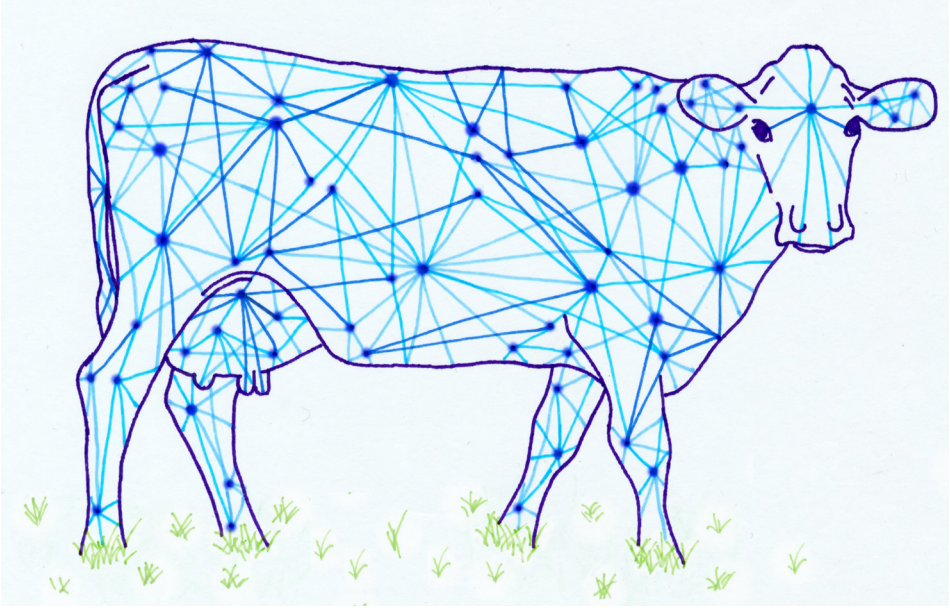
Acta Universitatis Agriculturae Sueciae

Doctoral Thesis No. 2021:5
Faculty of Veterinary Medicine and Animal Science

Indicators of mastitis and milk quality in dairy cows

Data, modeling, and prediction in automatic milking systems

Dorota Anglart



Indicators of mastitis and milk quality in dairy cows

Data, modeling, and prediction in automatic milking systems

Dorota Anglart
Faculty of Veterinary Medicine and Animal Science
Department of Clinical Sciences
Uppsala

SLU
SWEDISH UNIVERSITY OF AGRICULTURAL SCIENCES

DOCTORAL THESIS
Uppsala 2021

Acta Universitatis agriculturae Sueciae
2021:5

Cover: Drawing by Ingrid Bergman

ISSN 1652-6880
ISBN (print version) 978-91-7760-688-8
ISBN (electronic version) 978-91-7760-689-5

Uppsala
Print: SLU Service/Repro, Uppsala 2021

Indicators of mastitis and milk quality in dairy cows

Abstract

Methods for generating predictions of important and generally accepted indicators of udder inflammation and poor milk quality, such as somatic cell count (SCC) or changes in milk homogeneity, are few. The aim of this thesis was to investigate methods to identify indicators of mastitis and poor milk quality in dairy cows using data generated by automatic milking systems (AMS).

The first part of the project investigated the relationship between SCC and data regularly recorded by the AMS using models that could capture nonlinear associations between the explanatory variables and the outcome. This information could be used in modeling the SCC. Furthermore, three statistical methods, generalized additive model, random forest and multilayer perceptron, were compared for their ability to predict SCC using data generated by the AMS. The results showed that equally low prediction error was obtained using generalized additive model or multilayer perceptron for prediction of SCC based on AMS data.

The second part explored the dynamics of changes in milk homogeneity in cows milked in AMS using descriptive statistics for clots collected by inline filters, scored for density. Clots were found among certain cows and cow periods and appeared in new quarters over time. Models were fitted for detecting and predicting clots in single cow milkings as well as for detecting clots in milkings over a longer period. The models successfully distinguished periods of milking free of changes in milk homogeneity, although the detection and prediction performance was poor. The prediction target and severity grade of each density category is discussed.

Keywords: udder health, somatic cell count, milk homogeneity, generalised additive model, multilayer perceptron, random forest, machine learning

Author's address: Dorota Anglart, DeLaval International AB, P.O. Box 39, SE-147 21, Tumba, Sweden

Indikatorer för mastit och mjölkvalitet hos mjölkkor

Sammanfattning

Metoder för att prediktera viktiga och generellt accepterade indikatorer för juverinflammation som påverkar mjölkens kvalitet, så som celltal eller mjölkens homogenitet, är fåtaliga. Syftet med denna avhandling var att undersöka metoder för att identifiera dessa två viktiga indikatorer för mastit och mjölkkvalitet hos mjölkkor genom att användas data som genereras från automatiska mjölkningssystem (AMS).

I projektets första del undersöktes sambandet mellan celltal och data som genereras ur AMS med en generaliserad additiv modell, som kan fånga upp icke-linjära samband mellan variablerna och responsen. Denna information kunde sedan användas för att modellera celltal med AMS-data. Vidare jämfördes tre metoder generaliserad additiv modell, multilayer perceptron och random forest för att prediktera celltal. Resultatet visade på lika låga prediktionsfel för den generaliserade additiva modellen som för multilayer perceptron.

I den andra delen undersöktes dynamiken av homogenitetsförändringar i mjölken hos kor mjölkade i AMS genom att samla flockor med hjälp av filter monterade i mjölkledningen, samt poängsätta flockornas densitet på filtren. Flockor återfanns hos en begränsad grupp kor och ko-perioder samt förekom i nya fjärdedelar över tid. Flera modeller anpassades för att hitta och prediktera flockor i enskilda kors mjölkningar samt under en sammanhängande period. Modellerna lyckades mycket bra med att särskilja mjölkningar och perioder för kor som inte hade några flockor i sin mjölk medan prediktionen och detektionen av flockor var bristfällig. Säkerheten i mätmetoden av flockor samt graden av täthet på flockor och som bör kunna predikteras diskuteras.

Nyckelord: juverhälsa, celltal, mjölk homogenitet, generaliserad additiv modell, multilayer perceptron, random forest, maskininlärning

Address: Dorota Anglart, DeLaval International AB, P.O. Box 39, SE-147 21, Tumba, Sweden

Dedication

This thesis is dedicated to,

Pia 945, Krona 1082, Tekla 1296, Stella 1375, Lisa 1349, Lilja 1137, Ada 1368, Fagra 1377, Maja 1410, Krona 1426, and Fina 1381.

