

The use of artificial neural networks to diagnose mastitis in dairy cattle

M.G. López-Benavides, S. Samarasinghe and J.G.H. Hickford

1. Animal & Food Sciences Division

2. Natural Resources Engineering

P.O. Box 84

Lincoln University, Canterbury

New Zealand

Abstract - The use of milk sample categorization for diagnosing mastitis using Kohonen's Self-Organizing Feature Map (SOFM) is reported. Milk trait data of 14 weeks of milking from commercial dairy cows in New Zealand was used to train and test a SOFM network. The SOFM network was useful in discriminating data patterns into four separate mastitis categories. Several other Artificial Neural Networks were tested to predict the missing data from the recorded milk traits. A multi-layer perceptron (MLP) network proved to be most accurate ($R^2 = 0.84$, $r = 0.92$) when compared to other MLP ($R^2 = 0.83$, $r = 0.92$), Elman ($R^2 = 0.80$, $r = 0.92$), Jordan ($R^2 = 0.81$, $r = 0.92$) or linear regression ($R^2 = 0.72$, $r = 0.85$) methods. It is concluded that the SOFM can be used as a decision tool for the dairy farmer to reduce the incidence of mastitis in the dairy herd.

Keywords: ANN, SOFM, mastitis, dairy cattle

I. INTRODUCTION

In the dairy industry, mastitis is probably the most important disease. It not only affects the yield and composition of milk, but it also affects the welfare of the cow due to increased use of antibiotics and physical damage to the udder [1,2]. Early detection of sub-clinical symptoms is of major concern and has been the subject of research for many years, with elusive results. Various measures of mastitis have been proposed and some are currently used in the industry [10]. Of these, somatic cell count (SCC) is probably the most popular measure due to its high correlation with mastitis (0.7) [6]. Nevertheless, there are some doubts as to its use as an indicator of intra-mammary infection (IMI) [3]. Electrical conductivity (EC) is another measure that has proven to be effective in detecting changes in milk composition due to mastitis [7, 8]. This measure appears to be sensitive to changes in ion concentrations in milk when the animal is affected by mastitis, explaining why this measure is used with increasing frequency in the dairy industry. The use of combined traits may be a better way of understanding the health status of the cow at a determined time. One of the objectives of this research was to use both SCC and EC along with other milk traits to cluster 'quarter' milk samples into separate mastitis health categories using ANN.

One variation of artificial neural networks (ANN) is a self-organizing network, such as Kohonen's Self-Organizing Feature Map. In this type of ANN, the system is trained by showing examples of patterns that are to be classified, and the network provides its own output for classification [9]. The SOFM has been used in several fields including agriculture, but not as a classification tool in milk. It is believed that SOFM can be used as a reliable management tool for a quick identification of potentially mastitic animals, so that appropriate measures can be undertaken to reduce the impact of this disease in the dairy herd. Electrical conductivity can be measured easily as a cow-side test, while SCC measures must be done elsewhere. For this reason, several ANN and traditional linear regression analysis were tested to predict missing data necessary for the SOFM when some trait measures are unavailable.

II. MATERIALS AND METHODS

A. Animals and milk samples

Quarter milk from 112 crossbred (Holstein Friesian X Jersey) cows in their first to third lactations (LN) was taken at weekly intervals for a period of 14 weeks, from the start to the middle of lactation in the 2002-2003 season, as described earlier [5]. Each milk sample was analyzed for various milk traits including EC, SCC, fat percentage (FP) and protein percentage (PP).

B. Data analysis

Data was divided into two days-in-milking ranges (DIMR1 = 4 to 64 days & DIMR2 = 65 to 125 days) to account for differences in values due to lactation stage. All the data set was analyzed using SPSS for Windows 10.0 (SPSS Inc., Chicago, IL, USA) to obtain relationships among variables and to create new informative ones that could be used as inputs for the ANN. New informative variables included inter-quarter ratio (IQR), conductivity index (CI), linear score (SCS) and composite milk index (CMI). Briefly, IQR and CI are derived from the EC values and take into account the relationships between quarters in the udder. Linear score is a linear transformation of SCC, and CMI is a value obtained by the sum of raw and derived variables from the milk traits recorded (see Appendix).

