A large-scale study of indicators of sub-clinical mastitis in dairy cattle by attribute weighting analysis of milk composition features: highlighting the predictive power of lactose and electrical conductivity

Esmaeil Ebrahimie[1,2,3,4]*, Faezeh Ebrahimi[5], Mansour Ebrahimi[5], Sarah Tomlinson[6] and Kiro R Petrovski[6]

[1] School of Medicine, The University of Adelaide, Adelaide 5005, Australia
[2] Division of Information Technology, Engineering & Environment, School of Information Technology and Mathematical Sciences, University of South Australia, Adelaide, Australia
[3] Faculty of Science and Engineering, School of Biological Sciences, Flinders University, Adelaide, Australia
[4] Institute of Biotechnology, Shiraz University, Shiraz, Iran
[5] Department of Biology, University of Qom, Qom, Iran
[6] School of Animal and Veterinary Sciences, The University of Adelaide, Adelaide 5371, Australia

Sub-clinical mastitis (SCM) affects milk composition. In this study, we hypothesise that large-scale mining of milk composition features by pattern recognition models can identify the best predictors of SCM within the milk composition features. To this end, using data mining algorithms, we conducted a large-scale and longitudinal study to evaluate the ability of various milk production parameters as indicators of SCM. SCM is the most prevalent disease of dairy cattle, causing substantial economic loss for the dairy industry. Developing new techniques to diagnose SCM in its early stages improves herd health and is of great importance. Test-day Somatic Cell Count (SCC) is the most common indicator of SCM and the primary mastitis surveillance approach worldwide. However, test-day SCC fluctuates widely between days, causing major concerns for its reliability. Consequently, there would be great benefit to identifying additional efficient indicators from large-scale and longitudinal studies. With this intent, data was collected at every milking (twice per day) for a period of 2 months from a single farm using in-line electronic equipment (346 248 records in total). The following data were analysed: milk volume, protein concentration, lactose concentration, electrical conductivity (EC), milking time and peak flow. Three SCC cut-offs were used to estimate the prevalence of SCM: Australian ≥ 250 000 cells/ml, European ≥200 000 cells/ml and New Zealand ≥ 150 000 cells/ml. At first, 10 different Attribute Weighting Algorithms (AWM) were applied to the data. In the absence of SCC, lactose concentration featured as the most important variable, followed by EC. For the first time, using attribute weighted modelling, we showed that the concentration of lactose in milk can be used as a strong indicator of SCM. The development of machine-learning expert systems using two or more milk variables (such as lactose concentration and EC) may produce a predictive pattern for early SCM detection.

Keywords: Attribute Weighting, Expert system, Machine Learning, Mastitis, Milk Composition

Inflammation of the bovine udder, known as mastitis, is often associated with bacterial intra-mammary infections (IMI), and is subdivided into clinical (visible signs of inflammation of the udder or milk) and subclinical (inflammation without visible signs) mastitis. Mastitis is the most widespread and economically challenging disease of dairy cattle (Petrovski et al. 2006), as it can directly affect milk production and quality, costing up to £277 per infected cow (Halasa et al. 2007) and may affect up to 70% of a given herd (Bradley et al. 2007). In addition, it can be fatal for some infected cows (Addis et al. 2016).

*For correspondence; e-mail: esmaeil.ebrahimie@adelaide.edu.au

Clinical mastitis (CM) is classified as mild, moderate or severe and can be detected by simple observatory tests. Cows may show clinical signs such as fever, loss of appetite, dehydration, a swollen udder that is sensitive to touch, and reduced milk secretion with flakes, clots, and a watery appearance (Petrovski et al. 2006; Goncalves et al. 2016).

Subclinical mastitis (SCM) is up to 40 times more common than CM and far more difficult to detect, thus having a greater economic impact. To control SCM, accurate surveillance strategies are required. Somatic cell count (SCC) is currently the key SCM detection method (Bochniarz et al. 2016).

SCC primarily comprises white blood cells (leucocytes), including macrophages, lymphocytes and polymorpho-nuclear leucocytes (essentially neutrophils), which are produced by the cow's immune system in response to an infection. Upon bacterial invasion of the bovine mammary gland, leucocytes are recruited into the gland from the blood stream, increasing the SCC, measured in number of cells per mL of milk (Harmon, 1994). In a normal, healthy cow, SCC is around 70 000 cells/ml but when a cow has an IMI, it increases sharply. Since SCC increases with the severity of mastitis, it is used to indicate the IMI status at the time of sampling. Once SCC exceeds the selected cut-off (Australian cut-off ≥250 000 cells/ml, European Union cut-off ≥200 000 cells/ml, and New Zealand cut-off ≥150 000 cells/ml), a cow is considered to have an IMI.