– Thank you for giving me a new direction in life!

"Hard work beats talent when talent fails to work hard"

–Tim Notke

Contents

List of publications

This thesis is based on the work contained in the following papers, referred to by Roman numerals in the text:

I. Anglart, D.*, Hallén Sandgren, C., Waldmann, P., Wiedemann, M., Emanuelson, U. (2020). Modeling cow somatic cell counts using sensor data as input to generalized additive models. Journal of Dairy Research, 87 (3), 282–289.

II. Anglart, D.*, Hallén Sandgren, C., Emanuelson, U., Rönnegård, L. (2020). Comparison of methods for predicting cow composite somatic cell counts. Journal of Dairy Science, 103 (9), 8433–8442.

III. Hallén Sandgren, C., Anglart, D., Klaas, I.C., Rönnegård, L., Emanuelson, U. Homogeneity density scores of quarter milk in automatic milking systems. Manuscript.

IV. Anglart, D., Emanuelson, U., Rönnegård, L., Hallén Sandgren, C. Detecting and predicting changes in milk homogeneity using data from automatic milking systems. Manuscript.

*Corresponding author.

The contributions of Dorota Anglart to the papers included in this thesis were as follows:

I. Involved in formulation the research idea. Performed all data cleaning, analysis and modeling. Drafted and wrote the manuscript with regular input from the co-authors. Corresponded with the journal.

II. Involved in formulation the research idea. Performed all data cleaning, analysis and most parts of the modeling. Drafted and wrote the manuscript with regular input from the co-authors. Corresponded with the journal.

III. Actively involved in formulating the research idea. Responsible for the planning, execution and collection of data in the field trial. Prepared the data and performed the initial analysis. Wrote parts of the manuscript and contributed to other parts.

IV. Actively involved in formulating the research idea. Responsible for the planning, execution and collection of data in the field trial. Performed all data analysis and modeling. Drafted and wrote the manuscript with regular input from the co-authors.

List of tables

List of figures

Abbreviations

AIC	Akaike information criterion
AMS	Automatic milking system
CMC	Cow milk class
CMSCC	Cow composite somatic cell count
CMT	California Mastitis Test
CPSS	Cow period sum score
CPC	Cow period class
GAM	Generalized additive model
LDH	Lactate dehydrogenase
MDi	Mastitis detection index
MLP	Multilayer perceptron
MSe	Mean squared error
NPV	Negative predictive value
OCC	On-line somatic cell counter
PPV	Positive predictive value
QMS	Quarter milk score
QPSS	Quarter period sum score
RF	Random forest
SCC	Somatic cell count

1. Introduction

On many dairy farms, milking the cows is stressful and heavy work. Farmers are bound to late evenings and early mornings, which affects the social life and freedom of farming families (de Koning 2010). This, together with increased labor costs, has driven the need for milking automation (Rossing & Hogewerf 1997). Since the first automatic milking system (AMS) came to market in 1990, interest in milking cows automatically has only increased (Svennersten-Sjaunja & Pettersson 2008). In 2010, 8000 cows were already being milked automatically in 25 countries (de Koning 2010), and today, 10 years later, more than 50,000 AMS are milking cows in 50 countries (personal communication, Clara Secher, DeLaval International AB, Tumba, Sweden). This automation has changed not only the way of life of farmers, but also the nature of their labor (Lind *et al.* 2000). The role of the farmer, previously focusing on manual management in which milking the cows occupied most of the time, has shifted towards more flexible work and control activities, such as monitoring the outputs of sensor systems (de Koning 2010).

One challenge on dairy farms is to manage udder health and ensure that high-quality milk is delivered to dairies. Mastitis, a disease affecting udder health and consequently the milk quality, is among the most serious diseases in dairy cattle (Viguier *et al.* 2009) and has a major impact on farm profitability (Halasa *et al.* 2007; Hogeveen *et al.* 2011). In AMS, sensor systems that can correctly identify cows with mastitis and alert the farmers before a milking that would yield milk of unacceptable quality are thus very important.

1.1 Mastitis

Mastitis is an inflammation of the mammary gland, normally caused by bacteria. The condition can be divided into two main categories; clinical mastitis, a state with clinical signs in the milk with or without symptoms from the udder or general signs of illness, and subclinical mastitis, a state with no visible signs in the milk or udder (International Dairy Federation 2011; Pinzón-Sánchez & Ruegg 2011; Ruegg 2012). Furthermore, clinical mastitis can be classified as mild, moderate, or severe depending on the clinical signs (Ruegg 2012). Generally, severe cases are defined by systematic illness together with abnormal milk, while mild cases are defined only by signs of abnormal milk but without other signs of local or systemic illness. Moderate mastitis could be defined as abnormal milk together with signs of inflammation of the udder (Ruegg 2012). Mild cases are more common than moderate cases, while severe cases are even rarer (Wenz *et al.* 2001; Ruegg 2012). Mastitis could also be referred to as acute or chronic, depending on the duration of the condition (Viguier *et al.* 2009). Subclinical mastitis is often defined by an elevated somatic cell count (SCC), normally above 200,000, in the cows' composite milk, i.e., from all four quarters (Schukken *et al*. 2003; International Dairy Federation 2013). Subclinical mastitis is far more common than clinical mastitis (Viguier *et al.* 2009; Ruegg 2012), harder to monitor (Ruegg 2012), and can sometimes develop into clinical mastitis (Hovinen & Pyörälä 2011).

Mastitis and mastitis detection models are probably among the best-described topics in the dairy literature (e.g., Rutten *et al*. 2013). This is probably because of the impact of mastitis on so many different areas, such as animal welfare, antibiotic use, production loss, veterinary costs, culling rates (Halasa *et al.* 2007; Viguier *et al.* 2009), and milk quality (Politis & Ng-Kwai-Hang 1988; Barbano *et al.* 2006). Generally, prevention is better than cure, and early detection and intervention can prevent mastitis from progressing from mild to severe and/or from acute to chronic. Also, freedom from previous episodes of mastitis seems to have a protective effect in terms of ability to recover from clinical mastitis later in life (Pinzón-Sánchez & Ruegg 2011). It is therefore of great importance to find and prevent mastitis cases in a timely fashion.