ANOVA of means led to a grouping of data into four separate datasets, i.e. LN1&2-DIMR1, LN1&2-DIMR2, LN3-DIMR1 & LN3-DIMR2. Also, electrical conductivity and Composite Milk Index were chosen as the most important variables as inputs for the SOFM. Data was graphed using SigmaPlot for Windows 5.00 (SPSS Inc., Chicago, IL, USA).

C. Clustering of datasets into health categories

Data sets from defined LN and DIMR were clustered into separate health categories using NeuroShell@ 2 (Ward Systems Group, Frederick, MD, USA). Electrical conductivity and CMI were used as inputs. The unsupervised SOFM algorithm [4] was used to cluster data patterns from each set into four arbitrary health categories, i.e. healthy (H0), moderately ill (H1), ill (H2) and severely ill (H3). Clustering was also tested on each dataset using the K-means method.

D. Prediction of Composite Milk Index using ANN

Since EC can be easily measured as a cow-side test during milking, the prediction of CMI from EC and its derivatives (i.e. IQR and CI) was tested. Four different ANN and a linear regression method were tested to predict CMI. This included two multi-layer perceptron (MLP), a Jordan and an Elman networks [9]. The MLPs consisted of three layers: an input layer of 3 nodes of linear activation function (LiAF), a hidden layer of either 50 (MLP 1) or 100 (MLP 2) nodes of logistic activation function (LAF), and an output layer of 1 node (LAF). The Elman recurrent network consisted of an input of 3 nodes (LiAF), a hidden layer of 50 nodes (LAF), a recurrent context layer of 50 nodes and an output layer of 1 node (LAF). Finally, the Jordan recurrent network consisted of an input layer of 3 nodes LiAF, a hidden layer of 50 nodes (LAF), a context layer of 1 node and an output layer of 1 node (LAF).

The complete dataset used for prediction consisted of 5060 data patterns of which 60% were used for training, 20% for testing the network, and 20% as a validation set. The learning rate and momentum were 0.1 and the network was trained for 500 epochs.

III. RESULTS AND DISCUSSION

A. Health categories

The SOFM proved to be useful in clustering the data into statistically significant mastitis categories (Table 1). This is the first recorded use of an SOFM for this purpose. Previously, a back-propagation ANN was used to detect mastitis from dairy herd improvement records, with an overall accuracy of 0.86 [11]. But as the authors concluded, new sources of information would have made the categorization more accurate. This SOFM used several milk traits that were highly informative to mastitis

and therefore it can be suggested that the results are representative of the mastitic state of the cow. Moreover, since the differences between lactation number and days-in-milking were taken into account, it reflects the change in milk composition that the animal suffers throughout its productive life. K-means clustering of LN1&2-DIMR1 had a correlation of 0.89 ($P < 0.01$) with the one obtained with SOFM, while it was 0.91 ($P < 0.01$) for LN1&2-DIMR2. For LN3-DIMR1 and LN3-DIMR2 the correlations were 0.87 and 0.84 with the SOFM results, respectively.

Bacteriological analysis is usually considered a 'gold standard' for mastitis diagnosis [7]. Overall, only 9.55% of quarters were positive for any bacteriological growth. The SOFM network showed that the frequencies of positive cases were 7.37% (healthy), 8.94% (moderately ill), 10.11% (ill) and 50% (severely ill) (see Table 1). There were less than 20 cases with clinical symptoms such as clots or blood in the milk and about 50% of those were assigned to either the ill or severely ill categories. The other quarters were considered to be healthy or moderately ill by the network. Even though there was an increase in positive cases as quarters were considered increasingly ill, bacteriology was not a reliable standard for diagnosis. This may be due to cows being treated with antibiotics prior to sampling or because some cows developed a rapid and effective immune response to infection. It is notable that 7.37% of healthy cases were positive for bacteria, which suggests that in these cases, no detectable changes in milk traits were detected.