SCC (among other milk quality parameters) is measured during routine herd testing. The dairy industry worldwide has adopted this test-day SCC as its main mastitis surveillance tool (Schukken et al. 2003; Sharma et al. 2011). However, simple test-day SCC may not be stable enough to accurately monitor SCM, as SCC widely fluctuates between test days (Berglund et al. 2007).

Frequent (longitudinal) monitoring of SCC (Berglund et al. 2007) and/or finding additional SCC-independent predictors (such as milk quality parameters), could increase the robustness and early predictive power of SCM. Some milk composition features have high potential for use as SCM predictors. Infected cows have a decreased milk yield (volume), and composition changes such as reduced fat, protein and lactose (Bartlett et al. 1990; Leitner et al. 2003; Petrovski et al. 2006). However, Gyr cows with SCM had reduced lactose and solids, only, while the protein and fat content of the milk appeared unaffected (Malek dos Reis et al. 2013). Thus, key predictors of SCM need to be reliably identified.

Key predictors of SCM can be identified through mining a large number of milking records, to develop a robust and reliable statistical model of feature selection. Data mining (machine learning) is the process of extracting implicit, previously unknown, and potentially useful information from data. It can also be defined as data analysis for finding regularities and patterns in a given data set. Data mining outperforms the common multivariate statistical methods in large-scale studies. Great attention has been paid to the application of data mining in finance, medicine, drug discovery, agriculture, and biology as it can create accurate risk models and discover patterns in data (Ebrahimi et al. 2010; Ebrahimi et al. 2011; Vinod Bharat et al. 2016). Data mining is extensively used in the medical field for disease detection, and assessing drug outcomes, due to its predictive abilities and pattern recognition (KayvanJoo et al. 2014; Hande Küçükönder, 2015). Supervised feature selection has been largely used in data mining to analyse complex data.

For this study, the goal of feature selection was to find a subset of input features of milking parameters by removing the ones with little or no predictive value (Shekoofa et al. 2011). In a large-scale, longitudinal study, we analysed a range of milk composition variables by 10 attribute weighting models to find those linked to SCC and SCM.

Materials and methods

Data collection

Data was collected between July 2011 and June 2013 from a split-calving, commercial dairy herd with a year-round average of 2400 mixed-age, Holstein Friesian cows milked twice daily. The farm, located in Ongaonga, Hawkes Bay, New Zealand, supplements a 40% pasture-based diet with 50% mixed rations (palm kernel extract, tapioca, triticale, barley, oak, straw, grass silage, corn silage, hay), 10% food processing by-products (including apples, beetroot, turnips, brewer's grain) and 5 g yeast per cow per day, to promote high milk production in its cows. Cows were regularly sourced from various family lines and origins at a production-driven, high replacement rate. No experimental control group was used in this study. However, the farm was chosen for its strict enforcement of standard operating procedures to minimise confounding variables. Cows were exposed to the same environmental and management conditions, following standard operating procedures.

An automated, electronic monitoring system was used to analyse milk for volume, weight, fat, protein, lactose, electrical conductivity, milking time, peak flow and SCC. SCC was measured by CellSense®, an on-line SCC detector that collects milk samples from individual cows approximately one minute after the start of milking, and mixes them with a reagent, CellGel™, using the same principle as the rapid mastitis test, to give SCC, as described by (Whyte et al. 2004). CellSense® measures SCC with reasonably good precision, showing an overall correlation coefficient of 0·76 when compared with the Fossomatic technique (CombiFoss 5000, Foss Electric, Hillerød, Denmark), the most commonly used laboratory measure for herd testing globally (Kawai et al. 2013).

Automatically collected data was first tabulated in Excel to enable statistical processing. Incomplete records or those with recording errors were identified and removed from the data set before analysis.

Description of features

The following variables were recorded at each milking: SCC, milk volume, fat, protein, lactose, Electrical Conductivity (EC), milking time, and peak flow. Fifteen bails of the farm's 60 bail rotary parlour (a one in four ratio) were fitted with CellSense®, allowing data capture of an average three and a half SCC measurements per cow, per week.