1.2 Indicators of mastitis and milk quality

1.2.1 Visible indicators

Changes in milk homogeneity and color

Abnormal milk is often defined as milk that is deviant in terms of color, smell, or homogeneity. In particular, homogeneity changes, such as flakes or clots, are established indicators of clinical mastitis (Rasmussen, 2004; International Organization for Standardization, 2007; Kamphuis *et al*. 2013). Clots are also the most common deviation (Rasmussen *et al.* 2005). However, it is not entirely clear what makes the milk clot and the literature in this area is scarce. Some suggested reasons are, for instance, the agglutination of immunoglobulins with whey proteins, fat, or bacteria, changes in the casein fraction, and proteolytic activity due to the inflammation or infection process (Rasmussen & Larsen 2003).

Milk color is affected by several factors, such as breed, stage of lactation, and feed (Agabriel *et al.* 2007), but could also reflect udder health. A reddish color indicates blood in the milk. Hemorrhage, i.e., the passage of blood cells through capillary walls into the tissues, can occur at any stage of lactation, but more frequently after calving or due to trauma (Muhammad *et al.* 2015). Bacteriological infection could also cause hemorrhage, since some pathogens are associated with blood in milk (Pyörälä *et al.* 2011). As blood will change the color of the milk, it can also be measured by color-measuring sensors (Rasmussen & Bjerring 2005; Kamphuis *et al.* 2008a). Watery or yellow milk could indicate a bacteriological infection from, for example, *Escherichia coli* (Lohuis *et al*. 1990), however not confirmed with certainty.

A decrease in milk yield could be a sign of clinical mastitis (Hamann & Kr~smker 1997; Pyörälä 2003) or a consequence of subclinical mastitis, as indicated by the negative correlation between SCC and milk yield (Tyler *et al.* 1989; Koldeweij *et al.* 1999). However, a drop in milk yield is generally not a good indicator (Rasmussen 2004), since the problem is already established by the time the milk yield drops.

Behaviors such as reduced lying time (Siivonen *et al.* 2011; Medrano-Galarza *et al.* 2012; Fogsgaard *et al.* 2015) as well as kicking and leg lifting during milking (Fogsgaard *et al.* 2015) have been associated with mastitis, although the degree of pain might be the cause of such behaviors, which would therefore not indicate mild clinical mastitis cases (Medrano-Galarza *et al.* 2012). Behavioral changes are not widely investigated as input in

mastitis detection models, probably since the evaluation of behaviors is often based on quantitative measures captured by video recordings (Haidet *et al.* 2009). Sensors, such as accelerometers, have been used to measure cow movement for the assessment of lameness (Pastell *et al.* 2009), and information regarding milk cups kicked off during milking can be obtained from AMS data. The effect of adding this type of information to clinical mastitis detection models remains to be evaluated.

1.2.2 Non-visible indicators

Somatic cells

Somatic cells (leukocytes) are part of the immune system and are always present in cow milk to some extent (Leitner *et al.* 2012). Normal variability of somatic cells is a great part of the overall variability of a healthy cow's udder (Quist *et al.* 2008; Forsbäck *et al.* 2010). The variability can depend on many factors, such as day-to-day variation, stage of lactation, breed, milking interval, parity (Nyman *et al.* 2014), or stress and trauma (International Dairy Federation 2013). However, an elevated SCC in the mammary gland is usually a sign of inflammation, since somatic cells are primarily released from the blood into the mammary gland in response to an invasion of bacteria in the udder (Pyörälä 2003; Schukken *et al.* 2003). Day-to-day variation in SCC is also greater in infected cows (Chagunda *et al.* 2006b). The definition of a healthy udder is suggested to be an SCC below 200,000 cells/mL, while for a healthy quarter the corresponding level is suggested to be below 100,000 cells/mL (International Dairy Federation 2013). The different cutoffs reflect the fact that it is often only one quarter that is responsible for a high SCC, i.e., inflammation in more than two quarters concomitantly is rather rare (Forsbäck *et al.* 2009).

Elevated SCC levels also affect the quality of the milk (Politis & Ng-Kwai-Hang 1988; Barbano *et al.* 2006), and many dairies apply penalties for delivering milk with SCC levels above a certain threshold. To keep track of the udder health of individual cows, as well as bulk tank SCC, many farmers worldwide choose to participate in dairy herd improvement programs in which cows are sampled for cow composite SCC (CMSCC) (Schmidt & Smith 1986; Schukken *et al.* 2003). The frequency of samplings can vary greatly depending on the testing scheme, which could be, for example, monthly, bimonthly, or a selected number of times per lactation. Sampling and storage procedures are crucial to obtain a representative sample and a correct sampling result (International Dairy Federation 2013). To meet the

requirements of dairies, samples at the cow level as well as at the bulk tank level are analyzed in a laboratory, generally according to standard procedures (International Organization for Standardization 2005, 2008).

On the farm, the California Mastitis Test (CMT) can provide a quick and rough estimate of SCC at the quarter level by adding strip milk to a detergent that indicates the cell count by changing the viscosity to that of gel as an indication of SCC. The Wisconsin Mastitis Test is another option, suggested to be more precise but still only giving a rough indication of the SCC level (International Dairy Federation 2013). The use of fluoro-opto-electronic instruments, in which cells are fluoresced and counted using flow cytometry (Schmidt Madsen 1975; Kitchen 1981), is an established method to provide information regarding the CMSCC of cows' milk. This is an accurate and precise method and also the only standardized method for determining SCC (International Organization for Standardization 2006; International Dairy Federation 2013)

For AMS, online analysis equipment could be integrated in the milking station for on-farm analysis. One example is the online somatic cell counter (OCC™; DeLaval International AB, Tumba, Sweden), an AMS-integrated sensor with the capacity to measure CMSCC at every milking. A representative milk sample is collected throughout the milking, and image analysis is used to count the somatic cells. The concordance between the OCC measurements in the AMS and the results of sample analysis by the dairy herd improvement program procedure is around 80% (Nørstebø *et al.* 2019), which implies that this is an acceptably accurate and valuable tool for on-farm sampling. It is, of course, important that the costs of such a device do not exceed the gains from it in a particular milking system.

Conductivity

In a healthy udder, the major ions that contribute to the level of milk electrical conductivity are sodium, potassium, and chloride. The ion content can be measured by conductivity sensors. The electrical conductivity is temperature dependent and is between 4.0 and 5.5 mS/cm at 25°C in normal milk, with a variation of 0.1 mS/cm per degree Celsius (Wong *et al.* 1988). Low levels of sodium and chloride and a high level of potassium in milk are maintained by active cell metabolism. As an effect of a bacteriological infection, udder tissue is damaged and the levels of sodium and chloride will increase. To maintain the osmolarity, the level of potassium then decreases (Linzell & Peaker 1971). Because of this change in ion composition, the

electrical conductivity of the milk in the inflamed quarter will increase (Kitchen 1981).

