

# Detection of dairy cattle Mastitis: modelling of milking features using deep neural networks

**M. Khamaysa Hajaya**<sup>a</sup>, **S. Samarasinghe**<sup>b</sup>, **D. Kulasiri**<sup>c</sup> and **M. Lopez Benavides**<sup>d</sup>

<sup>a</sup> Assistant Professor, Department of Computer Science, Mutah University, Jordan,

<sup>b</sup> Professor & Head, Complex Systems, Big Data & Informatics Initiative (CSBII), Lincoln University, New Zealand,

<sup>c</sup> Professor & Head, Centre for Advanced Computational Solutions (C-fACS), Lincoln University, New Zealand,

<sup>d</sup> Technology Manager Milk Quality Animal Health, DeLaval, USA  
Email: [sandhya.samarasinghe@lincoln.ac.nz](mailto:sandhya.samarasinghe@lincoln.ac.nz)

**Abstract:** Dairy cattle Mastitis is one of the most notable and costly diseases in dairy industry worldwide. The total Mastitis cost to dairy industry in New Zealand is up near \$280 million a year; this cost includes drop in milk production, cattle treatment and other costs. This research includes the examination and analysis of data collected from a commercial robotic dairy farm, in order to design and build a computational model that can help efficient and accurate detection of Mastitis in dairy cattle herds.

Accurate Mastitis detection helps cut treatment costs, control the disease, retain milk production levels and maintain milk quality grade. In addition to cutting financial costs, efficient detection helps cows by protecting them and relieving pain caused by the disease. Computational models can help achieve these by helping farmers to adopt timely and suitable cattle treatment regime, and by preventing healthy cows from being infected.

For this study, robotic data have been collected for 12 months from a barn-based dairy farm at Makikihi in South Canterbury - New Zealand. At data collection time, that farm was the largest dairy farm in the world in terms of the number of milking-robots under one roof (24 milking robots in one barn). The collected dataset contains sensor data of more than 1,900 cows being milked more than 1.1 million times during the time of data collection.

Having about 29,000 milking instances fully labelled (healthy/sick), a deep neural network (DNN) was used to build, train and validate a classification model using variable combinations, including variables that have not been studied before. The model has shown the ability to perform detection tasks with a high level of accuracy; with Specificity (Sp) of 99%, and Sensitivity (Se) of 97%. With this high and stable Sp, and the relatively high Se, the proposed model avoids the problem of false positive alerts.

By using deep neural networks to build a Mastitis detection model, this study exploits the main characteristic that gives deep learning predominance compared with other techniques - representation learning, which means that the trained models can extract patterns that used to be ignored by other techniques, to present a robust definition of Mastitis, using real-world sensor data, generated by milking robots in a commercial dairy farm, including data for previously unexploited features. The results of this study allow viewing dairy cattle Mastitis detection from a different angle, which brings about a broader understanding of some of the signs and symptoms of Mastitis, leading to better control and management of the disease.

**Keywords:** Mastitis, deep neural networks, dairy, keras

## 1. INTRODUCTION

### 1.1. Cattle Mastitis

Mastitis is considered as one of the most prominent diseases in dairy industry worldwide, as described by Viguier *et al.* (2009) and Petrovski *et al.* (2006). This disease severely affects dairy cattle and results in costly bacterial treatment and causes a significant decrease in milk produced from sick cows. Mastitis also may cause other diseases that may affect the herd as a result of the bacterial infection, according to Halasa *et al.* (2007). The direct outcomes of Bovine Mastitis, according to Viguier *et al.* (2009), include: significant drop in milk production (permanent or temporary), low-grade milk quality due to the alteration in some components in milk, which leads to reduction in milk price, discarding milk produced from cows being treated by antibiotic, cattle treatment costs, cost of additional labour, additional costs of milk quality/ disease status monitoring, decrease of fertile-life of cattle, low quality/ price of cattle meat.

The National Mastitis Council defines Bovine Mastitis (Mast = breast, itis = inflammation) as the inflammation of the mammary gland (udder tissue) in cattle; this inflammation is a response to bacterial, chemical, thermal or physical injury. In dairy cattle, Mastitis is always caused by pathogens (bacteria) that attack the milk-producing tissues in the udder and produce toxins that immediately cause damage in the udder, as illustrated by Jones and Bailey (2009).