Although the differences in means between moderately ill or ill categories were statistically significant in all datasets, perhaps a combined category of these two, or an expansion to a further category could make these differences more noticeable. Conductivity Index, electrical conductivity and CMI were the most important indicators of mastitis status, while the SCC and fat percentage significance varied. As animals became more ill, SCC values tended to increase as well as Conductivity Index and CMI (Figure 1). It is generally accepted that an increase in SCC reflects the body's immune status, i.e. higher values mean a possible infection in the quarter, while lower values resemble a healthy one [3]. The only difference in SCC values was in LN3, where this tendency was not followed (Table 1). This is probably due to the lack of sufficient data needed for training the network. It is suggested that CMI be used as an indicator of mastitic status because it encompasses several milk traits into one single measure.

B. Prediction of Composite Milk Index using ANN

The training mean R^2 for all the prediction methods was around 0.70 and the correlation was about 0.82. In the training dataset, the Elman & Jordan networks proved to be better predictors of CMI than the MLP (Table 2). The IQR and

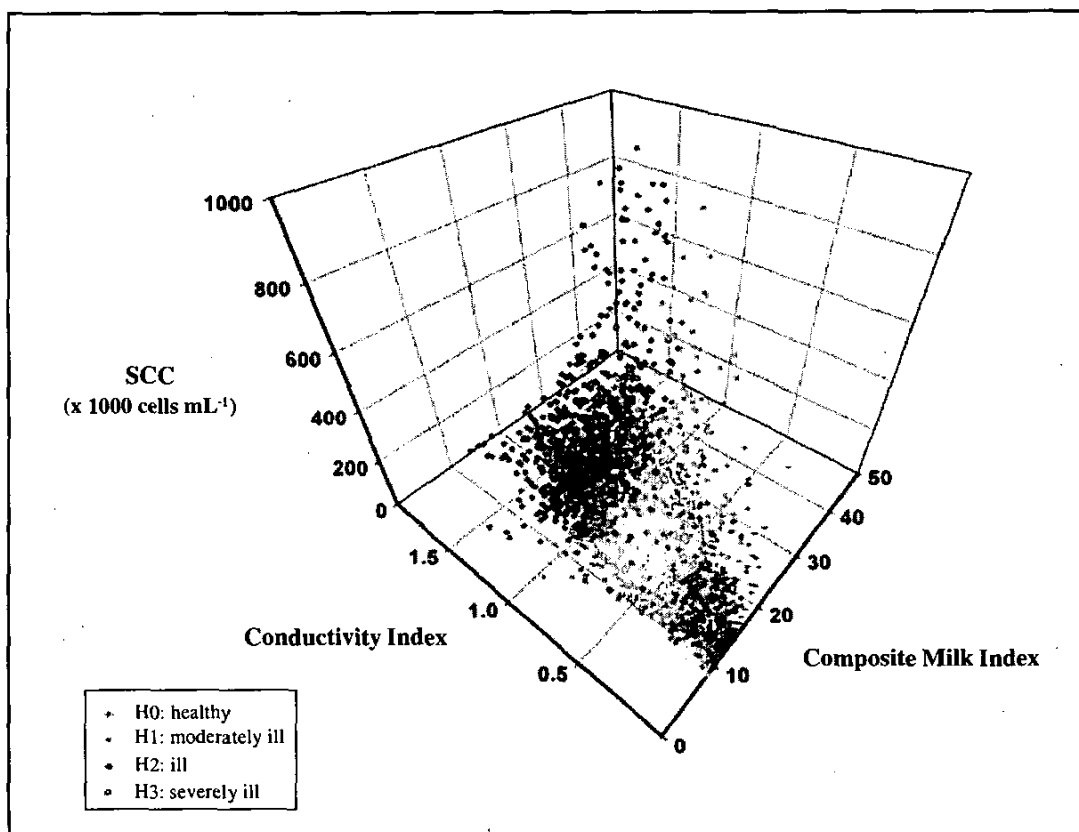


Figure 1. Scatter plot of SCC, Conductivity Index and CMI according to the mastitis categories