Measurements of differences in electrical conductivity between quarters say more about the inflammation status of a quarter than does the absolute electrical conductivity value alone (Nielen *et al.* 1995a; Kamphuis *et al.* 2008b; Khatun *et al.* 2018), so within-cow comparison of quarters could be valuable in prediction models (Hamann & Zecconi 1998). The milk fraction of the measurements may also play an important role when comparing electrical conductivity between quarters. Electrical conductivity in the infected quarter is usually higher in the foremilk, decreasing in the main milk phase (Woolford *et al.* 1998).

Biomarkers

Biological markers or "biomarkers" can be defined as measurements of a biological state or condition, for example, indicators of pathogenic processes such as mastitis. Biomarkers suggested as indicators of mastitis are mainly chosen based on a specific increased enzymatic activity or increased concentration in relation to mastitis or intramammary infection (Bogin & Ziv 1973; Chagunda *et al.* 2006b). One of the most commonly studied biomarkers is lactate dehydrogenase (LDH) (Chagunda *et al.* 2006a; Larsen *et al.* 2010; Åkerstedt *et al.* 2011). Early on, this enzyme was suggested by Bogin & Ziv (1973) as a mastitis indicator since it is one of the enzymes found to increase significantly in infected quarters. LDH is also positively correlated with SCC (Kitchen 1981). However, Nyman *et al*. (2014) found that LDH is affected by cow factors such as parity, days in milk, and period of sampling. The same observation was, to some extent, made for N-acetyl-β-D-glucosaminidase (NAGase) (Nyman *et al.* 2014), an enzyme whose activity increases in infected quarters (Chagunda *et al.* 2006b; Larsen *et al.* 2010; Åkerstedt *et al.* 2011) and that has predictive ability for some bacteriological infections (Emanuelson *et al.* 1987; Pyörälä & Pyörälä 1997).

1.3 Automatic milking and mastitis

Several aspects need to be considered when it comes to providing good milk quality as well as detecting cows with mastitis in AMS. Inspecting the milk for abnormalities is recommended (European Commission 2004), in addition to monitoring the bulk tank SCC and bacterial count. In systems with a

milker, the hygienic quality of the milk is partly monitored by pre-stripping before milking, although the method is not always applied (Wenz *et al.*2007; Nielsen & Emanuelson 2013). As no milker is required to be present during milking in AMS, the system must be able to detect deviations and issue alerts before milkings from potentially sick cows to prevent milk of unacceptable quality from ending up in the bulk tank. Many of the different changes in milk linked to mastitis could be detected using sensors. Interest in sensors was not new when AMS entered the market, and relevant sensors have been available for some time. The difference with the use of AMS is that the costs of deploying sensors can be decreased: since many cows are milked on the same unit, the number of units to be equipped is smaller than in parlors of the same capacity (Hogeveen & Ouweltjes 2002).

1.4 Detection models

Directly presented single-sensor values are rarely meaningful for the farmer (de Mol & Ouweltjes 2001), so combining information from several sensors using various algorithms is common. Mastitis detection systems based on data from different system-integrated sensors have been studied extensively since the introduction of the AMS (see, e.g., Hogeveen *et al.* 2010; Rutten *et al.* 2013). Since the first studies of mastitis prediction models (De Mol & Ouweltjes 2001), several approaches for this have been evaluated. Combining sensor information, such as electrical conductivity and SCC (Kamphuis *et al.* 2008b), electrical conductivity, milk yield, and milk color (Kamphuis *et al.* 2010), and electrical conductivity, milk yield, and LDH (Chagunda *et al*. 2006a), as well as adding cow information such as body weight (Jensen *et al.* 2016) has been attempted.

1.4.1 Types of methods

Some of the more common statistical methods investigated in udder health research combining sensor data to predict clinical mastitis are artificial neural networks (Nielen *et al.* 1995b; Cavero *et al.* 2008; Sun *et al.* 2010), multivariate regression (De Mol & Ouweltjes 2001), decision trees (Kamphuis *et al.* 2010), fuzzy logic (Cavero *et al.* 2006; Kamphuis *et al.* 2008b), moving averages or thresholds (Claycomb *et al.* 2009; Mollenhorst *et al.* 2010; Khatun *et al.* 2017), and logistic regression (Khatun *et al.* 2018). The methods investigated in this thesis are summarized here.

Artificial neural networks

The artificial neural network is designed to process information in a way similar to the human brain, basing decisions on patterns and relationships and learning from them (e.g., Agatonovic-Kustrin & Beresford 2000; Haykin 2009). They also generalize and are adaptive, as artificial neural networks can be retrained to deal with minor changes in the data (Haykin 2009). On the downside, the method needs a lot of training data to avoid overfitting. Also, as artificial neural networks are "black box" algorithms, they provide little information regarding how each variable influences the outcome variable.

The multilayer perceptron (MLP) is a classic feed-forward neural network based on the perceptron, which was introduced by Rosenblatt (1958). The perceptron is a single-layer feed-forward network that can accommodate multiple inputs while giving a single output. The MLP consists of linear classifiers with several perceptrons organized into layers. There are always at least three layers: one input layer, one hidden layer, and one output layer. Unlike a single perceptron, a MLP could potentially capture nonlinear relationships and interactions between explanatory variables in a flexible manner (Haykin 2009).

Artificial neural networks are very popular in udder-health related research and have been suggested in several areas, such as clinical mastitis detection (e.g., Nielen *et al.* 1995b; Sun *et al.* 2010; Ankinakatte *et al.* 2013) and pathogen prediction (Heald *et al.* 2000; Hassan *et al.* 2009). Notably, the MLP is one of the suggested methods for pathogen prediction (e.g., Heald *et al.* 2000; Sun *et al.* 2010).

Tree-based methods

A decision tree is exactly what it sounds like, a tree-based method similar to a flow chart consisting of a series of splitting rules that starts at the top of the tree, turning into branches and leaves. Each branch represents the outcome of a test, for example, "true" or "false," and the decision taken at the branch level is shown by the leaves, for example, the mean of the observed outcome values (James *et al.* 2013). Decision trees are easy to use, illustrate, and explain, and can also be used for regression as well as classification (James *et al.* 2013). The random forest (RF) is a method constructed of many decision trees, each trained on a different data sample by means of replacement (bagging), which prevents overfitting (Breiman 2001). Random

forests are easy to train and tune, so they are also popular for use in many areas (Hastie *et al.* 2009), such as medicine or economics. Furthermore, the importance of each variable can easily be determined (e.g., Wiener & Liaw 2002), meaning that it is easy to understand what variables are the most important in explaining the outcome variable. Although decision trees have been investigated for predicting clinical mastitis (Kamphuis *et al.* 2010) or the levels of SCC (Sitowska *et al.* 2017), the RF has not been widely explored as a prediction method in udder-health–related research.