As repeated bacteriological cultures were not available for individual cows, the accepted SCC cut-offs of ≥150 000 cells/ml in New Zealand, ≥200 000 cells/ml in the European Union and ≥250 000 cells/ml in Australia, were used to classify a cow's mastitis status (0 – non-mastitis and 1 – mastitis). In other words, SCM occurrence was determined as binomial distribution (Yes/No) based on New Zealand, European, and Australian cut-offs.

Since infection status was not confirmed bacteriologically, cows classified as having mastitis (1) were described as having an assumed intramammary infection (IMI).

Preparation of data sets

Two types of data sets were created:

(1) *Data set containing SCC variable (DCS)*: Herein, the sampling score (Australian, European Union or New Zealand threshold/cut-off) was set as the target (label) variable. In this data set, SCC was set as the predictive variable.
(2) *Data set without SCC variable (DwSCC)*: In this data set, mastitis was set as the target (label) variable.

The data sets were imported into Rapid Miner software (RapidMiner 5.0.001, Rapid-I GmbH, Stochumer Str. 475, 44227 Dortmund, Germany) and either the sampling score (Australian, New Zealand, or European Union threshold/cut-off) was set as the output (label or target) variable, and other variables (with/without SCC) as dependent inputs.

Descriptive statistics, univariate, and multivariate analysis

Descriptive statistics, univariate, and multivariate analyses were performed using Minitab 17 package (http://www.minitab.com/). Mean comparison between milk composition variables was carried out with 2-Sample *T*-TEST. Comparison of two proportions was performed by *Z*-test and Fisher exact test. Spearman correlation was used to calculate the correlation between numeric variable and mastitis, as a binomial (Yes/No) variable. Two Proportion tests were performed using *Z*-test and Fisher's exact test.

Attribute weighting algorithms

To determine the most important variables, various attribute weighting algorithms were applied to each data set (Ebrahimi et al. 2011):

- *Information gain attribute weighting*: Herein, the relevance of a factor by computing the information gain in class distribution.

- *Weighting by the information gain ratio*: This operator calculates the relevance of the attributes based on information gain and assigns weights to them accordingly.
- *Weighting by rule*: This operator calculates the relevance of a factor by computing the error rate of a model on the sample data set without the factor.
- *Weighting by deviation*: The operator creates weights from the standard deviations of all attributes. The average normalised the values, minimum or maximum of the attribute.
- *Weighting by the chi-squared statistic*: This operator calculates the relevance of a factor by computing, for each attribute in the input sample data set, the value of the chi-squared statistic with respect to the class attribute.
- *Weighting by the Gini Index*: This operator calculates the relevance of a factor by computing the Gini Index of the class distribution, if the given sample data set would have been split per the factor in question.
- *Weighting by uncertainty*: This operator calculates the relevance of an attribute by measuring the symmetrical uncertainty with respect to the class.
- *Weighting by Relief*: This operator measures the relevance of a factor by sampling the examples and comparing the value of the current factor with the nearest example of the same, and of a different, class. The resulting weights are normalised into the interval between 0 and 1.
- *Weighting by Support Vector Machine (SVM)*: This operator uses the coefficients of the normal vector of a linear SVM as feature weights.
- *Weighting by Principal Component Analysis (PCA)*: This operator uses the factors of the first principal component as feature weights. Data is normalised before running the models, so it is reasonable to expect that all weights will be presented as a digit between 0 and 1; showing the importance of each attribute for the target attribute.

To enable comparison between results of different attribute weighting models, the calculated weights were normalised in range from 0 to 1. As previously described (Ebrahimi et al. 2014), we defined an index based on intersection (agreement) of different weighting models. The features that were identified as important by most weighting models (weights >0·5, >0·75, or >0·95) were assumed the key indicators of mastitis.

