Founded 1988

APPLICATION OF ADVANCED NEURAL NETWORKS IN HYPOGLYCEMIA DETECTION SYSTEM

PHYO PHYO SAN

(*B.Eng.* & *M.Sc.* & *M.Eng.*)

THIS THESIS IS SUBMITTED

FOR THE REQUIREMENT OF DOCTOR OF PHILOSOPHY TO

FACULTY OF ENGINEERING & INFORMATION TECHNOLOGY

UNIVERSITY OF TECHNOLOGY, SYDNEY

APRIL 2013

Acknowledgments

First of all, I would like to express my appreciation to my supervisor, Dr. Steve Ling for his continuous advice, insightful discussions and remarks on many occasions that are extremely helpful in improving my research work as well as my thesis.

I am also very grateful to my co-supervisor, Professor Hung Tan Nguyen for his valuable suggestions, thoughtful guidance and support throughout my PhD studies.

My special thanks to all colleagues and staff members in the Center for Health Technology, University of Technology, Sydney. I also place on record my sense of gratitude to one and all who directly or indirectly having lent me a helping hand during my studies.

Latest but not least, I wish to express my love and gratitude to my parents, Mr. Pe Than and Mrs. Khin Win Yee for their unceasing encouragement and continuous support in various ways. Deepest appreciation to my husband, Mr. Wunna Tun and my son, Nyi Lin San for their understanding and continuous support through the duration of my studies.

Certificate of Authorship/Originality

I hereby declare that this thesis is my own work and effort and that it has not been submitted anywhere for any award. Any help that I have received in my research work and the preparation of the thesis itself has been acknowledged.

Signature Production Note:
Signature removed prior to publication.

Name PHYO PHYO SAN

Contents

Abstract

Hypoglycemia is the medical term for a state produced by lower levels of blood glucose. It represents a significant hazard in patients with Type 1 diabetes mellitus (T1DM) which is a chronic medical condition that occurs when the pancreas produces very little or no insulin. The imperfect insulin replacement places patients with T1DM at increased risk for frequent hypoglycemia. Deficient glucose counter-regulation in T1DM patients may even lead to severe hypoglycaemia even with modest insulin elevations. It is very dangerous and can even lead to neurological damage or death. Thus, continuous monitoring of hypoglycemic episodes is important in order to avoid major health complications.

Conventionally, the detection of hypoglycemia is performed by puncturing the fingertip of patients and estimate the blood glucose level (BGL) as well as the stage of hypoglycemia. However, the direct monitoring of BGL by extracting blood sample is inconvenient and uncomfortable, a more appealing preposition for preventing hypoglycemia is to monitor changes in relevant physiological parameters. Findings from

numerous studies indicate that sudden nocturnal death in type 1 diabetes is thought to be due to ECG QT prolongation with subsequent ventricular tachyarrhythmia in response to nocturnal hypoglycaemia. Though several parameters can be monitored, the most common physiological parameters to be effected from a hypoglycemic reaction are heart rate (HR) and corrected QT interval (QTc) of the ECG signal. Considering the real-time physiological parameters (HR and QTc) changes during hypoglycemia, a non-invasive monitoring of glycemic level is predicted for the hypoglycemia.

The topic of this thesis is covered by novel methodologies for the non-invasive hypoglycemia detection system by analyzing the behavioral changes of physiological parameters such as HR and QTc. These algorithms are comprised of three different classification techniques, i) variable translation wavelet neural network (VTWNN), ii) multiple regression-based combinational neural logic network (MR-NLN) and iii) rough-block-based neural network (R-BBNN). By taking the advantages of these proposed network structures, the performance in terms of sensitivity and specificity of non-invasive hypoglycemia monitoring system is improved.

The first proposed algorithm is VTWNN in which the wavelets are used as transfer functions in the hidden layer of the network. The network parameters, such as the translation parameters of the wavelets are variable depending on the network inputs. Due to the variable translation parameters, the proposed VTWNN has the ability

to model the inputoutput function with input-dependent network parameters. Effectively, it is an adaptive network capable of handling different input patterns and exhibits a better performance. With the adaptive nature, the network provides a better performance and increases the learning ability. For conventional wavelet neural network, a fixed set of weight is offered after the training process and fail to capture nonstationary nature of ECG signal. To overcome with this problem, VTWNN with multiscale wavelet function is firstly introduced in this thesis. With the variable translation parameter, the proposed VTWNN gives faster learning ability with better generalization.

The second algorithm, MR-NLN is systematically designed which is based on the characteristics of application. Its design is based on the binary logic gates (AND, OR and NOT) in which the truth table and K-map are constructed and it depends on the knowledge of application. Because the logic theory are used in the network design, the structure becomes systematic and simpler compared to other conventional neural networks (NNs) and enhance the training performance. Traditionally, the conventional NNs with the same structure are applied to handle different applications. The optimal performance may not always guaranteed due to different characteristics of applications. In real-world applications, the knowledge based-neural network that understands all the characteristics of practical applications are preferred for optimal performance. In conventional NNs, the redundant connections and weights of conventional neural networks make the number of network parameters unnecessarily large

and downgrades the training performance. But for neural logic network (NLN), the structure becomes simpler.

The third algorithm focuses on the hybridization technology using rough sets concepts and neural computing for decision and classification purposes. Based on the rough set properties, the input signal is partitioned to a predictable (certain) part and random (uncertain) part. In this way, the selected block-based neural network (BBNN) is designed to deal only with the boundary region which mainly consists of a random part of applied input signal and caused inaccurate modeling of data set. Due to the rough set properties and the adaptability of BBNN's flexible structures in dynamic environments, the classification performance is improved. Owing to different characteristics of neural network (NN) applications, a conventional neural network with a common structure may not be able to handle every applications. Based on the knowledge of application, BBNN is selected as a suitable classifier due to its modular characteristics and ability in evolving the size and structure of the network.

To obtain the optimal set of proposed network parameters, a global learning optimization algorithm called hybrid particle swarm optimization with wavelet mutation (HPSOWM) is introduced in this thesis. Compared to other stochastic optimization methods, the hybrid HPSOWM has comparable or even superior search performance for some hard optimization problems with faster and more stable convergence rates. During the training process, a fitness function which is characterized by the proposed network design parameters is optimized by reproducing a better fitness value.

Abstract

The proposed systems is validated using clinical trial conducted at the Princess Margaret Hospital for Children in Perth, Western Australia, Australia. A total of 15 children with 529 data points (ages between 14.6 to 16.6 years) with Type 1 diabetes volunteered for the 10-hour overnight for natural occurrence of nocturnal hypoglycemia. Prior to the application of the algorithms, the correlation between the measured physiological parameters, HR and QTc and the actual BGL for each subject were analyzed. The feature extracted ECG parameters, HR and QTc significantly increased under hypoglycemic conditions ($BGL \leq 3.3mmol/l$) according to their respective p values, HR ($p < 0.06$) and QTc ($p < 0.001$). The observation on these changes within the physiological parameters have provided the groundwork for model classification algorithms.

List of Figures

List of Tables

List of Symbols and Abbreviations

a	Dilation parameter of multi-wavelet
b	Translation parameter of multi-wavelet
n_{in}	Number of network inputs
n_h	Number of hidden nodes
n_{out}	Number of network outputs
$\psi_{a,b}(\cdot)$	Multi scaled wavelet function
$\psi(\cdot)$	Mother wavelet function
n_{para}	Total number of network parameters
$para^j_{\max}$	Maximum value (Upper boundary) of particle element
$para^j_{\min}$	Minimum value (Lower boundary) of particle element
$v_{\max}$	Maximum velocity of particle
$v_{\min}$	Minimum velocity of particle
$f(\cdot)$	Activation function
u_i	Input variables
v_{ij}	Weight between ith input and jth hidden nodes
w_{jl}	Weight between jth hidden and lth output nodes
b_j, b_l	Biases for hidden and output nodes
$logsig(\cdot)$	Logarithmic sigmoid transfer function
$tansig(\cdot)$	Hyperbolic tangent sigmoid transfer function
$pureline(\cdot)$	Liner transfer function
$\vee$	Maximum operator
o	Logic-AND operator
$\bullet$	Logic-OR operator
β	Regression parameters/coefficients
$\overline{I}(\cdot)$	Upper rough approximation region
$\underline{I}(\cdot)$	Lower rough approximation region
ζ_{wm}	Shape parameter of wavelet
μ_c	Probability of mutation

EEG	Electroencephalogram
ECG	Electrocardiographic
T1DM	Type 1 Diabetes Mellitus
QTc	Corrected QT Interval
HR	Heart Rate
CGMS	Continuous Glucose Monitoring System
VTWNN	Variable Translation Wavelet Neural Network
WNN	Wavelet Neural Network
FIS	Fuzzy Inference System
RBFNN	Radial Basis Function Network
FFNN	Flashforward Neural Network
WM	Wavelet Mutation
PSO	Particle Swarm Optimization
GA	Genetic Algorithm
BP	Back-propagation Learning
HPSOWM	Hybrid Particle Swarm Optimzation
HGAPSO	Hybrid PSO with GA mutation
ROC	Receiver Operating Characteristic
MR	Multiple Regression
NLN	Neural-Logic Network
BBNN	Block-Based Neural Network
BBNN	Block-Based Neural Network
R-BBNN	Rough-Block-Based Neural Network

Author's Publications

The contents of this thesis are based on the following papers that have been published, accepted, or submitted to peer-reviewed journals and conferences.

International Journal Papers:

1. Phyo Phyo San, Sai Ho Ling, and Hung T. Nguyen, "Industrial Application of Evolvable Block-Based Neural Network to Hypoglycemia Monitoring System", *IEEE Transactions on Industrial Electronics*, vol. 60, no, 12, pp. 5892-5901, 2013.

2. Phyo Phyo San, Sai Ho Ling, and Hung T. Nguyen, "Hybrid PSO-based Variable Translation Wavelet Neural Network and Its Application to Hypoglycemia Detection System", *Neural Computing and Applications*, In-press, October, 2012.

3. Sai Ho Ling, Phyo Phyo San and Hung T. Nguyen, "Non-invasive nocturnal hypoglycemia detection for insulin-dependent diabetes mellitus using genetic fuzzy logic method", *International Journal of Computational Intelligence and Applications*, In-press, 2012.

4. Phyo Phyo San, Sai Ho Ling, and Hung T. Nguyen, "Evolvable Rough-Block-Based Neural Network and Its Biomedical Application", *IEEE Transactions on Systems, Man, and Cybernetics B*, Under Revision, July, 2012.

5. Phyo Phyo San, Sai Ho Ling, and Hung T. Nguyen, "Application of Combinational Neural Logic System to Non-Invasive Hypoglycemic Monitor in Patients with T1DM", *IEEE Transactions on Systems, Man, and Cybernetics B*, Under Review, October, 2012.

International Conference Papers:

1. Phyo Phyo San, Sai Ho Ling, and Hung T. Nguyen , "Combinational Neural Logic System and Its Industrial Application", *8th IEEE Conference on Industrial Electronics and Applications*, Melbourne, Australia, 19-21 June, pp. 947-952, 2013. (Finalist of IEEE ICIEA 2013 Best Paper Award)

2. Phyo Phyo San, Sai Ho Ling, and Hung T. Nguyen , "Intelligent Detection of Hypoglycemic Episodes in Children with Type 1 Diabetes using Adaptive Neural-Fuzzy Inference System", *34th Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, San Diego, California USA, August 28-September 1, pp. 6325-6328, 2012.

3. Phyo Phyo San, Sai Ho Ling, and Hung T. Nguyen , "Optimized Variable Translation Wavelet Neural Network and Its Application in Hypoglycemia Detection

System", *7th IEEE Conference on Industrial Electronics and Applications*, Singapore, 18-20 July, pp. 534-538, 2012.

4. Phyo Phyo San, Sai Ho Ling, and Hung T. Nguyen , "Hybrid Particle Swarm Optimization Based Normalized Radial Basis Function Neural Network For Hypoglycemia Detection", *IEEE World Congress on Computational Intelligence*, Brisbane, Australia, 10-15 June, pp. 2718-2723, 2012.

5. Phyo Phyo San, Sai Ho Ling, and Hung T. Nguyen , "Block Based Neural Network for Hypoglycemia Detection", *33rd Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, Boston, USA, August 30-September 3, pp. 5666-5669, 2011.

6. Phyo Phyo San, Sai Ho Ling, and Hung T. Nguyen , "Non-invasive Detection of Hypoglycemic Episodes in Type1 Diabetes Using Intelligent Hybrid Rough Neural System", *35th Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, Osaka, Japan, July 3-7, Submitted, 2013.

Chapter in Book:

1. Sai Ho Ling, Phyo Phyo San and Hung T. Nguyen, "Hypoglycaemia Detection for Insulin-dependent Diabetes Mellitus: Evolved Fuzzy Inference System Approach", *Computational Intelligence and Its Applications: Evolutionary Computation, Fuzzy Logic, Neural Network and Support Vector Machine Techniques*, World Scientific, UK, pp. 61-85, March, 2011.

Chapter 1

Introduction

Hypoglycemia is a common side-effect of insulin therapy for patients with type 1 diabetes mellitus (T1DM) and is the major limiting factor to maintain tight glycemic control. The deficiency in glucose counter-regulation may even lead to severe hypoglycaemia. It is always threatening to the well-being of patients with T1DM since more severe hypoglycemia leads to seizures or loss of consciousness and the possible development of permanent brain dysfunction under certain circumstances. It has been confirmed that an increased risk of sudden death in young people is related to hypoglycemia and patients with type 1 diabetes. Thus, early detection and prevention on hypoglycemia becomes the number one treatment goal of diabetes.

Though tremendous development blood glucose monitoring systems are currently available for diabetic self-management and control, long-term continuous measurement of hypoglycemia still remains as a significant barrier for intensive therapy. This

thesis introduces novel methodologies for prediction on onset of hypoglycemia and providing early hypoglycemic alarms. This chapter begins with the background study of this research in Section 1.1. The thesis motivation, the objectives, contribution and organization are discussed in Sections 1.2, 1.3 and 1.4 respectively.

1.1 Background

Type 1 diabetes usually arises in children or young adults and is characterized by the inability to produce insulin. Although treatment regimens and types of insulin available have developed over time, individuals with Type 1 diabetes are still required to self inject with insulin to control their blood sugar levels. Insulin doses must be adjusted according to carbohydrate intake and the degree of physical activity being undertaken.

Hypoglycemia, or low blood sugar levels, is one of the most common side effects among individuals with Type 1 diabetes. It is also well know as a barrier for achieving and maintaining tight glycemic control. Individuals who experience frequent episodes of hypoglycemia and those who have been on insulin therapy for a long period of time may develop impaired awareness of hypoglycemia.

Early warning symptoms such as tremor and sweating may no longer occur and symptoms including drowsiness and lack of concentration become more prominent. Impaired hypoglycaemic awareness occurs in approximately 25 % of individuals with

type 1 diabetes. Severe hypoglycemia may lead to physical and psychosocial morbidities such as brain damage, loss of consciousness, depression and low self-esteem [Frederick2008]. It is also well known as death-in-bed syndrome. Several research works have been reported where patients with type 1 diabetes have been found dead in an undisturbed bed [Klonoff2001].

Since hypoglycemia can occur at any age due to many different conditions, early detection and prevention of hypoglycemia is more important for diabetes patients. To prevent or minimize hypoglycemic morbidities and mortalities, it is vital to return the blood glucose level to normal soon after a hypoglycemic event occurs. The patients' conditions are obviously need to know; when the event of hypoglycemia is initiated, and/or how the early warning is detected [Osareh2008]. In other words, the need for hypoglycemia detection device is essential in order for giving alarm alert to Type 1 diabetes patients or carer when low blood glucose level is detected.

Currently, the glucose meters are still using as standard techniques for monitoring the blood glucose level as well as detection of hypoglycemia [Graaff1999]. Efforts have been made in order to reduce the level of invasiveness by decreasing the blood sample volume to a few microlitres, and measuring areas of the body less sensitive to pain than fingertips. However, the needle-type glucose sensor used in the system can only give accurate measurements over a certain period of time and it is limited to used in critical nocturnal stage.

Due to discomfort and inconvenience of finger-stick methods, minimally invasive or non-invasive continuous methods are tested and introduced using technologies like reverse iontophoresis, polarimetry, metabolic heat conformation, ultrasound, thermal emission, electromagnetic, photoacoustic, Raman, light absorption and bioimpedance spectroscopy [Amaral2008]. Together with the choice of technique and sample regions are defined mostly by sweating, skin color, surface roughness, tissue thickness, breathing artifacts, blood flow and body movements. However, hypoglycemia preventive effect of continuous glucose monitoring has not been successfully established in practise. The technology advancement in the diabetes diagnostic testing and self-monitoring market is expected to have a non-invasive hypoglycemia monitoring system with the use of novel design concepts.

An alternative approach for detection of hypoglycemia is based on the considerable change of physiological factors during hypoglycemia. Findings from [Murphy2004] [Lee2004] confirmed that the electrocardiographic (ECG) abnormalities, especially in prolongation of corrected QT interval (QTc) is greatly affected by hypoglycemia. Based on ECG variations, numerous studies [Alexakis2003] [Alexakis2006] have been investigated on the detection of hypoglycemia by means of artificial neural network (ANN) and linear discriminant analysis (LDA). The detection of hypoglycemia has been investigated not only for ECG, but also for other physiological factors; electroencephalogram (EEG) signal [Juhl2010] [Iaione2005] and skin temperature [Johansen1986].

Considering correlation between physiological parameters of ECG signal and the status of hypoglycemia, non-invasive hypoglycemia monitoring systems have been tested by the use of appropriate computational intelligence technologies; fuzzy estimator [Ghevondian1997] [Hastings1998], fuzzy neural network estimator [Ghevondian2001]. By using fuzzy decision maker, a set of linguistic rules are generated. The status of hypoglycemia is predicted through the changes of heart rate (HR) and skin impedance parameters.

Similarly, the Bayesian neural network algorithm has been applied in non-invasive hypoglycemia monitoring systems due to its merit in generalization when addressing both nonlinear and the fuzzy nature of patients' data. To implement neural network-based monitoring systems, the physiological parameters of ECG signal such as heart rate (HR), corrected QT interval (QTc) and skin impedance were used as the main inputs and the status of hypoglycemia was detected [Nguyen2006] [Nguyen2007].

Furthermore, hybridization technologies such as genetic algorithm-based neural network system [Chan2010] [Chan2011], evolved fuzzy reasoning model [Ling2010] [Ling2012], swarm-based support vector machine [Nuryani2012] and statical multiple regression with fuzzy inference system [Ling2011], have been successfully applied to hypoglycemia monitoring systems. Even though satisfactory performances were found by the use of intelligent computational techniques, the overall accuracy of the proposed detection systems may not be enough to work as a good hypoglycemia detector. Much more effort is still needed to improve the designs and technologies of hypoglycemia

monitoring systems in order to avoid major health complications in insulin dependent Type 1 diabetes patients.

1.2 Motivation of Thesis

To analyze the blood glucose level (BGL) as well as the presence of hypoglycemia, many monitoring systems have been developed to be invasive or minimally invasive with the use of a needle-type probe. [Shichiri1982] [Kajiwara1993]. However, this kind of measurement can only provide an isolated glucose level which does not reflect variations occurring throughout the day and night. For serve hypoglycemia, monitoring BGL continuously becomes almost impossible task, especially during the stage of nocturnal hypoglycemia.

Other minimally invasive devices [Kimura1987] or non-invasive devices [Malin1994] [Heise1994] have been tested and introduced with the aid of current technologies such as reverse iontophoresis, suction effusion fluid and near-infrared spectroscopy. But they suffer from limitations in terms of measurement inaccuracy, high susceptibility to artefact noise, considerable time delays for obtaining results and long-term reproducibility. The complications which are inherent with these techniques limit their use as practical non-invasive hypoglycemic monitors.

An alternative idea and solutions are still needed for non-invasive continuous monitoring of BGL as well as for prediction of hypoglycemia. To get a hold of this

requirement, hypoglycemia detection systems have been investigated by the use of physiological parameters that are stimulated by falling blood glucose levels. For instance, Teledyne Sleep Sentry [Hansen1983] [Castano2000], multiparameter measurement system, Mini Med Medtronic CGMS [Maia2007] and Abbott Freestyle Navigator CGMS [Weinstein2007].

Teledyne Sleep Sentry is equipped with a sensor and measures falling skin temperature and skin resistance, but the device has high ratio of false positive alarms and results in a disrupted sleep cycle. In multiparameter measurement system, electroencephalogram (EEG), pulse and skin impedance are used to monitor the status of hypoglycemia, however the parameter, skin impedance has less correlation. For Mini Med Medtronic CGMS and Freestyle Navigator CGMS, it has been reported that the large mean absolute relative difference (MARD) is received 10 to 15% and 26.4% due to low sensor efficacy while detecting hypoglycemia.

Based on the strong relationship between physiological parameters and hypoglycemia, a non-invasive hypoglycemia monitor, HypoMon was tested and introduced from AIMedics Pty Ltd. This device measures the physiological parameters (heart rate (HR), corrected QT interval (QTc) and skin impedance) continuously and provides the detection of hypoglycemic episodes through the variations of those input parameters [Nguyen2006] [Nguyen2007]. By employing Bayesian neural network algorithm, the obtained sensitivity was found to be satisfactory, but the specificity need to be improved. More advanced neural network algorithms are still needed for obtaining

better testing sensitivity and specificity.

In addition, fuzzy estimator [Ghevondian1997] [Hastings1998], fuzzy neural network estimator [Ghevondian2001] were proposed for hypoglycemia detection systems. In these systems, the status of hypoglycemia was predicted through the changes in physiological parameters of the ECG signal. However, the membership functions which express the linguistic terms for fuzzy inference rules have to be defined. In practise, there is no formal approach for defining these rules and the optimization of these membership functions in terms of generalizing the data is also very important. Usually, tuning parameters of membership function is a time consuming task.

In [Ling2010] [Ling2012], an evolved fuzzy reasoning model was developed to recognize the presence of hypoglycemic episodes. The physiological parameters such as heart rate and corrected QT interval of the electrocardiogram (ECG) signal were continuously measured for early detection of hypoglycemic episodes in Type 1 diabetes mellitus (T1DM) patients. In evolved FRM, the fuzzy rules and fuzzy membership functions were optimized by an evolutionary algorithm called hybrid particle swarm optimization with wavelet mutation (HPSOWM). The results showed that the proposed algorithm performed well in terms of the clinical sensitivity and specificity. However, in evolved FRM, more parameters were need to be optimized. Hence, it is desirable to have an automatic adaption procedure with less design parameters

Further experiment were conducted by the use of neural network based rule discovery

system [Chan2010] [Chan2011]. The development was based on 420 data sets which were collected from 16 T1DM patients by using the genetic algorithm. It was found that the obtained sensitivity and specificity were reasonable. Apart from obtaining reasonable sensitivity and specificity, the neural network used in the developed detection system might not be able to improve further because its processing capability and adaptability is limited by fixed neural network structure.

Traditionally, conventional neural networks with fixed structures were applied to handle different applications. However, those network models offer a fixed single set of weights after training process that may not be good enough if data sets are distributed in a vast domain separately and/or the number of network parameters is too small. With these conventional neural networks with fixed structure, the optimal performance was not always guaranteed for the application with different characteristics. In some cases, the redundant connections and weights of conventional neural network makes the number of network parameters unnecessarily large and downgrades the training performance. Thus, the knowledge based neural network that understands all the characteristics of practical application is preferred for optimal performance.

In addition, those conventional neural networks with fixed structure offer fewer degrees of freedom (DOF) due to their network structure. Generally, the neural network with fixed structures might not be able to scale up to meet with the application requirements. Since the fixed structure NN were not structurally evolved during the training process, the degrees of freedom (DOF) were less. Conceptually, the number

of DOF allowed the network to adapt more with the characteristics of application. The neural network models were less accurate if less DOF were offered. The need for better sensitivity and specificity will only be achieved if NN with more DOF are able to be designed for different characteristics of applications. For the direction of achieving a truly non-invasive hypoglycemia detection system, the neural network structure with the most DOF that will be breaking the barriers of fixed structure NN become the most desirable step.

With a focus on improving sensitivity and specificity of the hypoglycemia detection system, in this thesis, three advanced neural network models such as variable translation wavelet neural network (VTWNN), multiple regression based combinational neural logic network (MR-NLN) and rough-block-based neural network (R-BBNN) were proposed in Chapter 3, 4 and 5 respectively.

Neural network training and learning are important issues in its implementation. The learning algorithm determines the rules for optimizing the weight and structure of the neural network within the training period. In the developed hypoglycemia detection systems [Nguyen2006] [Nguyen2007], the commonly used training algorithms- back propagation algorithms with variable learning rate [Haykin1999], and conjugate gradient algorithm [Moller1993] were only applied for training. However, these methods may only converge to a local minimum and are sensitive to the values of initial parameters. The function to be optimized needs to be differentiable and the learning method may only be good to some specific network structure.

To overcome local minimum problem, one of the most stochastic search algorithm, called particle swarm optimization (PSO) was introduced [Kennedy1995]. By the use of PSO algorithm, the error functions are less likely to be trapped in local minimum and need not to be differentiable. For proposed NNs in this thesis, an improved hybrid PSO with wavelet mutation (HPSOWM) [Ling2008] is selected as a suitable training algorithm. The training performance in HPSOWM depends only on the input-output data and the derivative information of the cost function is not required as needed in back propagation algorithms [Mazumdar2008]. Otherwise, the updating rule is required to derive each time for each different network structure. Besides, the hybrid HPSOWM is comparable with other stochastic optimization methods by achieving faster and more stable convergence rates [Ling2008]. The effectiveness of each proposed novel neural network based hypoglycemia detection system is discussed detail in Chapters 3, 4 and 5 respectively.

1.3 Objectives and Contributions

The main objective of this thesis is to develop three advanced neural network classifiers for hypoglycemia detection. In this work, the main objective is to accurately detect hypoglycemia episodes or measure the blood glucose level non-invasively and continuously without the need for taking blood samples from the fingertips of diabetic patients. Only physiological parameters of ECG signal are required.

For the prediction of hypoglycemia episodes, advanced neural network classifiers such as variable translation wavelet neural network (VTWNN), multiple regression based combinational neural logic network (MR-NLN) and rough-block-based neural network (R-BBNN) are proposed. For optimization of the proposed neural network structures and weights, a global learning optimization algorithm called hybrid particle swarm optimization with wavelet mutation (HPSOWM) is introduced in this thesis. In the proposed neural network models, the usefulness and advantages of existing NNs are preserved, removing the major limitations of fixed structure NN. Such advanced neural network models aimed to have advantages for all applications including the hypoglycemia detection system.

With the following contributions, advanced neural network classifiers are designed and developed for the detection of hypoglycemic episodes in order to obtain higher sensitivity and specificity. The contributions are:

- Introduce three different advanced neural networks namely variable translation wavelet neural network (VTWNN), multiple regression model based combinational neural logic network (MR-NLN) and rough-block-based neural network (R-BBNN) for the hypoglycemia monitoring system.

- The first proposed neural network, variable translation wavelet neural network (VTWNN) in which the wavelets are used as transfer functions in the hidden layer of the network. The network parameters, such as translation parameters of

the wavelets are variable depending on the network inputs. Due to the variable translation parameters, the proposed VTWNN has the ability to model the inputoutput function with input-dependent network parameters. Effectively, it is capable of handling different input patterns and provides better performance.

- The second neural network, multiple regression based neural logic network model (MR-NLN) is systematically designed based on its own characteristics of application. In this design, the binary logic gates (AND, OR and NOT) and their associated truth table and K-map are constructed depend on the knowledge of application. Since the logic theory is used in the network design, the structure becomes systematic and simpler compared to other conventional neural network (NN) and enhances the training performance.

- The third neural network, rough approximation block-based neural network (R-BBNN) is designed based on the hybridization technology using rough sets and neural network. With the use of rough set defined regions, lower region and boundary region, the applied input signal is partitioned to a predictable part (certain) and the random (uncertain) part. In this way, the selected block-based neural network (BBNN) is designed to deal only with the boundary region which mainly consists of a random part of applied input signal and caused inaccurate modeling of data set. Due to rough set properties and the structural flexibility of BBNN, the classification performance is improved.

- Introduce hybrid particle swarm optimization with wavelet mutation (HPSOWM) in order to obtain an optimal set of network parameters. For the proposed advanced neural networks, HPSOWM is selected as a suitable training algorithm because its training performance depends only on the input-output data while the derivative information of the cost function is not required as in back propagation algorithms. If the training algorithm needs the derivative information of cost function, the updating rule is required to derive each time for each different network structure. Besides, the hybrid HPSOWM is comparable with other stochastic optimization methods by achieving faster and more stable convergence rates in optimization problems.

- Investigate the proposed advanced neural network classifiers by the use of actual data sets of 15 T1DM children with ages 14.6 ± 1.5 years, which are collected at the Department of Health, Government of Western Australia. Validate the proposed algorithms as compared with other conventional algorithms such as feedforward neural network (FFNN), wavelet neural network (WNN), genetic algorithm (GA) based multiple regression with fuzzy interference system (MR-FIS), evolved fuzzy interference system (FIS) and multiple regression (MR). Several experiments showed that the proposed algorithms performed well in terms of sensitivity and specificity.

1.4 Structure of Thesis

The thesis is organized into seven chapters. Chapter 2 commences with a brief background into the definition of hypoglycemia, the glucose counterregulatory mechanism and its effect on the physiological parameters under hypoglycemic conditions. This chapter also includes the detailed methodology to obtain the physiological parameters, the current technologies behind the hypoglycemia monitoring system and several evolutionary neural network classification models. It is finalized by introducing three distinct proposed advanced neural network classifiers, such as variable translation wavelet neural network (VTWNN), multiple regression based combinational neural logic network (MR-NLN) and rough-block-based neural network (R-BBNN).

Chapter 3 initially introduces a detail design and methodology of optimized VTWNN for detection of hypoglycemic episodes. A brief account of all relevant theories such as wavelet theory, design of nonlinear activation function are also presented. A global learning or training algorithm, called particle swam optimization with wavelet mutation (HPSOWM) and its parameters selection methods are discussed. This chapter also includes the investigation on the prolongation of QT interval of ECG signal during spontaneous hypoglycemia with the aid of 15 children (aged 14.6 ± 1.5 years) with Type 1 diabetes. Extensive experimental results have shown that the proposed approach is effective for hypoglycemia detection by achieving better sensitivity and acceptable specificity.

Chapter 4 presents an integrated method, a neural logic network with multiple regression for the development of a noninvasive hypoglycemia monitoring system. Through ECG signal alternation, the episodes of hypoglycemia is firstly predicted by the use of knowledge based neural logic network (NLN) while the multiple regression model is used to enhance the sensitivity and specificity. This chapter also provides a systematic design of proposed NLN including the core inference mechanism with basic logic building blocks, namely neural-logic-AND, -OR, and -NOT gates and the truth table and K-map for decision making ability. The effectiveness of the proposed algorithms is validated through the use of other conventional neural classifiers.

Chapter 5 focuses on the development of an intelligent diagnostics system using the hybrid approach of rough-block-based neural network for recognition of hypoglycemia. In this hybrid system, the pre-processing stage is firstly carried out by defining the *lower region* and *boundary region* based on the rough set properties. An in-depth description of architecture, including the partitioning of the applied input signal, HR and QTc to the predictable (certain) part and random (uncertain) part are also given in this chapter. Besides, the topology of selected block-based neural network (BBNN) with four different internal configurations and its adaptability in dynamic environments are also presented before the performance evaluation of proposed R-BBNN is carried out with numerous comparison studies.

Chapter 6 gives a discussion on the three proposed neural networks (VTWNN, MR-NLN and R-BBNN), ranging from the proposed methodologies through to obtaining

experimental results and future recommendations. Based on the obtained experimental results, the operation and performance of the proposed three networks are compared and a conclusion will be drawn.

Chapter 7 summarizes the findings and results of this thesis and future research is presented.

Chapter 2

Literature Review

2.1 Hypoglycemia

Glucose is a primary source of energy in the body and provides energy to all cells, like brain cells and red blood cells for every bodily function. It is regarded as an essential fuel for the central nervous system as well as tissues including muscle, fat, and liver that use fatty acids and other substrates to satisfy their energy needs. The glucose in the bloodstream comes from the digested food and transported to the body cells to be used as an energy source. In order for glucose to enter the cell, a hormone made by the pancreas, insulin bonds to a receptor site on the outside of the cell and acts like a key to open a doorway into the cell through which the glucose can enter as described in Fig. 2.1.

With Type 1 diabetes, cells in the pancreas that produce insulin are the target of the body immune system and are eventually destroyed. For this reason, people with Type 1 diabetes produce no insulin and the body cells cannot turn glucose (sugar) into energy. The complete lack of insulin in the body results in Type 1 diabetes as presented in Fig. 2.2.

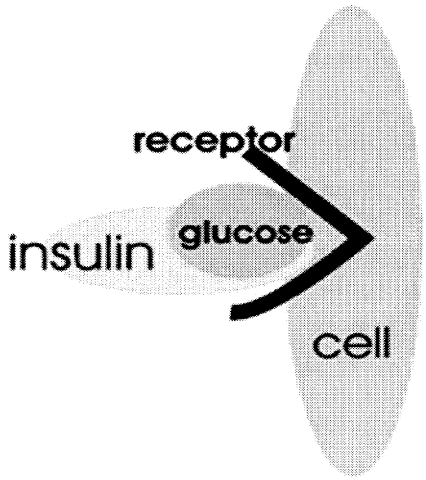

Figure 2.1: Action for glucose and insulin in normal subject

All patients with Type 1 diabetes will eventually require insulin treatment in order to maintain a near-normal level of blood glucose. Unless treated with daily injections of insulin, people with Type 1 diabetes accumulate dangerous chemical substances in the blood from the burning of their own fat as a substitute without insulin. In order to maintain blood glucose level close to a normal level, insulin replacement is required by daily injections for people with Type 1 diabetes.

Intensified lowering of blood glucose reduces the risk of chronic complications of Type 1 diabetes; however, clinical attempts to achieve these benefits are limited

by an increased risk of hypoglycemia which is induced by aggressive insulin therapy [Cryer1985]. It has been discussed that diabetic patients, especially for those who have been treated with insulin are at risk for developing hypoglycemia while it is less likely to occur in non-insulin-dependent patients who are taking sugar-lowering medicine for diabetes

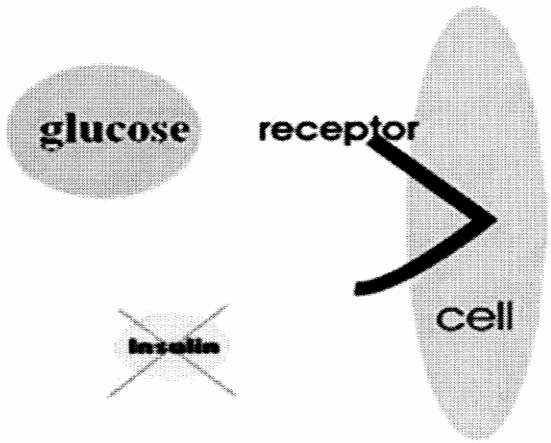

Figure 2.2: Action for glucose and insulin in Type 1 diabetic patient

Many different definitions have been provided to the concept of hypoglycemia [Frier2007]. Firstly, hypoglycemia is literally translated as a state produced by a lower level of blood glucose. It represents a significant hazard in patients with Type 1 diabetes mellitus (T1DM) which is a chronic medical condition that occurs when the pancreas produces very little or no insulin. In [Field1989] , another definition is given to hypoglycemia as the occurrence of a wide variety of symptoms in association with a plasma glucose concentration of 50 mg/dl or less. In clinical studies [Assoc2005], the events of hypoglycemia are classified ranging from mild to serve level after reviewing the background of hypoglycemia. Most surveys revealed that the tighter the

glycemic control in the younger patient, the greater frequency of both mild and serve hypoglycemia [Becker2000].

The level of blood glucose low enough to define hypoglycemia may be different for different people, in different circumstances, and for different purposes and occasionally has been a matter of controversy. Most healthy adults maintain fasting glucose levels above $70mg/dL$, $(3.9mmol/l)$ and develop symptoms of hypoglycemia when the glucose level falls below $55mg/dL$, $(3.0mmol/L)$. In [DCCT1995], it has been reported that the severe hypoglycaemic episodes are defined in those whose documented blood glucose levels is $50mg/dL$, $(2.8mmol/L)$ and the patient are advised to take necessary treatment.

A series of symptoms occur if there is not enough glucose supply to the brain since the brain and nervous system needs a certain level of glucose to function. The two typical symptoms of hypoglycemia arise from the activation of the autonomous central nervous systems (autonomic symptoms) and reduced cerebral glucose consumption (neuroglycopenic symptoms). Autonomic symptoms such as headache, extreme hunger, blurry or double vision, fatigue, weakness and sweating are activated before neuroglycopenic symptoms follow. The initial condition of the presence of hypoglycemia can only become obvious when autonomic symptoms occur which allow the patient to recognize correct ensuing episodes [DCCT1991] [Merbis1996]. The neuroglycopenic symptoms such as confusion, seizures, and loss of consciousness (coma) arise due to insufficient glucose flow to the brain [Cryer1999].

2.2 Physiological Disturbances in Hypoglycemia

As a consequence of blood glucose concentration falling to a certain level, pronounced physiological changes occurred in response to hypoglycemia [Frier2007].

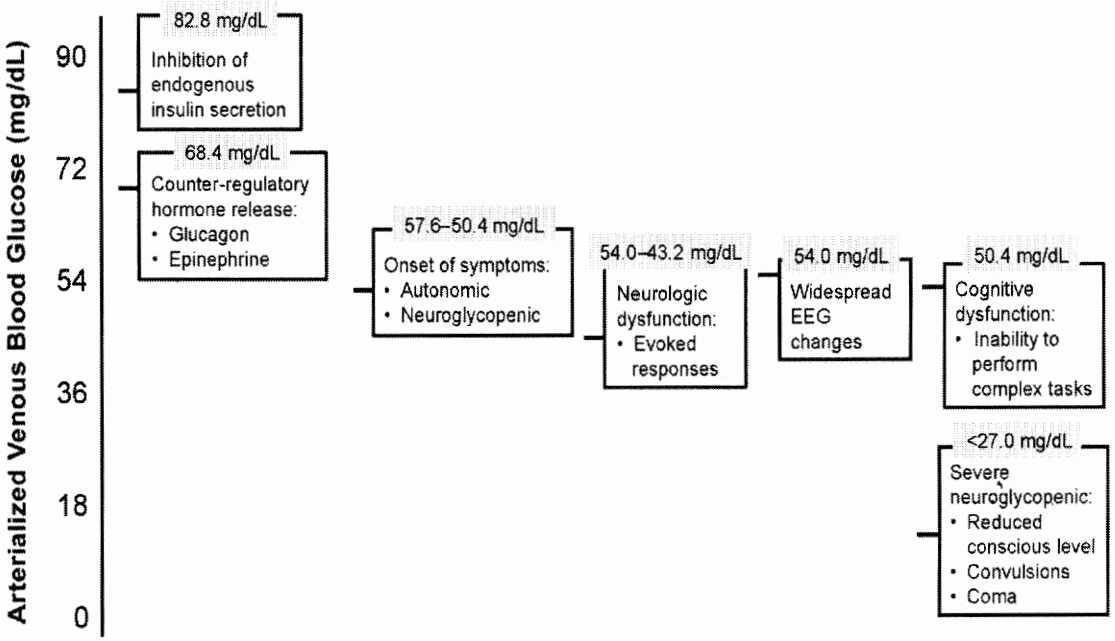

Figure 2.3: Hierarchy of Responses to Hypoglycemia [Wolpert2007]

As can be seen in Fig 2.3, falling glucose levels initiate a number of responses: insulin secretion is inhibited, followed by the release of counterregulatory hormones (glucagon and epinephrine) when glucose levels reach approximately 70 mg/dL; this is the basis for the consensus statement that proposes 70 mg/dL as the diagnostic criteria for hypoglycemia. Continued decline in glucose eventually leads to neurologic dysfunction, widespread electroencephalogram (EEG) changes, cognitive dysfunction and severe hypoglycemia. Some major physiological responses to hypoglycemia are

outlined in the following Sections: 2.2.1, 2.2.2 and 2.2.3 respectively.