Regression

Linear regression is a straightforward approach to predicting the variable y using the explanatory variable x and assuming that the outcome variable y is continuous. Categorical outcomes, on the other hand, can be predicted by classification models such as linear discriminant analysis or logistic regression (James *et al.* 2013). By using additive models, or more precisely, generalized additive models (GAM), both types of outcomes can be fitted. Additionally, no parametric form (i.e., equation) is assumed between the outcome and the explanatory variables using GAM which allows more flexibility and also provides information regarding the relationship between the explanatory variables and the outcome variable (Hastie & Tibshirani 1990). Generalized additive models have not been widely explored as prediction methods in udder-health–related research, but the studies conducted suggest that GAM perform better than neutral network when compared (Ankinakatte *et al*. 2013).

1.4.2 Evaluation of detection models

The performance of a clinical mastitis detection system is commonly described in terms of sensitivity and specificity values. The sensitivity is the proportion of milkings with abnormal milk correctly identified as abnormal by the detection system, while the specificity is the proportion of milkings with normal milk correctly identified as normal by the detection system. The International Organization for Standardization (2007) suggested that sensitivity should be >70% and specificity >99% for a system to be reliable and useful. Additionally, an sensitivity of ≥80% has been suggested to be an acceptable level in order to lower the number of false-positive alerts (Hillerton 2000; Hogeveen *et al.* 2010).

Generally, it is hard to compare the results between studies of clinical mastitis detection models due to the lack of a common "gold standard" definition (Hogeveen *et al.* 2010; Rutten *et al.* 2013). Thus, the definition of a "true" clinical mastitis case varies between studies, and even small changes in the gold standard definition affects the number of true cases and thereby the performance results of the model (Claycomb *et al.* 2009). As the main focus of detection models is clinical mastitis, hygienic parameters such as SCC or changes in milk homogeneity are not always included in the definition. Some examples of true case definitions are the results of bacteriological culturing, cases based on the treatment of clinical mastitis, or combinations of clinical signs together with an SCC of a certain level (Mein & Rasmussen 2008; Hogeveen *et al.* 2010).

However, there is a standardized method for evaluating mastitis detection systems in AMS, proposed in the International Organization for Standardization Annex C (2007). This standard was suggested, supposedly applicable to all types of milking systems (Rasmussen 2004), when AMS were first introduced to the market. Generally accepted and widely practiced ways to determine the quality of milk and to monitor udder health are to measure the SCC and inspect the foremilk before attaching the cluster (European Commission 2004; NMC 2013). The suggested evaluation is therefore based on changes in milk homogeneity in combination with SCC using the CMT. In short, the test should be performed at three farms with at least 20 samples yielding abnormal test results, that is, milk containing clots >2 mm and with a CMT value >3 (e.g., 150,000 to 300,000 cells/mL). The recommended method to detect changes in milk homogeneity is pouring the milk through a filter with a pore size of 0.1 mm and looking for deposits, i.e., clots (Rasmussen 2005; International Organization for Standardization 2007).

Challenges

The visual inspection of changes in milk homogeneity is suggested to be a universal and objective method to find sick cows and therefore also for evaluation of mastitis detection systems (Rasmussen 2005; Claycomb *et al.* 2009; Kamphuis *et al.* 2013). However, knowledge regarding the dynamics of clots on quarter level is scarce. This is probably due to that looking for clots in milk is very time consuming (Kamphuis 2010), since the information regarding the cases needs to be manually collected, visually inspected, and scored. This is also valid for SCC, for which sampling requires manual work,

additional costs, or both. Investment in sensor systems such as the OCC could of course overcome the practical problems, but such sensor systems cannot profitably be installed in all types of milking systems, i.e., systems with several milking units such as automatic milking rotaries. Furthermore, the more frequent sampling of SCC could improve the monitoring of udder health (Sørensen *et al.* 2016), since increased sampling decreases the risk of udder health misclassification (Quist *et al.* 2008), although the increased sampling frequency also increases costs.

In general, manual work is undesirable in AMS. Some types of manual executions are unsuitable or impractical due to safety concerns and work ergonomics, simply because AMS are not designed for them. Hence, methods for accurately predicting changes in milk, such as SCC or milk homogeneity monitoring, could play important roles in developing clinical mastitis detection models. Although predictions of SCC cutoff levels have been investigated (Mammadova & Keskin 2015; Sitowska *et al.* 2017; Ebrahimi *et al.* 2019), the usefulness of SCC cutoff levels has been questioned (Ruegg 2003). Furthermore, by using thresholds, important information, for example, changes such as increases or decreases in SCC, which can indicate an upcoming case or recovery, are overlooked.

2. Aims

The overall aim of the thesis was to investigate methods to identify indicators of mastitis and poor milk quality in dairy cows, specifically targeting data generated from AMS. The specific aims were therefore:

- To better understand how to utilize and combine data from different sources as input information for prediction models.
- To better understand the dynamics of milk homogeneity changes in cows milked in automatic milking systems.
- To detect and predict changes in milk using data generated by automatic milking systems.

3. Materials and Methods

A summary of the materials and methods used in papers I–IV is presented in this section. Papers I and II focus on CMSCC, while papers III and IV focus on changes in milk homogeneity. Data for papers I and II, and III and IV, respectively, were collected and prepared in the same ways, and the papers are summarized accordingly in this section. Detailed descriptions can be found in the corresponding sections of each paper. Some additional analysis (not presented in the papers) is included in the thesis and is duly introduced here.

3.1 Data collection

For papers I and II, data were collected during an eight-week trial milking of 372 Holstein-Friesian cows twice daily in an automatic rotary. The cows were sampled for CMSCC during milking once weekly with a milk sampler. Samples were analyzed for CMSCC according to International Organization for Standardization/IEC (2005). Animal information was extracted from the herd management system, together with information from each milking. The milking data at the quarter level included electrical conductivity (mS/cm), mean and peak milk flow (g/min), and incompletely milked quarters (yes/no). Examples of milking data collected at the cow composite level were milking duration (minutes) and an index capturing the likelihood of mastitis, i.e., the mastitis detection index (MDi), incorporating different phases of electrical conductivity during milking together with the presence of blood in milk.

