# The use of neural networks to detect minor and major pathogens that cause bovine mastitis

K.J. HASSAN, S. SAMARASINGHE and M.G. LOPEZ-BENAVIDES[1]

Natural Resource Engineering Group, Lincoln University, Canterbury, New Zealand

## ABSTRACT

Bovine mastitis is caused by a diverse range of bacteria, broadly categorised as major or minor pathogens. The objective of this research was to develop an unsupervised neural network (USNN) model for detecting major and minor bacterial pathogens present in milk, based on changes in milk parameters associated with mastitis. A database of 4852 quarter milk samples with records for milk parameters and bacteriological status was used to train and validate the USNN model. Correlations ($P<0.05$) were found between the infection status of a quarter and its somatic cell score (SCS, 0.86), electrical resistance index (ERI, -0.59) and protein percentage (PP, 0.33). Due to significant multicolinearity, the original variables were decorrelated using principle component analysis. Sensitivity of the model for correctly detecting major and minor infections was 80% and 89%, respectively. Specificity of the model for correctly detecting non-infected cases was 97%. The model is able to differentiate infected milk from non-infected based on milk parameters associated with mastitis. It is concluded that the USNN model can be developed and incorporated into milking machines to provide a reliable basis for mastitis control.

**Keywords**: mastitis; somatic cell count; electrical resistance; principle component analysis; self organising map; neural networks.

## INTRODUCTION

Mastitis is the costliest production disease of dairy cattle around the world (NMAC, 2006; NMC, 2006). Clinical mastitis cases require antibiotic treatment and withholding of infected milk from the vat for the duration of the treatment, creating negative economic and animal welfare repercussions for the farmer. Mastitis is caused by a diverse range of bacteria and broadly categorised into major and minor pathogens, based on physical and biochemical changes that they cause in milk (Harmon, 1994). Major pathogens include *Staphylococcus aureus, Streptococcus uberis, Streptococcus agalactiae, Streptococcus dysgalactiae, Escherichia coli* and *Klebsiella spp.,* while common minor pathogens include coagulase-negative staphylococci, *Corynebacterium bovis* and *Arcanobacterium pyogenes* (Brown *et al*., 1976; Harmon, 1994). Major pathogens are generally the cause of clinical mastitis, observable by the milker and characterised by obvious physical changes in milk appearance, such as clots, blood or flakes. In some cases, minor pathogens are also able to cause clinical mastitis. Detection of intramammary infection in the early stages (subclinical mastitis) of infection would be advantageous as timely management decisions could be made before clinical mastitis occurs.

Subclinical mastitis is harder to diagnose, as observable physical changes in milk are not as obvious, even though milk parameters are affected. Bacterial infection of the mammary gland causes changes in some milk parameters, including somatic cell count (SCC) (Heald *et al*., 2000), electrical resistance (ER) or electrical conductivity (Hamann & Zecconi, 1998; Norberg *et al*., 2004), milk protein (PP) and milk fat content (FP) (Sloth *et al*., 2003). Changes in milk parameters relative to normal values have been used before to detect mastitis with an acceptable degree of success (Woolford *et al*., 1998; Wang & Samarasinghe, 2005). Traditionally, bacteriology is considered the 'gold standard' for identification of mastitis pathogens, although this procedure requires culture of suspected milk in the lab, making routine bacteriological monitoring of glands expensive and time consuming. Another option is the use of milk parameters information obtained in-line to predict the causal agent of intramammary infection. Mathematical algorithms used to detect changes in individual cows must be robust and able to deal with complex interactions, such as days in milk (DIM), season, cow age and breed (Hamann & Zecconi, 1998) making artificial neural networks (ANN) an ideal tool for analysing these type of multivariate data.