2.2.1 Electrocardiography (ECG)

Hypoglycemia has long been known to affect the electrocardiogram (ECG) causing ST wave changes with lengthening of the QT interval and cardiac repolarization. These changes may increase the risk of cardiac arrhythmia; various abnormal heart rhythms, including ventricular tachycardia and atrial fibrillation [Benhorin1990]. These alternations are reflected by variations of ECG parameters, Q point, R peak and T wave peak including the commonly used QT interval (depolarization of the ventricles) which is the time taken from the start of the QRS complex to the end of the T wave (repolarization of ventricles) as presented in Fig. 2.4.

A number of studies have been carried out on the prolongation of QTc interval under both hyperinsulinemic hypoglycemia and spontaneous hypoglycemia and both conditions confirmed that QTc dispersion is increased in healthy individuals and in patients with Type 1 [Robinson2004] and Type 2 diabetes [Hallin1999].

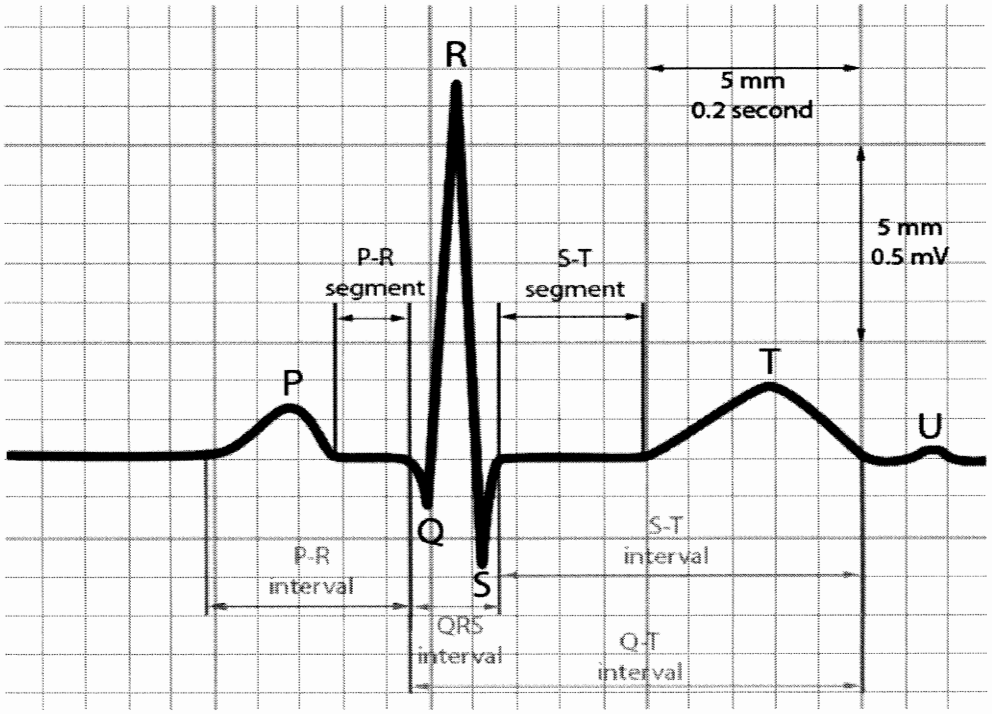

Figure 2.4: Illustration of the normal electrocardiogram (ECG) signal

In [Marques1997], the effect of insulin induced hypoglycemia on cardiac repolarization was analyzed using ECG of T wave and QRS complex morphology in 15 patients including 8 patients with insulin-dependent Type 1 diabetes and 7 patients with non-insulin dependent diabetes patients. All 15 subjects who underwent the glucose clamp studies had normal resting ECG and none of them showed microvascular complications. In this study, in order to analyze the ECG variations, ECG measurement was done by a high resolution method while the glucose clamp technique was applied for insulin-induced hypoglycemia.

The obtained results indicated that the degree of QTc lengthening during clamped

hypoglycemia (583 ms) was greater compared to the euglycaemic control period (429 ms) with $p<0.001$. An example of QT measurement from 1 subject under euglycaemia and hypoglycemia was given in Fig. 2.5 in which the QT intervals were marked using an on screen cursor by two independent observers. As can be seen in Fig. 2.5, the increase in heart rate can cause a reduction in the QT interval and the QT interval is corrected for differences in heart rate by Bazett correction, dividing QT by the square root of RR, ($QT_c = QT/\sqrt{RR}$) [Assoc2008].

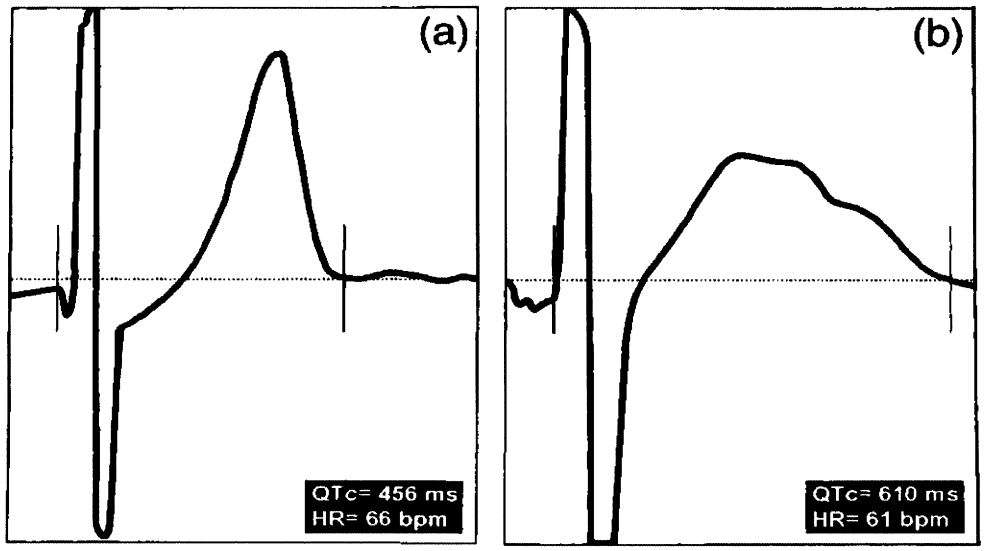

Figure 2.5: Typical QT measurement with on screen cursor placement from one subject during euglycemia: (a) showing a clearly defined T wave (b) showing prolonged repolarization [Marques1997]

Different studies [Robinson2003a] [Robinson2003b] have shown that experimental hypoglycemia were associated with significant lengthening of the corrected QT interval

(QTc) in all subjects with and without diabetes. For hyperinsulinemic clamps studies, 10 non-diabetic subjects were organized into study 1 group and study 2 group in [Robinson2003a] and 17 healthy males subjects (ages 18 to 40 years) [Robinson2003b] volunteered. Regular measurements of cardiac repolarization, QT dispersion and QTc were performed by tangent method. The measured experimental results confirmed that the QTc interval increased significantly from 406 ms to 480 ms during hypoglycemia.

Further investigation demonstrated that 8 adults with Type 1 diabetes had QTc lengthening (from 407 ms to 448 ms) during the experimental insulin-induced hypoglycemia [Lee2005]. In this experiment, hypoglycemic clamp studies were performed on two occasions, at leat 4 weeks apart. Measurements of QT interval were made using a custom-built high-resolution system with tangent method, called semiautomated tangent method and its corrected value, QTc is calculated by the Fridericia cube root formula, $QT_c = QT \Big/ \sqrt[3]{RR}$.

In [Iaione2005], 17 male non-diabetic subjects with age of 27 ± 6.4 years were underwent to study the ECG morphology, particularly for hypoglycemic QTc lengthening. With these subjects, a baseline ECG (blood glucose was around (5.0 $mmol/l$)) and hypoglycemic ECG (2.5 $mmol/l$) were recorded on a custom built system for high resolution ECG analysis. The QT measurement was done with two independent observers by using both tangent and non-tangent measuring methods in Fig. 2.6. In this study, the measured QT value was corrected for heart rate using the modified

Bazett's formula, $QT_c = QT \Big/ \sqrt{RR}$, where RR is 60/HR and it has shown to be sensitive to typical ECG morphology during hypoglycemia by giving p-value less than 0.0004.

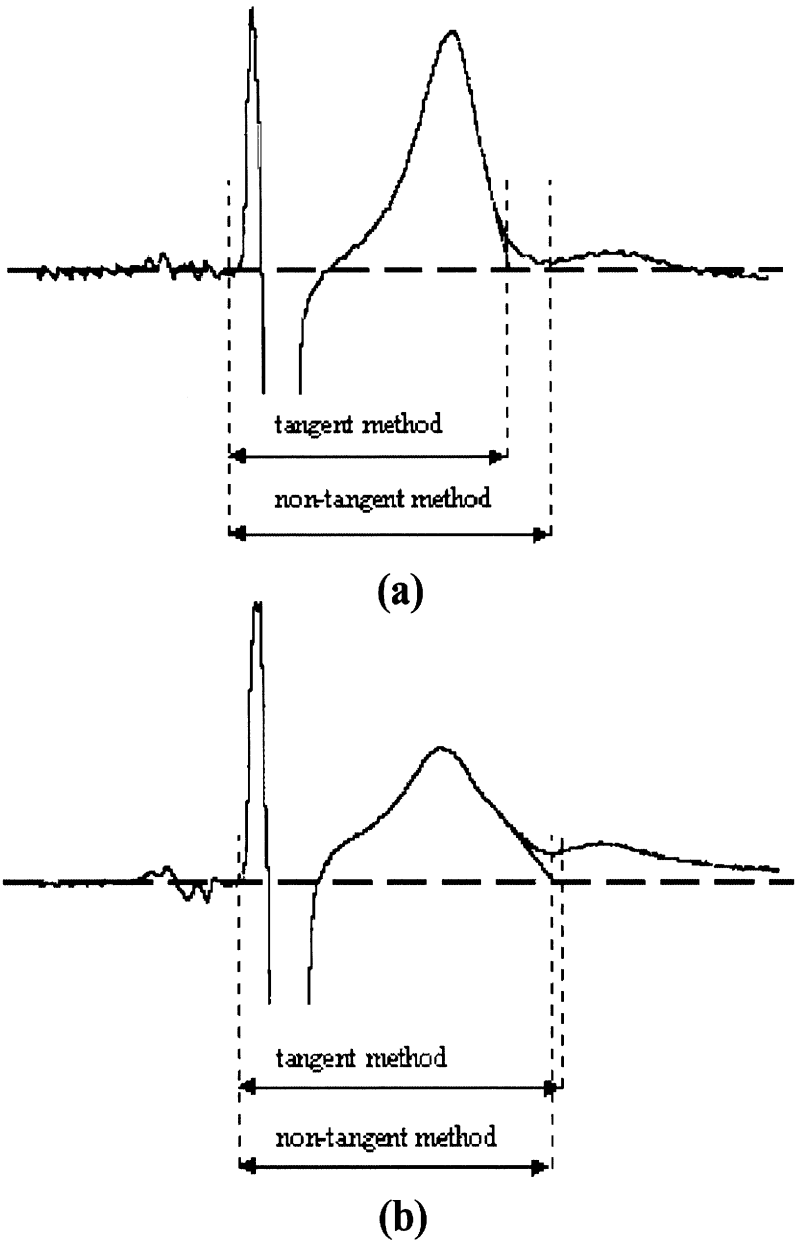

Figure 2.6: QT measurement with tangent and non-tangent methods (a) a baseline ECG (b) a hypoglycemic ECG [Ireland2000]

Another observation was performed with 18 healthy subjects and the effects of controlled hypoglycemia on cardiac repolarization especially on ECG descriptors of T wave and QRS complex morphology were evaluated. Several ECG variables characterizing repolarization were analyzed from digitized 12-lead ECG during euglycaemic and hypoglycemic clamp studies [Laitinen2008]. The results showed that hypoglycemia has marked effects on the cardiac electrical function because it prolonged QTc interval from 408 ms to 429 ms during the hypoglycemic phase.

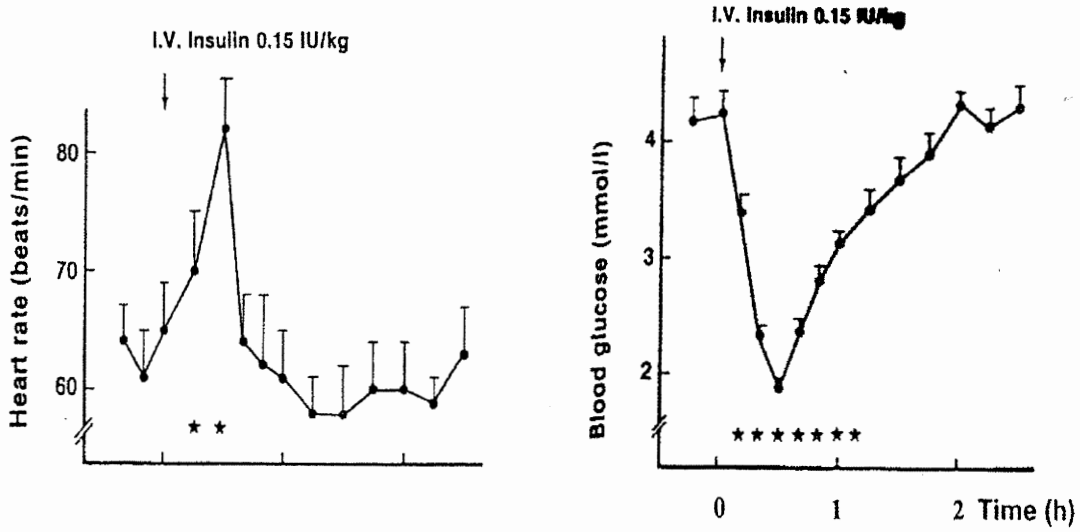

Figure 2.7: Increased heart rate during hypoglycemia [Hilsted1984]

A similar experiment was conducted [Hilsted1984] for observation of haemodynamic changes during insulin-induced hypoglycemia. Seven male subjects (ages 22 ± 0.6 years) volunteered for this experiment. In this experiment, the hypoglycemia was induced by IV injection of insulin and the cardiac output was measured. Fig. 2.7 shows the blood glucose profile and the significant increase in heart rate in response

to hypoglycemia after 30 minutes of insulin infusion. The study confirmed that the increase in heart rate was a significant occurrence during hypoglycemia.

A summary of the mean QTc lengthening during hypoglycemia is shown in Table. 2.1. It can be distantly seen that all the techniques produce a highly significant change of QTc for euglycemia and hypoglycemia conditions. Hence, the QTc prolongation during nocturnal hypoglycemia is of clinical interest importance because it may contribute to sudden death.

Subjects	**Euglycemia**		**Hypoglycemia**	
	QTc (ms)	ΔQTc (ms)	QTc (ms)	ΔQTc (ms)
8 T1DM and 7 Healthy [Marques1997]	429	18	583	162
10 Healthy [Robinson2003a]	400	20	450	70
17 Healthy [Robinson2003b]	406	16	480	90
8 T1DM [Lee2005]	407	16	448	57
17 Healthy [Ireland2000]	399	7	459	67
16 T1DM and 8 Healthy [Koivikko2007]	410	10	419	13
18 Healthy [Laitinen2008]	408	9	429	30

Table 2.1: Studies of changes in QTc during euglycemic and hypoglycemic studies

A potential clinical application of prolonged cardiac repolarization to the non-invasive detection of impending hypoglycemia at night would be a significant benefit to adults

and young children with diabetes. In Table. 2.1, QTc $(ms)^1$ measured at the end of euglycemic clamp, ΔQTc $(ms)^2$ is measured from baseline to the end of euglycemic clamp while QTc $(ms)^3$ is measured at the end of hypoglycemic clamp and ΔQTc $(ms)^4$ from baseline to the end of hypoglycemic clamp.

2.2.2 Sweating and Skin Impedances

During hypoglycemia, physiological parameter changes occur in patients due to the sympathetic nervous system, of which sweating is one of the predominant parameters. To analyze the effects of sweating in response to insulin-induced hypoglycemia, an experiment was conducted in [Ghevondian1998] to find out the relationship between blood glucose level and skim impedence/resistance. In this study, 12 volunteers (group A: 6 non-diabetic subjects and group B: 6 Type 1 IDDM patients) were monitored at regular intervals to analyze the transient behavior of sweating by effectively measuring the skin impedance.

The experimental results found that the skin impedance for group A and B were decreased to a mean of 276 ± 135 ohms and 306 ± 176 ohms from their base line values, 400 ± 96 ohms and 417 ± 22. The graph in Fig.2.8 showed that the skin impedance was lower in hypoglycemia. The study also found that the mean heart rate of healthy and diabetic subjects increased by 11 and 21 beats per minute and the increasing heart rated during hypoglycemia was confirmed. By monitoring skin impedance, the level of sweating can be measured with aid of other physiological

parameters which will detect the onset of hypoglycemia in Type 1 diabetic patients.

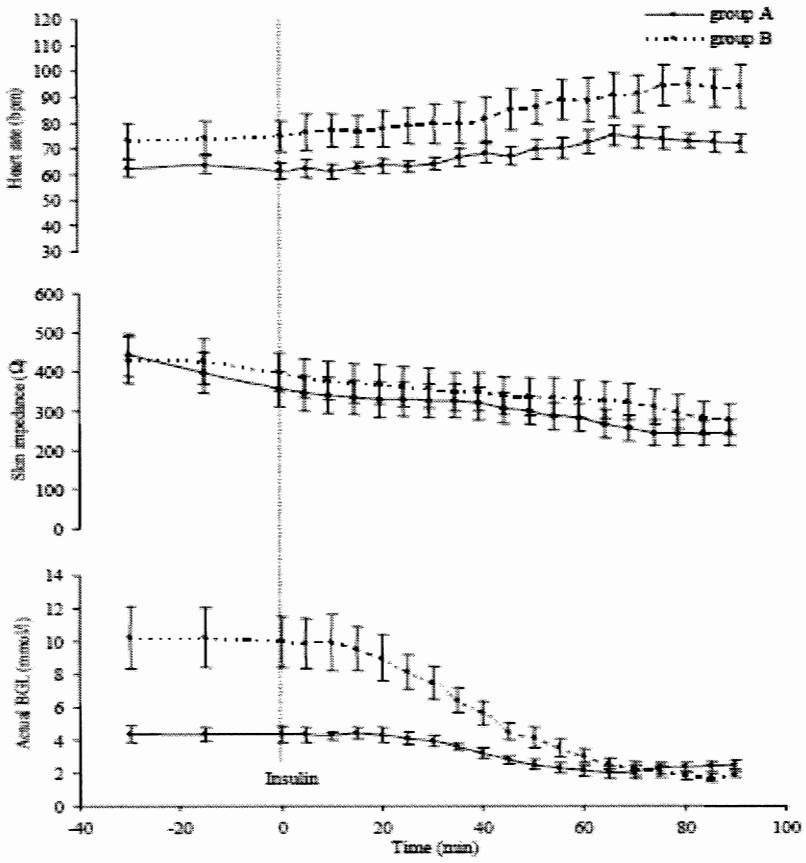

Figure 2.8: The effect of BGL on skin impedances [Ghevondian2001]

2.2.3 Electroencephalogram (EEG)

As discussed in Section 2.2, the brain depends on a continual supply of glucose and is vulnerable to any glucose deprivation. It is also one of the first organs affected by lowered blood glucose levels. Hypoglycemia develops when rates of glucose entry into the systematic circulation are continually declining, and eventually leads to neurologic dysfunction, widespread electroencephalogram (EEG) changes, cognitive dysfunction

and severe hypoglycemia. Experiments have shown that hypoglycemia can cause EEG abnormalities in subjects with and without diabetes [Pramming1988] [Harrad1985].

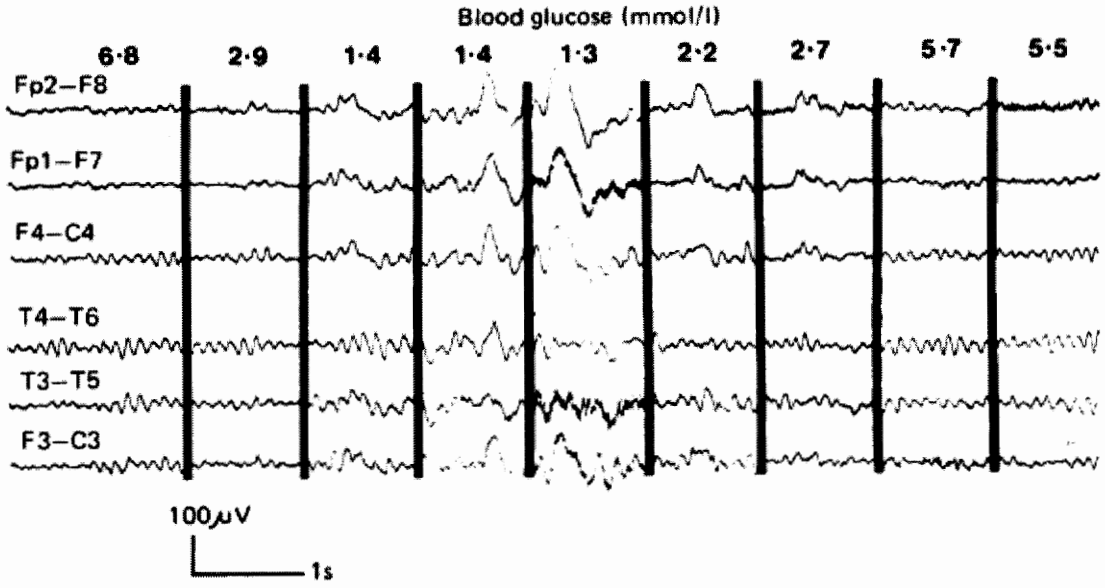

Figure 2.9: Electroencephalograms at different BGL from one patient [Pramming1988]

The effect of hypoglycemia on brain function was studied in 13 patients with insulin dependent diabetes [Pramming1988]. No changes were seen in EEG when the blood glucose concentration was above 3 mmol/l. At a median blood glucose concentration of 2 mmol/l, alpha activity decreased abruptly in the EEG concomitant with an increase in theta activity, reflecting neuronal dysfunction in the cortex. When the blood glucose concentration was further lowered, changes were observed in the EEG indicating that deeper brain structures were affected. Fig. 2.9 shows the EEG with different blood glucose concentrations for a patient.

In a similar experiment [Nguyen2010], a non-invasive hypoglycemia detection system was developed by measuring physiological responses derived from EEG. From a clinical study of six children with type 1 diabetes (T1D), associated with hypoglycemic episodes at night, their centroid (centre of gravity) alpha frequency reduced significantly ($p < 0.001$) and their centroid theta frequency increased significantly ($p < 0.02$). Using the optimal Bayesian neural network methodology, the sensitivity and the specificity was found to be 78% and 55% respectively. However, the corresponding specificity for hypoglycemia detection system was still low. Therefore, continuing research is still aiming to develop advanced intelligent algorithms for improving the overall accuracy of the hypoglycemia monitoring system.

2.3 Existing Technologies for Hypoglycemia Detection

Real-time continuous glucose monitoring has the potential to overcome diabetic complications and increase the likelihood of patients with diabetes by maintaining optimal glucose level without symptomatic hypoglycemia. In clinical practice recommendations, it has also been suggested that continuous glucose monitoring is especially useful in patients with Type 1 diabetes in order to detect hypoglycemia unawareness and/or frequent episodes of hypoglycemia. With ever improving advances in diabetes diagnostic technology, different blood glucose monitoring systems, invasive, minimally invasive and non-invasive are tested and introduced by the use of various techniques

with the choice of sample region as described in Fig. 2.10.

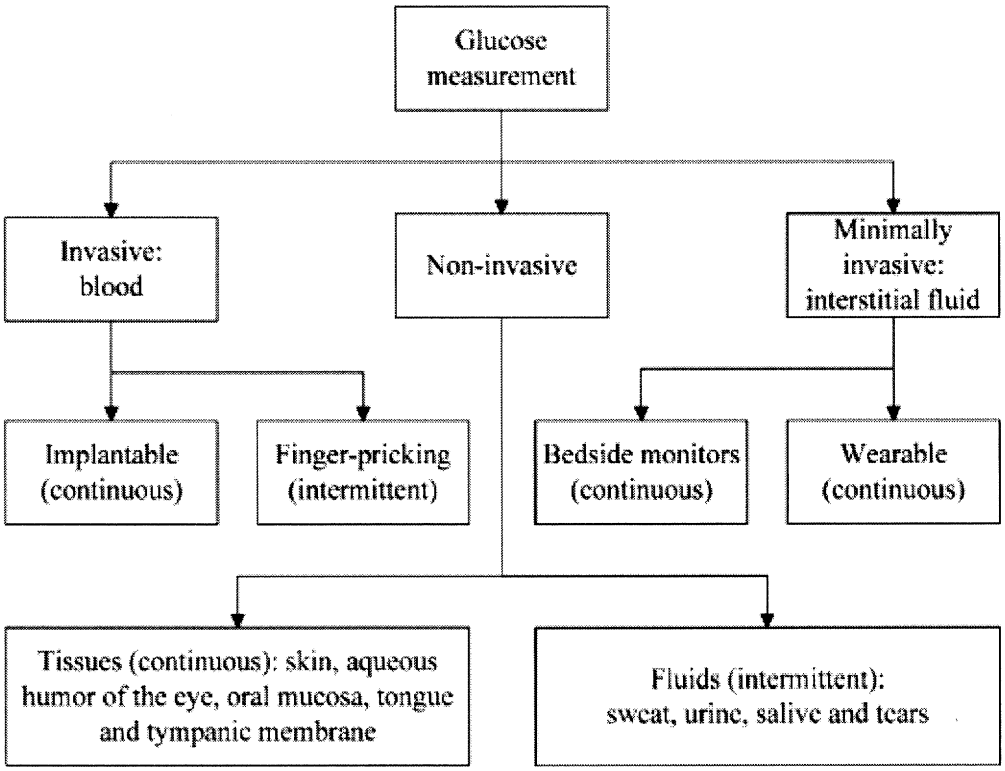

Figure 2.10: Overview of technologies for non-invasive blood glucose control: invasive, minimally invasive and non-invasive [Amaral2008]

2.3.1 Invasive Techniques

In the past few years, glucose meters in Fig.2.11 have been the method of choice for the measurement of blood glucose concentration for patients with T1DM. They use automatic lancet devices to prick the fingertip to take the blood sample, which is painful as the diabetic has to measure blood glucose very frequently i.e. more than

four times a day. Efforts have been made with reduced blood sample volume to a few microliters and minimized the painful aspects of piercing the fingertip by employing alternate sampling sites such as hand, upper arm, forearm or thigh with the use of 32 gauge lancet.

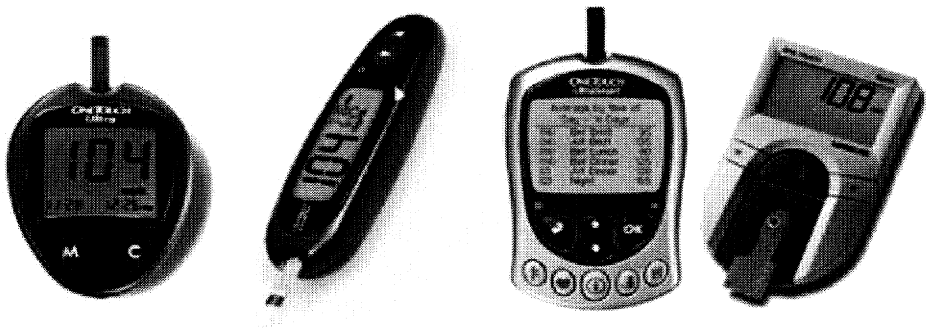

Figure 2.11: A number of different electrochemical glucose meters [Clarke2012]

However, the cost of strip and the boredom of making repeated measurements become the great barriers for Type 1 diabetic patients who frequently need to monitor BGL as well as episodes of hypoglycemia. Sometimes, the manual monitoring of blood glucose increases the risk of serve hypoglycemia due to their low efficiency in detecting frequent episodes of hypoglycemia. Due to discomfort and inconvenience of finger-stick methods, the minimally invasive approaches which sample the interstitial fluid (ISF) with subcutaneous sensors have been developed by using subcutaneous sensors to determine glucose concentration in interstitial fluid [Block2008]. But they suffer from limitations in terms of discomfort to patients, requirement of continuous calibration, and high susceptibility to biofouling. So far there are no reports or patents which show that such minimally invasive methods have improved accuracy as compared to

invasive procedures.

2.3.2 Non-invasive Techniques

The development of non-invasive continuous glucose monitoring systems (NCGM) is the only way for achieving painless control of blood glucose level and improving life quality of diabetes patients with better regulation of hypoglycemia episodes. Following with the advanced technologies such as reverse iontophoresis, fluorescence, ultrasound, electromagnetic sensing, bioimpedance, raman, oscular near-infrared and mid-infrared spectroscopy, non-invasive investigations have been carried out together with the choice of sample regions, sweating, skin color, surface roughness, tissue thickness, breathing artifacts, blood flow, body movements, ambient temperature which are sensitive for blood glucose measurement [Vashist2012].

The most promising instrument incorporating reverse iontophoresis technique was GlucoWatch G2 biographer in Fig. 2.12 which is in the form of a wrist-watch from Cygnus Inc., California, USA. In this device frequent measurement of blood glucose information was provided by extraction of interstitial fluid through the skin. This was done by conducting a constant low-level electric current (300μA) through the skin between two electrodes, anode and cathode, thereby migrating sodium and chloride ions from beneath of the skin to the cathode and anode respectively. The detection of glucose also takes into account the skin temperature and perspiration fluctuations by employing thermo transducers and conductivity sensors in the device.

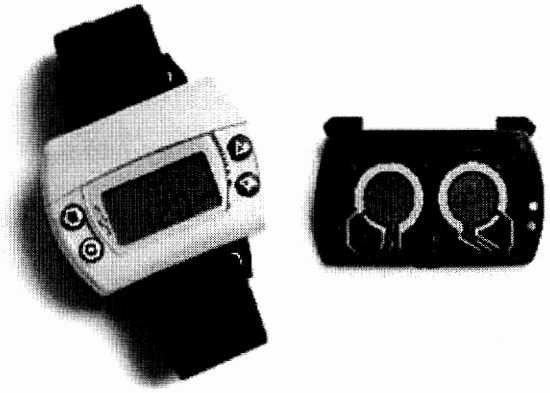

Figure 2.12: GlucoWatch biographer: non-invasive glucose monitoring device [Vashist2012]

However, some limitations such as the requirement of calibration using a standard blood glucose meter, replacement of disposable pad every 12 hours, taken for the warm-up period, inaccurate glucose measurements if the patient is moving, exercising, sweating or having rapid temperature changes, come along with the GlucoWatch G2 biographer. And also the device has also been found to cause skin irritation that becomes the most problematic factor for younger children where it was used for nocturnal monitoring. The device fails to detect hypoglycemia when it automatically shuts down in cases of sweating. Since sweating is a symptom of hypoglycemia, the device shuts down precisely when continuous monitoring is most needed. The clinical studies showed that the device performs better at high glucose levels but is not reliable in detecting low blood glucose levels [Tsalikian2004].

Another spectroscopical investigation has been done for OrSense NBM-200G device

from OrSense Ltd., Israel [Tierneya2001]. The device employs red NIR occlusion spectroscopy, which is based on detecting the red NIR optical signal of blood due to changes in the glucose concentration in blood vessels of the finger. The device is integrated with wireless telemetry and measures glucose in less than a minute. It also has the internal memory to store up to 500 readings. It measures glucose continuously for up to 24 hours and does not require frequent calibrations. It is completely safe for patients without any risk of contamination. However, the system has not been commercialized and is being utilized for investigation and market awareness purposes only.

On the other hand, a real-time continuous monitoring device, GlucoTrackTM was developed by Integrity Applications Ltd., Israel. This device determines the blood glucose concentration in the earlobe with the help of a personal ear clip (PEC) equipped with sensors and calibration electronics. The three non-invasive techniques i.e. ultrasonic, electromagnetic and heat capacity were employed for glucose measurements. The device comes with USB and IR connectivity for battery recharge, software for data processing, internal memory for storing up to 1000 readings per user and alerts for hypo- and hyperglycemia. But it has not been commercialized according to the need for improved performance in calibration and algorithm employed for data processing.

For measurement of blood glucose level as well as detection of hypoglycemia, an

attempt has been made to develop a wristwatch-like device, the Teledyne Sleep Sentry [Hansen1983]. It is similar to GlucoWatch G2 biographer, capable of detecting falling skin temperature as well as skin resistance during sweating. The efficacy and credibility of Teledyne Sleep Sentry for detecting hypoglycemia was studied during night-time with 22 adult insulin-treated diabetics [Johansen1986]. The evaluation results for functional performance showed that Sleep Sentry was able to detect BGL less than 3 *mmoll* with sensitivity of 67 and specificity of 27 %. However, using only skin impedance as the dominant input source becomes the considerable factor for modeling and estimation of actual blood glucose. The Teledyne Sleep Sentry cannot overcome the issues of skin impedance such as considerable time lag and non-liner characteristics of skin in response to hypoglycemia. Due to its poor reliability, low sensitivity and specificity in detecting hypoglycemia, further improvement for the device, especially the technique that is employed within the device is still needed.

Table 2.2 summarizes current non-invasive blood glucose monitors at prototype stage or if commercially available. In clinical practise recommendations, it has been suggested that the CGMS in Table 2.2 are useful for patients with hypoglycemia unawareness and/or frequent episodes of hypoglycemia. However, the hypoglycemia preventive effect of non-invasive continuous glucose monitoring have not been successfully established in reality.

CGMS Devices	Company	Target Site	Technology Employed
GlucoWatch	Cygnus Inc.	Wrist skin	Reverse iontophoresis
GlucoTrack	Integrity Applications Ltd.	Ear lobe skin	Ultrasound and electromag
OrSense NBM-200G	OrSense Ltd.	Fingertip skin	Occlusion NIR spectroscopy
Pendra	Biovotion AG	Wrist skin	Bioimpedance spectroscopy
Diasensor	Biocontrol Technology Inc.	Forearm skin	NIR spectroscopy
Glucoband	Calisto Medical Inc.	Wrist skin	Impedance spectroscopy
SugarTrac	LifeTrac Systems Inc.	Skin	NIR spectroscopy sensing
GluCall	KMH Co. Ltd.	Skin	Reverse iontophoresis
TouchTrak Pro 200	Samsung Fine Co. Ltd.	Fingertip skin	Electromagnetic sensing
Glu Control GC300	ArithMed GmbH Co. Ltd.	Fingertip skin	Electromagnetic sensing
Sleep Sentry		Skin temperature	Electric Current

Table 2.2: Current non-invasive blood glucose/hypoglycemia monitors

2.4 Intelligent Detection of Hypoglycemia Using Physiological Parameters (Non-invasive Technique)

Although the recent technological advances in Section 2.3.1 offer a certain accuracy in detection of hypoglycemia, there are still formidable technological problems to overcome. Since the detection of hypoglycemia is an important clinical problem, a reliable means of detecting hypoglycemia is still needed for maintaining optimum glycemic control. The ideal solution would be the development of non-invasive hypoglycemic monitors using intelligent technologies and physiological parameters of ECG signal.

Algorithms	Parameters
Fuzzy reasoning model [Ghevondian1997][Hastings1998]	HR
Fuzzy neural network [Ghevondian2001]	HR, QTc
Neural network [Nguyen2006]	HR, QTc
Bayesian neural network [Nguyen2007] [Nguyen2008]	HR, QTc
Neural network based rule discovery system [Chan2011]	HR, QTc
Support vector machine [Nuryani2012]	HR, QTc, TpTec, ToTec, RTpc, QTpc
Genetic algorithm based fuzzy reasoning model [Ling2011]	HR, QTc

Table 2.3: ECG parameters and intelligent methods for hypoglycemia detection

The development of non-invasive hypoglycemia detection system with a base of ECG

parameter changes in response to hypoglycemia is listed in Table 2.3. In the next sub-sections, the research works in fuzzy logic, neural networks and their applications to the hypoglycemia detection system be briefly reviewed.

2.4.1 Fuzzy Reasoning Model

By the use of the fuzzy reasoning algorithm, a portable microcontroller-based hypoglycemia monitoring system was developed based on the observation of physiological parameters such as sweating, snoring, heart rate (HR), ECG [Ghevondian1997] [Hastings1998]. In the system, fuzzy logic was used as the estimation algorithm due to its linguistic properties and its non-linear properties. It served as a software engine in a portable hypoglycemic monitoring device that has been developed in Fig. 2.13.

The system commences with the fuzzification process of the four physiological parameters using the triangle membership function which corresponds to three linguistic variables: normal, high and very high. By exploiting fuzzy decision making ability and generating a set of linguistic rules, the index of hypoglycemia is predicted through the defuzzification process.

In fuzzy-based hypoglycemia detection system [Ghevondian1997] [Hastings1998], the membership functions which express the linguistic terms for fuzzy inference rules have to be defined. However, more parameters need to be determined for fuzzy membership functions and rules. Larger memory, more computational time and learning data

are required to develop FIS based hypoglycemia detection. Even fuzzy rules which represent certain information from the models can be generated, the domains of inputs and outputs represented by the fuzzy rules are not systematic. Therefore, it can be difficult to make diagnosis decisions based on those fuzzy rules.

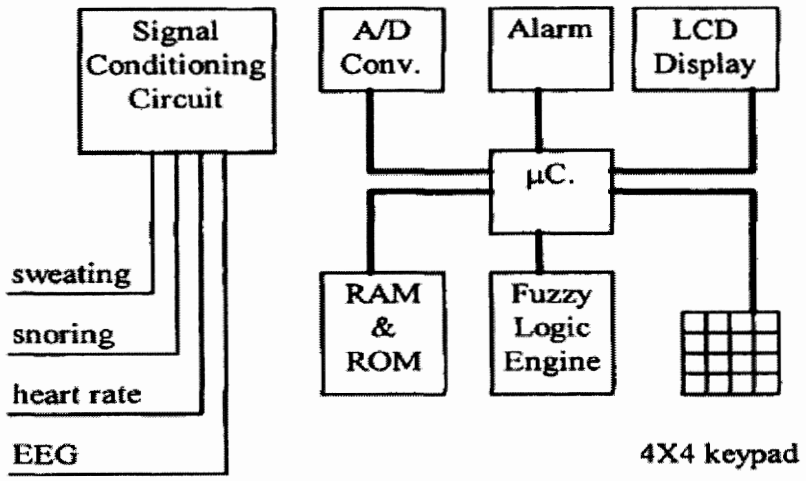

Figure 2.13: Block diagram for monitoring hypoglycemia in diabetic patients using fuzzy reasoning model [Ghevondian1997]

2.4.2 Fuzzy Neural Network

For predicting onset of hypoglycaemia in insulin-induced subjects, a more comprehensive study was continued by the use of a novel fuzzy neural network estimator algorithm (FNNE) [Ghevondian2001]. In this analysis, hypoglycaemia was firstly induced in 12 volunteers (group A: 6 non-diabetic subjects and group B: 6 Type 1 IDDM patients) using insulin infusion, and skin impedances, heart rates and actual blood glucose levels (BGL) were monitored at regular intervals. By analyzing the

static behavior of the measured physiological parameters of all 12 subjects, the mean heart rate of group A and group B was increased while skin impedance was decreased.