### 1.2. Background

At the present time, retaining high profit margins has lead dairy industry toward larger dairy farms, as stated by Huybrechts *et al.* (2014), and with the lack of skilled dairy farm labourers, Automatic Milking Systems (AMS) have been widely adopted by dairy farmers over the past few years to replace traditional farm labourers. Modern AMSs, like VMSs (Voluntary Milking Systems), are standardly equipped with embedded sensor-based Mastitis detection tools.

As shown in Rutten *et al.* (2013), among more than 35 published studies related to automatic Mastitis detection modelling, Se values have ranged from 55% to 89%, and Sp values have ranged from 56% to 99%.

Sensitivity (Se) and Specificity (Sp) are statistical indicators which are usually used to measure the performance of binary classification; (Se) measures the proportion of positive cases (sick cows) that were correctly detected by a model, while (Sp) measures the proportion of negative cases (healthy cows) that were correctly identified as such.

The performance of the sensor based Mastitis detection systems could be improved significantly by decreasing the number of false positive cases, as problem of false alerts is the main problem with these systems. This improvement could be achieved by two main means: improvement of the performance of sensors themselves, and improvement of the performance of models that are used to transform data from sensors into automatic Mastitis detection alerts, as discussed by Hogeveen *et al.* (2010).

This study examines a large-scale database, which allows flexible modelling, with Mastitis cases being well documented across 1,900 cows. The farm system keeps recording all milking readings for sick cows, even after clinical Mastitis is diagnosed and the sick cow's milk is diverted, which is considered priceless for Mastitis detection modelling.

While neural networks Mastitis detection models presented in previous studies were built using data collected from research dairy farms (conventional or robotic), and while most of previous studies were carried out using small herds (100-400 cows), this study introduces a model that is built and tested against data collected from a large-scale commercial dairy farm, which makes the produced model more realistic and representational.

### 1.3. Deep neural networks

A deep neural network (DNN) is described as a neural network that has three or more layers. Considering this definition, DNN is not a new concept, as this type of neural network architectures, like Multi-Layer Perceptron, was introduced a long time ago. However, multilayer neural networks became popular as deep neural networks after the approach of deep training for these networks was introduced by Hinton in 2006. This approach combines both unsupervised and supervised learning; it includes unsupervised feature extraction from different layers, followed by supervised training to tweak and optimize the model performance, as described and discussed by Hinton *et al.* (2006).

Deep neural networks are superior compared to "shallow" neural networks in two respects; the number of layers in DNN is higher (three layers or more) - this advantage allows large DNNs to learn complex features

from big data in short time, in other words, the many hidden layers in a DNN helps extracting features. In shallow neural networks, well-defined features have to be provided to the network so it can learn to map these features to outputs during the training phase; usually these features are designed and hand-crafted by humans. In DNNs, features are extracted and learned automatically by the network in the unsupervised training phase; this can improve the network performance, minimize human intervention and minimize learning time, as illustrated in Goodfellow *et al.* (2016). The second advantage of DNNs is related to the network activation function. Generally, in artificial neural networks, the activation function is the function that defines outputs based on a set of given inputs. In other words, the activation function is responsible for transferring inputs to outputs in a network. Rectified Linear Unit (ReLU) activation function is the function applied to inputs transferred through hidden layers in DNNs. ReLU activation function is applied to data in every hidden layer, and the outputs of the function are passed to the next hidden layer and so on. Equation (1) shows the form of ReLU activation function:

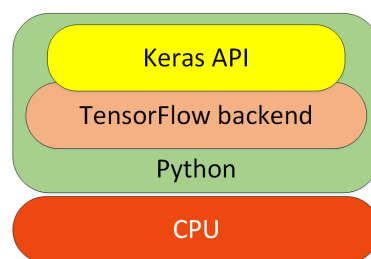
$$f(x)=\max(0, x) \tag{1}$$

ReLU activation function is widely used in DNNs due to its high performance in classification, which was reported by many research studies according to Lewis (2016). ReLU activation function is used in the proposed DNN model in this study; the model itself is described later in section 3.