Results

Distribution of milk composition variables and SCM occurrence

The distribution of SCM was compared between the 3 different scoring methods. As expected, due to the higher level of acceptable SCC, scoring based on the Australian cut-off resulted in the lowest SCM prevalence (19%) and scoring based on New Zealand's low cut-off had the highest prevalence (35%) with the European cut-off intermediate (24%: online Supplementary File Fig. S1).
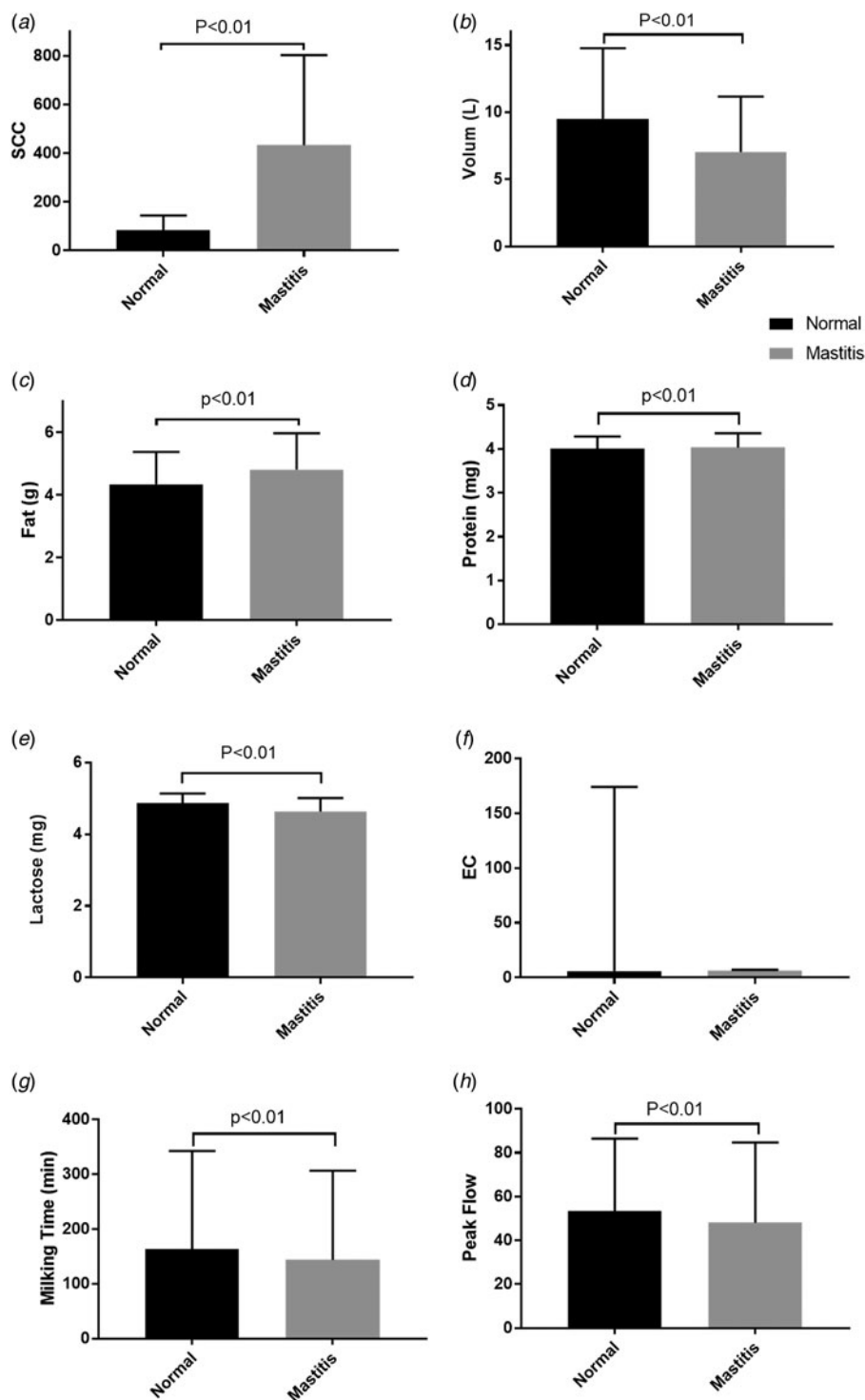
Fig. 1. Comparison of Somatic Cell Count (SCC), milking time, peak flow, electrical conductivity (EC), and milk composition variables between cows with and without Sub-Clinical Mastitis (SCM). (a) SCC, (b) Volume, (c) Fat, (d) Protein, (e) Lactose, (f) EC, (g) Milking Time, and (h) Peak Flow. SCM was determined by the Australian mastitis scoring system (cut-off ≥25 000 cells/ml). Two-Sample *t*-test was used for mean comparison.

Figure 1 shows the difference between SCC and milk composition variables between records with and without mastitis (using the Australian mastitis scoring system). As expected, cows with mastitis had significantly higher SCC ($83·9 \pm 60·6$ *vs.* $433·2 \pm 369·2$, $P < 0·01$) (Fig. 1). Mastitis significantly decreased lactose content, milking time (by

11·6%), peak flow (by 9·9%) and volume of milk by 26% (all $P < 0·01$). In contrast, fat and EC were increased ($P < 0·01$) (Fig. 1). While the mean comparison shows that mastitis increases the protein content of milk, this increase was not supported by median comparison (data not shown). Median of protein contents in cows with and without mastitis was 4·00 g. Milk EC was highly variable within cows without mastitis (Fig. 1).

