Application of Artificial Neural Networks in Early Detection of Mastitis from Improved Data Collected On-Line by Robotic Milking Stations

---

A Thesis Submitted in Partial Fulfillment of the

Requirements for the Degree of

Master of Applied Computing

by

Zhibin Sun

Lincoln University

New Zealand

2008

Abstract

Two types of artificial neural networks, Multilayer Perceptron (MLP) and Self-organizing Feature Map (SOM), were employed to detect mastitis for robotic milking stations using the preprocessed data relating to the electrical conductivity and milk yield.

The SOM was developed to classify the health status into three categories: healthy, moderately ill and severely ill. The clustering results were successfully evaluated and validated by using statistical techniques such as K-means clustering, ANOVA and Least Significant Difference. The result shows that the SOM could be used in the robotic milking stations as a detection model for mastitis.

For developing MLP models, a new mastitis definition based on higher EC and lower quarter yield was created and Principle Components Analysis technique was adopted for addressing the problem of multi-colinearity existed in the data. Four MLPs with four combined datasets were developed and the results manifested that the PCA-based MLP model is superior to other non-PCA-based models in many respects such as less complexity, higher predictive accuracy. The overall correct classification rate (CCR), sensitivity and specificity of the model was 90.74 %, 86.90 and 91.36, respectively. We conclude that the PCA-based model developed here can improve the accuracy of prediction of mastitis by robotic milking stations.

Keywords: Artificial Neural Network; Multilayer Perceptron; Self-organizing Feature Map; Principle Components Analysis;

Acknowledgements

First of all, I would like to express my heartfelt gratitude to my supervisor, Associate Professor Sandhya Samarasinghe, for her invaluable help, encouragement, useful suggestions and patience in providing my main supervision throughout the research project. Without her consistent guidance and support, this thesis may never have reached its present form.

I would also like to express my deepest gratitude to all professors, lecturers and staffs who have taught, instructed and helped me in the past two years: Professor Don Kulasiri, Professor Alan Mckinnon, Keith Unsworth, Dr Crile Doscher, Dr Brad Case, Dr Magdy Mohssen, Jane Swift, and Tracey Shields.

I must thank my beloved wife Jing Chen and my lovely daughter Feiman Sun, for their love, patience and support throughout my studies in New Zealand. I also must thank my parents, Baosan Sun and Yufeng Hu, and my brother, Zhiyong Sun, for their loving considerations and great confidence in me during my studies in New Zealand.

Table of Contents

List of Tables

List of Figures

Chapter 1

1 Introduction

Bovine mastitis is the most costly disease in the dairy industry and exists in every herd. Recent research conducted by Dairy NZ, a dairy research organization in New Zealand, shows that mastitis costs the dairy industry $180 million annually. The early detection of mastitis, therefore, is crucial for farmers' economic gain because it allows prompt treatment, a higher rate of recovery, reduce the risk of infection being passed onto other cows and help prevent the development of chronic infections (Bentley & Lacy-Hulbert, 2007).

Early detection of mastitis can be performed on-line by Robotic Milking Systems (RMS) which have been used for several years in dairy industry. However, their detection results can not be fully relied on due to the fact that these detection model algorithms are mainly based on Electrical Conductivity (EC), whose value as a mastitis detector has been argued for a long time as its value is easily influenced by a number of factors (Mein, Sherlock & Claycomb, 2004). Since other milk parameters such as quarter milk yield can be measured automatically during milking, it would be desirable if this information is incorporated into the algorithm of the models. Thus, mastitis would be detected not only based on changed EC values but also on the changes in quarter yield,

which would lead to higher predicting accuracy. In this study, Artificial Neural Networks (ANN) technique is adopted to develop such a model.

ANN consists of neurons that mimic the human brain to perform complex tasks. An advantage of ANN is that it can detect patterns in complex and non-linear data. Cows with mastitis have different milk traits pattern than healthy cows such as higher EC and lower quarter yield (Yamamoto, 1985; Aoki, 1992; Lake, 1992). By presenting these patterns, an ANN should be able to learn how to map them to their corresponding outputs (input-output mapping). A well-trained ANN should be able to classify new patterns correctly and thus provide reliable predictions for new situations. The outcomes of this research study may be applied to future development of on-line mastitis detection systems.

1.1 Objectives of the Research

- Find out proper data-processing methods for differentiating mastitic and non-mastitc quarters so that the ANN can be well trained.

- Develop a mastitis detection model with sufficient accuracy by using Multilayer Perceptron (MLP) network based on the EC and quarter yield

- Develop a mastitis detection model by using Self-Organizing Map (SOM) to categorize cows in terms of health states. These health states will be healthy, moderately ill and severely ill.

1.2 Structure of Thesis

The thesis is outlined as follows: Chapter 2 is a brief review of background including mastitis, RMS, and the usefulness of milk composition for detecting mastitis. The basic concept of ANN is introduced in Chapter 3. A brief literature review on performance of ANN in detecting mastitis is presented in this chapter as well. Methods employed to achieve objects of the research are detailed in Chapter 4, which includes two main stages: data preprocessing and model development. Chapter 5 illustrates the results and the main findings are discussed. Finally, conclusions are presented in Chapter6.

Chapter 2

2.1 Background of Mastitis

Mastitis is inflammation of the udder and is caused by bacteria that enter through the teat canal, multiply in mammary tissue, and produce toxins that set up infection (NMAC, 2001). According to degrees of infection, mastitis can be defined as clinical mastitis and sub-clinical mastitis. Clinical mastitis is when the signs of infection can be seen, such as swollen teat, clotted milk and discolouration of the milk. In Sub-clinical mastitis, there are no visible signs appearing in milk or the udder. It can only be detected by laboratory examination (Sharif et al., 1998).

Mastitis has two main influences on milk: milk yield reduction and change of milk composition. A clinical quarter produces less milk than those that are healthy. Figure 2.1 illustrates a cow with lower quarter yield in the infected quarter. John et al (1992) found that the relative drop in milk yield at the quarters during the infection period was 15.3% ± 2.5%. However, a drop in milk yield does not alwayys mean that the cow has mastitis.

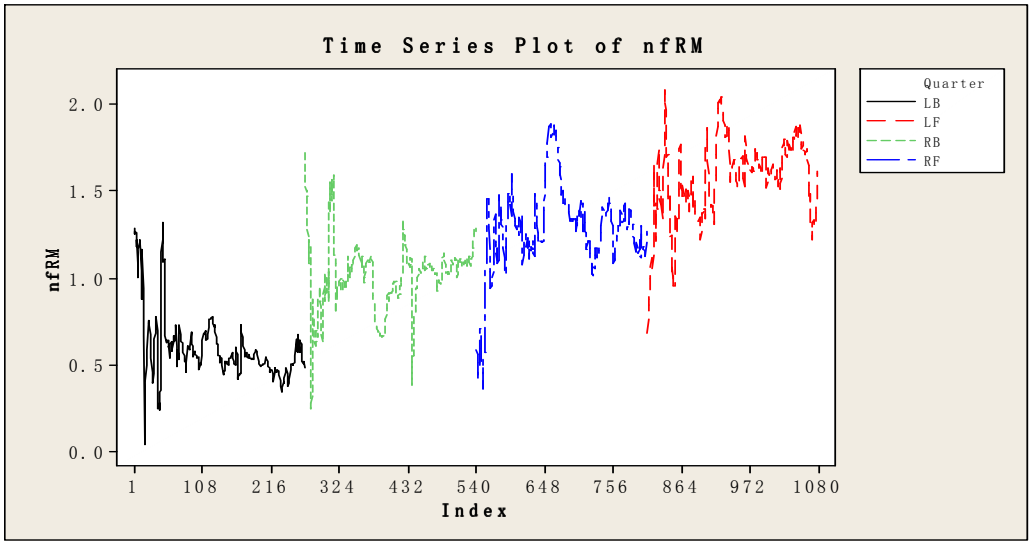


Figure 2.1 Normalized fraction of running mean quarter yield profiles for all 4 quarters of a cow with clinical mastitis in the LB. The quarters are named based on the locations of the quarter. B stands for back, F for front, L for left and R for right

Because of inflammation, the composition of milk from a sick quarter will be changed. This includes increased somatic cell counts (SCC), higher electrical conductivity (EC) and other components including fat and protein content (Auldist & Hubble, 1998). Somatic cells are mainly white blood cells sent to fight infection in the udder. When infection occurs and starts to damage the udder tissue, the immune system is called to action and very rapidly, large quantities of somatic cells are directed to the infection site (Leslie, Dohoo & Meek, 1983). A report form Dairy NZ states that a cow with SCC levels above 150,000 cells /ml is likely to be infected with mastitis.

As infection progresses, more cellular fluid enters into milk and the concentration of anions and cations in the milk increases. As a result, electrical

conductivity (EC) of the milk from the infected quarter is increased. In Figure 2.2 and 2.3, a cow with a high conductivity on the infected quarter is shown. Due to the correlation to mastitis, ease of measurement, and the low cost of recording, EC has been widely recognized as an important indicator for mastitis and employed in mastitis detecting system in the dairy industry (Barth, Fishcer & Worstorff, 2000). However, as shown in figure 2.4 not all the infected quarters have a highest EC. For such a case, EC alone is inadequate as an indicator for the detection of mastitis.

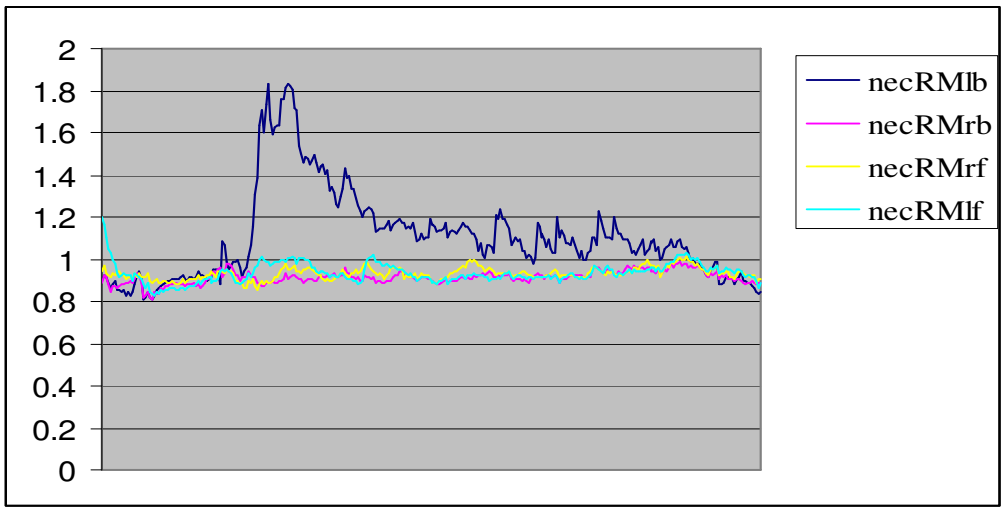


Figure 2.2 Normalized EC running mean profiles for all 4 quarters of a cow with clinical mastitis. The blue line indicates the EC profile of the infected quarter (lb). The quarters are named based on the locations of the quarter. b stands for back, f for front, l for left and r for right

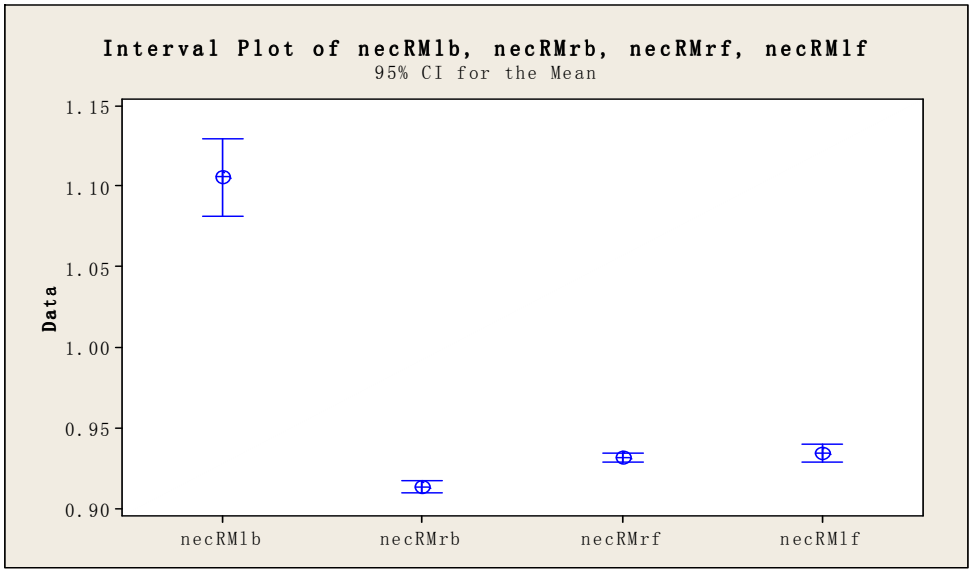


Figure 2.3 Mean Difference of electrical conductivity values for healthy and infected quarters. It shows that an infected quarter (necRMlb) has larger mean and variation than healthy quarters.

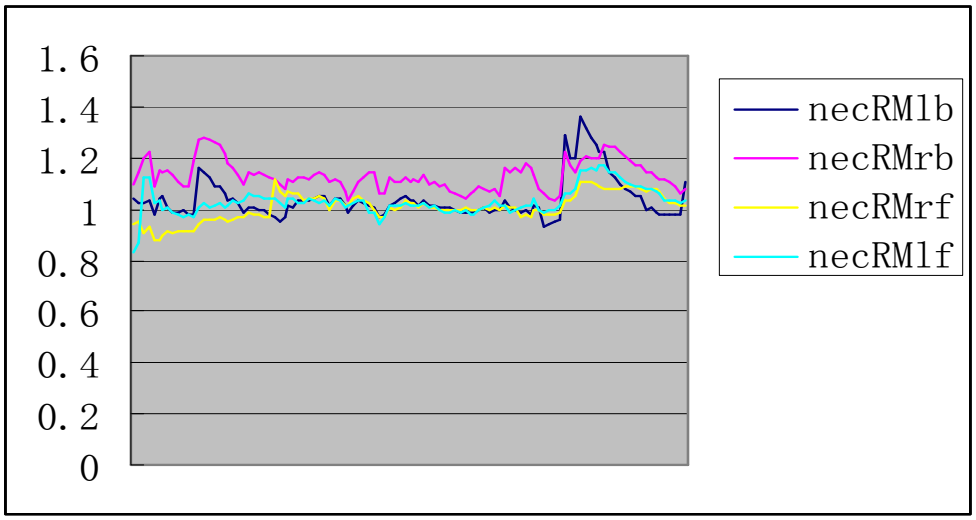


Figure 2.4 Normalized EC running mean profiles for all 4 quarters of a cow with clinical mastitis. This cow has mastitis on right front quarter (yellow line) but does not have the highest EC on the infected quarter.

2.2 Cost of Mastitis

Mastitis is one of the leading diseases and a serious problem in the dairy industry world wide. Past studies have found that the annual cost of mastitis per cow was around $200-300 in the USA and France (Blowey, 1986). In the U.K, the estimated cost of mastitis is around $150 to $200 million per year (Booth, 1988; Hillerton & Walton, 1991). Other studies of mastitis in Canada, Sweden, and the Netherlands have shown that dairy farmers suffer financial losses ranging from $125 to $250/yr per cow (Heuven et al., 1988; Miles et al., 1992; Monardes, 1994). The cost of mastitis for the average New Zealand dairy farmer is $36/cow. For the whole industry the figure amounts to more than $180 million per year (Bentley et al, 2006).

2.3 The Methods of Detection Used in Robotic Milking Systems (RMS)

In RMS, a detection model has been used to monitor the health status of the cows. It generates reports called Attention Lists that alerts the farmer to cows that may be sick. The algorithm of the model compare the EC value of the milk at each milking. If the EC value of a quarter is over 15% higher than the average of the two quarters with the lowest EC value, the quarter is detected as mastitis infected (Grennstam, 2005).

