorithm and prevents it from dropping to the actual minimum. If such a situation arises hidden layers can be added or reduced, nodes can be reduced or increased or try another starting point (randomize the network again). Other approaches to the BPN problem are based on alternative determination of MSE (least squares approximation and steepest-descent technique) [31], [32].

3.5 Chapter Summary

This chapter gives a model representation of basic human neurons as seen as artificial neural networks. The concept of training and the layers of neurons associated with the model are discussed. Back propagation network is used as an example to explain the method of adjusting nodes weights so that expected target can be met. SOM algorithm uses two layers namely input and output layers in its mapping process.

CHAPTER 4

SELF ORGANIZING MAPS (SOM)

4.1 Introduction

SOM is a data visualization technique invented by Professor Teuvo Kohonen which reduces the dimensions of data through the use of self-organizing artificial neural networks [33]. The problem that this technique attempts to solve is that humans simply cannot visualize high dimensional data so as to understand it. The way SOM goes about reducing dimensions is by producing a map of usually 1 or 2 dimensions which plot the similarities of the data by grouping similar data items together. Thus SOM accomplishes two things; reduce dimensions and display similarities.

The basic SOM can be visualized as a sheet-like neural-network array as shown in figure 4-1 where the cells (or nodes) become specifically tuned to various input signal patterns or classes of patterns in an orderly fashion. The learning process is competitive and unsupervised, meaning that no teacher is needed to define the correct output (or actually the cell into which the input is mapped) for an input. The goal is to group similar nodes close together in certain areas of the value range. The resultant maps are organized in such a way that similar data are mapped on the same node or to neighboring nodes on the map. This leads to a special clustering of similar input patterns in neighboring parts of the SOM and the clusters that appear on the map are themselves organized. SOM uses a distribution preserving property which has the ability to allocate more nodes to input

patterns that appear more frequently during the training phase of the network configuration. Thus the topology of the n-dimensional space is captured by the SOM and reflected in the ordering of its nodes. The input data is thus projected onto a lower dimension space while roughly preserving the order of the data in its original space. The learning process is unsupervised meaning that the training patterns have no category information that follows them. Figure 4-1 also shows the multi-dimensional dataset with input vectors x_i and the best matching unit (BMU) of a particular input vector marked red (winner node) on the node lattice. The neighboring nodes are indicated with colors that fade as the distance from the winner node varies. The lattice has dimensions that need initialization.

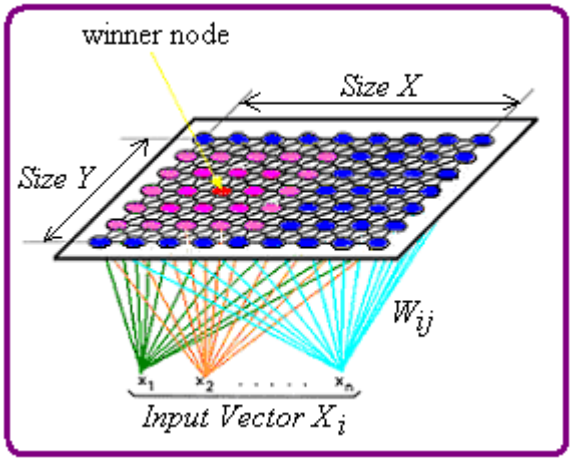


Figure 4- 1: Node Lattice

It is worth pointing out that during a simulation run; only one winner node is activated corresponding to each input. Similar inputs are mapped to the winner while inputs with similarities are mapped to neighboring nodes depending on their degree of similarities. Different clusters will be formulated using the same methods of simulations. The locations of the responses in the array tend to become ordered in the learning process

as if some meaningful nonlinear coordinate system for the different input features were being created over the network. Depending on the number of iterations decided, the clustering will through the training process converge to clearly defined clusters with boundaries between them well demarcated. It is through the advancement of computer technology that advantages like high speed and parallel processing can be used to make such simulations possible [Appendix A-1], [34].

Figure 4-2 illustrates the idea of what a SOM output looks like. The input samples are three dimensional data of colors red yellow and blue. The input colors are then grouped using SOM training process such that they can better be understood visually. The display shows like colors grouped together such as the greens are all in the upper left hand corner and the purples are all grouped around the lower right and right hand side.

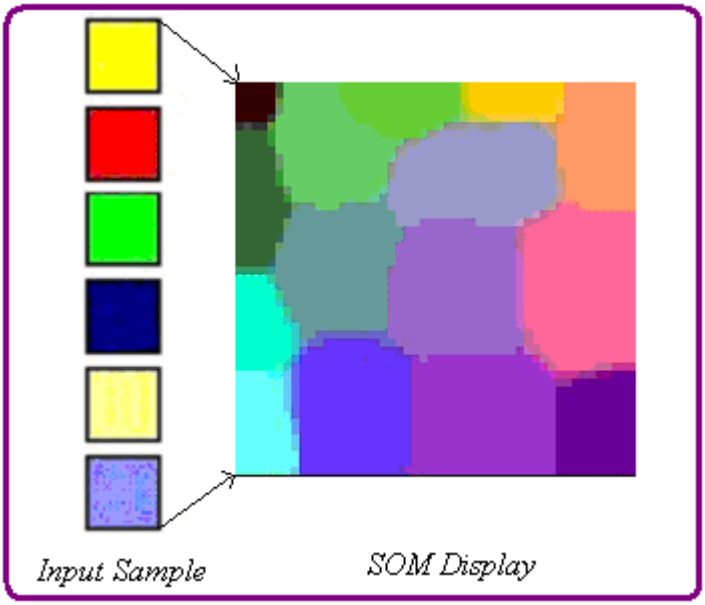


Figure 4- 2: Color grouping using SOM

4.2 Competitive learning

A competitive learning network comprises of feed-forward excitatory network(s) and lateral inhibitory network(s). The feed-forward network usually implements an excitatory Hebbian learning rule**;** when an input cell persistently participates in firing an output cell the input cell's influence firing that output cell is increased. The lateral competitive network is inhibitory in nature. The output layer of the network can be represented as a two-dimensional grid, also known as the competitive layer. The input values are continuous, typically normalized to any value between -1 and +1. The input vector is compared with the weight vectors leading to the competitive layer. The node with a weight vector most closely matching the input vector is called the winning node. The scheme is a winner takes it all process where the output unit receiving the largest input is assigned a full value (e.g. 1) whereas all the other units are suppressed to a 0 value [35].

4.3 SOM Algorithms

The Self Organizing Map is called a competitive algorithm because units compete to represent the input pattern. SOM algorithm chooses the winning unit by comparing the current input pattern against the weight vector of each of the output units. It provides a topology preserving mapping from the high dimensional space to map units (neurons). Map units usually form a two-dimensional lattice and thus the mapping is a mapping from high dimensional space onto a plane. The property of topology preserving means that the mapping preserves the relative distance between the points. Points that are near each other in the input space are mapped to nearby map units in the SOM. The SOM can thus serve as a cluster analyzing tool of high-dimensional data. Also, the SOM has the

capability to generalize. Generalization capability means that the network can recognize or characterize inputs it has never encountered before

The topology, or the structure, of the map has the neurons connected to each other via rectangular or hexagonal topology. Figure 4-3 shows the topological relations connected by lines between the neurons. One can also define a distance between the map units according to their topology relations. Immediate neighbors (the neurons that are adjacent) belong to the neighborhood N_c of the neuron M_c. The neighborhood function should be a decreasing function of time

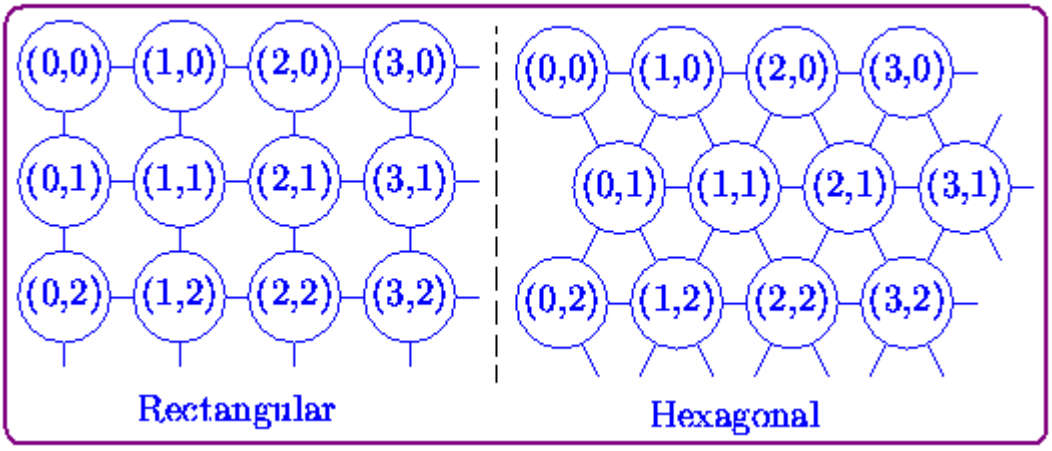


Figure 4- 3: Different SOM node topologies

Neighborhoods of different sizes in a hexagonal lattice are illustrated in Figure 4-4. In the smallest hexagon, there are all the neighbors belonging to the smallest neighborhood of the neuron in the middle belonging to a hexagonal lattice. The topological relations between the neurons are left out for clarity.

In the basic SOM algorithm, the topological relations and the number of neurons are fixed from the beginning. This number of neurons determines the scale or the granularity of the resulting model. Scale selection affects the accuracy and the generalization

capability of the model. It must be taken into account that the generalization and accuracy are contradictory goals. By improving the first, we lose on the second, and vice versa.

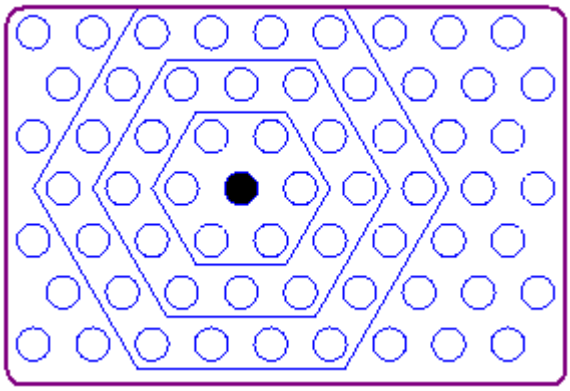


Figure 4- 4: Neighborhood of a given winner unit

4.31 Learning Vector Quantization

In a learning vector quantization (LVQ) machine, the input values are compared to the weight vector of each node. The weight vector of the BMU and those of nearby nodes are adjusted to be closer to the input vector by a certain step size. Nodes become trained to be individual feature detectors, and a combination of feature detectors can be used to identify large classes of features from the input space. LVQ training is a type of competitive learning rule [36].

Example: if the input vector is (0.35, 0.8), the winning node might have weight vector (0.4, 0.78). The learning rule would adjust the weight vector to make it even closer to the input vector. Only the winning node produces output, and only the winning node gets its weights adjusted. In more sophisticated models, only the weights of the winning node and its immediate neighbors are updated. Once a weight of the winner is determined, the neighbors of that weight are found and each of those neighbors in addition to the winning weight change to become more like the sample vector. If the

average distance were high, then the surrounding weights are very different and a dark color is assigned to the location of the weight. If the average distance is low, a lighter color is assigned. So in areas of the center of the blobs the colors are the same, so it should be white since all the neighbors are the same color. In areas between blobs where there are similarities it should be not white, but a light grey [37]. Areas where the blobs are physically close to each other, but are not similar at all are shaded black.

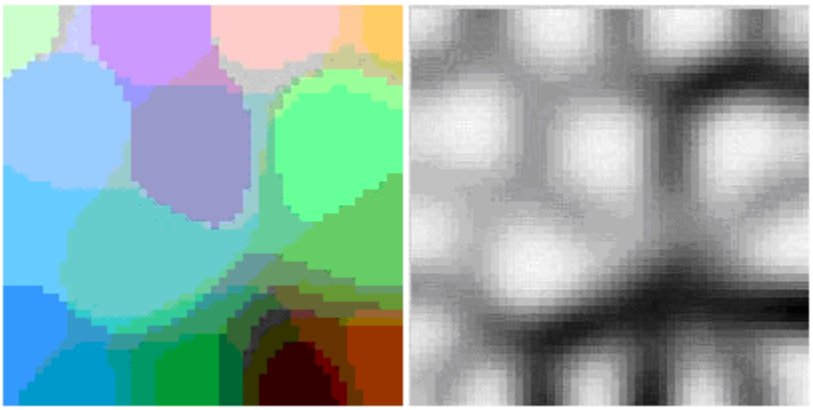


Figure 4- 5: Clustering colors using SOM

Referring to figure 4-5, the ravines of black show where the colors may be physically close to each other on the map, but are very different from each other when it comes to the actual values of the weights. Areas where there is a light grey between the blobs represent a true similarity. In the pictures above, in the bottom right there is black surrounded by colors which are not very similar to it. When looking at the black and white similarity SOM, it shows that black is not similar to the other colors because there are lines of black representing no similarity between those two colors. Also in the top corner there is pink and nearby is a light green which are not very near each other in reality, but near each other on the colored SOM. Looking at the black and white SOM, it

clearly shows that the two are not very similar by having black in between the two colors. With these average distances used to make the black and white map, we can actually assign each SOM a value that determines how good the image represents the similarities of the samples by simply adding these averages.

4.32 Basic Principles of SOM

Let's consider a d-dimensional dataset as a set of input vectors of d dimensions $\mathbf{x}_i = \{\mathbf{x}_1, \mathbf{x}_2, ...\mathbf{x}_n\}$, where n is the size of the dataset. The SOM training algorithm involves essentially two processes, namely vector quantization and vector projection. Vector quantization is to create a representative set of vectors called output vectors from the input vectors. Let's denote the output vectors as $\mathbf{m}_i = \{\mathbf{m}_1, \mathbf{m}_2, ...\mathbf{m}_k\}$ with the same dimension as the input vectors. In general vector quantization reduces the number of vectors, and this can be considered as a clustering process. Vector projection aims at projecting the k output vectors (in the d-dimensional space) onto a lower dimension with k neurons. In the vector projection each output vector is projected into a neuron where the projection is performed as such, "close" output vectors in the d-dimensional space will be projected onto neighboring neurons in SOM. This will ensure that the initial pattern of the input data will be present in the neurons. Usually the number of input vectors is greater than that of the output vectors, i.e. n>k and the size of SOM is the same as that of the output vectors.

SOM is trained iteratively with each training step sampling input vector $\mathbf{x}_i$ and the distance between it and all the weight vectors ($\mathbf{m}_i$) of the lattice are calculated joined by

scalar weights w_{ij}. The node whose weight vector is closest to the input vector is the BMU denoted here as c:

Euclidean norm of the vector **x** is defined as

$$\|\mathbf{x}\| = \sqrt{\sum_{i=1}^{n} \mathbf{x}_i^{\,2}} \tag{4.1}$$

Then, we can define the Euclidean distance in terms of the Euclidean norm of the difference between two vectors:

$$d_E(\mathbf{x}, \mathbf{y}) = \|\mathbf{x} - \mathbf{y}\| \tag{4.2}$$

The best-matching unit, usually noted as $\mathbf{m}_c$, is the codebook vector that matches a given input vector **x.** It is defined formally as the neuron

$$\|\mathbf{x} - \mathbf{m}_c\| = \min_i \{\|\mathbf{x} - \mathbf{m}_i\|\}, \tag{4.3}$$

where, $\mathbf{m}_i$ is the reference vector of each node on the lattice and $\mathbf{m}_c$ the winner node vector. After the winning node *c* is selected, the weights of the nodes in a neighborhood (defined) are adjusted so that similar input patterns are more likely to select this node again. This is achieved through computation:

$$\begin{aligned} &\mathbf{m_i}(t+1) = \mathbf{m_i}(t) + \alpha(t) h_{ci}(t)[\mathbf{x}(t) - \mathbf{m_i}(t)], && \text{for } i \in N_c(t) \\ &\mathbf{m_i}(t+1) = \mathbf{m_i}(t), && \text{for } i \notin N_c(t) \end{aligned} \tag{4.4}$$

where $\mathbf{x}(t)$ is a sample vector randomly taken from input vectors, $\mathbf{m}_i(t)$ is the output vector for any neuron *i* within the neighborhood $N_c(t)$, and $0<\alpha(t)<1$ and $h_{ci}(t)$ are the

learning rate and neighborhood kernel function around the winner unit *c* which is often taken to be Gaussian respectively.

$$h_{ci}(t) = \exp\left(-\left\|\mathbf{r}_i - \mathbf{r}_c\right\|^2 \Big/ 2\sigma^2(t)\right) \qquad (4.5)$$

where *t*, is the discrete-time index of the variables, $\mathbf{r}_i$ and $\mathbf{r}_c$, vectorial locations in the display grid and σ, the width of the neighborhood function which decreases monotonically with the regression steps. The learning rate is taken as a linear function $\alpha(t)$ = A/ (t + B), with 'A' and 'B' taken as suitably selected constants [38], [39].

The winning weight is rewarded with becoming more like the sample vector. The neighbors also become more like the sample vector. An attribute of this learning process is that the farther away the neighbor is from the winning vector, the less it learns. The rate at which the amount a weight can learn decreases with time and a choice of representation is a Gaussian function as shown in figure 4-6.

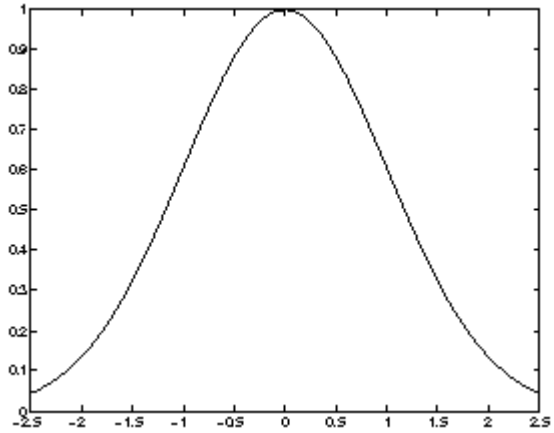


Figure 4- 6: Gaussian function representation.

This function will return a value ranging between 0 and 1. In the first iteration, the best matching unit will get a value α (*t*) of 1 for its learning function, so the weight will then

come out of this process with the same exact values as the randomly selected sample. Just as the amount a neighbor's weight falls, so is its learning rate.

4.4 Visualization of SOM

The Self-Organizing Map is an approximation to the probability density function of the input data. It can be used in visualization. In the next sections, we present common ways to visualize the Self-Organizing Map.

4.41 U-Matrix Representation

Unified distance matrix (U-Matrix) representation of the Self-Organizing Map visualizes the distances between the neurons. The distance between the adjacent neuons is calculated and presented with different colorings between the adjacent nodes. A dark coloring between the neurons corresponds to a large distance and thus a gap between the codebook values in the input space. A light coloring between the neurons signifies that the codebook vectors are close to each other in the input space. Light areas can be thought as clusters and dark areas as cluster separators. This can be a helpful presentation when one tries to find clusters in the input data without having any a priori information about the clusters.

In figure 4-7, we can see the neurons of the network marked as black dots. The representation reveals that these are a separate cluster in the upper right corner of this representation. The clusters are separated by a dark gap. This result was achieved by unsupervised learning, that is, without human intervention. Teaching a SOM and representing it with the U-Matrix offers a fast way to get insight of the data distribution.

The U-matrix makes the 2D visualization of multivariate data possible using SOM's code vectors as data source [40]. This is achieved by using topological relations property among nodes after the learning process.

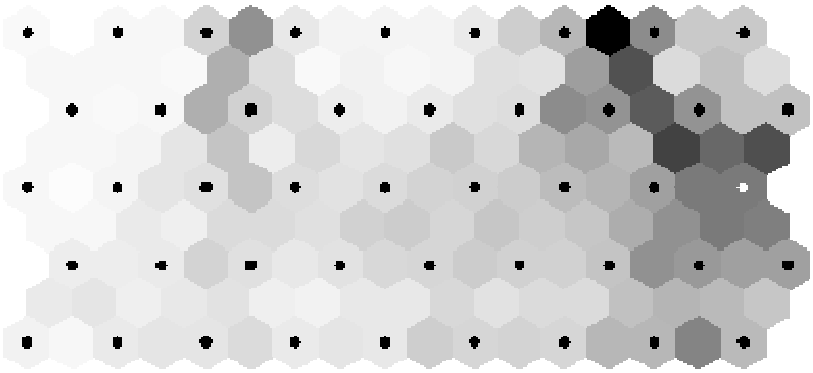


Figure 4- 7: U-Matrix representation of the Self-Organizing Map

The algorithm generates a matrix where each component is a distance measure between two adjacent nodes, thereby being able to visualize a multi-variant dataset in a two dimensional display.

4.41.1 U-matrix Algorithm

Clustering algorithms are methods to divide a set of n observations into g groups (called clusters) so that members of the same group are more alike than members of different groups. Since the main point about a cluster solution is that members of the same group are more alike than members of different groups, one has to have a means of measuring the likeness of such members. The most obvious measure of similarity (or dissimilarity) between two members is the distance between them. To state things more formally,

Let the clustering algorithm be a mapping q that assigns to each of n input vectors $\boldsymbol{x}$ a reproduction (codebook) vector $\hat{\mathbf{x}} = q(\mathbf{x})$ drawn from a finite reproduction alphabet $\hat{A} = \{\hat{\mathbf{x}}_i, i = 1,...,n\}$. Expected, is an algorithm that produces a mapping q for which the expected distortion caused by reproducing the input vectors $\boldsymbol{x}$ by codebook vectors $q(\boldsymbol{x})$is at least locally minimal. The expected distortion is usually estimated by using the average distortion D, where the most common distortion measure is the squared-error distortion (i.e. Euclidean distance) d:

$$D = \frac{1}{n}\sum_{j=0}^{n-1} d(\mathbf{x}_j, q(\mathbf{x_j}))$$
$$d(\mathbf{x}, \hat{\mathbf{x}}) = \sum_{o=0}^{k-1} \left| \mathbf{x}_o - \hat{\mathbf{x}}_o \right|^2 \qquad (4.6)$$

The average distortion D measures the total squared error of representing n samples $x_0,...,x_{n-1}$ by N codebook vectors (cluster centers) $\hat{\mathbf{x}}_1,...,\hat{\mathbf{x}}_N$. Both $\mathbf{x}$ and $\hat{\mathbf{x}}$ are of dimensions k. D is small if all the distances between cluster members and cluster centers within each cluster are small. A classical technique to achieve such a clustering is the K-means cluster (KMC) approach consisting of the following steps:

1. Initialization: Given N= number of codebook vectors, k = dimensionality of the vectors, n = number of input vectors, a training sequence $\{\mathbf{x}_j;\ j = 0,...n-1\}$ initial set $\hat{\mathbf{A}}_0$ of N codebook vectors $\hat{\mathbf{x}}$ and a discrete-time coordinate $t = 0,...,n-1$.

2. Given $\{\hat{\mathbf{A}}_t = \hat{\mathbf{x}}_i; i = 1,...,N\}$, find the minimum distortion partition $\{P(\hat{A}_t) = \{S_i; i = 1,...,N\}$ Compute $d(\mathbf{x_t}, \hat{\mathbf{x}}_i)$ for $i=1\ldots N$. If $d(\mathbf{x_t}, \hat{\mathbf{x}}_i) \leq (\mathbf{x_t}, \hat{\mathbf{x}}_l)$ for all l, then $\mathbf{x_t} \in S_i$

3. Update the codebook vector with the minimum distortion

$$\hat{\mathbf{x}}_{(t+1)}(S_i) = \hat{\mathbf{x}}_{(t)}(S_i) + \alpha[\mathbf{x}_{(t)} - \hat{\mathbf{x}}_{(t)}(S_i)] \quad (4.7)$$

Where α is a learning parameter to be defined by the user. Define $A_{(t+1)} = \hat{\mathbf{x}}(P(A_t))$, replace *t* by *t + 1*. If *t = n-1*, halt. Else go to step 2.

The only difference between the SOM-algorithm and KMC is the fact that the codebook vectors are the weight vectors of SOM's output units which are ordered either on a line or on a planar grid (i.e. in a one or two dimensional output space). The iterative procedure is the same as with KMC where Equation. (4.7) is replaced by;

$$\hat{\mathbf{x}}_{(t+1)}(S_t) = \hat{\mathbf{x}}_{(t)} + h[\mathbf{x}_t - \hat{\mathbf{x}}_{(t)}(S_t)] \quad (4.8)$$

and this update is not only computed for the $\hat{\mathbf{x}}_t$ that gives minimum distortion, but also for all the codebook vectors which are in the neighborhood of this $\hat{\mathbf{x}}_t$ on the line or planar grid. The degree of neighborhood and amount of codebook vectors which are updated together with $\hat{\mathbf{x}}_t$ that gives minimum distortion is expressed by *h*, a function that decreases both with distance on the line or planar grid and with time.

The neighborhood term *h* includes an additional learning parameter α since codebook vectors in close neighborhood in the one or two dimensional output space are always updated together, the result of the SOM algorithm is a set of topologically ordered codebook vectors, i.e. codebook vectors which distance is small in the low dimensional output space are also close to each other in the input space. If the degree of neighborhood is decreased to zero, the SOM-algorithm becomes essentially equal to the KMC-algorithm. Since clustering problems in general are plagued by a large number of local minima, the solutions obtained via SOM and KMC might nevertheless be different due to

different trajectories through the search space. Whereas local convergence is guaranteed for KMC (at least for decreasing α [41]), no general proof for the convergence of SOM with nonzero neighborhood is known. It should be noted that the last step of the SOM algorithm should be computed with zero neighborhood in order to guarantee "the most accurate density approximation of the input samples". One of the main problems in clustering data is to decide for the correct number of clusters (i.e. codebook vectors) [42]. Clearly *N*, the number of cluster centers or output units, should be equal to g, the number of clusters present in the data. In [7] it is argued that one should compute successive partitions of the data with an ever growing number of clusters *N*. If samples are really grouped into *g c*ompact, well separated clusters, one would expect to see any error function based on or between cluster variance (the same obviously holds for average distortion) decrease rapidly until $N = g$. Such error functions should decrease much more slowly thereafter until they reach zero at $N = n$.

4.42 Component Plane Representation

The various parameters that constitute the entire input data can be monitored individually using component plane representation. It can be thought of as a sliced version of the Self-Organizing Map. Each component plane has the relative distribution of one data vector component. By comparing component planes we can see if two components correlate. Component map gives a better view of the contribution that principal component has on the overall output SOM. If the outlook is similar, the components strongly correlate.

Colored component maps are an added advantage to the user. Figure 4-8 shows integrated component maps with various distinct color-coding for each component. Gray

represents the gray scale indicating the strength of the bonding between the various principal components. Gray scale is used here to indicate both strength of bonding between input parameters and the metabolic risk. The numerical value given in the gray scale shows how bonded the parameters are while the color indicates the metabolic risk. Principal component maps have their individual legend. The numerical value displayed on the SOM map represents the output of some logic. Taking for example BMI, value 4 is used to represent the component normalized range values ($1 \geq BMI \geq 0.8$). Other SOM displays can be explained by taking an example of a BMI SOM display of 3. This value represents normalized component input values in the range ($0.8 > BMI \geq 0.6$). The color shown on each principal component cell indicates the metabolic risk based on that input parameter. This approach gives the viewer a better visualization of the input data and hence an alternative method to decoding the input data.

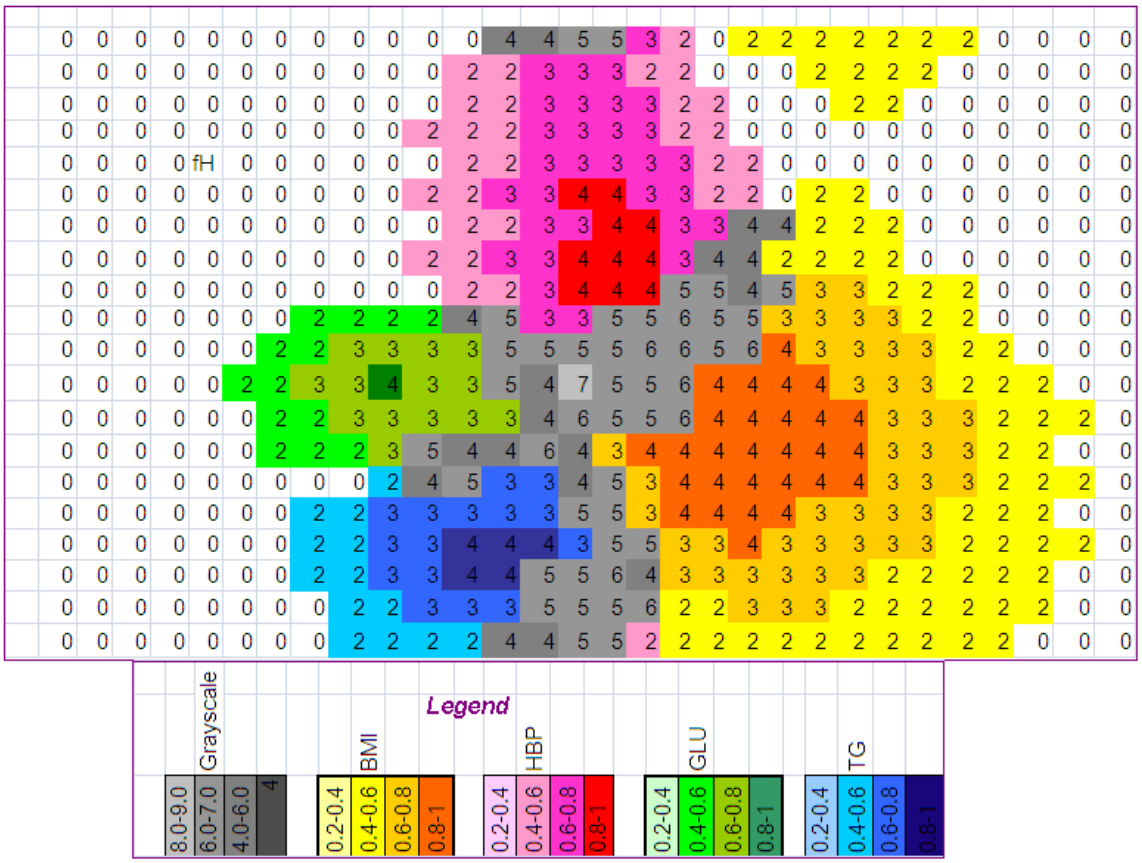


Figure 4- 8: SOM component map

4.5 SOM simulations

4.51 Introduction

A self-organizing feature map is a neural computing program and its output is used typically for categorizing complex multi-dimensional data. The program not only computes the category of objects in a dataset but also learns to compute the same. Typically the system is given an *n*-dimensional input, representing a set of objects in the dataset, which the program, through a process of regression, maps onto a two-dimensional surface. Each of the *n* dimensions refers to a property of the objects that either relates to features it may share with other objects in a given category together with

features that may distinguish it. The neural network usually has two layers: the *n* dimensional input layer and the two-dimensional output layer. The objects in the datasets are 'won over' by nodes in the output layer in a winner-takes-all manner, and one node may be associated with more than one object in the training dataset, and some nodes may not be able to win over any of the objects. If categorically similar objects occupy a neighborhood of the nodes, then the program clusters them [43]. Once the map is created, and presented with an object, particularly one it wasn't trained on, the system then assigns a category to the novel or unknown object. One of the key advantages of using the program is that it preserves the topology of the input data despite reducing it onto a surface. The category information is not directly available, in that, like many other neural computing systems, the output is not as discernible as may be the case for other learning systems.

Simulations based on SOM follows pre-defined steps as described in 4.52, 4.53 and 4.54.

4.52 Data pre-processing

The data fed to a SOM includes all the information that a network gets. If erroneous data is fed to the SOM, the result is also erroneous or of bad quality. Thus, Self-Organizing Map, as well as the other neural network models, follows the "garbage in - garbage out" principle. This is the motivation for data pre-processing. Especially, when analyzing real-life data, preprocessing is of paramount importance. If we are interested in a certain aspect or the subset of the input data, we should naturally use only that portion of the data. Almost needless to say, certain analysis focuses on a totally different subset than the others. Errors in the data must be removed. If the data is downloaded from a database as a query, the result is likely to include erroneous data because of the lack of

database integrity. Erroneous data must be filtered using a priori knowledge of the problem domain. For example, in databases, missing values are usually presented as zeros. Zeros are typical errors due to the lack of database integrity. These kinds of errors show up in the probability density function presentation as peaks at zero possibly outside the normal range of the variable. In the case of uncertainty, these kinds of values can be replaced with “don't care” values. In training of a SOM, input vectors with missing values can be used. Another approach is to remove the vectors from the training set if they have missing values. This has the negative side effect of reducing the training set size.

It is common that the components of the input data are scaled to have unit variance. This can be achieved by dividing the components by the square roots of their corresponding variances. This assures that for each component, the difference between two samples contribute approximately an equal amount to the summed distance measure between an input sample and codebook vector. Considerations should also be made on the similarity of components. Similarity measure usually loses identity of component differences via a summation, or treats all components equally. The components must contribute approximately as much to the similarity measure. Otherwise, a component with large variance would shadow components with small variance and thus only the components with large variance would contribute to the distance measure used as a similarity measure.

4.53 Initialization

Kohonen presents three different types of network initializations: random, initial samples, and linear initialization.

4.53.1 Random initialization

Random initialization means simply that random values are assigned to codebook vectors. This is the case if nothing or little is known about the input data at the time of the initialization.

4.53.2 Initial samples Initialization

Initial samples of the input data set can be used for codebook vector initialization. This has the advantage that the points automatically lie in the same part of the input space with the data.

4.53.3 Linear initialization

The codebook vectors are initialized to lie in the same input space that is spanned by two eigenvectors corresponding to the largest Eigen values of the input data. This has the effect of stretching the SOM to the same orientation as the input data.

4.54 Training SOM

Training is an iterative process through time. It requires a lot of computational effort and thus is time-consuming. The training consists of drawing sample vectors from the input data set and projecting them onto the SOM. The teaching consists of choosing a winner unit by the means of a similarity measure and updating the values of codebook vectors in the neighborhood of the winner unit. This process is repeated a number of times. In one training step, one sample vector is drawn randomly from the input data set.

This vector is fed to all units in the network and a similarity measure is calculated between the input data sample and all the codebook vectors. The BMU is chosen to be the codebook vector with greatest similarity with the input sample.

Consider training colors (RGB) using SOM trainer. We can think of colors as 3D points, each color component represented by an axis. Each axis is given a scale of 6. If we have chosen green which is of the value (0, 6, 0), the color light green (3, 6, 3) will be closer to green than red at (6, 0, 0).

Light green = Sqrt((3-0)^2+(6-6)^2+(3-0)^2) = 4.24

Red = Sqrt((6-0)^2+(0-6)^2+(0-0)^2) =8.49

So light green is now the best matching unit, but this operation of calculating distances and comparing them is done over the entire map and the weight with the shortest distance to the sample vector is the winner and the BMU. The computational effort consists of finding a BMU among all the neurons and updating the codebook vectors in the neighborhood of the winner unit. If the neighborhood is large, there are a lot of codebook vectors to be updated. This is the case in the beginning of the training process, where it is recommended to use large neighborhoods. In the case of large networks, relatively larger portion of the time is spent looking for a winner neuron. All these considerations depend on the time spent on each of these phases depending on particular software and hardware used [44].