In DNN training, an optimizer updates the network's parameters based on the loss function. The proposed DNN model uses ADAM optimizer which was introduced by Kingma and Ba (2014). ADAM is an SGD (Stochastic Gradient Descent) variant optimizer.

Keras API (Application Programming Interface) was used to build, train and validate the proposed DNN model within Python. Keras is a high-level framework for building deep neural networks, as described by Chollet (2015).

Keras works on top of TensorFlow, which is a neural network manipulation library that serves as the backend engine of Keras. Figure 1 illustrates the relationship between Keras and the other elements of the development environment.



**Figure 1.** the relationship between Keras, TensorFlow, Python and the hardware

## 2. FARM DATA

Research data for this paper were collected from Van Leeuwen Dairy Farm; a commercial Voluntary Milking System (VMS) dairy farm located in Makikihi in South Canterbury - New Zealand. This robotic dairy farm is the largest dairy farm in the world in terms of the number of milking robots (VMS units) under one roof. It has 24 DeLaval robots, milking about 1,900 cows in a single 13,000 square-meter barn. This high-tech farm started operation in late 2014. DeLaval VMS units are described in Heidelberg (2015).

Data collection started in June 2016, and ended in June 2017 (370 days). During that period, the total number of recorded milking instances was more than 1.1 million. The herd size was more than 1,900 cows, which is relatively very large compared to previous studies (few hundreds). This large-scale database allows extracting a reasonably large research dataset with a good number of sick milking instances (28%).

With a dataset of more than 1.1 million milking instances, data which are relevant to Mastitis detection need to be labelled and extracted. The farm treatment record contains the treatment log of different illnesses, including clinical Mastitis. This log was used to identify treated cows' IDs, and treatment dates for clinical Mastitis.

### 2.1. Data slicing

Figure 2 shows how data were sliced into sick days (the red part of the bar), and healthy days (the green part of the bar). For each clinical Mastitis case, 38 days of milking were sliced, in order to be examined and analysed. The 38 days consist of the following:

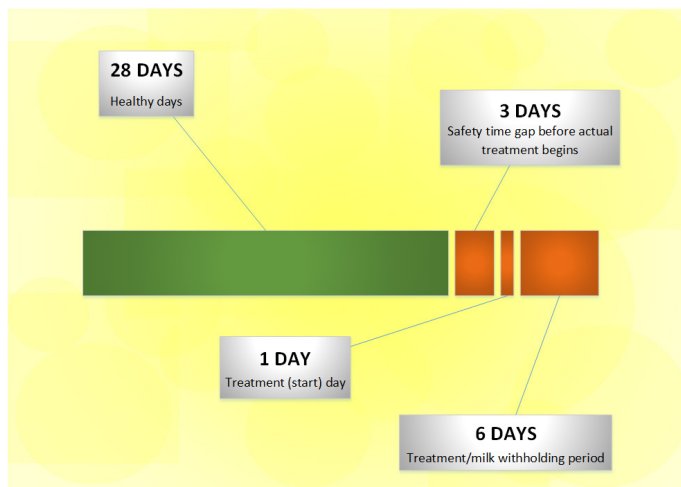
#### *Sick days*

Sick days include treatment start day (1 day), in addition to 3 days before that day as a safety time gap, this time gap covers any potential delay between the appearance of clinical symptoms and the application of the actual treatment. 6 days after the treatment start day are also counted as sick days- these are the treatment (milk

withholding) days, as in nearly all cases, no clinical Mastitis treatment lasts more than 6 days. It is worth mentioning that cows under treatment continues being milked normally with all milking data being collected as usual, but with produced milk being diverted to a different tank. Thus, 10 days are the total number of sick days.

### **Healthy days**

These are the days before sick days, which will be used to define negative milking instances. Deciding the number of these days was based on three factors; the total number of sick cows within the herd during data collection time, the total number of milking instances for each sick cow before her treatment begins and the number of sick milking instances for each sick cow. After trying different scenarios, going back 28 days before sick days begin (31 days before treatment day) was found to be the best scenario. Going back 4 weeks before each treatment is long enough, while preserving a reasonable sick-to-healthy ratio (2:5), which means that more than 28% of total milking instances analysed are sick instances.