2.4 The Predictability of Milk Traits for Mastitis

Electrical Conductivity (EC)

Reports from some studies on the ability of EC to detect mastitis showed that the sensitivity (correctly detecting mastitis) and specificity (correctly detecting healthy cases) were on average about 65% and 75%, respectively (Sheldrake, McGregor & Hoare 1983; Batra & McAllister 1984; Lmsbergen et al. 1994). They pointed out that EC alone might not be a good measure to accurately discriminate between clinical and healthy cases, because EC of milk is easily affected by a number of factors. To improve the performance of EC in detecting mastitis, other measurements such as inter quarter ratio (IQR) has been investigated. IQR is the ratio between the quarter with the highest and lowest EC quarter value of the same cow. A study (Norberg, et al, 2004) showed that IQR of EC provided a much better result than directly using EC value alone. By using this trait, 80.6% of clinical and45% of sub-clinical cases were classified correctly in their study. They also added that the combination of EC with other traits could improve the ability to classify cows into udder health categories.

Somatic Cell Count(SCC)

Previous studies showed that the ability of SCC to determine mastitis states varies greatly. The sensitivities and specificities of SCC from these studies ranged from 40 to 70% and 60 to 89%, respectively. (Fernando et al., 1982;

McDermott et al., 1982; Rindsig et al., 1979; Schultz, 1977; Sheldrake et al., 1983). One possible reason for this variation, they pointed out, could be attributed to threshold level because different levels could lead to different sensitivity and specificity. For example, setting a low level of SCC threshold could result in high sensitivity and reduce the false negative rate (a cow incorrectly classified as healthy cow), whereas setting a high threshold could lead to high specificity and reduce the false-positive rate (a cow incorrectly classified as infected cow). Another reason for the variation could be due to the fact that not only mastitis but also many other factors could result in a raised SCC, such as the age, lactation stage, milking equipment and season. Like EC, SCC alone may not be the best indicator, even though it is used throughout the world as an indicator of mastitis.

Chapter 3

3.1 Artificial Neural Networks

An ANN processes information through interactions of a large number of neurons and obtains knowledge through a learning process. The knowledge is stored within connections between neurons (Samarasinghe, 2006). For different purposes, a variety of ANN can be constructed based on differences in the arrangement of the layers, the interconnection of elements, and the learning methods. For purposes of this research study, two types of ANN, Multilayer Perceptron Networks (MLP) and Self-organizing Feature Map (SOM) Neural Networks, were employed. In the following two sections, the basic concepts of them are reviewed.

3.2 Multilayer Perceptron (MLP) Networks

The MLP are the most common neural network for nonlinear prediction and classification, in which the processing elements or neurons are grouped into an input layer, hidden layers, and output layer. Figure 3.1 shows structure of a three layer MLP. The neurons in one layer are connected to each neuron in the adjacent layer and the strengths of these connections are called weights which are the free parameter, and which can be positive or negative.

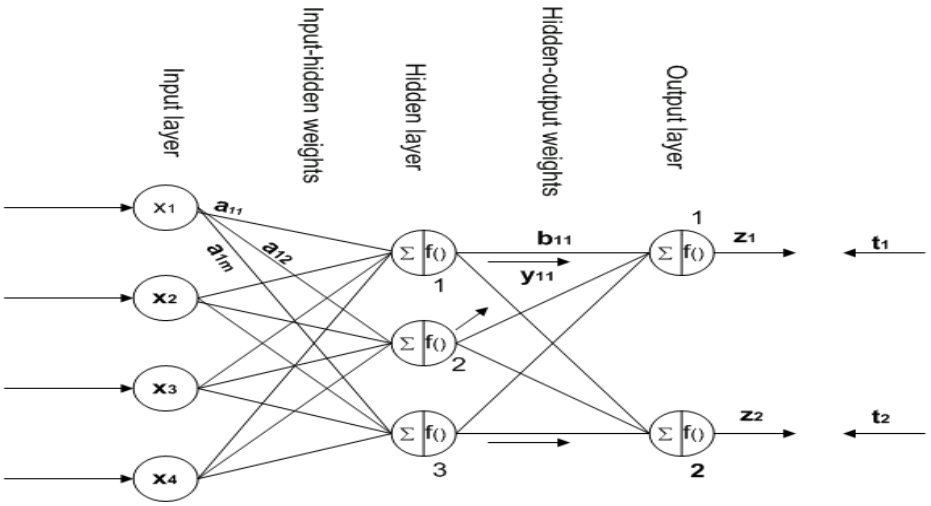


Figure 3.1 Architecture of a MLP, with four input neurons, three hidden neurons, two output neurons, and 18 weights.

The learning process of an MLP taking place during the training is a supervised process in which the target output is given for each input pattern. The goal of the training is to minimize the error by adjustting the weights. The error is the difference between the output generated by the network and the target output. MLP adopt back-propagation algorithm in which the delta rule is most often used to adjust the weights. Two repeated phases are involved in the application of this delta rule (Figure 3. 2).

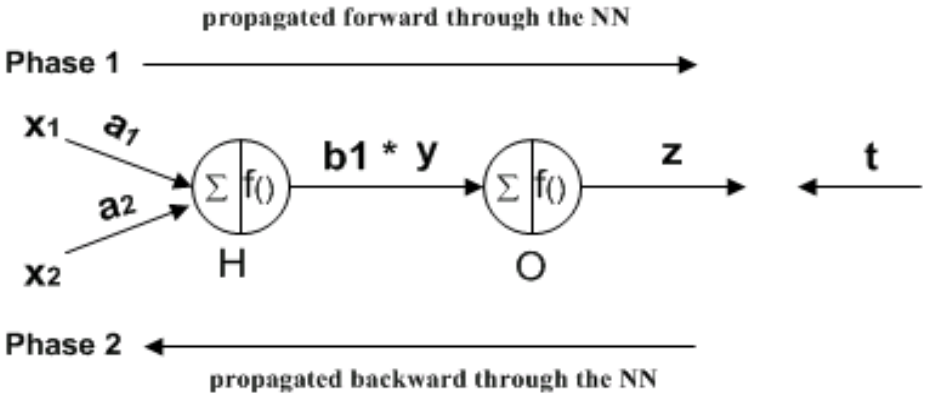


Figure 3.2 A simple network training example

During forward propagation phase of the NN, the input layer transmits input data (e.g. x1, x2) to the hidden neurons (H). The input data is weighted by the corresponding input-hidden weights (e.g. a1, a2) which initially are specified randomly. The effective input to a receiving neuron in the hidden layer is a weighted sum of the all inputs ($\sum$xiai). The hidden neurons then process this summed input value by processing it through an activation function (such as logistic, hyperbolic-tangent, Gaussian and sine function).

After the activations are calculated, the results are weighted by hidden-output weighs (bi) and then sent to output neurons which sum the weighted inputs ($\sum$**b**i**y**i) and pass them through activation function (f). The outputs of these neurons produce the network outputs (**z**). At this point, the errors are calculated (t-z), and then the second phase starts during which the errors are passed backwards through all hidden and input neurons. The adjustment of weights is in a gradient-descent fashion and can be performed after each presentation of an input-output data pattern (example by example learning), or after presentation of the entire or some portion of input-output data (batch learning). One pass of batch learning is called one epoch. Weight adjustment often is preferred to take place after every epoch (batch learning) because it generally provides stable solutions. In order to do this, the total squared sum of the error over these input-output pairs is calculated after each epoch. The average total squared sum of the error is called mean square error and can be calculated by

$$E = \frac{1}{2n}\left[\sum_{i=1}^{N}\left(t_i - z_i\right)^2\right] \qquad [3.1]$$

where n is the number of input-output pairs, t is the target output and z is the network output. The fraction ½ is arbitrary and used for mathematical convenience. This completes the forward propagation phase and error calculated is used to adjust weights in the back propagation phase.

The following equations defines the new weights (w_{m+1}) of a connection after mth epoch

$$w_{m+1} = w_m + \Delta w_m \qquad [3.2]$$

where $\boldsymbol{W}m$ is the older value of the same weight at mth epoch, ΔWm is the new increment of the weight change after epoch m and calculated as

$$\Delta w_m = -\varepsilon d_m \qquad [3.3]$$

where $\boldsymbol{\varepsilon}$ is learning rate, dm is total gradient for mth epoch and can be presented as

$$d_m = \sum_{N=1}^{N} \left[\frac{\partial E}{\partial W} \right]_n \qquad [3.4]$$

where n is the number of input-output pair, m is epoch number and E is mean square error. ∂E/∂Wm is the gradient of error with respect to weight in the mth epoch.

The process is repeated for weights in both layers and weights are adjusted. The training is finished when there is no error or it is acceptably small and the corresponding weights are the final weights in which all or maximum possible input-output data are correctly classified. If the model is well trained, it will be able to classify new input patterns. There are several improved variants of the delta rule learning algorithm, and these include adaptive learning rate, Newton's method and Lavenberg Marquardt method (Samarasinghe, 2006).

3.3 Self-organizing Feature Map Neural Networks

Self Organizing Feature Map (SOM or SOFM) involves a type of unsupervised learning in which the target outputs are not involved. An SOM is trained by showing examples of patterns (corresponding to input variables) that are to be clustered, and the network gradually learns to cluster these patterns into groups. The SOFM usually has two layers of neurons: an input layer and an output layer.

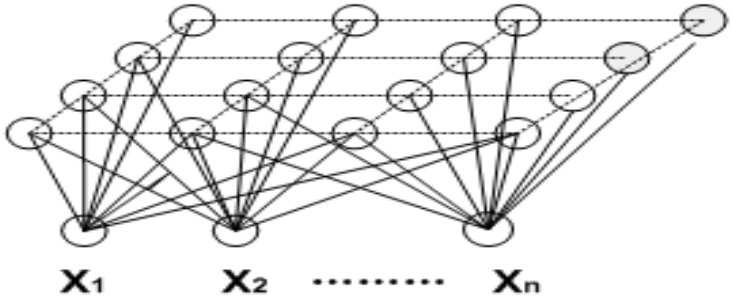


Figure 3.3 Configuration of a two dimensional SOM network

Figure 3.3 shows the structure of a two dimensional SOM network in which each output neuron is linked to the input neurons by corresponding weights.

A learning process of SOM is a type of competitive learning. Network weights are initially set to random values. The competition starts when the net input (weighted sum of inputs) is calculated by all the neurons in the network for a randomly presented input vector. Then each output neuron compares its net input or activation with each other and competes to be the winner. The neuron that has the highest activation is then defined as the winner. This competition can be implemented by using the concept of distance between an input and a weight vectors. Therefore, a winner can also be expressed as a neuron that has the smallest distance to input vector. The distance between an input and a weight vector (d *j*) can be calculated by

$$d_j = X - W_j = \sqrt{\sum_{i=1}^{n} \left(X_i - W_{ij}\right)^2} \qquad [3.5]$$

where **X** is the input vector, **W**$_j$ is weight vector associated with the jth output neuron. The x_i and w_{ij} are the i^{th} component of input vector and the j^{th} weight vector corresponding to input variable i.

The objective of the SOM learning is to adjust the weight vectors so that they, in repeated exposure to input vectors, respond appropriately reflecting the natural clustering in the training data. During the learning process, SOM not only adjusts the weight of the winner neuron but also the weights of neurons in a neighborhood of the winner neuron. SOM uses neighbor size and neighbor strength function to determine how much the neighbor neurons should adjust their weights. Neurons closer to the winner adjust weights more than those that are far from it. During the training process, the neighbor size and strength are decreased gradually until only the winner remains towards the end of training. This process can be expressed as

$$w_j(t) = w_j(t-1) + \beta(t) NS(d,t)\left[X(t) - w_j(t-1)\right] \qquad [3.6]$$

where W$_j$(t) is the weight update after t iterations, W$_j$(t-1) is the update after the previous iteration, β(t) is the learning rate which also gradually decreases with iterations, and NS(d,t) is the neighbor strength function, the X(t) is the input vector presented at the t^{th} iteration.

3.4 Use of ANN for Mastitis Diagnosis

In One study (Nielen et al, 1994), the ability of ANN to diagnose mastitis was explored. They used an ANN with the back-propagation technique and EC as the only input for training ANN. The network was trained with 17 healthy and 13 clinical mastitic quarters, all healthy and 12 of 13 mastitic quarters were classified correctly after training. They concluded that ANN was able to discriminate between normal and infected quarters without any correction for cow level. They also suggested that "further development should include the use of different input parameters for building more robust models."

In their later study carried out in 1995, they compared three analysis techniques for on-line detection of mastitis. These techniques were principal component analysis (PCA), linear logistic regression (LRM) and multi-layer perceptron (MLP). The variables employed in the study were improved from original data and included milk production, milk temperature, and electrical conductivity. The study showed that the ANN with three-layer back-propagation had a slightly higher sensitivity and specificity than other techniques did.

In research carried out by Yang et al. (1999), a number of variables were used for detection of mastitis including: SCC, lactation number, milk yield, days in milking, mean SCC for herd, herd size, season of calving, milk components. Three dataset were created based on the different ratio of clinical to healthy

records (1:1, 1:10 and 1:300). They used a 2*2 contingency table to assess the sensitivity and specificity of the model. The results of study indicated that network trained with a higher proportion of mastitic records (1:1 ratio) provided more superior recognition than those with a lower proportion of mastitic cases. They also found that higher proportion of healthy records leaded to more specificity. The network achieved 80% accuracy in distinguishing between clinical and healthy cows.

Nielen et al. (1995a) developed a back-propagation neural network for prediction of sub-clinical mastitis from on-line milking data. The variables used as inputs for training NN were EC per quarter, milk production per cow, parity groups (parity 1 and parity >1), and days in milking (DIM). Healthy periods were defined when a cow with four consecutive SCC measurements were < 200 × 10³cells/ml. The periods for the sub-clinical mastitis were defined as two levels. One was severe sub-clinical mastitis (SCC > 1000 × 10³cells/ml) and another was moderate sub-clinical mastitis (SCC between 500 and 1000 × 10^3 cells/ml). The neural network model achieved a sensitivity of 67% and specificity of 78%. They pointed out that the definition of the sub-clinical mastitis would have some influence on the sensitivity of model. As a result, the model could only be used for a herd with a fairly high incidence of sub-clinical mastitis.

In a study conducted by López-Benavides et al. (2003), a Self Organizing Feature Map (SOM) was developed to determine the health status of the cows in terms of the state of progression of mastitis. The variables used in the study were EC, protein percentage (PP), SCC, fat percentage (FP) and microbiological profile. They preprocessed original dataset and created new indicators for the model: Conductivity index (CI) and composite milk index (CMI). CI was derived from: CI = 2 + [(EC/100) – IQR]. IQR is the ratio of the EC of a quarter to sum total EC over all quarters. CMI was the sum of all the other variables. These two variables were used as input for the SOM and four health categories (healthy, moderately ill, ill, and severely ill) were defined. They suggested that CMI can be used as an indicator of mastitis status as it integrates several milk traits into one single measure. They concluded that the SOM model effectively clustered the cows into appropriate health categories.

Chapter 4

4.1 Methodology

This chapter provides the methods used to achieve the research objectives. It consists of two stages: data analysis and model development. Data analysis is an initial and important step in modeling. A well processed data will greatly enhance learning ability of ANN. In this study, scatter plots and correlation coefficients were employed to explore relationships and trends. Principle component analysis was used to account for multi-colinearity among variables. In the modeling stage, two kinds of ANN, MLP and SOM, were trained to detect the presence or absence of clinical mastitis. These two stages are described in detail next.