With these significant variations, the FNNE algorithm was employed based on a set of first order estimation functions for estimating the BGL profiles and consequently detect the onset of hypoglycemia episodes. In [Ghevondian2001], the FNNE model is constructed by distributing the input and output relationships to weights connecting neurons. The error is limited to a reasonable level via sample training and used for modification of each weight value to acquire the final weight value for connections between neurons. Despite this the approach is suitable for hypoglycemia detection, the accuracy of FNNE-based detection system is low and further experiments are still needed.

2.4.3 Neural Networks

Considering the correlation between physiological parameters and the status of hypoglycemia, a non-invasive hypoglycemia monitor named as HypoMon was developed from AIMedics Pty Ltd by the use of Bayesian neural network algorithm [Nguyen2006] [Nguyen2007]. In neural network based hypoglycemia monitoring systems, the physiological parameters such as heart rate (HR), corrected QT interval (QTc) and skin impedance are used as the main inputs and the status of hypoglycemia is detected through the variations of those input parameters. The developed hypoglycemia monitor, HypoMon consists of a battery-powered chest belt worn and was used to measure

the relevant physiological responses, while the actual blood glucose (BG) measurements were collected as reference. The sensitivity obtained by the hypoglycemia detection neural network is acceptable, but the overall accuracy still needs to be improved.

Furthermore, a neural network based rule discovery system which consists of a neural network based classification unit and a rule based extraction unit is proposed to perform diagnosis of hypoglycemic episodes in T1DM patients [Chan2011]. In this method, the neural network based classification unit is used for determining hypoglycemic episodes in T1DM patients using the specified physiological parameters while a set of rules which describe the domains of physiological parameters for which hypoglycemic episodes occur is extracted from the neural network classification unit by a rule based extraction unit.

The development was based on 420 data sets which were collected from 16 T1DM patients by using the genetic algorithm. Experimental results show that the proposed neural network based rule discovery system can achieve reasonable sensitivity and specificity. Apart from obtaining reasonable sensitivity and specificity, the neural network used in the developed detection system might not be able to improve further because its processing capability and adaptability were limited by fixed neural network structure.

In general, the neural network with fixed structures might not be able to scale up

to meet with the requirements because the processing capability of the network is restricted by the number of hidden neurons in the hidden layers. The need for better sensitivity and specificity will only be achieved if the different structures of neural network are designed for different characteristics of application. In other words, it will be more appropriate if the flexible structures of neural networks are employed based on the characteristics of application for good performance. Thus, in the direction of achieving a truly non-invasive hypoglycemia detection system, the flexible neural network structure that will be breaking the barriers of fixed structure become the most desirable step.

2.4.4 Swarm-based fuzzy support vector machine (SVM)

Another interesting method to detect hypoglycemia non-invasively is by means of physiological effects of hypoglycemia. In [Nuryani2012], a novel hypoglycemia detection strategy using swarm-based fuzzy support vector machine technique is developed. The proposed hypoglycemia detector system is a combination of two subsystems, namely fuzzy inference system (FIS) and support vector machine (SVM). Two most significant ECG parameters such as HR, QTc, $TpTe_c$, $ToTe_c$, RTp_c, and QTp_c are fed to SVM and classified to indicate the presence of hypoglycemia.

Analysis also showed that three and five membership functions are used for FIS while radial basis function (RBF), sigmoid and linear kernel functions are employed for mapping the inputs to high dimensional space in SVM. The performances of

fuzzy SVM with different kernel functions are found to be satisfactory. However, the disadvantages of SVM algorithm only covers the determination of the parameters for a given value of the regularization, kernel parameters and choice of kernel. In a way the SVM moves the problem of over-fitting from optimizing the parameters to model selection. Sadly kernel models can be quite sensitive to over-fitting the model selection criterion. Hence, a further study still needs to be conducted by the use of other computational intelligence technologies.

2.4.5 GA-based Fuzzy Reasoning Model

Following the methodology of hybridization, genetic algorithm based fuzzy reasoning model [Ling2010] [Ling2012] and statical multiple regression with fuzzy inference system [Ling2011] was developed based on heart rate (HR), corrected QT interval (QTc), change of HR (ΔHR) and corrected QT interval (ΔQTc).

In these two subsystems, the first subsystem, FIS plays a main role to approximate the correlation between the physiological parameters of ECG signal such as HR and QTc and the approximated hypo-index by using fuzzy rules. The larger approximated output gives the higher possibility of hypoglycemia. Based on the estimated hypo-index and the change of the HR and corrected QT interval, the other subsystem called multiple regression model is used to classify the presence of hypoglycemia. The global searching algorithm of GA is used to find the optimal fuzzy rules and membership functions of FIS and the model parameters of regression method.

The MR based FIS system [Ling2011] gives a significant result with less model parameters when detecting hypoglycemia episodes noninvasively and continuously from the real-time physiological responses in TIDM children. However, this model still lacks systematic procedures for defining fuzzy membership functions and rules; the tuning of those parameters are a time consuming task. By addressing these issues, an automatic adaption procedure, which is comparable to neural network structure is desirable and it is regarded as on-going research for the development of a non-invasive hypoglycaemia detection system.

2.5 The Proposed Methodologies for Hypoglycemia Detection

The proposed advanced neural network based hypoglycemia detection system is a type of consumer electronic device embedded in the digitized chest belt which encrypts and transmits the measured physiological parameters (heart rate and corrected QT interval) to a receiver computer using a wireless communication link. The collected data are applied to the proposed swarm based neural network classifiers (Fig. 2.14) to determine the level of hypoglycemia. An alarm sound is activated when the status of hypoglycaemia is detected thus warning the patient or physician in critical situations.

To monitor the status of hypoglycemic episodes in T1DM patients, advanced neural

network classifiers with 4 inputs and 1 output system is developed. The four psychological inputs are heart rate (HR) and corrected QT interval (QTc), change of heart rate and corrected QT interval (ΔHR and ΔQTc) and the output is the presence of hypoglycemia (h), $+1$ represents hypoglycemia and -1 is non-hypoglycemia. In this thesis, three different advanced neural network classifiers namely variable translation wavelet neural network (VTWNN), multiple regression model based combinational neural logic network (MR-NLN) and rough-block-based neural network (R-BBNN) are proposed.

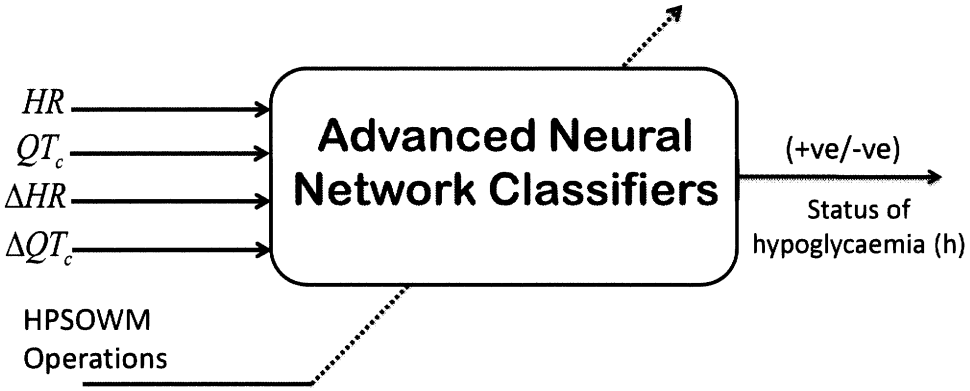

Figure 2.14: Proposed advanced neural networks based hypoglycemia detection system

The first proposed algorithm, VTWNN has adaptive network structure with its variable translation parameter and is excellent in capturing the nonstationary nature of ECG signal. The second MR-NLN is the knowledge based neural network that understands the characteristics of practical application and is suitable for optimal

performance. For the third R-BBNN algorithm, the rough set properties are introduced to the flexible network structure of BBNN which is enable to evolve internal structures and adaptability in dynamic environments.

In order to evaluate the performance of advanced neural network classifiers in detecting hypoglycemia, the two physiological parameters such as HR and QTc are used as the main inputs while ΔHR, ΔQTc are considered as additional inputs for improving sensitivity and accuracy of the proposed detection system. The effectiveness of proposed algorithms, VTWNN, MR-NLN and R-BBNN in hypoglycemia detection will be briefly discussed in the following Chapters 3, 4 and 5 respectively.

Chapter 3

Optimized Variable Translation Wavelet Neural Network For Non-Invasive Hypoglycemia Detection

3.1 Introduction

Neural networks (NNs) [Windrow1990] are useful tools for modeling and find the relationship between the inputs and outputs of complicated systems. The models are typically obtained based on the input-output data available for training and testing.

However, the structure of NN is an important design concern for good system modeling. Conventionally, the NN structure is determined based on expert knowledge to the target system, or is a fixed one assumed to be complicated enough to offer a good model, but the performance of NN cannot be guaranteed.

Currently, a fixed network structure with a fixed set of weights are widely applied in areas such as load forecasting, classification, system modeling and control. Those network models offer fixed structure and a single set of weights after training process that may not be good enough if data sets are distributed in a vast domain separately and/or the number of network parameters is too small.

To handle this kind of problem, variable translation wavelet neural network (VTWNN) is firstly presented. In VTWNN, the wavelets are used as transfer functions in the hidden layer of the network. The network parameters, such as the translation parameters of the wavelets are variable depending on the network inputs. Due to the variable translation parameters, the proposed VTWNN has the ability to model the inputoutput function with input-dependent network parameters. It works as if several individual neural networks are handling different sets of input data. Effectively, it is an adaptive network capable of handling different input patterns and exhibits a better performance.

One of the most important issues on neural network is learning or training because the optimal set of network parameters is obtained through the learning process. An

improved hybrid PSO which incorporates a wavelet-theory-based mutation operation is used to train the parameters of the proposed VTWNN. By applying wavelet theory in PSO, it can enhance the searching performance exploring more of the solution space to reach a better solution. A study of VTWNN for the detection of hypoglycemic episodes in type 1 diabetes mellitus (T1DM) will be given. It was found that the variable structure characteristic can lead to better results with increased learning ability than the conventional fixed neural network structures.

The chapter is organized as follows. In Section 3.2, the proposed VTWNN with its operation principle, network structure and parameter design is presented. The training of the parameters of the proposed network using HPSOWM will be presented in Section 3.3. In Section 3.4, an application of VTWNN to hypoglycemia detection in Type 1 diabetes mellitus (T1DM) is given to validate the effectiveness of the proposed network. Finally, the summary is given in Section 3.5.

3.2 Variable Translation Wavelet Neural Network (VTWNN)

The wavelet neural network is considered as a particular case of feedforward neural network and the neural network using wavelet basis function can provide faster convergence rates for approximation compared with conventional feedforward neural network. It has been applied in many research areas because of its excellent property in time-frequency localization of a given signal [Zekri2008] [Mallat1989].

By combining wavelet theory and neural network, wavelet neural network (WNN) has been developed in order to give better performance in function approximation and learning capabilities [Billings2005]. However, a conventional WNN fails to capture the characteristics of separately distributed input data. To overcome this, VTWNN with multiscale wavelet function is proposed in which the translation parameter (b) is depending on the neural network inputs. It offers the ability to model the input-output function as the input-dependent network parameters.

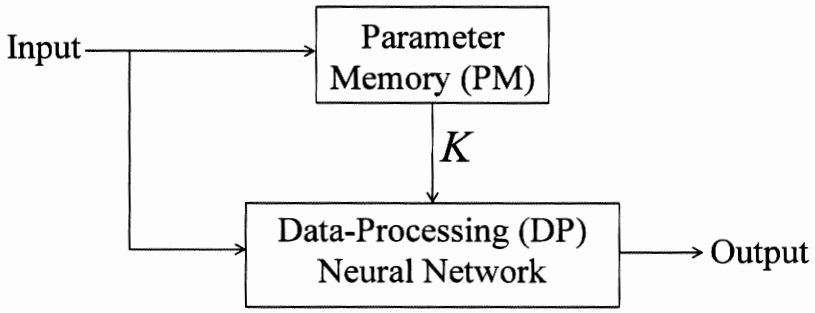

Figure 3.1: Proposed architecture of the neural network [Ling2008c]

Fig. 3.1, the architecture of the proposed VTWNN, which consists of two units, namely, the parameter memory (PM) and the data processing (DP) neural network. The PM stores some parameters (K) governing how the DP neural network handles the input data. By using this proposed neural network, some of the cases that cannot be handled by conventional neural networks with a limited number of parameters can now be tackled. For instance, it two set of data, $S1$ and $S2$ are separately distributed far from each other, the neural network with fixed structure can only model the data set S as shown in Fig. 3.2.

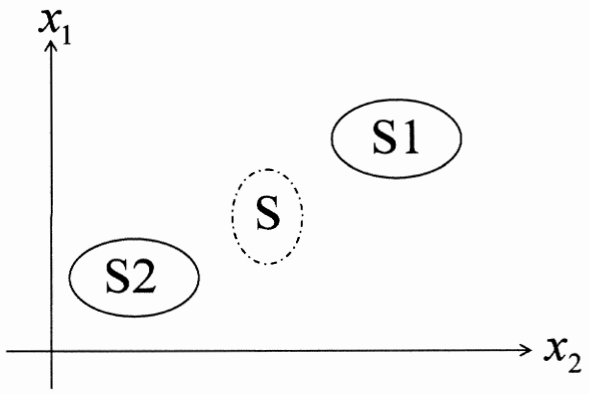

Figure 3.2: Diagram showing two sets of data in spatial domain [Ling2008c]

In the proposed VTWNN, if the input data belongs to $S1$, the PM will follow parameter set to drive the DP neural network to handle $S1$ data. Similarly, when the input data belongs to $S2$, the parameters corresponding to $S2$ will be employed to drive the DP neural network to handle these input data.

3.2.1 Basis Wavelet Theory

Certain seismic signals can be modeled by combining translation and dilation of an oscillatory function within finite duration called wavelet. A continuous function $\psi(x)$ is called a mother wavelet or wavelet if it satisfies the following properties:

Property I:

$$\int_{-\infty}^{+\infty} \psi(x)\, dx = 0 \tag{3.1}$$

In other words, the total positive momentum of $\psi(x)$ is equal to the total negative

momentum of $\psi(x)$. On the other hand, it is possible to show that the admissibility condition implies that $\hat{\psi}(0) = 0$, so that a wavelet must integrate to zero. Notice that ψ is the Fourier transform of wavelet ψ and the admissibility condition is defined as:

$$0 < C_\psi < +\infty \tag{3.2}$$

where

$$C_\psi = \int_{-\infty}^{+\infty} \frac{\left|\hat{\psi}(x)\right|^2}{|\nu|} d\nu \tag{3.3}$$

Property II:

$$\int_{-\infty}^{+\infty} |\psi(x)|^2 dx < \infty \tag{3.4}$$

where most of the energy of $\psi(x)$ is confined to a finite domain and is bounded. The Morlet wavelet in Fig. 3.3 [Daubechies1992] is an example of the mother wavelet which is expressed as follows:

$$\psi(x) = e^{\frac{-x^2}{2}} \cos(5x) \tag{3.5}$$

According to **Property I**, the Morlet wavelet integrates to zero while over 99% of the total energy of the function is contained in the interval of $-2.5 \leq x \leq 2.5$. In order to control the magnitude and position of $\psi(x)$, $\psi_{a,b}(x)$ is defined as:

$$\psi_{a,b}(x) = \frac{1}{\sqrt{a}} \psi\left(\frac{x-b}{a}\right) \tag{3.6}$$

where a is the dilation (scaling) parameter, and b is the translation (shifting) parameter. It should be noted that

$$\psi_{1,0}(x) = \psi(x) \tag{3.7}$$

As

$$\psi_{a,0}(x) = \frac{1}{\sqrt{a}}\psi\left(\frac{x}{a}\right) \tag{3.8}$$

it follows that $\psi_{a,0}(x)$ is amplitude-scaled version of $\psi(x)$.

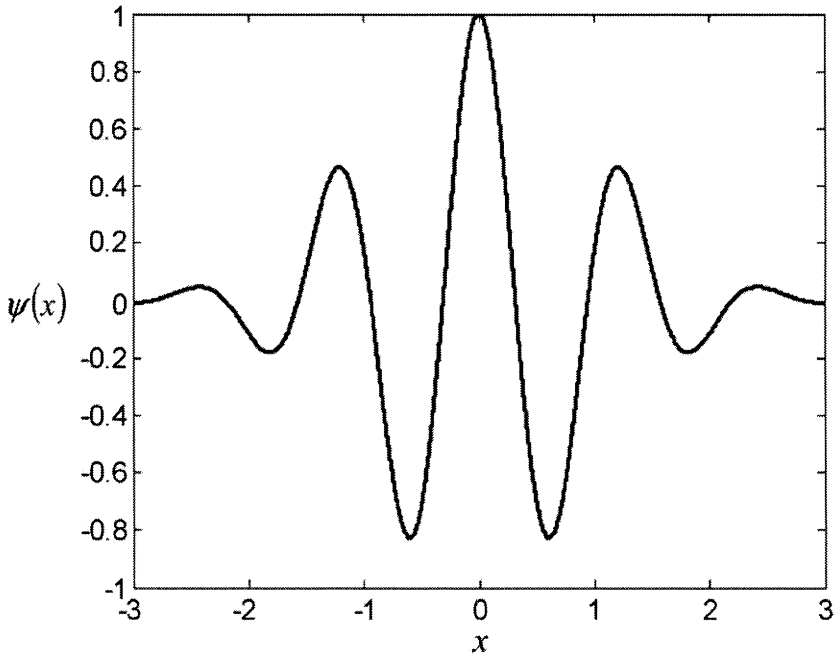

Figure 3.3: Morlet wavelet [Ling2008]

In Fig. 3.4, different dilations of Morlet wavelet are explained. The amplitude of $\psi_{a,0}(x)$ will be scaled down as the dilation parameter a increases. This property is used to do the mutation operation to enhance the searching performance.

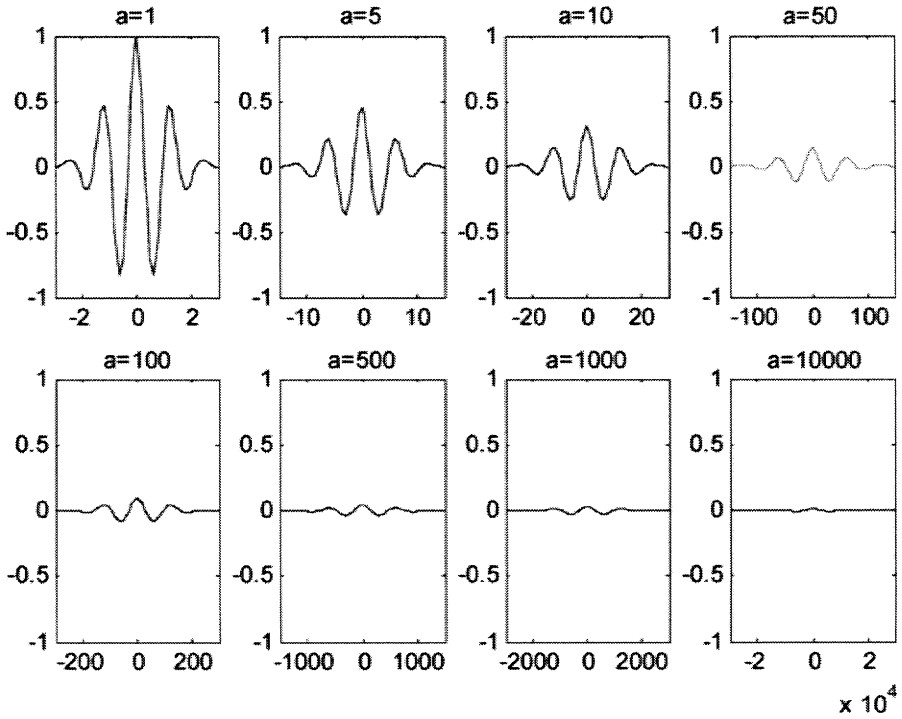

Figure 3.4: Morlet wavelet dilated by parameter a [Ling2008]

3.2.2 Architecture of VTWNN

The detail design and analysis of the variable translation wavelet neural network (VTWNN) will be discussed in this section. In the proposed VTWNN, the translation parameter, b is varying based on the neural network inputs. With the variable translation parameter b, the proposed VTWNN gives faster learning ability with better generalization compared with other conventional neural networks.

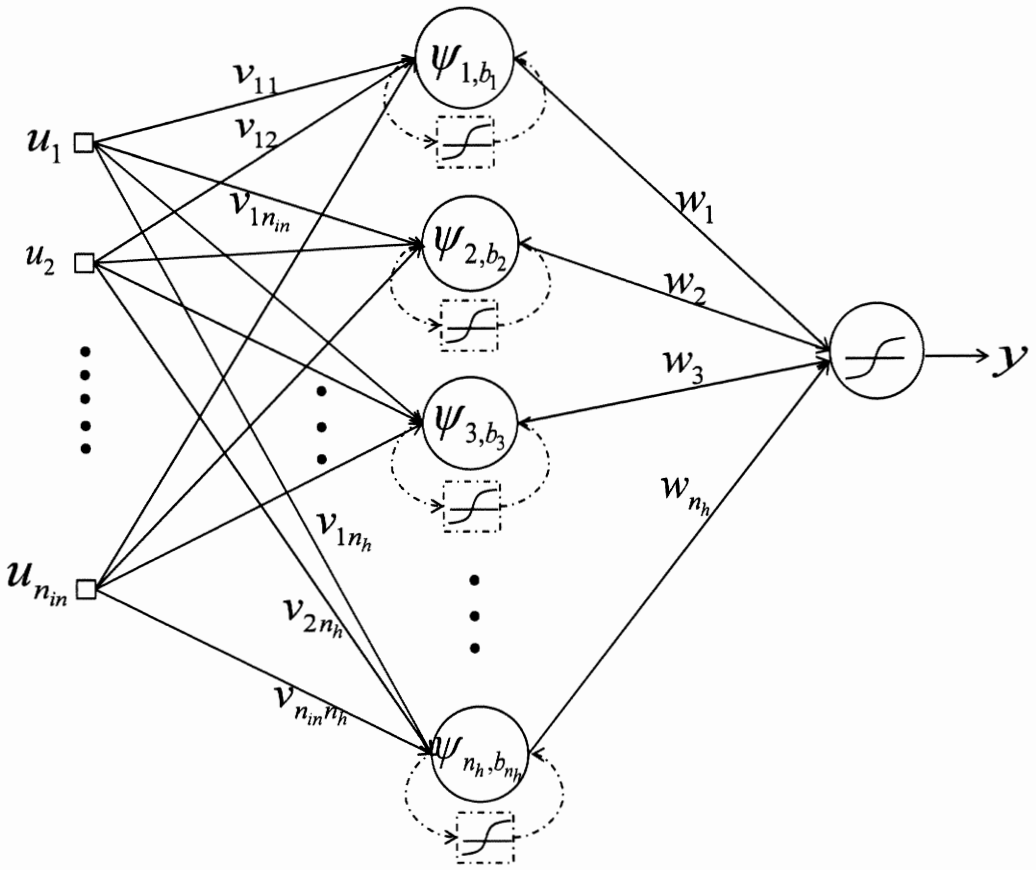

Figure 3.5: Structure of Variable Translation Wavelet Neural Network

The proposed VTWNN has a three-layer structure with n_{in} nodes in the input layer, n_h nodes in the hidden layer and n_{out} in the output layer as presented in Fig. 3.5. The input of hidden layer is calculated by:

$$S_j = \sum_{i=1}^{n_{in}} u_i v_{ij}, \quad j = 1, 2, \ldots, n_h \tag{3.9}$$

where $u_i (i = 1, 2, \ldots, n_{in})$ are the inputs, and v_{ij} denotes the weight between ith input and jth hidden nodes.

In order to control the magnitude and position of wavelet, multiscaled wavelet function, $\psi_{a,b}(x)$ in (3.6) is used as the hidden node transfer function. For instance, at the first and second hidden nodes at $j = 1, 2$, the dilation parameter a is set 1 and 2 and (3.6) becomes:

$$\psi_{1,b_1}(x) = \psi(x - b_1) \tag{3.10}$$

$$\psi_{2,b_2}(x) = \frac{1}{\sqrt{2}}\psi\left(\frac{(x - b_2)}{2}\right) \tag{3.11}$$

In (3.11), the output of the wavelet is scaled down by $1/\sqrt{2}$. Similarly, for the jth hidden node, the dilation parameter a is set as j and the VTWNN output of hidden layer is given by:

$$\psi_{j,b_j} = \frac{1}{\sqrt{j}}\psi\left(\frac{S_j - b_j}{j}\right) \tag{3.12}$$

In this proposed network, the Maxican Hat function in Fig. 3.6 is used as mother wavelet, $\psi(x)$ is denoted as follows:

$$\psi(x) = \left(1 - x^2\right)\exp\left(-\frac{x^2}{2}\right) \tag{3.13}$$

Different kinds of mother wavelets such as Mexican hat wavelet (normalized), Mexican hat wavelet have been considered. By trial and error through experiments, it was found that all wavelet functions are good, and there is not much significant difference in terms of the cost values among all wavelet functions. Finally, Mexican hat wavelet is used as the mother wavelet in wavelet neural network while Morlet wavelet is considered as the mother wavelet in wavelet mutation operation of HPSOWM.

From (3.12) and (3.13), the hidden nodes in the hidden layer of VTWNN is obtained by:

$$\psi_{j,b_j} = \frac{1}{\sqrt{j}}\left(1 - \left(\frac{S_j - b_j}{j}\right)^2\right) \exp\left(-\frac{\left(\frac{S_j - b_j}{j}\right)^2}{2}\right) \tag{3.14}$$

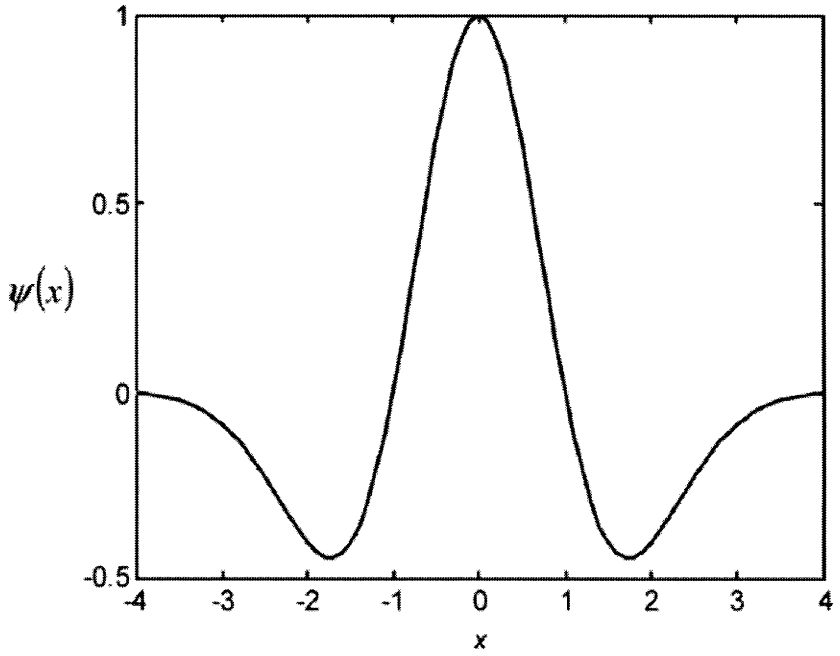

Figure 3.6: Maxican Hat Mother Wavelet

The translation parameter, b_j is depending on the input S_j and is governed by a nonlinear function $f^j(\cdot)$ as follows:

$$b_j = f^j(S_j) = 4j\left(\frac{2}{1 + e^{-K_j \times S_j}} - 1\right) \tag{3.15}$$

where K_j is a tuned parameter that is used to control the shape of nonlinear function, $f^j(\cdot)$. The value of K_j is selected from interval [0.3 1.5]. The value of K_j should not be too small or too large. It behaves as a threshold function when $K \to \infty$, and

it becomes a constant line when $K \to -\infty$. The effect of parameter K_j on the characteristics of nonlinear function $f^j(\cdot)$ in (3.15) is shown in Fig. 3.7.

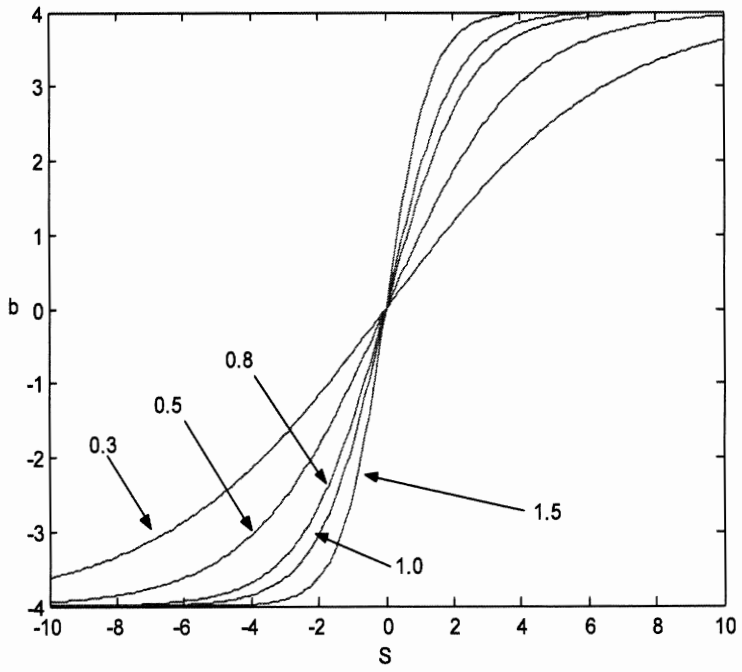

Figure 3.7: Nonlinear function with different values of parameter K

From (3.15) the value of translation parameters b_j is depending on the network related input S_j and the nonlinear function parameter K_j. By doing so, the neural network is handling with variable translation parameter b_j with respect to neural network input parameter S_j in (3.9). Because of tunable nonlinear function parameter K_j, the network increases flexibility and adaptability and provide better classification

performance. The output of proposed VTWNN is calculated by (3.9)-(3.15):

$$y_l = \sum_{j=1}^{n_h} \psi_{j,b_j}(S_j).w_{jl} \tag{3.16}$$

$$= \sum_{j=1}^{n_h} \psi_{j,b_j}\left(\sum_{i=1}^{n_{in}} z_i v_{ij}\right).w_{jl}, \quad l = 1, 2, \ldots, n_{out} \tag{3.17}$$

where w_{jl}, $j = 1, 2, \ldots, n_h$ denotes the weight between the jth hidden layer and lth output layer. The network parameters of VTWNN are v_{ij}, w_{jl} and K_j. The number of parameters for v_{ij} is equal to $n_{in} \times n_h$, the number of parameters for w_{jl} is equal to $n_h \times n_{out}$, and the number of parameters for K_j is equal to n_h.

3.2.3 Design Parameters of VTWNN

A. *Number of Hidden Nodes (n_h)*: The size of the hidden layer is a general question raised on designing multilayer FFNN for real-life applications. An analytical method to determine the number of hidden nodes is difficult to obtain owing to the complexity of the network structure and the undetermined nature of the training process. Hence, the number of hidden nodes is experimentally found. In practice, the number of hidden nodes (n_h) depends on the application and the dimension of the input space.

B. *Parameter (K)*: The parameter K is used to control the shape of the nonlinear function $f(\cdot)$ and govern the PM in Fig. 3.1 and shows the effect of tuned parameter K_j to b_j. Generally, the range of K is tuned within 0.3 to 1.5. It

shows that the function reduces to threshold function when $K \to \infty$. And also it becomes a constant line when $K \to -\infty$.

C. *Network Parameters (Weights)*: The network parameters weights are v_{ij} and w_{jl} in Section 3.2.2 and a search method called hybrid particle swarm optimization with wavelet mutation (HPSOWM) in Section 3.3 is used to find the optimal parameters.

D. *Total number of parameters (n_{para})*: The total number of network parameters are calculated by:

$$n_{para} = n_h \left(1 + n_{in} + n_{out}\right) \tag{3.18}$$

3.3 Hybrid Particle Swam Optimization with Wavelet Mutation (HPSOWM)

To find the optimized design parameters in Section 3.2.3, hybrid particle swarm optimization with wavelet mutation (HPSOWM) is introduced in this section. It is a directed random search technique that is widely used in optimization problems to find out the optimal solution globally over a domain [Ling2008]. It is especially useful for complex optimization problems where the number of parameters is large and the analytical solutions are difficult to obtain.

In addition, its training performance depends only on the input-output data, the

derivative information of the cost function is not needed as in back propagation algorithms in Appendix A.4.2. If the training algorithm needs the derivative information of cost function, the updating rule is needed to be derived each time for each different network structure. In addition, HPSOWM algorithm has comparable or even superior search performance for some hard optimization problems with faster and more stable convergence rates.

Conventionally, the standard PSO in Appendix A.5.3 is used to find the optimal set parameters. Though it works well in the early stage, it leads to the problem of stagnation when a particle's current position coincides with the global best position. To deal with this phenomenon, an alternative approach, HGAPSO in Appendex A.5.3 is used. In HGAPSO, a mutation process often used in GA into PSO is incorporated to the standard PSO to enhance the searching performance. Though this process allows the search to escape from local optima and search in different zones of the search space, it may not be the best approach by fixing the size of mutating space all the time along the search.

In the HGAPSO, the solution space can be explored by performing mutation operations on particles along the search and premature convergence is more likely to be avoided. However, the mutating space is kept unchanged all the time throughout the search and the space for the permutation of particles in the PSO is also fixed. It can be improved by varying the mutating space along the search. It can be further improved by incorporating a dynamic mutation operation space in which the mutating

space dynamically contracts along the search.

To achieve this, a hybrid PSO with dynamic mutating space is introduced in this thesis by incorporating a wavelet function which is called hybrid particle swarm optimization with wavelet mutation (HPSOWM). The detail process of HPSOWM is presented in Algorithm 3.3.1. In HPSOWM, the wavelet in Section 3.2.1 is used as a tool to model seismic signals by combining the dilations and the translations of a simple oscillatory function (the mother wavelet) of finite duration. Thus, the mutating space of PSO is dynamically varying along the search based on the properties of the wavelet function. The resulting wavelet mutation operation aids the hybrid PSO to perform more efficiently and provides faster convergence than the conventional PSO. The following advantages are offered by HPSOWM.

A. *Improve the solution stability*: The mother wavelet in Section 3.2.1 satisfies an admissibility criterion which is a kind of half-differentiability. As a result, the stability of the operation is improved.

B. *Fine-tuning ability*: By controlling the dilation parameter of the wavelet function, the amplitude of the function can be adjusted. This property can be used to realize a fine-tuning effect to the mutation operation by decreasing the amplitude of the wavelet function to constrain the searching space when the number of iterations increases. Thus, the solution quality can be improved.

From Algorithm 3.3.1, $X(t)$ is denoted as a swarm at the t-th iteration. Each particle

$\mathbf{x}^p(t) \in X(t)$ contains κ elements $x_j^p(t)$ at the t-th iteration, where $p = 1, 2, ..., \theta$ and $j = 1, 2, ..., \kappa$; θ denotes the number of particles in the swarm and κ is the dimension of a particle.

Algorithm 3.3.1: PSEUDO CODE FOR HPSO WITH WAVELET MUTATION(*HPSOWM*)

$t \leftarrow 0$

Initialize $X(t)$

output $(f(X(t)))$

while <not termination condition>

do

- $t \leftarrow t + 1$
- Update $v(t)$ and $x(t)$ based on (3.19) to (3.20)
- **if** $v(t) > v_{\max}$
  - **then** $v(t) = v_{\max}$
- **if** $v(t) < v_{\min}$
  - **then** $v(t) = v_{\min}$
- Perform wavelet mutation operation with μ_m
- Update $\bar{x}_j^p(t)$ based on (3.23) to (3.26)
- **output** $(X(t))$
- **output** $(f(X(t)))$

return $(\hat{\mathbf{x}})$

comment: return the best solution

To find the optimized design parameters in Section 3.2.3, a swarm $X(t)$ is constituted with number of particles and each particle is coded with network parameters, weight

parameters v_{ij} and w_{jl}, the nonlinear transfer function parameter K which is in the form of $\mathbf{x} = [v_{ij};\ w_{jl};\ K]$. The objective of HPSOWM is to maximize the fitness function (cost function) $f(X(t))$ in (3.31) of particles iteratively. The swarm evolves from iteration t to t+1 by repeating the procedures as shown in Algorithm 3.3.1.

The velocity $v_j^p(t)$(corresponding to the flight speed in a search space) and the position $x_j^p(t)$ of the j-th element of the p-th particle at the t-th generation can be calculated using the following formulae:

$$v_j^p(t) = k \cdot \{\{w \cdot v_j^p(t-1)\} + \{\varphi_1 \cdot r_1(\tilde{x}_j^p - x_j^p(t-1))\} + \{\varphi_2 \cdot r_2(\hat{x}_j - x_j^p(t-1))\}\} \quad (3.19)$$

and

$$x_j^p(t) = x_j^p(t-1) + v_j^p(t) \quad (3.20)$$

where $\tilde{x}^p = [\tilde{x}_1^p, \tilde{x}_2^p, \ldots, \tilde{x}_k^p]$ and $\hat{\mathbf{x}} = [\ \hat{x_1}\ \hat{x_2}\ , ...\hat{x_\kappa}\ ]$, $j = 1, 2, ..., \kappa$.

The best previous position of a particle is recorded and represented as $\tilde{x}$; the position of best particle among all the particles is represented as $\hat{x}$; w is an inertia weight factor; r_1 and r_2 are acceleration constants which return a uniform random number in the range of [0,1]; k is a constriction factor derived from the stability analysis of (3.20) to ensure the system to be converged but not prematurely [Eberhart2000]. Mathematically k is a function of φ_1 and φ_2 as reflected in the following equation:

$$k = \left(\frac{2}{|2 - \varphi - \sqrt{\varphi^2 - 4\varphi}|} \right) \quad (3.21)$$

where $\varphi = \varphi_1 + \varphi_2$ and $\varphi > 4$.

HPSOWM utilizes $\tilde{x}$ and $\hat{x}$ to modify the current search point to avoid the particles moving in the same direction, but to converge gradually toward $\tilde{x}$ and $\hat{x}$. A suitable selection of the inertia weight w provides a balance between the global and local explorations. Generally, w can be dynamically set with the following equation [Eberhart2000]:

$$w = w_{\max} - \left(\frac{w_{\max} - w_{\min}}{T} \right) \times T \tag{3.22}$$

where t is the current iteration number, T is the total number of iteration, $w_{\max}$ and $w_{\min}$ are the upper and lower limits of the inertia weight, $w_{\max}$ and $w_{\min}$ are set to 1.2 and 0.1 respectively.

In (3.19), the particle velocity is limited by a maximum value $v_{\max}$. The parameter $v_{\max}$ determines the resolution with which regions are to be searched between the present position and the target position. This limit enhances the local exploration of the problem space and it realistically simulates the incremental changes of human learning. If $v_{\max}$ is too high, particles might fly past good solutions. If $v_{\max}$ is too small, particles may not explore sufficiently beyond local solutions. From experience, $v_{\max}$ is often set at 10%-20% of the dynamic range of the element on each dimension. Next, the mutation operation in Section 3.3.1 is used to mutate the element of particles.