Attribute weighting analysis of mastitis data set containing SCC variable (DCS)

Table 1 presents the normalised weights for different measured features in this study with respect to mastitis (based on Australian cut-off) with SCC included as a feature. Attribute weighting results applied to data scored by the Australian cut-off, showed that the SCC variable gained the highest possible weights of 1·0 by 9 out of 10 weighting algorithms (PCA, SVM, Uncertainty, Gini Index, Chi Squared, Deviation, Rule, Info Gain Ratio, and Info Gain) and 0·65 by Relief method (Table 1). Other variables did not attain weights in the presence of SCC. When data with Australian and New Zealand scoring systems were analysed, the same results were derived, and the SCC variable was the sole important feature (Online Supplementary FileFigure S1).

Attribute weighting analysis of mastitis data set without SCC variable (DwSCC)

When the SCC variable was omitted from the variable list and the Australian mastitis scoring was appointed as the target or label variable, lactose was promoted to first place, followed by EC (Table 2). When analysing data with Australian mastitis scoring, three attribute weighting algorithms including Uncertainty, Chi Squared, and Info Gain Ratio gave the highest possible value of 1·0 to lactose (Table 2). EC received the weight of 1 by Gini Index and Info Gain algorithms. Milking time received two 1·0 weights by Gini Index and Info Gain (Table 2). In the data labelled with EU and New Zealand scoring, the same results were concluded, and again, lactose, EC, and milking time were the most important features (data not shown). Altogether, lactose content of milk is the key distinguishing attribute for mastitis prediction (Table 2), as identified by most of the attribute weighting models (intersection of different models).

Lactose content of milk: negative indicator of mastitis occurrence

Attribute weighting models selected lactose as the most important negative indicator of mastitis (when SCC variable was omitted from the variable list). We determined the frequency of mastitis at different concentrations of milk lactose (online Supplementary File Table S1). When the lactose content of milk decreased, the frequency of mastitis

Table 1. Results of 10 Attribute Weighting Algorithms on dataset with SCC feature in respect to mastitis prevalence (based on Australian cut-off of ≥250 000 cells/ml as a target variable). The number of weights higher than 0·5, or 0·95 allocated by each model is presented as index of importance

| Attribute | Weight_PCA | Weight_SVM | Weight_Relief | Weight_Uncertainty | Weight_Gini Index | Weight_Chi Squared | Weight_Deviation | Weight_Rule | Weight_Info Gain Ratio | Weight_Info Gain | Number of models announced feature as important feature | | |
											Cutoff >0·5	Cutoff >0·75	Cutoff >0·9
SCC	1·00	1·00	0·65	1·00	1·00	1·00	1·00	1·00	1·00	1·00	10	9	9
Lactose	0·00	0·10	0·56	0·13	0·09	0·18	0·00	0·00	0·22	0·08	1	0	0
EC	0·00	0·00	0·00	0·00	0·20	0·00	0·67	0·00	0·20	0·20	1	0	0
Fat	0·00	0·08	1·00	0·05	0·03	0·09	0·00	0·00	0·00	0·03	1	1	1
Milking Time	0·07	0·01	0·03	0·00	0·00	0·00	0·79	0·00	0·00	0·00	1	1	0
Protein	0·00	0·00	0·22	0·01	0·00	0·01	0·00	0·00	0·15	0·00	0	0	0
Milk volume	0·00	0·01	0·64	0·09	0·05	0·08	0·02	0·00	0·10	0·06	1	0	0
Peak Flow	0·01	0·00	0·03	0·00	0·03	0·00	0·14	0·00	0·14	0·03	0	0	0

SCC, Somatic Cell Count; EC, Electric Conductivity

Table 2. Results of 10 Attribute Weighting Algorithms on data set without SCC feature in respect to mastitis prevalence (based on Australian scoring as target variable). The number of weights higher than 0·5, or 0·95 allocated by each model are presented as index of importance