4.2 Data Analysis

4.2.1 Data and Variables

The data for this research study were supplied by Dairy NZ. It included two data files: treatment data file and milking data file. The treatment data included cowID, time of treatment, infected quarter, and SCC of all calls twice weekly. All this information was helpful to recognize mastitis cases in the milking data. The milking data file, which had been improved by another scientist in the Dairy NZ research team, had measurements taken by robots two or three times a day for each quarter of each cow during each milking. It contained 48,546

records (samples) representing 194 cows from end of July 2006 to early April 2007. The variables in this data file were Running Means of Normalized Quarter-yield Fraction (nyfRM) for each quarter; Running Means of Normalized Electrical Conductivity (necRM) for each quarter; Fractional Deviations from the smallest necRM value (necFD), cowID and Time of Milking. The explanations of each variable are presented below:

Running means of normalized quarter-yield fractions (nyfRM). This variable was created based on quarter yield and involved two steps of calculation: calculation of quarter-yield fraction (QYF) and calculation of running mean of QYF. The QYF is the ratio of milk yield of a quarter in a milking to the total milk yield from all four quarters in that milking. In the view of biology, it was calculated to account for the effect of biological differences on milk yield between each quarter of each cow. Running means of QYF was based on correction of QYF and it is illustrated below:

$$\chi_t^{'} = \chi_t \times \frac{1}{a} + \left(1 - \frac{1}{a}\right) \times \chi_{t-1} \qquad [4.1]$$

where $\chi_t^{'}$ is the running mean of QYF at time t, χ_t is the measured QYF at time t, χ_{t-1} is the measured QYF at time $t-1$ and a is a coefficient that represents 'running mean length'.

The running mean of QYF was calculated at two levels:

- *Herd level*. Value of 50 was assigned to a for this level. Thus, the herd running mean of QYF of the left-back quarter at milking t, for instance, approximates the average of the all left-back quarter QYF values through milking (t-50) to milking t. Since this contains the last 50 milkings, it covers the 50 cows that were consecutively milked prior to t
- *Quarter level*. This was calculated for individual quarters over their own history of milking. In this research, a value of 5 was adopted.

The Running Means of Normalized Quarter-yield Fraction was then calculated for each quarter of each cow at each milking by dividing the quarter running mean of QYF by its corresponding herd running mean of QYF at the milking. The reason for taking this particular normalization is that the herd normalization is able to provide some correction for machine problems which occur from time to time.

Running Means of Normalized Electrical Conductivity (necRM). This variable was calculated from electrical conductivity using the same procedure as that for the nyfRM. The only difference was, instead of QYF in Eq.4.1, x refers to the running mean of highest electrical conductivity of a quarter in a milking.

The fractional deviations from the smallest necRM value (necFD). This variable was defined as the relative deviation of *necRM* within quarters for each cow at each milking and was introduced to take in to account ratio of the three highest necRM values to the lowest one. It was calculated as follows:

$$necFD_i = \frac{necRM_i - necRM_{\min}}{necRM_{\min}} \qquad [4.2]$$

where $necFD_i$ is necFD of any of the four quarters of a cow, $necRM_i$ is normalized EC running mean ($necRM$) of the same quarter, and $necRM_{\min}$ is the smallest value of $necRM$ between four quarters. Reasoning behind the $necFD_i$ is as follows: Since not all four quarters are infected at a given time, it can be expected that the $necRM_{min}$ reflects a healthy state. Therefore, if a partiular quarter *i* becomes infected, its $necRM_i$ will be very high, thereby yielding a high $necFD_i$ compared to a healthy quarter whose fractional deviation according to Eq. 4.2 will be near zero.

4.2.2 Definitions of Mastitis and Healthy Quarters

The health state was defined on the basis of necFD, nyfRM and information on treatment data. According to the treatment data, 163 quarters of 43 cows in the milking data have received treatment and these quarters were defined as mastitic. In addition, it is very important to consider the health state before and after the treatment took place because the ability of early detection strongly depend on the length of the infection period around the date established for a

case of mastitic cow. There was no information about this period recorded in the treatment data. One literature (Mele et al, 2001) chose 7 days before and 7 days after for clinical and 10 days after and 10 days before for sub-clinical mastitis. De Mol et al (2001) took 10 days before and 7 days after for clinical mastitis. However, these were not the case in this study, in which the combination of two thresholds with respect to higher EC and reduced milk yield were used. John Bramley et al (1992) found that during the infection period, the reduction in milk yield at the clinical quarter was 15.3% ± 2.5%. The threshold of 12.8% was used in the present study as it ensured that most of the clinical quarters were recognized. The EC threshold was defined as the value of necRM that was over 15% higher than the average of the two quarters with the lowest necRM value. Therefore, all quarter milking samples, before or after the date treatment took place, was defined as mastitis if it had 12.8% drops in milk yield and 15% higher than average of the two lowest quarters in EC value. As a result, the infected period around the date treatment was performed varied from quarter to quarter. It was found by visually examining that the longest time interval was 19 days before an infected quarter was recorded in the treatment data and the shortest was 3 days. The quarters that met the two thresholds values but not treated, and therefore, not recorded in treatment data, were not adopted in this study.

The healthy quarter milking was defined as follows: the quarter never showed

on the treatment data, which means it had never been recorded as mastitic, and the weekly SCC value was always below 150,000cells/ml as this value was recommended by Dairy NZ as a proper threshold for predicting infected and non-infected quarters. Depending on mastitis definition, 895 clinical quarters and 3235 healthy quarters were found in the milking data. Therefore, a new data set was generated for the analysis in which there were a total of 4130 quarters and the ratios of healthy to sick samples approximated 4.6:1.

4.2.3 Correlation Scatter Plots

Scatter plots were created to get a better understanding of the data with respect to spread, trends and correlations among variables. Figure 4.1 shows plots of all input variables for relationship analysis. The off-diagonal scatter plots show how individual variables related to one another. Points lying on a line indicate a linear relationship; a dispersive set of points denotes a nonlinear relationship. It can be observed that necRM and necFD have a strong positive correlation, while nfRM has a weak correlation with both necRM and necFD. It also can be seen from the plots, such as the one: nfRM against necRM, that infected patterns (red points denoted by bacteriological state BS=1) generally have higher necRM value than non-infected patterns (black points denoting BS=0), which highlights the fact that a quarter with mastitis has higher electric conductivity than those are healthy. Furthermore, the figure reveals that infected and non-infected patterns partially overlapped. Another interesting

features is that although there are many healthy quarters (4.6:1 healthy to sick quarter ratio), the region of healthy data denoted by back in the plots are much more compact than that for sick data. This was also observed by Wang and Samarasinghe (2005). This consistent observation indicates that a stable region marked by healthy values exists and an infection makes these vales to change drastically beyond the healthy region.

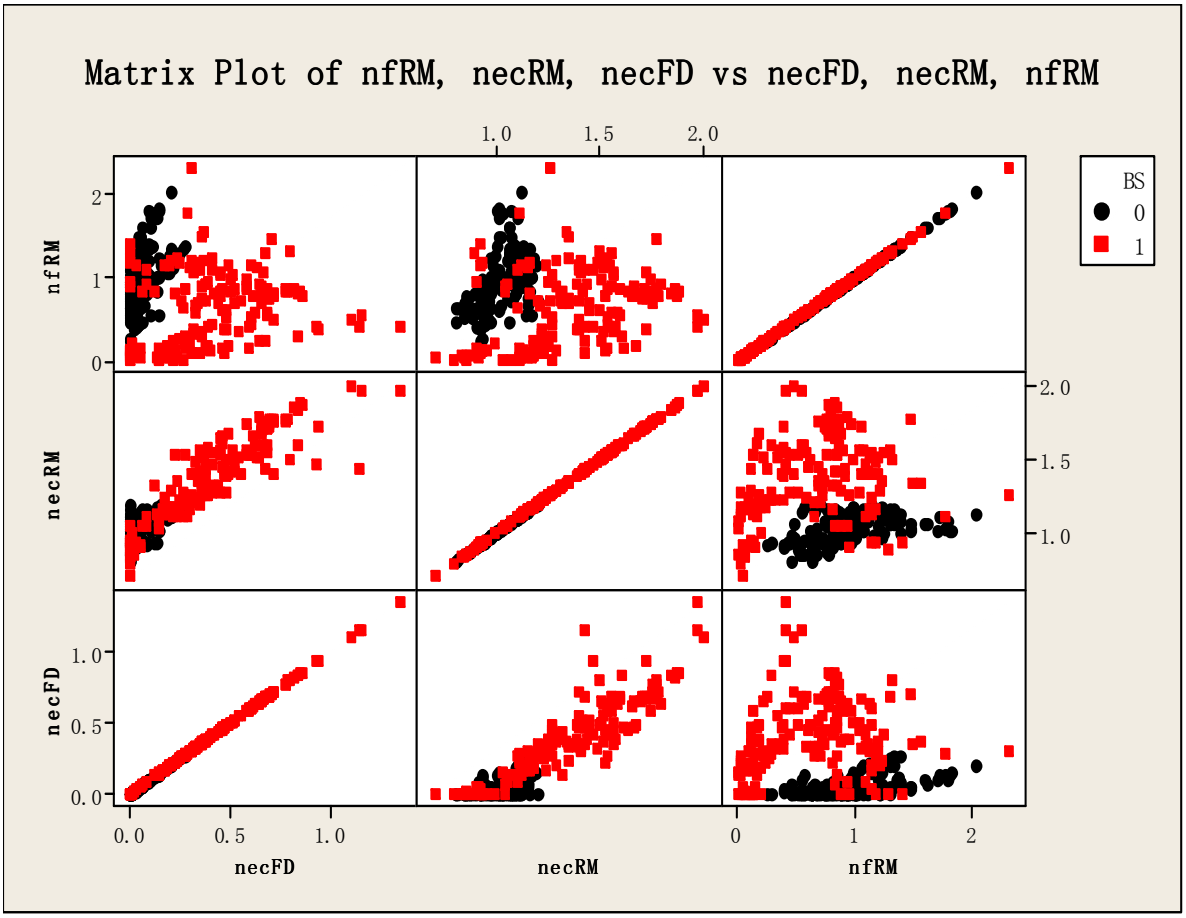


Figure 4.1 Correlation scatter plots of input variables. nfRM = Running Means of Normalized Quarter-yield Fraction; necRM = Running Means of Normalized Electrical Conductivity. necFD = The fractional deviations from the smallest necRM value. BS indicates mastitis State where 0 denotes healthy and 1 denotes sick.

4.2.4 Correlation between Variables

To measure the strength of relationship between variables, the coefficient of correlation was adopted. It indicates the linear relationship between two variables. When r gets closer to ± 1, the linear relationship between the two variables is stronger. When r is near 0, little or no linear relationship exists and it was calculated using equation below.

$$r = \frac{\sum_{i=1}^{N}(\chi_{1i} - \overline{\chi}_1)(\chi_{2i} - \overline{\chi}_2)}{\sqrt{\sum_{i=1}^{N}(\chi_{1i} - \overline{\chi}_1)^2 \sum_{i=1}^{N}(\chi_{2i} - \overline{\chi}_2)^2}} \qquad [4.3]$$

where r is the linear correlation coefficient, χ_{1i} is the ith value of the first variable, $\overline{\chi}_1$ is the mean of first variable, χ_{2i} is the ith value of the second variable, $\overline{\chi}_2$ is the mean of second variable, and N is the number of observations.

A correlation matrix was created to depict correlation between all variables and is presented in Table 4.1. It is symmetric with the diagonal values representing the correlation of a variable to itself. Off-diagonal values are the correlations between pairs of variables denoted by the labels indicated in the first row and column. Table 4.1 reveals that necRM and necFD are strongly and positively correlated at 0.869, whereas nyfRM has negative correlation with both necRM and necFD (-0.369 and -0.371, respectively). Furthermore, the necRM

Table 4.1 Correlation Matrix for the Input and Output Variables

	nyfRM	necRM	necFD	Mastitis Status
nyfRM	**1**	**-0.369**	**-0.371**	**-0.438**
necRM	**-0.369**	**1**	**0.869**	**0.580**
necFD	**-0.371**	**0.869**	**1**	**0.531**
Mastitis Status	**-0.438**	**0.580**	**0.531**	**1**

and necFD have a positive relationship with mastitis status (r = 0.58, r = 0.531), and nyfRM has a negative relationship with mastitis status (r = -0.438).

Based on data analysis it can be found that necRM and necFD are strongly correlated. This strong relationship should be carefully considered when developing ANN models because correlated variables provide redundant input dimensions to the network causing the computation more complicated. Furthermore, correlated variables can cause the problem of collinearity that lead to training problems such as overfitting, high prediction variance and ill conditioning. To deal with these problems, one possible way is to employ only one variable as a representative from the highly correlated variables. In this study, therefore, two combined data sets with different input variables were generated and used to train the models. The details are presented in the section 4.3.1. Another approach often used to solve problem of collinearity is Principle Component Analysis (PCA), which is discussed in the next section.

4.2.5 Principle Component Analysis (PCA)

To remove collinearity, PCA was carried out for predictive variables in classification of healthy and infected patterns. The main advantage of PCA is to reduce the number of dimensions, without much loss of information. This is achieved by transforming original variables into a new set of uncorrelated variables, which are ordered in terms of significance. Therefore, the first few principle components (PC) capture most of the variation present in all of the original variables.

Mathematically, the PCA can be represented by

$$\mathbf{COV\ x = y\ x} \qquad [4.4]$$

where **COV** is a covariance matrix of standardized original variables; **y** is scalar multiple of vector **x;** if the above equation holds true, then **x** is said to be the eigenvector of **COV**, representing PC and **y** is said to be the eigenvalue of **COV,** representing variance of the PC. By ordering Eigenvalue in descending manner, the first PC represented by the first eigenvector account for the largest amount of variation in the original data, each subsequent PC captures the largest amount of remaining variance, and so on.

The eigenvectors, eigenvalues, proportion of variance explained by each PC, cumulative percentages of variance were analyzed and the results are illustrated in Table 4.2 and Table 4.3. It can be observed from Table 4.2 that PC1 has the largest eigenvalue and account for 70% of total variance in the

Table 4.2 Results of Eigenvalue, Proportion and Cumulative Percentages of Variance

	PC1	PC2	PC3
Eigenvalue	**2.1147**	**0.7542**	**0.1311**
Proportion	**0.705**	**0.251**	**0.044**
Cumulative	**0.705**	**0.956**	**1.000**

Table 4.3 Eigenvector Matrix and Their Loadings Extracted from the COV Matrix of the Standardized Variables

Variable	PC1	PC2	PC3
nyfRM	**-0.425**	**0.905**	**0.001**
necRM	**0.640**	**0.301**	**-0.707**
necFD	**0.640**	**0.301**	**0.717**

data. The PC2 account for approximately 25% total variance. The total variance accounted for by the first two components is 96%.

According to Table 4.3, which is eigenvector matrix extracted from the COV matrix of the standardized variables, the PC1 strongly features both necRM and necFD (0.640 and 0.640 respectively) which suggests that these two variables are correlated as was found earlier. The PC2 strongly features nyfRM, indicating that it correlates much less with the other variables. The PC3 again

strongly features necRM and necFD and captures the remainder (4.4%) variance.

Because the purpose of PCA performed in the current study was for addressing the problem of collinearity, not for dimension reduction, all three PC were used. To obtain the transformed variables, the original mean-standardized variables are transposed (i.e. the data items were in each column, with each row holding a separate dimension) and then multiplied on the right by the transposed eigenvector matrix. This new PCA-transformed variables were uncorrelated and therefore not affected by the problem of collinearity.

4.3 Models Development

4.3.1 Development of Multilayer Perceptrons (MLP)

In order to perform the analysis, the feature of health status in original dataset was expanded into two new features (variables): state_sick and state_healthy, representing health states used as target output. For the infected quarters state_sick and state_healthy was set to 1 and 0 respectively. For the healthy quarters state_sick was 0 and state_healthy was 1. Four input-output pattern vectors extracted from the dataset with state_sick and state_healthy are displayed as an example of the data in Table 4.4

The reason why we had this feature expanded in this way rather than simply set

Table 4.4 A Sample of Four Records from the Dataset for Health States and Related variables (LB&RF = sick quarters; LF&RB = healthy quarters).

Quarter	nfRM	necRM	necFD	state_sick*	state_healthy*
LB	0.716559857	1.200276676	0.261971563	1	0
LF	1.287413784	0.88574183	0	0	1
RB	0.629766045	0.899254695	0.008375033	0	1
RF	1.390699242	1.001854109	0.045999956	1	0

***state_sick & state_healthy: Two new variables used as target output.**

infected quarters to 1 and healthy quarters to 0 is because neural network is made primarily for ordered sequences of data for each feature. If we assign sick quarters with 1 and healthy quarters with 0, for example, the in-between values (like 0.1-0.9) have no meaning (neither representing sick cow nor health cows) and it will just make things difficult for the neural network to make sense of it. Therefore, feature expansion was applied in this case as there were two separate classes: health and mastitis, that has no an implicit order.