4.55 Emergent SOM (ESOM)

Emergent SOM (ESOM) is a self organizing projection from the high dimensional data space onto a grid of neuron locations. The grid of neurons is usually embedded in a two dimensional manifold. This space is called a map with a geographical interpretation

in mind. The learning algorithm of the SOM is designed to preserve the neighborhood relationships of the high dimensional space on the map [45]. Therefore the map can be regarded as a roadmap of the data space.

The often used two dimensional plane as map space has the disadvantage that neurons at the borders of the map have very different mapping qualities than neurons in the center of the map. The reason for this is the different number of neighbors of center neurons than those at borders. This is important during the learning phase since the border projections are hampered by nature of the neuron structure arrangements. In many applications important clusters appear in the corners of such a planar map and due to the limitations of the learning phase, the data may be lost. The embedding into a borderless manifold, such as a torus, avoids such effects. To visualize such toroid maps, four instances of the grid are tiled and displayed adjacently [46, 47].

The U-Matrix is constructed on top of the map. Let *n* be a neuron on the map, NN (n) be the set of immediate neighbors on the map, w (n) the weight vector associated with neuron *n*, then

$$\text{U-height (n)} = \sum_{\text{m}\in\text{NN(n)}} d(\mathbf{w}(n) - \mathbf{w}(\text{m})) \tag{4.9}$$

where d(x, y) is the distance used in the SOM algorithm to construct the map. The U-Matrix is a display of the U-heights on top of the grid positions of the neurons on the map. A U-Matrix is usually displayed as a grey level picture or as three dimensional landscape as illustrated in figure 4-9. The U-Matrix has become the standard tool for the display of the distance structures of the input data on ESOM. A U-Matrix displays the local distance structure of the data set and delivers a “landscape” of the distance relationships of the input data in the data space.

Properties of the U-Matrix are:

- The position of the projections of the input data points reflect the topology of the input space, this is inherited from the underlying SOM algorithm
- weight vectors of neurons with large U-heights are very distant from other vectors in the data space
- Weight vectors of neurons with small U-heights are surrounded by other vectors in the data space
- Projections of the input data points are typically found in depressions.
- Outliers in the input space are found in “funnels”.
- “mountain ranges” on a U-Matrix point to cluster boundaries
- “valleys” on a U-Matrix point to cluster centers

The U-Matrix realizes the emergence of structural features of the distances within the data space. Outliers, as well as possible cluster structures can be recognized for high dimensional data spaces. The proper setting and functioning of the SOM algorithm on the input data can also be visually checked. Using the ESOM/U-Matrix methods for clustering has the advantage of a nonlinear disentanglement of complex cluster structures.

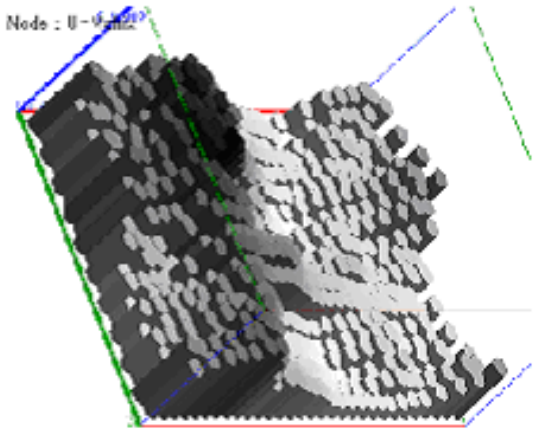


Figure 4- 9: U-Matrix representation of ESOM

4.6 Interpretation of SOM mapping

Figure 4-7 gave a U-Matrix representation of SOM where light grey sections of the map (zones) represented high density of input data. This would mean the zone had a high density of similar nodes. Dark grey zone had low density of similar nodes. Clusters can then be seen as light grey zones with dark grey boundaries.

Figure 4-8 showed a component map derived from SOM. It can be regarded as the input data belonging to a certain level of the class. Each class of input data is thus categorized into various zones. Taking an example of high blood pressure (HBP), there are those patients with high pressure measurements and others with low values. By having color coding for the zones decided for HBP, SOMs for each component can clearly be visualized. The way the component clusters relate to each other can vividly be monitored. Gray scaling is then added to indicate the strength of the bonds between clusters.

Figure 4-9 gave an ESOM representation of input data. Densely populated zones are the valleys while scarcely populated zones are the hilly areas. The steepness of the hilly zones indicate how unrelated the neighboring clusters are. The hilly zones become the boundaries of the clusters.

4.7 SOM types

4.71 Plane SOM

Plane SOM is a mode of projecting input dataset onto a 2D plane without considering the effect edge and corner nodes have on the overall representation of the

dataset. Consider the dataset of animals shown in figure 2-10 where 16 parameters (small, medium, large, 2legs, 4legs, hair, hooves, mane, feathers, hunt, run, fly, swim, nocturnal, herbivore, stripes) are used to distinguish them. Animals with parameter between 0 and 1 are given a normalized value of 0.5. Using SOM_PAK software developed by 'SOM Programming Team of the Helsinki University of Technology laboratory of Computer and Information Science FINLAND' to generate the relationships between the animals, U matrix SOM shown in figure 4-12 is realized where clustering the animals into various categories depending on the parameter relationships is evident.

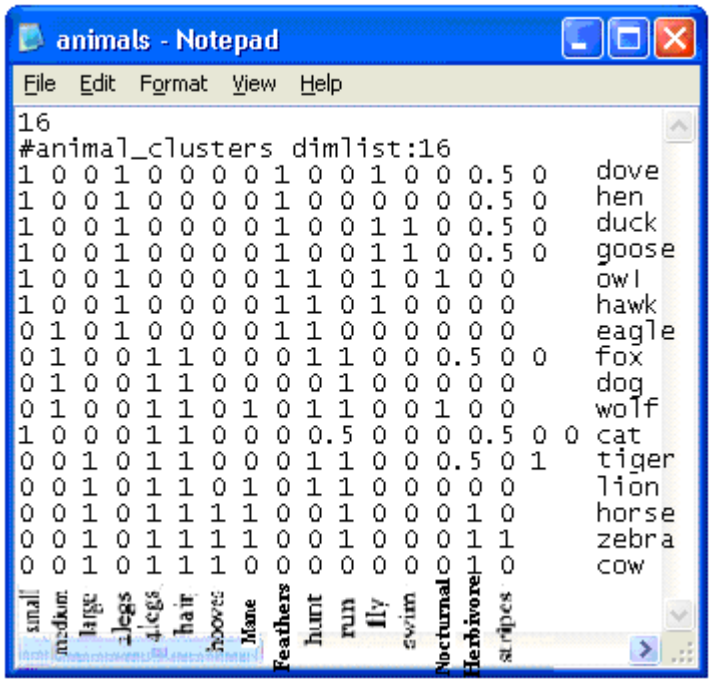


Figure 4- 10:Animal pre-processed dataset

4.71.1 SOM_PAK preview

The running of SOM_PAK software can be started by opening its executable file. The software prompts the user with a dialog window shown in figure 4-11. Before training

your dataset, SOM initialization is necessary. The following details give guidelines to the meaning of program parameters:

Mode of sampling the input data: Choose from random or linear.

-cin Name of the file from which the reference vectors are read (init.cod).

-din Name of the input data file (***.dat).

-cout Name of the file to which the reference vectors are stored (***.cod) after successive simulations. Type also the expected postscript (***.ps) file after simulations.

-rlen Running length (iteration count) in training.

-alpha Initial learning rate parameter. Decreases linearly to zero during training.

-radius Initial radius of the training area in SOM-algorithm. Decreases linearly to one during training.

-xdim Number of units in the x-direction.

-ydim Number of units in the y-direction.

-topol Topology type used in the map. Possible choices are hexagonal lattice (hexa) and rectangular (rect).

-neigh **N**eighborhood function type used. Possible choices are step (bubble) and Gaussian (Gaussian) functions.

-plane Component plane of the reference vectors that is displayed in the conversion routines.

-alpha_type Learning rate function type (in vsom and vfind). Possible choices are linear function (linear, the default) and inverse-time type function (inverse_t).

The linear function is defined as alpha (t) = alpha (0) (1.0 - t/rlen) and the inverse-time type function as alpha (t) = alpha (0) C/(C + t) to compute alpha (t) for an iteration step t.

In the package the constant C is defined C = rlen/100.0.

Select U-Matrix representation if required then press 'start [1]' button for simulation to run. If successful, a U-Matrix plane SOM output files are copied to the destination files.

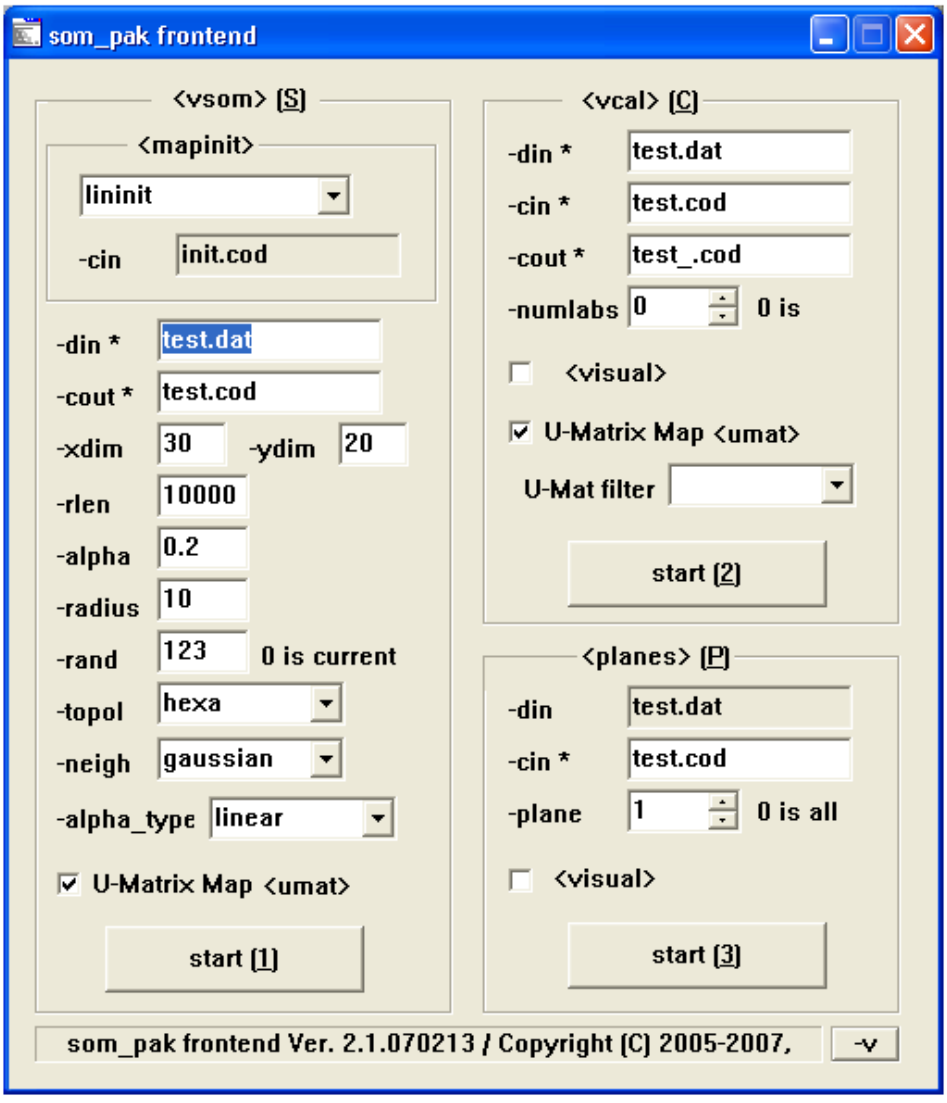


Figure 4- 11:SOM_PAK dialog window

A display of U-Matrix representation of animal data is shown in figure 4-12. If a different sample data is to be tested or calibrated using the trained output data (***.cod), then the new data need to be input under the <vcal> class. A new output file (****.cod) will be generated. You can use the new file for further analysis of the input data.

<planes> Component maps can be observed using class

In addition to above commands, it is always possible to give the -v n parameter (verbose parameter), which defines how much diagnostic output the program will generate. The value of n can range from 0 upwards, whereby greater values will generate more output; the default value is 1.

Figure 4-12 illustrates the similarities of different types of animals. Where the animals have more similar of the considered parameters, the lighter is the gray scaling. Animals with dissimilar parameters have dark gray areas between them. Thus by observing the intensity of the gray between animals, one is able to tell the similarities between them. The birds appear on the left section of the map **AD** while the other animals appear on the right **BC**. One important feature with plane SOM is that there are no cluster relations at the corners and edges of the map. If one considers a wider zone **AD,** consisting of large birds and the other zone **BC,** consisting of four legged, hair animals, the two zones have animals of diverse similarities and hence cannot be grouped together. This dissimilarity of clusters at the edges of the map becomes a drawback to its usage.

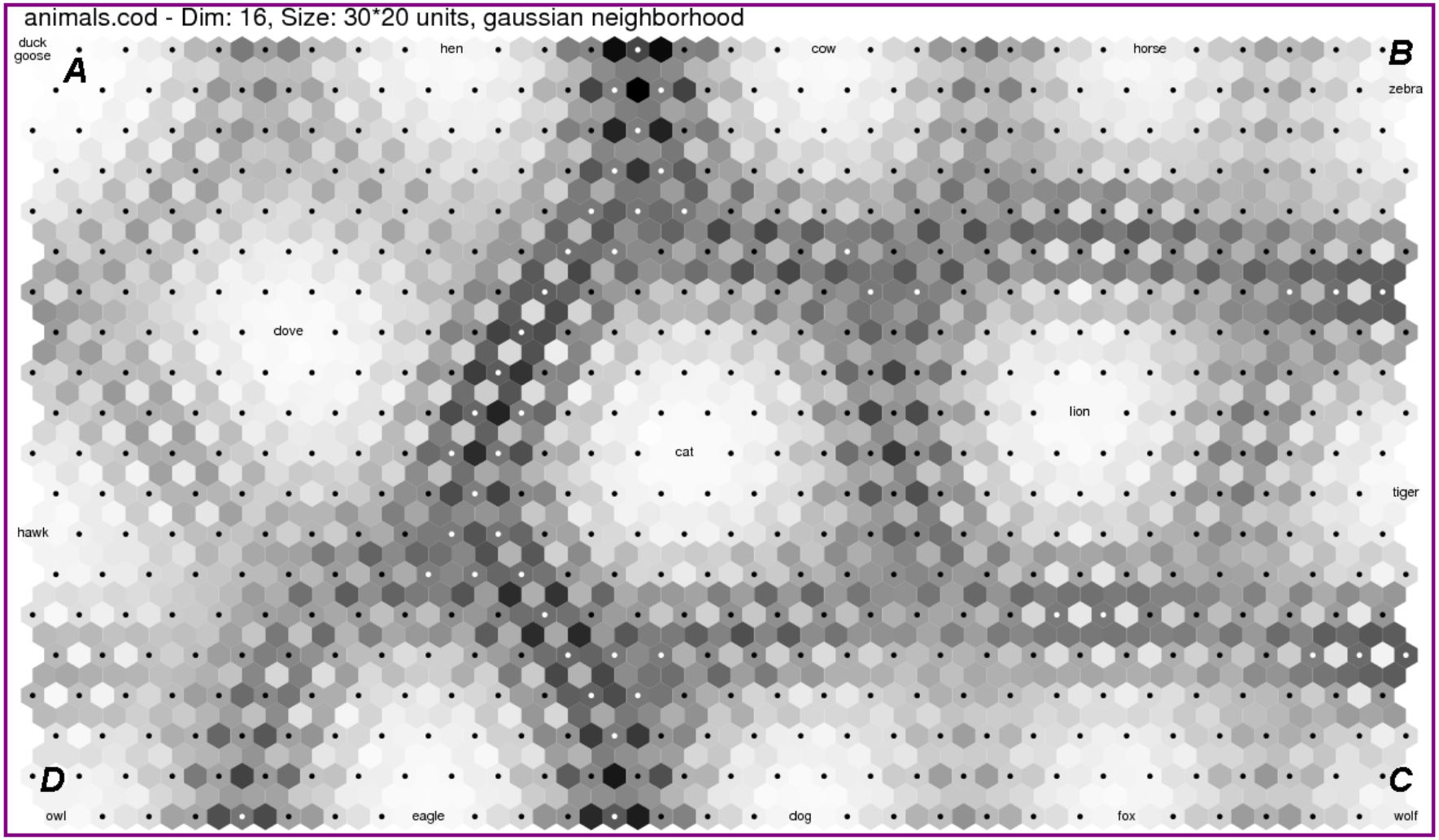


Figure 4- 12: Animal clustered SOM

4.72 Torus SOM

Overview of Mr. Torus SOM software

Torus SOM is evoked by double clicking the executable file icon. The software opens a dialog window as shown in figure 4-13. The following buttons when activated cause the software to perform various operations:

ALL The software opens another dialog window for the user to insert the input file (***.dat), code file (***.cod), the dimensions of the node lattice, learning rate, neighborhood, study count. Clicking the OK button causes the software to automatically train the map and generate the maps, and output files.

DATIN Reads the input file. If the file is not accessible, prompts an error of invalid path.

DATOUT Generates an output file from the input data.

DATDEL Delete data

DATEDIT Prepares a vectorial representation of input data using BMU

CODIN Used for inputting and displaying a map for an existing ***.cod file.

CODOUT Generates an output file as ***.cod.

NOCRE Creates nodes in a particular layer. **LER** causes the software to run for the node layer.

ADDLER Forbidden

LABEL The trained node lattice can be label by re-training the original trained SOM with the new ***.cod file with labels

Cod⇒Dat	Forbidden
EDIN	Reads input file
EDOUT	Outputs expected data
EDEST	Gives a detailed SOM map of the section marked X.
ONEEST	Component layout maps display

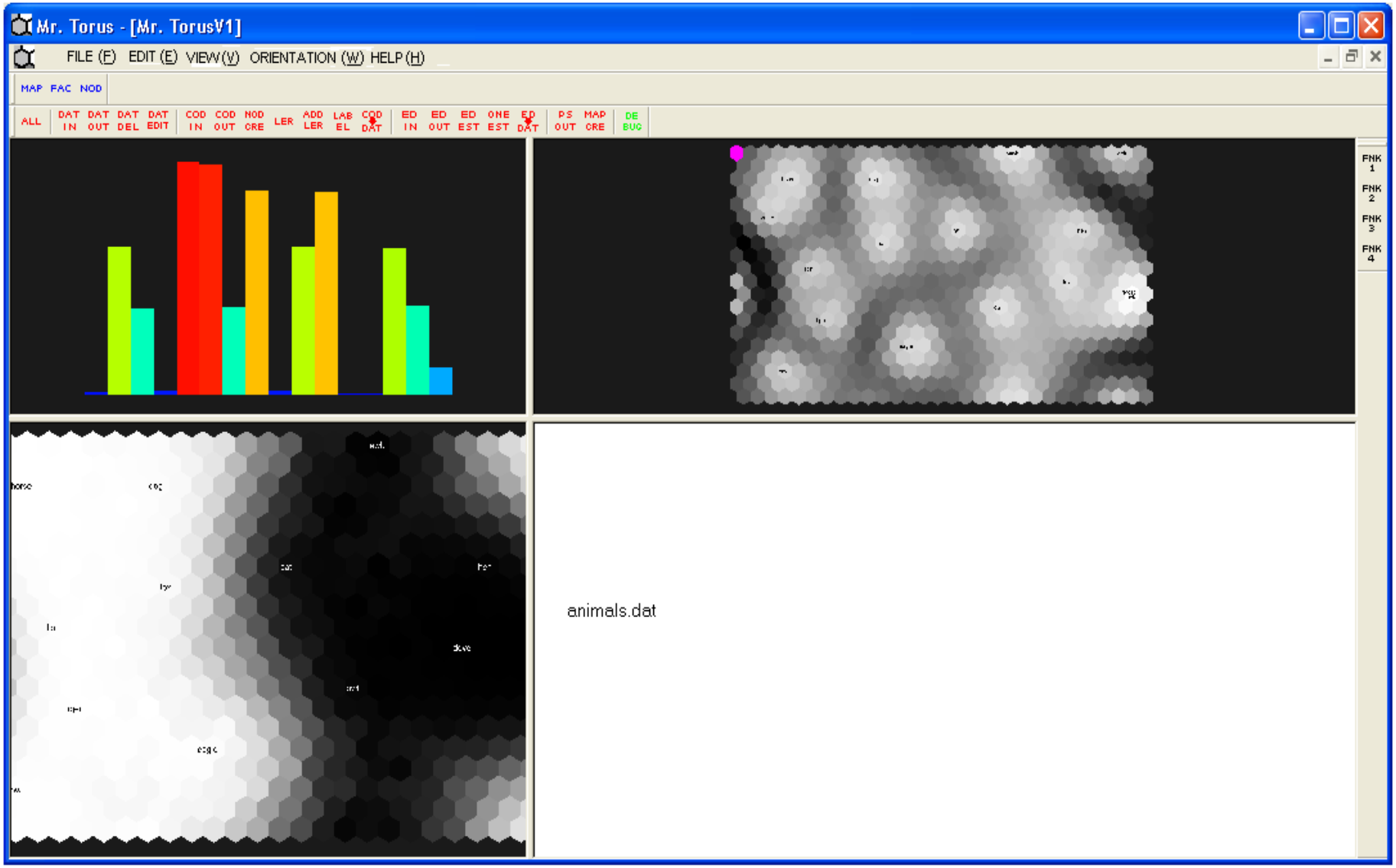


Figure 4- 13:Torus software dialog window

ED⇒DAT data taken to input layer and input vector shown. Able to run the trainer as it is.

PSOUT Postscript file generation

MAPCRE Calculate neighborhood distance and then make map. If this button is not pressed, the right labeling may not be given.

****Blue buttons (not shown) ****

MAP Display output layer

FAC Display element map

NOD Bar graph display of a particular node.

****Black Button****

FNK 1~4 Forbidden

****Shortcut keys****

Cursor Shift selected

Cntrl + cursor Map shift

F1 Changeover (color)

F2 **2D to 3D**

Mouse right click + move mouse + cntrl Shrinking or enlargement of map

Mouse right click + move mouse + shift Vector shrinking and enlargement.

****Only in 3D mode****

Mouse right click + move mouse Rotate map at origin

Figure 4-14 shows a U-matrix representation of the same animal dataset shown in figure 4-10 using Mr. Torus SOM simulation software. Dark grey areas represent less population of animals and can then be viewed as boundary relation status between the clusters. Lightest grey zones are densely populated regions of the clusters i.e. more animals.

Torus map gives an input data representation where each node on the map is adjacent to each other. This becomes evident particularly on the edge nodes and corner nodes. Zones emanating from corners **A, B, C** and **D** share the same cluster. Nodes at the edge **AB** are adjacent to nodes on edge **CD**. Similarly nodes on edge **BC** are adjacent to nodes **AD**. This mode of representation co-relates each node on the map to its neighbors without discontinuity. The analysis as well as the visualization of the complex input data projected onto the map becomes a better representation of the input than plane SOM. Cluster relations and their effects can be monitored to a higher degree of accuracy. It is equally important to note that the neighborliness and its effect seen on the map is the closest representation of the input data. The clear boundaries marked on the map indicate distinct behavioral patterns of the clusters. Mr. Torus IV simulator has ample features needed to simulate input data to Torus SOM.

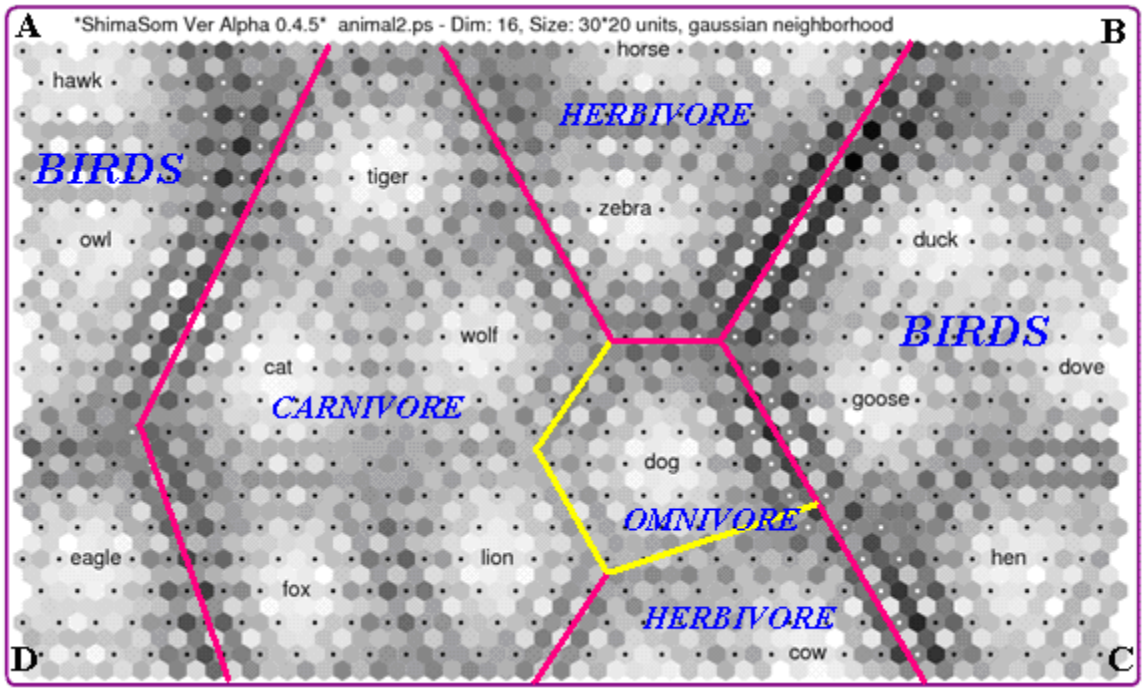


Figure 4- 14:U-Matrix animal Torus map

4.73 Spherical SOM

Spherical SOM (SSOM) is another mode of representing the input data, this time using a spherical map [48, 49]. Each zone on the map can clearly be related to its neighbors. The simulator "blossom" when evoked prompts a window shown in figure 4-15. When "blossom" is evoked the original sphere has untrained nodes. When pre-processed data is input to the trainer as a "***.dat" file and training icon pressed, training commences and its duration will depend on the volume of data, parameters (dataset) monitored, type of microprocessor and the clock frequency. The trainer prompts an icon (training finished) after training is complete. Figure 4-10 dataset is again used to obtain spherical SOM projections displayed in figure 4-16.

The user is able to perform the following operations:

- U-Matrix representation is one of the preferred maps for analytical purposes of the trained input data set. The user can rotate the map for better view of any location or area of interest on the sphere.
- Scaled colored map with low pre-processed data value zones colored dark blue till the highest values colored red.
- Component map gives the analyzer the contribution each component has on the overall appearance of the map.
- Construction of a dendrogram that helps the user to monitor the modularization of the input data set in a Top-Down format shown in figure 4-17.
- Other features of the simulator include saving developed files, reproducing map using code book vector table (***.scod) and matching data sets with labels.
- Glyptic display gives distortions of the sphere brought about by each input component. User is able to visualize the similarities of the clusters formed. The more a cluster has input data, the deeper is the valley and hence more distortion of the map in that zone. Figure 4-18 shows a distorted sphere with the depressions indicating the presence of a cluster and none depressed representing no cluster zones. One is the able to monitor the similarities of the animals by observing the deepness and gray color the cluster may be having.

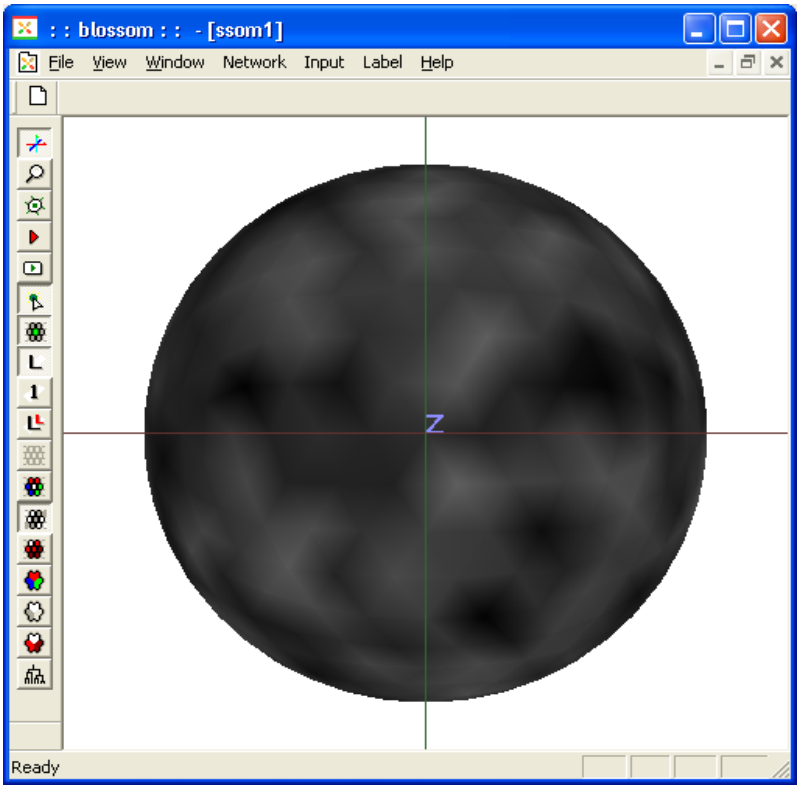


Figure 4- 15:"blossom": main window before input dataset training.

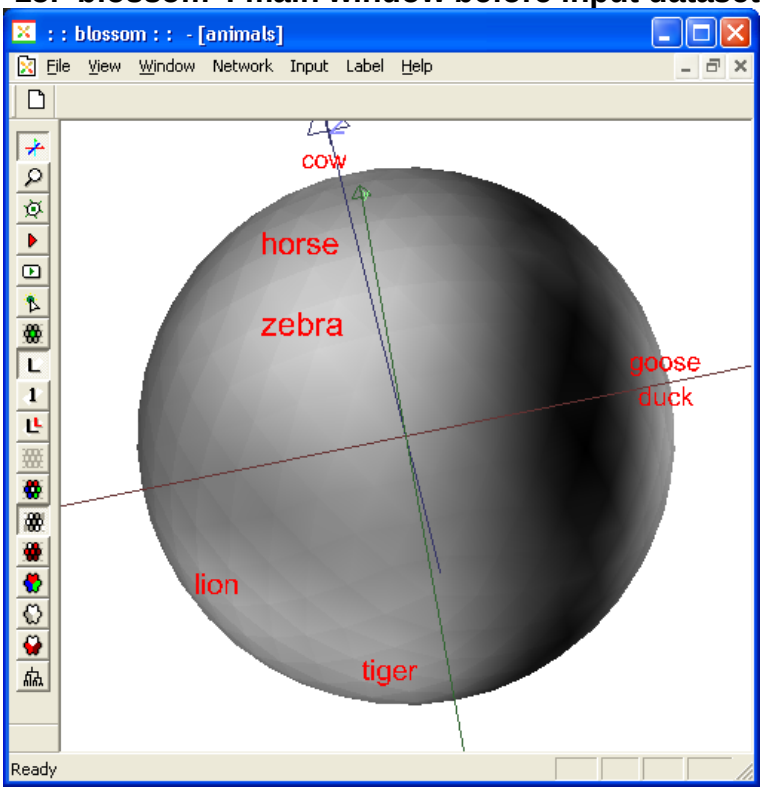


Figure 4- 16:U-Matrix animal clustering SSOM

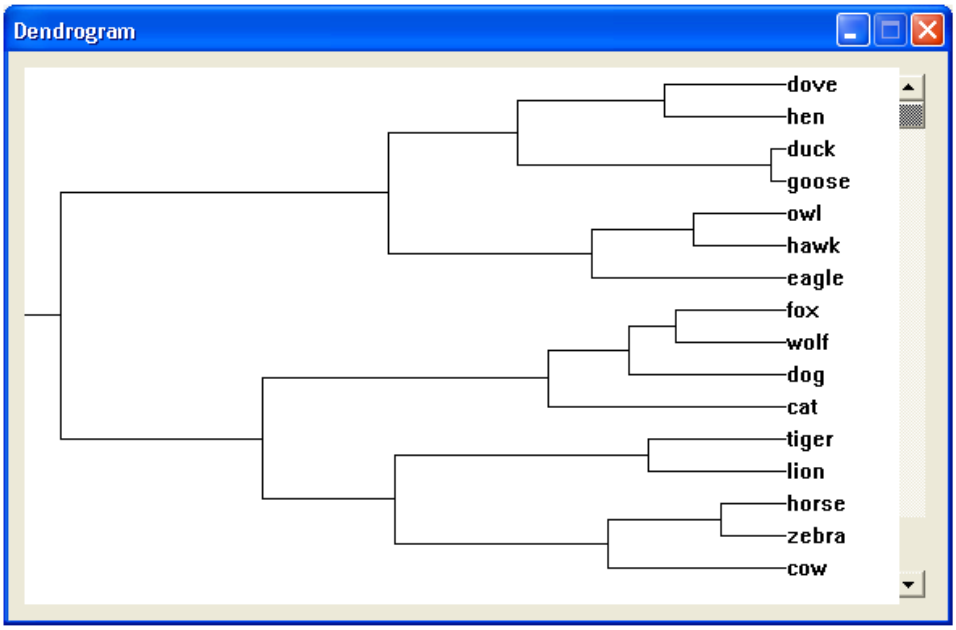


Figure 4- 17: Animal Dendrogram

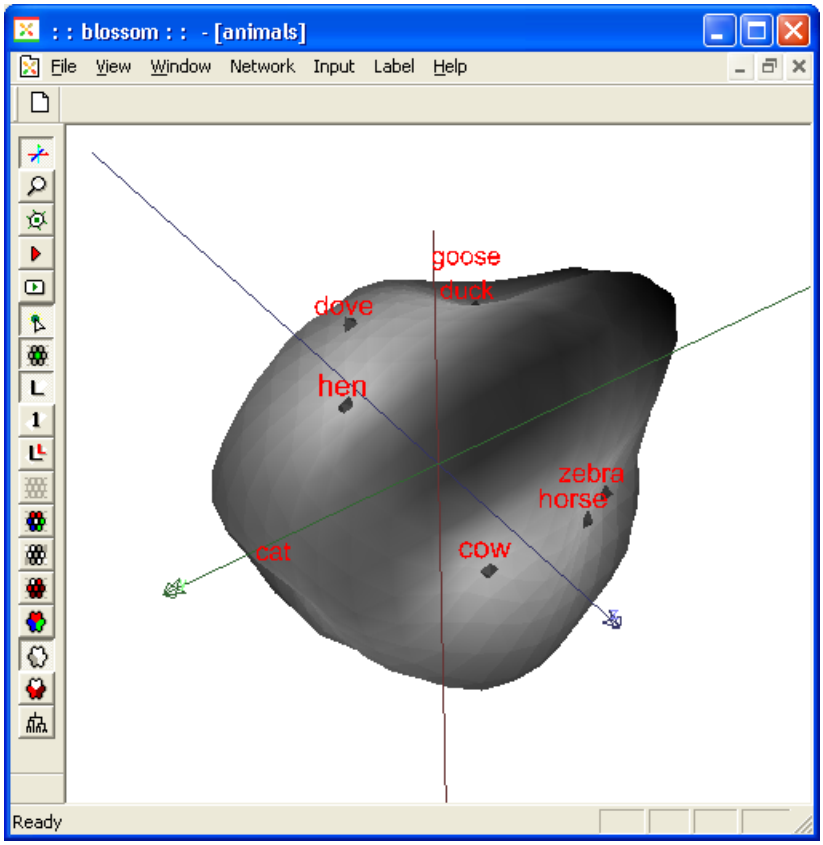


Figure 4- 18: Animal U-Matrix glyptic SOM

4.8 Chapter Summary

In this chapter, the concepts of SOM and the visualization approaches to the trained SOM are discussed. The procedures to simulate various forms of SOM outputs are presented. Focus is then drawn onto two forms of software developed during this research work namely Torus and Spherical SOM. Merits over plain SOM are expressed.