The ANNs are interconnected networks of artificial neurons that acquire knowledge by processing information in a manner analogous to the human brain. ANNs are of two types, supervised and unsupervised. For supervised ANN, the output for a given input pattern is used for the ANN learning process (Rumelhart *et al.,* 1994), whilst the unsupervised ANN learning paradigm

learns without being shown any target output (Kohonen, 1998). In previous studies, supervised ANN and milk parameters were used to diagnose clinical and sub-clinical mastitis (Nielen *et al.*, 1995a; Nielen *et al.*, 1995b; Yang *et al.*, 1999; Yang *et al.*, 2000). Heald *et al.* (2000) used SCC, milk yield, lactation number, DIM and field survey data to detect mastitis causing pathogens. Unsupervised ANN were employed to detect clinical mastitis from electrical conductivity and milk yield data (Wang & Samarasinghe, 2005). In another study, individual cow glands were classified into different health categories based on milk parameters by using unsupervised ANN (Lopez-Benavides, 2004).

Advances in milking technology offer the possibility of measuring several milk parameters during the milking process. Models that use these milk parameters to detect the presence of particular types of mastitis pathogens can provide valuable information in the management of mastitis. The objective of this study was to develop an USNN model that could be incorporated into milking machines for the detection of major and minor mastitis pathogens in milk of cow glands.

## MATERIALS AND METHODS

### Experimental Data

Data analysed in this research were collected at the Lincoln University dairy farm, in Canterbury, New Zealand, from August to November 2002. Starting at calving, quarter foremilk samples were collected weekly from 112 cows for a period of 14 weeks, resulting in 4852 samples with complete records for SCC, ER, FP, PP and bacteriological status (Lopez-Benavides, 2004). Cows were daughters of three different sires. In summary, SCC, FP and PP were measured using CombiFoss 500 (Foss Electric, Denmark) at Livestock Improvement Corporation Hamilton, New Zealand; ER was measured using the Draminski Mastitis Detector (Draminski, Warsaw, Poland), and bacteriological analyses of milk samples was performed according to the standard guidelines (NMC, 1999) at the Dexcel Mastitis Research Lab in Hamilton, New Zealand.

### Data pre-processing and partitioning

The following transformations were performed on the raw data before advanced analysis:

- SCC was transformed into Somatic Cell Score (SCS), as shown in equation [1] (Dabdoub & Shook, 1984).

$$SCS_i = \left[ \log 2 \frac{SCC_i}{100} \right] + 4 \qquad [1]$$

- Electrical Resistance Index (ERI) was developed to account for within and between cow variations as shown in equation [2]. Individual quarter ERI was derived from cow ER values, which is the relationship of the individual quarter ER to the total ER for a cow, and to the maximum ER of a quarter on the test day.

$$ERI = \frac{ER_i}{ER_{\max}} - \frac{ER_i}{\sum_1^4 ER_i} \qquad [2]$$

Where:

$ER_i$ is the ER of quarter $i$, $ER_{max}$ is the maximum ER observed from an individual cow, and;

$\sum_1^4 ER_i$ is the total ER from all four quarters on the test day.

To account for effects of stage of lactation on milk parameters due to DIM, data were divided into three ranges (range 1 = 1 to 30 DIM; range 2 = 31 to 57 DIM and range 3 = 58 to 125 DIM). Cow variation was accounted for by the sire family from which the cow originated and the lactation number of each cow. Bacteriological status (BS) was coded according to bacteriology of each case (no infection = NI; minor pathogen infection = MNI; major pathogen infection = MJI). The dataset was partitioned into training and validation sets. The training set was used to develop the USNN model and the validation set to assess its generalization capability. The training set was obtained by random extraction of an approximately equal amount of data from each BS using STATISTICA v7.1 (StatSoft, Tulsa, OK, USA 2006). This consisted of 50 NI, 43 MNI and 37 MJI cases. The remaining data containing 4411 NI, 301 MNI and 10 MJI cases were used for validating the model.