Attribute	Weight_PCA	Weight_SVM	Weight_Relief	Weight_Uncertainty	Weight_Gini Index	Weight_Chi Squared	Weight_Deviation	Weight_Rule	Weight_Info Gain Ratio	Weight Info Gain	Number of models identifying feature as important Cut-off >0·5	Cut-off >0·75	Cut-off >0·90
Lactose	0·00	0·17	0·48	1·00	0·43	1·00	0·00	0·00	1·00	0·42	3	3	3
EC	0·00	0·00	0·00	0·00	1·00	0·00	0·86	0·31	0·89	1·00	4	4	2
Fat	0·00	1·00	1·00	0·38	0·14	0·52	0·00	0·00	0·01	0·16	3	2	2
Milking Time	1·00	0·73	0·00	0·03	0·02	0·01	1·00	0·00	0·00	0·02	3	2	2
Protein	0·00	0·18	0·23	0·06	0·00	0·06	0·00	1·00	0·67	0·00	2	1	1
Milk Volume	0·01	0·01	0·20	0·65	0·24	0·42	0·03	0·00	0·47	0·29	1	0	0
Peak Flow	0·01	0·06	0·01	0·00	0·14	0·00	0·18	0·29	0·63	0·17	1	0	0

EC, Electric Conductivity

occurrence sharply increased. Spearman correlation showed a significantly (*P*-value < 0·01) negative correlation (−26·8%) between lactose and mastitis (Australian mastitis scoring). The prevalence of mastitis differed greatly between lactose <4·2 mg/100 ml (average of 73·5% mastitis prevalence) and lactose >4·4 mg/100 ml (average of 17·3% mastitis prevalence: online Supplementary File Fig. S2). Fisher's exact test and *Z* tests showed this difference to be significant at $P < 0·01$.

Discussion

In this study, we analysed a range of milk composition variables by 10 attribute weighting models, to find the ones linked to SCM. SCM was detected by on-line SCC measurements and a variety of threshold/cut-offs (i.e. Australian, New Zealand and European Union). Lactose, EC and their combination were the most accurate variables to detect mastitis on dairy farms equipped with in-line sensors. Both forms of mastitis (CM and SCM) influence the quantity and quality of milk and therefore are of major economic concern for the dairy industry. Early detection of mastitis (especially SCM) can improve animal welfare and the quality of milk production, and increase economic gains. Data mining, also known as knowledge discovery, is a relatively new technique to extract useful knowledge from data (compared to traditional multivariate analysis) (Shekoofa et al. 2014). The main strength of data mining resides in its ability to handle large data sets and to extract meaningful and easy-to-understand results, which is usually impossible by human calculations (Torkzaban et al. 2015). A few studies have looked into this field to improve early detection techniques by novel data mining tools (Sharifi et al. 2018).

In recent years, the application of sensors to collect data during automatic milking, has generated a considerable amount of data which needs efficient pattern recognition models to mine the data (Kamphuis et al. 2008). Decision tree algorithms have been employed to predict CM and SCM from sensor data with promising outcomes (Kamphuis et al. 2010b; Ebrahimie et al. 2018). Support vector machine (SVM) methods have been used to predict SCM in dairy cattle by constructing and examining a prediction model with 89% sensitivity, 92% specificity, and 50% error in mastitis detection (Mammadova & Keskin, 2013). In a recent study, neural network algorithms have been used to classify healthy and mastitic buffaloes, based on yield and milk quality parameters (Panchal et al. 2016).

This is the first large-scale and longitudinal study employing 10 different attribute weighting models to distinguish healthy from mastitic dairy cattle based on quality and quantity of milking parameters. Ten different attribute weighting models (PCA, Chi Squared, Rule, Relief, Info Gain, Info Gain Ratio, SVM, Uncertainty, Deviation and Gini Index) were used to find the most important features associated with SCM. In line

with previous findings SCC is one of the best indicators of SCM (Mitchell et al. 1986).

Lactose was the most important predictor of mastitis after SCC. Prevalence of mastitis was statistically higher when milk lactose was low. EC was another important predictor of SCM. EC of udder tissue increases during mastitis because of changes in ionic composition of milk: increased Na + and Cl- levels and decreases in other mineral substances. Computerised herd management systems have been implemented by modern enterprises, allowing variables, such as milk yield, flow rate and EC, to be automatically recorded during milking. Then, based on EC, each individual cow's mastitis status is evaluated and an alarm raised to indicate mastitis. However, these alarms may be heard at the wrong time, and thus, the importance of this feature for CM and SCM detection has been debated previously (Kamphuis et al. 2010a). Interestingly, there was no report found showing the relationship between SCC or mastitis with lactose or EC. Our results showed that lactose and EC are the second and third most important feature, and were given high values by some attribute weighting models. The results of this study should be further evaluated on more farms.