As it was found by data analysis in section 4.2.4 that the problem of collinearity existed between variables, two data sets were generated according to combinations of input variables (see dataset 1 and 2 in Table 4.5). The original data set (dataset 3) was used as well so that it could be explored that whether or not the multi- collinearity had any effect on model performance.

Table 4.5 Datasets with Different Input Variables for Supervised Neural Networks

MLP Model	Dataset	Input Variables	Output variables
1	**1**	**nyfRM; necRM**	**state_sick; state_healthy**
2	**2**	**nyfRM; necFD**	**state_sick; state_healthy**
3	**3**	**nyfRM; necRM; necFD**	**state_sick; state_healthy**
4	**4**	**PC1; PC2; PC3**	**state_sick; state_healthy**

In addition, a new dataset with three PC variables discussed in Section 4.2.5 was generated as well so that the performance of non-PCA and PCA-based models could be evaluated in terms of prediction accuracy. The best model was retained.

For each dataset in Table 4.5, the data was divided into training and validation subsets. The training set was used for training the neural network to develop the correlation between input and output variables and it contained 70% of patterns from both side of healthy and sick quarters. The rest of the patterns (30%) were assigned to validation set that was used to test the generalization of the system. i.e. how well it worked on data it has not been trained on.

All modeling was performed on the SYNAPSE software (Peltarion Synapse Version 1.25, 2006). One of the advantages of Synapse is that, by using a

Genetic Algorithm Optimizer, it automatically searches for optimal parameters (e.g. numbers of hidden neurons in the hidden layer, learning rate and momentum) that need to be estimated. Once the optimal parameters have been found, the optimum network is trained in the normal way for determining the (weights) coefficients of the network. .

Three-layer MLP with back-propagation was used for developing predication models. As mentioned in Chapter 2, the MLP processes information in a forward manner through the network while the prediction error is propagated backwards through the network. The input and output variables for each model were as detailed in Table 4.5. The performance accuracy was evaluated by the sensitivity, specificity and overall correct classification rate (CCR). In addition, a traditional statistical classifier, Linear Discriminant Analysis (LDA), was selected in the current study to contrast traditional statistical classifier with the ANN model. The same datasets used to train the ANN models were used for LDA.

4.3.2 Development of Self – Organizing Maps (SOM)

The data used for the development of the SOM models was the data set 3 in Table 4.5, but the output variables were not involved as SOM is a kind of unsupervised networks. The purpose of developing SOM is to cluster the health status into three categories (i.e., health, moderately ill and severely ill).

Thus, the network has three outputs. A SOM 15 x 15 with 225 neurons was trained and the results of the SOM were evaluated by using three statistical techniques: K-means clustering, ANOVA and Least Significant Difference.

Chapter 5

5.1 Results and Discussion of SOM

5.1.1 Results of SOM

A two dimensional SOM with 225 neurons in the output layer was trained to cluster the health status into three categories. The input variables were nyfRM, necRM and necFD. Thus there were three neurons in the input layer. The initial learning rate was 0.5 and Gaussian function was adopted as neighbour strength function.

The health categories clustered by the SOM are shown in top-left panel in the Figure 5.1, in which all the data patterns were well clustered into three groups representing different health states. The red colour represents severely ill, green colour represents moderately ill and blue colour represents healthy. The black circles inside the unit cells indicate how many data patterns are close to a particular neuron. Larger circle means more data. The rest of the panels are the individual input variables. The red colour indicates a higher value, and the blue represents a lower value.

Figure 5.1 shows that as the necFD and necRM increase, the health status of a quarter deteriorates gradually from healthy to severely ill. It is also revealed that with healthy state getting worse, the nyfRM decrease as well. For

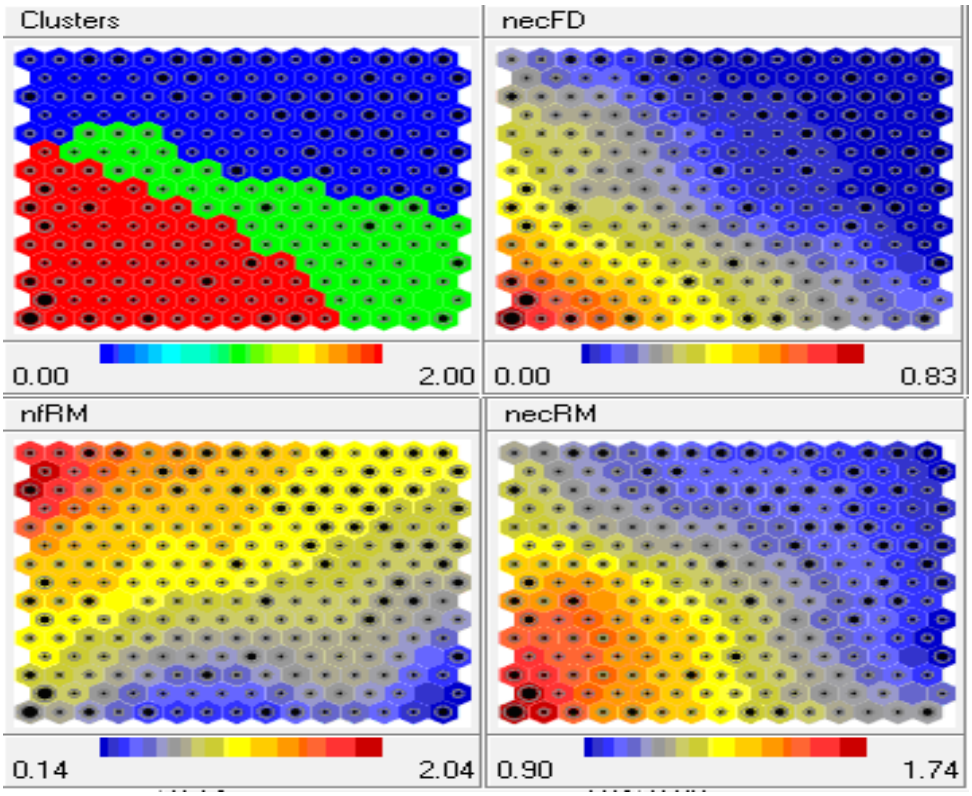


Figure 5.1 Mapping of 3 Dimensional data onto a two-dimensional SOM. The top-left panel shows the health states. Red = Severely Ill; Green = moderately Ill; Blue = Healthy. The other three panels present the input variables.

example, from the cluster panel and nyfRM panel it can be observed that all the infected quarters including moderately ill and severely ill have lower values of nyfRM than the healthy quarters. As discussed in the previous section, usually, an infected quarter has the higher than normal electrical conductivity and produces less milk. The SOM has well detected this trend.

Furthermore, Table 4.1 in Section 4.2.4 revealed that electrical conductivity (necFD and necRM) was more strongly correlated with health state than milk yield (nyfRM) did indicating that electrical conductivity plays a more leading

role in detecting mastitis status. Principal component analysis corroborated this evidence and quantified it by demonstrating that the very first PC was almost exclusively made of necFD and necRM capturing the largest amount (70%) of variation in data. The SOM in Figure 5.1 confirms this evidence by highlighting the fact that the 3 health states (first panel) are demarcated strongly according to the levels of necFD and necRM (top right and bottom right panels). However, the milk yield plays a meaningful role as shown by the bottom left panel in Figure 5.1. It shows without ambiguity that the yield drops progressively from healthy to marginally ill and drops further in the severely ill case. Owing to the high correlation between necFD and necRM, the panels representing these two (top and bottom right) show similar patterns of variation.

The SOM in Figure 5.1 further shed light on the distribution of cows in the spectrum of the three health states of healthy, moderately ill and severely ill. As stated previously, the black circles inside the neurons depicts the number of cows (i.e. quarters) belonging to the neuron. The larger the circle, the larger number of quarters represented by that neuron. On this basis, the cluster panel (top left) in Figure 5.1 shows that most of the healthy cows (blue cluster) are very healthy and a large proportion of the sick cows (red cluster) are very ill. This is because most of the larger black circles in the blue cluster are located towards the top right area away from the moderately ill (green) cluster and

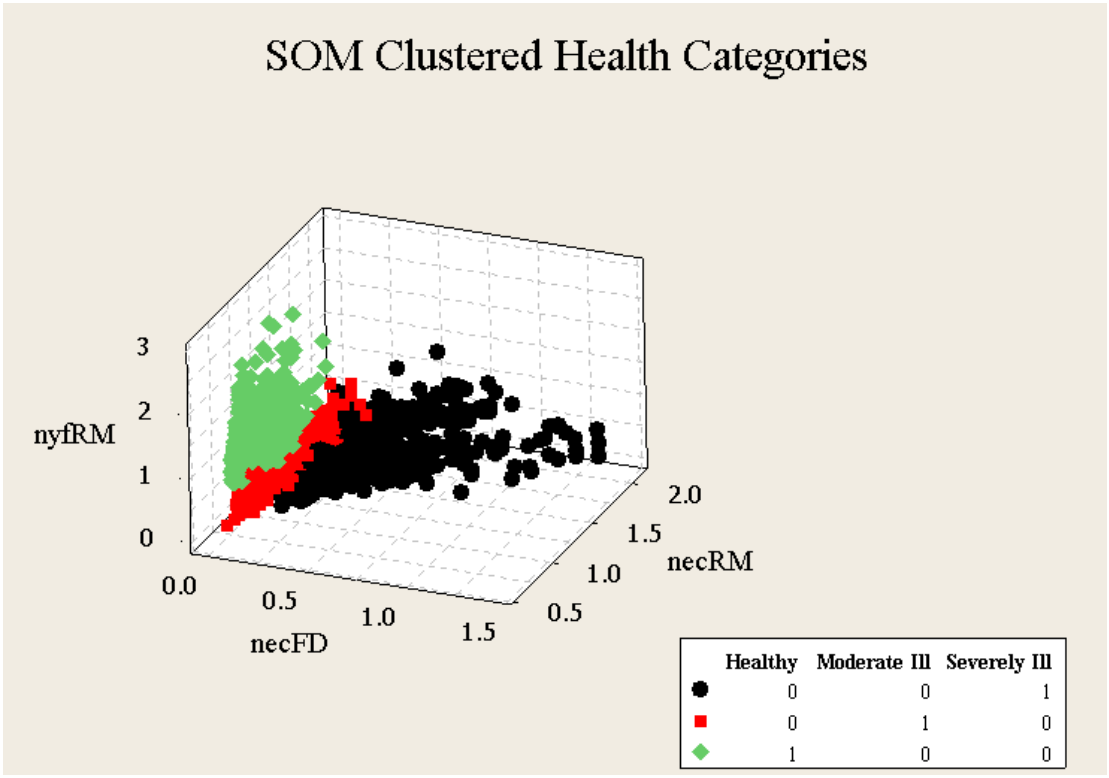


Figure 5.2 SOM Clustered Health Categories in 3-D Format

most of the larger circles in the red cluster are located in the bottom left area away from the moderately ill cluster.

Figure 5.2 also shows the results of the clustering in 3-D format that clearly illustrates the structure of the spatially organised data in the SOM network. It reveals meaningful cluster structures where healthy quarters (green cluster) all have high yield (nyfRM) and severely ill quarters (black cluster) all have very high conductivity (necFD and necRM). The moderately ill (red) cluster has in-between values for these variables.

5.1.2 Evaluation of SOM

For each cluster the mean values and the standard deviations of the three variables were analysed using One Way ANOVA, and Least Significant Difference (LSD) was used to test if the mean differences between the health categories were statistical significant. Table 5.1 shows the summary of the analyses and Table 5.2, which was extracted from SPSS software, illustrates the results of the LSD test. From these, it can be observed that all the health categories are statistically significant based on the mean values of the three input variables.

In terms of specifics, Table 5.1 presents the results of analysis of the three SOM based health clusters using ANOVA. It shows that all three variables are

Table 5.1 Mean and Standard Deviation of the Variables for Each Health Categories and the Statistical Significance of the Means between the Categories.

	Health Category			LSD(p<0.05)
Variable	**H0***	**H1***	**H2***	
nyfRM	1.12± 0.27	0.68 ± 0.22	0.46± 0.31	All
necRM	0.95 ± 0.91	1.03 ± 0.11	1.41 ± 0.19	All
necFD	0.04 ± 0.05	0.05 ± 0.08	0.43 ± 0.21	All

***H0: Healthy**
***H1: Moderately ill**
***H2: Severely ill**

statistically significant for the clusters ($p < 0.05$). It also indicates the interesting outcome that the yield drops sharply as healthy quarters deteriorates into moderately ill and it decreases much less dramatically from moderately ill to severely ill. This might indicate that the yield is quite sensitive to infection and is affected early on in the disease. The two conductivity related variables, in contrast, show a marked increase in the transition between marginally ill and severely ill states confirming the known biological evidence that electrical conductivity increase happens due to the breaking of blood-milk barrier at later stages of an infection. Of these two variables, necFD shows the most marked increase.

Table 5.2 Results of LSD Test. (1.00 stands for healthy, 2.00 for moderately ill and 3.00 for severely ill)

Statistical Significance of the Means between the Categories

LSD

Dependent Variable	(I) SOM_Clusters	(J) SOM_Clusters	Mean Difference (I-J)	Std. Error	Sig.	95% Confidence Interval	
						Lower Bound	Upper Bound
nyfRM	1.00	2.00	.44047*	.01018	.000	.4205	.4604
		3.00	.65653*	.01032	.000	.6363	.6768
	2.00	1.00	-.44047*	.01018	.000	-.4604	-.4205
		3.00	.21606*	.01174	.000	.1930	.2391
	3.00	1.00	-.65653*	.01032	.000	-.6768	-.6363
		2.00	-.21606*	.01174	.000	-.2391	-.1930
necRM	1.00	2.00	-.07731*	.00477	.000	-.0867	-.0680
		3.00	-.45169*	.00484	.000	-.4612	-.4422
	2.00	1.00	.07731*	.00477	.000	.0680	.0867
		3.00	-.37438*	.00550	.000	-.3852	-.3636
	3.00	1.00	.45169*	.00484	.000	.4422	.4612
		2.00	.37438*	.00550	.000	.3636	.3852
necFD	1.00	2.00	-.01791*	.00445	.000	-.0266	-.0092
		3.00	-.38695*	.00452	.000	-.3958	-.3781
	2.00	1.00	.01791*	.00445	.000	.0092	.0266
		3.00	-.36904*	.00514	.000	-.3791	-.3590
	3.00	1.00	.38695*	.00452	.000	.3781	.3958
		2.00	.36904*	.00514	.000	.3590	.3791

*. The mean difference is significant at the .05 level.

LSD test results presented in Table 5.2, where all possible mean differences are tested for significance, indicate that all mean differences are significant at the 0.05 level. None of the confidence intervals for the mean difference contains zero further highlighting that the two corresponding means belong to two different populations.

To assess if the clustering results were acceptable, k-means clustering was performed. The results obtained from the k-means clustering were compared to the results found by SOM. Table 5.3 (extracted from SPSS) shows the correlations between clusters obtained from SOM and K-means, respectively. It can be observed that the correlation between the clustering results from these two methods is 0.82 ($P<0.01$) which indicates that the results from the SOM are quite reasonable. Figure 5.3 shows the results of the K-means clustering in 3-D format. It can be seen clearly that it has a structure similar to that obtained from the result of SOM clustering. Considering the two methods however, k-means can be considered as a simpler linear version of SOM but without the neighbourhood preserving quality of SOM. These two aspects - ability to handle nonlinear cluster boundaries and preserving neighbourhood properties of the clusters give SOM extra advantage over k-means.