3.3 Hybrid Particle Swam Optimization with Wavelet Mutation (HPSOWM)

3.3.1 Operation of wavelet mutation (WM)

The mutation operation is used to mutate the elements of particles. In general, various methods like uniform mutations or nonuniform mutations [Michalewicz1994] are employed to realize the mutation operation. However, the proposed WM operation exhibits a fine-tuning ability and the details of the operation are presented as follows:

To being with hybrid PSO with wavelet mutation (HPSOWM), a probability of mutation, $\mu_m \in [0 \quad 1]$ is defined. For each particle element, a random number between 0 and 1 will be generated. If the generated number is less than or equal to μ_m, the mutation will take place on that element. For instance, if $x^p(t) = [x_1^p(t), x_1^p(t), \ldots, x_\kappa^p(t)]$ is the selected p^{th} particle and the element of particle $x_j^p(t)$ is randomly selected for mutation and its value are inside the particle element's boundaries, $\left[\rho_{\min}^j, \rho_{\max}^j\right]$.

The resulting particle, $\bar{x}^p(t) = [\bar{x}_1^p(t), \bar{x}_1^p(t), \ldots, \bar{x}_\kappa^p(t)]$ is governed by:

$$\bar{x}_j^p(t) = \begin{cases} x_j^p(t) + \sigma \times \left(\rho_{\max}^j - x_j^p(t)\right) & , \sigma > 0 \\ x_j^p(t) + \sigma \times \left(x_j^p(t) - \rho_{\min}^j\right) & , \sigma < 0 \end{cases} \tag{3.23}$$

where $j \in 1, 2, \ldots, \kappa$ and κ denotes the dimension of particles. The value of σ which is governed by Morlet wavelet function in (3.5) as mother wavelet which is obtained as:

$$\sigma = \psi_{a,0}(\varphi) = \frac{1}{\sqrt{a}}\psi\left(\frac{\varphi}{a}\right) = \frac{1}{\sqrt{a}}e^{\frac{-(\frac{\varphi}{a})^2}{2}}\cos\left(5\left(\frac{\varphi}{a}\right)\right) \tag{3.24}$$

Remark 3.1 *Different kinds of mother wavelets have been considered during the development of the algorithm, e.g., the Mexican hat wavelet (normalized), the Mexican hat wavelet, the Morlet wavelet, the Gaussian wavelet, and the Meyer wavelet [Ling2008]. By trial and error through experiments for good performance, various wavelet functions have been investigated in terms of cost values. Last, the Morlet wavelet in (3.5) is chosen as the mother wavelet in the WM operation because the selected wavelet function offers the best performance.*

If σ is positive and approaching 1, the mutated element of the particle will tend to the maximum value of $x_j^p(t)$. Conversely, when σ is negative ($\sigma \leq 0$) approaching to -1, the mutated element of particle will tend to minimum value of $x_j^p(t)$. A larger value of $|\sigma|$ gives a larger searching space for $x_j^p(t)$. When $|\sigma|$ is small, it gives a smaller searching space for fine-tuning. Referring to Property I of the wavelet, the sum of the positive σ is equal to the sum of the negative σ when the number of samples is large and φ is randomly generated, i.e.,

$$\frac{1}{N}\sum_{N}\sigma = 0 \quad for \quad N \to \infty \tag{3.25}$$

where N is the number of samples.

Hence, the overall positive mutation and the overall negative mutation throughout the evolution are nearly the same. This property gives better solution stability (a smaller standard deviation of the solution values upon many trials). As over 99% of the total energy of mother wavelet function contained in the interval [-2.5 2.5], φ can

be randomly generated from [-2.5a 2.5a].

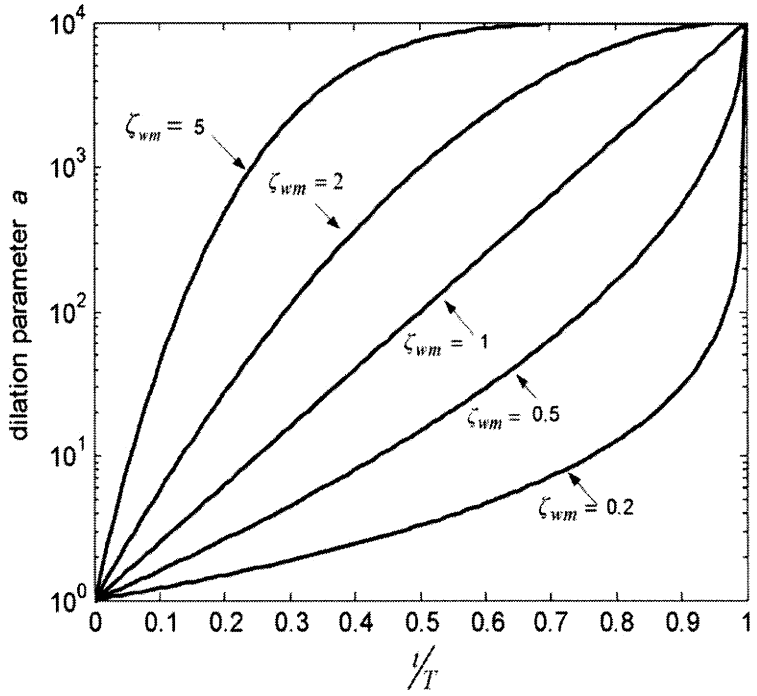

Figure 3.8: Effect of the shape parameter ζ_{wm} to a with respect to $\frac{t}{T}$

The value of the dilation parameter a is set to vary with the value of $\frac{t}{T}$ to meet the fine-tuning purpose, where T is the total number of iteration and t is the current number of iteration. In order to perform a local search when t is large, the value of a should increase as $\frac{t}{T}$ increases so as to reduce the significance of the mutation. Hence, a monotonic increasing function a which is governing a and $\frac{t}{T}$ is proposed in the following form:

$$a = e^{-ln(g)\times\left(1-\frac{t}{T}\right)^{\zeta_{wm}}+ln(g)} \tag{3.26}$$

where ζ_{wm} is the shape parameter of the monotonic increasing function, g is the upper limit of the parameter a. The effect of various values of the shape parameter ζ_{wm} and

the parameter g to a with respect to $\frac{t}{T}$ shown in Figs. 3.8 and 3.9 respectively. In Fig. 3.9, g is set as 10000. Thus, the value of a is between 1 to 10000.

Referring to (3.24), the maximum value of σ is 1 when the random number of $\varphi = 0$ and $a = 1$ at $t/T = 0$. If $\sigma = 1$, (3.23) becomes:

$$\bar{x}_j^p(t) = x_j^p(t) + 1 \times \left(para_{\max}^j - x_j^p(t)\right) = para_{\max}^j \tag{3.27}$$

It ensures a larger searching space for the mutated element. When the value $\frac{t}{T}$ is near to 1, the value of a is so large that the value of σ will be very small. For instance, at $\frac{t}{T} = 0.9$ and $\zeta_{wm} = 1$ the dilation parameter $a = 4000$. The value of $\sigma = 0.0158$, thus (3.24) becomes:

$$\bar{x}_j^p(t) = x_j^p(t) + 0.0158 \times \left(para_{\max}^j - x_j^p(t)\right) \tag{3.28}$$

A smaller searching space for mutated element is given in (3.28). Changing ζ_{wm} will change the characteristics of the monotonic increasing function of the WM. The dilation parameter a will take a value to perform fine-tuning faster as ζ_{wm} is increasing. It is chosen by trial and error which depends on the kind of optimization problem. When ζ_{wm} becomes larger, the decreasing speed of the step size σ of the mutation becomes faster.

In general, if the optimization problem is smooth and symmetric, the solution can be found easier by the searching algorithm and fine-tuning can be done in the early stages. Thus, a larger ζ_{wm} can increase the step size of mutating space. After the operation of the WM, an updated swarm is generated. The swarm with new particle

will repeat the same process. Such an iteration process will be terminated when a defined number of iterations are met.

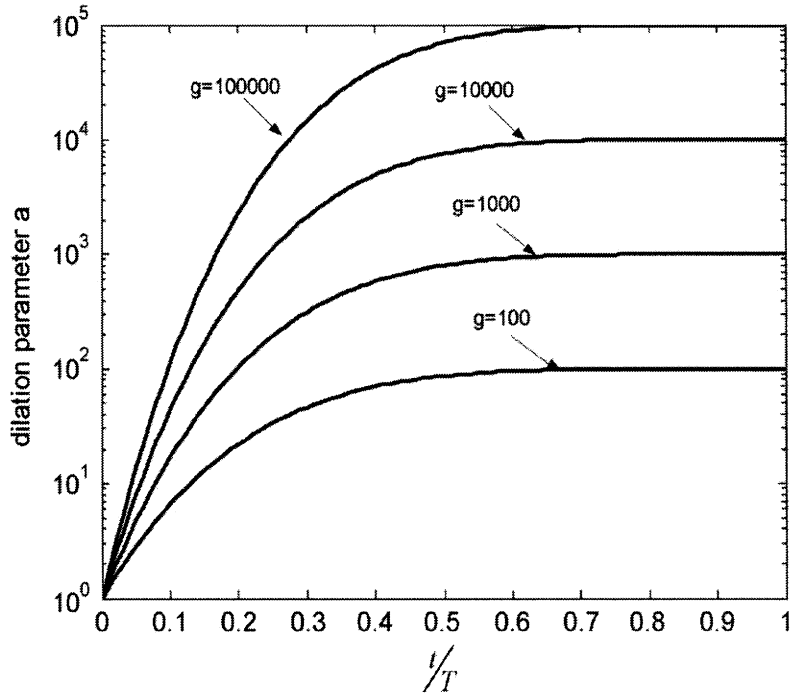

Figure 3.9: Effect of the parameter g to a with respect to $\frac{t}{T}$

3.3.2 Choosing parameters of HPSOWM

HPSOWM is seeking a balance between the exploration of new regions and the exploitation of the already sampled regions in the search spaces. This balance, which critically affects the performance of HPSOWM, is governed by the right choices of the control parameters: Swarm size (θ), the probability of mutation (μ_m), the shape parameter (ζ_{wm}) and parameter g of wavelet mutation. Some views about these parameters are given as follows:

(i) Increasing swarm size (θ) will increase the diversity of the search space, and reduce the probability that HPSOWM prematurely converges to a local optimum. However, it also increases the time required for the population to converge to the optimal region in the search space.

(ii) Increasing the probability of mutation (μ_m) tends to transform the search into a random search such that when $\mu_m = 1$, all element of particles will mutate. This probability gives us an expected number ($\mu_m \times \theta \times \kappa$) of element of particles that undergo the mutation operation. In other words, the value of μ_m depends on the desired number of element of particles that undergo the mutation operation. Normally, when the dimension is very low (number of element of particles is less than 5, μ_m is set at 0.5 to 0.8). When the dimension is around 5-10, μ_m is set at 0.3 to 0.4. When the dimension is in the range of 11 to 100, μ_m is set at 0.1 to 0.2. When the dimension is in the range of 101 to 1000, normally μ_m is set at 0.05 to 0.1.

Lastly, when the dimension is very high (number of element of particles is larger than 1000),μ_m is set at < 0.05. The rationale to set this select criterion is: when the dimension of input signal is high, μ_m should be set to be a smaller value; when the dimension is low, μ_m should be set to be larger value. It is because if the dimension is high and μ_m is set to be a larger number, then the number of elements of particles undergoing mutation operation will be large. It will increase the searching time and more importantly will destroy the current

information about the application in each iteration as all elements of particles are randomly assigned.

Generally speaking, by choosing the value of μ_m, the ratio of the number of elements of particles undergoing mutation operation to the population size can be maintained to prevent the searching process turning to a random searching one. Thus, the value of μ_m is based on this selection criterion and chosen by trial and error through experiments for good performance for all functions.

(iii) The dilation parameter a is governed by the monotonic increasing function (3.26), which is controlled by two parameters. They are called shape parameter ζ_{wm} and parameter g. Changing the parameter ζ_{wm} will change the characteristics of the monotonic increasing function of the wavelet mutation. The dilation parameter a will take a value so as to perform fine-tuning faster as ζ_{wm} is increasing. It is chosen by trial and error, which depends on the kind of the optimization problem.

When ζ_{wm} becomes larger, the decreasing speed of the step size (σ) of the mutation becomes faster. In general, if the optimization problem is smooth and symmetric, it is easier to find the solution and the fine-tuning can be done in early iteration. Thus, a larger value of ζ_{wm} can be used to increase the step size of the early mutation. Parameter g is the value of the upper limit of dilation parameter a. A larger value of g implies that the maximum value of a is larger.

In other words, the maximum value of ζ_{wm} will be smaller (smaller searching

limit is given). Conversely, a smaller value of g implies that the maximum value of a is smaller. In other words, the maximum value of σ will be larger (larger searching limit is given). In our point of view, fixing one parameter and adjusting another parameter to control the monotonic increasing function is more convenient to find a good setting.

3.4 Case Study in Hypoglycemia Detection System

An optimized variable translation wavelet neural network (VTWNN) with 4 inputs and 1 output system is developed for the detection of hypoglycemic episodes in T1DM patients as shown in Fig. 3.10.

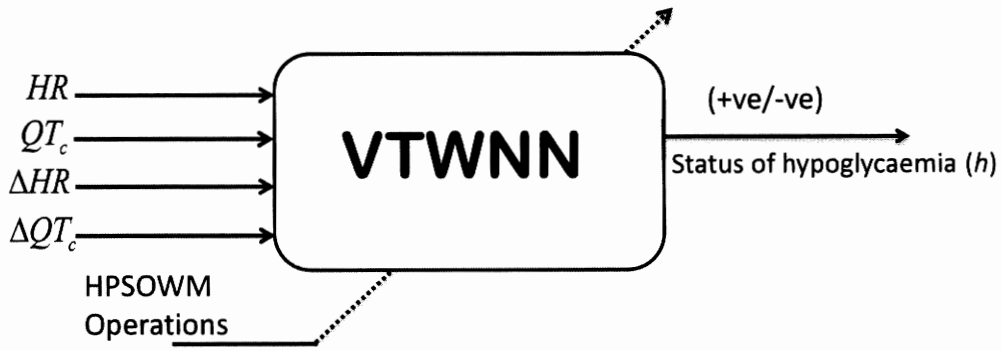

Figure 3.10: Hybrid PSO based VTWNN hypoglycemia detection system

The four psychological parameters of ECG signal are heart rate (HR), corrected QT interval (QTc), change of heart rate (ΔHR) and change of corrected QT interval in time (Δ QTc) are used as the input while the status of hypoglycaemia (h) (+1 is

hypoglycemia and -1 is non-hypoglycemia) represents as the output. The proposed VTWNN detection system is mainly used to find the relationship between physiological parameters of ECG signal (HR, QTc, ΔHR and Δ QTc) and the presence of hypoglycemia (h). As presented in Fig. 3.10, a global learning optimization algorithm, HPSOWM in Section 3.3.2 is used to optimize the parameters of system.

3.4.1 Analysis of electrocardiogram (ECG)

The ECG is a graphic recording of the electrical activity produced in association with the heart beat and measured on the body. Usually, a twelve lead system is used for the ECG recording which gives information about the activities of the heart. In ECG signal, the important points are the Q point, R peak and T wave peak as shown in Fig. 3.11.

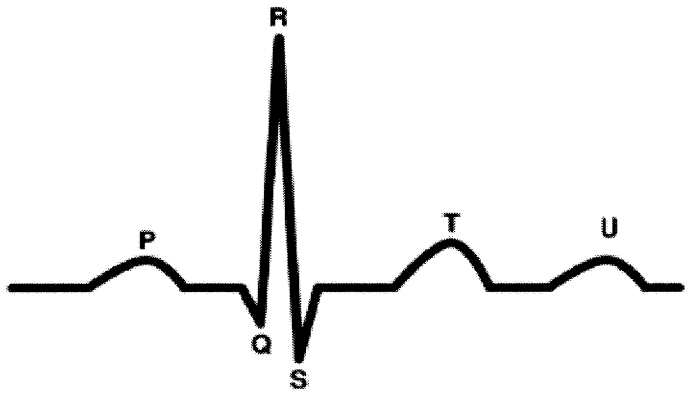

Figure 3.11: The normal ECG Signal

In order to analyze the signal, the characteristic of QRS wave is important and most of the useful information in the ECG is found in the intervals and amplitudes defined

by its features. The performance of ECG analyzing system depends on the accurate detection of QRS complex, QT interval of which the most important are the R peak detection, Q wave and T wave detection.

The typical QRS complex wave consist of compact continuous wave, the first downwards negative wave is namely Q wave, the next sharply upwards positive wave is R wave, and the latter downwards negative wave is S wave as presented in Fig. 3.11. The QRS wave is the potential difference signal between the signal generated by the bilateral ventricle and transmitted to the surface body in the process of depolarization. Its general duration is 0.04s to 0.1s. The abnormal Q wave is useful in medical diagnosis.

The ECG parameters under investigation involve the parameters in depolarization and repolarization stages of ECG signal. The important points of an ECG signal in Fig. 3.11 are the Q point, R peak and T wave peak, in which R peak is obtained by detecting the maximum amplitude which is higher than the defined threshold level. After obtaining R peak, it is regarded as a reference point and the Q waves, and peak of the T wave can be found by approximating their locations at the right and left sides of R peak. In normal ECG signal the Q wave and peak of T wave are located about 120 and 300ms intervals from R peak [Nguyen2008] [Ling2011].

Cumulating clinical and experimental evidence has shown in Section 2.2.1 that there is a marked change in ECG morphology, especially in QT interval and/or QRS complex

during hypoglycemia. In our recent study [Nguyen2012], the lengthening of the QT interval has been demonstrated in Type 1 Diabetic patients by measuring the QT interval precisely. Fig.3.12 shows the variation of HR and QTc from one patient in which the value of HR and QTc are increased under hypoglycemia conditions by giving the values of 85 bpm and 476ms whereas 69 bpm and 400ms are obtained under normal states.

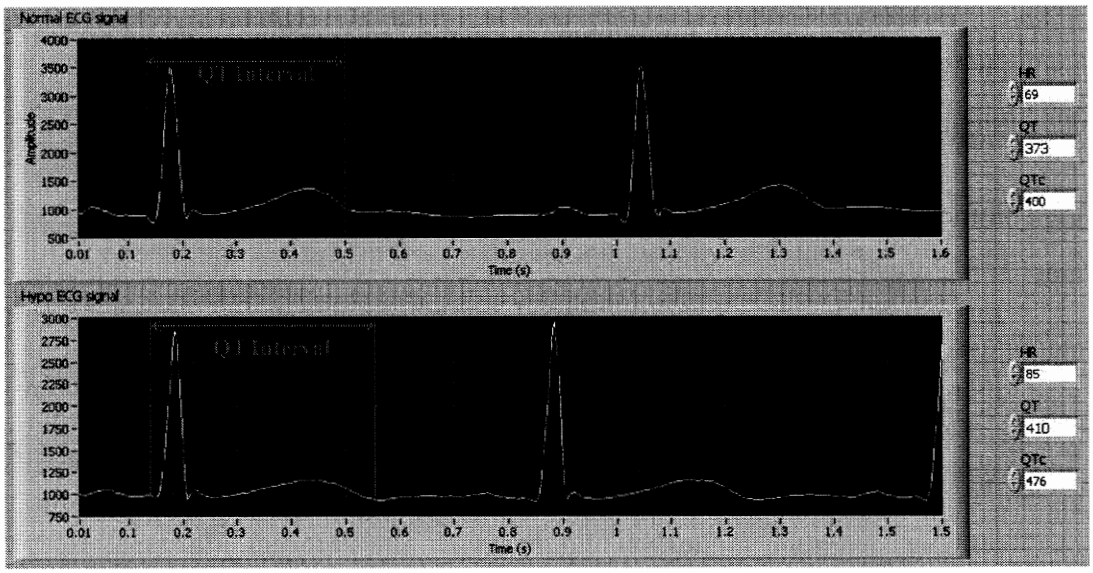

Figure 3.12: The lengthening of QTc interval under normal vs. hypoglycemic states

In this analysis of the ECG feature, R peaks are found by using a peak detection while the beginning and ends of P, Q, T waves are detected by the zero crossing method. Once R peaks are found, they are regarded as the reference and obtaining peaks of P waves, beginning of Q waves and T waves by scanning on both side of R peaks to get zero crossing points. As shown in Fig. 3.12, the QTc prolongation was associated with hypoglycemia and the QTc lengthening led to (>476ms) during hypoglycemia

against non-hypoglycemia.

3.4.2 Data Set

Detailed studies in 2.2.1 have shown that ECG waveforms are greatly affected during hypoglycemia including depression of the ST segment, flattening of the T wave and prolongation of QT interval. These changes are mostly investigated under insulin-induced hypoglycemia, i.e. the insulin injections are intentionally performed on healthy volunteers in order to investigate ECG alterations during insulin-induced hypoglycemia.

Instead of studying insulin-induced hypoglycemia, in this thesis, it is aimed to explore the effects of natural occurrence of hypoglycemic episodes on physiological responses. To perform, 15 children with T1DM (14.6 ± 1.5 years) volunteered for the 10-hour overnight hypoglycemia study at the Princess Margaret Hospital for Children in Perth, Australia. To measure the required physiological parameters, HypoMon (AiMedics) was used to measure the required physiological parameters, while the actual blood glucose (BG) levels were collected as reference using Yellow Spring Instruments (YSI) blood glucose analyzers. The measurement results revealed that a profound effect on QT interval prolongation and corrected heart rate affected the degree of prolongation during hypoglycemia and may allow the development of a hypoglycemia detection device.

All the collected data was obtained with the approval from Womens and Childrens Health Service, Department of Health, Government of Western Australia, and with informed consent. A comprehensive patient information and consent form was formulated and approved by the Ethics Committee. Each patient received the consent form which included the actual consent and a revocation of consent page at least two weeks prior to the start of the studies. He/she had the opportunity to raise questions and concerns with any medical advisers, researchers and the investigators. For the children participating in this study, the parent or guardian signed the relevant forms.

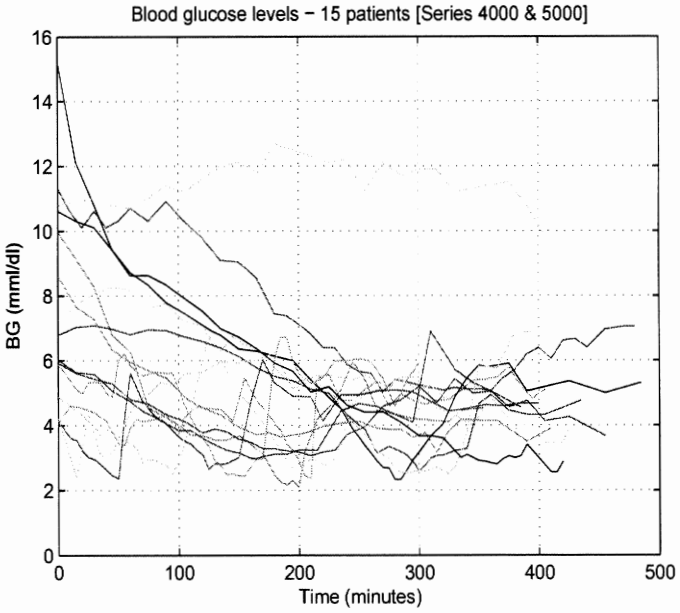

Figure 3.13: Actual blood glucose level profiles in 15 T1DM children

The main parameters used for the detection of hypoglycemia are HR and QTc. The actual blood glucose profiles for 15 T1DM children are shown in Fig. 3.13. Data were collected between 360 minutes to 480 minutes. The presence of hypoglycemia is

estimated at sampling period which is about 5 minutes and approximately 30 to 40 data points are used for each patient. The responses from the 15 T1DM children exhibit significant changes during the hypoglycemia phase against the non-hypoglycemia phase as shown in Fig.3.13.

Patient Number	No.of Hypo Episodes	Duration of Episodes(min)
1	0	0
2	8	40
3	7	45
4	10	50
5	4	20
6	14	75
7	9	45
8	11	55
9	17	105
10	0	0
11	0	0
12	9	45
13	9	85
14	13	70
15	7	70

Table 3.1: The 15 patients and their associated hypoglycemia and non-hypoglycemia events

Among 15 patients, 12 T1DM patients were found to experience hypoglycemic episodes as their BGL threshold level were less than ($3.3mmol/l$), while 3 T1DM patients did

not experience hypoglycemia events as presented in Table 3.1. This is because no natural occurrence of hypoglycemia episodes were presented during measurement or they may not have been valid with the defined BGL threshold level. For the clinical application in this study, BGL $< 3.3mmol/l$ was considered a hypoglycemic episodes. The relationships between overnight measurements of QT interval corrected for heart rate (QTc) and falling BGL were discovered by the use of statistical correlation analysis in the following Section 3.4.5.1.

3.4.3 Performance Evaluation

To measure the performance of a biomedical classification test, the sensitivity, ξ and specificity, η are introduced [Altman1994]. Sensitivity measures the proportion of actual positives which are correctly identified while specificity measures the proportion of negatives which are correctly identified.

$$Sensitivity\ (\xi) = \frac{N_{TP}}{N_{TP} + N_{FN}} \tag{3.29}$$

$$Specificity\ (\eta) = \frac{N_{TN}}{N_{TN} + N_{FP}} \tag{3.30}$$

where:

N_{TP} is the number of true positives which means the sick people are correctly diagnosed as sick

N_{FN} is the number of false negative which means the sick people are wrongly diagnosed as healthy

N_{FP} is the number of false positives which means the healthy people are wrongly diagnosed as sick

N_{TN} is the number of true negatives which means the healthy people are correctly diagnosed as healthy. The values of ξ and η are within 0 to 1.

The objective of the system is to maximize sensitivity and specificity. Thus, the fitness function in Section 3.4.4 is designed to maximize both values.

3.4.4 Fitness Function Formation

The formation of fitness function depends on the application. In general, the sensitivity (ξ) and specificity (η) in Section 3.4.4 are used to measure the performance of the biomedical classification test. In this clinical application, the sensitivity is more important than specificity because it identifies the actual hypoglycemic episodes in patients with T1DM.

Essentially, the fitness function is designed to optimize the sensitivity and the specificity. The objective of the proposed system is to maximize the fitness function which is equivalent to the maximization of sensitivity and specificity. In order to meet with the objective, the fitness function $f(\xi, \eta)$ is defined as follows:

$$f = \bar{\xi}_t + \bar{\eta}_t + \bar{\xi}_v + \bar{\eta}_v \tag{3.31}$$

where $\bar{\xi}_t$ and $\bar{\eta}_t$ are training sensitivity and specificity, $\bar{\xi}_v$ and $\bar{\eta}_v$ are validation sensitivity and specificity. They are defined as follows:

$$\bar{\xi}_t = \begin{cases} 1 & if\ \xi_t \geq \xi_l \\ 1 - \left(\frac{\xi_l - \xi_t}{\xi_l}\right) & if\ \xi_t < \xi_l \end{cases} \tag{3.32}$$

and

$$\bar{\xi}_v = \begin{cases} 1 & if\ \xi_v \geq \xi_l \\ 1 - \left(\frac{\xi_l - \xi_v}{\xi_l}\right) & if\ \xi_v < \xi_l \end{cases} \tag{3.33}$$

and

$$\bar{\eta}_t = \begin{cases} 1 & if\ \eta_t \geq \eta_l \\ 1 - \left(\frac{\eta_l - \eta_t}{\eta_t}\right) & if\ \eta_t < \eta_l \end{cases} \tag{3.34}$$

and

$$\bar{\eta}_v = \begin{cases} 1 & if\ \eta_t \geq \eta_l \\ 1 - \left(\frac{\eta_l - \eta_v}{\eta_t}\right) & if\ \eta_t < \eta_l \end{cases} \tag{3.35}$$

In some cases, the fitness function with constant values are used to control the balance between sensitivity and specificity [Ling2010]. However, this approach considered only training and testing data set which can increase the risk to the occurrence of overtraining. Thus, the validation strategy is introduced to the defined fitness function in (3.31) to reduce the risk of overtraining.

Fig. 3.14 shows the idea of defined fitness function. The x-axis represents the iteration number and T denotes the total iteration number. The y-axis represents the classification performance which is governed by sensitivity and specificity. The main function of validation strategy is to train the training set and test the trained results with the validation set with respect to a target region. The definition of target region varies due to the performance requirement of the system. The target region related to sensitivity and specificity, for example, $\xi > 0.6$ and $\eta > 0.4$.

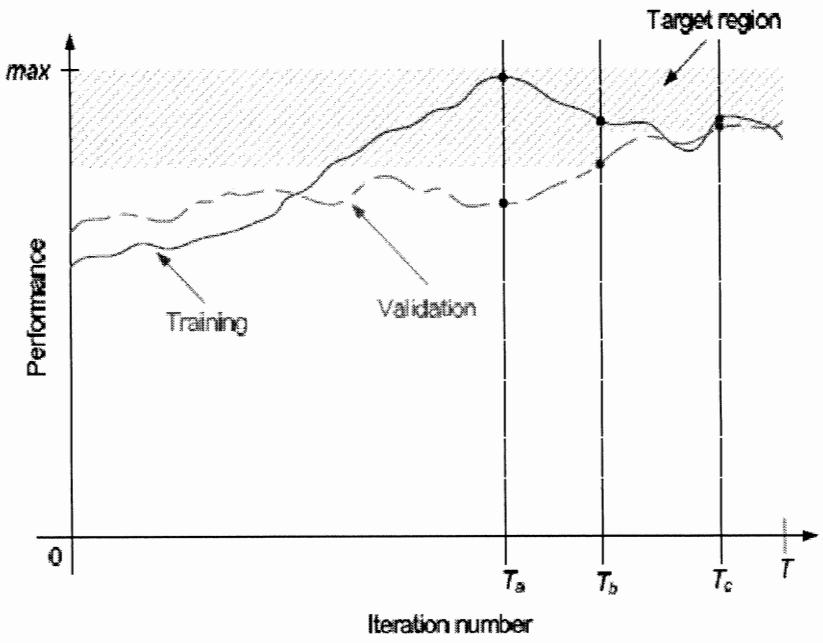

Figure 3.14: Fitness function with validation strategy

As shown in Fig. 3.14 at iteration T_a the training set performance has progressed to a good result. However, the performance of the validation set cannot meet the minimum requirement of the target region. As a result, the testing performance will

be out of the target region. It is an overtraining phenomenon, i.e., the model is over-fit with the training set and the generalization performance reduced.

At time T_b the performance of both the training set and the validation set are within the target region. As a consequence, the testing result has a good chance of meeting the performance requirement ($\xi > 0.6$ and $\eta > 0.4$). Once both sets have reached the target region, the target region is updated by increasing the performance requirement level, i.e., $\xi > 0.6$ to $\xi > 0.65$. At time T_c, when both the training and validation performance are in the updated target region, the model parameters will be updated.

Algorithm 3.4.1: ALGORITHM FOR FITNESS FUNCTION(f)

$\eta_l \leftarrow \eta_d$

$\xi_l \leftarrow \xi_d$

output $\left(\bar{\xi}_t, \bar{\eta}_t, \bar{\xi}_v, \bar{\eta}_v\right)$ *based on* (3.29) *and* (3.30)

output (f) *based on* (3.31)

if $(\xi_t > \xi_l \& \ \eta_t > \eta_l \ \& \ \xi_v > \xi_l \& \ \eta_v > \eta_l)$

  then $\xi_d = \xi_d + \lambda$

Based on the above concept, the process of formulating the fitness function is shown in Algorithm 3.4.1. First, a target region will be defined with the lower boundaries of sensitivity (ξ_l) and specificity (η_l) while the upper boundaries for them are defined as 1. In Algorithm 3.4.1, ξ_d and η_d are the initial values of lower boundaries sensitivity

and specificity. Second, based on (3.29) and (3.30), the training sensitivity (ξ_t), the training specificity (η_t), the validation sensitivity (ξ_v) and the validation specificity (η_v) are evaluated. Referring to Fig. 3.14, the training performance is the result of xi_t and η_t. Similarly, the validation performance is the result of ξ_v and η_v. Once xi_t, η_t, ξ_v and η_v are obtained, the fitness function is calculated as in (3.31).

In (3.31), the maximum fitness value is equal to 4 when $\bar{\xi}_t$, $\bar{\eta}_t$, $\bar{\xi}_v$, $\bar{\eta}_v$ equal to 1. It implies that both the training and validation performances have met the defined requirement. Finally, if the target region ($\xi_t > \xi_l$ and $\eta_t > \eta_l$ and $\xi_v > \xi_l$ and $\eta_v > \eta_l$) is reached, ξ_d will be updated by $\xi_d + \lambda$, where the value of λ is normally between 0.01 and 0.05. For finding out the optimal sensitivity with fixed specificity value, only ξ_d will be updated during the training process.

3.4.5 Experimental Results and Discussion

The responses from 15 T1DM children in Fig. 3.13 exhibit significant changes during the hypoglycemia phase against the non-hypoglycemia phase. Normalization was used to reduce patient-to-patient variability and to enable group comparison by dividing the patients heart rate, corrected QT interval by his/her corresponding values at time zero. The information for each patient in terms of number and duration of hypoglycemic events for hypoglycemic episodes was tabulated in Table 3.1.

To perform the classification test, the overall data set consisting of both hypoglycemia

data and non-hypoglycemia data points were organized into the training set (patient number 1 to 5 with 184 data points), the validation set (patient number 6 to 10 with 192 data points) and the testing set (patient number 11 to 15 with 153 data points). Referring to Table 3.1, each training, validation and testing set consists of four patients whose natural occurrence of hypoglycemia can be observed for the defined BGL $\leq 3.3mmol/l$ and one patient with no associated hypoglycemia events. Data sets that include both hypoglycemia data points and non-hypoglycemia data points were used to measure the classification performance in terms of sensitivity and specificity.

In clinical application, the proposed VTWNN is selected as a suitable classifier due to its excellent characteristics in capturing the nonstationary nature of ECG signal. It models the input-output function as the input-dependent network parameters because of input dependent translation parameter b in (3.15). Due to the variable translation parameters b, the network becomes an adaptive network and provides better classification performance.

The proper selection of the number hidden neurons (n_h) is vital for improving the network performance. In VTWNN, the selection on the number of hidden nodes is the same as the way that is chosen in conventional neural network structure. There is currently no mathematical theory that provides a definitive number of neurons for optimal network performance. Conventionally, the number of hidden neurons are chosen by a trial and error method until the optimal solution is obtained.

However, there is a trade off between too small number or too big number of hidden nodes. If too few hidden neurons are selected, the network is unable to model complex data and results in a poor fit. If too many hidden neurons are used, then training will become excessively long and the network may over fit. Over fitting occurs when the network begins to model random noise contained within the data, resulting in a failure to converge on a generalized solution. During the simulation studies, different number of hidden neurons such as 5, 10, 15 and 20 are tried. Among them, the 10 hidden neuron is given the best classification performance.

All the network parameters $[v_{ij} \quad w_{jl} \quad K_j]$ of VTWNN are given by $[-10 \quad 10]$ for v_{ij} and w_{jl} while K_j is selected between $[0.3 \quad 1.5]$. Since the number of parameters in VTWNN mainly depend on the number of hidden neurons (n_h) and 10 hidden nodes are experimentally selected for this application, the total number of VTWNN design parameters is about to be 66. Referring to (3.9)-(3.36), the output (y) of proposed 4 input 1 output system in Fig. 3.10 is obtained by:

$$y = \sum_{j=1}^{10} \psi_{j,b_j} \left(\sum_{i=1}^{4} z_i v_{ij} \right) . w_{jl} \tag{3.36}$$

In this application, the status of hypoglycemia h is positive when the output y is positive which is defined as follows:

$$h = \begin{cases} +1, & y \geq 0 \\ -1, & y < 0 \end{cases} \tag{3.37}$$

The optimal values of VTWNN parameters are obtained through the maximization of designed fitness function (3.31). During the training process, for a given set of

particle $\mathbf{x} = [v_{ij};\ w_{jl};\ K]$, HPSOWM evaluates the fitness value of each particle at each iteration step and searches for the optimum network parameters. In order to train VTWNN, the number of iteration is set at 1500.

All parameters of HPSOWM such as Swarm size (θ), the probability of mutation (μ_m), the shape parameter (ζ_{wm}) and parameter g of wavelet mutation were selected as specified in [Ling2008] apart from probability of mutation (μ_m) and shape parameters (ζ_{wm}). It should be noted that there were no default rules to find the best parameters. A trial and error approach was carried out through experimentation. In this application, the basic setting of the HPSOWM parameters were defined as:

- Swarm size θ : 50 [Ling2008];
- Constant c_1 and c_2 : 2.05 [Ling2008];
- Maximum velocity $v_{\max}$: 0.4 [Ling2008];
- Probability of mutation μ_m: 0.7 (by trial and error for optimal performance);
- The shape parameter of wavelet mutation, ζ_{wm} : 2;
- The constant value g of wavelet mutation : 10000 [Ling2008];
- Number of iteration T : 1000.

3.4.5.1 Statistical Correlation Analysis

The prevalence of prolonged QT interval (starting from Q point to the end of T point in Fig. 3.11) on the ECG signal during spontaneous hypoglycemia in 15 children with Type 1 diabetes was examined and explored for the relationships between overnight measurements of QT interval corrected for heart rate (QTc) and falling glucose level. In this clinical application, BGL $< 3.3 mmol/l$ was considered hypoglycemic episodes. The relationships between overnight measurements of QT interval corrected for heart rate (QTc) and falling BGL were discovered by the use of statistical correlation analysis. The changes in measured ECG parameters are summarized in Table. 3.2.

Parameters	**Hypolycemic State**	**Normal State**	**p-value**
HR	1.033 ± 0.242	1.082 ± 0.298	< 0.06
QTc	$1.0.31 \pm 0.086$	1.060 ± 0.084	< 0.001
Skin Impedances	1.111 ± 1.460	1.108 ± 1.277	< 0.984

Table 3.2: Changes in ECG parameters: HR and QTC under hypoglycemic conditions

Remark 3.2 *In statistical significance testing, p-value is generally defined as a short form for probability value as well as another way of saying significance value. It can also be simply said as the chance that the relationship under observation is simply carried out by pure chance. Commonly, its range is defined from 0 (no chance) to 1 (absolute certainty), i.e. p-value of 0.5 means a 50 % chance and 0.05 means*

5 % chance. In most statistical analysis, an experimental result is declared to be statistically significant if the p-value for the observed result is less than the level of significance (p-value < significance level=0.05).

The statistical results in Table. 3.2 have proven that the feature extracted ECG parameters, HR and QTc are significantly increased under hypoglycemic conditions with p-value, HR ($p<0.06$) and QTc ($p<0.001$). However, in this clinical study (natural occurrence), the reduction of skin impedances of the patients was not strong with p-value < 0.984). It should be noted that in our previous 4-hour glucose clamp study [Nguyen2007], skin impedances reduced significantly during hypoglycemic episodes.

Considering a higher correlation of HR and QTc to the status of hypoglycemia as presented in Table 3.2, they are used as the main inputs of hypoglycemia monitoring system while ΔHR and Δ QTc are used for fine-tuning of the proposed detection system Fig. 3.10. The design parameters are optimized by the use of HPSOWM in Section 3.3 for improving the classification performance.

3.4.5.2 Results Analysis

Three comparison approaches such as (1) variable translation wavelet neural network (VTWNN) in Section 3.2, (2) wavelet neural network (WNN) in Appendix A.2 and (3) feedforwrad neural network (FFNN) in Appendix A.1 are compared and analyzed. The performance of the proposed detection system is firstly analyzed by means of

a ROC curve in Fig. 3.15 in which the sensitivity (true positive rate) and the 1-specificity (false positive rate) are relatively plotted.