Table 1. Means (SEM) of milk traits according to health categories

Health category	N	CMI	S C C cells mL ⁻¹	FP (%)	CI	EC	Microbiology		
							n	Positive (%)	
LN 1 & 2 DIMR1	H0	723	12.9 ± 0.1	74 ± 6	4.3 ± 0.1	-0.02 ± 0.0	495 ± 1	36	4.98
	H1	1518	18.9 ± 0.1	91 ± 11	3.6 ± 0.0	0.65 ± 0.0	433 ± 0	113	7.44
	H2	1431	25.7 ± 0.3	203 ± 23	3.2 ± 0.0	1.29 ± 0.0	377 ± 1	136	9.50
	H3	25	108.7 ± 9.8	3140 ± 784	2.2 ± 0.3	5.33 ± 0.2	226 ± 5	9	36.00
LN 1 & 2 DIMR2	H0	288	12.9 ± 0.2	55 ± 7	3.9 ± 0.1	-0.03 ± 0.0	492 ± 1	36	12.50
	H1	558	16.9 ± 0.2	69 ± 19	2.9 ± 0.1	0.63 ± 0.0	448 ± 1	57	10.22
	H2	295	23.6 ± 0.7	368 ± 88	2.5 ± 0.1	1.24 ± 0.0	385 ± 1	38	12.88
	H3	3	129.7 ± 51.7	2319 ± 1213	0.9 ± 0.2	5.42 ± 1.1	227 ± 27	0	0.00
LN 3 DIMR1	H0	24	14.7 ± 0.6	55 ± 20	4.2 ± 0.4	0.21 ± 0.1	471 ± 5	2	8.33
	H1	54	19.8 ± 0.5	62 ± 12	3.7 ± 0.2	0.73 ± 0.0	428 ± 2	6	11.11
	H2	39	24.0 ± 0.5	42 ± 9	3.6 ± 0.2	1.25 ± 0.0	378 ± 1	5	12.82
	H3	18	36.0 ± 1.7	303 ± 78	5.4 ± 0.4	1.46 ± 0.1	361 ± 5	13	72.22
LN 3 DIMR2	H0	24	15.3 ± 0.7	52 ± 12	3.6 ± 0.2	0.27 ± 0.0	470 ± 3	4	16.67
	H1	17	19.2 ± 0.9	58 ± 13	3.5 ± 0.3	0.66 ± 0.0	428 ± 3	16	29.41
	H2	25	21.1 ± 0.8	40 ± 8	3.1 ± 0.2	1.11 ± 0.0	387 ± 3	2	8.00
	H3	18	32.3 ± 1.3	257 ± 43	4.5 ± 0.4	1.40 ± 0.1	376 ± 5	10	55.56

Conductivity Index variables tended to contribute more to the model than Electrical Conductivity (Table 2). In the validation dataset, the R² increased to about 0.80 and the correlation to 0.91 (Table 3). In this dataset, the MLP networks had a higher R² compared to the other ones. Among the networks, MLP 2 had the best results in predicting CMI, followed by the MLP 1, Jordan, Elman and linear regression networks.

Better results in estimating CMI could be obtained if more data were available for training the network. When SCC, fat and protein values are not available and only EC measures can be recorded, it could be possible to predict CMI from EC, IQR and CI. This method could prove to be a useful decision tool to the dairy farmer who needs to have instant data of cows for management decisions, such as preventive treatment or culling animals from the herd.