Main advantage of SOM over k-means is that in SOM, owing to the neighbourhood feature that it uses during training, input patterns that are

Table 5.3 Correlations between Clusters Obtained from SOM and K-means

Correlations

		SOM_ Clusters	K_ meansCluster
SOM_Clusters	Pearson Correlation	1	.819**
	Sig. (2-tailed)		.000
	N	4139	4139
K_meansCluster	Pearson Correlation	.819**	1
	Sig. (2-tailed)	.000	
	N	4139	4139

**. Correlation is significant at the 0.01 level (2-tailed).

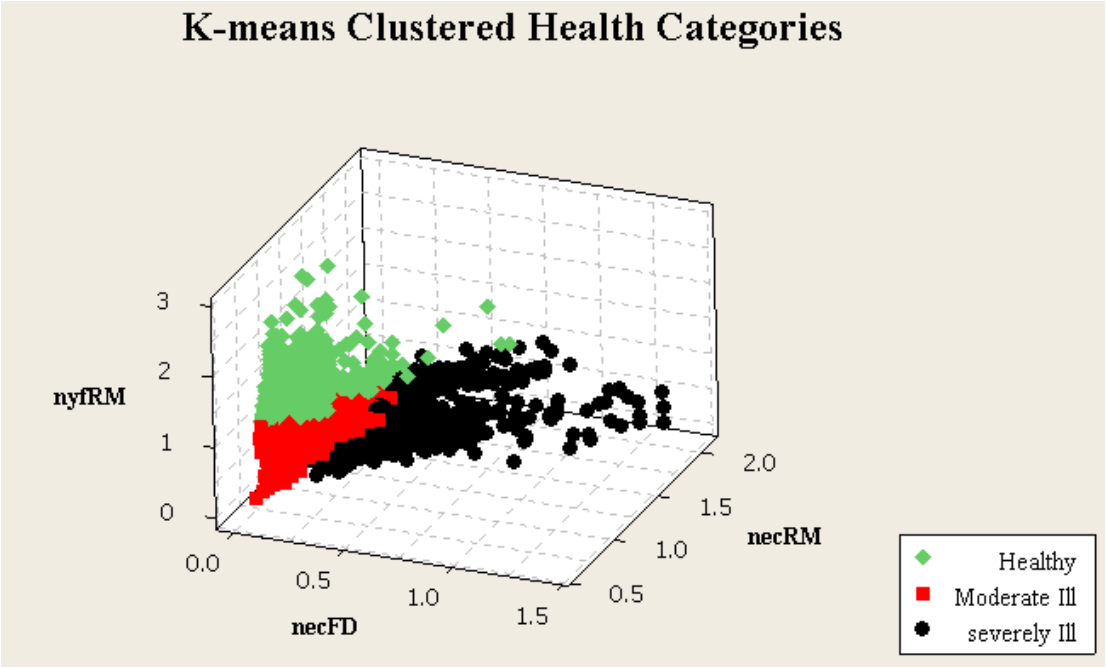


Figure 5.3 K-means Clustered Health Categories in 3-D Format

similar are spatially organized in close proximity to each other in such a way that cluster structures reveal spatial organization of the input patterns correctly. Furthermore, in SOM, clusters are found by clustering neurons of the trained map, i.e. after training. Thus, clusters can be more reliable. This is because clustering happens in 2 stages: in the first stage, input patterns are sorted

properly in the multi-dimensional space preserving the neighbourhood character and in the second stage, the clusters are formed on these spatially organized input patterns. The clustering of the map neurons were done in the Synapse by Ward clustering (Ward, 1963), a powerful statistical clustering method.

5.2 Results and Discussion of MLPs

5.2.1 Classifying Mastitis with MLP

Each MLP model was trained with the input and output variables depicted in Table 4.5. The training sessions were carried until the highest sensitivity on the validation dataset was achieved. For the hidden layer, the Tanh Sigmoid function was adopted and learning rate and number of neurons were automatically searched by using Genetic Optimizer function. Details of results on the optimal parameters in hidden layer for each model, validation mean square error (MES) and best results from each model in terms of sensitivity and specificity are given in Appendix 1. The best results from each model in predicting sick quarters in the validation datasets is illustrated in Fig.5.4, 5.5, 5.6, and 5.7. There are four lines in the plot: model output, target output (desired) and the higher and lower bounds of the confidence interval. The confidence interval was written on the top of the plot. It can be observed that the model that trained with PC as variables has a smaller confidence interval compared with others, which indicates that its prediction accuracy is slightly

higher than that of other models trained with a combination or all of the original variables.

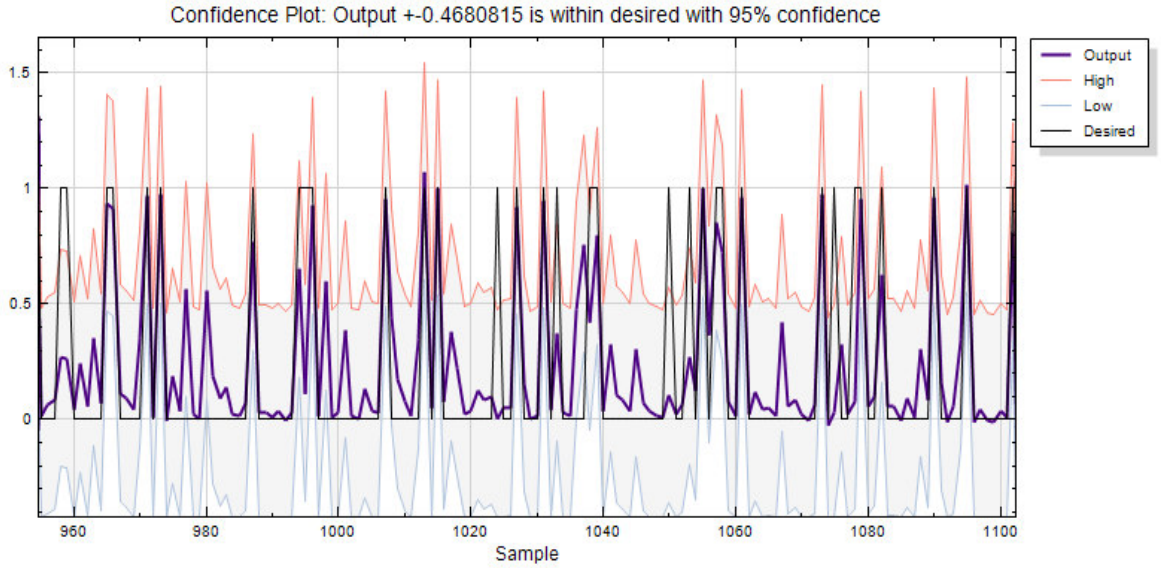


Figure 5.4 Prediction Performance of Model 1 (inputs: nyfRM, necRM).

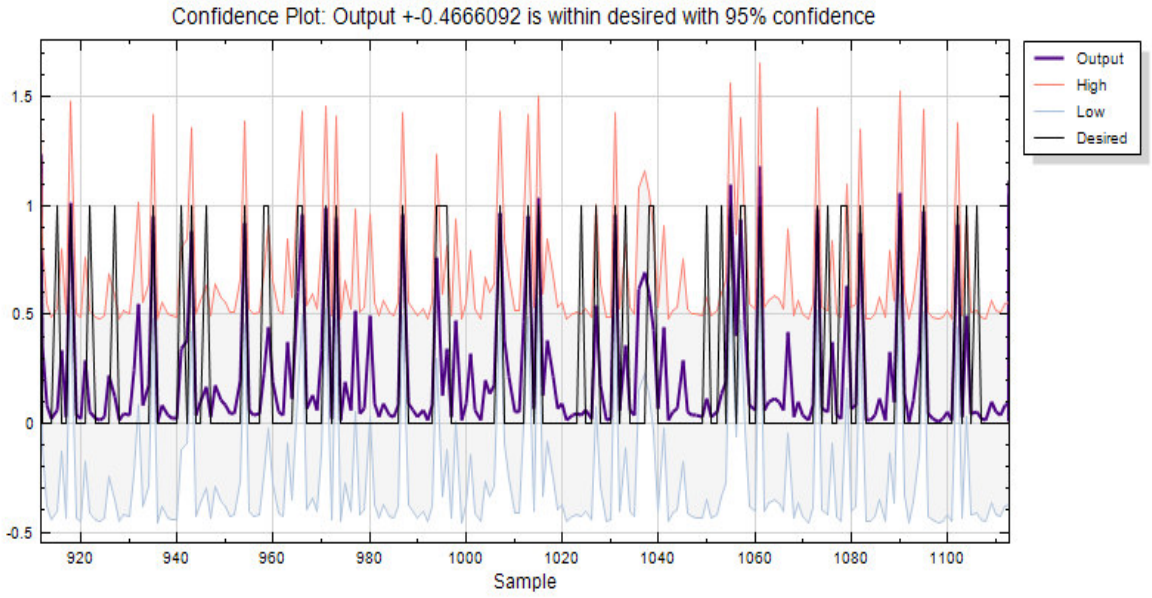


Figure 5.5 Prediction Performance of Model 2 (inputs: nyfRM and necFD).

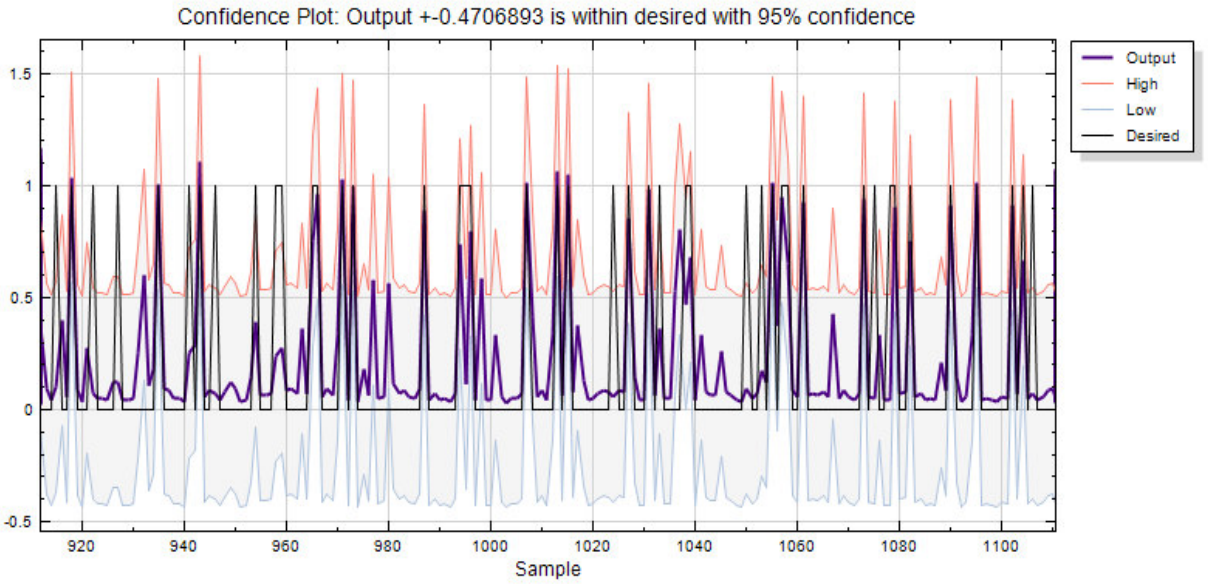


Figure 5.6 Prediction Performance of Model 3 (inputs: nyfRM, necRM, and necDV).

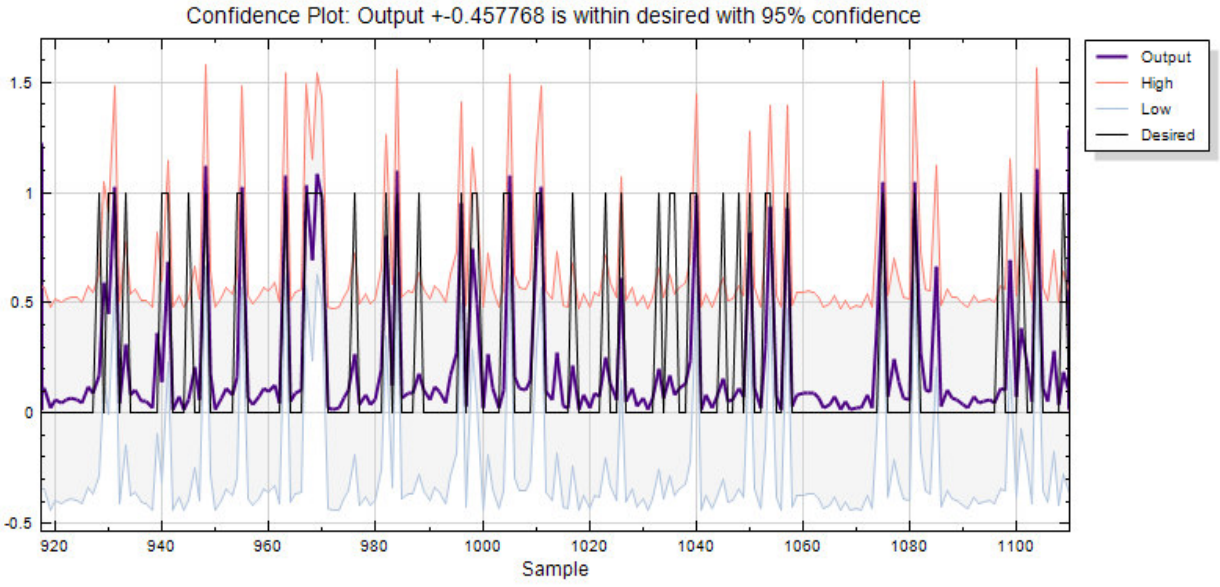


Figure 5.7 Prediction Performance of Model 4 (PCA-based: PC1, PC2, PC3).

To evaluate the predictive ability of each model, the sensitivity, specificity and correct classification rate (CCR) was employed as measurements. Table 5.4, which is summary of Appendix1, shows the optimal hidden neurons in hidden layer, the mean square error when best performance achieved, the specificity, sensitivity and overall correct classification rate on validation dataset from the four best models. The first observation that can be made is that PCA-based model has a better performance than non-PCA-based models. It classified the validation data set with an overall correct classification of 90.74%. Especially, the sensitivity of the model for correctly detecting infected cases is 86.9%, which is much higher than other non-PCA-based models. Although the specificity of model 1 is slightly higher than that of the PCA-based model, its overall CCR and sensitivity are worse than the PCA-based model. As it is of

Table 5.4 Predictive Abilities of the Four Best Models

Models	Input Variables	Neurons in Hidden Layer	Mean Square Error	Specificity (%)	Sensitivity (%)	(CCR)%*
1	nyfRM, necRM	12	0.170	92.00	81.23	89.46
2	nyfRM, necFD	7	0.171	90.00	83.15	89.77
3	nyfRM, necRM, necFD	10	0.173	91.31	78.93	87.21
4	PC1, PC2, PC3	6	0.169	91.36	86.90	90.74

*** Overall Correct Classification Rate**

great importance to correctly find cows with mastitis, the PCA-based model is more desirable than model 1. For non-PCA-based models, model 1 and model 2 has a similar prediction performance in overall CCR (89%), however, the specificity and sensitivity provided from these two models are different from each other. For model 1, the specificity and sensitivity are 92% and 81.23%, and for model 2 these are 90% and 83.15%, respectively. Model 2 with electrical conductivity deviation (necFD) seems to be slightly superior. The model 3, in which the overall CCR is 87.21%, is inferior to other models, especially; its sensitivity is much lower than others. Another observation is that the specificity is higher than sensitivity in the all models. This could be due to the different proportion of infected to non-infected cases in the training data. Because there was no other proportion investigated in the current study a firm conclusion about this finding can not be made. However, other researchers (Nielen et al, 1994; Yang et al. 1999) found out that higher proportion of healthy cases do increase the specificity of the predictions.

It also can be observed from Table 5.4 that the PCA-based model not only provided the best performance (CCR=90.74%) but also its architecture is less complex (i.e., less neuron numbers in hidden layer). It has the smallest validation mean square error (MSE) and it is reached with the smallest number of hidden neurons compared to the other 3 models. Therefore, this model has the least complexity, and therefore, the simplest structure. In the case of

non-PCA-based neural networks, it is seen that the model 1 with twelve hidden neurons is optimal and that the model 2 with the seven neurons in the hidden layer achieves the best predictive performance. For the model 3, 10 neurons in the hidden layer is optimal.