Metabolic syndrome analysis, visualization and trends monitoring tools will be developed based on the results obtained from the both Torus and Spherical SOM simulation outputs.

CHAPTER 5

ANN APPLICATIONS IN HUMAN HEALTH

5.1 Introduction

Human health is of prime importance to their everyday endeavors including the future projection of their ambitions. Health being a state of having a body and vigorous mind free from diseases contributes to all activities humans are able to perform. Mental health is one of the faculties of health and it so happens that it has a bearing to the physical health of the person.

Artificial Neural Networks (ANN) is currently an active research area in medicine and it is believed that it will receive extensive application to biomedical systems in the next few years. At the moment, the research is mostly on modeling parts of the human body and recognizing diseases from various scans (e.g. cardiograms, CAT scans, ultrasonic scans, etc.). Neural networks are ideal in recognizing diseases using scans since there is no need to provide a specific algorithm on how to identify the disease. Neural networks learn by example so the details of how to recognize the disease are not needed. What is needed is a set of examples that are representative of all the variations of the disease. The quality of examples is not as important as the 'quantity'. The examples need to be selected very carefully if the system is to perform reliably and efficiently.

SOM can be used to visualize and analyze the health behavior patterns of an individual. Clinical doctors can then use the charts to help the examinees visualize their

degree of health. The doctors can as well predict the consequences well in advance and hence save the member from being adversely affected.

5. 11 Modeling and Diagnosing the Cardiovascular System

Neural Networks are used experimentally to model the human cardiovascular system. Diagnosis can be achieved by building a model of the cardiovascular system of an individual and comparing it with the real time physiological measurements taken from the patient. If this routine is carried out regularly, potential harmful medical conditions can be detected at an early stage and thus make the process of combating the disease much easier.

A model of an individual's cardiovascular system must mimic the relationship among physiological variables (i.e., heart rate, systolic and diastolic blood pressures, and breathing rate) at different physical activity levels. If a model is adapted to an individual, then it becomes a model of the physical condition of that individual. The simulator will have to be able to adapt to the features of any individual without the supervision of an expert. This calls for a neural network [50].

5.12 Sensor fusion

ANN technology has also found other applications in sensor fusion where different sensor qualities and characteristics are combined to form a multi-sensing element. Sensor fusion enables the ANNs to learn complex relationships among the individual sensor values, which would otherwise be lost if the values were individually analyzed. In medical modeling and diagnosis, this implies that even though each sensor in a set may be sensitive only to a specific physiological variable, ANNs are capable of detecting

complex medical conditions by fusing the data from the individual biomedical sensors [51].

5.13 Electronic noses

ANNs are used experimentally to implement electronic noses. Electronic noses have several potential applications in telemedicine. Telemedicine is the practice of medicine over long distances via a communication link. The electronic nose would identify odors in the remote surgical environment [52]. These identified odors would then be electronically transmitted to another site where an odor generation system would recreate them. Because the sense of smell can be an important sense to the surgeon, telesmell would enhance present surgery.

5.14 Metabolic Syndrome

Metabolic syndrome is a symptom of body disorder (medical) that causes the various organs to malfunction. The malfunctioning organs can lead to diseases like cardiovascular or diabetes. The syndrome is believed to be associated with eating and little physical exercises habits, where the body weight to height factor is taken as a scalar quantity.

5.14.1 Metabolic syndrome diagnosis

There are no well-accepted criteria for diagnosing the metabolic syndrome. The criteria proposed by the National Cholesterol Education Program (NCEP) Adult

Treatment Panel III (ATP III), with minor modifications, are currently recommended and widely used.

The American Heart Association and the National Heart, Lung, and Blood Institute recommend that the metabolic syndrome be identified as the presence of three or more of these components:

- Elevated waist circumference
- Elevated triglycerides
- Reduced HDL ("good") cholesterol
- Elevated blood pressure
- Elevated fasting glucose

Other conditions associated with the syndrome include physical inactivity, aging, hormonal imbalance and genetic predisposition.

Some people are genetically predisposed to insulin resistance. Acquired factors, such as excess body fat and physical inactivity, can elicit insulin resistance and the metabolic syndrome in these people. Most people with insulin resistance have abdominal obesity. The biologic mechanisms at the molecular level between insulin resistance and metabolic risk factors aren't fully understood and appear to be complex [53].

The concept that big belly members of the society endanger themselves in relation to this syndrome cannot be overemphasized. Unfortunately many end up digging their grave through psychological implications and not due to their body weight height factor.

Analysis of this syndrome through SOM symbolizes the importance of getting more information of any patient and including him or her in the decision making process. In

effect the patient becomes a health conduit to solving any health issue and thereby preventing patients being affected by the secondary disease “psychology”.

5.14.2 Metabolic syndrome evaluation system parameters

A metabolic syndrome evaluation system using SOM was constructed based on health checkup data from examinees. Initially four parameters were used namely BMI, HBP, GLU and TG. Two additional parameters as syndrome evaluator were later considered namely LBP and HDL

5.2 Parameters effect on health (brief preview)

5.21 Body Mass Index

BMI is a statistical measurement which compares a person's weight and height. Though it does not actually measure the percentage of body fat, it is a useful tool to estimate a healthy body weight based on how tall a person is. Due to its ease of measurement and calculation, it is the most widely used diagnostic tool to identify obesity problems within a population. However it is not considered appropriate to use as a final indication for diagnosing individuals. It was invented between 1830 and 1850 by the Belgian polymath Adolphe Quetelet during the course of developing "social physics" [54]. Body mass index is defined as the individual's body weight divided by the square of his height. The formulas universally used in medicine produce a unit of measure of kg/m^2

Western societies have in the past used BMI as a simple numeric measure of a person's "fatness" or "thinness", allowing health professionals to discuss over- and under-weight problems more objectively with their patients. However, BMI has become controversial because many people, including physicians, have come to rely on its apparent numerical authority for medical diagnosis, but that was never the BMI's purpose. It is meant to be used as a simple means of classifying sedentary (physically inactive) individuals with an average body composition. For these individuals, the current value settings are as follows: a BMI of 18.5 to 25 may indicate optimal weight; a BMI lower than 18.5 suggests the person is underweight while a number above 25 may indicate the person is overweight; a BMI below 17.5 may indicate the person has anorexia or a related disorder; a number above 30 suggests the person is obese (over 40, morbidly obese).

$$BMI = \frac{Weight(Kg)}{Height^2(m^2)}$$

BMI may vary from time to time and country to country, making global, longitudinal surveys problematic. In 1998, the U.S. National Institutes of Health brought U.S. definitions into line with World Health Organization guidelines, lowering the normal/overweight cut-off from BMI 27.8 to BMI 25. This had the effect of redefining approximately 30 million Americans, previously "technically healthy" to "technically overweight". It also recommends lowering the normal/overweight threshold for South East Asian body types to around BMI 23, and expects further revisions to emerge from clinical studies of different body types.

One basic problem, especially in athletes, is that muscle is denser than fat. Some professional athletes are "overweight" or "obese" according to their BMI - unless the

number at which they are considered "overweight" or "obese" is adjusted upward in some modified version of the calculation [55]. In children and the elderly, differences in bone density and, thus, in the proportion of bone to total weight can mean the number at which these people are considered underweight should be adjusted downward.

5.22 Blood Pressure

Blood pressure is the force of the blood against the artery walls. High blood pressure (hypertension) and low blood pressure (hypotension) can both cause cardiovascular problems.

5.22.1 High blood pressure (HBP)

HBP is a condition in which a person's blood pressure is elevated. Blood pressure is the measure of the force of the blood pushing against the walls of the arteries – the blood vessels that carry blood from the heart to the rest of the body. If the high blood pressure has no known cause (more than 90 percent of cases), it is known as primary, essential or idiopathic hypertension. If it is caused by another condition, such as kidney disease, it is known as secondary hypertension. However, because of the complex variety of systems that influences blood pressure, these distinctions have blurred somewhat in clinical practice.

Blood pressure measurement is inexpensive and easily performed. Blood pressure is measured in two phases that correspond to the natural contractions of the heart. When the heart contracts (e.g., systole), the pressure of blood against arterial walls is known as

systolic pressure. When it relaxes (diastole), the pressure of blood against arterial walls is known as diastolic pressure.

Blood pressure is always expressed as systolic pressure over diastolic pressure. Normal blood pressure for adults is considered to be below 120/80 millimeters of mercury (mm/Hg). Generally, blood pressure above 140/90 is considered to be high for adults, and blood pressure under 90/60 is considered to be low for adults (hypotension). HBP may be diagnosed if an individual has any of the following three conditions:

- Has a blood pressure reading of 140/90
- Is taking antihypertensive medication
- Has been found twice by a physician to have high blood pressure

Researchers have also identified dozens of genes that contribute to high blood pressure. Though this implies that some people inherit a propensity for high blood pressure, the association is more complicated. Researchers believe that about 30 percent of essential hypertension can be traced back to genetic abnormalities that run in families. Most recent studies seem to indicate that inherited high blood pressure is the result of multiple gene expressions. There is currently no genetic test that consistently identifies people at risk for developing high blood pressure.

Whatever its cause, high blood pressure exacts a tremendous cost from society. High blood pressure is a major risk factor for heart attack, stroke and heart failure

Hypertensive patients are at increased risk of:

- Heart disease (e.g., heart failure, sudden cardiac death, cardiomyopathy) and arrhythmias.
- Stroke

- Accelerated coronary artery disease
- Aortic aneurysm (a weakness in the aortic wall where it balloons out to more than 1.5 times its normal size and is in danger of rupturing), often resulting in sudden cardiac death
- Kidney failure
- Retinopathy (eye disease that leads to loss of vision)

The risk of developing one or more of these serious health conditions increases as blood pressure rises. Although the cause of most cases of high blood pressure is unknown, researchers have uncovered evidence that blood pressure is associated with insulin resistance and/or elevated insulin levels. Both high blood pressure and insulin resistance are features of the metabolic syndrome, a cluster of abnormalities that includes obesity, elevated triglycerides and low HDL "good" cholesterol. A number of possible mechanisms have been proposed that would explain how insulin resistance contributes to hypertension. However, this link is still poorly understood [56].

5.22.2 Low Blood Pressure (LBP)

LBP occurs when the pressure on the blood vessel walls falls below normal limits. Low blood pressure that does not cause symptoms is generally considered to be a sign of good cardiovascular health because there is less stress on the heart, lungs and blood vessels. Low blood pressure can be a sign of good health in some people with no symptoms (e.g., athletes). However, there are a number of forms of low blood pressure that require diagnosis, evaluation and treatment. People may seek treatment for low blood pressure if they experience symptoms such as dizziness or syncope (fainting) from lack

of oxygen to the brain. Low blood pressure may be due to medications (e.g., blood pressure medications) or other causes, and changing medications or other treatments may be necessary.

Hypotension is the medical term for low blood pressure, which is considered to be under 90/60. There are a number of forms of low blood pressure that require medical diagnosis and treatment. The two most common are orthostatic hypotension and neurally mediated hypotension (NMH).

Signs and symptoms of low blood pressure:

- Dizziness or lightheadedness
- Blurry vision
- Lack of concentration
- Nausea or upset stomach
- Muscle weakness
- Fainting (syncope)

5.23 Blood sugar (blood glucose)

The body's main source of energy is a simple sugar called glucose. People get glucose by eating carbohydrates in foods. It is also manufactured in the body through the breakdown (metabolism) of protein and fats. After a meal the amount of glucose in the bloodstream rises. The pancreas responds by secreting a hormone called insulin which helps glucose be absorbed into the cells for use. Lack of insulin causes glucose to build up in the bloodstream, resulting in diabetes.

Once inside the cells, glucose – which is made up of carbon, hydrogen and oxygen – is broken down and converted into adenosine tri-phosphate, better known as ATP. This is the body's main fuel. Excess glucose in the bloodstream usually ends up in the liver, where it is stored as glycogen. When blood sugar is low, a hormone called glucagon stimulates the liver to convert glycogen back into glucose.

Consistently high blood glucose levels create a dangerous condition that can cause many associated complications, including:

- Kidney disease (diabetic nephropathy)
- Eye diseases (retinopathy, glaucoma, cataracts)
- Nerve disease (neuropathy)
- Heart conditions and stroke
- Poor circulation
- Foot problems and skin problems
- Gastroparesis (delayed stomach emptying)
- Sexual dysfunction
- Yeast infections and other infections
- Gum disease and tooth decay

It is important that people with diabetes maintain proper levels of glucose (blood sugar) to ensure optimal health [57]. For most people, unless otherwise instructed by their physician, blood glucose should fall into the following target ranges, with measurements given in milligrams per deciliter (mg/dL):

Fasting (upon waking): 70 mg/dL to 110 mg/dL

After meals: 70 mg/dL to 140 mg/dL

5.23.1 Hyperglycemia and hypoglycemia

Hyperglycemia: Abnormally high blood glucose occurs when the body has too little insulin or when the body cannot use insulin properly. All diabetic individuals occasionally have high blood glucose, but serious complications can develop when readings are unusually high or frequently high. Left untreated, hyperglycemia can lead to a coma.

Hypoglycemia: Abnormally low blood glucose occurs when a person's levels of glucose and insulin are unbalanced. People with hypoglycemia unawareness (difficulty sensing low glucose) are especially vulnerable. Mild cases of hypoglycemia can cause dizziness or weakness. Severe cases can lead to fainting, convulsions, brain damage or coma.

Medications and lifestyle changes such as nutritious diet, exercise and weight loss help those with diabetes to use insulin more effectively, thus controlling their glucose. This can help patients to live longer and healthier lives.

5.23.2 Insulin

Insulin is a hormone produced by the pancreas that helps move glucose into the cells. Diabetes occurs when the body cannot make or respond to insulin and glucose builds up in the blood. People with type-1 diabetes and some people with type-2 diabetes administer insulin to themselves daily. Methods of administration include syringe injections, insulin pumps and insulin pens.

5.23.3 Type-1 Diabetes

Type 1 diabetes occurs when the pancreas cannot produce insulin. Type 1 diabetes was once called juvenile diabetes because it is usually diagnosed in childhood. People with type 1 diabetes must supply insulin by injection, pump or other methods. Possible treatments include transplant of a pancreas or beta cells.

5.23.4 Type-2 Diabetes

Type 2 diabetes is the most common type of diabetes. It occurs when glucose builds up in the blood due to the body's inability to use insulin effectively. Type 2 diabetes was once called adult onset diabetes because it is usually diagnosed in adulthood. The disease may be prevented or controlled through diet and exercise, but some patients need insulin or other medications.

5.24 Cholesterol

Cholesterol is a waxy fat (lipid) that is present in all human beings. About 80 percent of the cholesterol in the body is manufactured by the liver. The rest is consumed through cholesterol-rich foods such as meat, eggs and dairy products. Cholesterol itself is vital for survival. However, it can also contribute to coronary artery disease. To understand how cholesterol is related to heart disease, it is necessary to understand how it is transported through the body. Cholesterol is carried in the bloodstream in specialized protein packages called lipoproteins. These are comprised of another building block called apolipoproteins [58].

A good analogy is to think of lipoproteins like vehicles on the road, while cholesterol represents the passengers. Some of the cars are sleek and fast, while others are cumbersome, large and slow. The nature of the lipoprotein package, or vehicles, ultimately determines what will happen to the cholesterol it carries. In some cases, excess cholesterol will be transported to the liver, where it is metabolized harmlessly. In other cases, excess cholesterol will penetrate the walls of arteries throughout the body, contributing to a disease called atherosclerosis.

Although there are many subclasses of lipoproteins, researchers generally focus on the following five types:

- High-density lipoproteins (HDL). “Good” cholesterol, HDLs move easily through the blood and are actually beneficial. They are stable and do not stick to artery walls. They help prevent heart disease by carrying cholesterol away from the arteries and back to the liver, where the process of its removal from the body begins. Liver damage, from alcohol abuse or other conditions, can undo the beneficial effects of HDLs.
- Low-density lipoproteins (LDL). “Bad” cholesterol, LDLs contain more fat and less protein than HDLs. LDLs are unstable and tend to fall apart. They are more likely to adhere to the walls of the artery and penetrate the protective inner lining of cells. Once cholesterol has migrated into the inner wall of the artery, it oxidizes and attracts other fatty substances (e.g., triglycerides), sticky blood-clotting materials (e.g., fibrin and platelets) and white blood cells. Together, these substances form the building materials for plaque deposits, which are the hallmark of "hardened arteries".

- Very low-density lipoproteins (VLDL). These are extremely harmful lipoproteins that carry triglycerides and cholesterol.
- Intermediate-density lipoproteins. Like VLDLs, these also carry triglycerides and cholesterol.
- Chylomicrons. These are very large particles that are rich in triglycerides

Cholesterol is vital to good health. The body uses cholesterol to:

- Assist in the manufacture of hormones
- Break down carbohydrates and proteins
- Help form a protective coating around nerves
- Build cell walls and produce bile (the word cholesterol is Greek for “bile solids”)

Cholesterol is a blood fat needed by the body in moderate amounts. However, high cholesterol levels can lead to atherosclerosis and coronary artery disease (CAD) and heart attack. Methods for increasing good cholesterol or lowering bad cholesterol levels are available as prescribed by a physician.

5.25 Triglycerides (TGs)

Triglycerides are a type of fat found in the blood. When one eats, the body converts any calories it doesn't need to use right away into triglycerides. TGs are major components of very low density lipoprotein (VLDL) and play an important role in metabolism as energy sources and transporters of dietary fat. In the intestine, triglycerides are split into glycerol and fatty acids (with the help of lipases and bile secretions), which are then moved into the cells lining the intestines. When the body requires fatty acids as an energy source, the hormone glucagon signals the breakdown of

the triglycerides by hormone-sensitive lipase to release free fatty acids. As the brain cannot utilize fatty acids as an energy source, the glycerol component of triglycerides can be converted into glucose, for brain fuel.

Triglycerides cannot pass through cell membranes freely. Special enzymes on the walls of blood vessels called lipoprotein lipases must break down triglycerides into fatty acids and glycerol. Fatty acids can then be taken up by cells via the fatty acid transporter (FAT).

Triglycerides are a form of fat in the bloodstream. People with high triglycerides often have high total cholesterol, high LDL (bad) cholesterol and low HDL (good) cholesterol level. Many people with heart disease also have high triglyceride levels. Several clinical studies have shown that people with above-normal triglyceride levels (greater than or equal to 200 mg/dL) have an increased risk of heart disease. People with diabetes or who are obese are also likely to have high triglycerides. Table 5-1 shows the standard values of triglycerides.

Table 5-1: Standard values of triglycerides

Triglyceride Level Classification	**Comments / Interpretation**
Less than 150 mg/dL	Normal
150-199 mg/dL	Borderline on the higher side
200-499 mg/dL	High
500 mg/dL	or higher ⇒Very high

5.25.1Ttriglycerides and Cholesterol

High triglycerides in the blood are often seen in overweight people. But even people who are not overweight may have stores of fat in their arteries as a result of insulin resistance. These triglycerides in the blood are the direct result of carbohydrates from the diet being converted by insulin. These triglycerides do not come directly from dietary fats. They are made in the liver from any excess sugars which have not been used for energy. Elevated triglycerides are one of the easiest problems to correct with appropriate diet; simple restriction of all sugars and grains.

If one just listens to the 'experts' change to another world on cholesterol issues, one would think that it is an evil substance and that most of us would benefit from lowering our cholesterol as low as possible. But it's not so. Cholesterol is a vitally important substance which is used for building our cell membranes and producing several of our hormones. If our cholesterol level drops too low, we are actually at increased risk for depression [59].

5.3 Metabolic Syndrome Evaluation using SOM

5.31 Physical Examination Data

Patients need physical examination particularly if they feel insecure or fall sick. For the sake of the metabolic syndrome, patients require medical and physical examinations for the physicians to carry out the analysis of the syndrome. In this work, employees of a certain company were done the examinations. To perform the syndrome analysis the following health parameter were first set as standards.

1) BMI: Over 25 Kg / m^2

2) HBP: Over 140mmHG

3) GLU: Over 110mg/dl

4) TG: Over 150mg/dl

5.32 Data Pre-Processing

Before the multi-dimensional data is applied to SOM, normalization of the data is done. This is due to the fact that the incoming data has different dimensions. Normalizing it causes the data to be taken as emerging from one source. For the normalization of the physical test data, let us take the minimum value of any input parameter as L, maximum value as H, actual data as X, and normalized value as Y.

Thus, if $(X < L)$;

$$Y = X/L \tag{5.1}$$

If $(L \le X \le H)$;

$$Y = 1 \tag{5.2}$$

If $(X > H)$;

$$Y = X/H \tag{5.3}$$

However after normalizing the data, some parameters seem to have high-normalized values causing the frequency distribution curve [60] for the normalized data to be more on the higher side. In some situations majority of the input data seems to be higher than

the normal. It is due to this abnormality that a ceiling value is decided for each input parameter. Any normalized value greater than the ceiling value is given the ceiling value.

Metabolic stage happens to be physical test data beyond the normal values. Taking the normal values to fall within those represented by equation (5.2), the four items BMI, HBP, GLU and TG, are re-normalized so that their values fall within the standard way of normalizing data. Thus Data (Y-1) is re-normalized again to a new Y. In this contribution equation (5.1) type of data is temporarily omitted.

Frequency distribution for every element was produced to obtain the ceiling value of the new Y of Equation (5.3). As shown in [60], ceiling values for all the elements were decided. Female ceiling scale values for BMI, HBP, GLU and TG were 1.25, 1.29, 1.55 and 2.3 respectively, while the males' were 1.2, 1.26, 1.6 and 2.9 respectively. Data obtained from various examinees is normalized as shown in table 5-2. Taking M1 as an example of a male examinee, normalized value Y will be in per unit basis by taking the Norm value have the denominator as the maximum difference (Y-1) of within the population being examined. All the parameters used in this analysis are normalized so that during input data training, the data will seem as coming from the same source. Table 5-3 shows samples of the normalized values of the examined members. It is worth noting that the scaling parameters for both male and female stand different due to population densities giving higher normalized values. After normalizing the data, the next stage is training the data using SOM trainers.

Table 5-2 Sample calculation of parameter normalization

	BMI	BMI	BMI	HBP
Male data Samples				
	Kg/(m)^2		Norm	mmHg
Upper limit	24.2			139
Lower limit	13			80
Ceiling scale	1.2			1.26
M1	27.4383	1.134	0.669	144
M2	27.8452	1.151	0.753	190
M3	23.5014	1.000	0	140
M4	21.5943	1.000	0	126
M5	23.7963	1.000	0	140

$$M1: Y = \frac{27.4393}{24.2} = 1.134 \quad Norm = \frac{1.134 - 1}{0.2005} = 0.668$$

Table 5-3: Male-Female normalized samples (Norm)

	BMI	BMI	BMI	HBP	HBP	HBP	GLU	GLU	GLU	TG	TG	TG	LBP	LBP	LBP	HDL	HDL	HDL
Male data Samples																		
	Kg/(m)^2		Norm	mmHg		Norm	mg/dl		Norm	mg/dl		Norm	mg/dl		Norm	mg/dl		Norm
Upper limit	24.2			139			109			150			89	89		200		
Lower limit	13			80			50			20			0	0		40		
Ceiling scale	1.2			1.26			1.6			2.9			1.14			1.25		
M1	27.4383	1.134	0.669	144	1.036	0.138	94	1	0	825	2	1	88	1	0	36	0.9	0.44444
M2	27.8452	1.151	0.753	190	1.26	1	317	1.6	1	777	2	1	100	1.124	0.8828	33	0.825	0.84848
M3	23.5014	1.000	0	140	1.007	0.028	142	1.3	0.505	773	2	1	100	1.124	0.8828	35	0.875	0.57143
M4	21.5943	1.000	0	126	1	0	113	1.04	0.061	769	2	1	84	1	0	49	1	0
M5	23.7963	1.000	0	140	1.007	0.028	98	1	0	673	2	1	104	1.14	1	36	0.9	0.44444
Female data Samples																		
	Kg/(m)^2		Norm		mmHg	Norm	mg/dl		Norm	mg/dl		Norm	mg/dl		Norm	mg/dl		Norm
Upper limit	24.2			139			109			150			89			200		
Lower limit	13.0			80			50			20			0			40		
Ceiling scale	1.25			1.29			1.55			2.3			1.12			1.14		
F1	24.7273	1.022	0.087	156	1.122	0.422	91	1	0	115	1	0	88	1	0	69	1	0
F2	24.7624	1.023	0.093	146	1.05	0.174	111	1.02	0.033	64	1	0	80	1	0	73	1	0
F3	24.1094	1	0	128	1	0	85	1	0	102	1	0	66	1	0	57	1	0
F4	19.2005	1	0	134	1	0	93	1	0	119	1	0	80	1	0	62	1	0
F5	19.6311	1	0	148	1.065	0.223	89	1	0	99	1	0	82	1	0	70	1	0

5.33 SOM Simulations

Originally there were 4007 female and 2450 male examinees. To remove healthy members from syndrome list, a re-normalization procedure was to be carried out. The re-normalization process takes two folds; No_cut and B0s02cut formats; No_cut metabolic syndrome members were those with any trace of increase from 0 to 1 (normalized values) in any of the four elements. This gave 2910 and 1764 respectively. The training of such data would both take time and not as accurate as that with lower population of examinees.

The B0s02cut case considered examinees with summation of three of the normalized values exempting BMI being less or equal to 0.2. This action reduced the examinees further to 2564 and 1375 respectively. To represent the whole spectrum of examinees, 20 healthy members were included in each input data.

5.34 Metabolic Syndrome Points

Weight factor for all the elements was taken as 1. Normally the weight of each element is based on the importance of the element to the health state.

Health mark point can be expressed by Equation (5.4), where a mark point is rounded off to the nearest 10-point units.

In equation (5.4), WV_n is the worst value of test data for particular parameter, NV normal value, $X_{ni,}$ the data of examinee, *n* the number of parameters being examined while 'i' is the count of metabolic examinee.

$$Health\ Mark\ point s(MK) = \frac{\sqrt{\sum_{j=1}^{n}\left(WV_j - NV\right)^2} - \sqrt{\sum_{j=1}^{n}\left(X_{ji} - NV\right)^2}}{\sqrt{\sum_{j=1}^{n}\left(WV_j - NV\right)^2}} \times 100$$

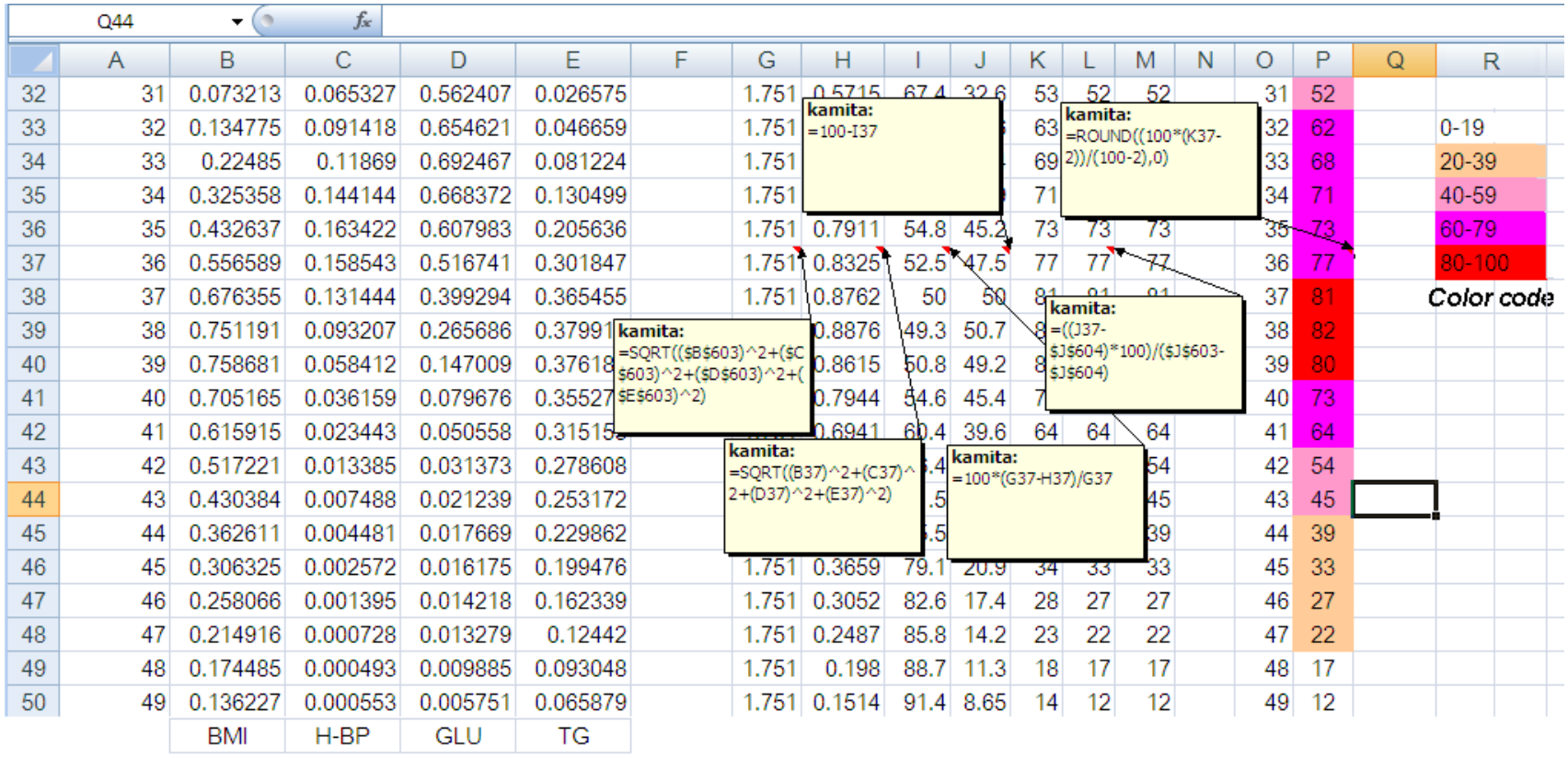
(5.4)

Metabolic Syndrome Points (MSP) for Torus SOM are (100-MK).

Table 5.4 shows a sample preview MSP data processing of equation (5.4) using excel software. The raw data (BMI, HBP, GLU and TG) is the simulation output data from Torus SOM trainer. Column P gives metabolic syndrome risk in percentages with the coding of the percentages as shown by the color code. Using the color code shown in the table, least MSP (clear or white) indicate healthy examinees while high MSP (red) have metabolic syndrome risk. The sample is taken from the trained six hundred (600) nodes. After obtaining the percentage metabolic data, a node lattice (30 X 20 nodes) can then be constructed. Figures 5-1 to 5-5 show the resultant colored SOM maps for both female and male examinees. It can be noted that the coloring portrays a grouping of the trained data. Healthy members are grouped (zoned) in clear sections while those at risk appear in sections marked red. Figures 5-1 and 5-2 show a similarity but with a rotate right behavior. Figure 5-3 represents the same figure 5-1 but this time having numbers 0-5max. Included in figure 5-3 are sampled examinees that medical experts gave for comparison purposes. Simulations were done for Torus and Spherical SOM with Emergent SOM (ESOM) maps added for clarity. The maps portray the metabolic health status for the entire sampled population. It can be observed that the user of the maps can gauge the extent of the risk at a glance, the ratios and hence the urgency of the threat posed over the population. The maps can be used to monitor other patients of interest to the physician.

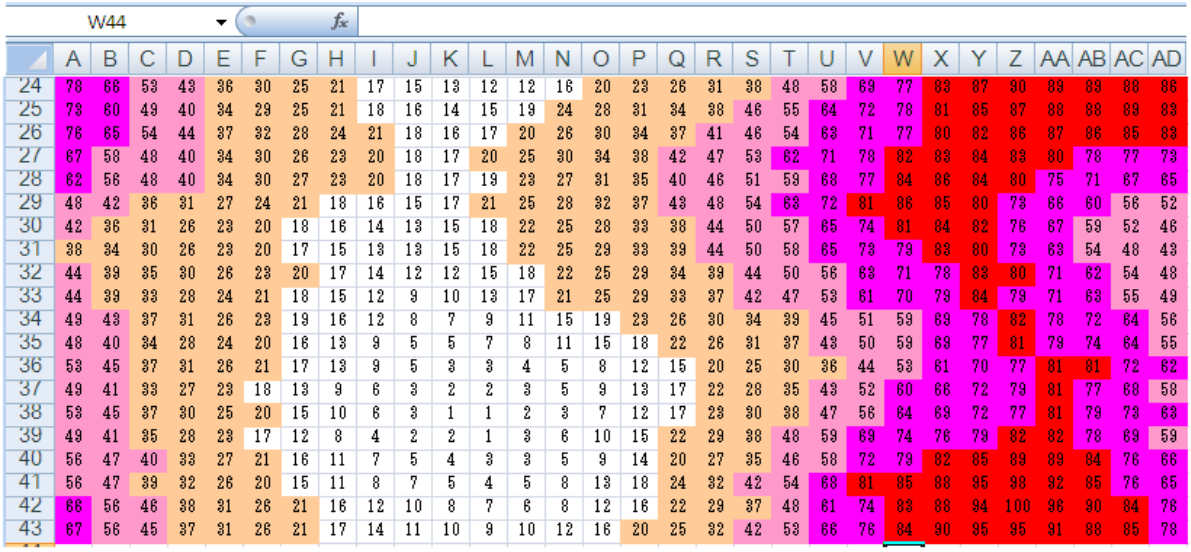
Figure 5-3 shows a map with specific examinees health status posted to their respective clusters (M25-M46). Referring to the same figure, some nodes appear to have two ratings. These are the boundary nodes between two ranked clusters. Physician's expert knowledge on the health metabolisms can automatically register worthwhile information about his patient's condition before diagnosis the tested health parameters.