**Figure 2.** Data slicing

Slicing milking instances of 38 days for each clinical Mastitis case resulted in about 29,000 milking instances (28,962 milking instances), more than 8,000 of them are associated with sick days, and more than 20,000 of them relate to healthy days. Extracting (slicing) healthy and sick days from the main dataset this way guarantees a balanced sick-to-healthy ratio that provides enough number of sick cases, while preserving a reasonable overall dataset size for DNN modelling.

## **2.2. Selected milking features**

### **Milk conductivity**

Milk conductivity is the ability of milk to conduct electrical current; it is measured in S/m<sup>3</sup> (Siemens per cubic meter) unit. Basically, as the disease progresses, blood-milk barrier is broken, and blood enters milk, thereby increasing ionic content in milk, which increases milk conductivity.

### **Milk yield**

Milk yield is the amount of milk produced by the end of a milking, measured in kg. 1 kg of milk approximately equals 1 litre of milk.

### **Last milking interval**

At VMS farms, the cows go to milking units voluntarily, as these milking robots are available 24/7. At every milking, the elapsed time between last milking and current milking is recorded. Last milking interval is measured in hours. The correlation between this variable and Mastitis has not been computationally analysed before. A longer last milking interval can be indicative of pain caused by Mastitis.

### **Milking duration**

At every milking, the milking duration (in minutes) is recorded. Milking duration is the actual time of milk suction; it doesn't include overhead time before or after the actual milk flow. The correlation between milking duration and Mastitis has not been computationally analysed in previous studies, while the herd average milking duration (min/cow) was used in some studies, like Mammadova and Keskin (2015). Milking duration can be considerably shorter for cows with Mastitis due to alterations in milk characteristics, such as having flakes and/ or clots in milk.

### 2.3. Data pre-processing

Some of the collected data variables were measured per quarter for each cow during each milking; these variables needed to be processed to be more discriminative for Mastitis detection.

**Table 1.** Pre-processed/ modelling-ready variables

Variable	Description
n_cond	Normalized milk conductivity (total of 4 quarters)
n_yield_dev	Normalized mean absolute deviation of 4 quarters milk yield
n_lmi	Normalized last milking interval
n_duration	Normalized milking duration (the duration of the whole milking)

Table 1 lists the modelling-ready variables that are used in designing the classification model for this study.

As the range of values for each variable varies, variables need to be scaled (normalized), so each variable contributes proportionately to different modelling stages. Equation (2) shows the min-max normalization, which was used to normalize different variables before starting modelling.

$$x = \frac{x - \min(x)}{\max(x) - \min(x)} \quad (2)$$

Where  $x$  is a value of a variable, and  $x$  is the normalized value between 0 and 1.

Since Mastitis affects the readings of some quarters while leaving other quarters giving normal readings for the same animal during the same milking, calculating the mean absolute deviation for the quarter-based variable (milking yield) results in more discriminative features that have more correlation with Mastitis.

Equation (3) calculates the mean absolute deviation for milk yield within the same cow, where  $y_k$  is the milk yield of quarter  $k$ .

$$\text{yield\_dev} = \frac{1}{4} \sum_{k=1}^4 \text{abs}(y_k - \text{avg}(y_1, y_2, y_3, y_4)) \quad (3)$$

As shown in Table 1, mean absolute deviation was not calculated for milk conductivity ( $n\_cond$ ), and the reason is because the VMS machine measures electrical conductivity for the whole milk produced in each milking (not for each quarter separately), so there was no need to calculate any deviation for that variable.

### 3. DNN MODELLING

Figure 3 illustrates the topology of the DNN Mastitis classifier, which is used in classifying clinical Mastitis cases.

This network consists of 3 fully-connected hidden layers, in addition to the output layer. The first hidden layer has 8 neurons, the second hidden layer has 16 neurons and the third hidden layer has 8 neurons. The output layer has one neuron since this classifier is built to perform binary classification tasks, i.e. healthy/sick classification. Different scenarios have been tried regarding the number of hidden layers in the network, and they have shown that adding more than 3 layers does not give any significant improvement in the network performance.