Method	Area under ROC curve(Training)
VTWNN	0.7082
WNN	0.6758
FFNN	0.6713

Table 3.3: Comparison studies: Area under ROC curve

In ROC curve analysis, the performance effectiveness is measured by the area under the ROC curve,i.e. the wider ROC curve area is defined as the better accuracy. The corresponding ROC curve areas for VTWNN, WNN and FFNN are compared and analyzed in Table 3.3. The study compared methods in VTWNN, WNN and FFNN found that the proposed VTWNN perform more accurately in detection of hypoglycemia (area under the curve = 0.7082) compared to WNN and FFNN (area under the curve = 0.6753 and 0.6695).

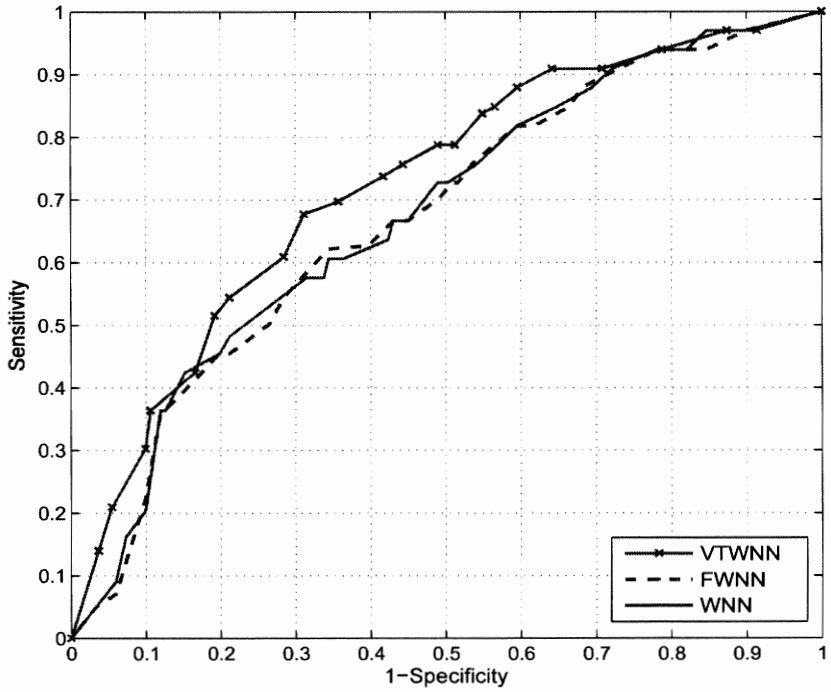

Figure 3.15: ROC curve analysis

In Fig. 3.3, the ROC curve indicates the relationship between sensitivity and specificity. It can be seen that the higher sensitivity is obtained when the specificity is lower and vice versa. For instance, if the cut-off point is fixed at $\eta_l = 20\%$ (1-specificity=0.8 (80%) in Fig. 5), the sensitivity is supposed to be about 90% or if the cut-off point is fixed at $\eta_l = 70\%$ (1-specificity=0.3 (30%) in Fig. 3.3), the sensitivity is supposed to be about 40%.

In this clinical application, the sensitivity which mainly identifies actual hypoglycemic episodes is more important than the specificity. Due to the importance of sensitivity, the cut-off point is defined in ROC curve (Fig. 3.15) at the lower boundary specificity

(η_l)= 0.6 ($1 - \eta_l$=60%) which is equivalent to maximum specificity, η_l=40% and find the optimized sensitivity at the defined specificity region.

Remark 3.3 *In this application, the common criteria of classification performance is defined to be the sensitivity $\geq$ 0.6 and specificity $\geq$ 0.4. To achieve this, the analysis starts at lower boundary specificity, η_l=40%. If the proposed detection system can correctly detect the status of hypoglycemia, the sensitivity will be higher, ie., the higher sensitivity represents the better performance of the proposed classifier.*

At the defined initial specificity (η_l=40%), the mean value of training, validation and testing results were presented in terms of the sensitivity and specificity in Table 3.4. All results are averaging over 20 runs. In Table 3.4, the proposed optimized VTWNN is found to be satisfactory by giving mean (average) testing sensitivity and specificity value of 77.41 and 47.42% while the other comparison methods WNN and FFNN are giving (71.39 and 44.37%) and (68.84 and 48.34%).

η_l			VTWNN	WNN	FFNN
40%	Training (%)	Sen(ξ)	88.40	84.12	83.64
		Spec(η)	40.80	40.63	40.50
		γ	69.36	66.72	66.38
	Validation (%)	Sen(ξ)	82.50	80.44	79.07
		spec(η)	41.25	40.94	41.38
		γ	66.00	64.64	63.99
	Testing (%)	Sen(ξ)	77.41	71.39	68.84
		Spec(η)	47.42	44.37	48.34
		γ	**65.41**	**60.58**	**60.64**

Table 3.4: Mean value of Training, Validation and Testing Results: Set specificity as η_l= 40%

In this application, it seems that the testing specificity (47.42%) is higher than the training specificity which is about 40.80%. It is because the nature of patients' data that is specified for testing set. As can be seen in Table 3.1, due to different nature of the patients' data to be used for the training set (patient number 1 to 5), validation set (patient number 6 to 10) and testing set (patients number 11 to 15), the higher testing specificity is obtained in Table 3.4. If the nature of testing data set is closely related to the training and validation sets, the testing specificity are expected to be similar to those in training and validation.

For evaluation purpose, γ analysis was introduced in (3.38). To meet with the minimum requirement of the proposed system (sensitivity $\geq$ 60% and specificity $\geq$ 40%), θ was set to 0.6 (which is equivalent to sensitivity 60%). By substituting θ in (3.38), the γ analysis becomes $\gamma = 0.6\xi + 0.4\eta$. The obtained mean sensitivity and specificity were evaluated in terms of γ analysis. As can be seen in Table 3.4, the proposed VTWNN outperformed other comparison methods by achieving gamma value 65.41%.

$$\gamma = \theta\xi + (1-\theta)\eta, \quad \theta \in [0.1,\ 1] \tag{3.38}$$

In order to prove that the optimized VTWNN gives better classification performance, the analysis is continuously carried out at different initial sensitivity and specificity at $\xi_l = 60\%$ and $\eta_l = 40\%$, $\xi_l = 40\%$ and $\eta_l = 60\%$ and $\xi_l = 20\%$ and $\eta_l = 80\%$. As presented in Table 3.5, the proposed VTWNN is working well for those defined target regions by achieving better testing sensitivity and acceptable specificity of 81.40 and 50.91 %. However, for WNN and FFNN, they can only give best testing sensitivity and specificity of (74.42 and 48.18 %) and (69.77 and 49.09 %). Similarity, at the defined specificity η_l = 60 and 80 %, the proposed VTWNN gives better sensitivity and specificity compared with other classifiers.

η_l		VTWNN	WNN	FFNN
40%	Sen (ξ)	**81.40**	74.42	69.77
	Spec (η)	**50.91**	48.18	49.09
	γ	**69.20**	63.92	61.50
60%	Sen (ξ)	**53.49**	48.84	44.88
	Spec (η)	**68.18**	72.73	70.00
	γ	**59.34**	58.40	54.93
80%	Sen (ξ)	**37.91**	23.16	20.93
	Spec (η)	**84.55**	90.17	93.28
	γ	**56.57**	49.96	49.87

Table 3.5: Best Testing Result for Hypoglycemia Detection with Different Approaches as η_l= 40%, 60%, and 80%

From Table 3.6, it can be seen that the optimized VTWNN achieves better testing sensitivity and specificity compared with other neural network classifiers (WNN, FFNN), evolved fuzzy interference system (FIS) and a genetic algorithm (GA) based multiple regression with fuzzy interference system (MR-FIS). As can be seen in Table 3.6, the optimized VTWNN gives the best testing sensitivity of 81.40 % while the specificity is kept around 50 %.

Based on the common criteria of clinical classification (sensitivity $\geq$ 0.6 and specificity $\geq$ 0.4), from Table 3.4 - 3.6, it can be distinctly seen that the optimized VTWNN is

effective and suited for detection of hypoglycemic episodes. In summary, the effectiveness of optimized VTWNN detection system can be distinctly seen in Table 3.4 - 3.6 with the better classification performance (81.40 % of Sensitivity and 50.91 % of acceptable testing specificity).

Methods	Sensitivity	Specificity	γ
VTWNN	**81.40**	**50.91**	**69.20**
MR-FIS [Ling2011]	75.00	50.00	65.00
WNN	74.42	48.18	63.92
FIS [Ling2010]	70.45	55.14	64.32
FFNN	69.77	49.09	61.50
MR	65.12	57.27	61.98

Table 3.6: Best Testing Result for Hypoglycemia Detection as η_l= 40%

3.5 Conclusion

For the detection of hypoglycemic episodes in T1DM patients, a hybrid particle swarm optimization based variable translation wavelet neural network (VTWNN) has been proposed in this chapter. An improved hybrid PSO is used to train the VTWNN parameters such as weights and parameters of nonlinear function. In the VTWNN, the translation parameter vary according to the input data. The network is therefore

made adaptive to the contingent changes of the environment. This adaptive characteristics makes the VTWNN better for problems with large sets of input data in a vast domain. Thanks to the variable property of translation parameter, the learning and generalization abilities of the proposed VTWNN have been enhanced. The comparison results with other conventional classifiers (WNN, FFNN, FIS, MR-FIS and MR) showed that the optimized VTWNN detection system gives better testing sensitivity of 81.40% and specificity of 50.91%.

Chapter 4

MR-based Combinational Neural Logic System and Its Application in Hypoglycemia Detection

4.1 Introduction

For hypoglycemia detection, a variable translation wavelet neural network (VTWNN) is presented in Chapter 3. Thanks to variable property, the performance of the VTWNN is better than conventional feedforward neural network structures. However, one main concern of the VTWNN is the computational demand which is reflected in the possibly larger number of network parameters than other networks. In this chapter, an integrated method, a neural logic network with multiple regression is

applied on the development of non-invasive hypoglycemia monitoring system.

Generally, conventional neural networks with the same structure were used to handle different applications. However, optimal performance was not always guaranteed due to different characteristics of applications. In addition, the redundant connections and weights of conventional neural network makes the number of network parameters unnecessarily large and downgrades the training performance [Leung2003]. In real-world application, the knowledge based neural network that understands all the characteristics of practical application is preferred for optimal performance.

Conventionally, the statistical regression model [Seber2003] [Freedman2005] has been widely carried out for classification [Chang2009] [Chu2008]. The learning process between a set of independent and dependent variables are modeled based on their relationship. However, it is not suitable to use if the patients' data is irregular and/or if there is no relationship between dependent and independent variables.

The modeling accuracy is only possible over the range of data where the model was developed. If it is used to model the irregular patients' data, the resulted regression model will have an unnaturally too wide possibility range. Despite both neural network and multiple regression models show advantages in dealing with classification, the limitations on each algorithm still remain to obtain the best classification model.

To overcome individual limitations of a conventional neural networks model and a

statistical regression model, a modified technique based on neural logic network combined with multiple regression model is proposed in this Chapter. The proposed network is targeted to dedicated application since its design based on the binary logic gates (AND, OR and NOT) [Lam2009], in which truth table and K-map are constructed depending on the knowledge of application. Because the logic theory is used in the network design, the structure becomes systematic and simpler compared to other conventional neural network and enhances the training performance. For a neural logic network with different structures for different applications, HPSOWM in Section 3.3 is selected as a suitable training algorithm.

In this hybrid system, through ECG signal alternation, the episodes of hypoglycemia were firstly predicted by the use of knowledge based-neural logic network (NLN) while the multiple regression model is used to enhance the sensitivity and specificity. By applying the proposed method, it is found that a combinational neural logic network can achieve better results with more understanding of the characteristics of application.

The Chapter is organized as follows: combinational neural logic system and multiple regression model, plus their training procedures by the use of HPSOWM are discussed in Sections 4.2 and 4.3 respectively. To show the effectiveness of proposed MR-based neural logic system, an application example will be given in Section 4.4. Finally, a chapter conclusion will be drawn in Section 4.5.

4.2 Design of MR-based Combinational Neural-Logic Network (NLN)

An integrated method based on neural logic network with multiple regression model is presented in this Section. As presented in Fig.4.1, there are two steps to develop the proposed system. For the first step, the approximated output is obtained by using the property of the combinational neural logic network structure. A statistical multiple regression is used in the second step for learning more about the relationship between the combined inputs, i.e., approximated output of first sub-system and remaining inputs and the overall system outputs. By applying the proposed method, it is found that a combinational neural-logic network can achieve better results with more understanding of the characteristics of application.

The design of NLN is based on the binary logic gates (AND, OR and NOT) [Lam2009], in which truth table and K-map are constructed depending on the knowledge of application. By using logic theory, some redundant connections and weights of combinational neural logic networks are reduced. The network structure becomes simpler compared to other conventional neural network and enhances the training performance.

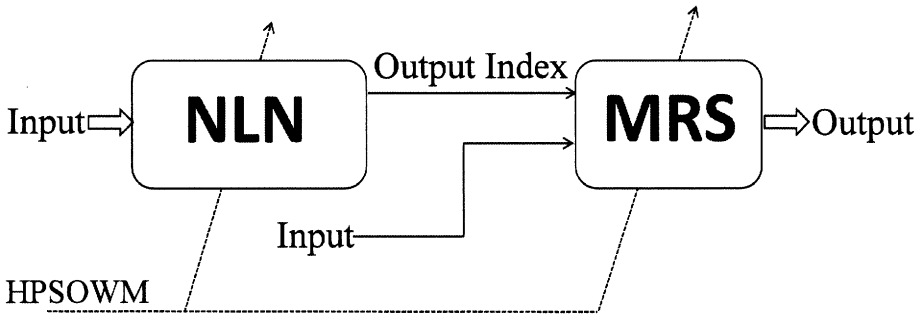

Figure 4.1: The proposed MR-based combinational neural logic system

4.2.1 Combinational Neural-Logic Network (NLN)

In this section, a combination neural logic network consisting of *rule based logic operation* and *neural network operation* will be discussed. As presented in Fig. 4.2, the proposed neural logic network mainly consists of *(1) rule-based logic-AND, -OR and -NOT operations, (2) neural network operation and (3) combination function.*

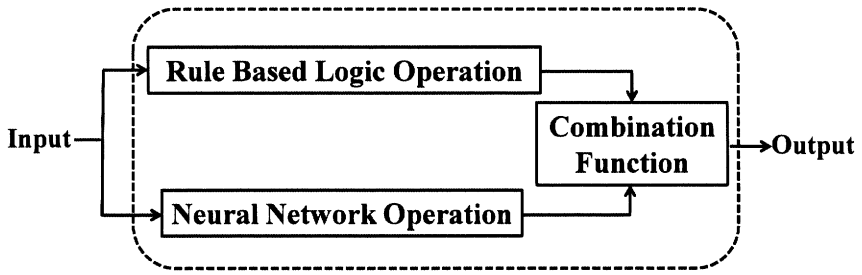

Figure 4.2: Internal Structure of Combinational Neural Logic System

In this hybrid system, the given system input will firstly undergo the *rule-based logic operation* which stores the boundary conditions and the properties of the logic-gates. The *neural network operation* is used to introduce the nonlinearity to the logic gates

for error correction. The neural logic network output is obtained via the *combination function* through the use of both rule-based logic operation and the neural network outputs.

4.2.1.1 Rule-Based AND Gate

For *rule based logic operation* in Fig. 4.2, a two-input-single-output logic AND gate is firstly proposed in this sub-section. Referring to Table 4.1, the neural logic AND operation is defined as:

$$u_1(t) \otimes u_2(t) = 0 \vee (u_1(t) + u_2(t) - 1) \in [0 \quad 1] \tag{4.1}$$

which is actually the bounded product of the input $u_1(t)$ and $u_2(t)$, where t denotes the current number of input vector which is a nonzero integer. $\vee$ and $\otimes$ denote the maximum operator AND operator. The inputs lie between 0 and 1 inclusively. The output of the neural-logic-AND operator will be fed to the rule base which guarantees the boundary conditions and the property of binary AND gates as shown in Table 4.1. Throughout this paper, the neural-logic-AND operator is denoted by a "o", e.g., $u_1(t)$ AND $u_2(t)$ is written as $u_1(t) \circ u_2(t)$.

$A \circ 0 = 0$	$A \circ 1 = A$	$A \circ \bar{A} = 0$	$A \circ A = A$
$0 \circ 0 = 0$	$0 \circ 1 = 0$	$1 \circ 0 = 0$	$1 \circ 1 = 1$

Table 4.1: Boundary Condition and Properties of The Neural Logic AND Gates

4.2.1.2 Rule-Based OR Gate

A two input single output logic-OR gate is presented for *rule based logic operation* in Fig. 4.2. Similarly, the inputs and output of rule-based OR gate lie between 0 and 1 inclusively. Referring to Table. 4.2, the logic-OR gate is defined as:

$$u_1(t) \oplus u_2(t) = (u_1(t) + u_2(t)) \wedge 1 \in [0 \quad 1] \tag{4.2}$$

which is actually the bounded sum of the inputs $x_1(t)$ and $x_2(t)$. $\oplus$ denotes OR operator. Its operation follows the properties of binary OR gates as shown in Table 4.2. The symbol, $\bullet$ is used as the OR operator throughout this paper, e.g., $u_1(t)$ OR $u_2(t)$ is written as $u_1(t) \bullet u_2(t)$.

$A \bullet 0 = A$	$A \bullet 1 = 1$	$A \bullet A = A$	$A \bullet \bar{A} = A$
$0 \bullet 0 = 0$	$0 \bullet 1 = 1$	$1 \bullet 0 = 1$	$1 \bullet 1 = 1$

Table 4.2: Boundary Condition and Properties of The Neural Logic OR Gates

4.2.1.3 Neural Network Operation

For *neural network operation* in Fig.4.2, a three layer fully connected neural network in Appendix A.1 is employed. As discussed, the input vector for proposed neural network is denoted as $U(t) = \left[u_1(t), u_2(t), \ \ldots \ u_{n_{in}(t)}\right]$ in which n_{in} is the number of inputs, for a two input-input-singale-output neural logic-AND gate n_{in} is equal to 2 ; v_{ij}, $j = 1, 2, \ldots, n_h$ denotes the weight between the input layer and the hidden

layer; n_h is the number of hidden neuron; w_{jl}, $l = 1, 2, \ldots, n_{out}$ denotes the number of output nodes. For two-input-single-output neural logic-AND gate, the n_{out}=1; b_j and b_l are the biases for hidden and output nodes.

The input-output relationship of a fully connected feedforward neural network in Fig.A.1.1, Appendix A.1 is calculated as follows:

$$y_l = f_l^2 \left(\sum_{j=1}^{n_h} w_{jl} f_j^{n_{in}} \left(\sum_{i=1}^{n_{in}} w_{ij} u_i - b_j \right) - b_l \right) \quad l = 1, 2, \ldots, n_{out} \tag{4.3}$$

The logarithmic sigmoid transfer function in (A.1.2) and hyperbolic tangent sigmoid transfer function in (A.1.3) are used in the hidden layer and output nodes activation functions $f_j^1(\cdot)$ and $f_l^2(\cdot)$. From (4.1) and (4.3), the input-output relationship of the neural logic-AND gate is defined as:

$$\delta_{AND}(t) = (0 \vee f_{AND}(u_1(t) \otimes u_2(t), y_l(t))) \wedge 1 \tag{4.4}$$

Consequently, the output of neural logic-OR gate can also be obtained from (4.3) and (4.2) as follows:

$$\delta_{OR}(t) = (0 \vee f_{OR}(u_1(t) \oplus u_2(t), y_l(t))) \wedge 1 \tag{4.5}$$

Subject to the properties of Table 4.1 and 4.2, the symbol $\wedge$ and $\vee$ denote the minimum and maximum operators. The combination function $f_{AND}(\cdot)$ and $f_{OR}(\cdot)$ for neural logic-AND gate and neural logic-OR gate to be designed based on the application requirements. They satisfy the boundary conditions and exhibit the properties of the binary AND and OR gates. Referring to Fig. 4.2, the output of rule based logic expression and neural network operation are defined as v_1 and v_2.

Similarly, by the use of the characteristics of binary logic NOT gate, the output of neural-logic-NOT gate can also be defined as:

$$\delta_{NOT}(t) \in [0 \quad 1] = 1 - u(t) \tag{4.6}$$

In (4.6), it can be seen that the characteristics of a binary logic NOT gate are still retained, i.e., NOT 0=1 and NOT 1=0. The neural-logic-NOT operator is denoted by a "bar". For instance, NOT *u(t)* is written as $\bar{u}(t)$

4.2.2 Design Example on Combinational Neural-Logic Network

In this section, the design of the combinational neural-logic system will be illustrated using a simple example. First, a truth table which governs the relationship among some linguistics stages, i.e., high (H) and low (L) are constructed. It should be noted that the linguistic variables, H and L are different from those in binary logic systems. H and L in binary logic systems refer to crisp 1 and 0. In combinational neural logic system, H and L refer to a state which is about 1 and 0. If a combinational neural logic circuit have two input (z_1 and z_2) and one output (δ), the two linguistic states for each input are H and L. A truth table for two input system is given in Table 4.3:

z_1	z_2	δ
L	L	H
L	H	H
H	L	H
H	H	L

Table 4.3: Truth Table: Design Example

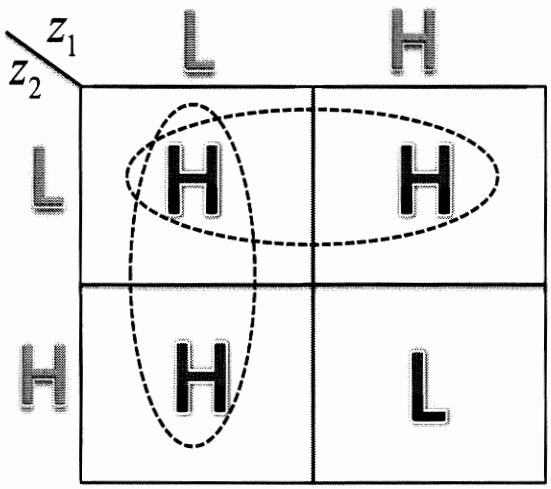

Figure 4.3: K map: Design Example

The two linguistic states such as H and L divide the input into two subregions 0 to 1. Taking the first row one of Table 4.3 as an example, it can be interpreted as if z_1 is about L and z_2 is about L, then δ is about H. These rules are determined based on human knowledge about the problem to be handled. Based on the truth table, a combinational neural logic system can be designed.

With reference to the state H, taking z_1 as an example, $\bar{z}_1$ denotes the neural logic

NOT of z_1. It means that z_1 is about H and $\bar{z}_1$ is about L in the neural logic function respectively. Referring to the truth tables in Table 4.3, the Karnaugh map (K-map) can be obtained as shown in Fig. 4.3.

By the use of the K-map in Fig. 4.3, the output δ can be obtained:

$$\delta = (\bar{z}_1 \circ \bar{z}_2) \bullet (\bar{z}_1 \circ z_2) \bullet (z_1 \circ \bar{z}_2) \tag{4.7}$$

Referring to the property of neural logic gates, (A $\bullet$ A = A) tabulated in Tables 4.1 and 4.2, (4.7) can be written as:

$$\begin{aligned} \delta &= (\bar{z}_1 \circ \bar{z}_2) \bullet (\bar{z}_1 \circ \bar{z}_2) \bullet (\bar{z}_1 \circ z_2) \bullet (z_1 \circ \bar{z}_2) && (4.8) \\ &= (\bar{z}_1 \circ (\bar{z}_2 \bullet z_2)) \bullet ((\bar{z}_1 \bullet z_1) \circ \bar{z}_2) && (4.9) \\ &= \bar{z}_1 \bullet \bar{z}_2 && (4.10) \end{aligned}$$

Remark 4.1 *The above application illustrates the idea of designing a two-input-single output combinational neural logic system. The idea can readily be extended to design a multiple-input-multiple-output combinational neural-logic system. It should be aware that the design of NLN, such as the construction of truth table and K-map in Table 4.4 and Fig. 4.5, the definition of combination function in Fig. 4.2, are largely depended on the characteristics of application.*

4.2.3 Multiple Regression Model

A multiple regression model [Freedman2005] introduced in this section is used to find the relationship between the system's inputs and outputs. Referring to Fig. 4.1, the inputs of regression model are the output of first subsystem (δ) and the rest of the inputs based on the design requirement. In general, multiple regression model procedures will be estimated in the following form:

$$G_i = \beta_0 + \beta_1 X_i + \beta_2 X_i^2 + \ldots + \beta_\eta X_i^\eta \tag{4.11}$$

where X_i and G_i are the system inputs and outputs, i represents the input-output dimensions, β denotes the parameters of the regression model, and η is the number of order. In order to find the optimized model parameters, HPSOWM in Section 3.3 is used.

4.2.4 Design Parameters of MR-based Neural Logic Network

A. *Number of Hidden Nodes (n_h)*: The size of the hidden layer is a general question raised on designing multilayer FFNN for real-life applications. An analytical method to determine the number of hidden nodes is difficult to obtain owing to the complexity of the network structure and the undetermined nature of the training process. Hence, the number of hidden nodes is experimentally found. In practice, the number of hidden nodes (n_h) depends on the application and the dimension of the input space.

B. *Network Parameters (Weights)*: The network parameters, weights and biases are v_{ij}, w_{jl}, b_j and b_l and a search method hybrid particle swarm optimization with wavelet mutation (HPSOWM) in Section 3.3 is used to find optimal vales.

C. *Parameter (β)*: The regression parameters/coefficients [β_0 β_1 β_2 ... β_η are nonrandom but they are unknown quantities. In general, these parameters are predicted and substituted in the MRS equation in (4.17) for model estimation. It is important for correct determination of *beta* in order to find out the relationship between dependent variable (G) and independent variables (X).

D. *Total number of parameters (n_{para1})*: The total number of network parameters are calculated by:

$$n_{para1} = (n_{in} + 1) \times n_h + (n_h + 1) \times n_{out} \tag{4.12}$$

E. *Total number of parameters (n_{para2})*: For regression model, the total number of parameter is largely depend on the number or order η, i.e., for second order, η is 2 and there are three tunable parameters [β_0 β_1 β_2] to be designed. The n_{para2} are calculated by:

$$n_{para2} = (2 \times n_{in}) + 1 \tag{4.13}$$

4.3 Training of MR-based NLN using HPSOWM

The parameters of the MR-based combinational neural logic system are optimized by HPSOWM in Section 3.3 using the same fitness function in (3.31) in Section 3.4.4. The optimized MR-NLN parameters are obtained through the maximization of defined fitness function in (3.31). During the training process, for a given set of particles $\mathbf{x} = [v_{ij};\ w_{jl};\ b_j;\ b_l;\ \beta]$, HPSOWM evaluates the fitness value of each particle at each iteration step and searches for the optimum network parameters. In order to train MR-NLN, the number of iterations is set at 1000.

4.4 Case Study in Hypoglycemia Detection System

To monitor the status of hypoglycemic episodes in T1DM, the proposed integrated method, a neural logic network with multiple regression in Fig. 4.4 is applied on the development of a non-invasive hypoglycemia monitoring system.

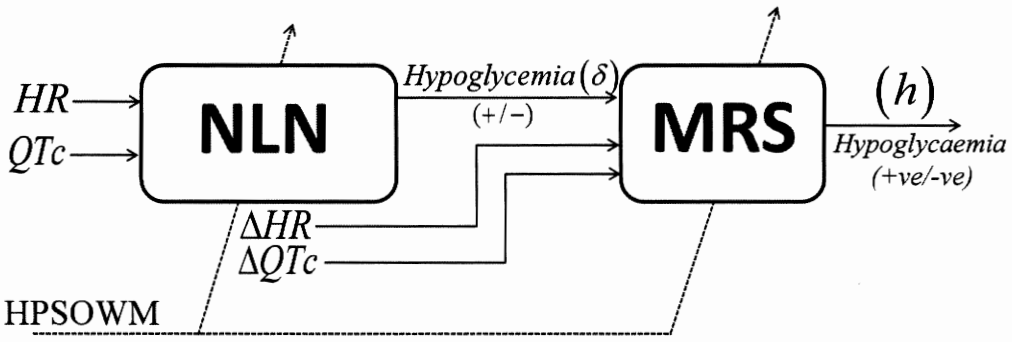

Figure 4.4: The proposed combinational neural logic network for hypoglycemia monitoring system

The inputs are four psychological parameters of ECG signal, the heart rate (HR) and the corrected QT interval (QTc), change of heart rate and change of corrected QT interval (ΔHR and ΔQTc) while the output represents the presence of hypoglycemia (h), $+1$ represents hypoglycemia and -1 is non-hypoglycemia. In this hybrid system, through alternation the alternation of physiological parameters such as HR and QTc, the episodes of hypoglycemia were firstly predicted by the use of knowledge based neural logic network (NLN) while the multiple regression model is used to enhance sensitivity and specificity by the use of ΔHR and ΔQTc. After the training process, the NLN gaves the predicted hypo-status (δ) and the multiple regression model continuously detects the presence hypoglycemia based on δ value.

To illustrate the design procedure and the merits of the proposed approach, the proposed MR-based combinational neural-logic system is employed to perform detection of hypoglycemia. Its design is based on expert knowledge of the task to be handled. A truth table is firstly set up which governs the relations among some linguistic states i.e., H for high and L for low. The H and L in binary logic systems refer to crisp 1 and 0 respectively. In the subsystem's first part, NLN subsystem in Fig. 4.4, the physiological parameters of ECG signal such as HR and QTc are used as the inputs and the δ is defined as the approximated output.

HR	**QTc**	δ
H	H	H
H	L	L
L	H	L
L	L	L

Table 4.4: Truth Table: Hypoglycemia Detection

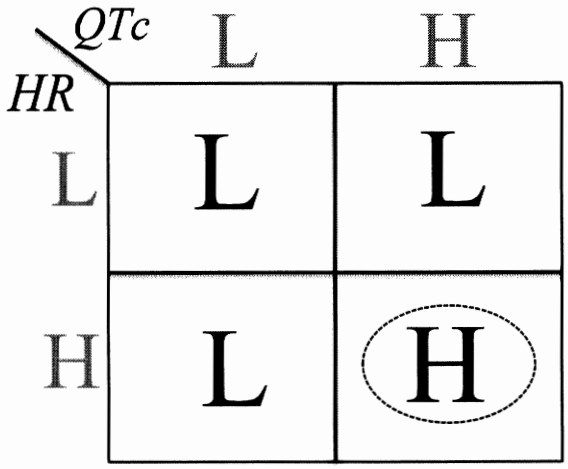

Figure 4.5: K map: Hypoglycemia Detection

The two linguistic states for each variable are H and L given in the truth table, Table 4.4. These two linguistics states divide the input into two subregions, 0 to 1. Depending on the characteristics of the applications, for a value of input HR, it takes as logic condition 1 (H) if $HR \in [1.01 \sim 2.575]$ whereas 0 (L) if $HR \in [0.479 \sim 1.0019]$. The same condition applied for QTc, it is defined as 1 (H) if $QTc \in [1.034 \sim 1.41]$ else 0 (L) if $QTc \in [0.7 \sim 1.033]$. For both HR and QTc, the values

which are out of these boundary conditions are taken as don't care condition (X). Picking up the first row in Table 4.4 as an example, it can be interpreted that if HR is about H (increased HR)and QTc (prolonged QTc) is about H, then δ is H. These rules are determined based on human knowledge about the problem to be handled.

With the help of the truth table in Table 4.4, the K-Map is constructed as shown in Fig. 4.5. Based on its own characteristics of this clinical application, the proposed NLN is designed to exactly follow the properties of logic-AND gates which rules are defined from Table 4.1. Thus, the following logic expression is obtained as the output of the first combination neural logic subsystem (δ) in Fig. 4.4 which is governed by the AND properties:

$$\delta = HR \circ QTc \tag{4.14}$$

In order to map the inputs to be $[-1 \;\; 1]$, the combination function in Fig. 4.2 is selected as "tansig", i.e., $f_{AND}(v_1, v_2) = tansig(v_1 + v_2)$ in which v_1 and v_2 are defined as the outputs of rule based logic operation and neural network operation. Other functions which could be replaced depend on the application. In this application, they are chosen experimentally to give a satisfactory result.

After the approximated output δ in Fig. 4.4 is obtained, the second subsystem called multiple regression model in Section. 4.2.3 is used to classify the presence of hypoglycemia based on the approximated output δ, ΔHR and ΔQTc. This model is used for fine-tuning of hypoglycemic detection performance due to the slight correlation of

ΔHR and ΔQTc. The advantages of the regression model are that simple structure and few parameters are needed.

4.4.1 Experimental Results and Discussion

In this study group of 15 patients with Type 1 diabetes, the same data are used as that in Section 3.4.2. Similarly, BGL $< 3.3mmol/l$ was considered a hypoglycemic episodes and the relationships between overnight measurements of QT interval corrected for heart rate (QTc) and falling BGL were revealed by the use of statistical correlation analysis given in sub-Section 3.4.5.1. In this analysis, the significant increment in HR, QTc in response to episodes of hypoglycemia gives p values of (1.033 ± 0.242 vs. $1.082 \pm 0.242 \pm < 0.06$) for HR and ($1.031 \pm 0.086$ vs. $1.060 \pm 0.084 < 0.001$) for QTc.

Table 3.1 discussed that among 15 patients, 12 T1DM patients were found to have their natural occurrence of hypoglycemic episodes for the defined BGL threshold level ($< 3.3mmol/l$) while 3 of them had no associated hypoglycemia events. It is because there were no naturally occurring episodes of hypoglycemia presented during measurement or they may not have been valid with the defined BGL threshold level.

Based on the patients information in Table 3.1, the overall data set is organized into the training set (patient number 1 to 5 with 184 data points), the validation set (patient number 6 to 10 with 192 data points) and the testing set (patient number

11 to 15 with 153 data points) which is the same condition in 3.4.5, Chapter 3.

In this clinical study, an MR-combinational neural logic network is systematically designed to incorporate the developed hybrid neural logic network structure to the characteristic of application. As presented in Fig.4.4, there are two steps to develop the proposed system for hypoglycemia detection. The first step is to determine approximated output (δ) based on the alternation of ECG parameters such as HR and QTc. In the first sub-system only two inputs are used because the design of the combinational neural logic network has two inputs and one output structure.

As discussed in the earlier section, for logic operation in the first sub-system, the logic-AND gates are selected based on their own characteristics of application. Referring to (4.4), the input-output relationship of the neural logic-AND gate is defined as:

$$\delta_{AND}(t) = (0 \vee f_{AND}(u_1(t) \otimes u_2(t), y(t))) \wedge 1 \tag{4.15}$$

As combinational neural logic system has two inputs and one output structure, y in (4.15) is defined from (4.3) as follows:

$$y = f_l^2\left(\sum_{j=1}^{n_h} w_{jl} f_j^{n_{in}}\left(\sum_{i=1}^{n_{in}} w_{ij}u_i - b_j\right) - b_l\right) \tag{4.16}$$

For $f_l^2(\cdot)$ and $f_j^1(\cdot)$, *logsig* (A.1.2) and *tansig* transfer functions (A.1.3) are used in the hidden layer and output nodes activation functions in order to map the output between 1 to -1.

Once the approximated output (δ_{AND}) is obtained, a multiple regression model is developed for fine-tuning of hypoglycemia detection system with the inputs (ΔHR, ΔQTc and δ) and output G. With these inputs and output, the multiple regression model in Section 4.2.3 becomes:

$$G_i = \beta_0 + \beta_1 X_i + \beta_2 X_i^2 \tag{4.17}$$

In this application, the input parameters X_i is defined as $X_i = [\delta_i \;\; \Delta HR \;\; \Delta QTc]$, $i = 1, 2, \ldots, n_{in}$. The output $G_i \geq 0$ represents positive hypoglycemia ($h = +1$) and $G_i < 0$ represents negative hypoglycemia ($h = -1$) which is expressed as:

$$h = \begin{cases} +1, & G_i \geq 0 \\ -1, & G_i < 0 \end{cases} \tag{4.18}$$

As discussed in Section4.3, all the design parameters of MR-based combinational neural logic system will be taken as the element of particles to perform the HPSOWM process. Different numbers of hidden nodes (7, 11 and 15) are tried for the feedforward neural network in neural network operation in Fig. 4.2. Finally, the optimal performance was obtained when the number of hidden nodes, n_h is 11. The lower and upper bound of the network parameters [v_{ij} w_{jl} b_j b_l] are given [-10 10]. For MR parameters, the lower and upper bound are given [-75 75]. The total number of tuned parameters for the first and second subsystem are about 45 and 7 parameters. In order to train MR-NLN, the number of iteration is set at 1000. The setting of HPSOWM parameters are the same that is chosen in subSection 3.4.5. In the following section, the best trained of MR-based combinational neural logic system with the

best set of parameters will be given.

4.4.1.1 Results Analysis

To evaluate the effectiveness of the proposed hypoglycemia monitoring system, five approaches such as (1) multiple regression based neural logic network (MR-NLN) (2) neural logic network (NLN) [Lam2007] (3) wavelet neural network (WNN) in Appendix A.2 (4) feedforward neural network (FFNN) in Appendix A.1 and (5) multiple regression (MR) in Section 4.2.3 are compared and analyzed.

In this clinical application, it is important to keep a higher value of sensitivity because it mainly identifies actual hypoglycemic episodes in patients with T1DM. Due to the importance of sensitivity, the fitness function (3.31) was designed and find the optimized sensitivity while the specificity was kept at acceptable value, i.e., sensitivity $\geq$ 60% and specificity $\geq$ 40% for this specific application.

To analyze this, the initial specificity (η_l) in (3.31) was set at 0.6 ($1 - \eta_l$=60%) which is equivalent to specificity, η_l=40%. At the defined initial condition, η_l=40%, the average (mean) value of training, validation and testing results were analyzed in terms of the sensitivity and specificity in Table 4.5. The best set of parameters among 20 runs will be employed to develop a hypoglycemia monitoring system. If the proposed MR-NLN can correctly detect the status of hypoglycemia, the sensitivity will be higher, ie., the higher sensitivity represents the better performance of proposed

classifier.

		MR-NLN (4 Inputs)	NLN (2 Inputs)	WNN (4 Inputs)	FFNN (4 Inputs)	MR (4 Inputs
Training	Sen (ξ)	90.91%	90.23%	84.12%	83.64%	81.82%
	Spec (η)	40.47%	41.28%	40.63%	40.50%	41.32%
Validation	Sen (ξ)	90.62%	86.81%	80.44%	79.07%	82.19%
	Spec (η)	40.29%	40.00%	40.94%	41.38%	40.75%
Testing	Sen (ξ)	78.81%	74.93%	71.39%	68.84%	64.65%
	Spec (η)	51.45%	53.83%	44.37%	48.34%	53.09%
	Gamma (γ)	67.86%	66.49%	60.58%	60.04%	60.03%

Table 4.5: Mean Value of Training, Validation and Testing Results as η_l= 40%

As presented in Table 4.5, the average (mean) testing result of proposed MR-NLN was found to be satisfactory with better testing sensitivity and specificity, (78.81 and 51.45%) compared to the other traditional classifiers such as NLN (74.93 and 53.83%), WNN (71.39 and 44.37%), FFNN (68.84 and 48.34%) and MR (64.65 and 53.09%). By comparing with MR-NLN and NLN, MR-NLN gives a better sensitivity which implied that the sensitivity of the monitoring system was enhanced by the inputs of ΔHR and ΔQTc.