Table 5-4: Sample of percentage metabolic points calculations



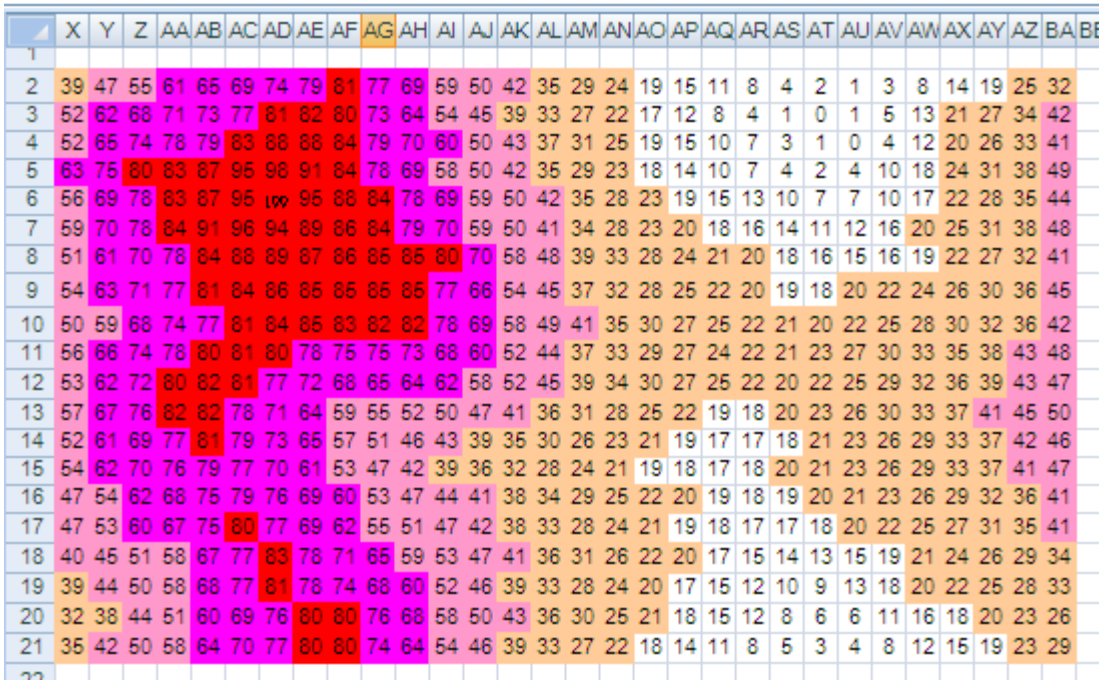
	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R
32	31	0.073213	0.065327	0.562407	0.026575		1.751	0.5715	67.4	32.6	53	52	52		31	52		
33	32	0.134775	0.091418	0.654621	0.046659		1.751				63				32	62		0-19
34	33	0.22485	0.11869	0.692467	0.081224		1.751				69				33	68		20-39
35	34	0.325358	0.144144	0.668372	0.130499		1.751				71				34	71		40-59
36	35	0.432637	0.163422	0.607983	0.205636		1.751	0.7911	54.8	45.2	73	73	73		35	73		60-79
37	36	0.556589	0.158543	0.516741	0.301847		1.751	0.8325	52.5	47.5	77	77	77		36	77		80-100
38	37	0.676355	0.131444	0.399294	0.365455		1.751	0.8762	50	50	81	81	81		37	81		Color code
39	38	0.751191	0.093207	0.265686	0.37991			0.8876	49.3	50.7	8				38	82		
40	39	0.758681	0.058412	0.147009	0.37618			0.8615	50.8	49.2	8				39	80		
41	40	0.705165	0.036159	0.079676	0.35527			0.7944	54.6	45.4	7				40	73		
42	41	0.615915	0.023443	0.050558	0.31515			0.6941	60.4	39.6	64	64	64		41	64		
43	42	0.517221	0.013385	0.031373	0.278608				5.4				54		42	54		
44	43	0.430384	0.007488	0.021239	0.253172				.5				45		43	45		
45	44	0.362611	0.004481	0.017669	0.229862				.5				39		44	39		
46	45	0.306325	0.002572	0.016175	0.199476		1.751	0.3659	79.1	20.9	34	33	33		45	33		
47	46	0.258066	0.001395	0.014218	0.162339		1.751	0.3052	82.6	17.4	28	27	27		46	27		
48	47	0.214916	0.000728	0.013279	0.12442		1.751	0.2487	85.8	14.2	23	22	22		47	22		
49	48	0.174485	0.000493	0.009885	0.093048		1.751	0.198	88.7	11.3	18	17	17		48	17		
50	49	0.136227	0.000553	0.005751	0.065879		1.751	0.1514	91.4	8.65	14	12	12		49	12		
		BMI	H-BP	GLU	TG													

Using the data in column P, a 30x20 node lattice of the trained data can be constructed as shown in figures (5-1 to 5-3)



	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z	AA	AB	AC	AD
24	78	66	53	43	36	30	25	21	17	15	13	12	12	16	20	23	26	31	38	48	58	69	77	83	87	90	89	89	88	86
25	73	60	49	40	34	29	25	21	18	16	14	15	19	24	28	31	34	38	46	55	64	72	78	81	85	87	88	88	89	88
26	76	65	54	44	37	32	28	24	21	18	16	17	20	26	30	34	37	41	46	54	63	71	77	80	82	86	87	86	85	83
27	67	58	48	40	34	30	26	23	20	18	17	20	25	30	34	38	42	47	53	62	71	78	82	83	84	83	80	78	77	73
28	62	56	48	40	34	30	27	23	20	18	17	19	23	27	31	35	40	46	51	59	68	77	84	86	84	80	75	71	67	65
29	48	42	36	31	27	24	21	18	16	15	17	21	25	28	32	37	43	48	54	63	72	81	86	85	80	73	66	60	56	52
30	42	36	31	26	23	20	18	16	14	13	15	18	22	25	28	33	38	44	50	57	65	74	81	84	82	76	67	59	52	46
31	38	34	30	26	23	20	17	15	13	13	15	18	22	25	29	33	39	44	50	58	65	73	79	83	80	73	63	54	48	43
32	44	39	35	30	26	23	20	17	14	12	12	15	18	22	25	29	34	39	44	50	56	63	71	78	83	80	71	62	54	48
33	44	39	33	28	24	21	18	15	12	9	10	13	17	21	25	29	33	37	42	47	53	61	70	79	84	79	71	63	55	49
34	49	43	37	31	26	23	19	16	12	8	7	9	11	15	19	23	26	30	34	39	45	51	59	69	78	82	78	72	64	56
35	48	40	34	28	24	20	16	13	9	5	5	7	8	11	15	18	22	26	31	37	43	50	59	69	77	81	79	74	64	55
36	53	45	37	31	26	21	17	13	9	5	3	3	4	5	8	12	15	20	25	30	36	44	53	61	70	77	81	81	72	62
37	49	41	33	27	23	18	13	9	6	3	2	2	3	5	9	13	17	22	28	35	43	52	60	66	72	79	81	77	68	58
38	53	45	37	30	25	20	15	10	6	3	1	1	2	3	7	12	17	23	30	38	47	56	64	69	72	77	81	79	73	63
39	49	41	35	28	23	17	12	8	4	2	2	1	3	6	10	15	22	29	38	48	59	69	74	76	79	82	82	78	69	59
40	56	47	40	33	27	21	16	11	7	5	4	3	3	5	9	14	20	27	35	46	58	72	79	82	85	89	89	84	76	66
41	56	47	39	32	26	20	15	11	8	7	5	4	5	8	13	18	24	32	42	54	68	81	85	88	95	98	92	85	76	65
42	66	56	46	38	31	26	21	16	12	10	8	7	6	8	12	16	22	29	37	48	61	74	83	88	94	100	96	90	84	76
43	67	56	45	37	31	26	21	17	14	11	10	9	10	12	16	20	25	32	42	53	66	76	84	90	95	95	91	88	85	78

Figure 5- 1: Male_nocut (Torus_SOM)



	X	Y	Z	AA	AB	AC	AD	AE	AF	AG	AH	AI	AJ	AK	AL	AM	AN	AO	AP	AQ	AR	AS	AT	AU	AV	AW	AX	AY	AZ	BA	BB
1																															
2	39	47	55	61	65	69	74	79	81	77	69	59	50	42	35	29	24	19	15	11	8	4	2	1	3	8	14	19	25	32	
3	52	62	68	71	73	77	81	82	80	73	64	54	45	39	33	27	22	17	12	8	4	1	0	1	5	13	21	27	34	42	
4	52	65	74	78	79	83	88	88	84	79	70	60	50	43	37	31	25	19	15	10	7	3	1	0	4	12	20	26	33	41	
5	63	75	80	83	87	95	98	91	84	78	69	58	50	42	35	29	23	18	14	10	7	4	2	4	10	18	24	31	38	49	
6	56	69	78	83	87	95	99	95	88	84	78	69	59	50	42	35	28	23	19	15	13	10	7	7	10	17	22	28	35	44	
7	59	70	78	84	91	96	94	89	86	84	79	70	59	50	41	34	28	23	20	18	16	14	11	12	16	20	25	31	38	48	
8	51	61	70	78	84	88	89	87	86	85	85	80	70	58	48	39	33	28	24	21	20	18	16	15	16	19	22	27	32	41	
9	54	63	71	77	81	84	86	85	85	85	85	77	66	54	45	37	32	28	25	22	20	19	18	20	22	24	26	30	36	45	
10	50	59	68	74	77	81	84	85	83	82	82	78	69	58	49	41	35	30	27	25	22	21	20	22	25	28	30	32	36	42	
11	56	66	74	78	80	81	80	78	75	75	73	68	60	52	44	37	33	29	27	24	22	21	23	27	30	33	35	38	43	48	
12	53	62	72	80	82	81	77	72	68	65	64	62	58	52	45	39	34	30	27	25	22	20	22	25	29	32	36	39	43	47	
13	57	67	76	82	82	78	71	64	59	55	52	50	47	41	36	31	28	25	22	19	18	20	23	26	30	33	37	41	45	50	
14	52	61	69	77	81	79	73	65	57	51	46	43	39	35	30	26	23	21	19	17	17	18	21	23	26	29	33	37	42	46	
15	54	62	70	76	79	77	70	61	53	47	42	39	36	32	28	24	21	19	18	17	18	20	21	23	26	29	33	37	41	47	
16	47	54	62	68	75	79	76	69	60	53	47	44	41	38	34	29	25	22	20	19	18	19	20	21	23	26	29	32	36	41	
17	47	53	60	67	75	80	77	69	62	55	51	47	42	38	33	28	24	21	19	18	17	17	18	20	22	25	27	31	35	41	
18	40	45	51	58	67	77	83	78	71	65	59	53	47	41	36	31	26	22	20	17	15	14	13	15	19	21	24	26	29	34	
19	39	44	50	58	68	77	81	78	74	68	60	52	46	39	33	28	24	20	17	15	12	10	9	13	18	20	22	25	28	33	
20	32	38	44	51	60	69	76	80	80	76	68	58	50	43	36	30	25	21	18	15	12	8	6	6	11	16	18	20	23	26	
21	35	42	50	58	64	70	77	80	80	74	64	54	46	39	33	27	22	18	14	11	8	5	3	4	8	12	15	19	23	29	
22																															

Figure 5- 2: Male_B0s02cut (Torus_SOM)

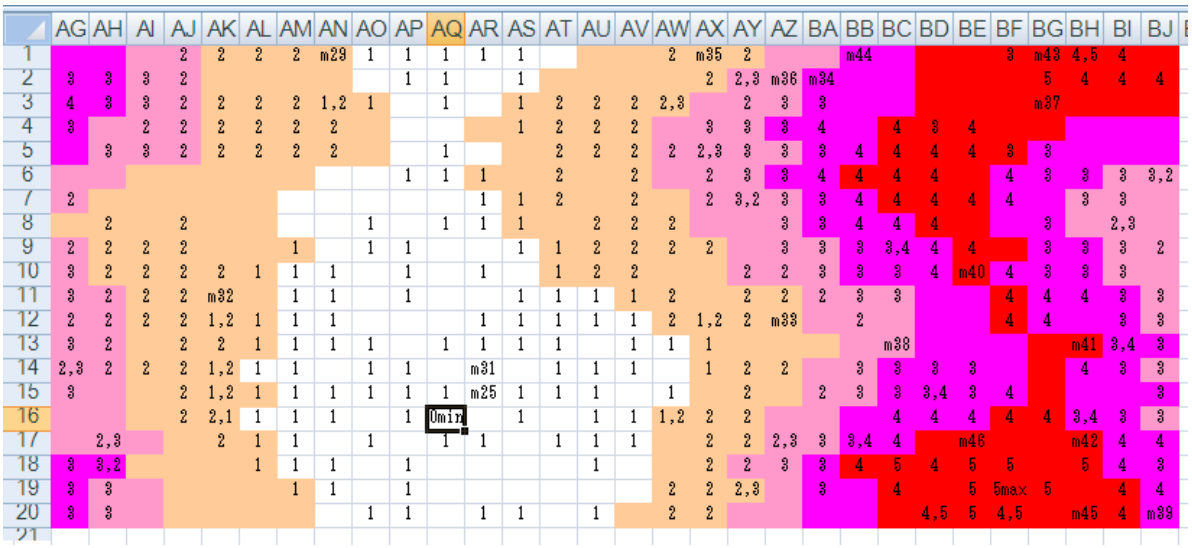


Figure 5- 3: Male_nocut recalibrated (Torus)

5.4 Analysis and Visualization of Metabolic Syndrome

Figures 5-4 and 5-5 are the U-Matrix representation of the map shown in figure 5-1. An overview of examinees' status can clearly be noted without going to the specific parameters affecting the health. Figure 5-4 gives boundaries between clusters whereas the same can clearly be noted by the ridges in figure 5-5. The dark grey areas or zones marked **A-D** are the metabolic risky areas whereas the light grey (hence white) zones are the healthy zones. It is worth noting that dark grey zones **A-D** in a normal SOM map represents the less populated areas or boundaries between clusters, whereas in this application, the zones become the metabolic syndrome zones.

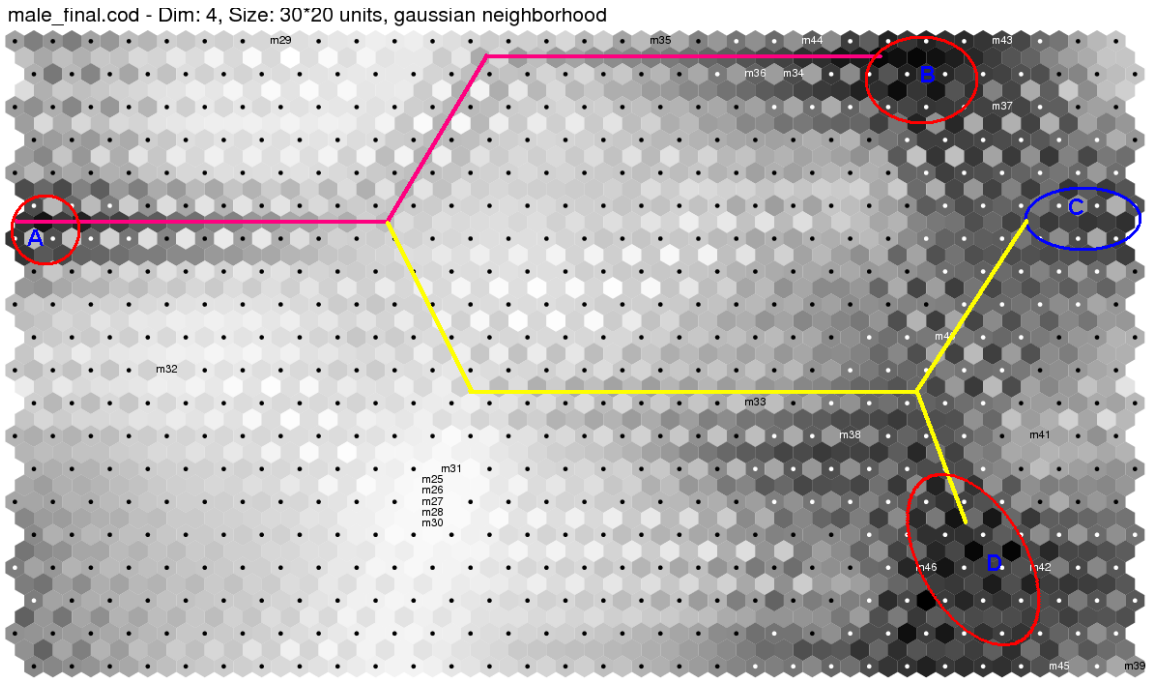


Figure 5- 4: U-Matrix Torus map

Figure 5-5 gives a three dimensional representation of the same dataset. The hilly zones and hence risky areas to venture in if you thought yourself as a mountain climber, are the metabolic syndrome highest risk zones. The planer zones are examinees with the symptom but not acute. The chances of them joining the hilly zones are a matter of the type of parameters causing them to gain the risk. If remedies are taken, they can then join the deep valleys-healthy zones.

5.41 Component Maps

SOM maps have greatly been used to decode the input dataset into various clusters. The trainer monitors the various cluster features (parameters) from the input dataset and any member picked with similar values as the BMU is lumped to that node. The neighborhood function is used to include the neighbors to that particular node. When all

members are grouped to their respective clusters, the computing and hence the training is said to be complete.

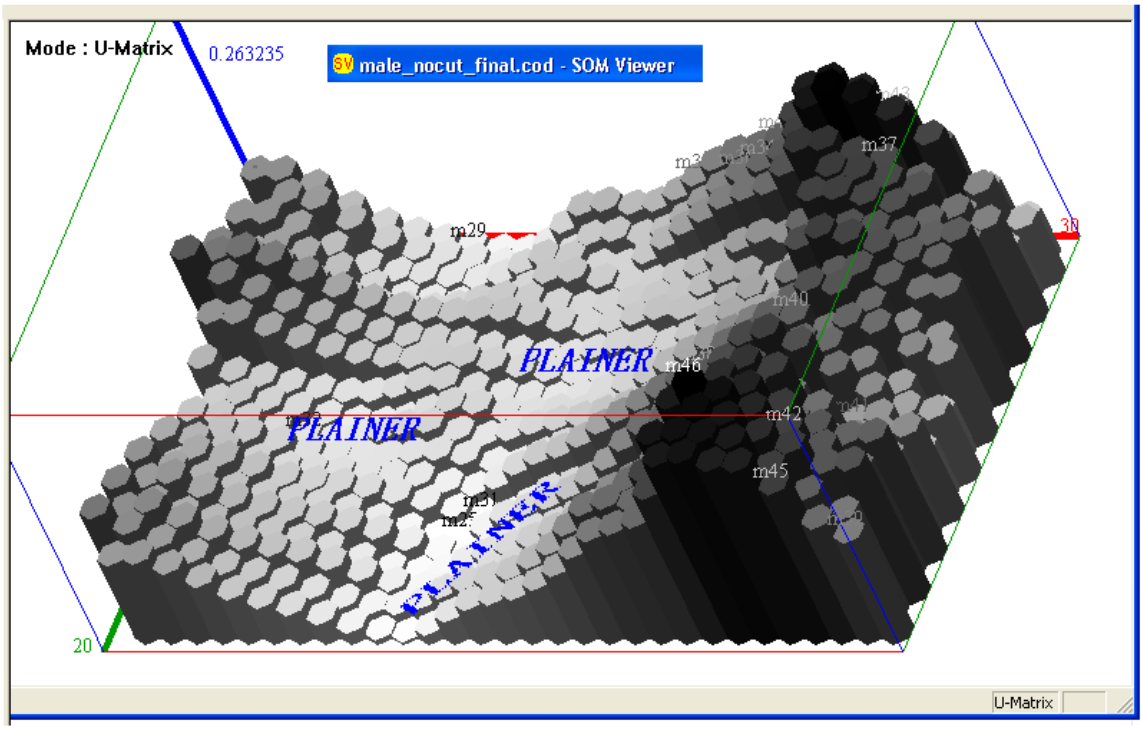


Figure 5- 5: Emergent SOM (recalibrated)

Boundaries between clusters become the differences between them and hence the characteristic features distinguishing the clusters. When one observes the maps shown in figures 5-1 to 5-3, one notices that they give an overall picture of the input dataset. The contribution each component has on the overall metabolic syndrome is not evident. A component map becomes a hardy tool to show the contribution each component has on the overall output SOM. Component maps are an added advantage to the user.

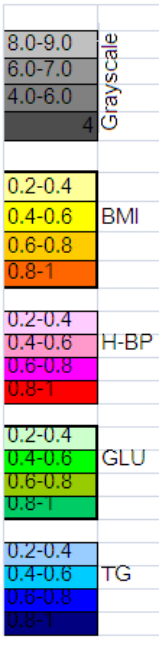
Figure 5-6 displays an integrated component map that shows component inter-relationships. This type of map helps the observer to visualize the most probable risk factor or factors. Displayed items show the contribution each of the four parameters has

to the syndrome. Cases are there where an individual parameter seems the only cause of the metabolic.

Figure 5-6 cannot be drawn without a component sorter shown in Table 5-5. Column **A** gives the specified node whereas columns **B-E** display the trained data matching that node. Each parameter is given a weighting from 0-1 as shown by the legend besides the sorter. Columns **G-J** gives an equivalent weighting (0-4) of the same input parameter weighting. Column **P** gives the resultant component map with gray scaling showing another form of weighting where more than one input parameter contributes to the resultant outcome. Grey represents the gray scale indicating the strength of the bonding between the various principal components.

Table 5-5: Component map sorter

A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q
267	0.01458	0.29	0.07254	0.15732		0	1	0	0	267	0	0	0	1	1	0
268	0.01233	0.32361	0.10387	0.12308		0	1	0	0	268	0	0	0	1	1	0
269	0.01022	0.36826	0.14583	0.08192		0	1	0	0	269	0	0	0	1	1	0
270	0.00992	0.42343	0.19512	0.05135		0	2	0	0	270	1	2	0	1	2	0
271	0.01217	0.59659	0.15523	0.03131		0	2	0	0	271	1	2	0	1	2	0
272	0.01845	0.69005	0.17839	0.04187		0	3	0	0	272	1	3	0	1	3	0
273	0.02826	0.7783	0.17722	0.07157		0	3	0	0	273	1	3	0	1	3	0
274	0.04468	0.8227	0.1599	0.13564		0	4	0	0	274	1	4	0	1	4	0
275	0.07439	0.81235	0.14139	0.25738		0	4	0	1	275	1	4	0	2	5	5
276	0.11086	0.74818	0.11845	0.42468		0	3	0	2	276	2	5	5	2	5	5
277	0.12541	0.63839	0.09121	0.56869		0	3	0	2	277	2	5	5	2	5	5
278	0.11866	0.50277	0.07097	0.66354		0	2	0	3	278	2	5	5	2	5	5
279	0.0878	0.35353	0.05603	0.73096		0	1	0	3	279	1	3	0	2	4	4
280	0.06216	0.22277	0.04207	0.77438		0	1	0	3	280	1	3	0	2	4	4
281	0.04944	0.13028	0.03395	0.77882		0	0	0	3	281	1	3	0	1	3	0
282	0.04026	0.07517	0.02898	0.73618		0	0	0	3	282	1	3	0	1	3	0
283	0.0292	0.03982	0.02686	0.65621		0	0	0	3	283	1	3	0	1	3	0
284	0.01943	0.01927	0.02208	0.56715		0	0	0	2	284	1	2	0	1	2	0
285	0.01404	0.0111	0.01639	0.48393		0	0	0	2	285	1	2	0	1	2	0
286	0.01003	0.00865	0.0122	0.41526		0	0	0	2	286	1	2	0	1	2	0
287	0.00728	0.00807	0.00911	0.36615		0	0	0	1	287	0	0	0	1	1	0
288	0.00578	0.00739	0.00639	0.33136		0	0	0	1	288	0	0	0	1	1	0
289	0.00611	0.00936	0.00482	0.30503		0	0	0	1	289	0	0	0	1	1	0
	BMI	H-BP	GLU	TG												



This approach gives the viewer a better visualization each component has on the input data and hence an alternative method to interpret the input data. SOM Viewer is another

SOM trainer developed with added advantage of displaying U-Matrix, 3D representation and component maps of individual parameters [61]. Figure 5-8 and figure 5-9 show male B0s02cut Torus maps using SOM_Viewer software. All these displays can be viewed under the same roof. The user is able to monitor any node or nodes globally just by flipping from one map to the other. All the above software tools were used to help analyze the metabolic syndrome patterns of the population.

5.42 Analysis and Visualization of male examinees SOM outputs

Referring to figures, 5-2 and 5-3, the results of the trained data for male No_cut and B0s02cut examinees in form of percentage MSP are pasted. It can be observed that the SOM structure drifts to the right and downwards from No_cut to B0s02cut. Input data is elaborately represented in the B0s02cut structures.

Referring to No_cut SOM output-figures (5-1 to 5-5), the syndrome risk is mainly due to HBP, TG and GLU with TG and GLU being the main risky elements. It is worth noting from figures 5-6 and 5-7 that TG is a prime contributor in any cluster where the syndrome appears high.

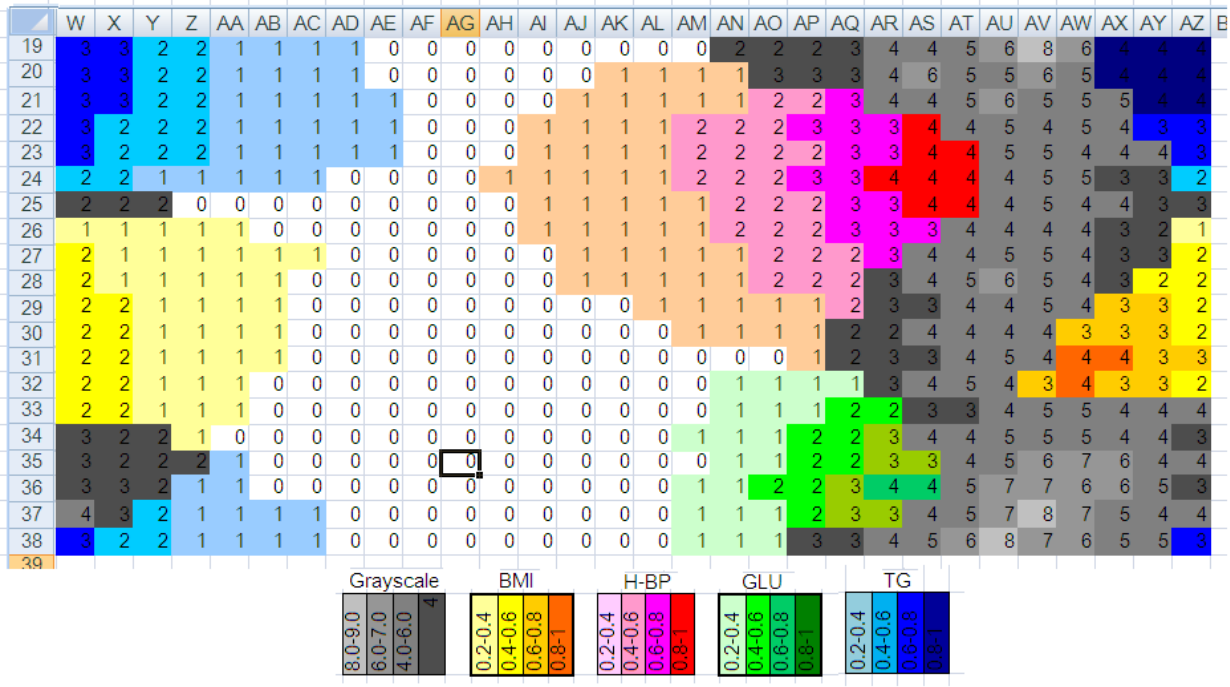


Figure 5- 6: Component map male_nocut (Torus)

Figure 5-2 and 5-7 display male B0s02cut Torus maps. The maps confirm that the parameters contributing greatly to the syndrome risk are HBP, GLU and TG. Figure 5-8 gives three syndrome risk zones **A, B** and **C** with patients like M36, M37 and M42. Male examinees syndrome risk trends to fall in the following categories:

- Male examinees show two main risky trends one due to HBP and TG and the other due to GLU and TG. Hence male risky element is TG.
- BMI element can be seen to have less impact to the metabolic risk trends in male examinees.
- Mostly affected members are in the 40 years and above range.

	AH	AI	AJ	AK	AL	AM	AN	AO	AP	AQ	AR	AS	AT	AU	AV	AW	AX	AY	AZ	BA	BB	BC	BD	BE	BF	BG	BH	BI	BJ	BK
1																														
2	2	2	3	3	5	5	5	5	5	4	4	3	2	2	1	1	1	1	0	0	0	0	0	0	0	0	0	1	1	1
3	2	3	4	4	6	5	5	5	4	4	4	3	3	2	1	1	1	0	0	0	0	0	0	0	0	0	1	1	1	2
4	2	3	3	4	5	6	6	7	6	5	4	4	3	2	2	2	1	0	0	0	0	0	0	0	0	0	1	1	1	2
5	3	4	4	6	5	7	7	7	6	5	5	3	3	2	2	1	1	0	0	0	0	0	0	0	0	1	1	1	2	2
6	3	3	4	5	7	7	9	7	6	4	4	4	3	3	2	1	1	1	1	0	0	0	0	0	0	0	1	1	1	2
7	3	4	4	5	6	8	7	7	6	5	4	3	3	2	2	1	1	1	1	1	0	0	0	0	0	0	1	1	2	2
8	3	3	5	6	6	6	8	6	6	4	4	4	3	3	2	2	1	1	1	1	1	1	0	0	0	0	0	1	1	2
9	4	4	4	6	7	5	6	5	5	4	4	4	3	2	2	2	1	1	1	1	1	1	0	0	0	0	1	2	2	2
10	3	3	4	4	5	6	6	5	4	5	4	4	3	3	2	2	1	1	1	1	1	1	1	0	1	1	1	1	1	2
11	2	3	3	4	5	5	5	5	4	4	3	3	3	2	2	2	1	1	1	1	1	1	0	1	1	1	1	2	2	2
12	2	3	3	4	4	5	4	4	4	4	4	3	3	2	2	2	1	1	1	1	1	1	0	1	1	1	1	2	2	2
13	3	3	4	4	4	4	4	4	4	3	3	2	2	2	1	1	1	1	1	1	0	0	1	1	1	1	1	2	2	2
14	2	3	3	4	4	4	5	4	4	3	3	2	2	2	2	1	1	0	0	0	0	0	0	1	1	1	1	2	2	2
15	2	3	3	4	5	4	4	4	3	3	3	1	1	1	1	1	1	1	0	0	0	0	1	1	1	1	1	2	2	2
16	2	2	3	3	4	4	5	4	4	3	3	2	2	2	1	1	1	1	1	1	0	0	0	1	1	1	1	1	1	2
17	2	2	3	4	4	5	5	4	3	2	2	2	2	2	1	1	1	1	1	1	0	0	0	1	1	1	1	1	1	2
18	2	2	3	3	4	5	6	5	4	3	3	2	2	2	2	1	1	1	1	1	0	0	0	0	1	1	1	1	1	1
19	2	3	3	3	4	5	5	4	3	3	3	2	2	2	1	1	1	1	1	0	0	0	0	0	1	1	1	1	1	1
20	1	1	2	3	5	5	4	5	4	4	3	3	2	2	2	1	1	1	1	0	0	0	0	0	0	0	1	1	1	1
21	2	3	3	3	4	5	5	4	4	3	3	2	2	2	1	1	1	1	0	0	0	0	0	0	0	0	0	0	0	1
22																														

Figure 5- 7: Male B0s02cut Torus

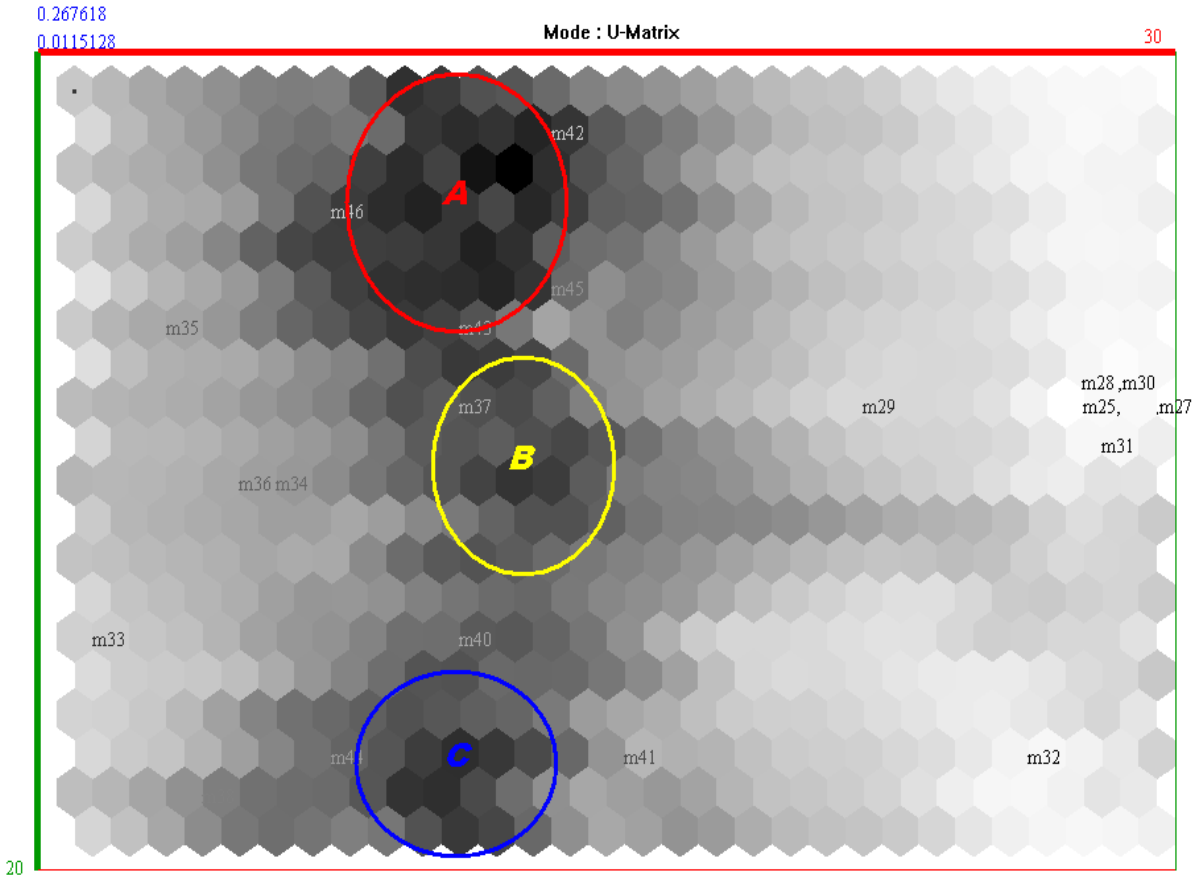


Figure 5- 8: U-Matrix male B0s02cut Torus

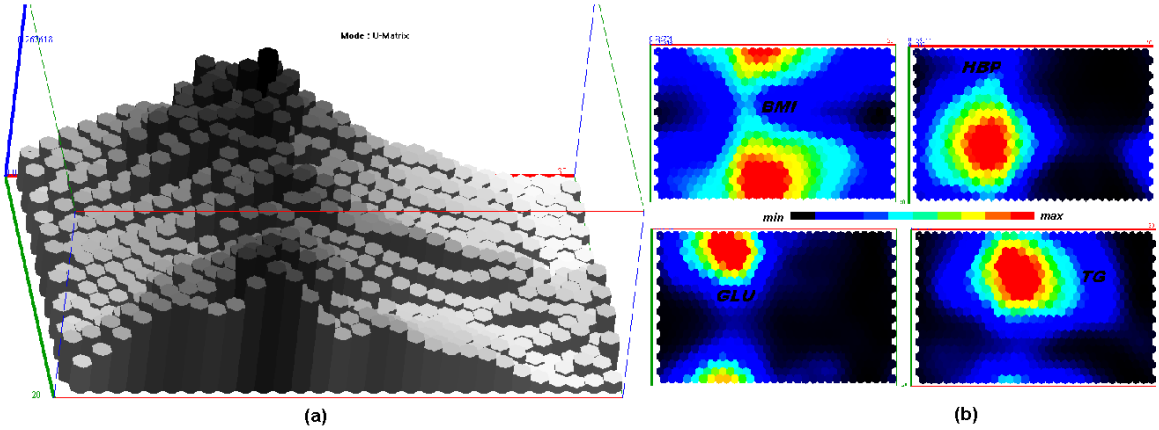


Figure 5- 9: (a) Male Enhance Matrix B0s02cut Torus.
(b) Male B0s02cut component maps

5.43 Analysis and Visualization of female examinees SOM outputs

Figures 5-10 to 5-13 show the female No_cut and B0s02cut Torus. The B0s02cut map transverse upwards with BMI having least drift. Input data is elaborately represented in the B0s02cut structures. More attention will be given to the B0s02cut Torus maps during analysis and visualization of the syndrome. Integrated Principal component map for female examinees is shown in figure 5-11. Female examinees risk elements are BMI, HBP and TG with the main risky element being BMI.