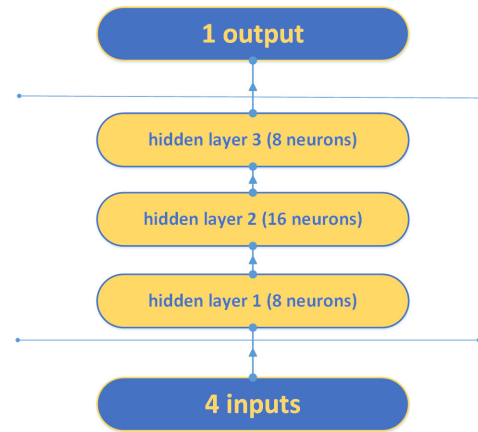
ReLU activation function is used in all three hidden layers, while sigmoid activation function is used in the output layer, which is ideal for binary classification (sick/healthy), as sigmoid activation function monotonically varies from 0 and 1 and is suited to representing a simple threshold response based on nonlinear relations.

For modelling, 60% of the corresponding dataset were used for training, and the remaining 40% were used for validation. Normally, 70:30% split or 80:20% split are used for training/validation, but the relatively large size of the current dataset allows allocating more data for the validation part, which enhances the validation process.

#### 4. MODEL RESULTS

Table 2 demonstrates the model performance (validation results) through different combinations of variables: two-variable combinations, three-variable combinations, and finally, combining all four variables together in one combination for training the DNN. Training and validating the DNN using different scenarios (combinations) have shown that **n\_cond** (milk electrical conductivity) and **n\_duration** (actual milking duration) are the two most influential variables among the four variables analysed, in terms of discriminating sick milking instances. The next most influential variable is **n\_yield\_dev** (milk yield deviation), which has shown a good discriminating capability but less than the above mentioned two variables. While **n\_lmi** (last milking interval) comes fourth in terms of discriminating power among other variables, combining all four variables and training the DNN using that combination have shown the most robust and stable performance in validation, in terms of detecting sick milking instances correctly (Se of 97%), and detecting healthy milking instances correctly (Sp of 99%).

Other combinations have varied in performance, with Sp of 97-99%, and Se of 78-97%, this variation is due to the number of variables used, and also the discriminating power of those variables.



**Figure 3.** The topology of the DNN classification

**Table 2.** Model results using different variable combinations

Combination	Sp	Se
n_cond   n_yield_dev	99%	85%
n_cond   n_lmi	98%	81%
n_cond   n_duration	99%	88%
n_yield_dev   n_lmi	97%	78%
n_yield_dev   n_duration	99%	83%
n_lmi   n_duration	98%	80%
n_cond   n_yield_dev   n_lmi	99%	91%
n_cond   n_yield_dev   n_duration	99%	94%
n_cond   n_lmi   n_duration	99%	92%
n_yield_dev   n_lmi   n_duration	98%	89%
<b>n_cond   n_yield_dev   n_lmi   n_duration</b>	<b>99%</b>	<b>97%</b>

This study has introduced a highly accurate Mastitis detection classification model using deep neural networks; the proposed model is capable of detecting clinical Mastitis cases with Sp of up to 99% and Se of up to 97%. While meeting ISO (International Standards Organization) minimum limit of Sp (99%) and Se (80%) for Mastitis detection systems, and while outperforming other neural network models that were proposed in previous studies, such as Hassan *et al.* (2008), and Sun *et al.* (2009), the proposed model has avoided the problem of false positive alerts, while maintaining a very good capability of detecting clinical cases within the herd.

The proposed model has validated the discriminating power of variables that were previously used in Mastitis detection modelling, namely, milk electrical conductivity and milk yield. This model has also explored and utilized variables that have not been used before in computational Mastitis detection, namely, last milking interval and actual milking duration.

The superior results of the proposed DNN model suggest that representation learning can significantly enhance Mastitis detection. Representation learning means that a model solves a problem while learning how to represent that problem, and due to the expansive structure of the deep neural network, the learned representation includes the nuances and intricacies of the problem, and this seems to have enhanced Mastitis detection.

## ACKNOWLEDGMENTS

Collecting data for this study was carried out with the help of the technical support at DeLaval, given by Adrian Garner. The authors acknowledge contributions from Mr. Garner and other persons at Van Leeuwen Dairy Farm in Makikihi, Canterbury, New Zealand.