For papers III and IV, data were collected at four commercial dairy farms milking 624 cows in a total of 10 AMS. Farms were selected based on additional sensor equipment available, such as OCC units at all farms and

activity meters and LDH measuring equipment available at two farms. The cows involved in this data collection were mainly Holstein-Friesian. Each farm was visited on three occasions within 2–4 weeks; the author collected samples at all farms and was responsible for training two support persons. To identify clots in milk, visual milk inspection of all cow quarters milked during 30 consecutive hours (henceforth, "periods") was conducted on each visit. The milk inspection was performed using a meshed filter installed along the milk tube (Figure 1), and sampling was performed separately for each quarter by inspecting filters for clots after each cow milking. If there were no signs of clots, the filter was cleaned with water and put back into the holder. If clots or suspected clots were visible on the filter, the filter was put in a holder in a black box and photographed together with information such as cow number and time (Figure 1). Data from the AMS used in the analysis in paper IV covered each period as well as 48 hours before each period. The data contained similar AMS information as the data collected for papers I and II, with some system-specific exceptions, and in addition also contained information on, OCC, cows' hourly activity, and LDH.

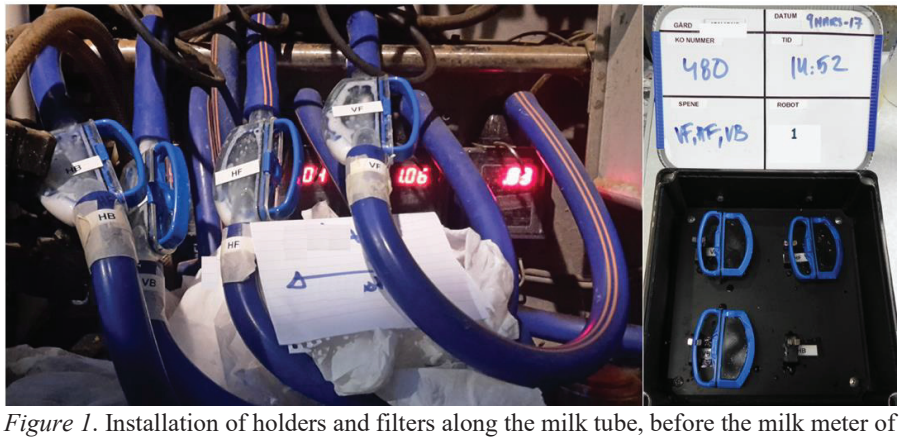


Figure 1. Installation of holders and filters along the milk tube, before the milk meter of each teat cup (left) and the box with holders where filters were photographed together with the information regarding the sample (right).

3.2 Data preparation

Data preparation and cleaning as well as statistical analyses for papers I, II, and IV were performed in R (R Development Core Team, 2018). Details of all R packages used for each analysis are found in papers I, II, and IV. Statistical analyses for paper III were performed using SAS version 9.4 (SAS Institute Inc., Cary, NC, USA).

3.2.1 Data for modeling and predicting CMSCC (I & II)

Data from cows not included in the weekly CMSCC sampling were removed. The cows were grouped according to parity 1, 2, or ≥3 and all milking events for cows during the first week of lactation were removed. Observations of CMSCC without a complete setup of explanatory variables were removed, as were explanatory variable outliers. In total, <1% of the data were removed due to cleaning.

The CMSCC values were transformed to a log10 scale, further referred to as log10CMSCC. To analyze the log10CMSCC outcome variable at the composite level together with the potential explanatory variables at the quarter level, new variables were created from the quarter variables and named with suffixes as follows, for instance: "max," i.e., highest value of a variable within cow and milking session; "diff," i.e., difference of a variable between quarters; "var," i.e., variance of a variable within cow between quarters; and "min," i.e., lowest quarter value within cow and milking. Past-period records (lags) were added to all variables for seven days, i.e., 14 milking sessions before each of the eight CMSCC sampling events. For paper II, the numerical variables and the outcome variable log10CMSC were also scaled, i.e., normalized with a mean value of zero and standard deviation of one.

3.2.2 Data to describe and predict changes in milk homogeneity (III & IV)

Density scoring of clots on filters

The collected images of clots on filters were scored according to a scale ranging from 0 to 5: a score of 0 was defined as no signs of clots; a score of 1 was defined as a trace; a score of 2 was defined as a mild case of clots; a score of 3 was defined as a moderate case of clots; a score of 4 was defined