As it was found in data analysis in the previous section, all input variables were related to some extent, in particular, necRM and necFD were strongly correlated to each other and can cause the problem of collinearity. This is clearly highlighted by the inferior results of model 3 as shown in Table 5.4. The improvement achieved by PCA-based model proves that collinearity indeed exists within the input data and PCA is a suitable option to deal with this issue as the predictive performance can be significantly improved by PCA-based model.

In mastitis research, definitions of mastitis are usually defined based on SCC. However, in the current study mastitis was defined on the basis of two thresholds relating to higher EC and lower quarter yield. Owing to the relatively low threshold defined for quarter yield, it is possible that more ill quarters are involved, thus, the proportion of ill cases was high leading to a relatively low error rate i.e. the model gains more relevant knowledge about mastitis and in turn could more accurately detect infected quarters. In addition, on visual exploration of misclassified infected quarters in relation to whole data, it was found that the misclassified ones are those that are closer to or

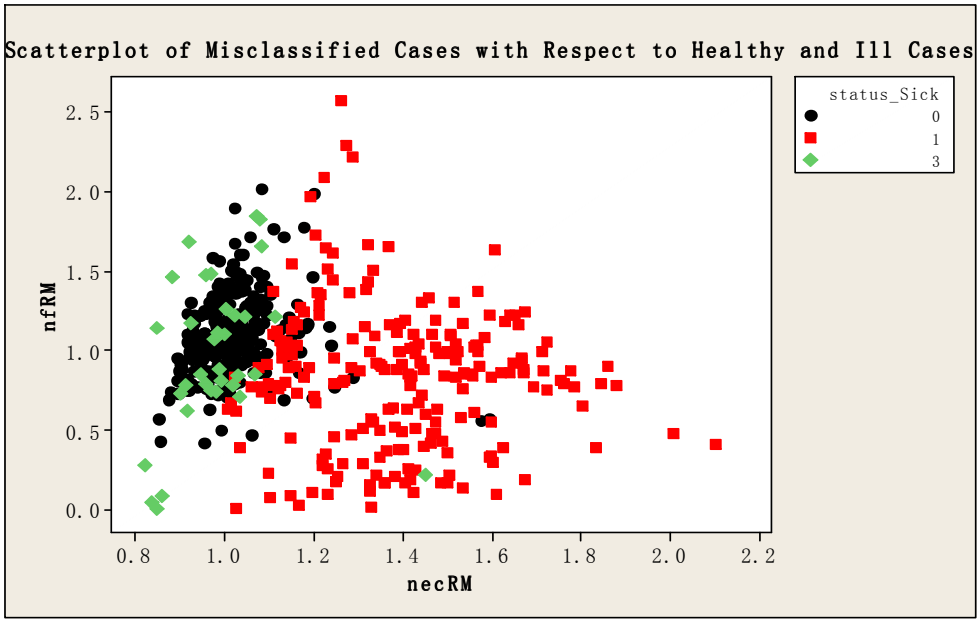


Figure 5.8 Relation of misclassified infected cases to healthy and ill cases for the validation data set. Black dots represent healthy cases. Red squares represent infected cases. Green diamond squares represent clinical cases that are wrongly detected as healthy cases.

within the healthy region. Figure 5.8, which is scatter plot of necRM and nyfRM from validation dataset of model 3, is presented here as an example to illustrate the relations of misclassified infected cases to healthy and ill cases. In Figure 5.8, the misclassified infected cases that are superimposed on the data are represented by green diamond squares, the black dots represent healthy cases, and red squares represent infected cases. It clearly shows that the most misclassified infected cases are those that have neither high conductivity (necRM) nor low milk yield (nyfRM). As discussed early, the EC values vary from cow to cow and not all clinical cows have very high EC or even have a reduced milk production on the ill quarter. For such cases, it is difficult or would be impossible for a model to detect them correctly.

5.2.2 Comparing MLP and LDA

LDA was adopted to contrast the traditional statistical method with ANN in the current study. To compare these two methods, the same datasets used to train MLPs were employed in LDA. Details of the result are given appendix 2 and Table 5.5 shows the classification results in terms of the specificity and sensitivity on the validation datasets from LDA models for the four datasets in Table 4.5

It can be observed that the PCA-based linear model, which has 88.7% overall CCR, is the best one in classifying the infected and non-infected quarters. The sensitivity (84.2%) is dramatically higher than other non-PCA- based models.

Table 5.5 Predictive Performance of LDA

Linear Models	Variables	Specificity (%)	Sensitivity (%)	Overall Correct Classification Rate (CCR)%
1	**nyfRM, necRM**	**89.9**	**77.2**	**87.1**
2	**nyfRM, necFD**	**89.6**	**79.6**	**87.5**
3	**nyfRM, necRM, necFD**	**89.5**	**76.3**	**86.7**
4	**PC1, PC2, PC3**	**89.9**	**84.2**	**88.7**

Among the non-PCA-based linear models, the model 2 has higher predictive performance in correctly classifying mastitis cases (79.6%) and its overall CCR (87.5%) is the highest as well compared to other non-PCA-based models. However, the specificity of model 2 is slightly lower than that of model 1, of which the specificity is 89.9%. The model 3, which was trained with nyfRM, necRM and necFD, has the lowest overall CCR, specificity and sensitivities (86.7%, 76.3% and 89.5%, respectively).

By comparing Table 5.4 and Table 5.5, several observations can be made: firstly, the fact that both PCA-based models from two methods have the best predictive performances, and that both models using nyfRM, necRM and necFD as input variables have the worse predictive performance, again emphasizing the issue of multi-collinearity within data. When inputs are highly correlated, there can be over compensation due to redundancy. This can lead to model overfitting, low predictive capability, less robustness and high variance in the predictions (Samarasinghe, 2006). As for robustness, when variable are correlated, the redundancy means that there can be more than one model configuration that suit the data due to overcompensation of one variable over the other. This leads to non-unique model parameters and less robust or unstable outputs. However, when the variables are independent in the model, it ensures uniqueness of the model configuration thereby ensuring the uniqueness of model parameters and enhancing stability and robustness of the

model outcomes.

Another observation from Table 5.4 and 5.5 is that the ANN methods perform better compared with the LDA for all the datasets. For example, in the dataset with PC as input variables, 86.9% infected cows were correctly classified by PCA-based ANN model compared with 84.2 % by PCA-based linear model. In the dataset with three variables, 78.93% infected cows were correctly classified by ANN model compared with 76.3% of correctly classified by LDA. The ANN and the LDA have similarities and dissimilarities in classification. Both are methods that minimize the error between the actual and desired outputs. However, for LDA, certain assumptions about the input parameters are usually required and it is based on linear combination of inputs. In contrast, ANN do not make assumptions about the data, incorporate nonlinear interactions and have the capability to learn from the input data to produce an optimal output within a changing data environment. In addition, LDA cannot draw multiple partitions. Only one partition in a sample is possible. Also, it cannot draw a nonlinear partition. In contrast, a neural network is capable of drawing any number and types of partitions as long as a sufficient number of hidden neurons are provided (Lippmann, 1987).

Due to the difference in mastitis definition and data properties, it is complicated to compare the model performance with other studies. However,

by comparing models performed in this particular study it has been clearly shown that the performance of the neural network can be improved by using three principal components as neural network inputs. PCA-based model is superior to other models in many respects such as less complexity, higher predictive accuracy, and also in terms of addressing the problem of collinearity.

Chapter 6

6.1 Conclusions

In this study, the self organizing map (SOM) and multilayer perceptron (MLP) were developed for mastitis detection using the preprocessed data relating to the electrical conductivity and milk yield. Also, the LDAs were performed on each dataset developed for ANN models to compare with the ANN in predictive performance. The Principle Components Analysis technique was adopted for addressing the problem of multi-colinearity existed in the data. A new mastitis definition based on higher EC and lower quarter yield was created to distinguish between the infected and non-infected quarters. Based on this new definition, the PCA-based MLP model manifested to be superior to other non-PCA-based models. The overall correct classification rate (CCR), sensitivity and specificity of the model was 90.74 %, 86.90 and 91.36, respectively.

The other 2 models, one involving yield and conductivity as inputs and the other with yield and fractional deviation of conductivity had lower prediction accuracy than the PCA-based model but were still reasonably high at 89.36% and 89.47 for overall CCR, respectively. The last model with all 3 input variables had lower performance than the above 3 indicating the undesirable influence of multi-correlinearity among variables, in this

case, the correlation between conductivity and its fractional deviation. With such high accuracy, the models such as the PCA-based model developed here can improve the accuracy of prediction of mastitis by robotic milking stations.

The results of comparison between the two methods of the ANN and LDA indicate that the ANN is superior to LDA for all the datasets. The advantage of using ANN over LDA for classifying problems is that ANN can learn to improve performance while employing nonlinear capabilities to find multiple clusters.

The SOM was developed to classify the health status into three categories: healthy, moderately ill and severely ill. These categories were meaningful and clear in terms of their regions of spread and the mean of the clusters. The clustering results were successfully evaluated and validated by using statistical techniques such as K-means clustering, ANOVA and Least Significant Difference. Results indicated that yield drop is prominent in the early stages and conductivity increase is dominant in the later stages of an infection. It can be concluded that the SOM can be employed by a robotic milking station as a detection model for mastitis.

Due to the limited number of mastitis indicators, the results of this study

may not be optimal. Therefore, in the future research more informative milk traits related to mastitis should be added so that the detection model would be improved and optimized.

Reference

Auldist, M.J., Hubble, I.B., (1998). Effects of mastitis on raw milk and dairy products. *The Australian Journal of Dairy Technology*, 53: p. 28-36.

Aoki, Y., Notsuki, I. & Ichikawa, T. (1992). Variation in patterns of mastitis indicators during milking in relation to infection status (summary in English). Anim. Sci. Technol. (Jpn.), 63: 728-735.

Barth, K., Fishcer, R., & Worstorff, H. (2000). Evaluation of Variation in Conductivity During Milking to Detect Subclinical Mastitis in Cows Milked by Robotic Systems, pp: 89-96, Robotic Milking: Proceedings of the International Symposium, Wageningen Press, The Netherlands.

Batra, T. R., & McAllister, A. J. (1984). A comparison of mastitis detection methods in dairy cattle. Can. J. Anim. Sci. 64305.

Bentley, R., & Lacy-Hulbert, J. (2007). Minimizing Mastitis. *Dairy Exporter Great Farming Guides Series*, 32, 9

Blowey, R.W. (1986). An assessment of the economic benefits of a mastitis control scheme. Veterinary Record 119.551-553,.

Booth, J. M. (1988). Control measures in England and Wales. How have they influenced incidence and aetiology. British Vet. J. 144(4): 316-322

De Mol, R.M., Ouweltjes, W. (2001). Detection model for mastitis in cows milked in an automatic milking system. Prev. Vet. Med. 49, 71-82.

Fernando, R. S., Rindsig, R. B., & Spahr, S. L. (1982). Electrical conductivity of milk for detection of mastitis. J. Dairy Sci. 65659.

Grennstam, N. (2005) *On Predicting Milk Yield and Detection of Ill Cows.* Retrieved July 23, 2007, from http://www.ee.kth.se/php/modules/publications/reports/2005/IR-RT-EX-0519.pdf

Heuven, H. C. M., Bovenhuis, H., & Politiek, D. (1988). Inheritance of somatic cell count and its genetic relationship with milk yield in different parities. Livest. Prod. Sci. 18(2): 115-127.

Hillerton, J. E., & Walton, A. W.(1991). Identification ofsubclinical mastitis with a handheld electrical conductivitymeter. Vet. Rec. 128(22): 513-515.

John, B., Hamman, H., & Frank, H. D., (1992). Insight Books 1992. Milking and Lactation, editor A.

Lake, J.R., Hillerton, J.E., Ambler, B. & Wheeler, H.C., (1992). Trials of a novel mastitis sensor onexperimentally infected cows. J. Dairy Res., 59: 1 l-l 9.

Leslie,K.E., Dohoo, I. R., &.Meek, A. H. (1983). Somatic cell counts in bovine milk. Comp.Cont.Ed. 5: s601.

Lippmann, R. P. (1987). An introduction to computing with neural nets. IEEE ASSP Magazine. 4:4–22

Lmsbergen, L. M. T. E., Nielm, M., Lam, T. J. G., Pmngov, M. A., Schukken, Y. H., & Maatje, K. (1994). Evaluation of a prototype on-line electrical conductivity system for detection of subclinical mastitis. J. Dairy Sci. 77: 1 132.

López-Benavides, M. G., Samarasinghe, S., & Hickford, J. G. H. (2003). the use of artificial neural networks to diagnose masitits in dairy cattle: Lincoln University. Canterbury. New Zealand. Unpublished.

Mein, GA, Sherlock, RA., & Claycomb, RW. (2004). Making Sense of In-Line Sensing for Milk Conductivity. pp. 252-53. Automatic milking - a better understanding: Proceedings of the International Symposium, Wageningen Academic Publishers, The Netherlands.

Mele, M., Secchiari, P., Serra, A., Ferruzzi, G., Paoletti, F., Biagioni, M. (2001). Application fo the 'tracking signal' method to the monitoring of udder health and oestrus in dairy cows. Livest. Prod. Sci. 72, 279-284

McDermott. M. P., Erb, H. N., & Natzke, R. P. (1982). Predictability by somatic cell counts related to prevalence of intnmmary infection within herds. J. Dairy Sci. 65: 1535.

Miles, H., W. Lesser, and P. Sears. (1992). The economic implications of bioengineered mastitis control. J. Dairy Sci.75 (2): 596-605

Monardes, H. G. (1994). Somatic cell counting and genetic improvement of resistance to mastitis. In Proc. XXXI Annual Meeting Brazilian Soc. Anim. Prod., Maringa, Parana, Brazil. Parana, Brazil: State University

of Maringa.

Nielen, M., Spigt, H. M., Schukken, Y. H., Deluyker, H. A., Maatje, K., & Brand, A. (1994). Application of a neural network to analyse on-line milking parlour data for the detection of clinical mastitis in dairy cows.

Nielen, M., Schukken, Y. H., Brand, A., Haring, S., & Ferwerdavanzonneveld, R. T. (1995). Comparison of Analysis Techniques for online Detection of Clinical Mastitis. *Journal of Dairy Science,* 78(5), 1050-1061.

NMAC. (2001). Managing Mastitis – A practical guide for New Zealand dairy farmers (3rd). Hamilton, New Zealand: Livestock Improvement.

Norberg, E., Hogeveen, H., Korsgaard, I. R., Friggens, N. C., Sloth, K., & Lovendahl, P. (2004). Electrical conductivity of milk: Ability to predict mastitis status. Journal of Dairy Science, 87(4), 1099-1107.

Rindsig, R. B., Rodewald, R. G., Smith, A. R., Thornsen, N. K., & Spahr, S. L.(1979). *Mastitis history. California mastitis test. and somatic cell counts tor identiQing cows for treatment in a selective dry cow rhenpy prognm*. J. Dairy Sci. 62: 1335.

Samarasinghe, S., (2006). *Neural networks for applied sciences and engineering: from fundamentals to complex pattern recognition.* Boca Raton,FL: Auerbach

Schultz, L. H. (1977). *Somatic cell counting of milk in production testing*

progmms as a mastitis control technique. J. Am. Vet. Med. Assoc. 170: 1244.

Sharif, S., Mallard, B.A., Wilkie, B.N., Sargeant, J. M., Scott, H. M., Dekkers, J.C.M., et al.(1998). Associations of the bovine major histocompatibility complex DRB3(BoLA-DRB3) alleles with occurrence of disease and milk somatic cell score in Canadian dairy cattle. *Animal Genetice*, 29(3), 185-193.

Sheldrake, R.F., McGregor, G. D., & Hoare, R. J. T. (1983). *Somatic cell count, Electrical conductivity and serum albumin concentration for detecting bovine mastitis.* J. Dairy Sci. 66:548.