In this application, it seems that the testing specificity in Table 4.5 (51.45%) is higher

than the training specificity which is about 40.47%. It is because of the nature of patients' data that is specified for testing set. As can be seen in Table 3.1, due to different nature of patients' data are used for the training set (patient number 1 to 5), validation set (patient number 6 to 10) and testing set (patient number 11 to 15), the higher testing specificity is obtained. If the nature of testing data set is closely related to the training and validation sets, the testing specificity are expected to be similar to those in training and validation.

	No.of Inputs	Sen (ξ)	Spec (η)	Gamma (γ)
MR-NLN	4	79.07%	53.64%	68.90%
NLN	2	76.74%	54.55%	67.86%
MR-FIS [Ling2011]	4	75.00%	50.00%	65.00%
WNN	4	74.42%	48.18%	63.92%
FIS [Ling2010]	4	70.45%	55.14%	64.32%
FFNN	4	69.77%	49.09%	61.50%
MR	4	65.12%	57.27%	61.98%

Table 4.6: Best Testing Result for Hypoglycemia Detection as η_l= 40%

Generally speaking, if more patients' data (up to 100 patients) could be collected, the testing specificity may come closer to the training specificity. However, all the actual data were collected at the Princess Margaret Hospital for Children in Perth, Western

Australia, Australia and obtaining more patients data is currently not possible due to the limited number of volunteered children with Type 1 diabetes and the limited funding supports. It should be considered as the limitation of this application.

The γ analysis was introduced as discussed in Section 3.4.5.2, i.e., $\gamma = \theta\xi + (1-\theta)\eta$, ($\theta \in [0.1, 1]$) for evaluation of proposed system performance. By Substituting the minimum requirement of proposed system which is around sensitivity $\geq$ 60% and specificity $\geq$ 40%, the γ analysis becomes $\gamma = 0.6\xi + 0.4\eta$. Based on this γ analysis, as can be seen in Table 4.5, the proposed MR-NLN outperformed other comparison methods by achieving γ value 67.86%.

After the training process, the optimized MR-NLN achieved the better testing sensitivity, 79.09% and acceptable specificity, 53.64%. It is because the network was systematically designed to incorporate the characteristics of the application into the structure of combinational neural logic system. As can be seen in Table 4.6, the proposed MR-NLN gives better sensitivity and specificity compared with the conventional neural networks (WNN, FFNN), evolved fuzzy interference system (FIS) [Ling2010], genetic algorithm (GA) based multiple regression with fuzzy interference system (MR-FIS) [Ling2011] and MR.

To validate the effectiveness of the proposed method, the analysis were continuously carried out by setting the initial specificity, η_l at 50%. As presented in Table 4.7,

	No.of Inputs	Sen (ξ)	Spec (η)	Gamma (γ)
MR-NLN	4	69.77%	60.91%	66.23%
NLN	2	62.79%	62.73%	62.76%
WNN	4	62.79%	55.85%	60.05%
FFNN	4	60.47%	60.00%	60.28%
MR	4	65.12%	57.27%	61.98%

Table 4.7: Best Testing Result for Hypoglycemia Detection as η_l= 50%

the proposed MR-NLN gave larger gamma (γ) value, 66.23% compared to other conventional neural networks. In short, the proposed MR-NLN can successfully detect the status of hypoglycemia by achieving best sensitivity and specificity at 79.07 and 53.64%.

4.5 Conclusion

In this paper, multiple regression based-neural logic network is developed for hypoglycemia monitoring system by considering changes in physiological parameters of ECG signal during hypoglycemia and non-hypoglycemia reactions. Traditionally, neural networks with the same structure were applied to handle every application and may not give the optimal solution due to different characteristics of applications.

In these two subsystems, the approximated hypoglycemia status, δ is firstly obtained

through the relationship between the inputs HR and QTc. Once the approximated output is obtained through the training process, a multiple regression model is used to detect the status of hypoglycemia with the inputs δ, ΔHR and ΔQTc. The results in Section 5.4.1.1 indicate that the hypoglycemia episodes in T1DM children can be efficiently detected non-invasively and continuously by achieving better sensitivity (79.07%) and specificity (53.64%).

Chapter 5

Evolvable Rough-Block-Based Neural Network for Non-Invasive Hypoglycemia Detection

5.1 Introduction

On the last two Chapters 3 and 4, the variable translation wavelet neural network (VTWNN) and a neural logic network with multiple regression have been presented. With these improved networks, the classification performance was improved over conventional neural networks. However, those neural network models offer far fewer degrees of freedom due to their network structure. Conceptually, the number of degree of freedom allows the network to be more adaptive with the characteristics of

application. If there are less degrees of freedom, the model becomes less accurate.

Therefore, the choice of an appropriate architecture of neural network plays an important role. In [Nielsen1990], it has been proposed that several neural network architectures depend on the characteristics of their applications. For instance, the specialized structure of recurrent neural networks choose certain connections between neurons which are particularly relevant to given data sets [Kim1998]. Using such specialized structures it is possible to increase the efficiency and learning process of neural networks.

As different structures of neural networks are needed for different applications with different characteristics, the neural network should be designed based on the characteristics of applications in order to obtain optimal performance [Lippmann1987] [Kinoshita1987]. Though neural network itself has advantages in self-adaption, better generalization capability and ability to work in noisy environments, there still remains some problems in managing the architecture of the network and accelerating the training of the network [Lingras1998].

The introduction of rough set theory has been regarded as an effective approach in order to deal with modeling vagueness and uncertainties. It has proved that its soundness and usefulness has been successfully applied in many fields, such as machine learning, data mining, data analysis, expert systems, knowledge acquisition and pattern recognition [Pawlak1982] [Pawlak1991]. It is also a useful tool for pre-processing

data by applying its concept of approximation with a pair of set, lower and upper approximation of the set [Hassan2010].

Both rough set theory and neural networks show advantages in dealing with imprecise and incomplete knowledge. However, neural networks have complex structures when dimensions of input data are large while rough sets have the advantage of decreasing redundance among input data sets [Liu2004]. On the other hand, rough sets have poor generalization and weak tolerance whereas neural networks have better generalization ability and self-organization performance. By simplifying the input of neural network with the use of rough set algorithm, the number of training samples is reduced. Consequently, the size of the whole network structure is reduced and obtains faster convergence [Jagielska1999].

Dealing with individual limitations, a hybrid intelligent system is created by integrating rough set techniques in the architecture of the neural network. This is due to the ability of rough sets to discover patterns in ambiguous data and to provide tools for data and pattern analysis [Jagielska1998]. In [Hassan2002], a hybrid method is developed by the integration of rough set theory to a neural network system which is specifically used for decision making and classification. In terms of the advantages of rough sets and neural networks, the hybrid system has been widely applied in the field of pattern recognition, image retrieval [Jaroslaw2003] and medical diagnosis [Zhang2009].

In [Zhao2011] a new classifier, rough set based back propagation neural network is presented for bearing a fault diagnostic system. It has been illustrated that the hybrid system is able to classify the data effectively and achieves the goal of bearing fault classification. To estimate highway traffic, a rough neural network with upper bound neuron and lower bound neuron are introduced [Chandana2005]. In addition, a medical diagnostic support system, especially working as a cancer detection system is designed using extended rough neural network and multiagent [Yamaguchi2008].

Alternatively a hybridization methodology involving neural, fuzzy and rough approximation concepts, so called neuro-fuzzy inference systems has been developed [Chandana2007]. It is tested on the problem of approximating theoretical aerodynamic equations of various parameters. The results are also confirmed that the number of training epochs are considerably reduced by the use of a rough approximation based neuro-fuzzy inference system. Although many hybrid systems based on rough sets and neural networks have been proposed, there are still some problems in feature selection and classification performances.

In this Chapter, a new rough approximation block-based neural network (R-BBNN) is designed based on the concept of rough set properties, partitioning the applied input signal to a predictable part (certain) and the random (uncertain) part. Based on input data information, the *lower region* is defined with predictable data while the *boundary region* is considered with randomness and uncertainty of data. By the use of this lower and boundary regions, the block-based neural network is designed

to deal only with the *boundary region* which mainly consists of the random part of the applied input signal. Such approach improves the classification accuracy.

Owing to different characteristics of neural network (NN) applications, a traditional neural network with a common structure may not be able to handle every application. Based on the knowledge of application, BBNN is selected as a suitable classifier due to its modular characteristics and ability in evolving the size of the network by adding more basic blocks. The structure of the network simply corresponds to determining signal flows between each block [Moon2001]. The design parameters of proposed R-BBNN are learned by the use of HPSOWM in Section 3.3. Case studies on hypoglycemia detection are employed to demonstrate the better performance which is achieved by rough-block-based neural network (R-BBNN).

The organization of this Chapter is as follows: in Sections 5.2 and 5.3, the design and learning of the proposed rough-block-based neural network (R-BBNN) by the use of HPSOWM is presented. The detection of nocturnal hypoglycemic episodes in T1DM by the use of the proposed R-BBNN method is discussed in Section 5.4. To show the effectiveness of the method, several experiments are conducted, compared and analyzed with different kinds of neural networks in subSection 5.4.1. Finally a conclusion is written in Section 5.5.

5.2 Design and Architecture of Rough-Block-Based Neural Network

The hybridization technology using rough sets concepts and neural computing for decision and classification are presented in this section. Neural network has been widely used as a universal function approximator due to its excellent property in dealing with complex function approximation [Jain1996]. However longer training time and complexity in neural computation become limiting factors when handling neural networks in every application [Yuliang2009]. Recently much attention has been given to accelerate the process of neural network learning and reduce complexity by carrying out a pre-process stage before the network training.

As presented in Fig. 5.1, in order to carry out the pre-processing stage, the *lower region* and *boundary region* are defined based on the rough set concepts. By the use of defined rough regions, the input signal is partitioned to predictable (certain) part and random (uncertain) part. In this way the neural network is designed to deal only with the *boundary region* which mainly consists of the random part of applied input signal. Such architecture has the ability to reduce the size of NN input and scale down the whole structure of the network [Pawlak2002]. It satisfies the fact that a smaller network usually requires shorter learning time.

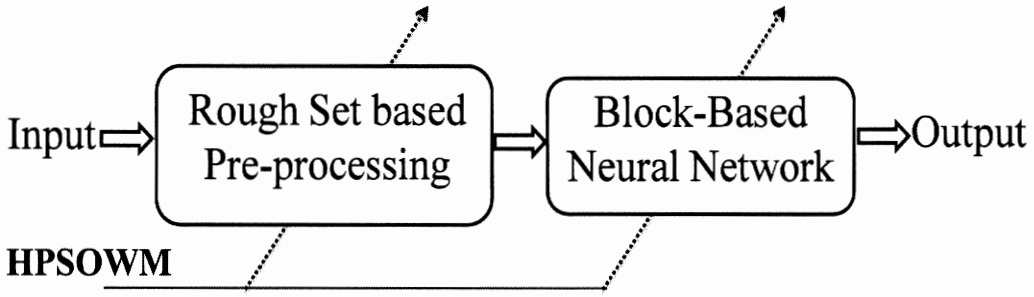

Figure 5.1: Proposed Rough-Block-Based neural network (R-BBNN)

In order to unite the characteristics of application to the structure of neural network, block-based neural network (BBNN) is selected as a suitable classifier due to its ability in evolving internal structures and adaptability in dynamic environments [Jiang2004]. The proposed system in Fig. 5.1 is systematically designed to incorporate the characteristics of application to the structure of hybrid rough-block-based neural network (R-BBNN). A global training algorithm, HPSOWM is introduced for parameter optimization of proposed R-BBNN.

5.2.1 Rough Set Preliminaries

The rough sets theory predominantly deals with the classification analysis of imprecise, uncertain or incomplete information [Lingras1996]. It offers effective methods that are applicable in many branches of artificial intelligence (AI) technologies. Its soundness and usefulness has been proven in many real life applications.

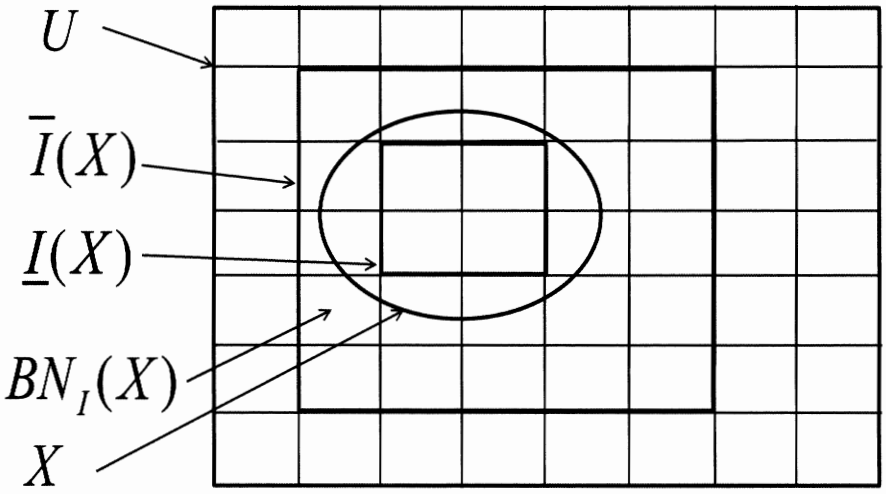

Figure 5.2: Rough Set Approximation

As shown in Fig. 5.2, the universe, (U) be a finite, non-empty set with, I, being an equivalence relation called the indiscernibility relation on U. $I(x)$ would then be described as an equivalence class of the relation I containing the element x. The concept of an indiscernibility relation brings about the fact that not all elements in the universe, U can be discerned given the information available. Further, such an indiscernibility relation is used to determine the *lower*, *upper* and *boundary* approximations which are expressed as:

$$\overline{I}(X) = \{x \in U : I(x) \subseteq X\} \tag{5.1}$$

$$\underline{I}(X) = \{x \in U : I(x) \cap X \neq 0\} \tag{5.2}$$

$$BN_I(X) = \overline{I}(X) - \underline{I}(X) \tag{5.3}$$

where $\overline{I}(X)$ and $\underline{I}(X)$ are defined as the *lower* and *upper region* of approximation

region while the *boundary region* is denoted as the disjoint between lower and upper approximation. Decision rules describe data patterns which are represented in the form: $IF\ \alpha\ THEN\ \beta$, where α is the condition part and β is the predicted class.

The optimized lower and upper boundary parameters are obtained by using a global learning algorithm, HPSOWM in which elements of particle are organized with the boundary parameters, $\delta \in \left[\overline{I}(X),\ \underline{I}(X)\right]$ and give optimal values for those elements of particle. A detailed discussion on HPSOWM is given in Section 3.3.

5.2.2 Topology of Block Based Neural Network

A block-based neural network (BBNN) is a network design that is more flexible for changing the structure depending on the signal flow between blocks. It can be represented by a 2-D array of blocks and each individual neuron block works as a basic signal processing unit that is composed of a feedforward neural network having four variable input/output nodes [Moon2002].

As shown in Fig. 5.3, the structure of BBNN is organized with m rows and n columns of blocks which are labeled as B_{ij}, in which $(i = 1, \ldots m)$ and $(j = 1, \ldots, n)$. The first column of blocks, $B_{11}, B_{21}, B_{31} \ldots, B_{m1}$ is determined as the input layer of BBNN network structure while $B_{1n}, B_{2n}, B_{3n} \ldots, B_{mn}$ is represented as output layer. The output of BBNN $(y = y_1, y_2, y_3, \ldots, y_m)$ is a function of summation of weighted inputs and a bias which is characterized by feedforward neural network architecture.

A constant input value is given to redundant input nodes and un-used output nodes are ignored.

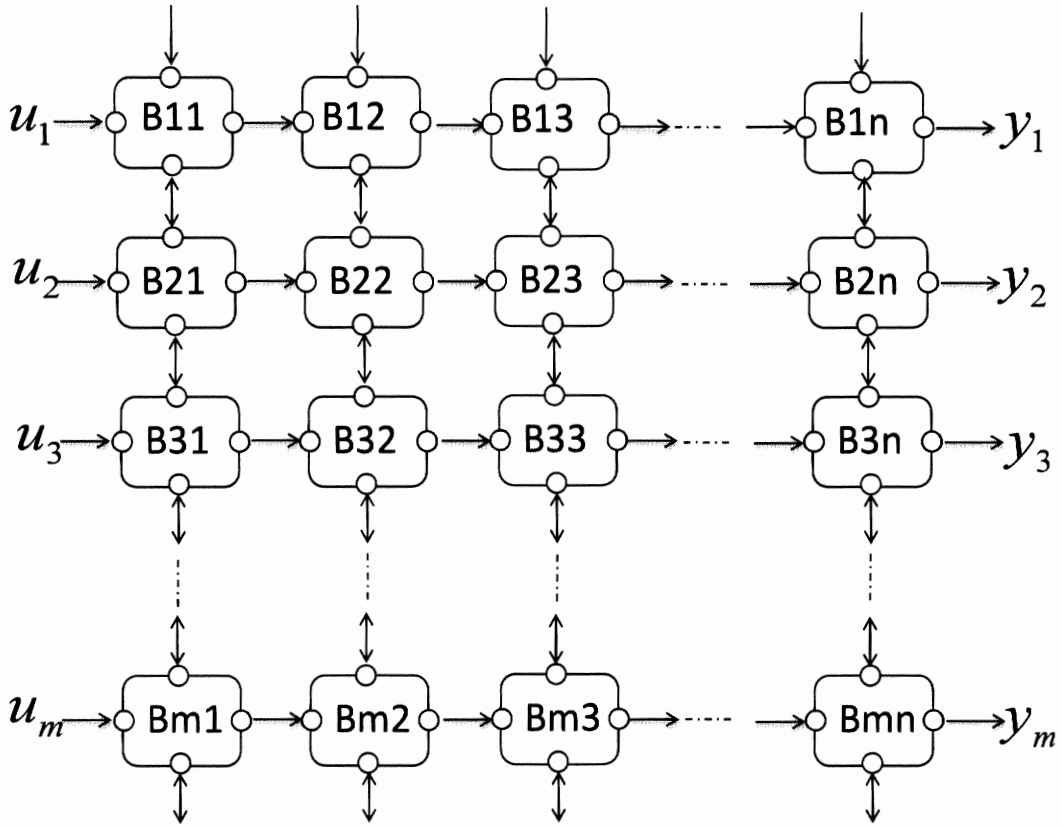

Figure 5.3: Structure of Block-Based Neural Network

Due to modular characteristics, BBNN can be easily expanded to be a larger network. In Fig. 5.3, a block is connected to its neighboring blocks with signal flow represented by arrows: $\downarrow$ represent as 0 while $\uparrow$ and $\rightarrow$ assign as 1. In this work, the connections between layers will only be considered as feedforward configuration in a forward direction. The structure of BBNN is determined by automatically internal configuration or input-output connection of basic blocks.

5.2.2.1 Four Different Internal Configurations of BBNN

According to the input-output connections of the network structure, the block has four different types of internal configuration. Fig. 5.4 (a) and (d) represent two inputs and two outputs with different internal configurations while (b) and (c) correspond to one input-three outputs and three-output configurations. The capability of generalization is improved through various internal configurations of a block. Even though each basic block should be one input and three outputs (1/3), three inputs and one output (3/1), two inputs and two outputs (2/2), the extreme cases of all input nodes (4/0) or all output nodes (0/4) are considered as invalid configurations.

For each basic block consists of four nodes and all node inside the block are connected with each other through the connection of weights, w_{ij}. A block can have up to a maximum of six connection weights including the biases. For the case of two inputs and two outputs in Fig. 5.4 (b) and (d), there are four weights and two biases. It is the same as in Fig. 5.4 (c), one input and three outputs structure which has three weights and three biases.

For three inputs and one output in Fig. 5.4 (a), it belongs to three weights and one bias. Each node of a block can be an input or an output according to internal configuration which is determined by the signal flow. An incoming arrow to a block is defined as the input node and its associated outputs are considered as the block outputs with outgoing arrows. The output of each block is connected to another block

as the input signal.

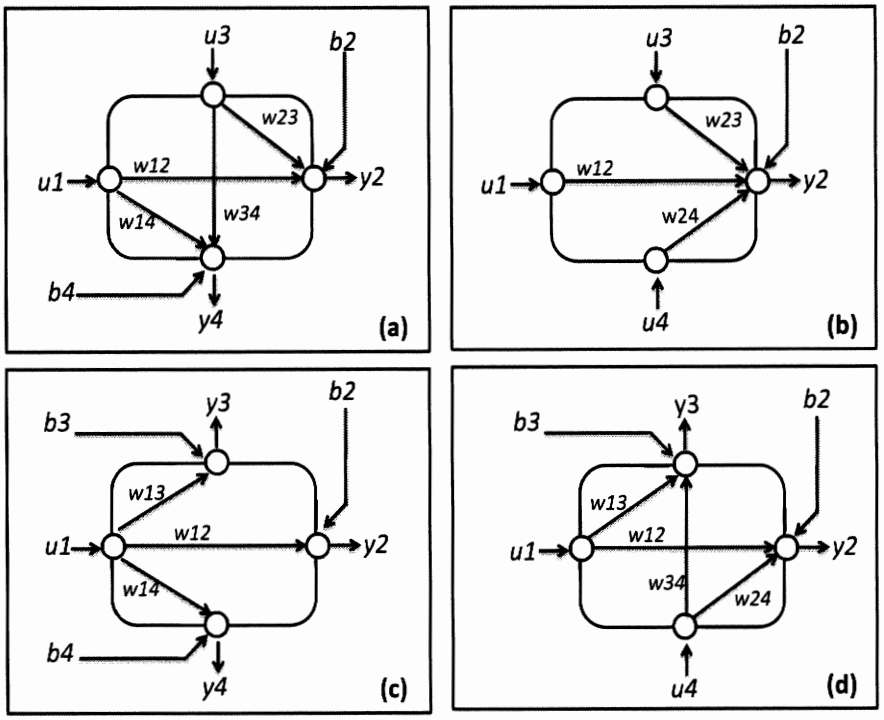

Figure 5.4: Internal configurations of block-based neural network

The output of the block is calculated by the summation of weighted inputs and biases corresponding to a feedforward neural network in Section A.1 as follows:

$$y_j = \sum_{i \in I} w_{ij} u_i + b_j, \quad j \in J \tag{5.4}$$

where u_i is the input to node i, w_{ij} is a weight connection from node i to node j. b_j is the bias, I and J are the two index set for input and output nodes. For (1/3) block $I = 1$ and $J = 2, 3, 4$. For block type of (3/1), $I = 1, 3, 4$ and $J = 2$ while the (2/2) block type has $I = 1, 4$ and $J = 2, 3$. For each node, it is characterized by the

following activation function, $\varphi(\cdot)$:

$$\varphi(u) = \alpha\left(\frac{2}{1+e^{-\beta u}} - 1\right) \tag{5.5}$$

where α and β are the transfer function parameters. The capability of network generalization is improved through various internal configurations i.e., two inputs and two outputs (Fig. 5.4 (a) and (d)), three inputs and one output (Fig. 5.4 (b)) and one input and three outputs (Fig. 5.4 (c)).

The assignment of I and J in (5.4) depend on internal network configurations. For instance, in (2/2) configuration in Fig. 5.4 (a) and (d), $I = 1,3$ and $J = 2,4$ and $I = 1,4$ and $J = 2,3$, while (3/1) and (1/3) combination in Fig. 5.4 (b) and (c) have $I = 1,3,4$ and $J = 2$ and $I = 1$ and $J = 2,3,4$. The extreme cases of four input nodes (4/0) or four output nodes (0/4) are considered as invalid configurations [Jiang2007].

For instance, the output y_j of block in Fig. 5.4 (a) having the two inputs, u_1 and u_3 and the two outputs, y_2 and y_4 is calculated as follows:

$$\begin{aligned} y_2 &= w_{12}u_1 + w_{32}u_3 + b_2 \\ y_4 &= w_{14}u_1 + w_{34}u_3 + b_3 \end{aligned} \tag{5.6}$$

Likewise, the output y_j of other type of configurations in Fig. 5.4 (b), (c) and (d) can be calculated according to their respective value of I and J. A basic building block consists of four nodes which have cascaded each other through a connection weights, w_{ij}. For each block, up to six connection weights and biases are available. In

this study, the connection between layers will only be considered in forward direction which is automatically determined by internal configurations or input-output signal flows ($\downarrow$ represents as 0 while $\uparrow$ and $\rightarrow$ assign as 1).

5.2.3 Design Parameters of Rough-Block-Based Neural Network

A. *Number of Hidden layers (n_l)*: The selection of the number of hidden layer is the same way in which conventional feedforward neural networks are chosen. In a conventional neural network, the number of hidden nodes is chosen by a trial and error method until the optimal solution is obtained. The same conditions are applied, so an optimal number of hidden nodes is selected by a trial and error process till the desired condition is met. However there is a trade off between too small a number or too big a number of hidden layers, because a small number of hidden layers could be insufficient to reach the optimal solution and a larger number of hidden layers may increase the number of parameters to be tuned and would require more computation time.

B. *Network Parameters (Weights)*: The network parameters are weights, biases and transfer function parameters which are w_{ij}, b_j, α and β. A search method hybrid particle swarm optimization with wavelet mutation (HPSOWM) in Section 3.3 is used to find optimal values.

C. *Rough boundary parameters (δ)*: Based on the nature of the application data,

the parameter δ are defined for positive lower region, $\underline{I}_{+}(X)$ and negative lower region, $\underline{I}_{-}(X)$ in (5.1)-(5.3). The correct determination of those boundary regions is important because partitioning of the input data to certain (predictable) or uncertain (random) parts is largely depend on them.

D. *Total number of parameters (n_{para})*: The total number of R-BBNN parameters are calculated by:

$$\begin{aligned} n_{para} &= ((n_{in}-1)\times n_l)+((7\times n_{in})\times n_l)+(4\times n_{in})+4 \\ &= n_l(8\times n_{in}-1)+4\times(n_{in}+1) \end{aligned} \tag{5.7}$$

5.3 Training of Rough-Block-Based Neural Network

Similarly, the parameters of the R-BBNN are optimized by HPSOWM in Section 3.3 using the same fitness function in (3.31) in Section 3.4.4. The parameters of R-BBNN are obtained through the maximization of defined fitness function in (3.31). During the training process, for a given set of particle $\mathbf{x} = [w_{ij};\ \alpha;\ \beta;\ \delta]$, HPSOWM evaluates the fitness value of the each particle at each iteration step and searches for the optimum network parameters. In order to train R-BBNN, the number of iteration is set at 200.

5.4 Case Study in Hypoglycemia Detection System

A case study in hypoglycemia detection system will be given in this section. The proposed rough-block-based neural network (R-BBNN) is employed to perform hypoglycemia detection in Type 1 diabetes mellitus (T1DM). Based on the knowledge of application, the hypoglycemia detection system is designed as 4 inputs and 1 output system as shown in Fig. 5.5. The four physiological inputs are heart rate (HR), corrected QT interval (QT_c), change in heart rate (ΔHR) and change in corrected QT interval (ΔQT_c) while the output represent the status of hypoglycemia ($+1$ is presence of hypoglycemia and -1 represent non-hypoglycemia).

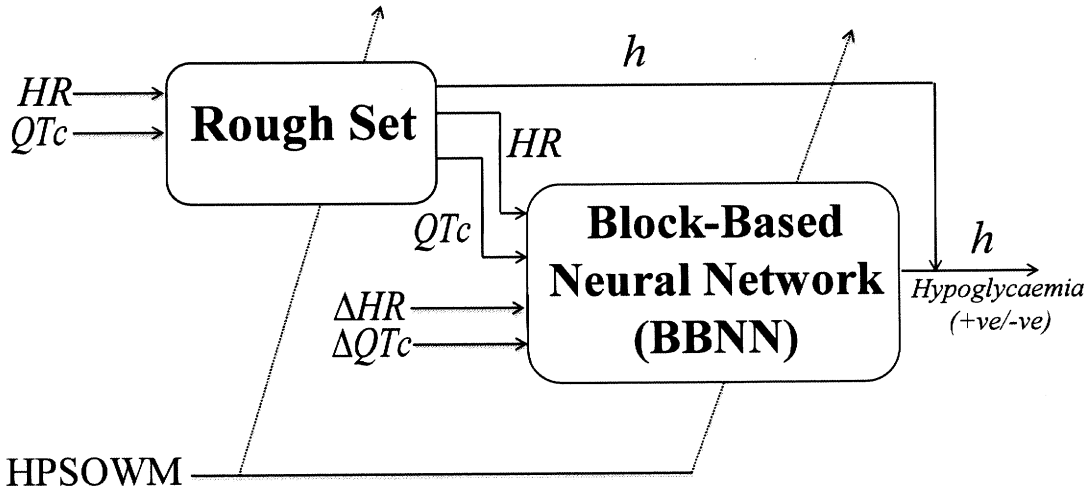

Figure 5.5: Hypoglycemia detection system using R-BBNN

As shown in Fig. 5.6, in each rough region either (a) or (b), the first half region is defined as the *lower region* which mainly captures definite or certain parts of applied input signal. It is also responsible for output approximation. The second half

region is the *boundary region* which mainly deals with the random part of applied input signal. Once lower and boundary segregation have taken place, the block-based neural network (BBNN) will deal with the *boundary region* consisting of randomness and uncertainty, to reduce inaccurate modeling of data. The construction process of proposed R-BBNN is presented as follows:

1. Partitioning of the input data to certain (predictable) or uncertain (random) parts based on rough regions which are made up of a combination of lower region and boundary region, (5.1)-(5.3). In order to meet with the objective of clinical application, lower region (5.1) is defined as positive lower region, $\underline{I}_{+}(X)$ and negative lower region, $\underline{I}_{-}(X)$ in order to perform the output prediction.

2. Prediction of output (status of hypoglycemia $(+/-)$ by using $(+/-)$ lower region) simplifies the neural network input i.e., if the applied input signal is within the positive lower region, the approximated the output is given as $+$ hypoglycemia whereas it approximates as $-$ hypoglycemia if the applied input signal is within the negative lower region.

3. The rest of the elements which are associated with randomness and uncertainty are categorized under the boundary region (5.3) and a simultaneous classification task is performed by block-based neural network (BBNN) Section 5.2.2.

4. Network training is performed by the use of HPSOWM to obtain optimized parameters. Each particle $x_j^p(t)$ in the swarm $X(t)$ is encoded with rough

boundary parameters and network parameters as $\mathbf{x} = [\delta\ s_l\ w_k^l\ b_k^l\ \alpha\ \beta]$.

5. After the training process, the optimal rough boundary parameters, δ from (5.1)-(5.3), BBNN parameters, structure, weight and biases, s_l, w_k^l and b_k^l from (5.4), transfer function parameters, α and β from (5.5) are obtained for the proposed R-BBNN detection system. The process is repeated until no improvement in classification performance has been achieved.

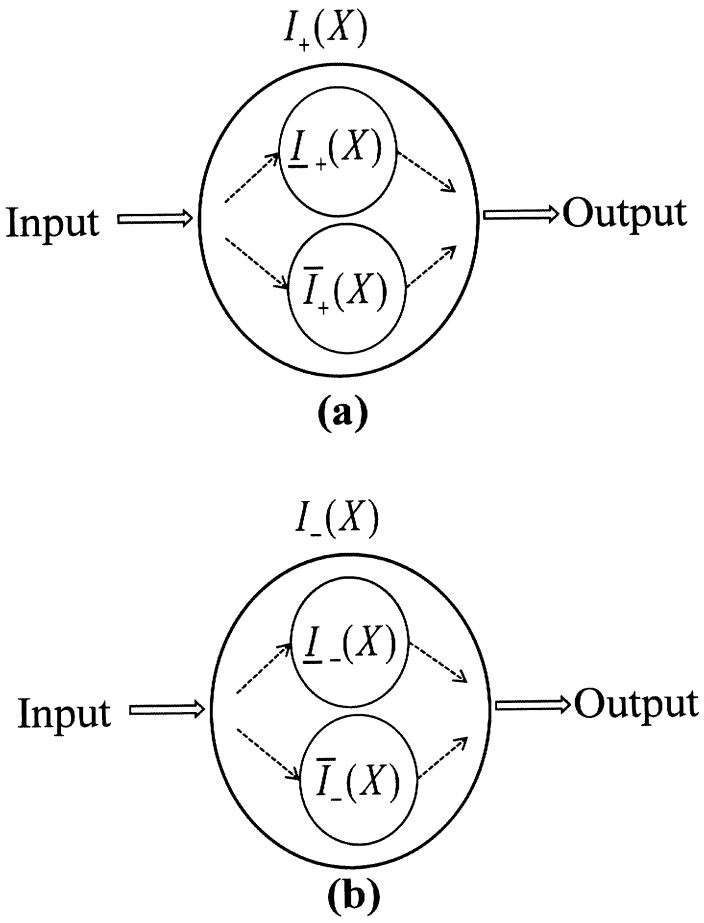

Figure 5.6: (a) Positive lower boundary regions (b) Negative lower and boundary regions

5.4.1 Experimental Results and Discussion

In subsection 3.4.5.1, the statistical results on normalized heart rates, correct QT interval and their association with hypoglycemia has significant increases in heart rate and corrected QT interval by giving p value of $HR < 0.06$ and $QTc < 0.001$. From 15 patients' data in Table 3.1, the training set (patient number 1 to 5 with 184 data points), the validation set (patient number 6 to 10 with 192 data points) and the testing set (patient number 11 to 15 with 153 data points) are organized as in Sections 3.4.5 and 4.4.1 respectively. In this section, hypoglycemia episodes (BGL $\leq 3.3 mmol/l$) are detected using the proposed hybrid rough set based block neural network as presented in Fig. 5.5.

Firstly, the applied input signal such as HR, QT, ΔHR and ΔQTc will undergo rough set operation in order to partition certain (predictable) or uncertain (random) parts. Based on the characteristics of this clinical application,i.e., based on the nature of the applied inputs (HR, QT, ΔHR and ΔQTc) information, the boundary parameters for rough set operation are defined. For the input HR, the upper and lower positive regions are defined as $\overline{I}_{HR^+} \in [0.5068 \sim 2.4736]$ and $\underline{I}_{HR^+} \in [0.980054 \sim 2.31798]$, whereas $\overline{I}_{HR^-} \in [0.5068 \sim 2.4736]$ and $\underline{I}_{HR^-} \in [0.609376 \sim 0.94962]$ are defined for negative lower and upper regions.

similarly, the *lower* and *upper* region for input QT is defined $\overline{I}_{QT^+} \in [0.5068 \sim 2.4736]$ and $\underline{I}_{QT^+} \in [0.980054 \sim 2.31798]$, $\overline{I}_{QT^-} \in [0.5068 \sim 2.4736]$, $\underline{I}_{QT^-} \in [0.609376 \sim 0.94962]$

in which $\overline{I}_{QT+}$ and $\underline{I}_{QT-}$ represent as defined positive and negative regions. In this application, the inputs ΔHR and ΔQTc are used to improve sensitivity and specificity because they have information regarding time changes of HR and QTc. By feeding them to BBNN as inputs, the network obtain more information and can do classification tasks accurately.

The output prediction is carried out based on those rough regions, i.e., if the applied input signal is within the *positive lower region* ($\underline{I}_{+}(X)$), the approximated output is given as + hypoglycemia whereas it approximates as − hypoglycemia if the applied input signal is within the *negative lower region* ($\underline{I}_{-}(X)$). For instance, For the input HR, if it is within the *positive lower boundary* ($\underline{I}_{HR+}$), it is approximated as + whereas it is defined as − if it is within the *negative lower boundary* ($\underline{I}_{HR-}$).

A similar way to input QTc, in which the boundary output is defined as + and - if they are under their respective *positive lower boundary* $\underline{I}_{QT+}$ and *negative lower boundary* $\underline{I}_{QT-}$. The rest of the elements which are neither in the *positive lower region* ($\underline{I}_{+}(X)$) nor *negative lower region* ($\underline{I}_{-}(X)$), are categorized under the boundary region (5.3) and a simultaneous classification task is performed by block-based neural network (BBNN).

Since there are four inputs (HR, QTc, ΔHR and ΔQTc) are used in this application, the structure of BBNN in Section 5.2.2 is organized with $m = 4$ rows while the number of layers (n_l) is determined through the training process. As discussed in Section 5.2.3,

the selection of number of layers (n_l) is carried out by a trial and error method that is the same chosen for the number of hidden nodes in conventional feedforward neural network. In a hypoglycemia detection system, the best classification performance is obtained when n_l is equal to 2 which has the best structure combination as shown in Fig. 5.7.

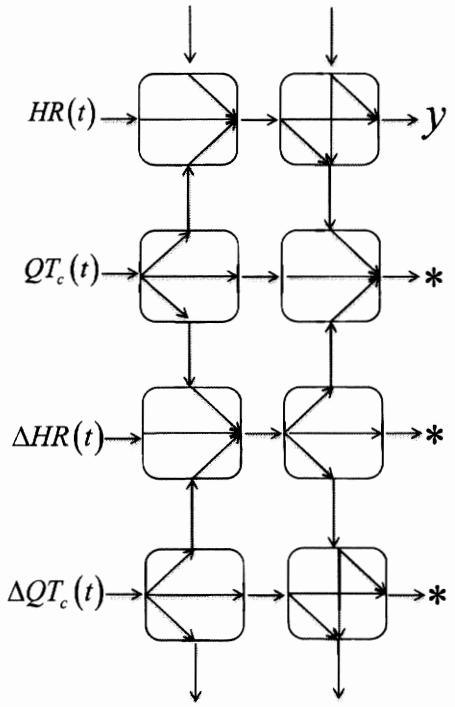

Figure 5.7: Best Rough-block-based neural network (R-BBNN) structure

Based on the obtained structure through the training process, the output of each block is calculated by (5.4) and (5.5). The overall output y in Fig.5.7 is defined as

positive when the status of hypoglycemia h is positive which expressed as:

$$h = \begin{cases} +1, & y \geq 0 \\ -1, & y < 0 \end{cases} .$$

Without the introduction of rough regions, only for the BBNN model, the number of layers (n_l) needs to be up to 6 layers for the best performance as shown in Fig. 5.8.

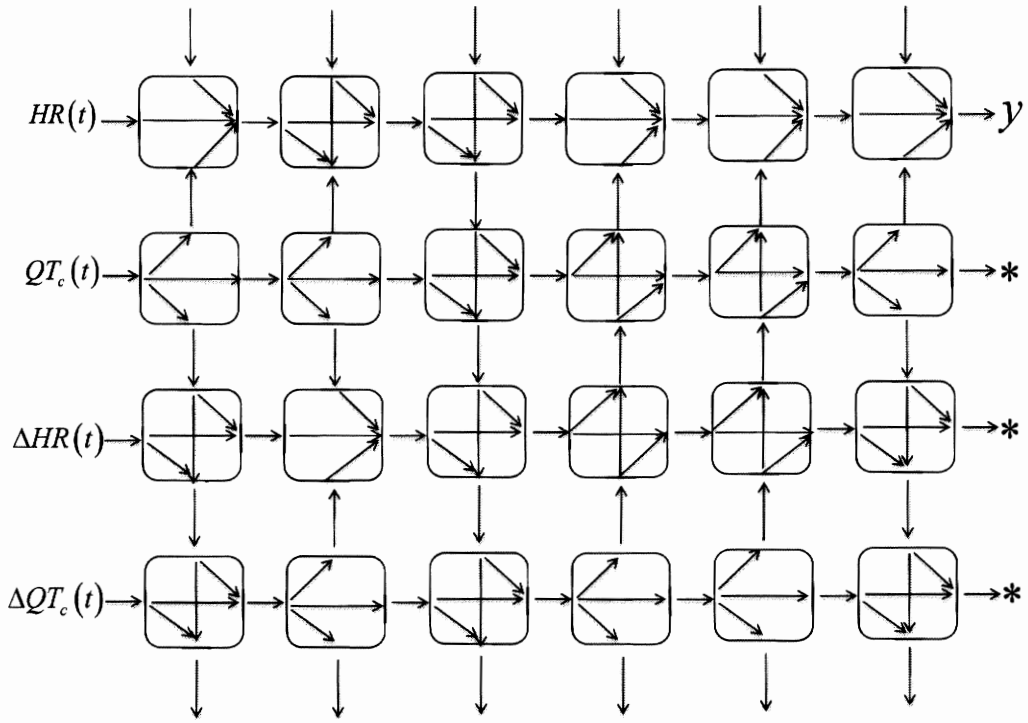

Figure 5.8: Best evolved block-based neural network structure (BBNN)

Compared with Fig. 5.7 and 5.8, it can be seen that the number of layers (n_l) is reduced to 2 layers. As a consequence of the reduced number of network layers, the number of network parameters is also decreased , i.e., less design parameters in R-BBNN4 (102 parameters) whereas 258 parameters in BBNN4. In addition, the total

number of training epoches required to obtain a stable system has been reduced to 200 using rough-block-based neural network. All the design parameters of rough-block-based neural network (R-BBNN) are encoded with particles of HPSOWM within the parameter boundaries,i.e., $s_l \in [0 \quad 1]$, $w_k^l \in [-7.55 \quad 7.55]$, $b_k^l \in [-7.55 \quad 7.55]$, $\alpha \in [1.55 \quad 10], \beta \in [0.55 \quad 2.5]$. The setting of HPSOWM parameters are the same that is chosen in subsection 3.4.5. In the following section, the best performance results of the proposed R-BBNN will be discussed.