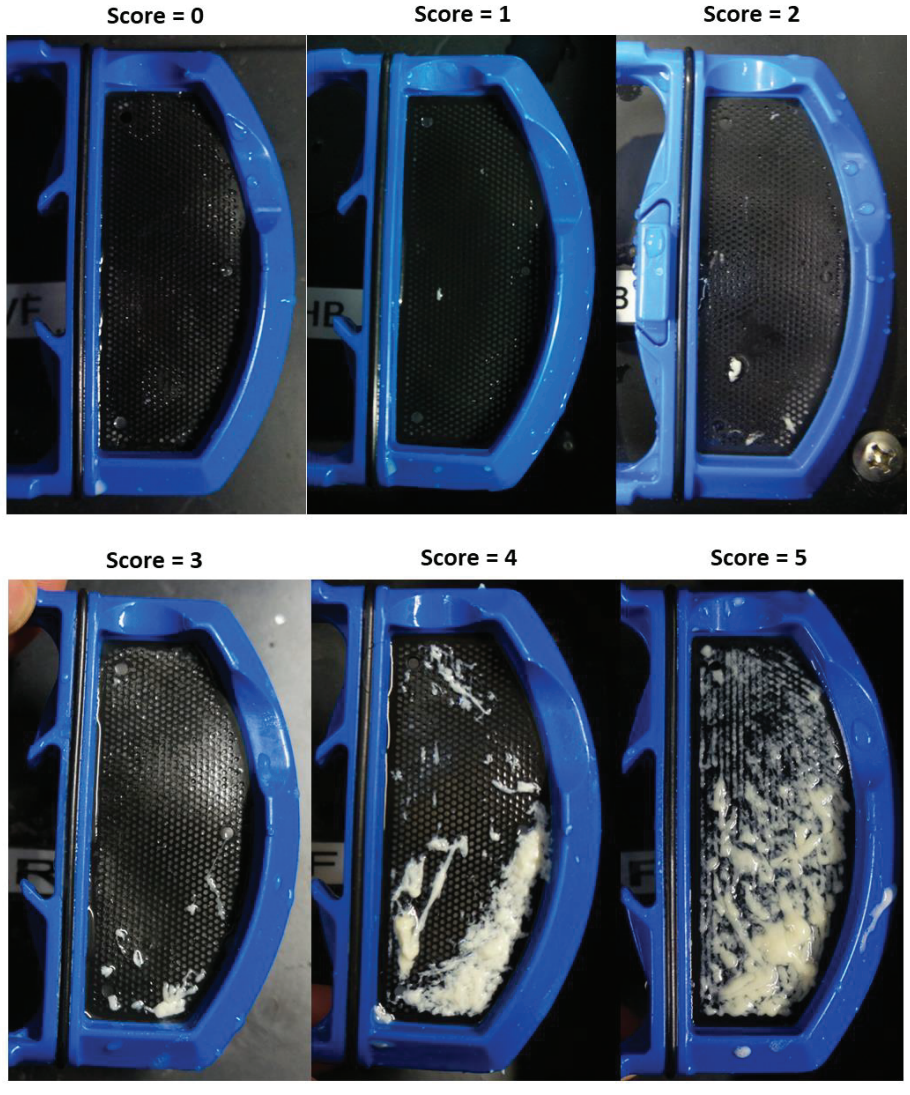


Figure 2. Score, definition and aggregate area of deposits on the filters scored for density:
0 = no sign of clots, none
1 = trace, Ø < 3 mm
2 = mild, Ø ≥ 3 mm
3 = moderate, Ø ≥ 5 mm or approximately 10% covered
4 = heavy, between 10% and 50% covered
5 = very heavy assemble of clots, more than 50% covered

as a heavy case of clots; and a score of 5 was defined as a very heavy case of clots (Figure 2). Three assessors scored each image individually so that each image received three scores. If at least two out of three assessors' scores agreed, that score was set as the quarter milking score (QMS), otherwise the score was removed from the dataset. Scorer agreement was assessed using Fleiss kappa with three raters (Fleiss 1971).

Creating combined cow scores from the quarter density scores

The QMS were combined into several different cow or quarter scores to suit the analysis for papers III and IV. The abbreviations of each are summarized in Table 1.

In paper III, data from cows with at least two milkings during a sampling period were included in the analyses. Three basic definitions of scores were used: QMS, i.e., the score of the quarter milk sample from a single milking; the quarter period sum score (QPSS), i.e., the sum of all QMS per period and quarter; and the cow period sum score (CPSS), i.e., the sum of all QMS per period and cow. In addition, ΔCPSS was created for statistical modeling by subtracting the QMS from the CPSS to which it contributed at each quarter milking observation.

In paper IV, the cow milk class (CMC) was computed for each cow at each milking; it was 1 if any QMS ≥2, and otherwise 0. The cow period class (CPC) was computed for each cow period by summing all QMS (1 to 5) in the period, dividing the sum by the number of quarters, and dichotomizing the result by setting two thresholds; periods when no quarter received a QMS ≥3 or periods when no quarter received a QMS ≥4, respectively, were set to 0 and thus excluded from the positive category. These two outcomes were labelled CPC.3 and CPC.4, respectively. Thus, each cow obtained one CMC for each milking and two CPC for each period, with a value of 0 corresponding to a negative outcome and a value of 1 to a positive outcome.

Table 1. *List of abbreviations of combined scores created for papers III and IV*

Abbreviation	
QMS	Quarter milk score
QPSS	Quarter period sum score
CPSS	Cow period sum score
CMC	Cow milk class
CPC	Cow period class

Preparing system data for model input

For paper IV, AMS data from all farms were merged and cows received a farm-specific number. The cows' hourly activity values were recalculated to two variables: a daily mean and a daily coefficient of variation. As the LDH values were available more sparsely due to the sampling scheme, the latest sampled LDH value within the 48 hours before each period was used for the analysis. Explanatory variables corresponding to quarters set as "not to be milk" were considered faulty and removed. Categorical variables with missing values were assigned an additional level indicating the missing value. All numerical explanatory variables were normalized and missing values were set to 0, except missing OCC values, which were imputed using RF imputation (Stekhoven 2013) and log transformed. Factor explanatory variables, such as cow number, parity, and breed, were converted to dummy variables. Data from three milkings before the milk inspection were used to create past-period variables (lags) for all explanatory variables. Data were merged with the data containing the computed cow scores and thereafter divided into 70% test and 30% training datasets using random sampling.

3.3 Data analysis

3.3.1 Potential explanatory variables and modeling CMSCC (I)

Generalized additive models (Hastie & Tibshirani 1990) implemented in the mgcv package (Wood 2011) in R were used for the analyses. The AMS data collected for papers I and II were analyzed using log10CMSCC as outcome variable *y*. Each of the 934 explanatory variables was analyzed individually in an initial variable scanning together with the confounding variables lactation number (LN, factor), days in milk (DIM, linear variable), and Cow (random factor). Each individual explanatory variable, *X*, was fitted using a smooth function (non-parametric spline), unless *X* was a class variable, in which case it was fitted as a factor. The model for variable scanning also included Cow as a random effect. Hence, the following model was fitted for each *X*:

$$y_i = \beta_0 + LN\beta_{LN} + DIM\beta_{DIM} + \alpha_{Cow} + f[\![(X)]\!]_i + \varepsilon_i$$

$$\alpha_{\mathrm{Cow}} \sim N\left(0, \sigma^2_{\mathrm{Cow}}\right),\ \epsilon_i \sim N(0, \sigma^2_{\epsilon})$$

where i is an index for each cow, β_0 an intercept term, β_{LN} a fixed effect of lactation number, β_{DIM} a fixed effect of days in milk, α_{Cow} a random cow effect (assumed to be independent and identically distributed, iid), $f[\![(X)]\!]_i$ the non-parametric spline function of the explanatory variable, and ε_i a residual term (iid).

From the variable scanning, all explanatory variables for which $P < 0.001$ were kept for further analysis. The variables were corrected for multiple comparisons using Bonferroni correction and tested for multicollinearity using the variance inflation test. The remaining explanatory variables were included in a final multivariable GAM:

$$y_i = \beta_0 + LN\beta_{LN} + DIM\beta_{DIM} + \alpha_{Cow} + \sum_{j=1}^{p} f_j[\![(X)]\!]_{ij} + \varepsilon_i$$

$$\alpha_{\mathrm{Cow}} \sim N\left(0, \sigma^2_{\mathrm{Cow}}\right), \qquad \epsilon_i \sim N(0, \sigma^2_{\epsilon})$$

where the same notation is used as in the model for variable scanning. Here p is the number of explanatory variables included in the final multivariable GAM and j is an index for each explanatory variable.