Wang, E., & Samarasinghe, S. (2005). *On-line detection of mastitis in dairy herds using artificial neural networks.* Paper presented at the International Congress on Modeling and Simulation, Melbourne, Australia.

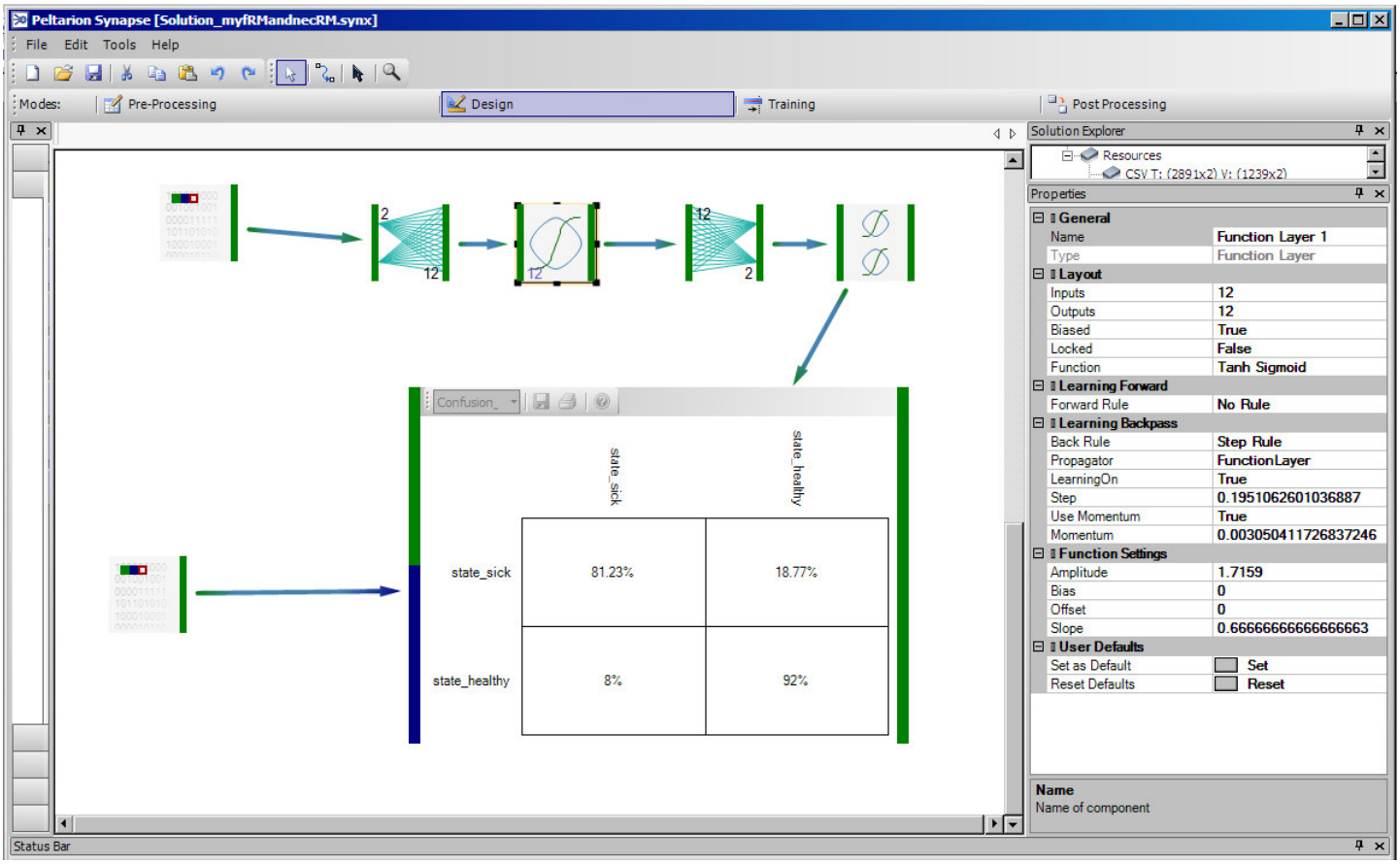
Ward, J. h. Jr., (1963). Hierarchical trouping to optimize an objective function, *Journal of the American Statistical Association,* 58, 263.

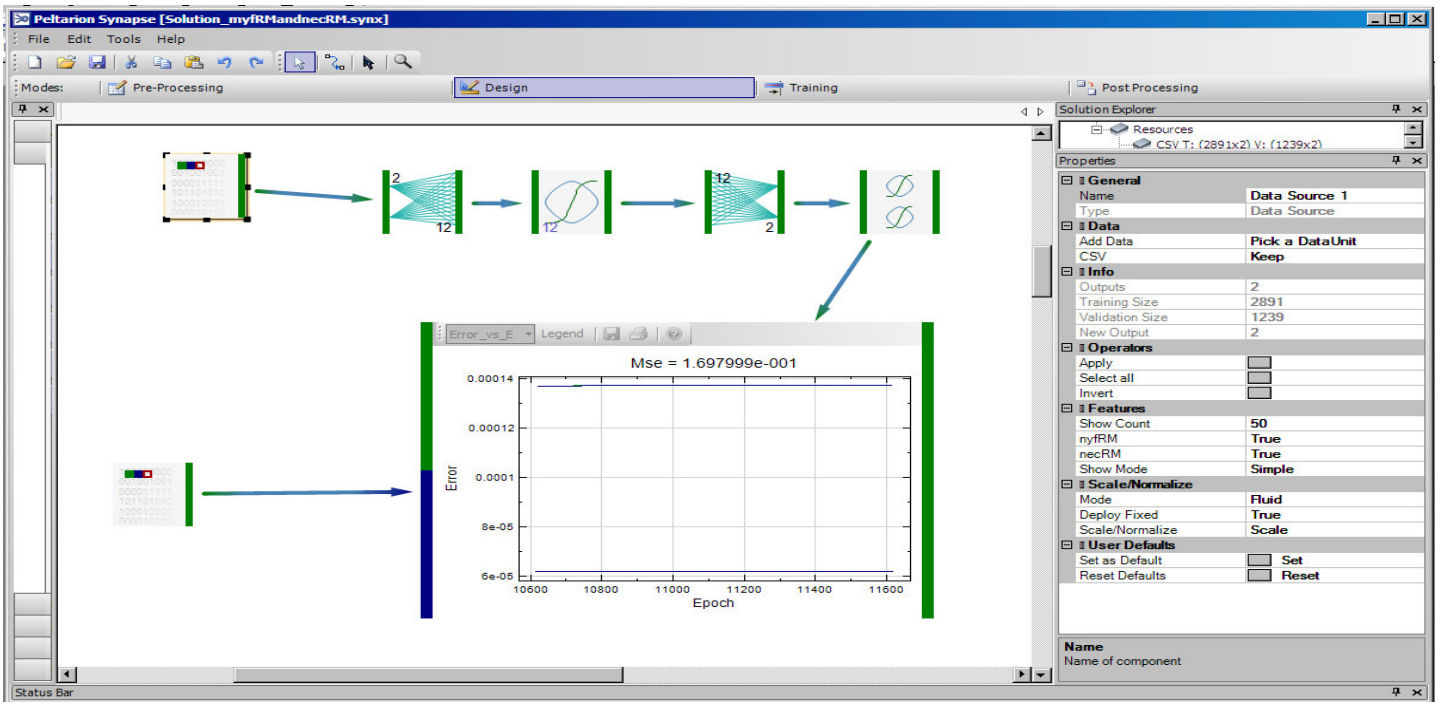
Yamamoto, M., Kuma, T., Nakano, M., Obara, Y., Kudo, T., Ichikawa, T. & Notsuki, I., (1985).Automatic measurement of electrical conductivity for the detection of bovine mastitis. Kiel. Milchwirtsch. Forschungsber., 37: 364-369.

Yang, X.Z., Lacroix, R & Wade, K. M., (1999). *Neural detection of mastitis from dairy herd improvement records.* Transactions of the Asae, 42(4), 1063-1071

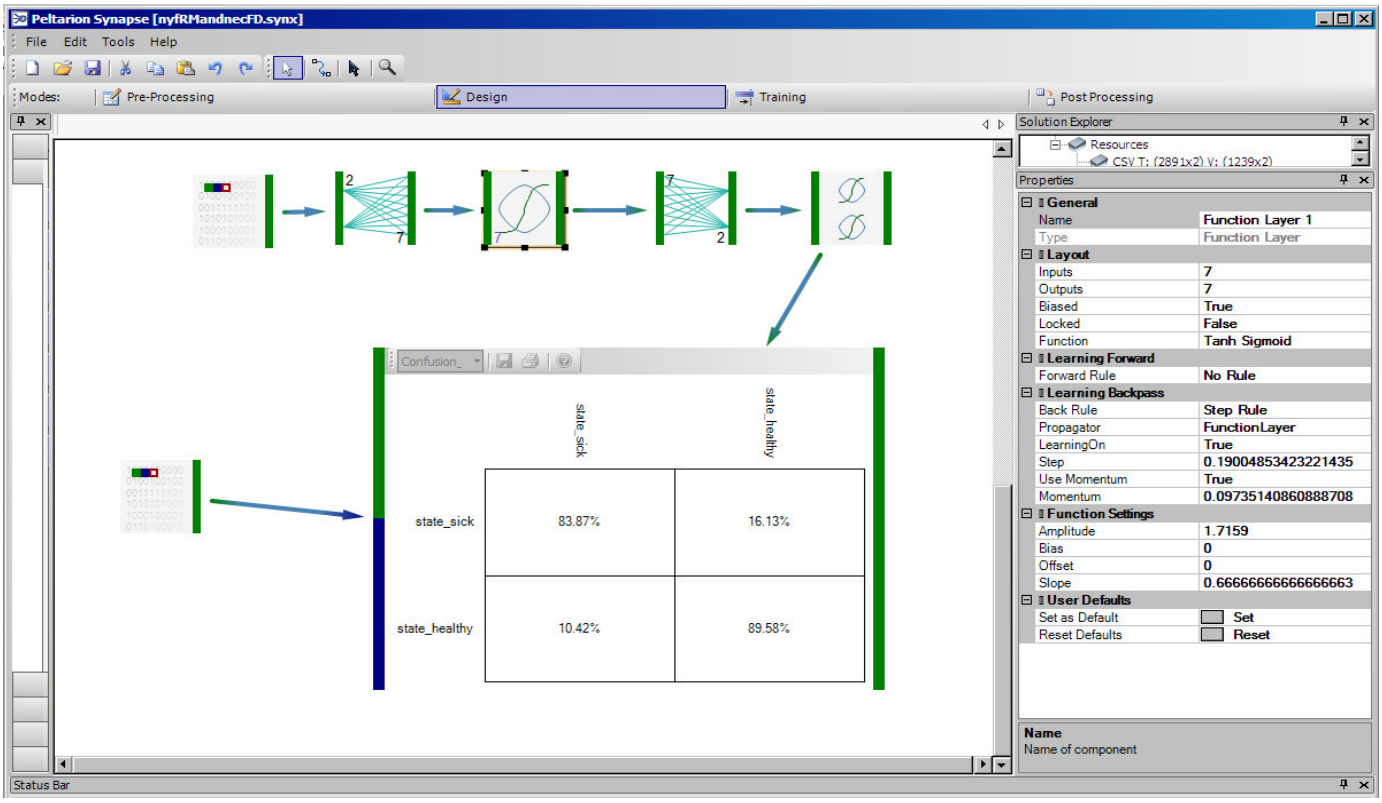
Appendix 1

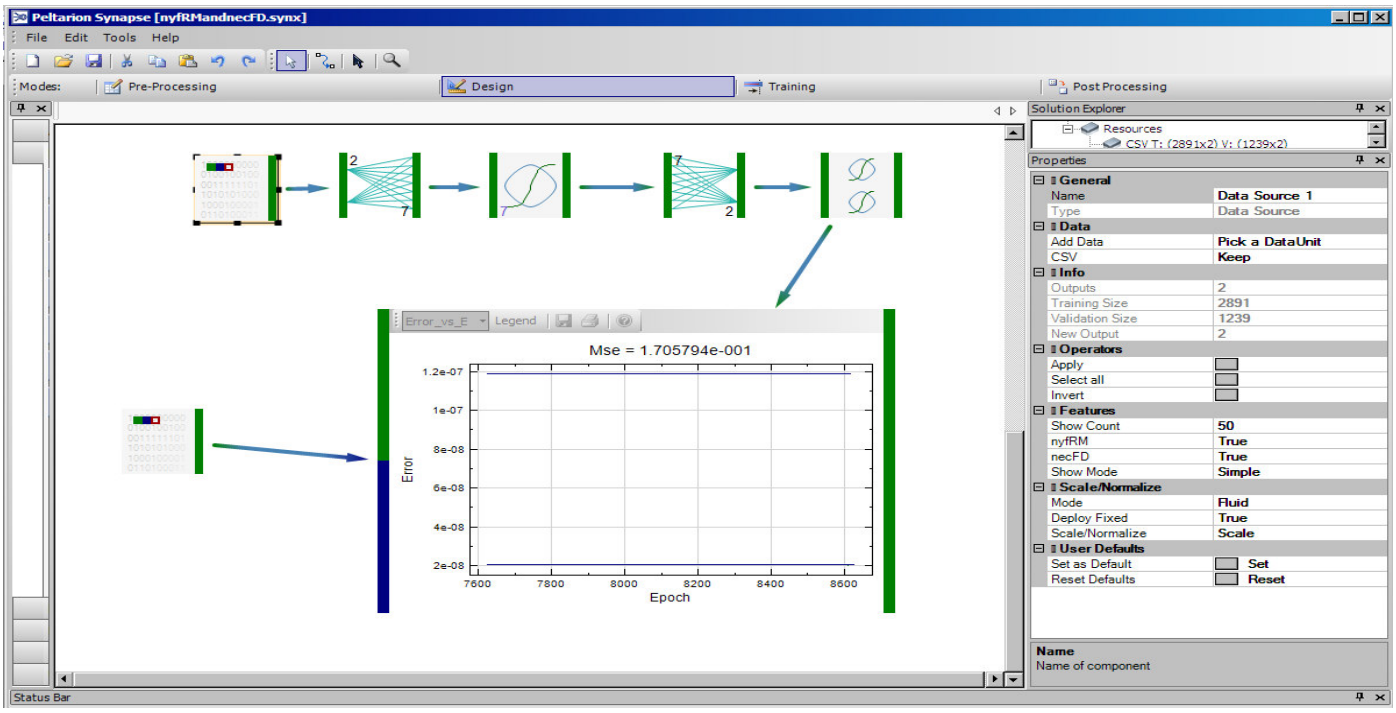
Figures show the predictive performance of model 1 (input: nyfRM and necRM) on the validation data set and MSE when model achieved the best performance. (The optimal parameters selected by genetic optimizer for the hidden layer are shown in the right part of the figure)