### Data analysis and model development

Relationships between variables were analysed using STATISTICA v7.1. Pearson correlation analysis between variables was conducted to identify the most influential input variables to train the USNN model and principle component analysis was used to eliminate multicollinearity between the input variables. The proportion of variance (PV) explained by each principle component (PC) was calculated. The USNN learning paradigm (Kohonen, 2001) was used to develop the model using Viscovery-SOMine (Eudaptics software GmbH, Vienna, Austria). The model was trained using the training data set containing four PCs, BS, DIM, lactation and sire numbers. The model was validated using the validation data set containing all the variables, with the exception of BS. The input data set was used to train 66 neurons in the

output layer of the model, to learn the intrinsic relationship between the input variables. When the training error was reduced to a minimum level, the training was considered as complete. Trained neurons were grouped into three clusters, representing the three BS, using the SOM-Ward clustering method that uses a combination hierarchical cluster algorithm of Ward and SOM methods.

## RESULTS

### Data Analysis

A strong correlation was observed between SCS and BS ($r = 0.86$; $P<0.05$) and ERI and BS ($r = -0.59$; $P<0.05$). Moderate correlation was observed between PP and BS ($r = 0.33$; $P<0.05$), while no significant correlation was observed between FP and BS. However, a strong correlation existed between ERI and SCS ($r = -0.65$; $P<0.05$), and moderate correlations were observed between PP and SCS ($r = 0.32$; $P<0.05$), and PP and ERI ($r = -0.35$; $P<0.05$). Weak correlations were observed between ERI and FP ($r = 0.18$; $P<0.05$), DIM and ERI ($r = 0.19$; $P<0.05$), and DIM and PP ($r = -0.22$; $P<0.05$). No significant correlations were observed for sire and lactation numbers.

The strong multicolinearity that existed between SCS, ERI, PP and FP was resolved using principle component analysis, resulting in the extraction of four PCs: PC-1 (PV=39%), PC-2 (PV=27%), PC-3 (PV=23%) and PC-4 (PV=12%). It was found that SCS and ERI contributed most towards PC-1 and PP and ERI towards PC-2. The FP, which was not strongly correlated to any of the SCS, ERI or PP, had maximum contribution towards PC-3. Finally, PC-4 explained the remaining variance of SCS and ERI.

### Unsupervised neural network model

The classification results of the USNN model for the validation data set are shown in Table 1. Overall, the specificity or correct identification of NI cases was 97% (4289/4411) and sensitivities for correctly identifying minor and major pathogens were 89% (269/301) and 80% (8/10) respectively. Misclassification rate was 3% for NI, 11% for MNI and 20% for the MJI cases. For NI, 2% (86/4411) were misclassified as MNI and 1% (36/4411) as MJI. For MNI cases, 1% (4/301) were misclassified as NI and 9% (28/301) as MJI. For MJI cases, 20% (2/8) were misclassified as MNI and none as NI. In the MNI category 87% of coagulase-negative staphylococci spp. (131/150) and 92% of *C. bovis* cases (138/150) were classified correctly. In the MJI category, 71% of *S. uberis* (5/7), 100% of *S. dysgalactiae* (2/2) and 100% of *S. agalactiae* (1/1) were correctly classified.

Differences in mean values of milk `parameters` and PCs were analysed for each BS cluster using one way ANOVA (Table 2). As BS changed from NI to MNI, the mean values of SCS and FP increased ($P<0.001$) and ERI decreased slightly ($P<0.001$). The PP change was not significant. Although SCS was the most important parameter for discriminating between NI and MNI, some overlapping regions existed, as some MNI cases had high ERI as well as high SCS. Milk parameters differences were more pronounced between NI and MJI. In general, mean SCS was elevated ($P<0.001$) and ERI and FP means were reduced ($P<0.001$), while change in PP was not significant. The relationship between the MNI and MJI cases was also evaluated; for the MJI, mean SCS was higher ($P<0.001$), but ERI and FP ($P<0.001$) were lower, while change in PP was not significant. For discriminating between BS clusters, a negative linear trend was observed for PC-1, where values decreased as BS changed from NI to MNI and MJI. A positive linear trend was observed for PC-2 and PC-3, where values increased as BS changed from NI to MNI and MJI. Although the differences between BS clusters were significant for PC-4, no linear trend was evident.