5.4.1.1 Results Analysis

For comparison and analysis purposes,(1) a hybrid rough-block-based neural network (R-BBNN),(2) block-based neural network (BBNN) [Jiang2004],(3) a hybridization methodology involving rough set and conventional feedforward neural network, rough feedforward neural network (R-FFNN),(4) conventional wavelet neural network (WNN) in Appendix A.2 and (5) conventional feedforward neural network (FFNN) in Appendix A.1 with 4 inputs and 1 output are employed. All approaches are trained by HPSOWM.

The performance of the proposed rough set-based neural network model with four inputs (R-BBNN4), different types of neural network models, block-based neural network with four and two inputs (BBNN4) and (BBNN2), wavelet neural network with four inputs (WNN4) and feedforward neural network (FFNN4) are firstly evaluated in terms of mean values (averaging over 20 runs) as tabulated in Table 5.1.

		R-BBNN4	BBNN4	BBNN2	WNN4	FFNN4
Training	**Sen**(ξ)	82.43 %	86.45 %	84.85 %	84.12 %	83.64 %
	Spec(η)	40.03 %	40.73 %	40.35 %	40.63%	40.50 %
Validation	**Sen** (ξ)	81.25 %	92.50 %	87.34 %	80.44 %	79.07 %
	Spec(η)	41.23 %	40.29 %	40.13 %	40.94 %	41.38 %
Testing	**Sen**(ξ)	**82.09 %**	79.30 %	76.28%	71.39 %	68.84 %
	Spec(η)	**50.91 %**	50.00 %	50.40 %	44.37 %	48.34 %
	Gamma(γ)	***69.62***%	67.58 %	65.56 %	60.58 %	60.04 %

Table 5.1: Mean Value of Training Validation and Testing Results as η_l= 40%

As discussed earlier, the sensitivity of the detection system is more important than the specificity because it mainly represents the performance of the classifier. Based on the importance of sensitivity, the fitness function (3.31) was designed to find the optimized sensitivity while the specificity was kept at an acceptable value, i.e., sensitivity $\geq$ 60% and specificity $\geq$ 40% for this specific application.

For analysis, the initial specificity (η_l) in (3.31) was set at 0.6 ($1 - \eta_l$=60%) which is equivalent to maximum specificity, η_l=40%. At the defined initial condition, η_l=40%, the average (mean) value of training, validation and testing results were analyzed in terms of the sensitivity and specificity in Table 5.1.

From Table 5.1, it is found that the average (mean) testing result of proposed R-BBNN

is satisfactory by achieving testing sensitivity and specificity (82.09 and 50.91%). Compared with other approaches, BBNN4 (79.30 of sensitivity and 50.00% of specificity), BBNN2 (76.28 of sensitivity and 50.40% of specificity), WNN4 (71.39 of sensitivity and 44.37% of specificity) and FFNN4 (68.84 of sensitivity and 48.34% of specificity), the proposed hybrid system gives a better result in terms of sensitivity and specificity.

As introduced in the earlier subSections 3.4.5 and 4.4.1, the similar gamma analysis ($\gamma = \theta\xi + (1-\theta)\eta$, $\theta = 0.1$ to 1) is defined to evaluate the performance of the proposed R-BBNN4 detection system. The value of θ is set to 0.6 since the minimum requirement of the hypoglycemia detection system is 60 % of sensitivity and 40 % of specificity. In terms of γ analysis, the proposed R-BBNN is found to be satisfactory which gives γ value of 69.62 % while BBNN4, BBNN2, WNN4 and FFNN4 have γ value of 67.58 %, 65.93 %, 64.96% and 60.04 %.

In order to prove ΔHR and ΔQTc are useful inputs to enhance the sensitivity of the proposed detection system, the evolvable BBNN is tested with 4 inputs (BBNN4) and 2 inputs (BBNN2). In Table 5.1, the improvement can be satisfactorily found by achieving, 79.30% of sensitivity in BBNN4 compared with BBNN2 giving (76.28%) of sensitivity.

Since the prediction of the hypoglycemia status, $(+1/-1)$ in (5.8) is performed using $(+/-)$ lower regions, i.e., if the applied input signal is within the positive lower region,

the approximated output is given as + hypoglycemia. On the other hand, if the applied input signal is in the negative lower region, − hypoglycemia is approximated.

The optimized boundary parameters are obtained after the training process which can simplify network inputs and achieves better sensitivity as tabulated in Table 5.2. The proposed R-BBNN4 is efficient in obtaining better testing sensitivity, 83.72% and acceptable specificity, 51.91% which outperforms the other classifiers (BBNN4, BBNN2, MR-FIS4, WNN4, FIS4 and FFNN4) whose respective testing sensitivity and testing specificity are 79.09% and 51.82%, 76.74% and 52.73%, 75.00% and 50.00%, 74.42% and 48.18%,70.45% and 55.14% and 69.77% and 49.09%. The comparison studies are continuously carried out by means of γ analysis in which the proposed R-BBNN4 is found to be satisfactory with a γ value of 71.00%.

Methods	**Sen**(ξ)	**Spe**(η)	**Gamma**(γ)
R-BBNN4	***83.72*** %	***51.91*** %	***71.00***%
BBNN4	79.09 %	51.82 %	68.18 %
BBNN2	76.74 %	52.73 %	67.14 %
WNN4	74.42 %	48.18 %	63.92 %
FFNN4	69.77 %	49.09 %	61.50 %

Table 5.2: Best Testing Results as η_l= 40%

As can be seen in Figs. 5.7 and 5.8, the proposed R-BBNN needs only 2 network layers instead of using 6 layers in BBNN4. In addition, the 2 layers conventional

BBNN with no rough pre-processing can only give sensitivity, 60.47% and specificity, 42.73% whereas R-BBNN with 2 layers achieves 83.72% of sensitivity and 51.91% of specificity. It should be noted that the proposed methodology should not only be efficient in obtaining better sensitivity but also achieve faster learning rates with a reduced number of network parameters.

5.5 Conclusion

The innovative approach is based on the merit of rough set and block-based neural network which is integrated and introduced in this paper. Firstly, a rough approximation property is adopted by defining *lower* and *boundary regions* and implementing those defined regions within a neural network framework. In order to integrate the characteristics of application to the proposed methodology, block-based neural network is selected as a suitable classifier due to its ability in performing simultaneous optimization of both structures and weights. To enhance the performance of the proposed R-BBNN4 system, a hybrid particle swarm optimization with wavelet mutation (HPSOWM) operation is used to optimize the design parameters.

To illustrate the effectiveness of the proposed approach, R-BBNN4 is practically realized in detection of hypoglycemia episodes in patients with Type 1 diabetes mellitus (T1DM). The performance of the proposed hybrid system is compared with some

of the existing neural networks, BBNN4, BBNN2, WNN4, FFNN4. The comparison results show that the performance of proposed R-BBNN4 method is superior to other comparative classifiers by achieving higher sensitivity 83.72% and acceptable specificity 51.91%.

From results in Section 5.4.1, it is quite evident that hypoglycemia episodes in T1DM children can be efficiently detected non-invasively and continuously from the real-time physiological responses. The proposed architecture has been designed in such a way that that rough set method can greatly accelerate the network training time and improve its prediction accuracy.

It should be noted that the proposed methodology should not only be efficient in obtaining better sensitivity but also achieve a faster learning rate with reduce number of network parameters, i.e., only 82 parameters are needed for R-BBNN4 while 128 parameters are needed for BBNN4. It also revealed that the proposed approach improves classification performance and results in early convergence of the network.

Chapter 6

Summary and Discussion

For the non-invasive hypoglycemia detection system, this thesis is covered by three novel methodologies namely variable translation wavelet neural network (VTWNN), multiple regression based combinational neural logic network (MR-NLN) and rough-block-based neural network (R-BBNN) which are presented in Chapters 3, 4 and 5 respectively. For an optimal set of proposed network parameters, a global learning optimization algorithm called hybrid particle swarm optimization with wavelet mutation (HPSOWM) is used. It has been proven that, by exploiting the advantages of each algorithm, the performance of the non-invasive hypoglycemia monitoring system is improved. In this Chapter, the comparison studies between algorithms in terms of network operation principle, network complexity and learning ability will be given.

6.1 Summary on the Characteristics of Network Topologies

In this Section a general description regarding the characteristics of proposed network topologies is discussed. Each topology is best suited for use in specific application offering its own advantages and disadvantages.

The first topology is the VTWNN which uses multi-scaled wavelet function in the hidden layer. The network parameters, such as the translation parameters of the wavelets are variable depending on the network inputs. Because of variable translation parameters, the network becomes adaptive network, able to model the input-output function with input-dependent network parameters. Due to the adaptive nature, the network provides better performance and increased learning ability. It will be beneficial if the VTWNN handles different input data with different values of the network parameters.

The second topology is MR-NLN in which the binary logic gates (AND, OR and NOT) are cooperatively designed by the use of a truth table and K-map. Because the logic theory is used in the network design, the structure becomes systematic and simpler compared to other conventional neural network and enhances the training performance. Traditionally, conventional NN with the same structure are applied to handle different applications. The optimal performance may not always be guaranteed due to different characteristics of applications. In real world applications, the knowledge based NN that understands all the characteristics of practical application

is preferred for optimal performance. In conventional NN, the redundant connections and weights makes the number of network parameters unnecessarily large and downgrades the training performance. But for NLN, the structure becomes simpler as compared to conventional NN because some redundant connections are removed.

The third topology is R-BBNN which is a hybrid intelligent system designed by integrating rough set techniques in the architecture of the neural network. In this system, design improvements are made by the use of rough set properties in which partitioning of the applied input signal to a predictable part (certain) and a random (uncertain) part is carried out by using lower and boundary regions. Once rough set pre-processing is finished, a simultaneous classification task is performed by block-based neural network (BBNN) in order to deal only with the random (uncertain) part of the applied input signal. In BBNN model, the network structure is simply determined by signal flows between each blocks. Due to rough set properties and the adaptability of BBNN's flexible structures in dynamic environments, the classification performance is improved.

For hypoglycemia detection, the VTWNN in Chapter 3 gives better performance compared with conventional neural network structures such as wavelet neural network (WNN), feedforward neural network (FFNN) and radial basis function network (RBFNN). However, the main concern of the VTWNN is its computational demand which is reflected in the possibly larger number of network parameters. In Chapter 4, an integrated method, a neural logic network with multiple regression is applied

to the development of noninvasive hypoglycemia monitoring system. In general, conventional neural networks with the same structure were used to handle different applications. However, optimal performance was not always guaranteed due to different characteristics of applications.

The structural difference between VTWNN and MR-NLN is one of the important concerns. Though adaptive VTWNN is best suited for different input data with different values of the network parameters, their structure is less dedicated to the nature of application. Moreover, in fully-connected VTWNN, some links in the network could be practically redundant while the redundant connections are reduced in MR-NLN. In general, the redundant connections and weights of conventional NN makes the number of network parameters unnecessarily large and downgrades the training performance. Due to the reduction of those redundant connections, the NLN network structure becomes simpler.

With these above improved networks, VTWNN and MR-NLN, the classification performance was improved over conventional neural networks. However, both VTWNN and MR-NLN offers for less degree of freedom due to their network structure. It is because their network designs are not structurally evolving during the training process. Conceptually, the degree of freedom that allowed for the network is less adaptive with the characteristics of application. As a consequence, the model becomes less accurate.

In R-BBNN, BBNN classifier is selected as a suitable classifier due to its modular

characteristics and ability to evolve the size of the network by adding more basic blocks. With the flexible network structure, the degree of freedom is increased and the learning and generalization abilities are improved.

6.2 Comparison Studies on Experimental Results

Experimental results of the application on hypoglycemia detection system will be given in this section. Three proposed networks, VTWNN, MR-NLN and R-BBNN are compared and analyzed.

Methods		**Sensitivity** (ξ)	**Specificity** (η)	**Gamma** (γ)
R-BBNN	Training	82.43%	40.03%	65.47%
	Validation	81.25%	41.23%	65.24%
	Testing	**82.09%**	**50.91%**	**69.62%**
MR-NLN	Training	90.91%	40.47%	70.73%
	Validation	90.62%	40.29%	70.49%
	Testing	**78.81%**	**51.45%**	**67.86%**
VTWNN	Training	88.40%	40.80%	69.36%
	Validation	82.50%	41.25%	66.00%
	Testing	**77.41%**	**47.42%**	**65.41%**

Table 6.1: Mean value of Training, Validation and Testing Results: Set maximum specificity, $\eta_{\max}$= 40%

The best average (mean) training, validation and testing results are summarized in Table 6.1. Comparing the three proposed methods, the average (mean) testing sensitivity and specificity of R-BBNN (82.43 and 50.91%) is the best. In terms of Gamma (γ) analysis, the R-BBNN gives better γ value which is about 69.62% whereas VTWNN and MR-NLN gives 65.41% and 67.86% respectively.

In Table 6.1, it seems that the results for R-BBNN in training, validation and testing are similar while results for MR-NLN/VTWNN have quite a big variation i.e. the training and validation results are much better than testing. As the data partitioning in training, validation and testing for these three different neural networks are the same, the training and validation results should be similar. However, the training and validation results of R-BBNN is less than the others. It is because BBNN needs to handle only for the data that are out of the defined rough lower boundary regions. Since BBNN needs to handle less amount of data, less background information is given to BBNN. It may lead to less sensitivity and specificity of training and validation. It is confirmed from Table 5.1 that sensitivity and specificity for the training and validation of BBNN (without rough set pre-processing) are about (86.45 and 40.73 %) and (92.50 and 40.29 %) which are similar as obtained in MR-NLN and VTWNN.

The best testing results of the hypoglycemia detection system by all the proposed networks are tabulated in Table 6.2. In this table, the best testing results are compared and analyzed in terms of sensitivities, specificities and gamma values. The total number of training epoches required to obtain a stable system are also presented. In

comparison studies between three proposed methods, the performance of R-BBNN is excellent by achieving testing sensitivity, specificity and gamma at 83.72, 51.91 and 71.00%. It can be seen that the number of design parameters required for each network look similar (68, 66 and 52 parameters), but the total number of training epoches required to obtain a stable system has been reduced to 200 in R-BBNN while MR-NLN and VTWNN needs 1000 and 1500.

Methods	**Sens**(ξ)	**Spec**(η)	**Gamma**(γ)	**Iteration Number**
R-BBNN	83.72%	51.91%	71.00%	200
MR-NLN	79.07%	53.64%	68.90%	1000
VTWNN	81.40%	50.91%	69.20%	1500

Table 6.2: Best Testing Results: Set maximum specificity, $\eta_{\max}$= 40%

As can be seen in Table 6.2, the proposed R-BBNN and MR-NLN obtain better sensitivity and specificity with faster time i.e. iteration about 200 and 1000 due to less complexities and more flexibilities of network structure. Because the structure of R-BBNN is simplest among three methods, it achieves better sensitivity and specificity with faster computation time, number of iteration is about 200. Generally speaking, the network may take longer time for stable and optimal results as the network structural complementarities increases. In this application, the optimal iterations number for each proposed methods are set by trial and error method.

With a first look at VTWNN and MR-NLN in Table 6.2, it seems that the performance

of VTWNN is better than MR-NLN with best testing sensitivity of 81.40%. However, in terms of solution stability, the MR-NLN performs better by giving mean testing sensitivity and specificity at 78.81 and 51.45%. Overall, among the three proposed methodologies, the R-BBNN is efficient in obtaining better sensitivity with faster learning rates.

6.3 Summary

A summary of comparisons among three proposed methodologies in terms of topology complexities, learning ability and overall performance rating is given in Table 6.3. From Table 6.3, it can be seen that the complexities of network topologies are generally less in R-BBNN and MR-NLN due to the introduction of rough set and binary logic properties. Because of the evolvable BBNN network structure, the R-BBNN offers a greater degree of freedom compared with MR-NLN and VTWNN. For MR-NLN and VTWNN, the design of networks are not structurally evolving during the training process. As a consequence, the degree of freedom is less and these models become less accurate.

Compared with other conventional neural networks such as feedforwrad neural network (FFNN) and wavelet neural network (WNN) in Appendix A.2 in Appendix A.1, the proposed R-BBNN and MR-NLN are less complexities due to introductory of rough set and binary logic properties. By using those properties, the R-BBNN and

MR-NLN only need to handle only small amount of input data, where as conventional NNs in Appendix A.2 and A.1 are handling all the input data. But for VTWNN, the network complexities is similar to FFNN and WWN since no other properties such as rough set pre-processing and binary logic properties are incorporated in network structure. However, VTNN has advantages in handling the distributed data while FFNN and VTWNN may not be good enough.

The proposed NNs are developed in the Matlab (R2010b) platform using a Dell Optiplex 780 PC with an Intel Core Duo (E8400) Processor and 4 GB DDR3 SD RAM. It takes about 1 hour of computational time until the termination condition is met after 1500 iterations. Based on the knowledge and experience, the program developed in Matlab takes more computation time than the C language. It should be aware that the difference in computation time may come from different termination conditions, software and hardware implementation. In terms of computation time, the proposed R-BBNN and MR-NLN take less time for training which is about 30 minutes and 15 minutes, whereas VTWNN and other conventional NNs take longer than 45 minutes.

In addition, for R-BBNN and MR-NLN, their design is more dedicated to the application while VTWNN has less understanding on the knowledge of application. Generally, conventional neural networks with the same structure were used to handle different applications. However, optimal performance may not be always guaranteed due to different characteristics of applications. Thus, in real world applications, the

knowledge based neural network such as R-BBNN and MR-NLN that understands all the characteristics of practical application is preferred for optimal performance.

The learning performance of all networks, R-BBNN, MR-NLN and VTWNN are found to be satisfactory. For this specific application in hypoglycemia detection, the performance of R-BBNN offers about 20 % improvement over the MR-NLN and VTWNN. To conclude, the overall performance of proposed three methodologies such as R-BBNN, MR-NLN and VTWNN are better than other conventional classifiers, WNN, FFNN and MR in terms of sensitivity and specificity .

	R-BBNN	**MR-NLN**	**VTWNN**
Topology Complexities	Simple	Simple	Complex
Network Flexibility (Degree of freedom)	More	Less	Less
Dedicated Application	More	More	Less
Learning Algorithm	HPSOWM	HPSOWM	HPSOWM
Learning Ability	Satisfactory	Satisfactory	Satisfactory
Overall Performance	Best	Good	Acceptable

Table 6.3: Summary of the R-BBNN, MR-NLN and VTWNN

Chapter 7

Conclusions and Further Research

7.1 Conclusions

In this thesis, novel methodologies for a non-invasive hypoglycemia detection system have been developed by analyzing the behavioral changes of physiological parameters, HR and QTc. These methodologies are comprised of three different classification techniques, namely variable translation wavelet neural network (VTWNN), multiple regression based combinational neural logic network (MR-NLN) and rough-block-based neural network (R-BBNN). Thanks to these advanced neural network classifiers, the learning, generalization abilities and classification accuracies are improved. The detail of three proposed methodologies have been discussed in Chapter 3, 4 and 5 respectively.

The variable translation wavelet neural network (VTWNN) algorithm in Chapter 3 was the first of the three proposed algorithms to be used in the hypoglycemia detection system. In VTWNN, every hidden-layer activation function is characterized by multiscaled wavelet function in which the translation parameter is variable depending on the network inputs. One distinct advantage of VTWNN is that the input data are manipulating in the wavelet activation function and the output is estimated with input-dependent hidden layer activation function. Hence, the output of the VTWNN changes smoothly with respect to the nature of input data. By using VTWNN, some of the cases that cannot be handled by conventional neural networks with a limited number of parameters can be easily solved.

As the second algorithm, a hybrid technique based on neural logic network combined with multiple regression model (MR-NLN) is proposed in Chapter 4. In MR-NLN, its network structure is systematically designed based on the characteristics of application. The network structure is targeted to dedicated application. The network is designed by the use of binary logic gates (AND, OR and NOT) in which a truth table and K-map are constructed depending on the knowledge of application. Because logic theory is used in the network design, the structure becomes systematic and simpler compared to other conventional neural network and enhance the training performance. Compared to VTWNN, the number of network parameters in MR-NLN is smaller due to its simple structure. An important concern of the VTWNN is its computational demand reflected on the possibly larger number of network parameters.

The redundant connections and weights of VTWNN makes the number of parameters unnecessarily large and may downgrade the training performance.

The third algorithm, rough approximation block-based neural network (R-BBNN) focuses on the hybridization technology using rough sets concepts and neural computing for decision and classification. Based on the rough set properties, the input signal is partitioned to a predictable (certain) part and random (uncertain) part. In this way, the neural network is designed to deal only with the boundary region which mainly consists of a random part of applied input signal causing inaccurate modeling of the data set. Based on the knowledge of application, block-based neural network (BBNN) is selected as a suitable classifier due to its ability in evolving internal network structures and adaptability in dynamic environments. The performance results obtained by the R-BBNN aid in providing best classification accuracy.

With improved networks, VTWNN and MR-NLN, the classification performance was improved over conventional neural networks. However, both VTWNN and MR-NLN offers less degrees of freedom due to their network structure. In VTWNN and MR-NLN, the network designs are not structurally evolving during the training process ,i.e., the degree of freedom allowed for the network is less adaptive with the characteristics of application. Conceptually, fewer degrees of freedom make the model less accurate. In R-BBNN with flexible network structure, the degree of freedom is increased and the learning and generalization abilities are improved.

To formally analyze the natural occurrence of hypoglycemia episodes based on the pre-defined BGL, three advanced neural network classifiers were used. A general comparison between the different proposed topologies are discussed in Chapter 6. In short, the performance of R-BBNN is the best among the neural network proposed in this thesis. It can be concluded that the performance results obtained by the three proposed algorithms (VTWNN, MR-NLN and R-BBNN) aid in improving the sensitivity and specificity of hypoglycemia detection system compared with conventional algorithms (FFNN, WNN and RBFNN).

7.2 Future Works

Some possible future research directions are given in this section. In terms of the algorithms that are employed for hypoglycemia detection, more systematic neural network architectures could be further designed and developed. The conventional networks with fixed architecture may not be good enough for every task. Those networks usually start with a defined architecture and the final parameters and/or structure under this architecture are determined through the learning process.

Besides, the degree of freedom allowed for the neural networks should be more adaptive with the characteristics of application. More degrees of freedom make the networks more adaptable with the nature of application and provide better accuracy. For the networks with flexible structure, the degrees of freedom are increased and

generalization abilities are improved. Thus, the overall network architecture should be properly and systematically designed.

Concerning with ECG parameters, the heart rate (HR) and corrected QT interval (QTc) are currently used as the inputs for hypoglycemia detection. From clinical and experimental studies, as shown in Section 2.2.1, there is a marked change in ECG morphology, especially in QT interval and/or QRS complex during hypoglycemia. There should be other forms of ECG parameters that are significantly changing under hypoglycemic conditions. A future work might need to investigate more ECG parameters for hypoglycemia detection.

In terms of proposed algorithms, in order to validate the effectiveness, other clinical studies such as breast cancer detection [Singh2011], Parkinson's disease detection [Handojoseno2012] and other applications such as short-term traffic flow forecasting [Chan2012] should be further examined.

Appendix A

Artificial Neural Networks

Artificial neural networks (ANNs) are computational models inspired by the functioning of the human brain. They consist of simple but highly interconnected computing devices, each of which imitates the biological neurons. Research in the field of artificial neural networks has attracted increasing attention in recent years. After the first model of artificial neurons [Ash1994] had been proposed, new and more sophisticated neural network models have been made from decade to decade. A very important feature of artificial neural networks is their adaptive nature which makes such computational models very appealing in applications, especially where there is little or incomplete understanding of the problem to be solved but where training data is readily available.

Artificial neural networks have been widely used in pattern recognition and machine

learning. They possess strong adaptability and learning ability in addition to precision in data classification and prediction which is extremely high. Unlike classical statistical methods, the neural network can be applied to data analysis of small samples, without satisfying normal distribution. Because of the nonlinear transformation involved in the hidden layer, neural networks can work well with complex data. It offers the following advantages compared with conventional processing techniques:

- It has a strong self-adoption learning ability
- It can quickly recognize the abilities of parallel-distributed information storage and processing
- There exists strong error-tolerance, and it is able to recognize input pattern with noise

In all neural network models, a single neuron is the basic building block of the network. The operation of a single neuron is modeled by mathematical equations and the individual neurons are connected together as a network. Each neural network has its learning laws according to which it is capable of adjusting parameters of the neurons [Ham2001]. In most neural networks model, the operation of a single neuron can be divided into two separate parts; weighted sum and an output function as shown in Fig. A.0.1.

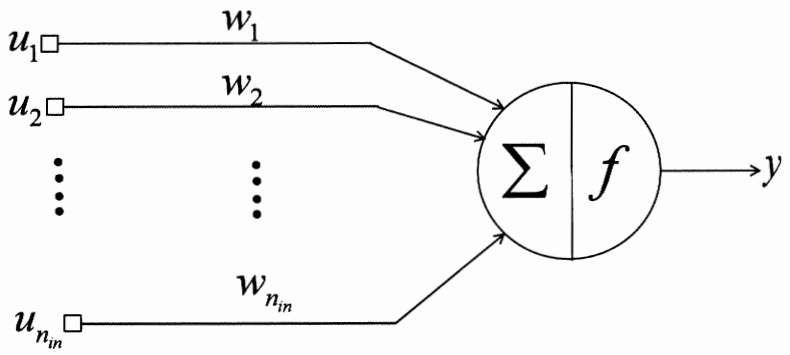

Figure A.0.1: Model of artificial neuron

For each input u_i, $i = 1, 2, \ldots, n$ has its associated weights w_j, $j = 1, 2, \ldots, m$ which can be modified to model synaptic learning. The node computes the activation function $f(\cdot)$ of weighted sum of inputs in the following form:

$$y = f\left(\sum_{i=1}^{n} w_i x_i\right) \tag{A.0.1}$$

A.1 Feed Forward Neural Network (FFNN)

A multi layer feedforward neural network becomes the most famous due to its good approximation in any smooth and continuous nonlinear separation function in a compact domain to arbitrary accuracy [Windrow1990].

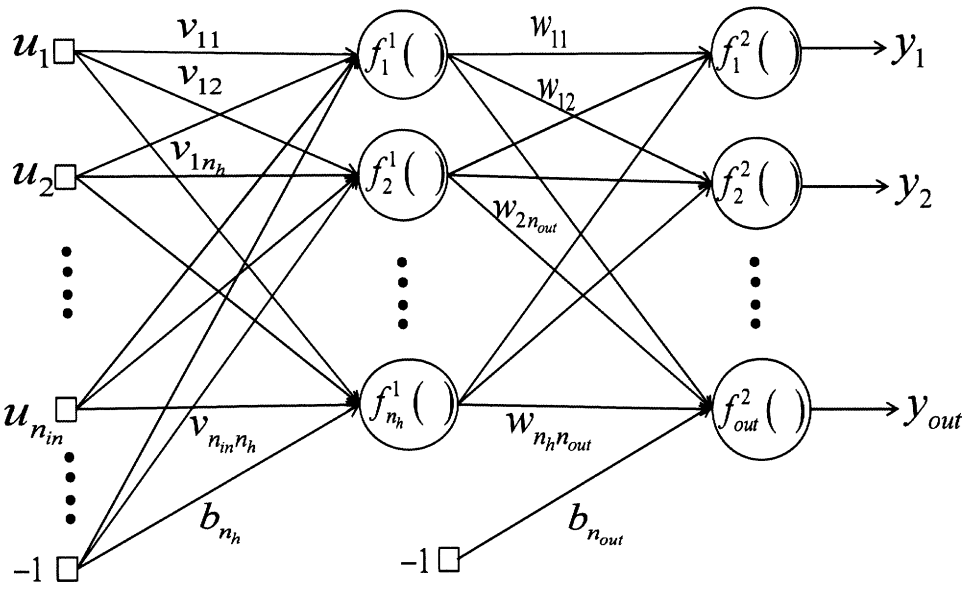

Figure A.1.1: Structure of three layer feedforward neural network (FFNN)

For each input has its associated weights and it can be modified to model synaptic learning. The input-output relationship of a fully connected feedforward neural network in Fig. A.1.1 is calculated as follows:

$$y_l = f_l^2\left(\sum_{j=1}^{n_h} w_{jl} f_j^{n_{in}}\left(\sum_{i=1}^{n_{in}} w_{ij}u_i - b_j\right) - b_l\right) \quad l = 1, 2, \ldots, n_{out} \tag{A.1.1}$$

where $u_i,\ i = 1, 2, \ldots, n_{in}$ are the input variables, n_{in} is the number of inputs while n_h denotes the number of hidden notes. $v_{ij},\ j = 1, 2, \ldots, n_h$ is defined as the weight which is linked between the ith input and jth hidden nodes, w_{jl} is the weight of link between the jth hidden and lth output nodes, b_j and b_l are the biases for hidden and output nodes. The total number of tuned parameters are $(n_{in} + 1) \times n_h + (n_h + 1) \times n_{out}$.

The transfer functions $f_j^1(\cdot)$ and $f_l^2(\cdot)$ are used as the activation function in the

hidden nodes and output nodes. The following transfer functions, logarithmic sigmoid transfer function (A.1.2), hyperbolic tangent sigmoid transfer function (A.1.3) and linear transfer function (pureline) (A.1.4) are commonly used in the neurons of hidden and output layers [Looney1997]:

$$logsig\,(\theta) = \frac{1}{1+e^{-\theta}} \in [0 \quad 1],\ \theta \in \Re \tag{A.1.2}$$

$$tansig\,(\theta) = \frac{2}{1+e^{-2\theta}} - 1 \in [0 \quad 1],\ \theta \in \Re \tag{A.1.3}$$

$$pureline\,(\theta) = \theta,\ \theta \in \Re \tag{A.1.4}$$

A.2 Wavelet Neural Network (WNN)

The wavelet neural network is considered as a particular case of feedforward neural network and the neural network using wavelet basis function can provide faster convergence rates for approximation compared with conventional feedforward neural network. In addition, the wavelet has been applied in many research areas because of its excellent property in time-frequency localization of a given signal [Zekri2008] [Mallat1989].

By combining wavelet with neural network, wavelet neural network (WNN) has been developed in order to give better performance in function approximation and learning

capabilities [Billings2005]. It can also be considered as a particular case of feedforward neural network (FFNN) apart from using multi-scaled wavelet activation function, $\psi_{a,b}(x)$ in the hidden layer.

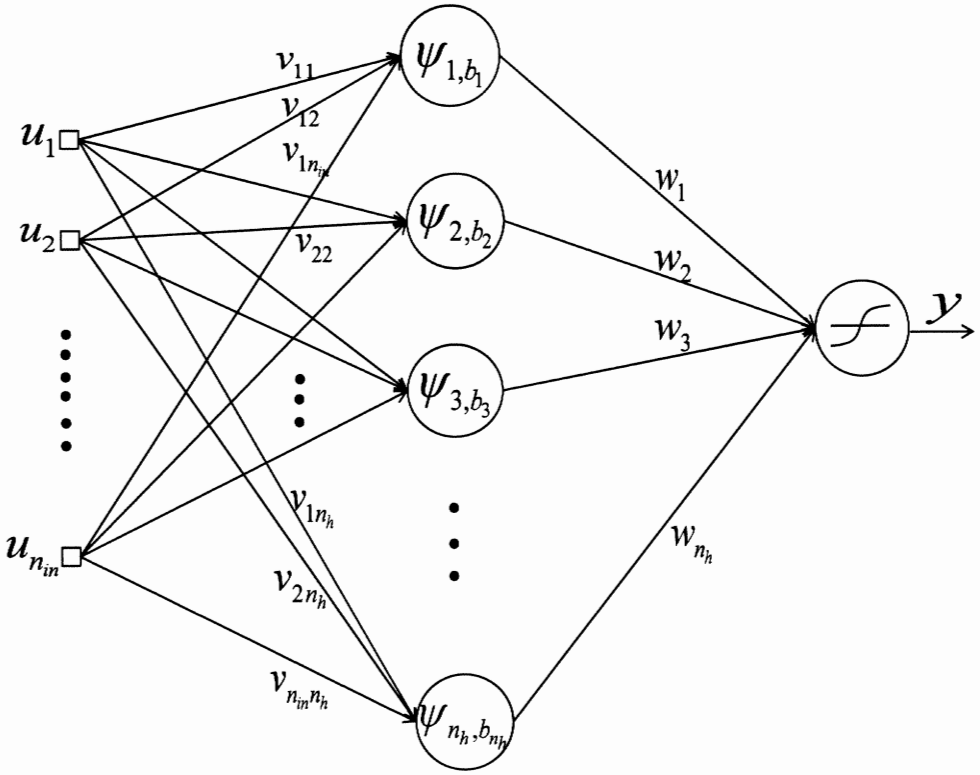

Figure A.2.1: Structure of wavelet neural network (WNN)

By the use of wavelet as activation functions of the hidden layer, the dilation parameters (a) and translation parameter (b) of the wavelet is variable and definable with any real positive number. The input-output relationship of the WNN is given by:

$$y_l = \sum_{j=1}^{n_h} \psi_{j,b_j} \left(\sum_{i=1}^{n_{in}} u_i v_{ij} \right) w_{jl}, \quad l = 1, 2, \ldots, n_{out} \tag{A.2.1}$$

where u_i, $i = 1, 2, \ldots, n_{in}$ are the input variables, n_{in} is the number of inputs; n_h denotes the number of hidden nodes; v_{ij} is the weight of the link between ith input

and jth hidden node; j denotes the dilation parameter and b_j is the translation parameter of the multiscale wavelet function; $\psi(\cdot)$ is mother wavelet. The total number of parameters for WNN is $n_h \times (n_{in} + n_{out} + 1)$.

A.3 Radial Basis Function Network (RBFNN)

RBFNN has been widely used in pattern recognition tasks [Korrek2010] due to its fast learning and good approximation ability. However, the standard RBFNN has local respective regions property, the approximation effect is poor when it is applied for function estimation outside of the training data. In order to cope with this problem, a normalized RBFNN is developed as shown in Fig. A.3.1 for modeling and design of a non-invasive hypoglycemia monitor with physiological responses.

In general, the structure of normalized RBFNN (NRBFNN) is the same for RBFNN apart from normalization of radial basis function in the hidden layer. Due to the extrapolation property of normalized radial basis function neural network (NRBFNN), a smaller number of hidden neurons is needed to capture the entire input space. Thus, compared with standard RBFNN, normalized RBFNN provides better generalization ability.

As can be seen in Fig. A.3.1, each neuron in the hidden layer has n Gaussian membership function in the following form:

$$g_{ij} = \exp\left(-\frac{(u_i - c_{ij})^2}{\sigma_{ij}^2}\right), \quad j = 1, 2, \ldots, m \tag{A.3.1}$$

where u_i, $i = 1, 2, \ldots, n_{in}$ are the input variables, n_{in} is the number of inputs; n_h is defined as the number of hidden neurons; g_{ij} is the ith membership function in the jth neurons, c_{ij} and σ_{ij}^2 are the center and the width of the membership function respectively.

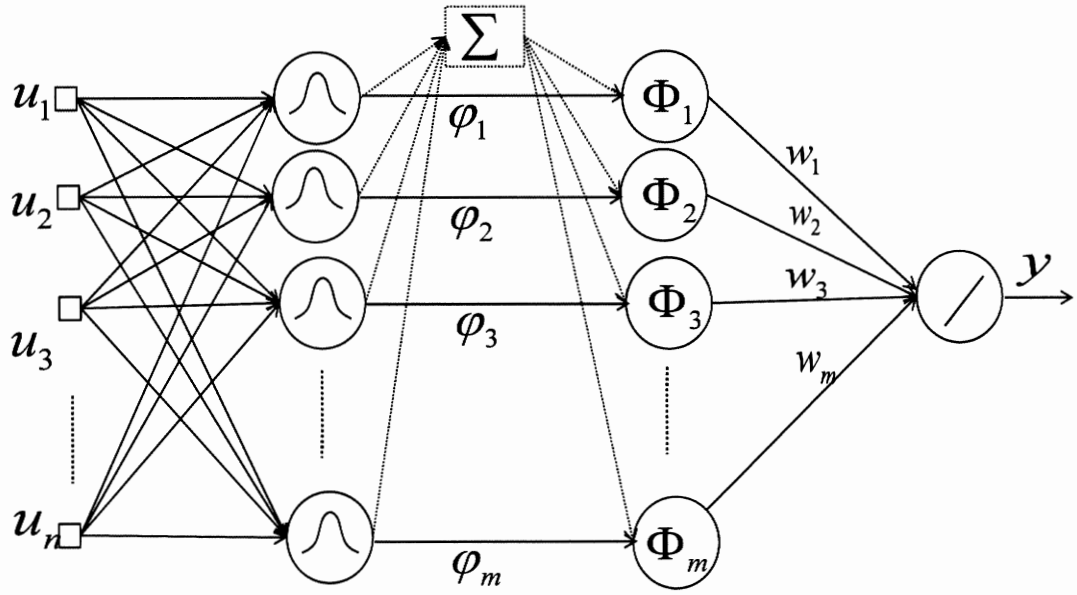

Figure A.3.1: Structure of radial basic function neural network (RBFNN)

The output of jth hidden neuron is computed from the multiplication of the inner n membership functions which denotes the firing strength of jth hidden neuron and

expressed as follows:

$$\varphi_j(u) = \prod_{i=1}^{n} g_{ij}(u_i) = \exp\left(-\sum_{i=1}^{n} \frac{(u_i - c_{ij})^2}{\sigma_{ij}^2}\right) \quad . \tag{A.3.2}$$

Although normalization operation increases the computing complexity, it can give better interposition performance. The normalized output $\Phi_j(u)$ is calculated as:

$$\Phi_j(u) = \frac{\varphi_j}{\sum_{j=1}^{n_h} \varphi_j} = \frac{\exp\left(-\sum_{i=1}^{n} \frac{(u_{it}-c_{ij})^2}{\sigma_{ij}^2}\right)}{\sum_{j=1}^{n_h} \exp\left(-\sum_{i=1}^{n} \frac{(u_{it}-c_{ij})^2}{\sigma_{ij}^2}\right)} \quad . \tag{A.3.3}$$

If the basis functions of the RBF network are Gaussian functions, the normalized output is calculated by:

$$y_l = \sum_{j=1}^{n_h} w_j \Phi_j(u) \ , \tag{A.3.4}$$

where w_{jl} is the connection weight between the jth normalized neuron and $l$the output neuron in the output layer in which $j = 1, 2, \ldots, n_h$ and $l = 1, 2, \ldots, n_{out}$. The total number of tuned parameters of RBFNN is $n_h(n_{in} + n_{out} + 1)$.