- Female examinees show different trends in that the two main risky trends are due to elements BMI and HBP, and the other HBP and TG. The main risk element being BMI.
- The blood glucose level seems to have less impact to the metabolic risk trends.
- TG and HBP have a bigger share in the metabolic risk.
- The examinees with high BMI factors seem to have fewer complications from the other parameters.
- Mostly affected examinees are in the 40 years and above range.

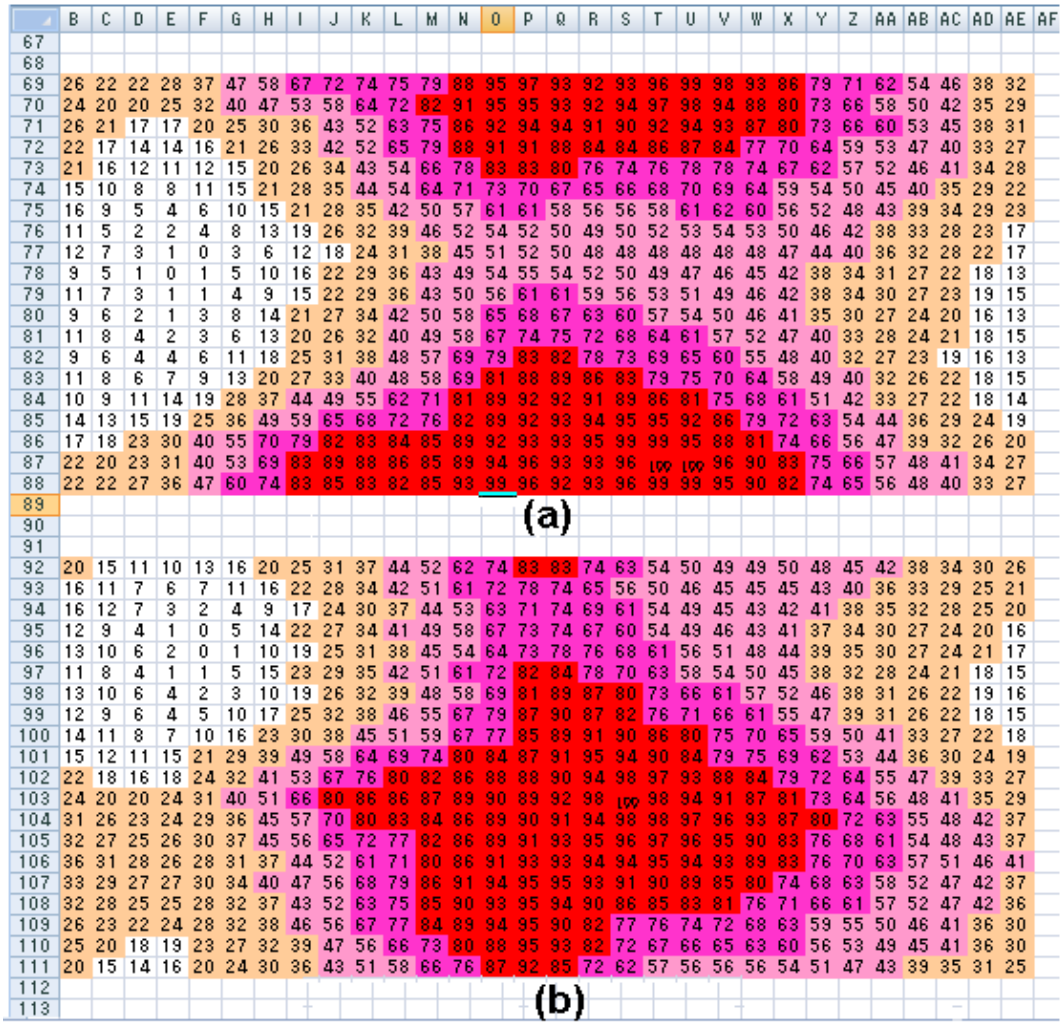


Figure 5- 10: (a) Female_nocut Torus. (b) Female_B0s02cut Torus

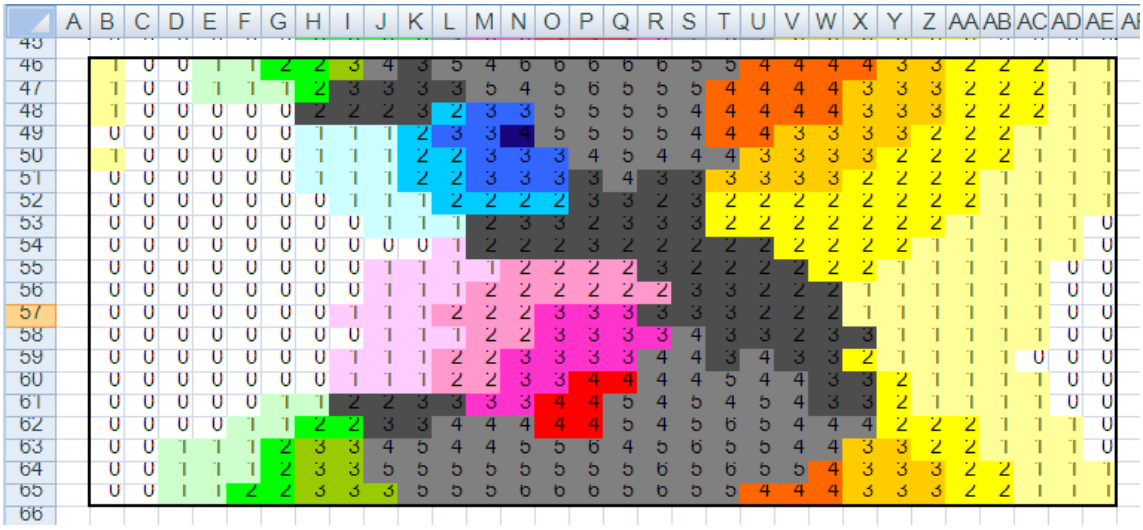


Figure 5- 11: Female No_cut component map Torus

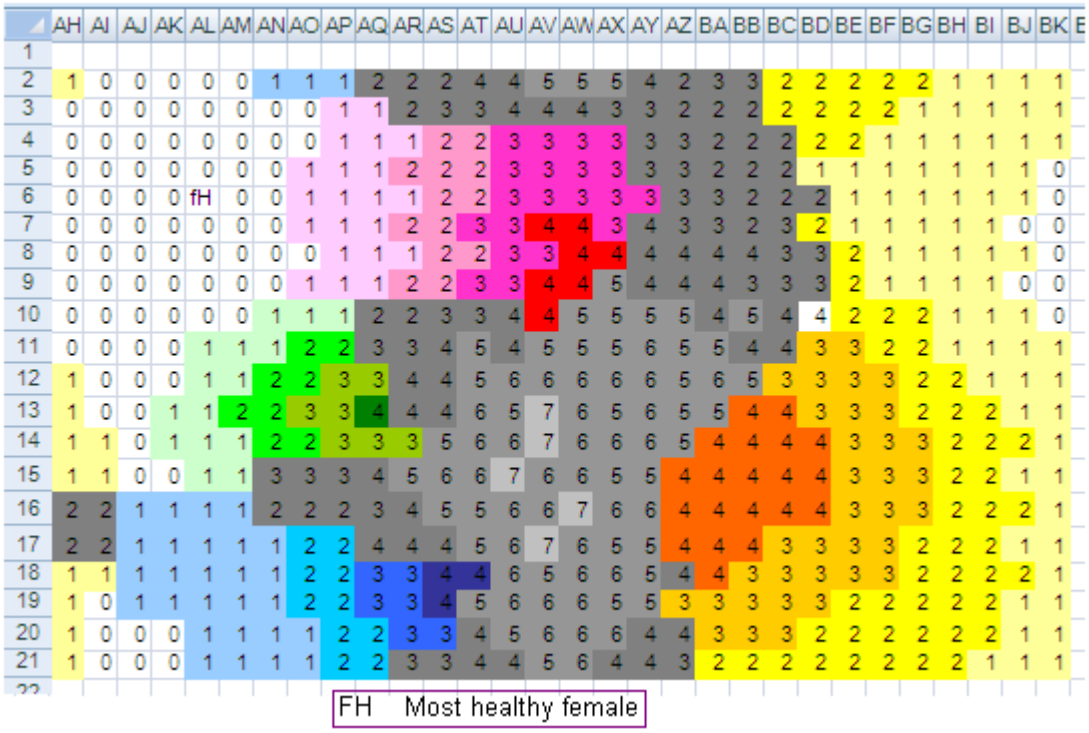


Figure 5- 12: Female B0s02cut component map

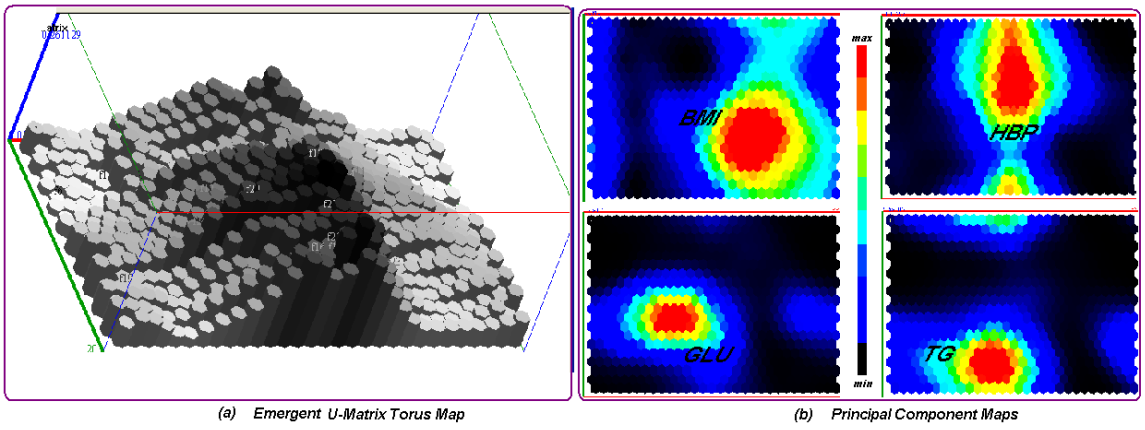


Figure 5- 13: Female Torus maps

5.44 Six (6) health parameters metabolic syndrome evaluation

The research work presented so far considered four (4) parameters. The choice of the parameters was aimed at investigating the effect the commonly known parameters had on the syndrome. It should be noted that the more the parameters, the complex the

analysis. The global definition of at least three of the parameters was fulfilled by taking the four parameters. The analysis and visualization of the syndrome was satisfactory compared to the outcome of the diagnosis of the medical experts shown later [Table 5-6 pp.125]. It is from these findings that an evaluation tool was designed.

The analysis and visualization of the syndrome could not be complete without considering the other two parameters namely low blood pressure (LBP) and high-density lipoproteins (HDL). According to the medical experts, the identified parameters that are associated with metabolic syndrome total to the six considered. The ceiling values for the two parameters were taken to be 1.14 and 1.25 for male and 1.12 and 1.14 for female examinees respectively [Table 5-3 pp.104].

Figures 5-14 and 5-16 were used to identify the most probable zones that will cater majority of examinees under risk. With the help of principal component maps shown in Figures 5-15 and 5-17, the contribution each input parameter had on the overall metabolic risk was identified. This approach gave an indication of the most probable trends the syndrome was taking. Figure 5-18 shows the trends each zone took and the parameters associated with them.

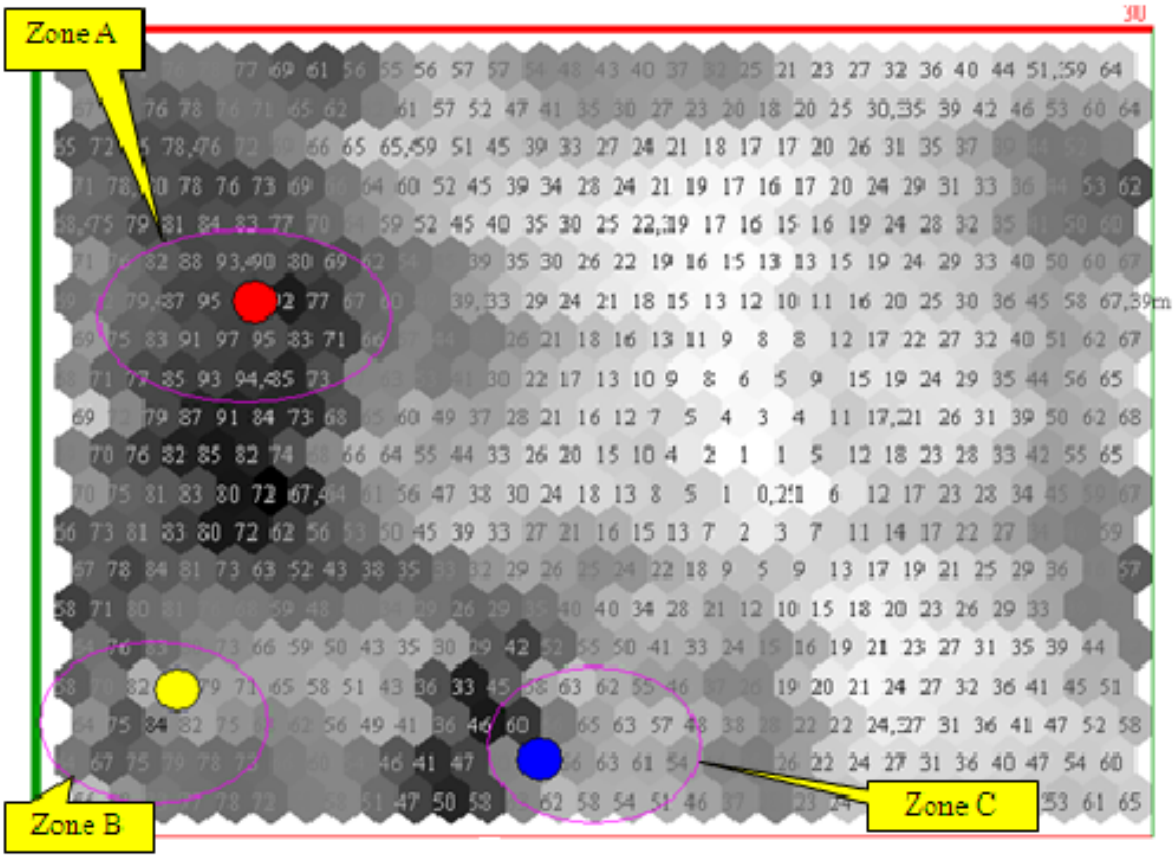


Figure 5- 14: Male six parameter torus map

The trends represent the current status of the examinees. If any of the parameters worsen within the group or zone, the status will change to a new combination. This behavior can be interpreted as the members joining a new cluster within the SOM.

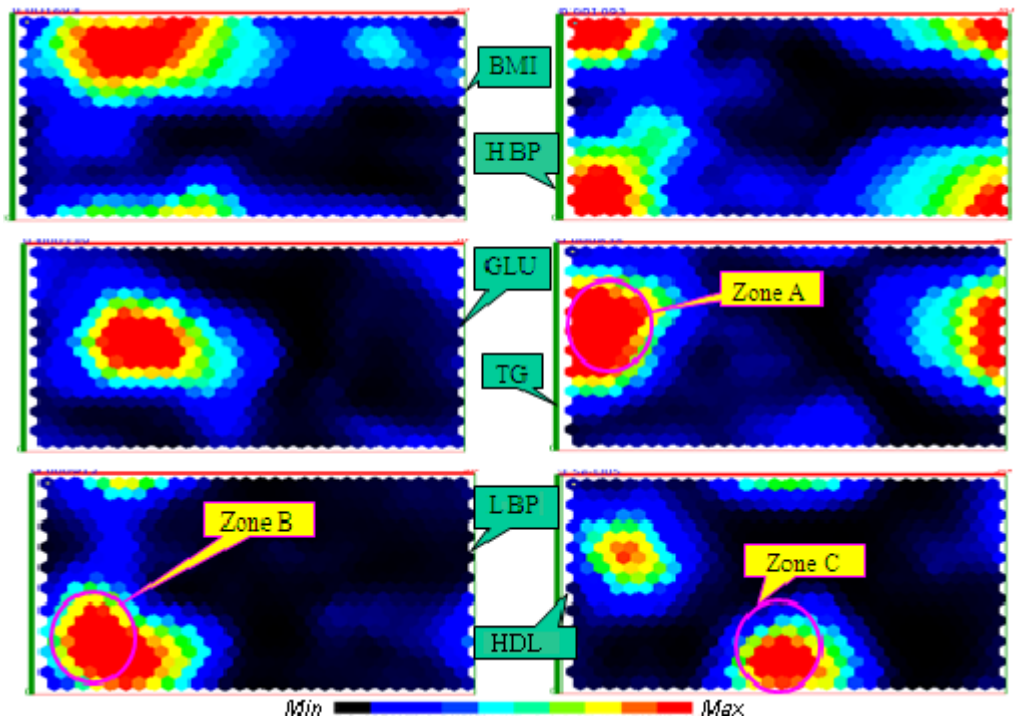


Figure 5- 15: Male six principal component maps

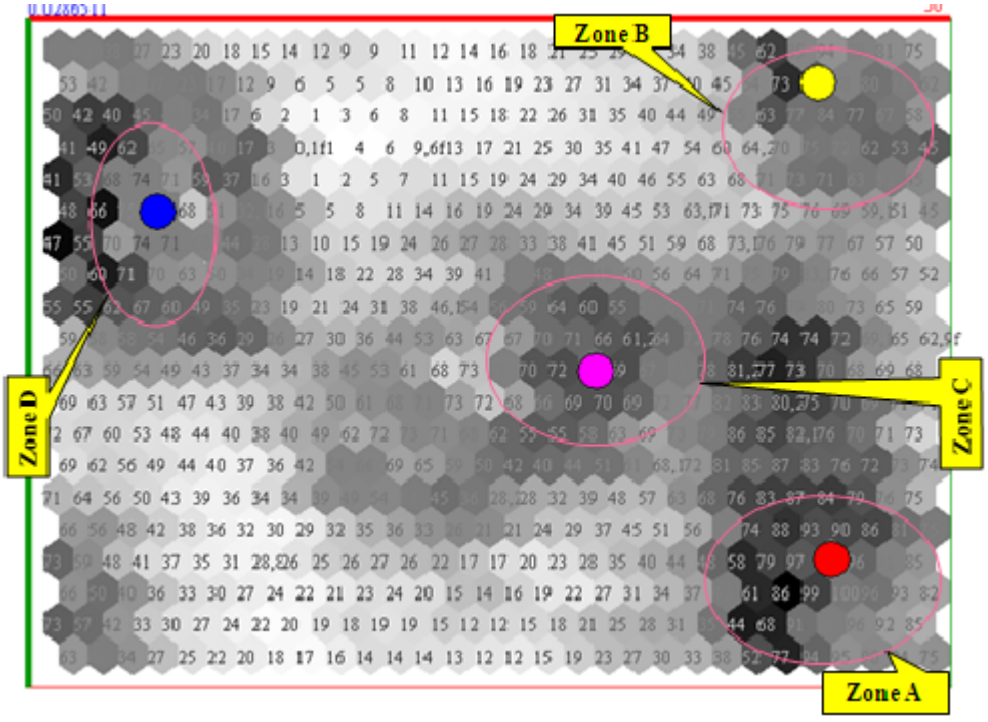


Figure 5- 16: Female six parameter torus map

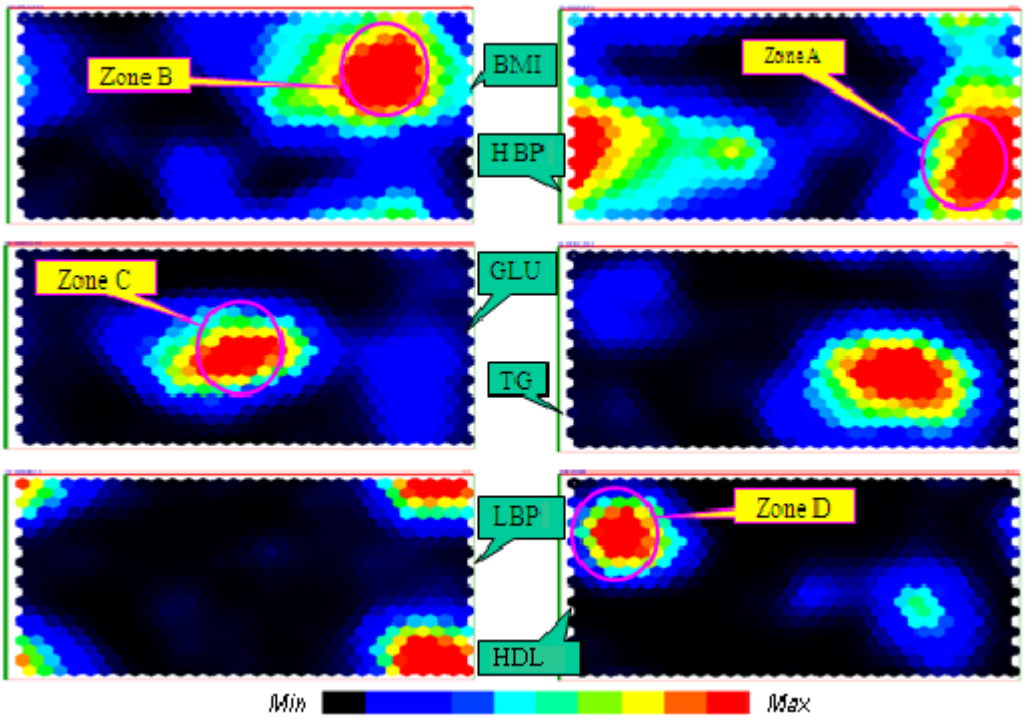


Figure 5- 17: Female six principal component maps

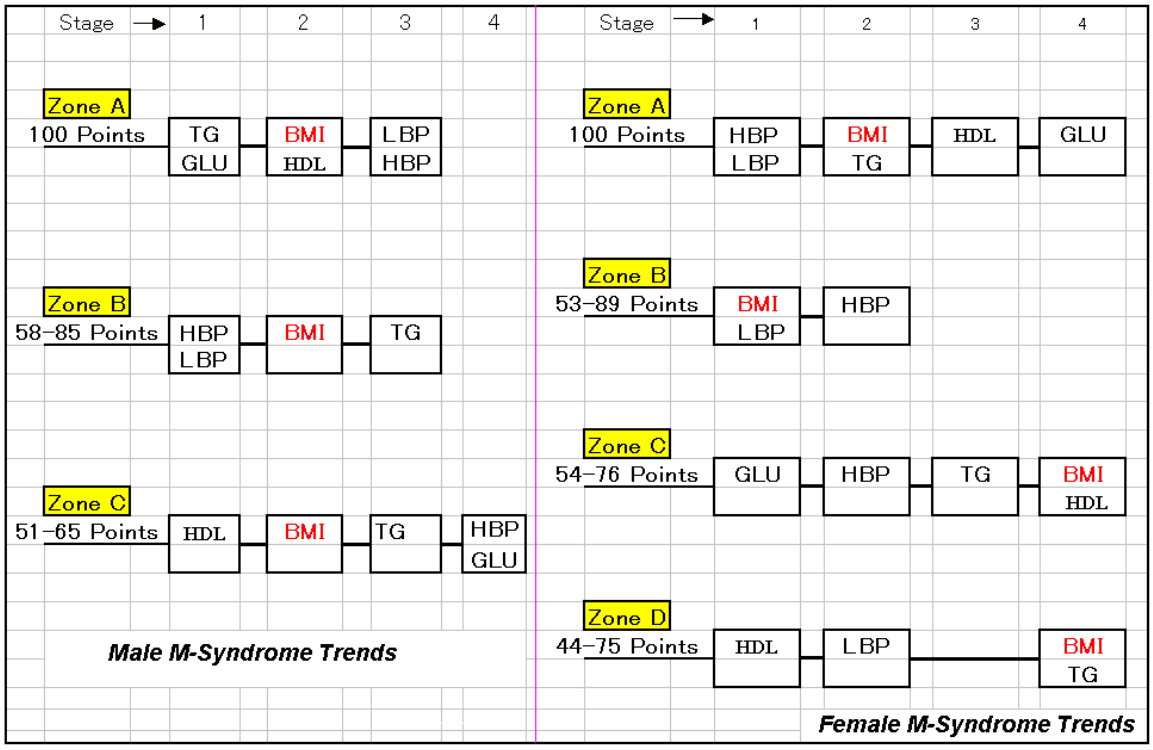


Figure 5- 18: Male-Female Metabolic syndrome trends chart.

Figure 5-18 gives a trend chart obtained from both visualization and analysis of all examinees.

It was observed that:

- Male examinees show risk parameters in the priorities HBP, TG, BMI, LBP, HDL and then GLU.
- Females' trends have BMI, HBP and LBP, TG, HDL and then GLU.
- GLU seems to have its unique effects on the metabolic syndrome.
- Taking the whole spectrum of examinees, the priorities go by order TG, HBP and then BMI respectively.
- HDL stands more independent but has relations with TG and HBP for the males.

Table 5-6 gives a comparative table of SOM output analysis and actual physician results on samples of both male and female patients. The legend column has metabolic syndrome risk given in percentages as well as color coding. Green color indicates the difference between the analysis and the diagnosis. Two opinions have been given; one using four parameters and the other six. SOM simulations outputs considered for both cases were B0s02cut metabolic.

Four parameter analysis:

The differences had some cases that appeared extreme with the example being patient f9, f10 and f11.

Six parameter analysis:

There appeared differences but on the marginal.

The comparisons show that the differences shown may be marginal in terms of percentages but the simulation results are able to give results to both precision and the

degree of syndrome risk. There is a great difference between verbal expression from the physician and an image to express the same. As the saying goes “A picture gives more information than a thousand words” the physician would be able to give his expert knowledge in conjunction with the maps much easily and precisely. Adding the degree of risk in the display has a more satisfying solution to both the physician and the examinees.

Table 5-6: Male-Female Metabolic Check Tabulation

	SEX	Age	BMI	HBP	TG	GLU	Meta_doc	Meta_SOM	Nocut	B0s02cut	Girth of ab	LBP	HDL	Meta_doc	Meta_SOM	B0s02cut	Legend
							Meta_syndrome=1		meta4	meta4				Meta_syndrome=1		meta6	
f1	0	56	24.2	122	37	103	0	0	0	0	77.5	78	69	0	0	3	
f2	0	69	17.9	128	77	105	0	0	0	0	69.5	70	80	0	0	3	80-100
f3	0	64	20.4	100	112	90	0	0	0	0	68	72	64	0	0	3	70-79
f4	0	38	20.4	116	131	84	0	0	0	0	82.5	80	53	0	0	3	50-69
f5	0	60	20.1	130	67	90	0	0	0	0	70.5	80	92	0	0	3	0-49
f6	0	79	25	120	120	88	0	0	12	11	90	70	59	0	0	8	Difference
f7	0	69	21.1	130	93	109	0	0	2	0	70	70	79	0	0	3	f- Female
f8	0	72	20.3	154	85	88	0	0	34	35	73.5	80	49	0	0	26	f - SEX=0
f9	0	66	26	168	92	108	0	1	83	70	101.5	50	76	0	1	56	m-Male
f10	0	71	22.9	148	769	82	0	1	88	84	85	88	68	0	1	65	m- SEX=1
f11	0	71	27.7	156	76	98	0	1	70	75	98	78	87	0	1	53	
f12	0	75	22.7	134	51	146	0	1	60	66	83.5	80	59	0	0	48	
f13	0	57	30.9	142	146	100	1	1	98	94	98.5	78	38	1	1	68	
f14	0	64	30.4	160	167	101	1	1	100	100	101	78	57	1	1	82	
f15	0	74	29.2	134	159	102	1	1	86	83	92	70	52	1	1	63	
f16	0	55	22.7	160	462	120	1	1	91	88	95	80	50	1	1	67	
f17	0	70	25	158	261	127	1	1	88	88	90	80	41	1	1	67	
f18	0	72	27	164	157	129	1	1	93	89	95.1	90	82	1	1	71	
f19	0	75	25.5	140	166	92	1	0	26	28	90	74	38	1	0	34	
f20	0	65	26	130	205	124	1	1	52	61	94	80	58	1	0	46	
f21	0	63	28.7	178	370	139	1	1	97	95	99	84	39	1	1	79	
f22	0	74	29.7	156	528	132	1	1	94	95	98	90	37	1	1	85	
f23	0	53	29.1	144	150	110	1	1	86	83	97.5	88	39	1	1	68	
f24	0	67	25	140	182	123	1	0	43	37	99	70	39	1	0	33	
m25	1	74	22.4	110	99	102	0	0	0	0	84.5	80	51	0	0	0	
m26	1	63	22.5	128	70	88	0	0	0	0	79.5	74	42	0	0	0	
m27	1	28	20.5	122	70	93	0	0	0	0	67	68	63	0	0	0	
m28	1	53	21.5	106	51	104	0	0	0	0	78.5	70	78	0	0	0	
m29	1	40	22.8	126	185	88	0	0	20	20	83.5	78	46	0	0	20	
m30	1	43	22.2	130	91	93	0	0	0	0	83.9	80	81	0	0	0	
m31	1	78	21.5	140	84	95	0	0	1	0	79	60	60	0	0	2	
m32	1	68	25.8	124	55	92	0	0	25	24	90	74	44	0	0	22	
m33	1	72	25.6	150	59	100	0	0	36	20	86	74	61	0	0	29	
m34	1	69	24.3	164	87	138	0	1	64	59	86	88	54	0	1	63	
m35	1	68	24	150	86	125	0	0	30	26	82.5	70	57	0	0	30	
m36	1	69	24.1	160	105	130	0	1	54	50	90	80	73	0	1	54	
m37	1	71	26.3	170	572	94	1	1	87	85	89	84	46	1	1	75	
m38	1	67	26.5	150	77	125	1	1	52	51	89.5	78	60	1	0	39	
m39	1	71	25.6	124	327	111	1	1	77	70	99.5	76	53	1	1	73	
m40	1	68	29.1	178	226	97	1	1	84	83	99.5	90	47	1	1	81	
m41	1	49	29.6	134	150	106	1	1	80	76	97	86	40	1	1	73	
m42	1	65	37.2	126	295	131	1	1	84	84	116	84	33	1	1	93	
m43	1	65	26.1	154	515	132	1	1	89	86	86.5	80	52	1	1	77	
m44	1	83	25.4	160	150	157	1	1	68	71	94.2	80	57	1	1	63	
m45	1	41	26.5	132	390	121	1	1	88	84	89	70	36	1	1	85	
m46	1	71	26.4	138	220	170	1	1	84	83	95.5	56	29	1	1	87	

5.5 Spherical SOM Simulations

Trained data can be mapped onto a spherical surface instead of a two dimensional surface. One of the contributions Torus SOM had in the representation of trained SOM was to make data mapped on the edges of the plane to have relationship with those on other edges. Spherical SOM maps the input data onto a spherical surface which has a uniform phase expression. The displayed data can be viewed on any part of the sphere by rotation. This method of representing the data makes the user to have glance at any part of the sphere. It is possible to mark the position of interest on the sphere particularly when dealing with component maps.

Dendrogram can be defined as a branching diagram showing the interconnections between things. The parameters being monitored can be interconnected since they are taken to originate from the same source. This concept of interconnection was used to derive the trends the input parameters follow. In doing so, the root of the branching emerges to be the main contributor to the metabolic syndrome. This can be interpreted to mean that the frequency to which that particular parameter occurred within the trained data was the highest. The branching can be seen as a tree with the root being the dominant parameter. A healthy tree is thus one of possible dendrograms.

Spherical SOM was used to train the same data (B0s02cut) that was used to train Torus SOM with "blossom" software [48] as the tool to construct the SOM. Figure 5-19 is a trained spherical SOM from male B0s02cut data. Gray scaling is used for the population density of the examinees. Figure 5-19(a) is colored light grey indicating examinees in this cluster are healthy members. Figure 5-19(b) has a dark grey zone

marked 5MAX. Examinees falling on this cluster have a highest metabolic risk. On the smooth face of the sphere, nodes are implanted. These are the matching patterns from selected male data with their degree of metabolic risks ranked in percentage basis. Figure 48 illustrates the Principal component of the trained female examinees' data, where individual's health can be analyzed using this "blossom" tool. It is good to note that the location of an examinee on the spherical SOM need to be constant as one varies the type of component to display. This is illustrated by a red spot on all maps. Each member can visualize the degree of risk one may be in by observing each principal component used in the training.