## DISCUSSION

This study explored the use of individual quarter milk `parameters` along with bacteriology data to train an USNN for differentiating quarter milk samples into three distinct bacteriological infection clusters, NI, MNI and MJI. A plethora of literature has shown that SCC/SCS reflects the immune response of a cow to bacterial infections,

**Table 1:** Matching matrix of observed and predicted bacteriological status clusters for the validation data set using the USNN model. Specificity for NI was 97%. Sensitivities were 89% for MNI and 80% for MJI.

| | | USNN prediction (n) | | | |
|---|---|---|---|---|---|
| **Bacteriological status** | **Observed (n)** | **NI** | **MNI** | **MJI** | **Total** |
| Not Infected (NI) | 4411 | 4289 | 86 | 36 | 4411 |
| Minor Infected (MNI) | 301 | 4 | 269 | 28 | 301 |
| Major Infected (MJI) | 10 | 0 | 2 | 8 | 10 |
| Total | 4722 | | | | 4722 |

**Table 2:** Milk parameter means (±SEM) for each bacteriological status cluster for the validation data set using the USNN model. Milk samples were clustered into not-infected (NI), infected by minor pathogens (MNI) or infected by major pathogens (MJI).

| | Bacteriological status cluster | | | |
|---|---|---|---|---|
| **Milk Parameters** | **NI** | **MNI** | **MJI** | **(p<0.001)** |
| Somatic cell score | 3.14 ± 0.02 | 5.31 ± 0.07 | 7.06 ± 0.15 | All clusters |
| Electrical resistance index | 0.72 ± 0.00 | 0.71 ± 0.00 | 0.57 ± 0.00 | All clusters |
| Milk protein (%) | 3.57 ± 0.00 | 3.56 ± 0.01 | 3.57 ± 0.03 | NS |
| Milk fat (%) | 3.44 ± 0.02 | 4.03 ± 0.07 | 2.88 ± 0.17 | All clusters |
| Principle component-1 | 0.15 ±0.01 | -0.76 ± 0.04 | -2.66 ± 0.09 | All clusters |
| Principle component-2 | -0.03 ± 0.01 | -0.31 ± 0.05 | 2.46 ± 0.11 | All clusters |
| Principle component-3 | -0.08 ± 0.02 | 0.43 ± 0.05 | 1.61 ± 0.11 | All clusters |
| Principle component-4 | -0.08 ± 0.01 | 0.88 ± 0.05 | -0.49 ± 0.11 | All clusters |

which was also observed in this study. The high correlation between SCS and BS observed in this study agrees with Berning & Shook (1992), although they found that SCS did not differentiate between the MNI and MJI cases. In this study we found that SCS can effectively discriminate between NI, MNI and MJI cases (Table 2). Another useful milk parameter for discriminating infected vs. not infected milk samples was electrical conductivity, and more specifically ERI. Electrical conductivity measures the ionic changes resulting from damage caused by bacteria to the alveoli (Kitchen *et al.*, 1980). We found that ERI was lowest for MJI (Table 2), suggesting that tissue damage caused by major pathogens was noticeably higher compared to minor pathogens. The mean difference in the ERI values between the NI and MNI cases was very low, suggesting little damage of mammary tissues by minor pathogens.