A.4 Local Learning Algorithms

Learning or training is one of the most important issue in neural networks since the optimal set of network parameters is obtained through the learning process. It is a process by which the parameters of a neural network are adapted through a process of stimulation by the environment in which the network is embedded. The network learning is affected by two factors: the network structure and network parameters

(weights). The learning algorithm provides a rule to optimize the weight values while the structure affects the non-linearity of the network function. In the following sections, the local and global learning algorithms are discussed further for the training of network structure and parameters.

Neural networks have different types and every type has its own learning rule. Each method of learning mainly changes the network parameters according to its learning rule to accommodate the network characteristics to its desired pattern. In order to train network parameters, some local learning algorithms such as hebbin learning rule in Appendix. A.4.1, back-propagation learning rule in Appendix A.4.2 are generally used.

A.4.1 Hebbian Learning Rule

The Hebbian Learning Rule is a learning rule that specifies how much the weight of the connection between two units should be increased or decreased in proportion to the product of their activation. The rule builds on Hebbs's learning rule [Hebb1949] which states that the connections between two neurons might be strengthened if the neurons fire simultaneously. The Hebbian Rule works well as long as all the input patterns are orthogonal or uncorrelated. A basic Hebbian adaption (Hebbian learning) has been mathematically described as follows:

$$w_{ji}^{new} = w_{ji}^{old} + \eta x_i y_i \tag{A.4.1}$$

where $i = 1, 2, \ldots, n$, $j = 1, 2, \ldots, p$, η is a constant that represents an adaption rate, x_i is the value of ith activation function in the input layer and y_j is the value of the jth activation function and w_{ji} is the connection weights between two activation functions.

A special case of Hebbian adaption is the delta rule, also sometimes called the Widrow-Hoff rule in which the amount of weight adjustment is proportional to the delta between the target activation value b_{kj} and the actual activation value y_{kj} which is described as:

$$w_{ji}^{new} = w_{ji}^{old} + \eta \delta_{kj} a_{ki} \tag{A.4.2}$$

where $\delta_{kj} = b_{kj} - y_{kj}$ and η is adaption coefficient which typically takes on values between 0 to 1. The detail implementation of delta rule is discussed in [Eberhart2007].

A.4.2 Back-propagation (BP) Learning Rule

Back-propagation technique [Manic2002] [Zurada1992] [Ham2001] is one of the most popular training algorithms for neural networks. Basically, back-propagation is a gradient descent technique to minimize the error which is defined by the output error across all the output activation functions and all input patterns [Rumelhart1986]. In this algorithm, the weight adjustment is performed based on the gradient information of error function (A.4.3) and updates the weights by moving them along the gradient

descendent direction and reduce the mean square error over all input patterns.

$$\Delta w_{ij}(n) = -\varepsilon \frac{\partial E}{\partial w_{ij}} + \alpha \Delta w_{ij}(n-1) \tag{A.4.3}$$

where ε and α are two non negative constant parameters called learning rate and momentum. With this learning algorithm, it is obviously known that a kind of continuous activation functions; *sigmoid*, *hyperbolic* and *tangent* need to be used instead of using step function in the hidden layer of NN. Since this method requires computation of the gradient of the error function at each iteration step, the continuity and differentiability of the error function are important.

In order to improve the learning rate, different back propagation algorithms such as the back propagation algorithms with momentum [Haykin1999], and conjugate gradient algorithm [Moller1993] have been proposed in which the derivative information of error function is evaluated with regard to weights and biases in the network. However, the drawbacks such as the requirement of the derivative information, trapping in local optimum still come with this algorithm.

A.5 Global Learning Algorithm

To overcome the local convergence problem in solving complex nonlinear optimization problems, evolutionary computation (EC) [Yu2010] has been reported for obtaining globally optimal solutions. These EC methods rely much on expert knowledge or a

trial-and-error approach to determine the choice of the EC algorithm, its parameter setting and the criterion for terminating the evolution. In this section, famous EC algorithms such as genetic algorithm (GA) and particle swarm optimization are discussed detail in Appendix A.5.1 and A.5.3

A.5.1 Genetic Algorithm (GA)

Genetic algorithm (GA) is one of the evolutionary computation techniques that is especially useful for complex optimization problems with a larger number of parameters [Davis1991] [Ling2010b]. Unlike BP algorithm in Section. A.4.2, the GA searches for global solutions and does not require the objective function to be differentiable. Since error surfaces for even simple problems can be quite complex with many local optima, the GA seems to be better suited for this type of search. This algorithm is more suitable in a large, complex domain. They are also good to train different types of neural network regardless if they are feedforward neural network or any other structure of neural networks. This generally saves a lot of human effort in developing training algorithms for different types of networks.

The GA searches from one population of points to another focusing on the area of the best solution while continuously sampling the total parameter space. In Algorithm. A.5.1, the standard pseudo code for the process of GA is given. In Algorithm. A.5.1, P is defined as $P = [p_1 \quad p_2 \quad \dots \quad p_{pop\_size}]$ in which pop_size is the number of chromosomes in the population. Each chromosomes p_i contains some genes p_{i_j}. The

population evolves from generation t to $t+1$ by repeating the procedures: selection, crossover and mutation. The selection operation is used to select the chromosomes from the population with respect to some probability distribution based on fitness values. The crossover mutation is used to combine the information of the selected chromosomes (parents) and generates the offspring. The mutation operation is used to change the offspring genes. The Algorithm. A.5.1 stop either a ceratin condition is met or predefined number of iteration have been reached.

Algorithm A.5.1: PSEUDO CODE FOR GENETIC ALGORITHM (GA)

$t \leftarrow 0$

Initialize $P(t)$

output $(f(P(t)))$

while <not termination condition>

do

- $t \leftarrow t+1$
- Select 2 parents $p1$ and $p2$ from $p(t-1)$ using the selection operation
- Perform the cross over operation
- Perform the mutation operation
- **output** $(P(t))$
- **output** $(f(P(t)))$

A.5.2 Particle Swarm Optimization (PSO)

Particle swarm optimization is another instance of the EC techniques that tackle complex optimization problems. It is a population-based stochastic optimization algorithm which is inspired by the social behaviors of animals like fish schooling and bird flocking [Kennedy1995]. It considers a number of particles that constitute a swarm. Each particle transverses the search space looking for global optimum. The construction of standard PSO with construction and inertia weight factors is shown in Algorithm. A.5.2.

From Algorithm A.5.2, $X(t)$ is denoted as a swarm at the t-th iteration. Each particle $\mathbf{x}^p(t) \in X(t)$ contains κ elements $x_j^p(t)$ at the t-th iteration, where $p = 1, 2, ..., \theta$ and $j = 1, 2, ..., \kappa$; θ denotes the number of particles in the swarm and κ is the dimension of a particle. First, the particles of the swarm are initialized and then evaluated by a defined fitness function. The objective of HPSOWM is to minimize the fitness function (cost function) $f(X(t))$ of particles iteratively. The swarm evolves from iteration t to t+1 by repeating the procedures as shown in Algorithm A.5.2. The operations are discussed as follows.

The velocity $v_j^p(t)$(corresponding to the flight speed in a search space) and the position $x_j^p(t)$ of the j-th element of the p-th particle at the t-th generation can be calculated using the following formulae:

$$v_j^p(t) = k \cdot \{\{w \cdot v_j^p(t-1)\} + \{\varphi_1 \cdot r_1(\tilde{x}_j^p - x_j^p(t-1))\} + \{\varphi_2 \cdot r_2(\hat{x}_j - x_j^p(t-1))\}\} \quad \text{(A.5.1)}$$

and

$$x_j^p(t) = x_j^p(t-1) + v_j^p(t) \tag{A.5.2}$$

where $\tilde{x}^p = [\tilde{x}_1^p, \tilde{x}_2^p, \ldots, \tilde{x}_k^p]$ and $\hat{\mathbf{x}} = [\ \hat{x_1}\ \hat{x_2}\ , ...\hat{x_\kappa}\ ]$, $j = 1, 2, ..., \kappa$. The best previous position of a particle is recorded and represented as $\tilde{x}$; the position of best particle among all the particles is represented as $\hat{x}$; w is an inertia weight factor; r_1 and r_2 are acceleration constants which return a uniformly random number in the range of [0,1]; k is a constriction factor derived from the stability analysis of (A.5.2) to ensure the system to be converged but not prematurely [Eberhart2000]. Mathematically k is a function of φ_1 and φ_2 as reflected in the following equation:

$$k = \left(\frac{2}{|2 - \varphi - \sqrt{\varphi^2 - 4\varphi}|} \right) \tag{A.5.3}$$

where $\varphi = \varphi_1 + \varphi_2$ and $\varphi > 4$.

PSO utilizes $\tilde{x}$ and $\hat{x}$ to modify the current search point to avoid the particles moving in the same direction, but to converge gradually toward $\tilde{x}$ and $\hat{x}$. A suitable selection of the inertia weight w provides a balance between the global and local explorations. Generally, w can be dynamically set with the following equation [Eberhart2000]:

$$w = w_{\max} - \left(\frac{w_{\max} - w_{\min}}{T} \right) \times T \tag{A.5.4}$$

where t is the current iteration number, T is the total number of iteration, $w_{\max}$ and $w_{\min}$ are the upper and lower limits of the inertia weight, $w_{\max}$ and $w_{\min}$ are set to 1.2 and 0.1 respectively.

Algorithm A.5.2: PSEUDO CODE FOR STANDARD PSO ($SPSO$)

$t \leftarrow 0$
Initialize $X(t)$
output $(f(X(t)))$
while <not termination condition>
do
- $t \leftarrow t + 1$
- Update $v(t)$ and $x(t)$ based on $(A.5.1)$ to $(A.5.2)$
- **if** $v(t) > v_{\max}$
  - **then** $v(t) = v_{\max}$
- **if** $v(t) < v_{\min}$
  - **then** $v(t) = v_{\min}$
- **output** $(X(t))$
- **output** $(f(X(t)))$

return $(\hat{\mathbf{x}})$
comment: return the best solution

In (A.5.1),

the particle velocity is limited by a maximum value $v_{\max}$. The parameter $v_{\max}$ determines the resolution with which regions are to be searched between the present position and the target position. This limit enhances the local exploration of the

problem space and it realistically simulates the incremental changes of human learning. If $v_{\max}$ is too high, particles might fly past good solutions. If $v_{\max}$ is too small, particles may not explore sufficiently beyond local solutions. From experience, $v_{\max}$ is often set at 10%-20% of the dynamic range of the element on each dimension. Next, the mutation operation is used to mutate the element of particles.

Comparing with other population-based stochastic optimization methods such as evolutionary algorithms, the PSO has comparable or even superior search performance for many hard optimization problems with faster and more stable convergence rates. However, observations revealed that the PSO sharply converges in the early stages of the searching process, but usually presents the problems on reaching the near optimal solution. It behaves like the traditional local searching methods that trap in the local optima. As a result, it is hard to obtain any significant improvements by examining neighboring solutions in the later stages of the search.

Recently, different types of hybrid PSOs have been proposed to overcome the drawback of trapping in the local optima. The hybrid PSO was firstly proposed, in which a standard selection mechanism of PSO is integrated with the gradient descent information, namely hybrid gradient descent PSO (HGPSO) [Angeline1998]. Though it is aimed to achieve faster convergence without getting trapped in the local minima, the computational demand of the HGPSO is increased by the process of the gradient descent [Noel2004].

A.5.3 Hybrid PSO with GA mutation (HGAPSO)

Another hybrid PSO algorithm named HGAPSO which incorporates a genetic algorithm evolutionary operations of crossover, mutations and reproduction to hybrid PSO is proposed [Juang2004]. In the HGAPSO, the solution space can be explored by performing mutation operations on particles along the search and premature convergence of standard PSO in Algorithm A.5.2 is more likely to be avoided. In HGAPSO, the GA mutation operation in (A.5.5) which starts with a randomly chosen particle in the swarm is introduced. The pseudo code of the hybrid PSO with the mutation operation is given in Algorithm. A.5.3 in which the mutation operation is performed after updating their velocities and positions. To perform mutation operation Algorithm. A.5.3, the following mutation operation is used:

$$\begin{aligned} mut\,(x_j) &= x_j - \omega, \quad r < 0 \\ mut\,(x_j) &= x_j + \omega, \quad r \geq 0 \end{aligned} \tag{A.5.5}$$

where x_j is a randomly chosen element of particle from the swarm and ω is randomly generated within the range $[0,\ 0.1\ \times\ para^j_{\max} - para^j_{\min}]$ which represent one tenth of the length of the search space. r is a random number between $+1$ and -1 and $para^j_{\max}$ and $para^j_{\min}$ are the upper and lower boundaries of each particle element respectively. However, the mutating space is kept unchanged all the time throughout the search and the space for the permutation of particles in the PSO is also limited by ω in (A.5.5). Fixing the size of mutating space all the time along the search may not be the best approach.

Algorithm A.5.3: PSEUDO CODE FOR PSO WITH GA MUTATION (*HGAPSO*)

$t \leftarrow 0$

Initialize $X(t)$

output $(f(X(t)))$

while <not termination condition>

do

- $t \leftarrow t + 1$
- Perform the process of PSO (shown in Algorithm. A.5.2)
- Perform wavelet mutation operation with p_m in $(A.5.5)$
- **output** $(X(t))$
- **output** $(f(X(t)))$

return $(\hat{\mathbf{x}})$

comment: return the best solution

Bibliography

[Alexakis2003] C. Alexakis, H. O. Nyongesa, and R. Saatchi. Feature extraction and classification of electrocardiogram (ecg) signals related to hypoglycaemia. *Proceeding of Computers in Cardiology*, 30:537–540, 2003.

[Alexakis2006] C. Alexakis, H. O. Nyongesa, and R. Saatchi. A knowledge-based electrocardiogram monitoring system for detection of the onset of nocturnal hypoglycaemia in type 1 diabetic patients. *Proceeding of Computers in Cardiology*, 1:5–8, 2006.

[Altman1994] D. G. Altman and J. M. Bland. Statistics notes: Diagnostic tests 1: sensitivity and specificity. *Clinical chemistry*, (308):1552–1552, 1994.

[Amaral2008] C. E. F. Amaral and B. Wolf. Current development in non-invasive glucose monitoring. *Medical Engineering and Physics*, 30(5):541–549, 2008.

[Angeline1998] P. J. Angeline. Detection of nocturnal hypoglycemic episodes using eeg signals. In *IEEE World Congress on Computational Intelligence*, pages

84–89, 1998.

[Ash1994] T. Ash and G. W. Cottrell. *A review of learning algorithms that modify network topologies.* San Diego, USA, 1994.

[Assoc2005] A. D. Association. Defining and reporting hypoglycemia in diabetes. *Diabetes Care*, 28(5):1245–1249, 2005.

[Assoc2008] A. D. Association. Standards of medical care in diabetes. *Diabetes Care*, 31:12–54, 2008.

[Becker2000] D. J. Becker and C. M. Ryan. Hypoglycemia: A complication of diabetes therapy in children. *Trends in Endocrinology and Metabolism*, 11(5):198–202, 2000.

[Benhorin1990] J. Benhorin, M. Merri, and M. Alberti. Long qt syndrome: new electrocardiographic characteristics. *Circulation*, 82(5):521–527, 1990.

[Billings2005] S. A. Billings and H. L. Wei. A new class of wavelet networks for nonlinear system identification. *IEEE Transactions Neural Networks*, 16(4):862–874, 2005.

[Block2008] C. D. Block and J. Vertommen. Minimally-invasive and non-invasive continuous glucose monitoring systems: indications, advantages, limitations and clinical aspects. *Current Diabetes Reviews*, 4(3):159–168, 2008.

[Castano2000] J. Castano, Y. Z. Wang, and D. Chuang. Blood glucose dependence of visual flicker threshold. *Diabetes Technology and Therapeutics*, 2(1):31–43, 2000.

[Chan2010] K. Y. Chan, S. H. Ling, T. S. Dillon, and H. T. Nguyen. Classification of hypoglycemic episodes for type 1 diabetes dellitus based on neural networks. In *IEEE Congress on Evolutionary Computation*, pages 1–5, 2010.

[Chan2011] T. S. Dillon K. Y. Chan, S. H. Ling and H. T. Nguyen. Diagnosis of hypoglycemic episodes using a neural network based rule discovery system. *Expert Systems with Applications*, 38(8):9799–9808, 2011.

[Chan2012] K. Y. Chan, S. Khadem, and T. S. Dillon. Selection of significant on-road sensor data for short-term traffic flow forecasting using the taguchi method. *IEEE Transactions on Industrial Informatics*, 8(2):255–266, 2012.

[Chandana2005] S. Chandana and R. V. Mayorga. The new rough neuron. In *International Conference on Neural Networks and Brain*, pages 13–18, 2005.

[Chandana2007] S. Chandana and R. V. Mayorga. Rough approximation based on neuro-fuzzy inference systems. In *5th International Conference on Hybrid Intelligent Systems*, pages 518–520, 2007.

[Chang2009] C. L. Chang and M. Y. Hsu. The study that applies artificial intelligence and logistic regression for assistance in differential diagnostic of pancreatic cancer. *Expert Systems with Applications*, 36:10663–10672, 2009.

[Chu2008] A. Chu, H. Ahn, B. Halwan, and A. Kumar. A decision support system to facilitate management of patients with acute gastrointestinal bleeding. *Artificial Intelligence In Medicine*, 42(3):247–259, 2008.

[Clarke2012] S. F. Clarke and J. R. Foster. A history of blood glucose meters and their role in self-monitoring of diabetes mellitus. *British Journal of Biomedical Science*, 69(2):83–92, 2012.

[Cryer1985] P. E. Cryer and J. E. Gerich. Glucose counterregulation, hypoglycemia, and intensive insulin therapy in diabetes mellitus. *Journal of Medicine*, 313(4):232–241, 1985.

[Cryer1999] P. E. Cryer. Symptoms of hypoglycemia, thresholds for their occurrence and hypoglycemia unawareness. *Endocrinology and Metabolism Clinics of North America*, 28(3):495–500, 1999.

[DCCT1991] DCCT Research Group. Epidemiology of severe hypoglycemia in the diabetes control and complication trial. *The American Journal of Medicine*, 90(4):450–459, 1991.

[DCCT1995] DCCT Research Group. Adverse events and their association with treatment regimens in the diabetes control and complications trial. *Diabetes Care*, 18(11):1415–1427, 1995.

[Daubechies1992] I. Daubechies. *Ten lectures on wavelets.* Philadelphia, PA, 1992.

[Davis1991] L. Davis. *Handbook of genetic algorithms.* Van Nostrand Reinhold, NY, 1991.

[Eberhart2000] R. C. Eberhart and Y. Shi. Comparing inertia weights and constriction factors in particle swarm optimization. In *Proceedings of IEEE Congress on Evolutionary Computing*, pages 84–88, 2000.

[Eberhart2007] R. C. Eberhart. *Computational intelligence: concepts to implementations.* Morgan Kaufmann, Boston, 2007.

[Field1989] J. B. Field. Hypoglycemia: Definition, clinical presentations, classification, and laboratory tests. *Diabetes Research Laboratory*, 18(1):27–43, 1989.

[Frederick2008] L. G. Frederick and J. Zrebiec. Detection of hypoglycemia by children with type 1 diabetes 6 to 11 years of age and their parents: a field study. *Pediatric*, 121(3):e489–e495, 2008.

[Freedman2005] D. A. Freedman. *Statistical models : theory and practice.* Cambridge University Press, NY, 2005.

[Frier2007] B. M. Frier and M. Fisher. *Hypoglycaemia in clinical diabetes.* John Wiley and Sons, NJ, 2007.

[Ghevondian1997] N. Ghevondian and H. T. Nguyen. Using fuzzy logic reasoning for monitoring hypoglycemia in diabetic patients. In *19th Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, pages 1108–1111, 1997.

[Ghevondian1998] N. Ghevondian and H. T. Nguyen. Skin impedance measurements used for early onset of hypoglycemia in diabetic patients. In *20th Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, pages 1610–1613, 1998.

[Ghevondian2001] N. Ghevondian and H. T. Nguyen. A novel fuzzy neural network estimator for predicting hypoglycaemia in insulin-induced subjects. In *23rd Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, pages 1657–1660, 2001.

[Graaff1999] J. C. Graaff, G. J. Hemmes, and T. Bruin. Influence of repetitive finger puncturing on skin perfusion and capillary blood analysis in patients with diabetes mellitus. *Clinical Chemistry*, 45(12):2200–2206, 1999.

[Hallin1999] L. L. Hallin, A. Englund, and U. Adamson. Increased qt dispersion during hypoglycaemia in patients with type 2 diabetes mellitus. *Journal of Internal Medicine*, 246(3):299–307, 1999.

[Ham2001] F. M. Ham. *Principles of neurocomputing for science and engineering.* McGraw Hill, NY, 2001.

[Handojoseno2012] A. M. A. Handojoseno, J. M. Shine, and H. T. Nguyen. The detection of freezing of gait in parkinsons disease patients using eeg signals based on wavelet decomposition. In *34th Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, pages 69–72, 2012.

[Hansen1983] K. A. Hansen and S. C. Duck. Teledyne sleep sentry: evaluation in pediatric patients for detection of nocturnal hypoglycemia. *Diabetes Care*, 6(6):597–600, 1983.

[Harrad1985] R. A. Harrad, C. S. Cockram, and A. P. Plumb. The effect of hypoglycaemia on visual function: a clinical and electrophysiological study. *Clinical Science*, 69(6):673–679, 1985.

[Hassan2002] Y. Hassan, E. Tazaki, S. Egawa, and K. Suyama. Rough neural classifier system. *IEEE International Conference on Systems, Man and Cybernetics*, pages 1–6, 2002.

[Hassan2010] Y. F. Hassan. Rough sets for adapting wavelet neural networks as a new classifier system. *Applied Intelligence*, 35(2):260–268, 2010.

[Hastings1998] G. Hastings, N. Ghevondian, and H. T. Nguyen. A self-organising fuzzy estimator for hypoglycaemia monitoring in diabetic patients. In *20th*

Annual International Conference of the IEEE Engineering in Medicine and Biology Society, pages 1371–1374, 1998.

[Haykin1999] S. S. Haykin. *Neural networks : a comprehensive foundation.* Prentice Hall, Upper Saddle River, NJ, 1999.

[Hebb1949] D. O. Hebb. *The organization of behavior: a neuropsychological theory.* Wiley, NY, 1949.

[Heise1994] H. M. Heise, R. Marbach, and T. Koschinsky. Noninvasive blood glucose sensors based on near-infrared spectroscopy. *Artificial Organs*, 18(6):439–447, 1994.

[Hilsted1984] J. Hilsted, F. B. Petersen, and M. B. Nrgaard. Haemodynamic changes in insulin-induced hypoglycaemia in normal man. *Diabetologia*, 26(5):328–332, 1984.

[Iaione2005] F. Iaione and J. L. Marques. Methodology for hypoglycaemia detection based on the processing, analysis and classification of the electroencephalogram. *Medical and Biological Engineering and Computing*, 43(4):501–507, 2005.

[Ireland2000] R. H. Ireland, R. T. Robinson, and S. R. Heller. Measurement of high resolution ecg qt interval during controlled euglycaemia and hypoglycaemia. *Physiological Measurement*, 21(2):295–303, 2000.

[Jagielska1998] I. Jagielska. Hybrid rough sets/neural network approach to the development of a decision support system. *IEEE World Congress on Computational Intelligence*, 1:24–28, 1998.

[Jagielska1999] I. Jagielska, C. Matthews, and T. Whitfort. An investigation into the application of neural networks, fuzzy logic, genetic algorithms, and rough sets to automated knowledge acquisition for classication problems. *Neurocomputing*, 24:37–54, 1999.

[Jain1996] A. K. Jain, J. Mao, and K. M. Mohiuddin. Artificial neural networks: a tutorial. *IEEE Computer Society*, 29(3):31–44, 1996.

[Jaroslaw2003] S. Jaroslaw and K. Katarzyna. Hybrid classifier based on rough sets and neural networks. *Electronic Notes in Theoretical Computer Science*, 82(4):228–238, 2003.

[Jiang2004] W. Jiang and G. Seong Kong. Block-based neural networks for personalized ecg signal classification. *IEEE Transactions on Neural Networks*, 18(6):1750–1761, 2004.

[Jiang2007] W. Jiang, S. G. Kong, and G. D. Peterson. Ecg signal classification using block-based neural networks. *IEEE Transactions on Neural Networks*, 18(6):1750–1761, 2007.

[Johansen1986] K. Johansen, S. Ellegaard, and S. Wex. Detection of nocturnal hypoglycemia in insulin-treated diabetics by a skin temperature-skin conductance meter. *Journal of Internal Medicine*, 220(3):213–217, 1986.

[Juang2004] C. F. Juang. A hybrid of genetic algorithm and particle swarm optimization for recurrent network design. *IEEE Transactions on Systems, Man, and Cybernetics: Part B*, 34(2):997–1006, 2004.

[Juhl2010] C.B. Juhl, K. Hojlund, and R. Elsborg. Automated detection of hypoglycemia-induced eeg changes recorded by subcutaneous electrodes in subjects with type 1 diabetes. *Diabetes Research and Clinical Practice*, 88(1):22–28, 2010.

[Kajiwara1993] K. Kajiwara, T. Uemura, and H. Kishikawa. Noninvasive measurement of blood glucose concentrations by analysing fourier transform infrared absorbance spectra through oral mucosa. *Medical and Biological Engineering and Computing*, 31(8):S17–S22, 1993.

[Kennedy1995] J. Kennedy and R. Eberhart. Particle swarm optimization. In *International Joint Conference on Neural Networks*, pages 1942–1948, 1995.

[Kim1998] S. S. Kim. Time-delay recurrent neural network for temporal correlations and prediction. *Neurocomputing*, 20(1):253–263, 1998.

[Kimura1987] J. Kimura, N. Ito, and T. Kuriyama. A novel blood glucose monitoring method. *Chemical Sensors*, 87(9):327–333, 1987.

[Kinoshita1987] J. Kinoshita and N. G. Palevsky. Computing with neural networks. *Neurocomputing*, 7(5):24–32, 1987.

[Klonoff2001] D. C. Klonoff. The need for hypoglycemia detection and prevention in type 1 diabetes. *Diabetes Technology and Therapeutics*, 3(4):567–570, 2001.

[Koivikko2007] M. L. Koivikko, M. Karsikas, and P. I. Salmela. Effects of controlled hypoglycaemia on cardiac repolarisation in patients with type 1 diabetes. *Diabetologia*, 51(3):426–435, 2007.

[Korrek2010] M. Korrek and B. Dogan. Ecg beat classification using particle swarm optimization and radial basis function neural network. *Expert Systems with Applications*, 37(12):7563–7569, 2010.

[Laitinen2008] M. D. T. Laitinen and T. L. Laitinen. Electrocardiographic alterations during hyperinsulinemic hypoglycemia in healthy subjects. *Annals of Noninvasive Electrocardiology*, 13(2):97–105, 2008.

[Lam2007] H. K. Lam and F. H. F. Leung. Design and training for combinational neural-logic systems. *IEEE Transactions on Industrial Electronics*, 54(1):612–619, 2007.

[Lam2009] H. K. Lam and H. F. Leung. Design and training for combinational neural-logic systems. *IEEE Transactions on Industrial Electronics*, 54(1):612–619, 2009.

[Lee2004] S. P. Lee, L. Yeoh, and N. D. Harris. Influence of autonomic neuropathy on qtc interval lengthening during hypoglycemia in type 1 diabetes. *Diabetes*, 53(6):1535–1542, 2004.

[Lee2005] S. P. Lee, N. D. Harris, and R. T. Robinson. Effect of atenolol on qtc interval lengthening during hypoglycaemia in type 1 diabetes. *Diabetologia*, 48(7):1269–1272, 2005.

[Leung2003] F. H. F. Leung, H. K. Lam, S. H. Ling, and P. K. S. Tam. Tuning of the structure and parameters of neural network using an improved genetic algorithm. *IEEE Transactions on Neural Networks*, 14(1):79–88, 2003.

[Ling2008] S. H. Ling, H. C. C. Iu, K. Y. Chan, and F. H. F Leung. Hybrid particle swarm optimization with wavelet mutation and its industrial applications. *IEEE Transactions on systems, man, and cybernetics:Part B*, 38(3):743–763, 2008.

[Ling2008c] S. H. Ling, F. H. F. Leung, and K. Y. Chan. Improved hybrid particle swarm optimized wavelet neural network for modeling the development of fluid dispensing for electronic packaging. *IEEE Transactions on Industrial Electronics*, 55(9):3447–3460, 2008.

[Ling2010] S. H. Ling, N. Nuryani, and H. T. Nguyen. Evolved fuzzy reasoning model for hypoglycaemic detection. In *32th Annual International Conference of*

the IEEE Engineering in Medicine and Biology Society, pages 4662–4665, 2010.

[Ling2010b] S. Ling. *Genetic algorithm and variable feed-forward neural networks.* Saarbrcken, Germany, 2010.

[Ling2011] S. H. Ling and H. T. Nguyen. Genetic-algorithm-based multiple regression with guzzy inference system for detection of nocturnal hypoglycemic episodes. *IEEE Transactions on Information Technology in Biomedicine*, 15(2):308–315, 2011.

[Ling2012] S. H. Ling and H. T. Nguyen. Natural occurrence of nocturnal hypoglycemia detection using hybrid particle swarm optimized fuzzy reasoning model. *Artificial Intelligence In Medicine*, 55(3):177–184, 2012.

[Lingras1996] P. Lingras. Rough neural networks. In *Internatinal Conference on Information Processing and Management of Uncertainty in Knowledge-based Systems*, pages 1445–1450, 1996.

[Lingras1998] P. Lingras. Comparison of neofuzzy and rough neural networks. *Information Sciences*, 110(3):207–215, 1998.

[Lippmann1987] R. P. Lippmann. An introduction to computing with neural nets. *IEEE ASSP Magazine*, 4(2):4–22, 1987.

[Liu2004] H. Liu, H. Tuo, and Y. Liu. Rough neural network of cariable precision. *Neural Processing Letters*, 19(1):73–68, 2004.

[Looney1997] C. G. Looney. *Pattern recognition using neural networks : theory and algorithms for engineers and scientists.* New York, Oxford University Press, 1997.

[Maia2007] F. F. R. Maia and L. R. Arajo. Efficacy of continuous glucose monitoring system (cgms) to detect postprandial hyperglycemia and unrecognized hypoglycemia in type 1 diabetic patients. *Diabetes Research and Clinical Practice*, 75(1):30–34, 2007.

[Malin1994] S. F. Malin, T. L. Ruchti, and T. B. Blank. Noninvasive prediction of glucose by near-infrared diffuse reflectance spectroscopy. *Clinical Chemistry*, 45(9):1651–1658, 1999.

[Mallat1989] S. G. Mallat. A theory for multiresolution signal decomposition: The wavelet representation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 11(7):674–693, 1989.

[Manic2002] M. Manic and B. Wilamowski. Robust algorithm for neural network training. In *Proceedings of the 2002 International Joint Conference on Neural Networks*, pages 1528–1533, 2002.

[Marques1997] J. L. Marques. Altered ventricular repolarization during hypoglycaemia in patients with diabetes. *Diabet Medicine*, 14(8):648–654, 1997.

[Mazumdar2008] J. Mazumdar and R. G. Harley. Recurrent neural networks trained with backpropagation through time algorithm to estimate nonlinear load harmonic currents. *IEEE Transactions on Industrial Electronics*, 55(9):3484–3491, 2008.

[Merbis1996] M. A. E. Merbis, F. J. Snoek, K. Kanc, and R. J. Heine. Hypoglycaemia induces emotional disruption. *Patient Education and Counseling*, 29(1):117–122, 1996.

[Michalewicz1994] Z. Michalewicz. *Genetic algorithms + data structures = evolution programs (2nd, extended ed.)*. Springer-Verlag, 1994.

[Moller1993] M. F. Moller. A scaled conjugate gradient algorithm for fast supervised learning. *Neural Networks*, 6(4):525–533, 1993.

[Moon2001] S. W. Moon and S. G. Kong. Block-based neural networks. *IEEE Transactions on Neural Network*, 12(2):307–317, 2001.

[Moon2002] S. W. Moon and S. G. Kong. Pattern recognition with block-based neural networks. *International Joint Conference on Neural Networks*, pages 992–996, 2002.

[Murphy2004] N. P. Murphy, M. E. Ford-Adams, and K. K. Ong. Prolonged cardiac repolarisation during spontaneous nocturnal hypoglycaemia in children and adolescents with type 1 diabetes. *Diabetologia*, 47(11):1940–1947, 2004.

[Nguyen2006] H. T. Nguyen, N. Ghevondian, and T. W. Jones. Neural-network detection of hypoglycemic episodes in children with type 1 diabetes using physiological parameters. In *28th Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, pages 6053–6056, 2006.

[Nguyen2007] H. T. Nguyen, N. Ghevondian, S. T. Nguyen, and T. W. Jones. Detection of hypoglycemic episodes in children with type 1 diabetes using an optimal bayesian neural network algorithm. In *International Conference of IEEE Engineering in Medicine and Biology Society*, pages 3140–3143, 2007.

[Nguyen2008] H. T. Nguyen, N. Ghevondian, and T. W. Jones. Detection of nocturnal hypoglycemic episodes (natural occurrence) in children with type 1 diabetes using an optimal bayesian neural network algorithm. In *30th Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, pages 1311–1314, 2008.

[Nguyen2010] H. T. Nguyen and T. W. Jones. Detection of nocturnal hypoglycemic episodes using eeg signals. In *32nd Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, pages 4930–4933, 2010.

[Nguyen2012] L. L. Nguyen, S. Su, and Hung T. Nguyen. Identification of hypoglycemia and hyperglycemia in type 1 diabetic patients using ecg parameters. In *34th Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, pages 2716–2719, 2012.

[Nielsen1990] R. Hecht-Nielsen. *Neurocomputing.* Addison Wesley, MA, 1990.

[Noel2004] M. M. Noel and T. C. Jannett. Simulation of a new hybrid particle swarm optimization algorithm. In *Proceedings of the 36th IEEE Southeastern Symposium on System Theory*, pages 150–153, 2004.

[Nuryani2012] N. Nuryani, S. H. Ling, and H. T. Nguyen. Electrocardiographic signals and swarm-based support vector machine for hypoglycemia detection. *Annals of Biomedical Engineering*, 40(4):934–945, 2012.

[Osareh2008] A. Osareh and B. Shadgar. Intrusion detection in computer networks based on machine learning algorithms. *International Journal of Computer Science and Network Security*, 8(11):15–23, 2008.

[Pawlak1982] Z. Pawlak. Rough sets. *International Journal of Information and Computer Science*, 11:341–356, 1982.

[Pawlak1991] Z. Pawlak. *Rough sets: theoretical aspects of reasoning about data.* Kluwer Academic Publishers, London, 1991.

[Pawlak2002] Z. Pawlak. Rough set theory and its applications. *Journal of Telecommunication and Information Technology*, 3:710, 2002.

[Pramming1988] S. Pramming, B. Thorsteinsson, and B. Stigsby. Glycaemic threshold for changes in electroencephalograms during hypoglycaemia in patients with insulin dependent diabetes. *British Medical Journal*, 296(6623):665–667, 1988.

[Robinson2003a] R. T. Robinson, N. D. Harris, and R. H. Ireland. Mechanisms of abnormal cardiac repolarization during insulin-induced hypoglycemia. *Diabetes*, 52(6):1469–1474, 2003.

[Robinson2003b] R. T. Robinson, N. D. Harris, and R. H. Ireland. Comparative effect of human soluble insulin and insulin aspart upon hypoglycaemia-induced alterations in cardiac repolarization. *British Journal of Clinical Pharmacology*, 55(3):246–251, 2003.

[Robinson2004] R. T. Robinson, N. D. Harris, and R. H. Ireland. Changes in cardiac repolarization during clinical episodes of nocturnal hypoglycaemia in adults with type 1 diabetes. *Diabetologia*, 47:312–315, 2004.

[Rumelhart1986] D. E. Rumelhart, G. E. Hinton, and R. J. Williams. Learning internal representations by error propagation. *Parallel distributed processing: explorations in the microstructure of cognition*, 1:318–362, 1986.

[Seber2003] G. A. F. Seber. *Linear Regression Analysis.* New York, Wiley, 2003.

[Shichiri1982] M. Shichiri, Y. Yamasaki, and R. Kawamori. Wearable artificial endocrine pancreas with needle-type glucose sensor. *The Lancet*, 2(8308):1129–1131, 1982.

[Singh2011] S. Singh and P. R. Gupta. Breast cancer detection and classification using neural network. *International Journal of Advanced Engineering Sciences And Technologies*, 6(1):4–9, 2011.

[Tierneya2001] M. J. Tierneya, R. O. Pottsa, and L. Jovanovicb. Clinical evaluation of the glucowatch biographer: a continual, non-invasive glucose monitor for patients with diabetes. *Biosensors and Bioelectronics*, 16(9):621–629, 2001.

[Tsalikian2004] E. Tsalikian, C. Kollman, and N. Mauras. Glucowatch g2 biographer alarm reliability during hypoglycemia in children. *Diabetes Technology and Therapeutics*, 6(5):559–566, 2004.

[Vashist2012] S. K. Vashist. Non-invasive glucose monitoring technology in diabetes management: A review. *Analytica Chimica Acta*, 750:16–27, 2012.

[Weinstein2007] R. L. Weinstein. Accuracy of the 5-day freestyle navigator continuous glucose monitoring system: comparison with frequent laboratory reference measurements. *Diabetes Care*, 30(5):1125–1130, 2007.

[Windrow1990] B. Windrow and M. A. Lehr. 30 years of adaptive neural networks: Perception, madaline and backpropagation. *Proceeding of The IEEE*, 78(9):1415–1442, 1990.

[Wolpert2007] H. A. Wolpert. Use of continuous glucose monitoring in the detection and prevention of hypoglycemia. *Journal of Diabetes Science and Technology*, 1(1):146–150, 2007.

[Yamaguchi2008] D. Yamaguchi, F. Katayama, and M. Takahashi. The medical diagnostic support system using extended rough neural network and multi-agent. *Artificial Life and Robotics*, 13(1):184–187, 2008.

[Yu2010] X. Yu. *Introduction to evolutionary algorithms.* New York, Springer, 2010.

[Yuliang2009] Q. Yuliang. Integrated method of rough set and neural network and its application study. In *6 th International Conference on Fuzzy Systems and Knowledge Discovery*, pages 539–543, 2009.

[Zekri2008] M. Zekri and S. Sadri. Adaptive fuzzy wavelet network control design for nonlinear systems. *IEEE Transactions on systems, man, and cybernetics:Part B*, 20:2668–2695, 2008.

[Zhang2009] Z. Zhang, Y. Shi, and G. Gao. A rough set-based multiple criteria linear programming approach for the medical diagnosis and prognosis. *Expert Systems with Applications*, 36(5):8932–8937, 2009.

[Zhao2011] L. Wang Y. Zhao, H. Jin and S. Wang. Rough neuron network for fault diagnosis. *International Journal of Image, Graphics and Signal Processing*, 3(2):51–58, 2011.

[Zurada1992] J. M. Zurada. *Introduction to artificial neural systems*. St. Paul, West, 1992.