Table 2. Training dataset of the methods used to predict CMI

Prediction Method	SEM	R ²	r	Relative Factor Contribution		
				EC	IQR	CI
MLP 1	33.25	0.742	0.86	0.25	0.37	0.38
MLP 2	33.43	0.741	0.86	0.33	0.35	0.32
Elman	34.03	0.736	0.86	0.22	0.42	0.37
Jordan	34.01	0.736	0.86	0.21	0.32	0.31
Linear regression	35.94	0.722	0.85			

Table 3. Validation dataset of the methods used to predict CMI

Prediction Method	SEM	R ²	r
MLP 1	33.68	0.826	0.92
MLP 2	31.85	0.836	0.92
Elman	39.47	0.796	0.92
Jordan	37.89	0.805	0.92
Linear regression	35.94	0.722	0.85

ACKNOWLEDGEMENTS

This research was funded by Fonterra Co-Op Group Ltd. Special acknowledgements for the people that helped carry out the sampling, the Lincoln University dairy farm staff, Livestock Improvement Corporation Ltd, Dexcel and Shoof International.

REFERENCES

- [1] M.J. Auld and I.B. Hubble, "Effect of mastitis on raw milk and dairy products," *The Australian J. Dairy Tech.*, vol. 53, pp. 28-36. 1998.
- [2] C. Fourichon, H. Seegers, N. Bareille and F. Beaudeau, "Effects of disease on milk production in the dairy cow: a review," *Preventive Veterinary Medicine*, vol. 41, pp. 1-35. 1999.
- [3] M.E. Kehrli and D.E. Shuster, "Factors affecting milk somatic cells and their role in health of the bovine mammary gland," *J. Dairy Sci.*, vol. 77, pp. 619-27. 1994.

- [4] T. Kohonen, *Self Organizing Maps*. New York, USA: Springer Verlag. 1995.
- [5] M.G. López-Benavides, S. Samarasinghe and J.G.H. Hickford, "The use of combined milk traits to determine mastitis status using artificial neural networks," unpublished.
- [6] R.A. Mrode and G.J.T. Swanson, "Genetic and statistical properties of somatic cell count and its suitability as an indirect means of reducing the incidence of mastitis in dairy cattle," *Animal Breeding Abstracts*, vol. 64, pp. 847-57. 1996.
- [7] M. Nielsen, H. Deluyker, Y.H. Schukken and A. Brand, "Electrical conductivity of milk: measurement, modifiers, and meta analysis of mastitis detection performance," *J. Dairy Sci.*, vol. 75, pp. 606-14. 1992.
- [8] M. Nielsen, Y.H. Schukken, A. Brand, S. Haring and R.T. Ferwerda-van Zonneweld, "Comparison of analysis techniques for on-line detection of clinical mastitis," *J. Dairy Sci.*, vol. 78, pp. 1050-61. 1995.
- [9] J.C. Principe, N.R. Euliano and W.C. Lefebvre, *Neural and Adaptive Systems*. New York, USA: John Wiley & Sons, Inc. 2000.
- [10] K.R.S. Rao, "Mastitis milk - physical and chemical tests for detection," *Indian J. Dairy & Bio-Sciences*, vol. 8, pp. 57-60. 1997.
- [11] X.Z. Yang, R. Lacroix and K.M. Wade, "Investigation into the production and conformation traits associated with clinical mastitis using artificial neural networks," *Canadian J. Animal Sci.*, vol. 80, pp. 415-425. 2000.

APPENDIX

1. Inter-Quarter Ratio (IQR)

$$IQR_i = \frac{\sum_{j=1}^4 EC_j}{EC_i} \quad \text{where } i = \text{quarter number} \\ j = \text{iteration number for quarter}$$

2. Conductivity Index (CI)

$$CI_i = 2 + \left[\left(\frac{EC_i}{100} \right) - IQR_i \right]$$

3. Linear Score (SCS)

$$SCS_i = \log_2 \left(\frac{SCC_i}{100} \right) + 3$$

4. Composite Milk Index (CMI)

$$CMI_i = IQR_i + FP_i + PP_i + SCS_i + CI_i + A_i + B_i + C_i + D_i + E_i$$

where:

$$A_i = \ln (PP_i * FP_i); B_i = (CI_i * SCS_i); C_i = (CI_i * FP_i); D_i = (CI_i * PP_i); E_i = (A_i * B_i)$$