Mastitis-related changes in protein and fat content are not conclusive as literature results are contradictory or at odds. Mastitis may cause milk protein content to increase (Auldist *et al.*, 1995), decrease (Lee *et al.*, 1991) or even stay unchanged (Rogers & Mitchell, 1989). In our study we found little change in PP for infected and non-infected milk samples (Table 2), which agrees with the findings of Rogers & Mitchell (1989). In a similar manner, mastitis also causes changes in the milk fat content, but without a definitive trend, as it may increase (Mitchell *et al.*, 1986), decrease (Kitchen, 1981) or remain unchanged (Rogers & Mitchell, 1989). In this study, a slight non-significant increase in FP was observed for MNI cases, but a significant decrease for MJI (Table 2). This indicates that FP may discriminate between the NI and MJI cases but may not be optimal for discriminating between the NI and MNI cases.

The USNN model was able to cluster the data with a 95% degree of accuracy. These results are superior to previous models developed by Heald *et al*. (2000), where the degree of accuracy ranged from 57 to 71%. Nevertheless, the input variables and learning paradigm used by Heald *et al*. (2000) were different to those used in the present study. In their study, the classification of pathogens was based on their mode of transmission (contagious, environmental & others), whereas in our study, milk samples were clustered based on their impact on the mammary gland, which gave better results.

In some instances the USNN model assigned milk samples into the wrong BS cluster. A few NI cases were classed as MNI or MJI by the model probably because they had high SCS and low ERI values. One reason for this result could be that these were late infections, where the cow's immune system had already eradicated the causal pathogens. Another reason may be that the measurement devices failed to give correct readings for the milk parameters of these cases. Due to the low prevalence of major pathogens in this herd, there were fewer cases for the MJI cluster. The results of the model for MJI may be inconclusive as more data are needed to make the model more robust. In this study, only internal validation was performed and it is suggested that external validation of the model should be carried out before its application at the farm level.

It is concluded that milk parameters associated with mastitis can be used to build robust ANN models for detecting the type of infection caused by either minor or major pathogens. The incorporation of such models in the in-line milking systems may improve the efficacy of detecting mastitis pathogens in milk before any clinical manifestations occur, which may form the basis for managing and controlling mastitis at the farm level.

## ACKNOWLEDGEMENTS

The authors thank the New Zealand's International Aid and Development Agency, for financially supporting this research. The resources at Lincoln University were used to conduct this

research. The authors also acknowledge the people and organisations that were instrumental in the development of the database used in this research (Lincoln University Dairy Farm staff, Livestock Improvement Corporation Ltd, Fonterra Ltd, and Dexcel Ltd). The authors also thank the reviewers for their constructive comments.

## REFERENCES

Auldist, M.J.; Coats, S.; Rogers, G.L.; McDowell, G.H. 1995: Changes in the composition of milk from healthy and mastitic dairy cows during the lactation cycle. *Australian Journal of Experimental Agriculture* **35**(4): 427-436.

Berning, L.M.; Shook, G.E. 1992: Prediction of mastitis using milk somatic cell count, N-acetyl-β-D-glucosaminidase, and lactose. *Journal of Dairy Science* **75**: 1840-1848.

Brown, R.W.; Battista, W.J.; Schultz, L.H.; Johnston, R.P. 1976: Variation in somatic cell counts in dairy herd improvement milk samples. *Journal of Dairy Science* **59**: 1119.

Dabdoub, S.A.M.; Shook, G.E. 1984: Phenotypic relations among milk yield, somatic cell count and clinical mastitis. *Journal of Dairy Science* **67**: 163-164.

Hamann, J.; Zecconi, A. 1998: Evaluation of the electrical conductivity of milk as a mastitis indicator. *Bulletin of the International Dairy Federation, Brussels, Belgium* **334**: 5-22.

Harmon, R.J. 1994: Symposium - Mastitis and Genetic Evaluation for Somatic-Cell Count - Physiology of Mastitis and Factors Affecting Somatic-Cell Counts. *Journal of Dairy Science* **77**(7): 2103-2112.