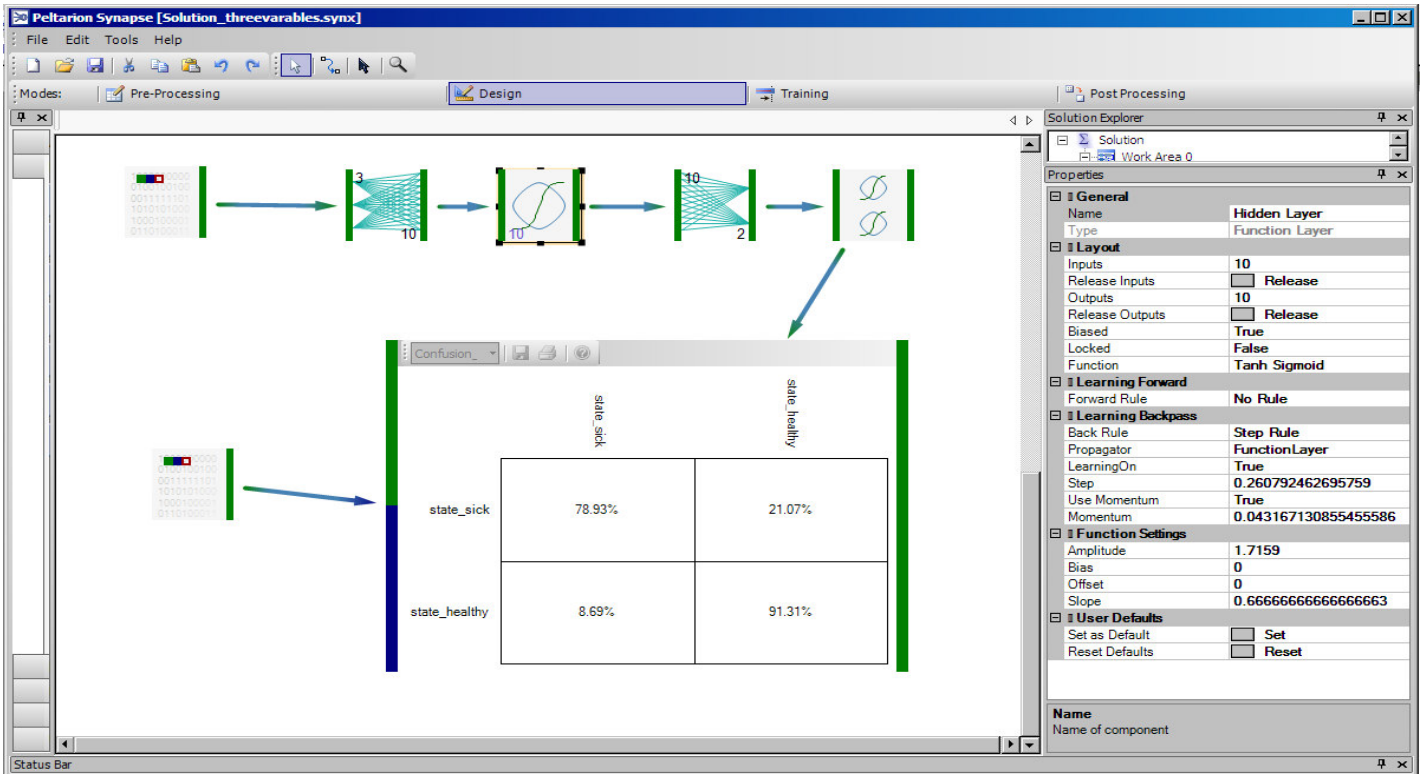


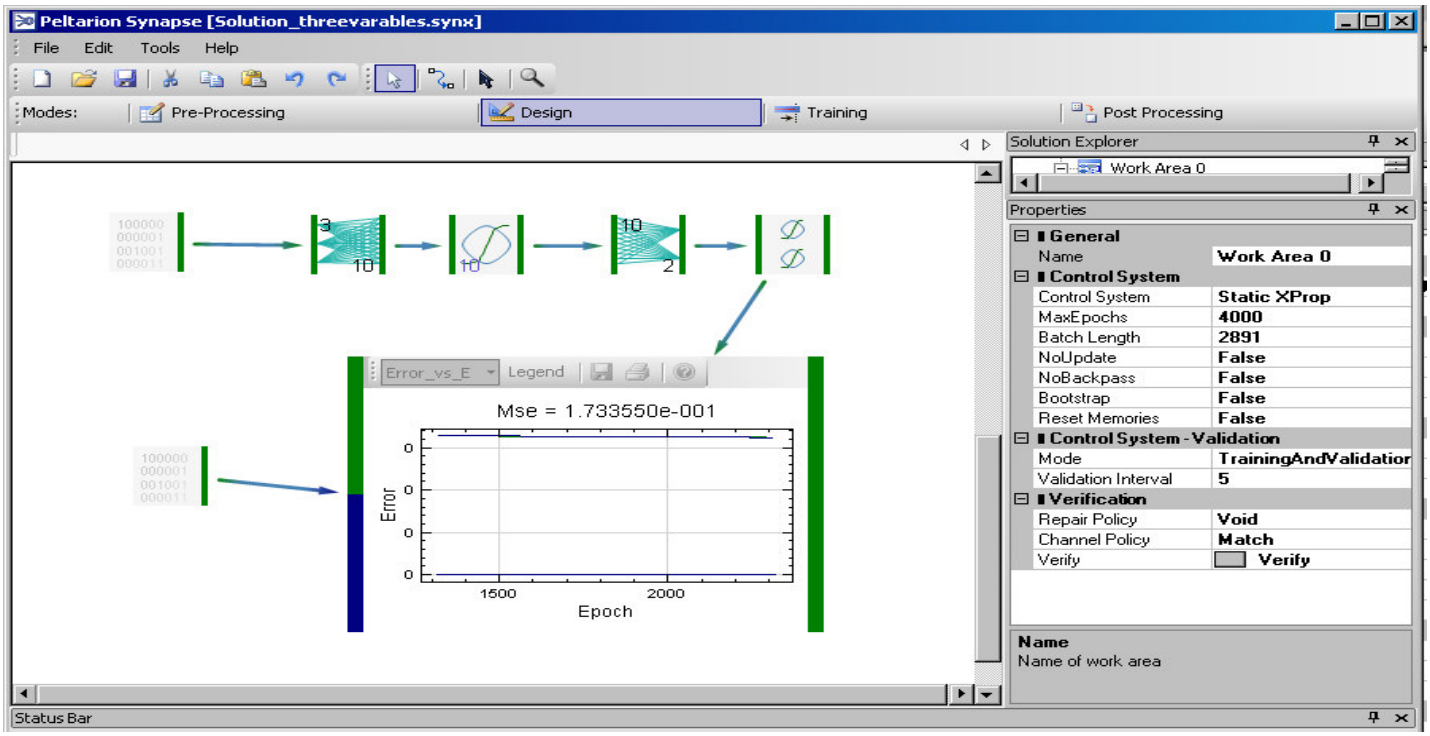
Figures show the predictive performance of model 2 (input: nyfRM and necFD) on the validation data set and MSE when model achieved best performance. (The optimal parameters selected by genetic optimizer for the hidden layer are shown in the right part of the figure)



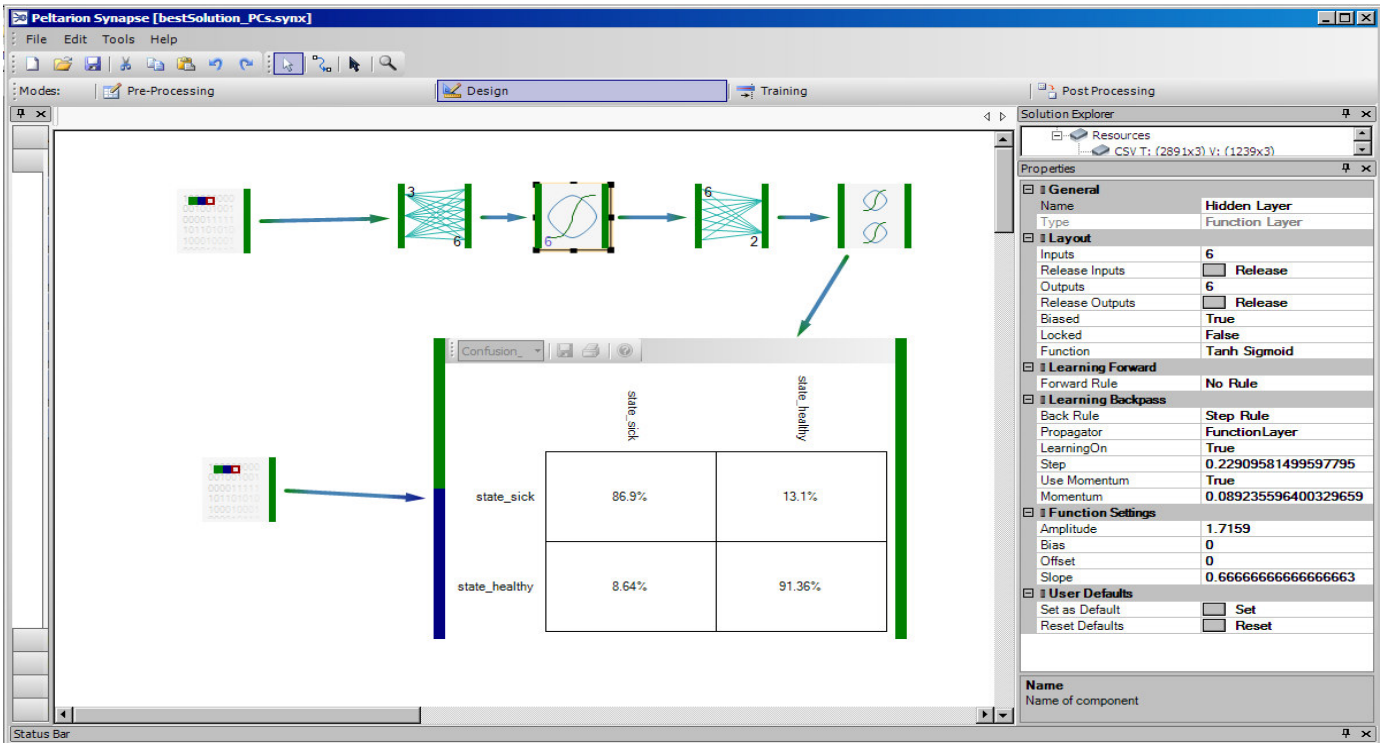


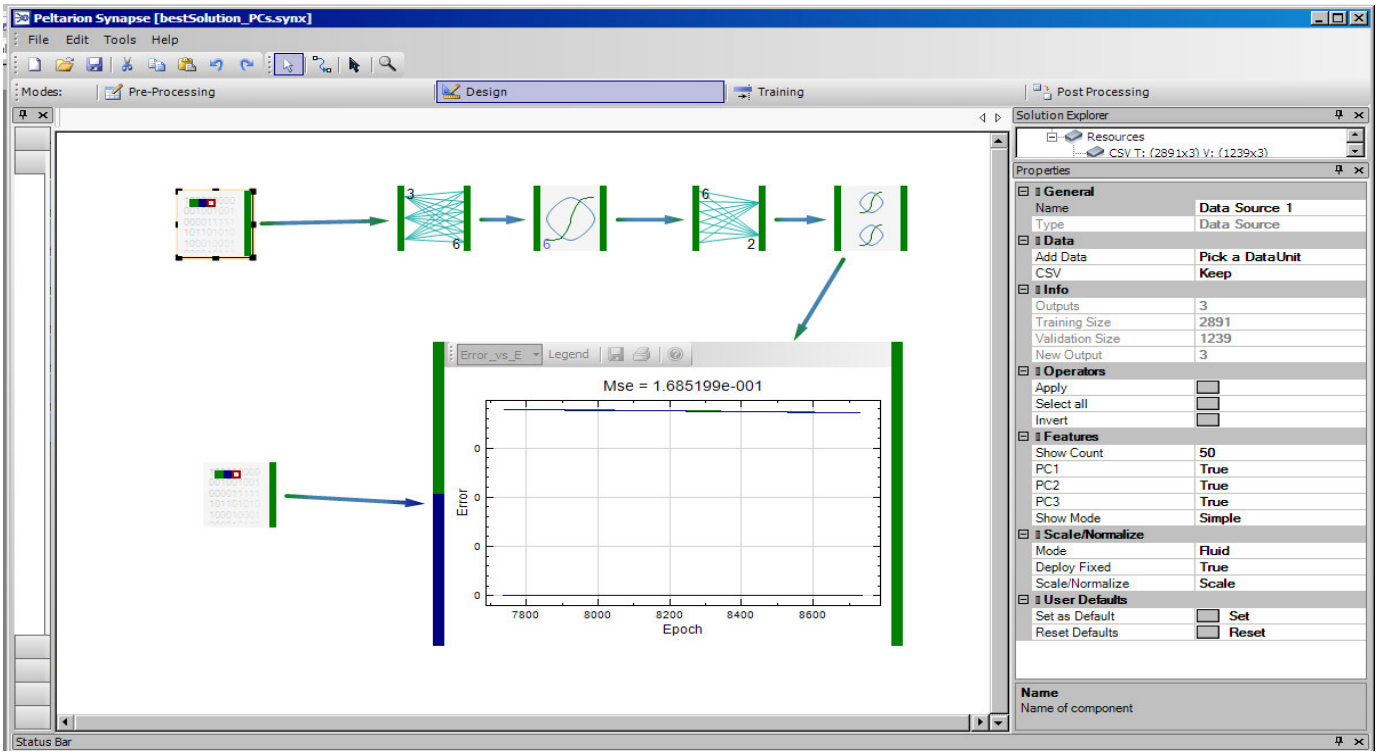
Figures show the predictive performance of model 3 (input: nyfRM, necRM and necFD) on the validation data set and MSE when model achieved the best performance. (The optimal parameters selected by genetic optimizer for the hidden layer are shown in the right part of the figure)





Figures show the predictive performance of model 4 (input: PCs) on the validation data set and MSE when model achieved the best performance. (The optimal parameters selected by genetic optimizer for the hidden layer are shown in the right part of the figure)





Appendix 2

Classification Results of LDA for the Dataset1 (nyfRM and necRM) (a, b)

				Predicted Group Membership		Total
			healthy_states	.00*	1.00*	.00
Cases Selected*	Original	Count	.00	2063	223	2286
			1.00	115	507	622
		%	.00	90.2	9.8	100.0
			1.00	18.5	81.5	100.0
Cases Not Selected*	Original	Count	.00	854	96	950
			1.00	62	210	272
		%	.00	**89.9**	10.1	100.0
			1.00	22.8	**77.2**	100.0

a 88.4% of selected original grouped cases correctly classified.

b 87.1% of unselected original grouped cases correctly classified.

* Cases selected = Training Dataset

* Cases Not Selected = Validation Dataset

* .00 = Healthy

* .1.00 = Mastitis.

Classification Results LDA for the Dataset 2 (nyfRM and necFD) (a, b)

				Predicted Group Membership		Total
			healthy states	.00	1.00	.00
Cases Selected	Original	Count	.00	2021	233	2254
			1.00	145	479	624
		%	.00	89.7	10.3	100.0
			1.00	23.2	76.8	100.0
Cases Not Selected	Original	Count	.00	880	102	982
			1.00	55	215	270
		%	.00	**89.6**	10.4	100.0
			1.00	20.4	**79.6**	100.0

a 86.9% of selected original grouped cases correctly classified.

b 87.5% of unselected original grouped cases correctly classified.

* Cases selected = Training Dataset

* Cases Not Selected = Validation Dataset

* .00 = Healthy

* .1.00 = Mastitis.

Classification Results of LDA for the Dataset3 (nyfRM, necRM and necFD) (a, b)

				Predicted Group Membership		Total
			healthy_states	.00*	1.00*	.00
Cases Selected*	Original	Count	.00	2076	214	2290
			1.00	117	520	637
		%	.00	90.7	9.3	100.0
			1.00	18.4	81.6	100.0
Cases Not Selected*	Original	Count	.00	847	99	946
			1.00	61	196	257
		%	.00	**89.5**	10.5	100.0
			1.00	23.7	**76.3**	100.0

a 88.7% of selected original grouped cases correctly classified.

b 86.7% of unselected original grouped cases correctly classified.

* Cases selected = Training Dataset

* Cases Not Selected = Validation Dataset

* .00 = Healthy

* .1.00 = Mastitis.

Classification Results of LDA for the Dataset 4 (PC1, PC2 and PC3) (a, b)

				Predicted Group Membership		Total
			healthy_states	.00	1.00	.00
Cases Selected	Original	Count	.00	2050	213	2263
			1.00	132	509	641
		%	.00	90.6	9.4	100.0
			1.00	20.6	79.4	100.0
Cases Not Selected	Original	Count	.00	875	98	973
			1.00	40	213	253
		%	.00	**89.9**	10.1	100.0
			1.00	15.8	**84.2**	100.0

a 88.1% of selected original grouped cases correctly classified.

b 88.7% of unselected original grouped cases correctly classified.

* Cases selected = Training Dataset

* Cases Not Selected = Validation Dataset

* .00 = Healthy

* .1.00 = Mastitis.