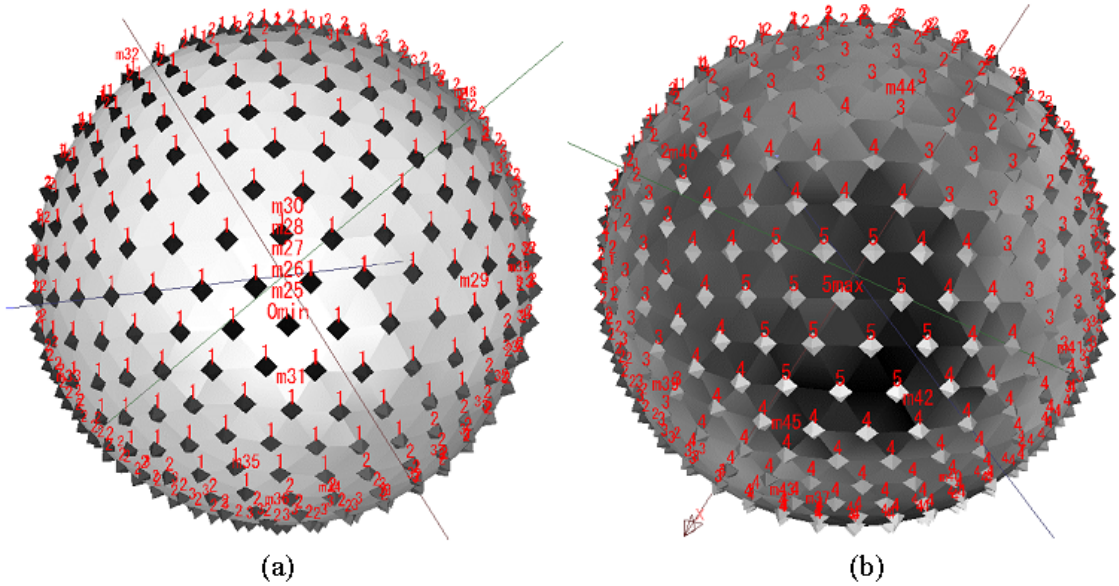


Figure 5- 19: U-Matrix male spherical SOM

Referring to the component maps shown in figure 5-20, examinees like F23 and F15 have high values of BMI but low values HBP and GLU and even very low values of TG. Referring to figure 5-21, some female examinees have very high BMI but with fewer complications with the other health parameters. It can then be pointed out that some of

the examinees don't suffer from diseases that are associated with BMI such as cardiovascular disorders or diabetes.

Taking issue on the common beliefs that BMI is the root cause of the syndrome, we note that female examinees belonging to this cluster have fewer complications to other risk factors.

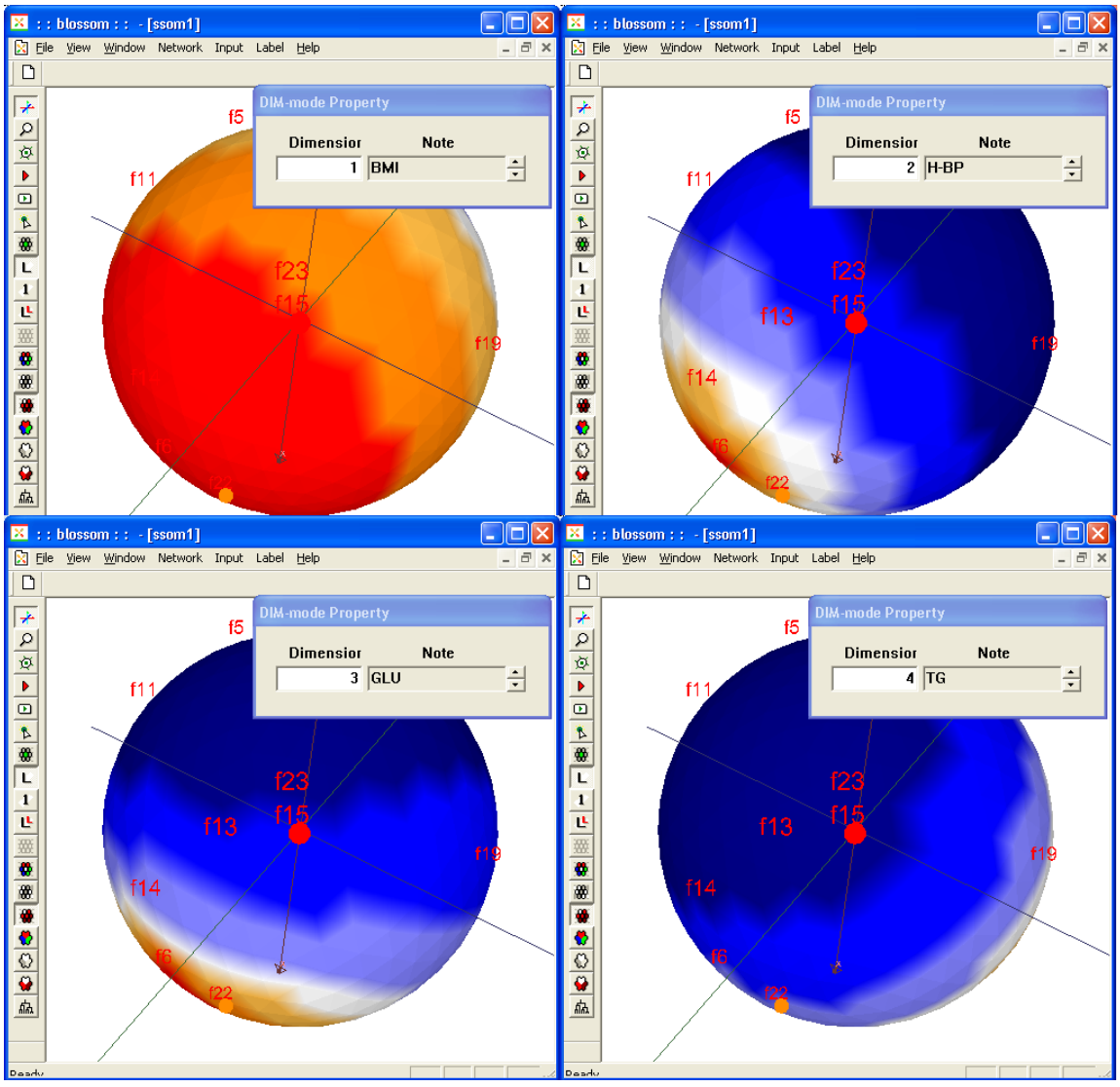


Figure 5- 20: Female principal component maps (4 parameters)

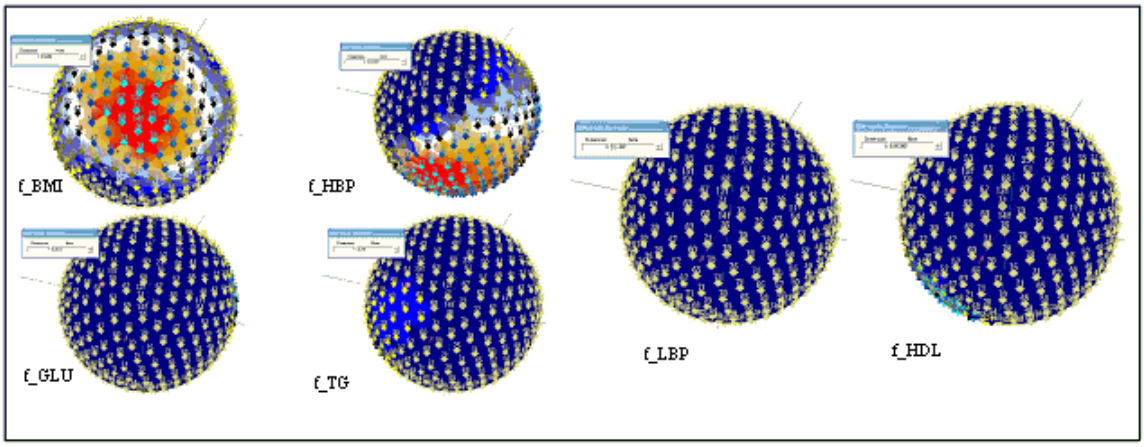


Figure 5- 21: Female principal component maps (6 parameters)

Figure 5-22(a) shows a glyptic SOM representation of male examinees' health parameter data where a big proportion of the volume of examined members is affected by TG. The other parameter having a share in the syndrome risk is HBP. BMI happens to have fewer members having the syndrome risk. Figure 5-22(b) gives an examinees metabolic evaluation dendrogram where two distinct categories emerge, healthy affiliated examinees and those affiliated to metabolic related parameters. Using similar approach female examinees metabolic syndrome trends were analyzed and Figure 5-23(b) displays their general pattern dendrogram.

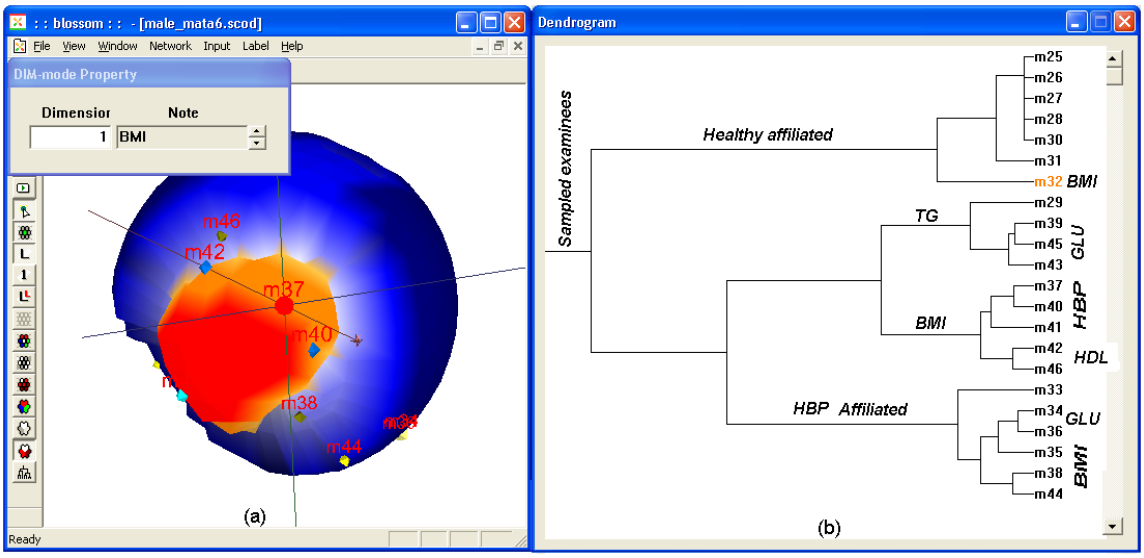


Figure 5- 22: Male Glyptic SOM and Dendrogram

Male examinees have one major metabolic parameter HBP while females have BMI. The capillary branches in the sampled male and female data show diseases that are affecting their health. Some members may have traces of particular diseases, an indication showing the observer need to check on that parameter.

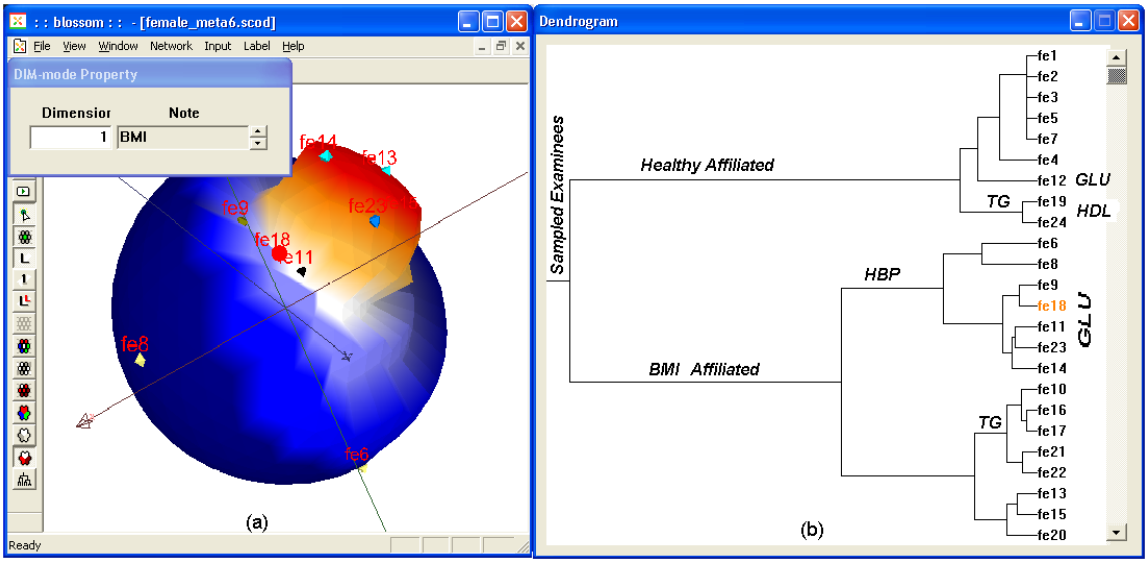


Figure 5- 23: Female glyptic SOM and Dendrogram

5.6 Metabolic syndrome SOMs considering age clusters

The syndrome trends may not be conclusive without considering the age bracket. It is a known fact that the defense mechanism of human body to diseases changes depending on various parameters like environment, age and malnutrition to mention a few. To monitor the contribution age had on the metabolic syndrome, the examinees were clustered into age which varied from 19 to 93 for males and 17 to 94 years for the females. The clusters were decided as:

Male Examinees 19 to 30, 31 to 40, 41 to 50, 51 to 60, 61 to 70, 71 to 80 and 81 to 93 years

Female examinees 17 to 30, 31 to 40, 41 to 50, 51 to 60, 61 to 70, 71 to 80 and 81 to 94 years.

Each cluster has some members sampled at random for mapping to the overall spherical map. The clusters are then analyzed, visualized with an intention of observing the most probable causes of metabolic trends within each age cluster. It is worth noting that some clusters had huge population of members. This situation necessitated cluster SOM training before sampling was attempted. The trends are then used to foresee the overall behavior patterns of the gender examinees.

5.61 Male Clusters

19-30 years cluster

Figure 52 shows a spherical glyptic component map of the cluster. It can be observed that any member with the syndrome is due to BMI parameter only. Majority of the members are in the healthy zone as shown in figure 53 (b). There are traces of TG and HDL in the sampled examinees, but they are at very low quantities.

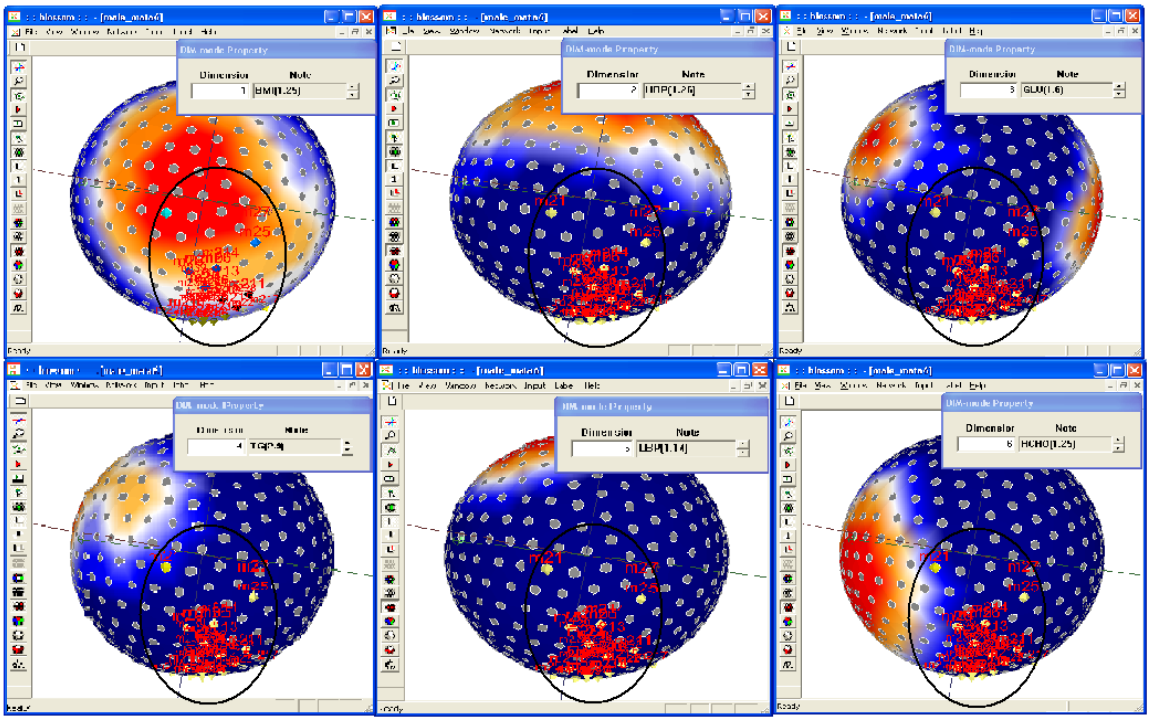


Figure 5- 24: Male 19-30 cluster's spherical component maps

As an observation, the physician may look at the scenario differently. A percentage metabolic syndrome map is shown in Figure 54 where mark point 1 represent percentage syndrome points 0 to less than 20 while 5 is 80 to less than 100 percent. Mark '5max' represent 100 percent metabolic.

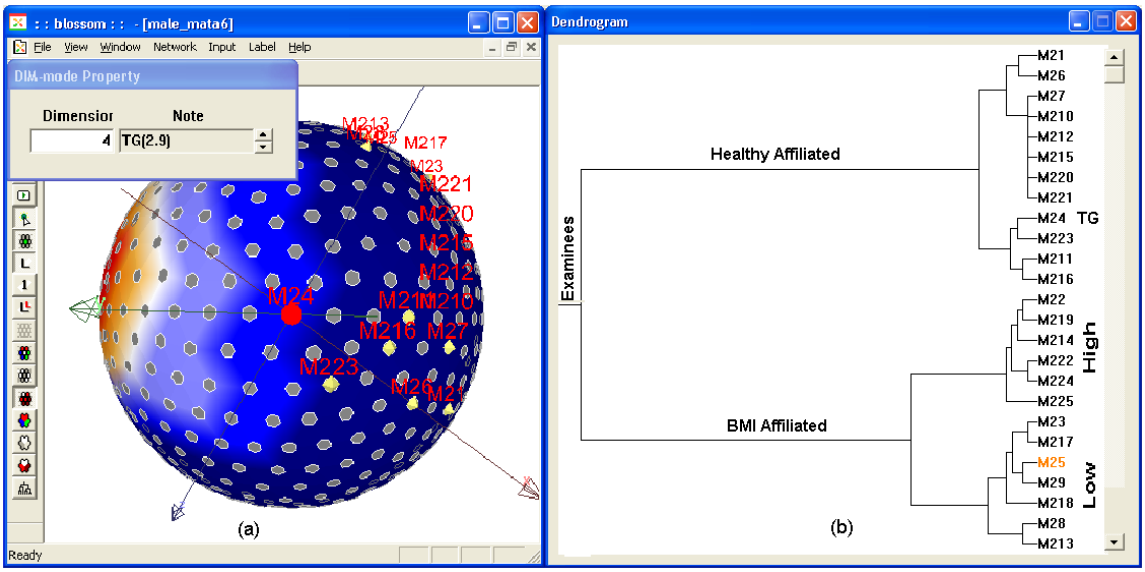


Figure 5- 25: Male 19-30 spherical component map and dendrogram

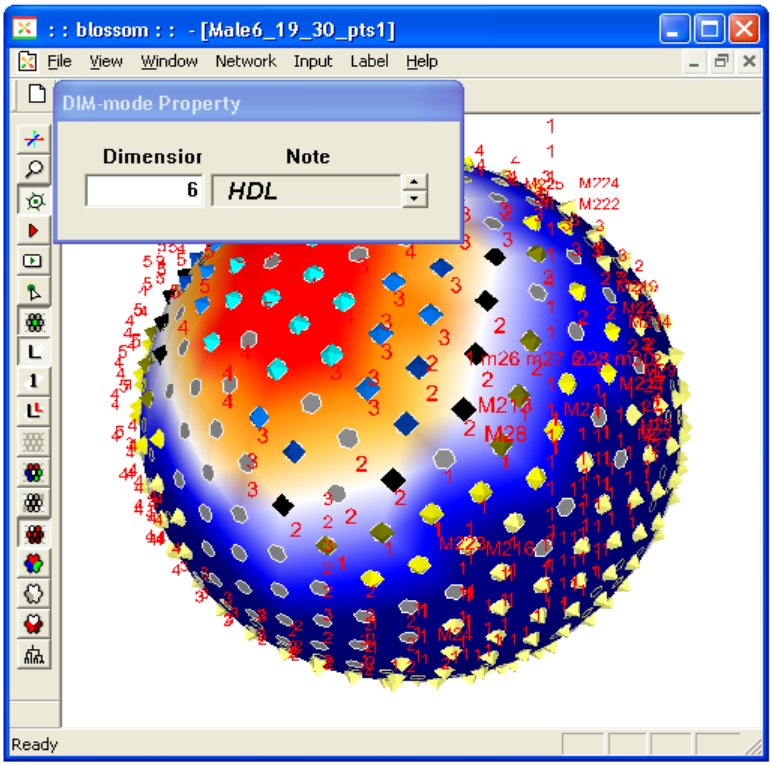


Figure 5- 26: Male spherical SOM mark points (1-5)

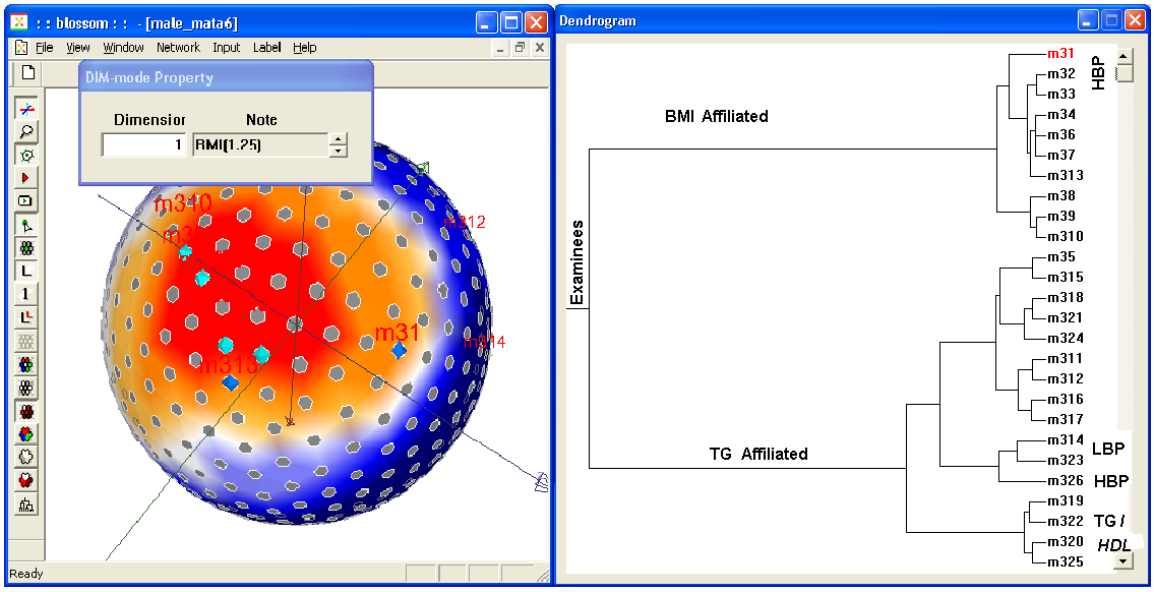


Figure 5- 27: Male 31-40 examinees' component map and dendrogram

31-40 years cluster

Major parameters affecting this group of examinees are BMI and TG. HBP, LBP and HDL are influencing the degree of metabolic to the affected members though in low traces.

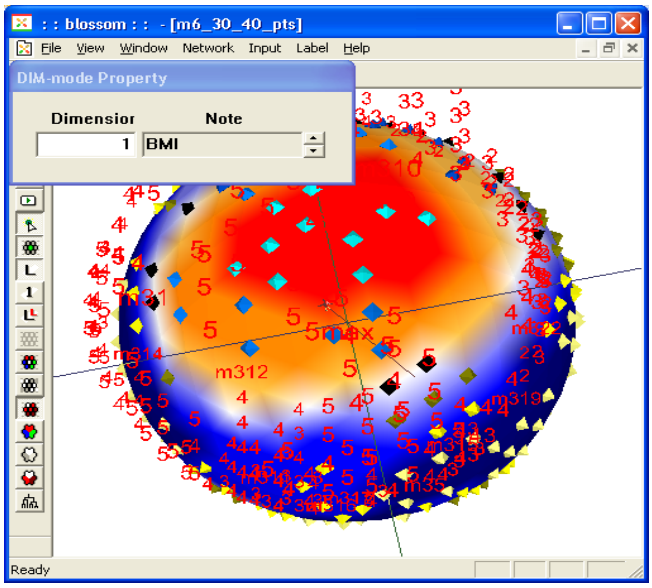


Figure 5- 28: **Male 31-40 years mark-points metabolic SOM**

Figure 5-28 gives a percentage representation of metabolic spherical SOM of the examinees. The same members are considered in figure 55 with a difference of notation. The percentage map has members like m**3**12 with reference '**3**' for cluster range 31-40 years.

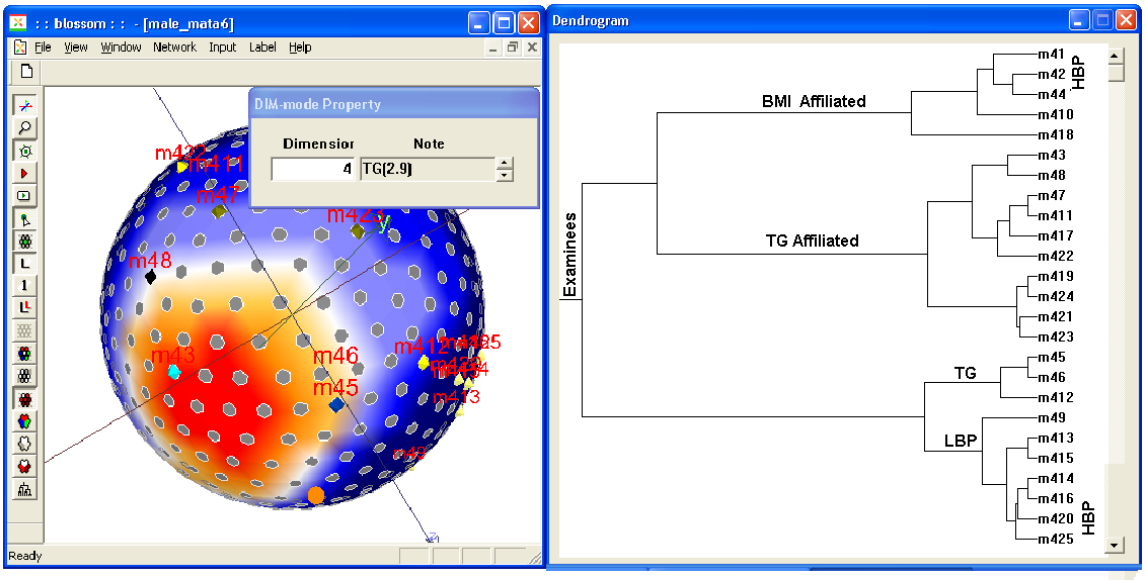


Figure 5- 29: 41-50 Spherical component map and dendrogram

41-50 years cluster

Referring to figure 5-29, the age cluster shows the examinees mainly affected by parameters TG, BMI and LBP. HBP is observed to be affecting fewer members. It is also noted that some members were affected by a combination of LBP and HBP.

51-60 years cluster

This cluster as shown in figure 5-30 has a combination of all the parameters affecting the members. There are two main branches namely BMI and HBP /LBP/TG. Blood pressure parameters have a larger share in its effect on the examinees with some members having a combination of the blood pressures.

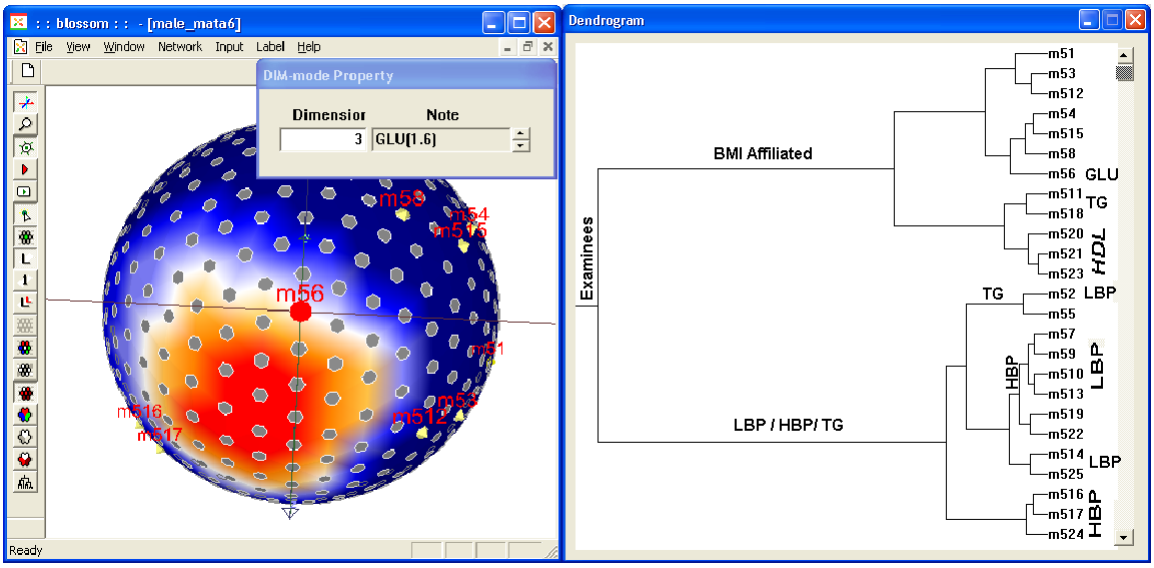


Figure 5- 30: Male 51-60 spherical component and dendrogram

61-70 and 71-80 clusters

The two clusters have a large population of examinees. The sampling would not adequately represent the entire population and as a result each of the clusters is trained using "blossom" software. Figure 5-31 gives the male 61-70 years of age component map as well as a dendrogram chart. It can be observed from the figure that BMI and HBP affect a higher population of examinees than the other parameters. TG is also noted to be the third in hierarchy leaving the other parameters to have very low impact though not to be ignored, to the health of the examinees. Figure 5-32 displays the 71-80 years cluster for the male examinees. Metabolic syndrome risk is mainly dominated by BMI and HBP. The other parameters having traces of effects to the examinees are TG and LBP, but at low risk values. Figure 5-33 gives metabolic syndrome highest risk zone of the cluster members (71-80 years). It is observed that the examinees under risk have TG and HDL parameters quite high and some members with BMI parameter.

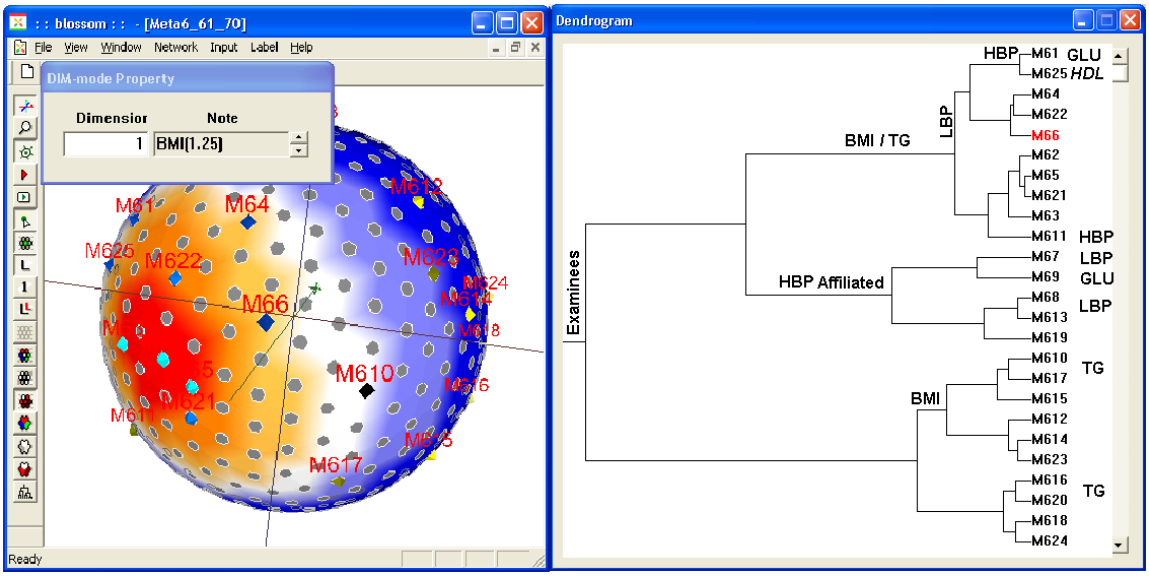


Figure 5- 31: Male 61-70 spherical component map and dendrogram

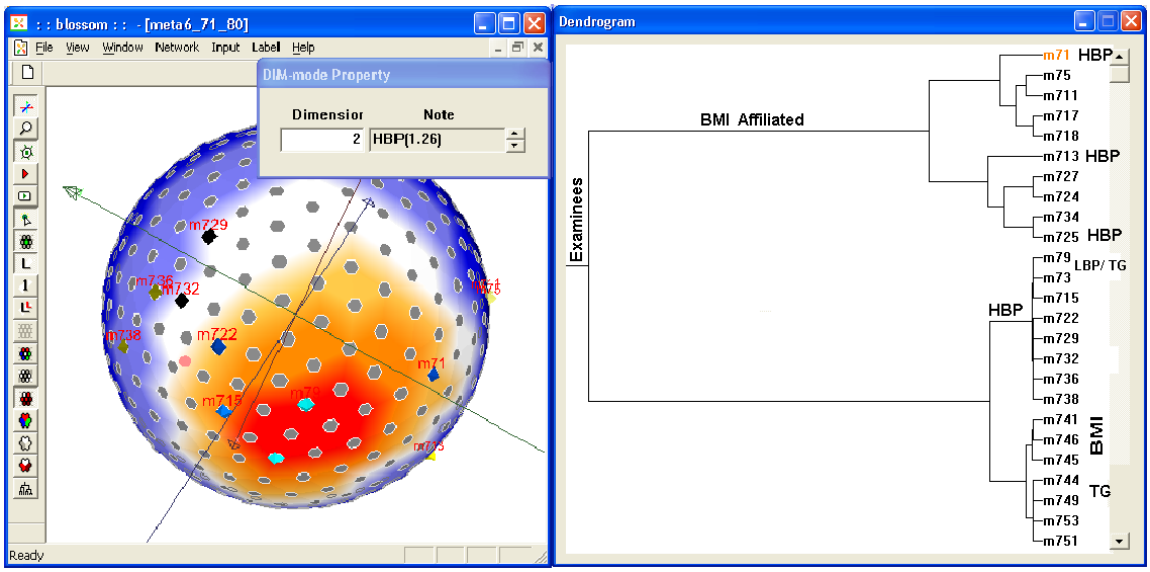


Figure 5- 32: Male 71- 80 spherical component map and Dendrogram

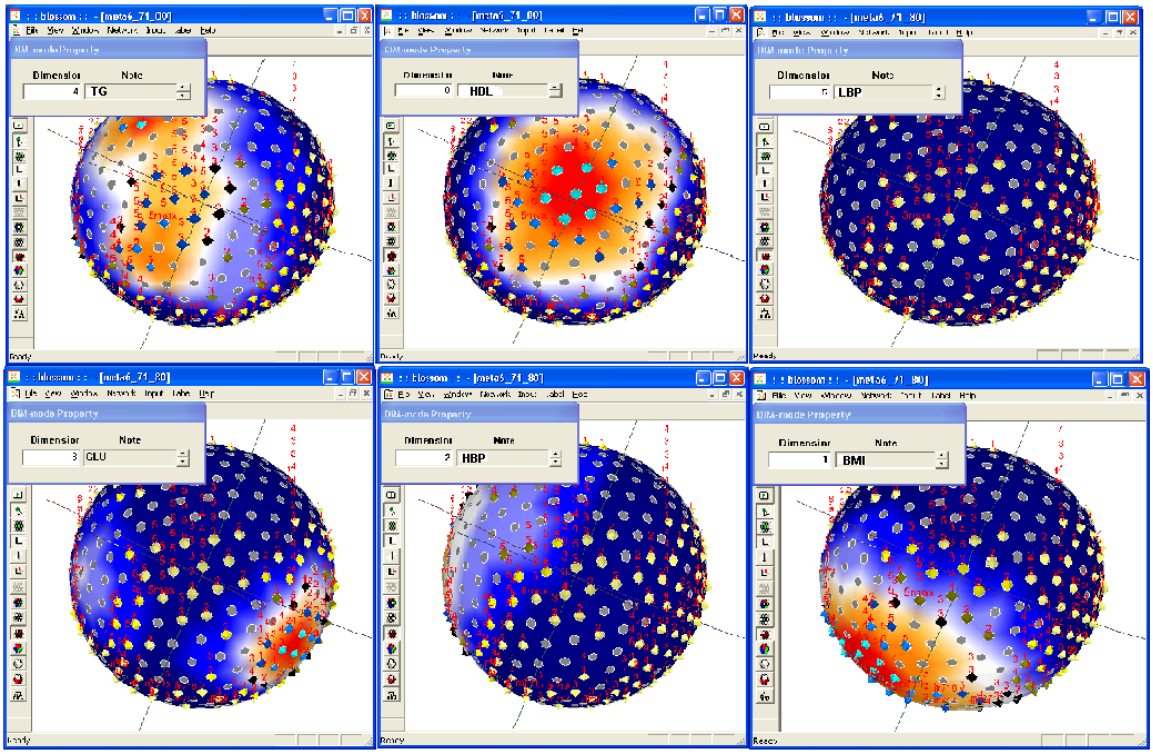


Figure 5- 33: Male 71-80 metabolic points

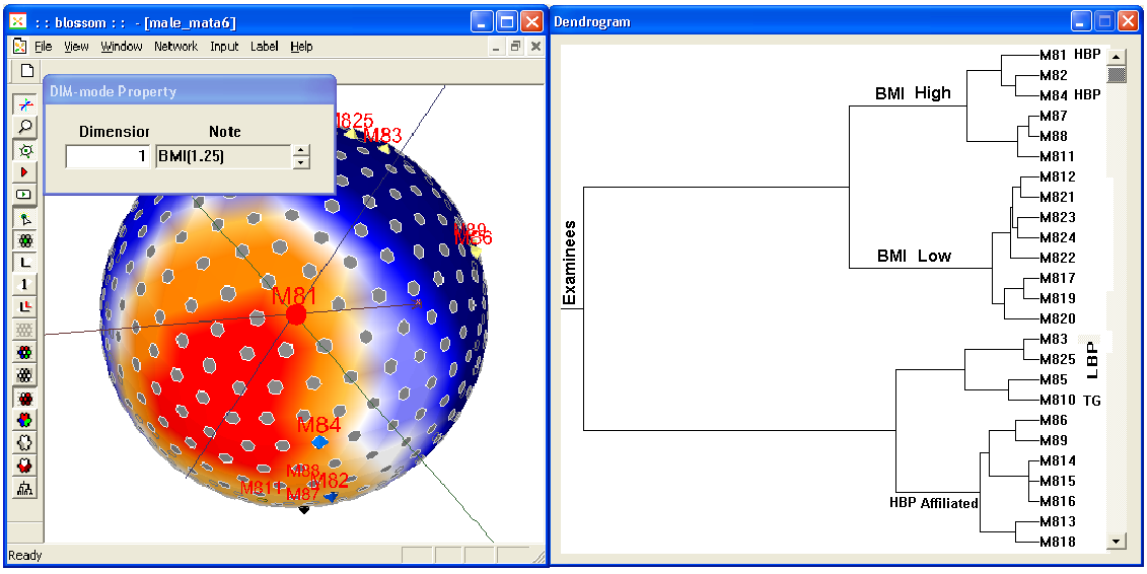


Figure 5- 34: Male 81- 93 spherical component map and Dendrogram

81-93 cluster

The cluster has BMI and HBP parameters affecting the sampled members as shown in figure 5-34. The sample being a good representation of the examinees indicates that high blood pressure is a major contributor of metabolic syndrome for this age group.