Heald, C.W.; Kim, T.; Sischo, W.M.; Cooper, J.B.; Wolfgang, D.R. 2000: A computerised mastitis decision aid using farm-based records: An artificial neural network approach. *Journal of Dairy Science* **83**(4): 711-720.

Kitchen, B.J.; Middleton, G.; Durward, I.G.; Andrews, R.J.; Salmon, M.C. 1980: Mastitis diagnostic tests to estimate mammary gland epithelial cell damage. *Journal of Dairy Science* **63**(6): 978-983.

Kitchen, B.J. 1981: Review of the progress of dairy science: bovine mastitis: milk compositional changes and related diagnostic tests. *Journal of Dairy Research* **48**(1): 167-188.

Kohonen, T. 1998: The self-organizing map. *Neurocomputing* **21**(1-3): 1-6.

Kohonen, T. 2001: *Self-Organising Maps* (3rd ed.): Berlin, Heidelberg, New York. Springer.

Lee, S.C.; Yu, J.H.; Jeong, C.L.; Back, Y.J.; Yoon, Y.C. 1991: The influence of mastitis on the quality of raw milk and cheese. *Korean Journal of Dairy Science* **13**: 217-223.

Lopez-Benavides, M.G. 2004: *BoLA-DQA2 haplotypes and resistance to bovine mastitis. PhD thesis.* Lincoln University, Canterbury, New Zealand.

Mitchell, G.E.; Rogers, S.A.; Houlihan, D.B; Tucker, V.C.; Kitchen, B.J. 1986: The relationship between somatic cell count, composition and manufacturing properties of bulk milk. I. Composition of farm bulk milk. *Australian Journal of Dairy Technology* **41**(1): 9-12.

Nielen, M.; Schukken, Y.H.; Brand, A.; Deluyker, H.A.; Maatje, K. 1995a: Detection of Subclinical Mastitis from Online Milking Parlor Data. *Journal of Dairy Science* **78**(5): 1039-1049.

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

NMAC. 2006: The Cost of Mastitis - Final Report, by The National Mastitis Advisory Committee, New Zealand. Retrieved Jan 10, 2007, from http://www.dexcel.co.nz/data/usr/Costs%20of%20Mastitis%20Report%20Final.pdf.

NMC. 1999: Laboratory handbook on bovine mastitis. National Mastitis Council. Madison, WI, USA.

NMC. 2006: National Mastitis Council USA - A global organization for mastitis control and milk quality. Retrieved May 14, 2006, from: http://www.nmconline.org/dhiscc.htm.

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.

Rogers, S.A.; Mitchell, G.E. 1989: The relationship between somatic cell count, composition and manufacturing properties of bulk milk. IV. Non-protein constituents. *Australian Journal of Dairy Technology* **44**(2): 53-56.

Rumelhart, D.E.; Bernard W.; Michael A.L. 1994: The basic ideas in neural networks. *Association for computing machinery. Communications of the ACM.* ***37***(3): 86-92.

Sloth, K.H.M.N.; Friggens, N.C.; Lovendahl, P.; Andersen, P.H.; Jensen, J.; Ingvartsen, K.L. 2003: Potential for Improving Description of Bovine Udder Health Status by Combined Analysis of Milk Parameters. *Journal of Dairy Science* **86**(4): 1221-1232.

Wang, E.; Samarasinghe, S. 2005: *On-line detection of mastitis in dairy herds using artificial neural networks.* Paper presented at the International congress on modelling and simulation, Melbourne, Australia.

Woolford, M.W.; Williamson, J.H.; Henderson, H.V. 1998: Changes in electrical conductivity and somatic cell count between milk fractions from quarters subclinically infected with particular mastitis pathogens. *Journal of Dairy Research* **65**(2): 187-198.

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.

Yang, X.Z.; Lacroix, R.; Wade, K.M. 2000: Investigation into the production and conformation traits associated with clinical mastitis using artificial neural networks. *Canadian Journal of Animal Science* **80**(3): 415-426.