## REFERENCES

- Chollet, François. (2015). *Keras*. 2015, <https://keras.io/>.
- Goodfellow, Ian, Yoshua Bengio, and Aaron Courville. (2016). *Deep Learning*. The MIT Press, 2016, <http://www.deeplearningbook.org/>.
- Halasa, T., K. Huijps, O. Østerås, and H. Hogeveen. (2007). “Economic Effects of Bovine Mastitis and Mastitis Management: A Review.” *Veterinary Quarterly*, vol. 29, no. 1, pp. 18–31, doi:10.1080/01652176.2007.9695224.
- Hassan, K. J., S. Samarasinghe, and M. G. Lopez-Benavides. (2008). “Use of Neural Networks to Detect Minor and Major Pathogens That Cause Bovine Mastitis.” *Journal of Dairy Science*, vol. 92, pp. 1493–99, doi:10.3168/jds.2008-1539.
- Heidelberg, Roy L. (2015). “The Power of Knowing the Rules - DeLaval Online Cell Counter OCC for VMS.” *American Review of Public Administration*, pp. 1–17, doi:10.1177/0275074015575353.
- Hinton, Geoffrey E., Simon Osindero, and Yee-Whye Teh. (2006). “A Fast Learning Algorithm for Deep Belief Nets.” *Neural Computation*, vol. 18, no. 7, pp. 1527–54, doi:10.1162/neco.2006.18.7.1527.
- Hogeveen, Henk, Claudia Kamphuis, Wilma Steeneveld, and Herman Mollenhorst. (2010). “Sensors and Clinical Mastitis-the Quest for the Perfect Alert.” *Sensors*, vol. 10, no. 9, pp. 7991–8009, doi:10.3390/s100907991.
- Huybrechts, T., K. Mertens, J. De Baerdemaeker, B. De Ketelaere, and W. Saeys. (2014). “Early Warnings from Automatic Milk Yield Monitoring with Online Synergistic Control.” *Journal of Dairy Science*, vol. 97, no. 6, Elsevier, pp. 3371–81, doi:10.3168/jds.2013-6913.
- Jones, G. M., and T. L. Bailey. (2009). “Understanding the Basics of Mastitis.” *Virginia Cooperative Extension*, vol. 404, no. 233, pp. 1–5, doi:10.1108/02621711111098406.
- Kingma, Diederik P., and Jimmy Ba. (2014). *Adam: A Method for Stochastic Optimization*. pp. 1–15, doi:<http://doi.acm.org.ezproxy.lib.ucf.edu/10.1145/1830483.1830503>.
- Lewis, Nigel Da Costa. (2016). *Deep Learning Made Easy with R: A Gentle Introduction for Data Science*. 2016.
- Mammadova, Nazira M., and Ismail Keskin. (2015). “Application of Neural Network and Adaptive Neuro-Fuzzy Inference System to Predict Subclinical Mastitis in Dairy Cattle.” *Indian Journal of Animal Research*, vol. 49, no. OF, pp. 671–79, doi:10.18805/ijar.5581.
- Petrovski, K. R., M. Trajcev, and G. Buneski. (2006). “A Review of the Factors Affecting the Costs of Bovine Mastitis.” *Journal of the South African Veterinary Association*, vol. 77, no. 2, pp. 52–60.
- Rutten, C. J., a G. J. Velthuis, W. Steeneveld, and H. Hogeveen. (2013). “Invited Review: Sensors to Support Health Management on Dairy Farms.” *Journal of Dairy Science*, vol. 96, no. 4, Elsevier, pp. 1928–52, doi:10.3168/jds.2012-6107.
- Sun, Zhibin, Sandhya Samarasinghe, and Jenny Jago. (2009). *Detection of Mastitis and Its Stage of Progression by Automatic Milking Systems Using Artificial Neural Networks*. doi:10.1017/S0022029909990550.
- Viguier, Caroline, Sushrut Arora, Niamh Gilmartin, Katherine Welbeck, and Richard O’Kennedy. (2009). “Mastitis Detection: Current Trends and Future Perspectives.” *Trends in Biotechnology*, vol. 27, no. 8, pp. 486–93, doi:10.1016/j.tibtech.2009.05.004.