5.62 Female clusters

Figures 5-35, 5-36, 5-37 and 5-38 display the trends and metabolic percentage mark points used for the analysis of female examinees metabolic syndrome. Similar approaches as that of the male examinees have been used to derive the maps.

17-30 years cluster

Figure 5-35 segment 'A' displays the trends of the cluster examinees. The main parameter affecting the members is BMI. TG and HDL affect some few members.

30-40 years cluster

Figure 5-35 segment 'B' shows BMI being the major parameter followed by TG. HBP / LBP and HDL have a contribution towards the metabolic risks.

41-50 years cluster

Figure 5-35 segment 'C' gives a representation of the age cluster of examinees' parameters trends. The main parameters affecting the members in this cluster are TG and BMI. The cluster members have a more complicated combination of the analyzed parameters with BMI taking a bigger share of the combination. It is worth noting that the TG tree bares more metabolic affected patients than that of BMI. Almost all the sampled members are metabolic (MK 2- 5max) except members, f48 and f418.

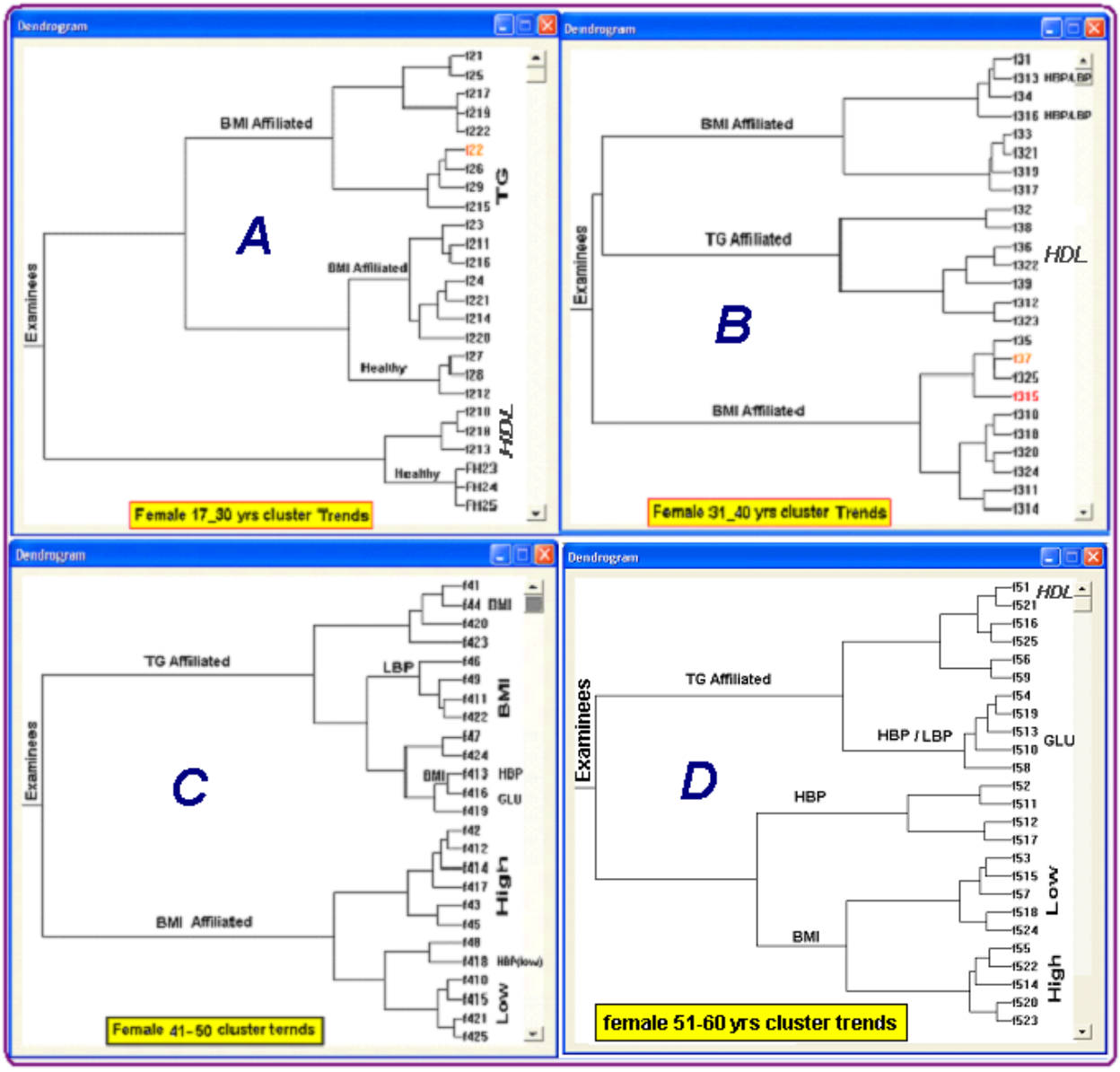


Figure 5- 35: Age clusters examinees' parameter trends

51-60 years cluster

Figure 5-35 segment 'D' shows the parameter trends for the examinees. The main parameters affecting these cluster members are TG, HBP and BMI. GLU and HDL are affecting isolated cases. The sampled members are all metabolic (MK 2- 5max).

61-70 years cluster

The population of the examined members was large (735). This prompted the training of this group of members independently. Using the trained map a sample was taken equal to the other cluster's samples. Figure 5-36 segment 'E' gives the pattern the monitored health parameters are taking. The cluster has HBP, TG and then GLU as the major contributors to the metabolic risks. BMI affects isolated cases as is HDL health parameters.

71-80 years cluster

The age cluster was trained as that of 61-70 years due to its large population of the examinees. The cluster health parameters trends are shown in figure 5-36 segment 'F'. Parameters affecting this cluster can be prioritized as TG, HBP and BMI. GLU parameter is affecting a big population of the sampled members.

81-94 years cluster

Figure 5-36, segment 'G' shows the health parameters trends of the age cluster. The members in this cluster had HBP as the main parameter affecting the sampled members. Some examinees had complications as some disease related parameters like HBP and LBP had a combined affect on them.

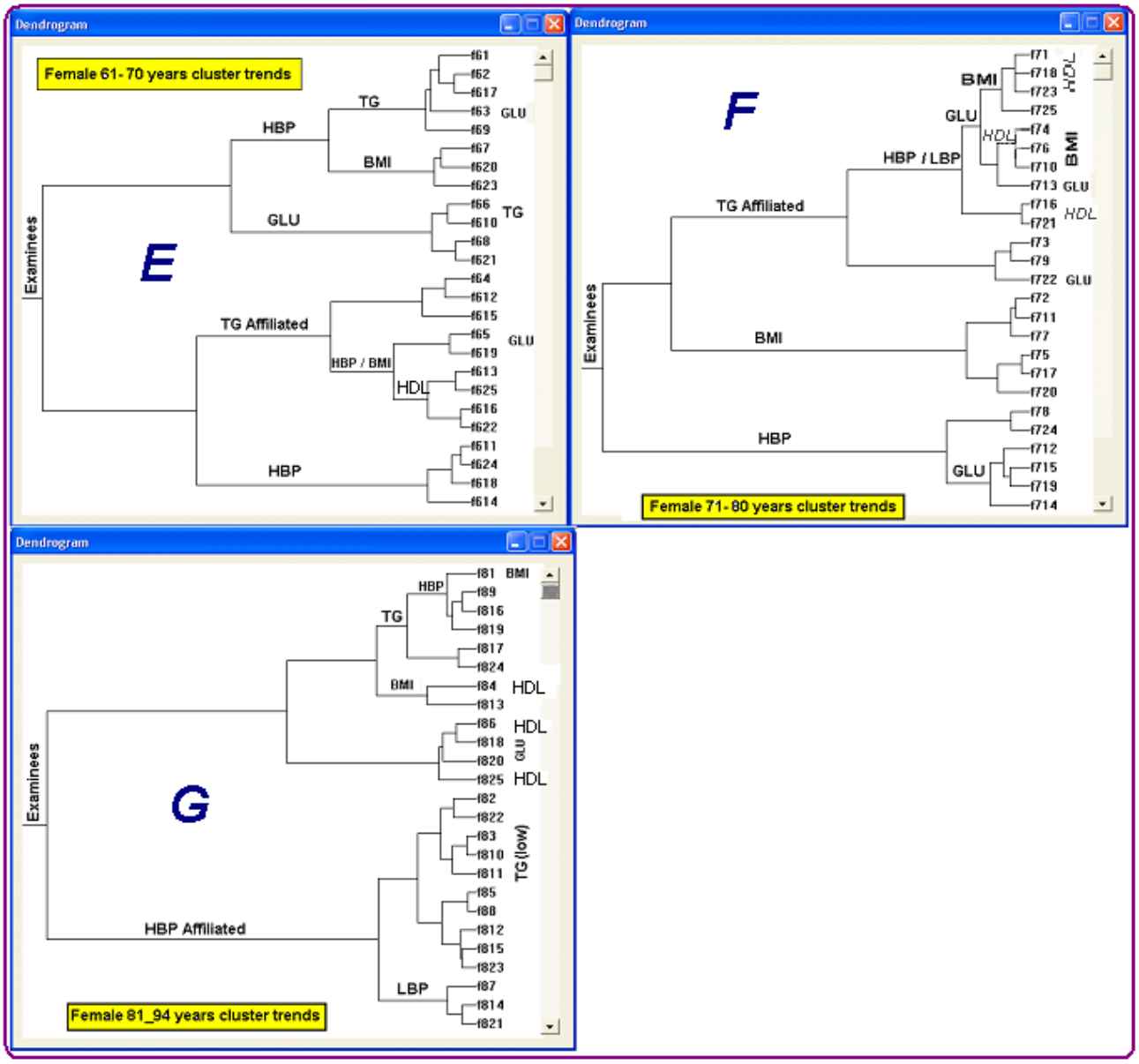


Figure 5- 36: Age clusters examinees' parameter trends

SSOM Metabolic Mark-points

Figure 5-37 and 5-38 show the mark-points of the various age clusters. Spherical component maps are used except segment 'M' that gives a U-Matrix representation of the mark-points.

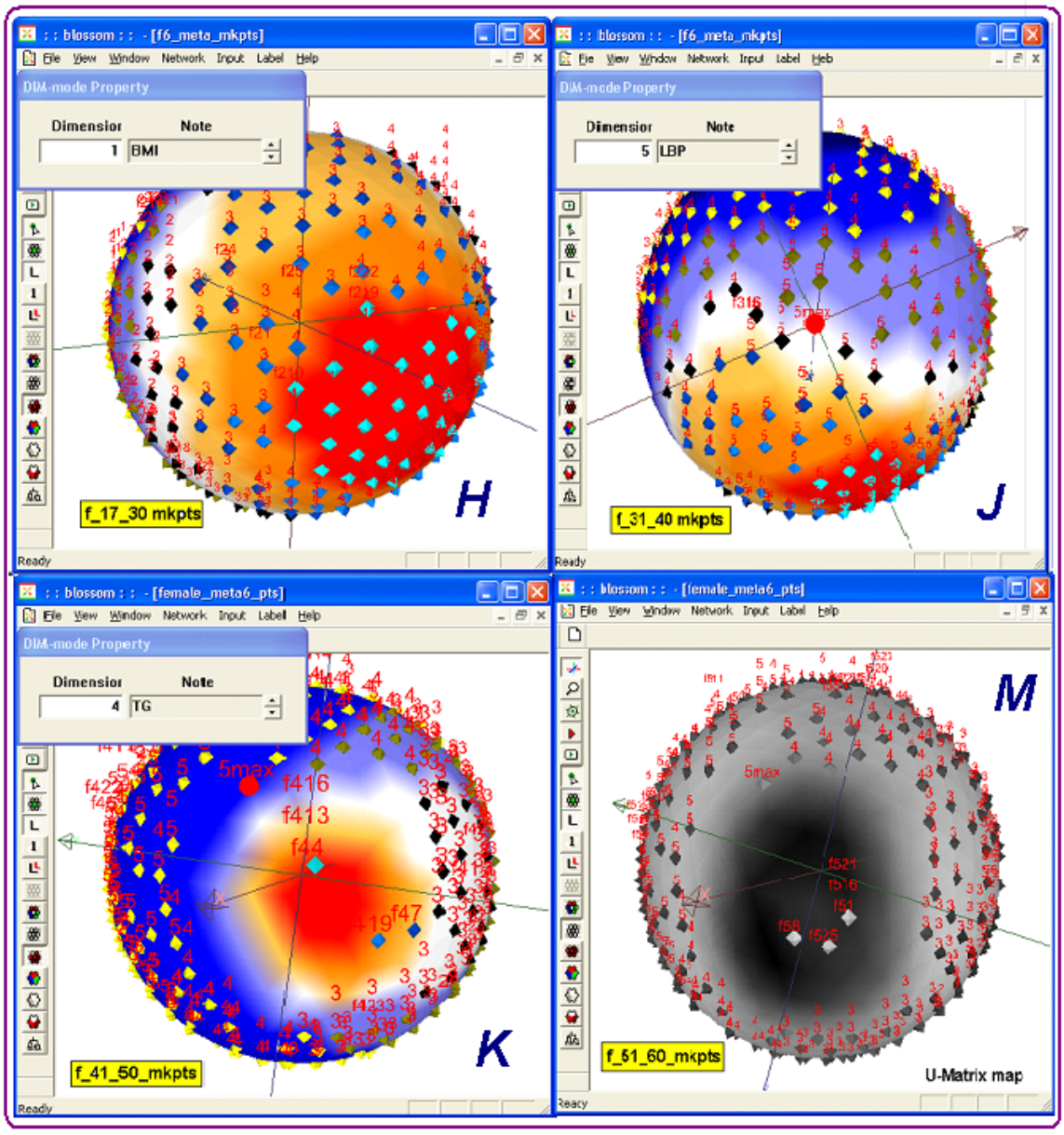


Figure 5- 37: 17- 60 years cluster metabolic mark-points SSOM maps

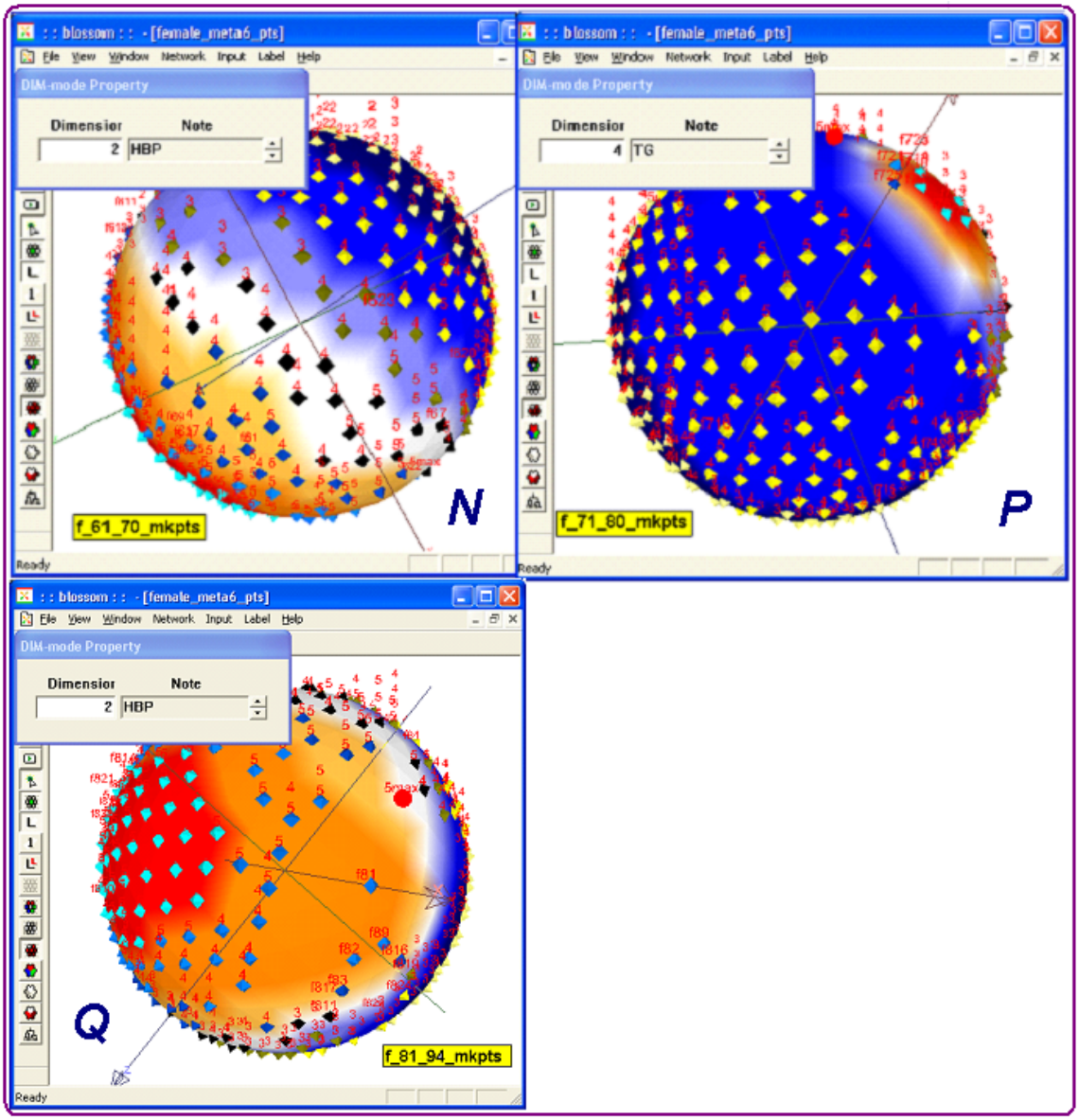


Figure 5- 38: 61- 94 years cluster metabolic mark-points SSOM maps

5.7 Metabolic syndrome points visualization Tools

In the process of developing metabolic syndrome visualization tools, the examinees' data underwent training using SOM algorithms. The data was then analyzed with the help of physicians so that a professional input as well as guidelines could be injected to the overall conclusion of the intended metabolic tools [62, 63, and 64]. Having the results

of the analysis of metabolic syndrome using SOM trainers concur with the diagnosis of the syndrome, the development of the visualization tools were given a nod by medical experts. It was noted that the visualization tool would go along way in helping the physicians decode the health behavior / patterns of their patients. Physicians as well as the patients can use the charts to visualize the degree of health. Physicians can as well predict the consequences well in advance and hence save the member from being affected. Analysis of this syndrome through SOM symbolizes the importance of getting more information of any patient and including him or her in the decision making process thereby enjoining him / her in the healing process.

Figure 5-39 and 5-40 shows a visualization approach to the syndrome where examinee information is displayed as well as the metabolic points.

Figures 5-40 to 5-42 show four parameters (BMI, HBP, GLU and TG) checkup tools developed from the results of the trained data for male and the female examinees. Referring to figure 5-39, an examinee parameter values are first entered. The tool matches the individual's parameters to those of the trained SOM and then it displays the cluster the member joined. In the example given, the examinee joins the highest risk zone (red). Displayed also is the risk in terms of percentage points, age, year of check-up and name of examinee. Clear color is taken as normal members.

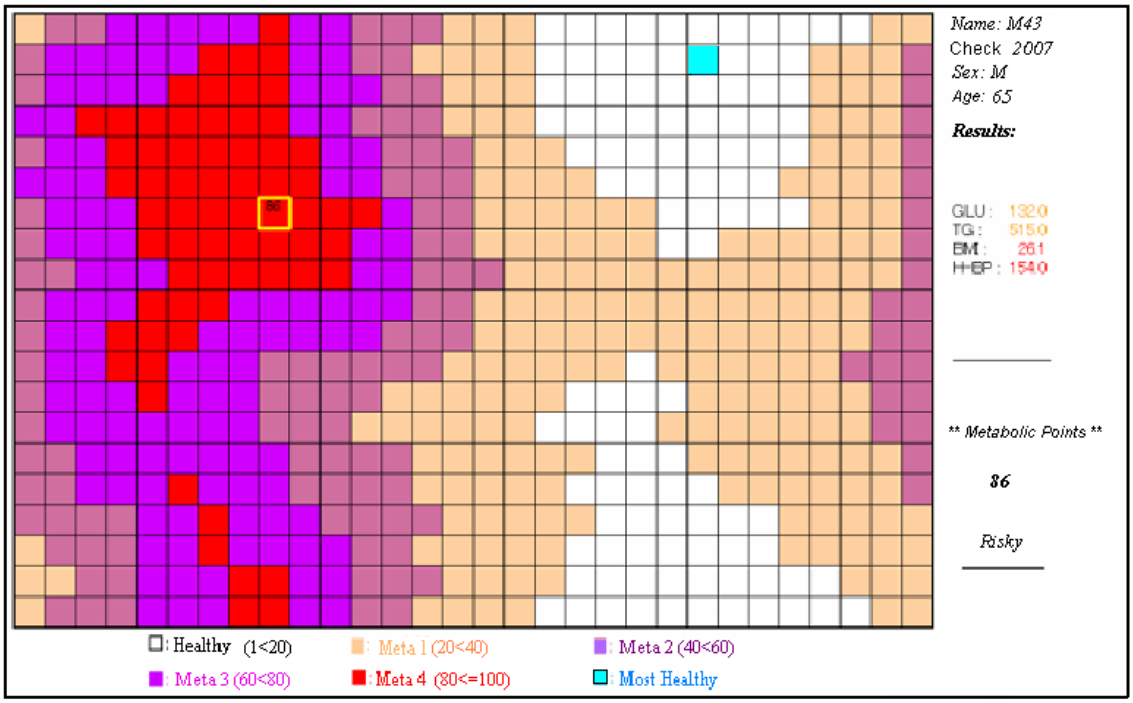


Figure 5- 39: Male metabolic visualization tool

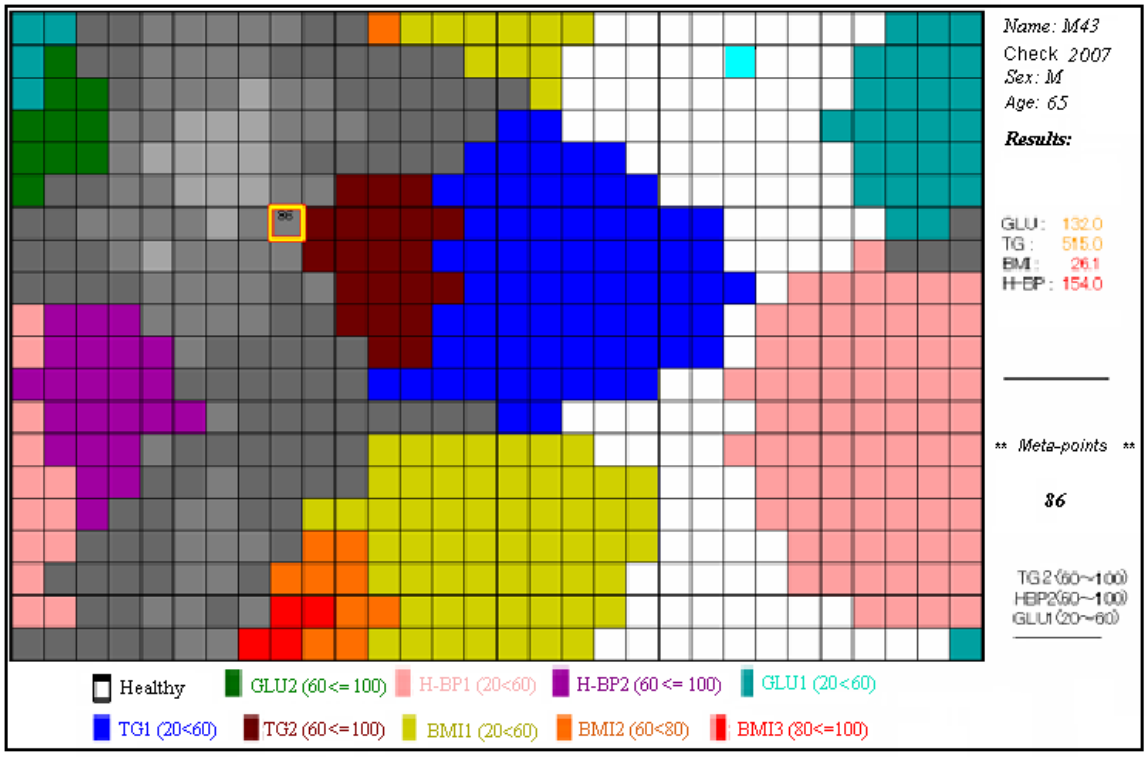


Figure 5- 40: Male metabolic component map

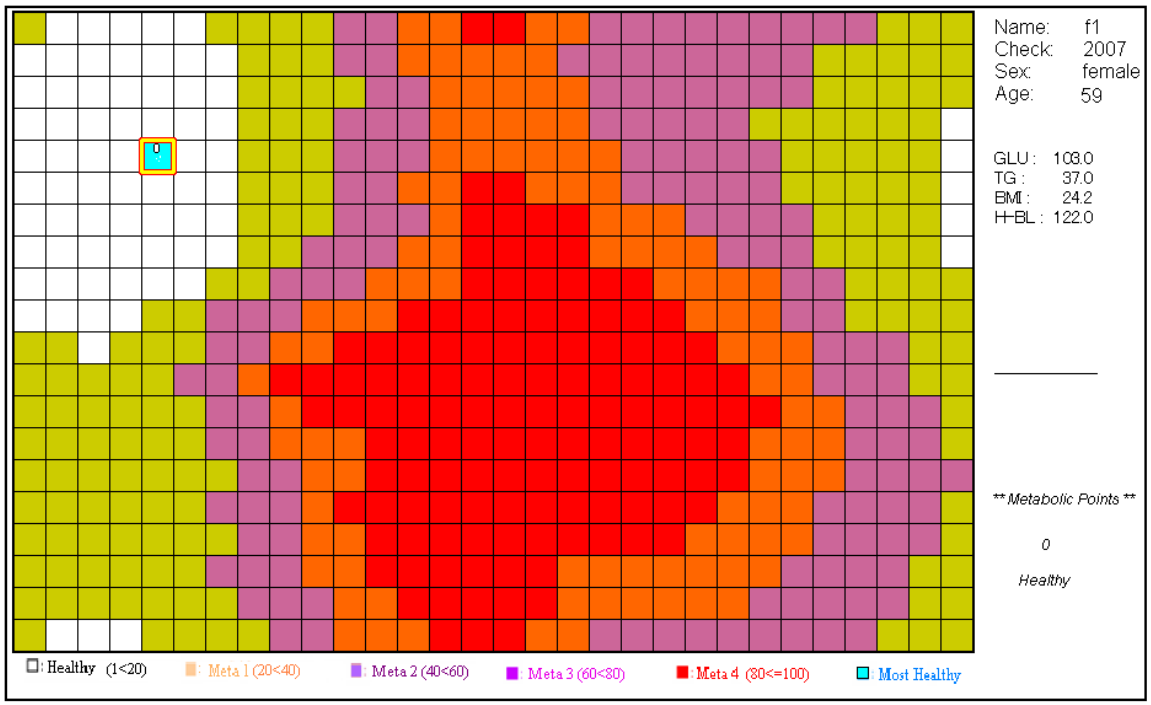


Figure 5- 41: Female metabolic visualization tool

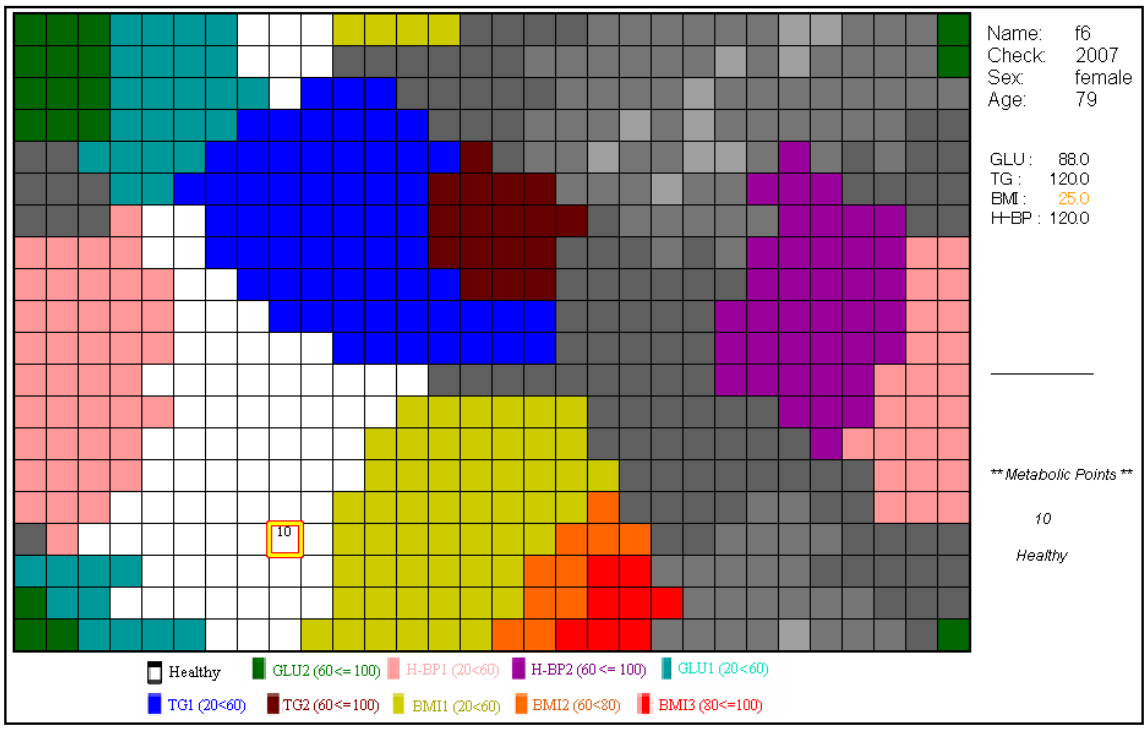


Figure 5- 42: Female Metabolic component map

Figures 5-40 and 5-42 display male and female component maps derived from SOM trained data. The interrelationship of the principal components is evident. These types of maps help the observer to visualize the most probable risk factor or factors. Displayed items show the contribution each of the four parameters has to the syndrome. Cases are there where an individual parameter may be the due cause of the metabolic.

5.8 Metabolic syndrome risk prediction tools

The Metabolic evaluation tools may not be complete without introducing a prediction option where patients health data can be stored in a database and be referred to anytime the physician deems necessary. The database can as well be used to predict metabolic status based on the trend the patient's health has been fluctuating.

Construction of the SOM can be described as follows: The value in the next year after the last year can be estimated by inserting the checkup data of all years. Take an example, there are 3 years data of $M(t_1), M(t_2), and\ M(t_3)$ and then, the following year $M(t_4)$ can be estimated by equation (5.5) [65].

$$M(t_4) = M(t_3) + \frac{M(t_2) - M(t_1)}{t_2 - t_1} + \frac{M(t_3) - M(t_2)}{t_3 - t_2} \tag{5.5}$$

The SOM display also includes all the medical examination data of the examinee at the predicted status. How far away the patient is from health status can be given as a shortest distance like Traveling Salesman Problem (TSP) method. The results are displayed and the state of the movement of the data of the examinee can be easily

understood on inspection (see the results of figure. 5-43). Also, if an item especially with a bad value from the blood test is to be returned to the normal value, the position would move in the health area on the SOM map.