Conclusion

The high variability of SCC, particularly in test-day SCC, reinforces the need to identify additional predictive indicators for SCM, derived from longitudinal, large-scale studies, mined by highly efficient pattern recognition models. This study employed ten different attribute weighting models (data mining) with different statistical backgrounds and selected key features based on the agreement of these algorithms. Results showed that after SCC, lactose concentration and EC gained the highest weight by all attribute weighting models. We suggest that a combination of lactose concentration and EC can provide a predictive tool for mastitis.

Supplementary material

The supplementary material for this article can be found at https://doi.org/10.1017/S0022029918000249

References

Addis MF, Tedde V, Puggioni GM, Pisanu S, Casula A, Locatelli C, Rota N, Bronzo V, Moroni P & Uzzau S 2016 Evaluation of milk cathelicidin for detection of bovine mastitis. *Journal of Dairy Science* **99** 8250–8258

Bartlett PC, Miller GY, Anderson CR & Kirk JH 1990 Milk production and somatic cell count in michigan dairy herds. *Journal of Dairy Science* **73** 2794–2800

Berglund I, Pettersson G, Ostensson K & Svennersten-Sjaunja K 2007 Quarter milking for improved detection of increased SCC. *Reproduction in Domestic Animals* **42** 427–432

Bochniarz M, Adaszek L, Dziegiel B, Nowaczek A, Wawron W, Dabrowski R, Szczubial M & Winiarczyk S 2016 Factors responsible for subclinical mastitis in cows caused by Staphylococcus chromogenes and its susceptibility to antibiotics based on bap, fnbA, eno, mecA, tetK, and ermA genes. *Journal of Dairy Science* **99** 9514–9520

Bradley AJ, Leach KA, Breen JE, Green LE & Green MJ 2007 Survey of the incidence and aetiology of mastitis on dairy farms in England and wales. *Veterinary Record* **160** 253–257

Ebrahimi M, Ebrahimie E & Shamabadi N 2010 Are there any differences between features of proteins expressed in malignant and benign breast cancers? *Journal of Research in Medical Science* **15** 299–309

Ebrahimi M, Lakizadeh A, Agha-Golzadeh P, Ebrahimie E & Ebrahimi M 2011 Prediction of thermostability from amino acid attributes by combination of clustering with attribute weighting: a new vista in engineering enzymes. *PLoS ONE* **6** e23146

Ebrahimi M, Aghagolzadeh P, Shamabadi N, Tahmasebi A, Alsharifi M, Adelson DL, Hemmatzadeh F & Ebrahimie E 2014 Understanding the underlying mechanism of HA-subtyping in the level of physic-chemical characteristics of protein. *PLoS ONE* **9**(5) e96984

Ebrahimie E, Ebrahimi F, Ebrahimi M, Tomlinson S & Petrovski KR 2018 Hierarchical pattern recognition in milking parameters predicts mastitis prevalence. *Computers and Electronics in Agriculture* **147**(C) 299–309

Goncalves JL, Tomazi T, Barreiro JR, Beuron DC, Arcari MA, Lee SH, Martins CM, Araujo Junior JP & dos Santos MV 2016 Effects of bovine subclinical mastitis caused by Corynebacterium spp. on somatic cell count, milk yield and composition by comparing contralateral quarters. *Veterinary Journal* **209** 87–92

Halasa T, Huijps K, Osteras O & Hogeveen H 2007 Economic effects of bovine mastitis and mastitis management [a review]. *Veterinary Quarterly* **29** 18–31

Hande Küçükönder FÜ, Ceyhan A and Cinar M 2015 Determination of the effect of somatic cell count on udder measurements and subclinical mastitis with data mining method. *Pakistan Veterinary Journal* **35** 441–445

Harmon RJ 1994 Physiology of mastitis and factors affecting somatic cell counts. *Journal of dairy science* **7** 2103–2112

Kamphuis C, Pietersma D, van der Tol R, Wiedemann M & Hogeveen H 2008 Using sensor data patterns from an automatic milking system to develop predictive variables for classifying clinical mastitis and abnormal milk. *Computers and Electronics in Agriculture* **62** 169–181

Kamphuis C, Mollenhorst H, Heesterbeek J & Hogeveen H 2010a Data mining to detect clinical mastitis with automatic milking. *In Proceedings of the Proceedings of the 5th IDF Mastitis Conference: Mastitis Research Into Practice*, pp. 568–572. Christchurch, New Zealand, 21–24 March 2010 (Ed J Eric Hillerton). Wellington, N.Z: Vetlearn, 2010.

Kamphuis C, Mollenhorst H, Heesterbeek JAP & Hogeveen H 2010b Detection of clinical mastitis with sensor data from automatic milking systems is improved by using decision-tree induction. *Journal of Dairy Science* **93** 3616–3627

Kawai K, Hayashi T, Kiku Y, Chiba T, Nagahata H, Higuchi H, Obayashi T, Itoh S, Onda K, Arai S & Sato R 2013 Reliability in somatic cell count measurement of clinical mastitis milk using DeLaval cell counter. *Animal Science Journal* **84** 805–807

KayvanJoo AH, Ebrahimi M & Haqshenas G 2014 Prediction of hepatitis C virus interferon/ribavirin therapy outcome based on viral nucleotide attributes using machine learning algorithms. *BMC Research Notes* **7** 565

Leitner G, Chaffer M, Caraso Y, Ezra E, Kababea D, Winkler M, Glickman A & Saran A 2003 Udder infection and milk somatic cell count, NAGase activity and milk composition—fat, protein and lactose—in Israeli-Assaf and Awassi sheep. *Small Ruminant Research* **49** 157–164

Malek dos Reis CB, Barreiro JR, Mestieri L, Porcionato MADF & dos Santos MV 2013 Effect of somatic cell count and mastitis pathogens on milk composition in Gyr cows. *BMC Veterinary Research* **9** 67

Mammadova N & Keskin İ 2013 Application of the support vector machine to predict subclinical mastitis in dairy cattle. *The Scientific World Journal* **2013** 3603897 doi: 10.1155/2013/603897

Mitchell GE, Rogers SA, Houlihan DB, Tucker VC & Kitchen BJ 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** 9–12

Panchal I, Sawhney IK, Sharma AK & Dang AK 2016 Classification of healthy and mastitis Murrah buffaloes by application of neural network models using yield and milk quality parameters. *Computers and Electronics in Agriculture* **127** 242–248

Petrovski KR, Trajcev M & Buneski G 2006 A review of the factors affecting the costs of bovine mastitis: review article. *Journal of the South African Veterinary Association* **77**(2) 52–60

Schukken YH, Wilson DJ, Welcome F, Garrison-Tinofsky L & Gonzales RN 2003 Monitoring udder health and milk quality using somatic cell counts. *Veterinary Research* **34** 579–596

Sharifi S, Pakdel A, Ebrahimi M, Reecy JM, Farsani SF & Ebrahimie E 2018 Integration of machine learning and meta-analysis identifies the transcriptomic bio-signature of mastitis disease in cattle. *PLoS ONE* **13**(2) e0191227

Sharma N, Singh N & Bhadwal M 2011 Relationship of somatic cell count and mastitis: an overview. *Asian-Australasian Journal of Animal Sciences* **24** 429–438

Shekoofa A, Emam Y, Ebrahimi M & Ebrahimie E 2011 Application of supervised feature selection methods to define the most important traits affecting maximum kernel water content in maize. *Australian Journal of Crop Science* **5** 162

Shekoofa A, Emam Y, Shekoufa N, Ebrahimi M & Ebrahimie E 2014 Determining the most important physiological and agronomic traits contributing to maize grain yield through machine learning algorithms: a new avenue in intelligent agriculture. *PLoS ONE* **9**(5) e97288

Torkzaban B, Kayvanjoo AH, Ardalan A, Mousavi S, Mariotti R, Baldoni L, Ebrahimie E, Ebrahimi M & Hosseini-Mazinani M 2015 Machine learning based classification of microsatellite variation: an effective approach for phylogeographic characterization of olive populations. *PLoS ONE* **10** (11) e0143465

Vinod Bharat BS, Khandelwal K, Navsare S, HOD Computer Department D. Y.Patil School of Engineering Academy Ambi, Pune, Maharashtra, India B.E Computer & Pune, Maharashtra, India 2016 A review paper on data mining techniques. *International Journal of Engineering Science and Computing* 6 6268–6271

Whyte D, Orchard R, Cross P, Frietsch T, Claycomb R & Mein G 2004 An on-line somatic cell count sensor. *Automatic Milking: A Better Understanding*, pp. 235–240 (Eds A Meijering, H Hogeveen and CJAM de Koning). Wageningen, the Netherlands: Wageningen Academic Publishers