Even though some of the displayed figures are shown by red characters indicating bypassing normal values, 100 point marks are shown in the map. This is the result where '100' is displayed in the case of the point mark equal to or more than 95 (rounding, after being computed by equation (5.4)).

Metabolic data used to draw the male examinee' syndrome trends is retrieved from his data bank. It is evident that the physician can monitor, predict the metabolic pattern and hence have a second opinion to his analysis. Some examinees may show deterioration while others improvement. The client is also able to visualize the degree of risk he/she may be in.

The examinee can also observe the projected patterns and hence be ready to accommodate the suggestions laid down. Referring to random sampled examinee's data shown in figure 5-43, the examinee has deteriorated due to HPB and TG. Figure 5-44 refers to female examinee whose syndrome has eased giving one a sigh of relief. The predicted value of metabolic is zero (0); giving an indication that the examinee would be the healthiest if the trends followed currently are maintained.

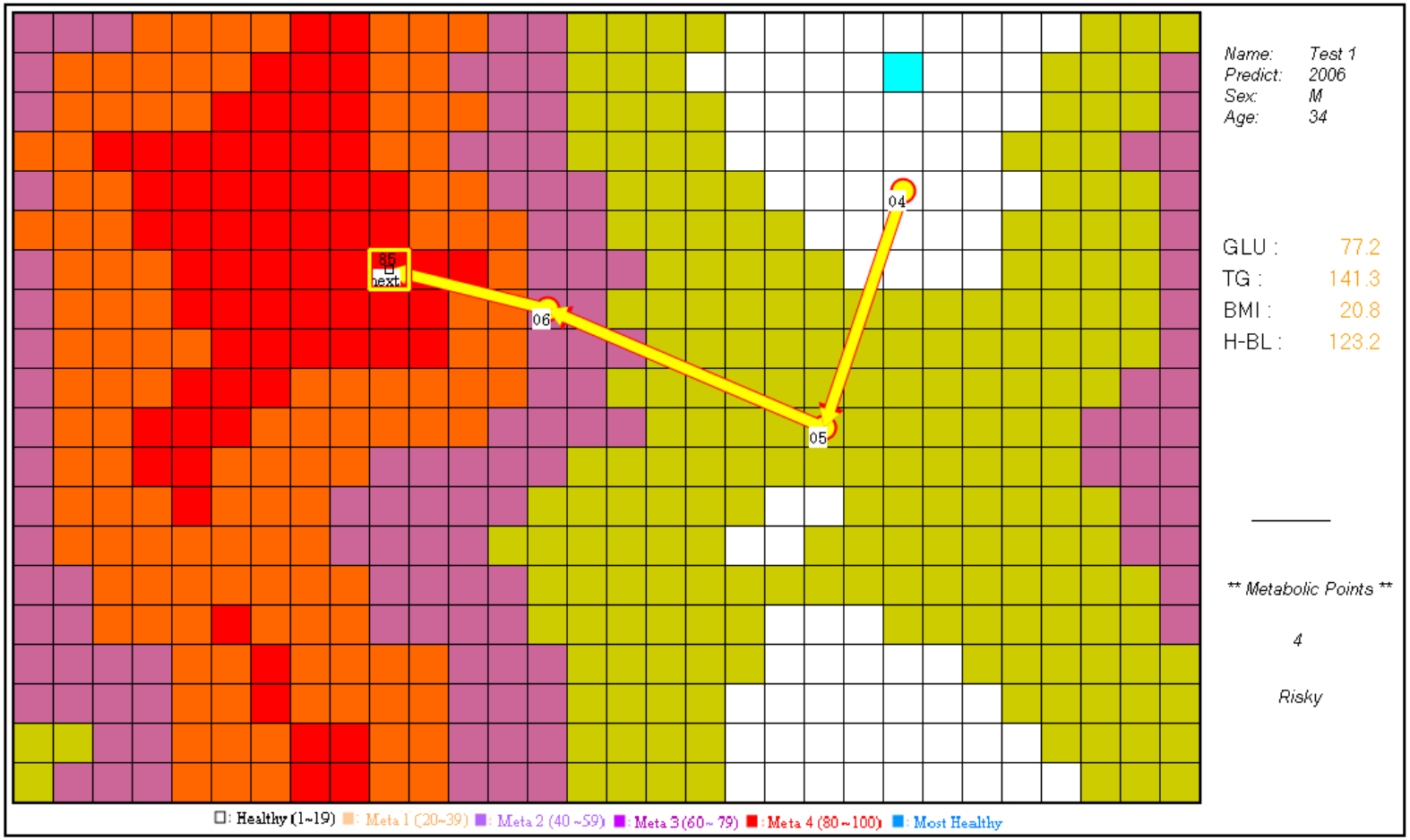


Figure 5- 43: Male Metabolic Syndrome Individual Prediction

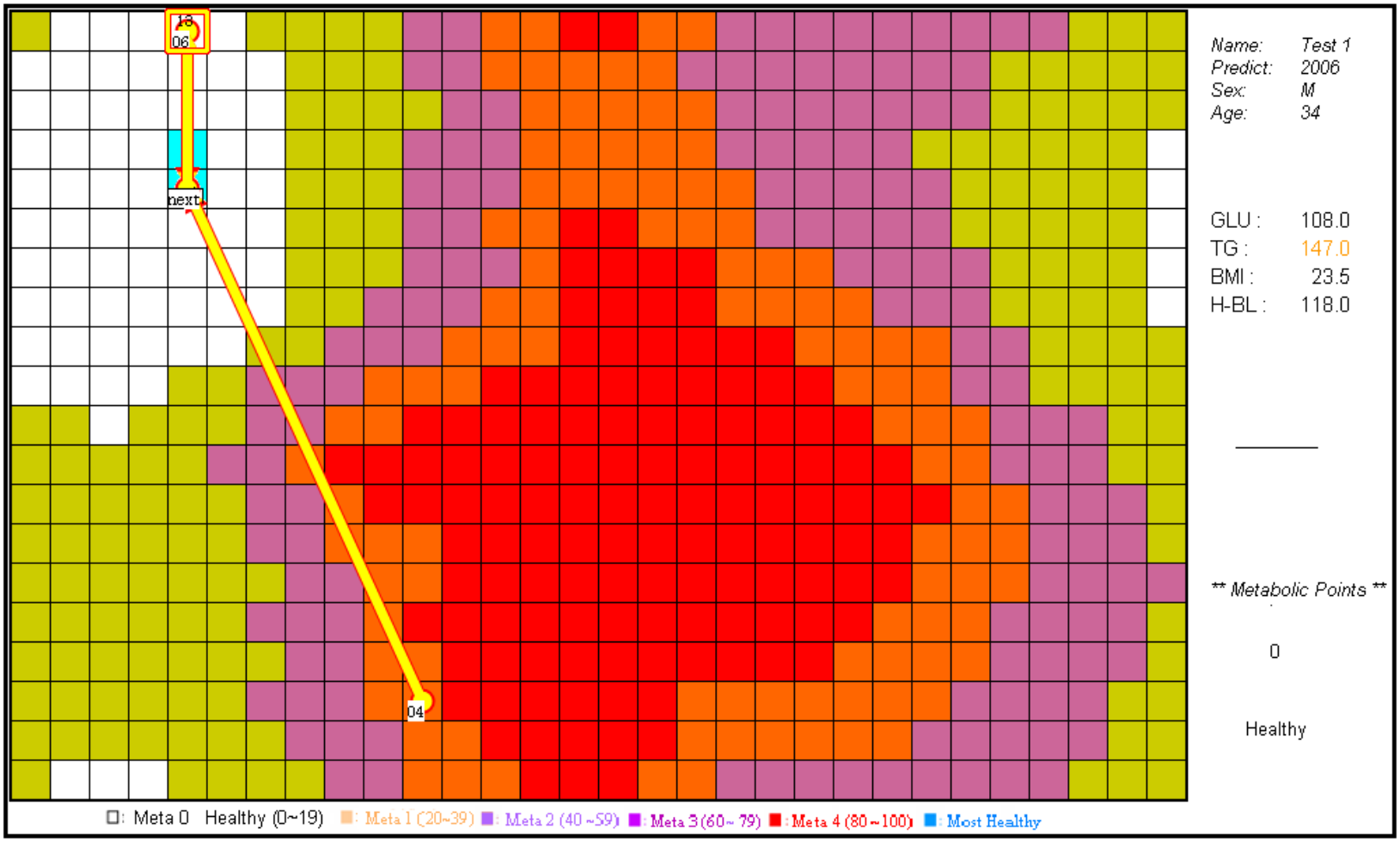


Figure 5- 44: Female Metabolic prediction

5.9 Analysis and Interpretation of overall Results

Male:

- **1546** out of **2450** examined members were in the metabolic bracket (**63% of the total**)
- Grouping the examined population into age clusters as shown in figures 5-24 to 5-38 pp 133-145 has evoked different metabolic trends with youth mainly affected by BMI while the aged have HBP parameter.
- Generally the health parameters affecting the male population can be grouped from the most prevalent to the least as TG, HBP and then BMI with respect to contribution to metabolic syndrome (refer figures 5-14 and 5-15 pp 121-122) .
- Referring to figure 5-15, HDL parameter stands more independent but has relation to TG and HBP.
- Blood sugar (GLU) has lesser threat to male examinees though a disease needing care to the affected few.
- Examinees in the age bracket 19 to 40 as shown in figure 5-25 and 5-27 pp 133 and 134 have two major parameters affecting their health and hence joining the metabolic risk zones namely BMI and TG. Figures 5-29 to 5-34 show the HBP parameter affecting most members from 40 years and above. More noticeable is the combinations of parameters affecting the health of age groups 50 to 80 years. Diseases become more irresistible and some members belonging to these clusters being more prone to more than one disease.

- Referring to figure 5-34 pp139, the age cluster 81 to 93 seem to be affected more by HBP parameter.
- LBP is also a threat to members of age clusters 41 to 70 years. Some situations as shown in figures 5-29 and 5-30 pp 136 and 137 arise where both blood pressures affect the members.

Female:

- Out of **4007** examinees checked for metabolic syndrome **1935** appeared on the metabolic bracket (**48% of the total**)
- Referring to figure 5-35 A pp 141, examinees between 17 and 40 years of age have BMI as the major parameter affecting their health enjoining them to metabolic risk zones.
- Examinees in the age bracket 41 to 60 as shown in figure 5-35 pp 141 have TG parameter seeming as the route cause of the various ailments affecting them. Members between 61 to 80 years seem to have a compounded effect of TG where disease related parameters are seen to be in combination with TG.
- HBP parameter affects mainly examinees of age bracket 61 to 94 years.
- The examinees with high BMI factors seem to have fewer complications to the other parameters as evidenced by the by figure 5-21 pp 130.

Chapter Summary

In this chapter the construction of metabolic syndrome SOM maps from the examinees health checkup data are developed. U-matrix SOM maps helped to display the examinees in form of zones or clusters. My colored U matrix as well as enhanced U-matrix representation of the trained data gave a better presentation of the syndrome. Metabolic risk zones in terms of color coding and percentages are also considered in the displays presented. Colored component maps are developed to aid in identifying the most probable cause (s) of the syndrome.

Analysis of the syndrome based on the output of the trained data displayed in the various formats is done with an intension of developing the metabolic trends. Spherical SOM generates U-matrix, component maps as well as dendrogram based on the input variables considered and the sampled population of the data. This development helps in observing the various trends the trained variables may be taking.

The analysis and development of visualization displays of the trained data and in conjunction with the medical experts, visualization tools that could help physicians in their quest of diagnosing the syndrome were developed.

Age clustering as a method of decoding the syndrome is considered using spherical SOM trainer. Two dominant input parameters persist namely BMI for the examinees below 40 years and HBP for those above 40 years of age.

CHAPTER 6

CONCLUSIONS AND RECOMMENDATIONS

6.1 Introduction

Analysis and visualization of metabolic syndrome using SOM has been realized. The research was divided into two groups namely male and female for an effective analysis and visualization of the syndrome. Screening of healthy examinees was done so as to work with examinees having traces of the metabolic risk. Few healthy members were added after screening to allow for a healthy distribution curve with all members represented. Four parameters were initially selected for investigation. This was necessitated by global definition of the syndrome that anybody with at least three of the medically accepted parameters is said to have metabolic syndrome. Medical experts itemize the six parameters that were further investigated as the parameters needed to be monitored when diagnosing metabolic syndrome namely BMI, HBP, GLU, TG, LBP and HDL. The tools used were Torus and Spherical SOM, Excel and Visual C++ software for the analysis, visualization and development of metabolic syndrome visualization tools.

6.2 Conclusions

The SOM tools gave trained data that could be visualized using gsview32 and Som_Viewer even though Torus and Spherical SOM were complete with their viewing software. The SOM maps obtained demonstrated how metabolic syndrome parameters

clustered themselves, the bonding and barriers between clusters, and the trends the metabolic syndrome was taking on the selected population.

The effects the considered parameters had on metabolic syndrome and the trends the syndrome took on various patients can be summarized as follows:

- **Female examinees** were mainly affected by BMI, HBP and TG with the most risky one being BMI. Blood glucose had less impact on the metabolic risk. One major observation was that the examinees with high BMI had fewer complications on the other considered parameters.
- **Male examinees** had the common zones between HBP and TG clusters and that between GLU and TG clusters being metabolic. It was observed that HBP was the major risk factor towards metabolic syndrome amongst the examinees. BMI was noted to have less impact to the metabolic risk trends.

The concluded analysis found out that the male syndrome was due to risk parameters in the order HBP, TG, BMI LBP, HDL and then GLU. The female counterparts had BMI, HBP, LBP, TG, HDL and then GLU. The overall syndrome trends were mainly due to TG, HBP, and BMI respectively.

Examinees age brackets cannot be overlooked in the syndrome investigation. It was observed that 19 to 40 years clusters had metabolic syndrome mainly due to BMI for females and BMI, TG for males. 40 to 60 age groups were mainly affected by TG for females and HBP for male counterparts. Examinees above 60 years had compounding links between the parameters with the females having their base risky parameter as TG while the male examinees had HBP. The most elderly (above 81 years) seemed to be

most affected by HBP. All examinees with BMI as the main risk factor seemed to have fewer complications to the other monitored parameters.

Referring to the above analysis, 1546 out of 2450 male examinees (63% of the total) tested were in the metabolic syndrome bracket with the majority initially occurring due to BMI then TG, and finally to HBP. The 4007 female examinees tested, 1935 were metabolic (48% of the total). The general female trend starts with BMI, then HBP and finalizes with HBP. It can thus be concluded that the major parameter affecting the examinees is HBP. BMI seems to be dormant parameter within the youth (below 40 years).

The analysis cited above could not be possible without the help of a simulator able to decode the complex input data into planer output that is easy to understand and visualize by both examinees and the medical experts. Having SOM displays and formulating component maps as additional display tools, I was able to analyze and visualize to details the contribution each input parameter had on the overall metabolic status of the examinees.

6.3 Contribution of this work

The research findings documented in this report were arrived at and compared well with those found by physicians. I can thus conclude that the information required when medical experts are diagnosing examinees health status may further need simulation tools to help in the analysis and visualization of the expected diagnostic results so that the affected members are given precise information about their health condition. With this type of visualization method, physicians can have an alternative approach to giving their

patients a sight of their health status and also their health trends. This in effect would make the patient share with the medical experts any health decision needed. The recovery period may as well be improved. Additionally, most of the affected members are in their prime age which is very vital to the well being of the economy of their respective countries.

Spherical SOM tools were used in this research to give us a second opinion on the expected results. The package has added features that helped us decode the effect age had on the metabolic risk, the trends metabolic syndrome was taking with regard to the monitored parameters and a vision on how the entire health was being encroached by each parameter. The user is also given a sense of feel as user scrolls around the entire metabolic syndrome globe.

Secondary effects based on psychological impact cannot be overlooked. Many examinees may have joined the risky zone due to the beliefs that go with the syndrome. It is important to point out that psychological effects should be addressed in conjunction with such visualization techniques for a more comprehensive diagnosis of the syndrome. The common beliefs that big belly members of our society carry the highest metabolic risk may need an alternative definition and approach so that the patients are given enough details of their health if possible.

Having used SOM tools for the analysis and visualization of the syndrome and obtaining results that compared well with those diagnosed by the physicians, it was found necessary to develop a metabolic evaluation tool using SOM trainers. The tool evaluates the examinee's health data and displays ones position on the metabolic map. The probable parameters causing the metabolic, degree of risk, name, and age also appear on

the same display. It is also possible to use the simulator to predict the syndrome trends over the years which become an added advantage to the doctors since access to a database will be automatic.

6.4 Recommendation for further research work

The analysis and visualization approach to metabolic syndrome using SOM techniques has revealed a lot of characteristic relationships between the considered parameters. One is able to relate and view the metabolic trends at a glance.

There are a lot of hidden details towards man's health that need to be decoded so that doctors can come up with better methods of diagnosing the various ailments affecting it. Further analysis of the syndrome need to introduce psychologically oriented symptoms like stress as an input parameter. The design of metabolic syndrome evaluation tool that decodes all the inputs parameters considered so far would be of much help to the physicians. The visualization tool should be a SOM map of choice (SOM or component map display) that is user friendly. The user can request to input required parameters if new or update the database. The tools will be able to seek proper identification before any alterations of the database is done. The tools will be designed to give easy to use instructions, interactions that doctors will happily embrace.

REFERENCES

[1] Allen Newell; McDermott; Forgy, “Artificial Intelligence: A Self Paced Introductory Course", September 1977

[2] The age of intelligent machines: http://mitpress.mit.edu/catalog/item/default.asp

[3] Bose; Bimal K., “Modern Power Electronics and AC drives”, (2002), Prentice Hall.

[4] Kazuo Tanaka, “An Introduction to Fuzzy logic for Practical Applications”. (1997), Springer New York.

[5] Riccardo Poli, “On Fitness Proportionate Selection and the Schema Theorem in the presence of stochastic effects”:

http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.36.6731

[6] Genetic Algorithm: www.cseiq.org/pdf_files/1230102614_Genetic.pdf

[7] T. Kohonen, “Self-Organizing Maps”, Springer series in information Sciences, Vol. 30, (2001), Reading Mass.

[8] Metabolic Syndrome:
http://www.mayoclinic.com/health/metabolic%20syndrome/DS00522

[9] Metabolic syndrome: http://www.aafp.org/afp/20040615/2875.html

[10] American Heart Association:

http://www.americanheart.org/presenter.jhtml?identifier=4756

[11] Health and metabolic syndrome:

http://www.medicinenet.com/metabolic_syndrome/article.htm

[12] P.K. Kihato; H. Tokutaka; M. Ohkita; K. Fujimura; K. Kotani; Y. Kurozawa; Y. Maniwa, "Analysis and Visualization of Metabolic Syndrome Trends using SOM", Biomedical Soft Computing and Human Sciences, Vol. 13, No. 2, pp. 89-95, 2008.

[13] P.K. Kihato; H. Tokutaka; M. Ohkita; K. Fujimura; K. Kotani; Y. Kurozawa;and Y. Maniwa, "Spherical and Torus SOM approaches to Metabolic Syndrome Evaluation", ICONIP,2007, Part II, LNCS, 4985, pp. 274-284, 2008

[14] P.K. Kihato; H. Tokutaka; M. Ohkita; K. Fujimura; K. Kotani; Y. Kurozawa;and Y. Maniwa, "Focus on Metabolic Syndrome Trends using SOM", Journal of Japanese society for Fuzzy Theory and Intelligent informatics, Vol. 19, No. 6, pp. 634-641, 2007

[15] H. Tokutaka; Y. Maniwa; P.K. Kihato; K. Fujimura; M. Ohkita, "Application of SOM in a Health Evaluation System", Workshop on Self Organizing Maps in Bielefeld Germany, 2007

[16] Per Brodal, "Central Nervous System", (1992), Oxford University Press.

[17] J. L. Mcgaugh; N.M. Weinberger; G. Lynch, "Brain Organization and Memory: Cells, Systems, and Circuits", (1990), Oxford University Press.

[18] Gary A. Rosenberg, "Brain Fluids and Metabolisms", (1990), Oxford University Press.

[19] H.S. Bachelard, "Brain Biochemistry", (1984), Chapman and Hall, London.

[20] http://serendip.brynmawr.edu/bb/kinser/Structure1.html#cerebrum

[21] Neuron Action Potential: http://en.wikipedia.org/wiki/Action_potential

[22] Synaptic junction: http://www.flyfishingdevon.co.uk/salmon/year1/neurotr.htm#synapse

[23] Phil Picton, “Introduction to Neural Networks”, (1994), Macmillan publishers.

[24] L. Fausett, “Fundamentals of Neural Networks Architecture, Algorithms, and Applications”, (1994), Prentice Hall.

[25] J. A. Freeman, D. M. Skapura, “Neural Networks algorithms, Applications, and Programming Techniques”, (1991), Addison-Wesley.

[26] Perceptrons: An Associative Learning Network: http://www.ei.cs.vt.edu/~history/Perceptrons.Estebon.html

[27] G.F. page; J.B. Gomm; D. Williams, “Application of Neural Networks to Modelling and Control”, (1993), Chapman and Hall.

[28] Dayhoff; Judith E, “Neural Network Architectures: An introduction”, Van Nostrand Reinhold (Australia), (1990).

[29] Ultsch, A.; Korus, D, “Integration of neural networks with knowledge-based systems”, Proc. IEEE Int. Conf. on NN, vol.4, pp 1828-1833, 1995.

[30] Xiao-Feng Gu; Lin Liu; Jian-Ping Li; Yuan-Yuan Huang; Jie Lin, “Data Classification based on Artificial Neural Networks”, Int. IEEE Conf. on Apperceiving Computing and Intelligence Analysis, pp 223 – 226, 2008,

[31] Chang-Shian Chen; Chao-Chung Yang; Chin-Hui Liu, “The Interval Estimation of Parameters for Back-Propagation Network to Flood Discharge Forecasting”, Int. IEEE Conf. on NN, pp 3729 – 3735, 2006.

[32] Back propagation Network: http://www.itee.uq.edu.au/~cogs2010/cmc/chapters/BackProp/index2.html#Background

[33] Self-Organizing Map (SOM): http://www.cis.hut.fi/jhollmen/dippa/node9.html

[34] H. Ritter, “Self-Organizing Maps on non-Euclidean Spaces”, Kohonen Maps, Editors, E. Oja, and S. Kaski, Elsevier, 1999, pp.95-110

[35] Ultsch, A., "U*-Matrix: a Tool to visualize Clusters in high dimensional Data", Technical Report No. 36, Dept. of Mathematics and Computer Science, University of Marburg, Germany, (2003)

[36] Learning Vector Quantization: www.cis.hut.fi/panus/papers/dtwsom.pdf

[37] A. Flexer, "On the use of self-organizing maps for clustering and visualization", Intelligent Data Analysis, Vol. 5, PP 373-384, 2001

[38] Ultsch, A., "Data mining and Knowledge Discovery with Emergent Self-Organizing Feature Maps for Multivariate Time", Kohonen Maps, (1999), pp. 33-46

[39] Adrian A. Hopgood, "Knowledge-Based Systems for Engineers and Scientists", (1993), CRC Press.

[40] U-Matrix generation: http://www.cis.hut.fi/jhollmen/dippa/node23.html

[41] Chi Kin Chow; Shiu Yin Yuen, "Signal Self Organizing Map", Proc. of Int. IEEE Conf. on NN, pp 213 – 218, 2007

[42] K. Obermayer and T.J. Sejnowski, "Self Organizing Map Formation Foundations of Neural Computation", (2001), MIT Press.

[43] Kumar, D.; Rai, C.S.; Kumar, S.; "Face Recognition using Self-Organizing Map and Principal Component Analysis", Int. IEEE Conf. on NN and Brain, vol. 3, pp 1469 – 1473, 2005

[44] Ultsch, A., "Maps for the Visualization of high-dimensional Data Spaces", Int. Proc. Workshop on Self-Organizing Maps (WSOM 2003), Kyushu, Japan, (2003), pp. 225-230

[45] Neighbor nodes: http://davis.wpi.edu/~matt/courses/soms/#Determine%20Neighbors.

[46] Ultsch, A.; Halmans, G.; “Data normalization with self-organizing feature maps”, Int. Conf. on NN, vol. 1, pp 403-406, 1991.

[47] Benabdeslem, K.; Lebbah, M.; “Feature Selection for Self-Organizing Map”, Int. Conf. on Information Technology Interfaces, pp 45-50, 2007

[48] http://www.somj.com/

[49] D. Nakatsuka, M. Oyabu, “Application of Spherical SOM in Clustering”, Proceedings of Workshop on Self-Organizing Maps (WSOM'03), 2003, pp. 203-207

[50] Development kit: http://dkc3.digikey.com/PDF/T091/0628-0629.pdf

[51] J. Gámez Garcia, A. Robertson, J. Gómez Ortega and R. Johansson, “Force and acceleration sensor fusion for compliant robot motion control”, IEEE proceedings on Robotics and Automation, 2005, pp 2709-2714

[52] Abdel-Aty-Zohdy, H.S.; “Artificial neural network electronic nose for volatile organic compounds”, Proc. of Great Lakes Symposium on VLSI, pp 122-125, 1998

[53] Insulin resistance: http://www.medicinenet.com/insulin_resistance/article.htm

[54] Social physics: http://en.wikipedia.org/wiki/Adolphe_Quetelet

[55] Body Mass Index: http:/en.wikipedia.org/wiki/Body_mass_index#cite_note-1

[56] Blood pressure: http://www.yourtotalhealth.ivillage.com/blood-pressure

[57] Diabetes: http://en.wikipedia.org/wiki/Diabetes_mellitus

[58] Triglycerides: http://www.medic8.com/healthguide/articles/triglycerides.html

[59] Cholesterol Control: http://www.dietspotlight.com/cholesterol-control-without-dieting-review/

[60] H. Kurosawa, Y. Maniwa, K. Fujimura, H. Tokutaka and M. Ohkita, "Construction of checkup system by Self-Organizing Maps", Proceedings of workshop on Self-Organizing Maps (WSOM'03), 2003, pp. 144-149

[61] Umamaheswari, K.; Sumathi, S.; Sivanandam, S.N.; Anburajan, K.K.N.; "Efficient Finger Print Image Classification and Recognition using Neural Network Data Mining", International Conference on Signal Processing, Communications and Networking, pp 426-432, 2007

[62] M. Ohkita Research Team, "SOM viewer Software" 2007

[63] R. M. Jones, "Introduction to MFC Programming with Visual C++", (2000), Prentice Hall PTR.

[64] Max Fomitchev, "Enterprice application development with visual C++ 2005", (2007), Charles River Media.

[65] The Traveling Salesman Problem: http://www.hermetic.ch/misc/ts3/ts3demo.htm

PUBLICATIONS FROM THIS WORK

[1] P.K. Kihato, H. Tokutaka, M. Ohkita, K. Fujimura, K. Kotani, Y. Kurozawa, and Y. Maniwa, "Analysis and Visualization of Metabolic Syndrome Trends using SOM, Biomedical Soft Computing and Human Sciences, Vol. 13, No. 2, (2008), pp. 89-95

[2] P.K. Kihato, H. Tokutaka, M. Ohkita, K. Fujimura, K. Kotani, Y. Kurozawa, and Y. Maniwa "Spherical and Torus SOM approaches to Metabolic Syndrome Evaluation", ICONIP,2007, Part II, LNCS, 4985, pp. 274-284, 2008

[3] P.K. Kihato, H. Tokutaka, M. Ohkita, K. Fujimura, K. Kotani, Y. Kurozawa, and Y. Maniwa, "Focus on Metabolic Syndrome Trends using SOM", Journal of Japanese society for Fuzzy Theory and Intelligent informatics, Vol. 19, No. 6, pp. 634-641, Dec. 2007

[4] H. Tokutaka, Y. Maniwa, P.K. Kihato, K. Fujimura and M. Ohkita, "Application of SOM in a Health Evaluation System", Workshop on Self Organizing Maps in Bielefeld Germany, Sept. 2007

APPENDIXES

A- 1: HARDWARE FOR SOM

Many applications of Self-organizing feature Maps (SOMs) need a high performance hardware system in order to be efficient. Because of the regular and modular structure of SOMs, a hardware realization is obvious. Based on the idea of a massively parallel system, several chips have been designed, manufactured and tested by the authors.

A-1.1 Parallel processing operation

SOM-implementation consists of processing elements that are working in single instruction multiple data (SIMD) manner that are controlled by an external controller. Nearly all calculations are performed in parallel on all processing elements. A bidirectional bus is used for data transfers to dedicated elements and for broadcasting data to groups of processor elements (or to all processor elements). Single elements and groups of elements are addressed by row and column lines that are connected to the two-dimensional matrix. The externally generated instructions are transferred to the processor elements via an additional control bus. Two more signals are used for status messages to and from the controller. The architecture is able to simulate virtual maps, i.e. it is possible to simulate maps that are larger than the array of processor elements that is implemented. Apart from the typical two dimensional grid of neurons, any other kind of network topology can be implemented (e.g. one dimensional or toroidal map). Because the values for the adaptation factors are provided by an external controller, any adaptation function and neighborhood function may be realized with the proposed hardware (without any changes in the field programmable gate array (FPGA) configuration). To date,

supercomputers in the market are using multiple instructions multiple data (MIMD) approach to data manipulation. The technology is such that the computer is designed to have a cluster of machines, each of which implements new architectures SIMD instructions.

A-1.2 Hardware requirements

SOM software designers and implementers can use an ordinary computer with the following specifications:

- **Processor:** Minimum PC Required: Pentium4 and above
- **RAM Memory:** 512MB Minimum on both PC. If buying a new one, 1 GB is recommended.
- **Hard Drive:** Minimum of 40 GB.
- **USB Drive:** 512 MB USB drive or larger for downloading large files/presentations.
- **DVD/ROM:** DVD/ROM for curricular application distribution.
- **Display Resolution:** A display capable of a minimum resolution of at least 1024 x 768 pixels.
- **Sound Card:** Ability to play sound.
- **Operating Systems** Windows XP or Vista.
- **Productivity Tools:** PC Minimum Requirement: Word, Excel, PowerPoint, Microsoft Office XP, Office 2003 or Office 2007, Visual C++ or any other equivalent programming software

A-1.11 Matlab SOM Toolbox

The SOM Toolbox is built using the MATLAB script language. Since structures and N-dimensional matrices are used, it requires Matlab 5. The basic SOM Toolbox built under UNIX environment does not require any other toolboxes to work, just the basic Matlab. However, some of the contributed functions may require for example Statistics Toolbox.

A-1.3 Standalone Processors

Stand alone massive parallel data processors have been developed. Traditionally, hardware engineers have used FPGA technology with programming tools tailored to embed experts' knowledge. However, as FPGAs become faster and more affordable, engineers and scientists with little or no digital hardware design expertise are looking to take advantage of FPGAs to create custom solutions. To address this growing interest, vendors are creating higher-level tools that make it easier to program FPGAs and deliver the benefits of FPGA technology to new applications.

In order to limit the hardware requirements for the implementation, the original SOM algorithm has been simplified. In particular the Manhattan distance is used for calculating the distance between the input vector and the model vectors to avoid multiplications as required for the Euclidean distance (which is typically used in SOM implementations). The internal precision is set to 16 bit and the accuracy of the input vector components and of the model vectors is set to eight bit. Restricting the values of the neighborhood function to negative powers of two gives us the opportunity to replace the multiplications that are required for adaptation by shift operations. Of course these simplifications do not come without shortcomings (disadvantages) (e.g. the convergence

time may increase in some cases), but it has been shown that the simplified algorithm is well suited for a lot of applications. Furthermore the actual generation of Xilinx FPGAs comes with integrated multipliers and our implementations on these chips will thus be able to use Euclidean distance instead of Manhattan distance with no loss in performance.

Data pre- and post-processing is a crucial and often ignored aspect in neural-computer design. The use of reconfigurable hardware enables us to implement optimally fitting hardware implementations not only for neural networks but also for pre- and post-processing. The visualization of component maps and pattern position maps is supported as well as all kind of distance matrices like the U-Matrix. Implementing these algorithms in hardware dramatically reduces communication and thus enables a more efficient utilization of the hardware accelerator.

A-1.31 NBISOM_25 IC:

The NBISOM_25 is an integrated circuit that contains 25 processing elements in a 5 by 5 array. Due to the scalability of the chips a VME-bus board was built with 16 ICs on it. The controller for the VME-bus and the SOM hardware are realized using Field Programmable Gate Arrays (FPGA). The system runs SOM applications with up to 400 elements in parallel mode (20 by 20 map). Each weight vector can have up to 64 weights of 8 bit accuracy.

A-1.4 Further information

[1] M. Porrmann, M. Franzmeier, H. Kalte, U. Witkowski and U. Rückert, “A Reconfigurable SOM Hardware Accelerator”, ESANN’2002 proceedings, 2002, pp. 337-342

[2] PhyCORE®-ARM7/LPC2294 Rapid Development Kit:

http://www.phytec.com/products/rdk/ARM-XScale/phyCORE-ARM7-LPC229x.html

A- 2: CUTTINGS FROM NEWS PAPER

SUNDAY NATION ONLINE

Sunday, July 08, 2007

=premium content

Ads by Google Kenya Nation Kenya Times Air Kenya Ke

TODAY
Home
Today's News
News
Politics
News Analysis
Special Report
World News
News Extra
Opinion
Saturday Review
Business
Sports
Cutting Edge
Daily Magazine
Yesterday's News

SERVICES
Archives
NationMobile
Financial Reports
Digital Paper

NEWS

How dangerous lifestyles are sending young Kenyans to an early grave

Story by DAGI KIMANI and ARTHUR OKWEMBA
Publication Date: 2007/07/08

The famous Kavirondo Study by British missionaries early in the last century found no trace of hypertension among the Luo. The first African heart attack was recorded in 1967.

Today, 6 0 per cent of heart ailments are among young people, in the prime of their lives. For more than a decade since the mid-1990s, Aids was the leading killer. Now, a deadly cocktail of lifestyle diseases is quietly taking the lead. And doctors are frightened that before anyone takes any notice, it might be too late.

Roasting meat in a grill. The popular nyama choma is not good for your health. Photos/ELVIS OGINA

Lifestyle diseases will soon overtake all other diseases such as malaria and Aids as the leading killer of young people unless the Government — and you — do something about it, doctors warn.

More young Kenyans are falling victim to

Cutting 1: Kenya Newspaper

SUNDAY NATION ONLINE

Sunday, July 08, 2007

=premium content

Ads by Google Kenya TV Kenyan Sport Family Media

TODAY

Home

Today's News

News

Politics

EDITORIALS

Sad outcome of bad eating and lack of exercise

Publication Date: 2007/07/08

It is one of the cruel ironies of life. Prosperity will kill you, just as will poverty. The poor cannot afford good nutrition or medical care, so they die.

Cutting 2: Kenya Newspaper