In paper I, two main model variations were created by either including all potential explanatory variables from the variable scanning, or by excluding MDi variables and investigating the impact of using observed values of electrical conductivity rather than values derived by the MDi algorithm. To evaluate how well the models would perform on milking data from restricted time periods before the CMSCC sampling event, six additional variations of both models were fitted with potential explanatory variables from various time periods (close to or distant in time from the CMSCC sampling). In paper I, the corrected Akaike information criterion (AIC) and proportion of variance explained were used for model evaluation.

3.3.2 Method comparison to predict CMSCC (II)

In paper II, three methods for predicting CMSCC were compared: the GAM, representing regression models; the RF, representing decision trees; and the MLP, representing artificial neural networks. In addition to the confounding variables, explanatory variables as model inputs were chosen based on

findings in paper I, and the explanatory variable setup comprised the MDi, different electrical conductivity variables, and low peak milk flow. The outcome variable in each model was weekly log10CMSCC. Four explanatory variable setups were evaluated for each of the three modeling methods: data with seven-day lags, data with three-day lags, and removing Cow as the explanatory variable from both day-lag variations.

Constructing the models

The GAM was fitted the same way as the multivariable GAM in paper I, but the random effect of Cow was set to zero in cases in which cow was missing. Since the expectation of a random effect without any information is zero, we could make predictions for all cows regardless of whether or not a particular cow was sampled

The RF as well as the MLP were initially fitted using the seven-day-lag data for hyperparameter tuning, i.e., parameters to be assigned before training the models. For the RF, the "randomForest" package in R (Wiener & Liaw 2002) was used, fitting several regression models while considering the number of decision trees between 250 and 2000. The lowest mean squared error (MSe) was obtained by the model using 1000 trees. The default number of variables selected in each tree was applied, since the MSe was very similar between models comparing different variables selected using the tune RF function (Wiener & Liaw 2002).

The MLP was constructed with Keras for R (Chollet 2017), using the Keras model sequential. Two hidden layers were applied and the number of units in each layer was determined by running several models, with the validation split set to 0.2. In searching for the lowest validation MSe, 50–500 units were evaluated; the optimal choice was found to be 200 units in the first layer and 100 in the second. Since the model was constructed for a regression problem, a single output layer was constructed using one unit. The default linear activation function (relu), which is suitable for regression, was applied in the first and second layers. For model compilation, ADAM, the stochastic optimization method (Kingma & Ba 2015) was used, since this optimizer works well even with little tuning of the hyperparameters. The loss function, showing the difference between the observed and predicted values, was set to MSe. For model training, the default number of times for full forward and backward propagation (10) was used with a batch size of 64.

Model evaluation

Comparison of the three methods' performance was evaluated in two ways: five-fold cross-validation by random sampling, with 80% of data used for model training and the remaining 20% for model testing, and prediction on future data, i.e., dividing the dataset to use data associated with milk sampling events 1–6 for training and with milk sampling events 7 and 8 for testing. The metric used for method comparison in paper II was the MSe. The MLP is a stochastic machine learning algorithm and some randomness is used during learning. Hence the results will be slightly different each time and were calculated as mean MSe over ten runs.

3.3.3 Dynamics of density scores (III)

Descriptive statistics were used to describe the variation in CPSS, QPSS, and QMS values. In a dynamic analysis of QPSS between periods, only QMS from cows meeting the inclusion criteria, i.e., with three periods and CPSS ≥4 in at least one period, were used. To assess the likelihood of a single quarter being positive or negative in a period, a logistic regression model, with QMS as the outcome variable and the explanatory variables ΔCPSS, cow (repeated factor), breed, parity, period, quarter location, QMS status in previous period, days in milk, quarter milk yield (kg), and milking interval (time from previous milking) was fitted. GAM was used to investigate the functional form of the continuous variables, before inclusion in the logistic regression model. Results of the model in paper III are presented as predicted probabilities based on marginal means.

3.3.4 Models to detect and predict changes in milk homogeneity (IV)

In paper IV, several model variations were created to explore the ability to detect or predict changes in milk homogeneity. Two model variations were created using CMC as the outcome variable: a detection model, including data from the milking when clots were observed, and a prediction model, excluding data from the milking when clots were observed, i.e., using only the data from the three milkings before the milk inspection. Two model variations using CPC as the outcome variable were created, one for each level of clot density (i.e., CPC.3 and CPC.4, as defined above). The models included data from the first milk inspection of the period together with data from three milkings before the milk inspection.

Additional analysis

To investigate the effect of including additional sensor data on model performance, one additional detection model using CMC as the outcome variable enlarged with the explanatory variables LDH, activity mean, and coefficient of variation of the daily cow activity, was created, using data from only the two farms from which these data were available. For comparison, the original CMC model was re-run on data from these two farms.

Constructing the models

The algorithm used in paper IV was the MLP, again constructed using the Keras model sequential (Chollet, 2017). One hidden layer was applied and the number of units in the hidden layer was 50, determined by running several CMC.D models with 5–500 units and evaluating the accuracy and loss of each model. The model was customized for a binary classification problem, so the output layer was constructed with two units, using an activation function that normalizes the model output into a probability distribution (softmax). Binary accuracy was chosen as the metric, calculating how often the predicted values equal the actual values. As in paper II, ADAM (Kingma & Ba 2015) was chosen for configuring the learning process and, additionally, the weight regularization kernel regularizer l2 was used to prevent overfitting.

The CMC models were fitted with the default number of times (10) for full forward and backward propagation and the default batch size (i.e., 32). By setting the regularizer option to 0.005, the difference between validation loss and training loss was minimized. The parameters were again tuned for the CPC models, resulting in changing the number of epochs to 20 and setting the regularizer option to 0.05. Finally, each model variation was run 10 times on the training dataset. The performance of each of the ten model runs was evaluated on the test data by comparing the predicted and observed values and calculating the sensitivity, specificity, positive predictive value (PPV), and negative predictive value (NPV), presented as the median over ten runs.

4. Results

This section gives an overview of the most important results of papers I–IV. Detailed information on the results can be found in the corresponding sections in each paper. Some additional results (not presented in the papers) are also presented here.

4.1 Modeling and predicting CMSCC (I & II)

4.1.1 Important explanatory variables (I)

The strongest statistical association with CMSCC was found for different electrical conductivity variables or variables incorporating electrical conductivity (e.g., the MDi). The MDi was also the most significant variable when included in the main model. When the MDi was excluded, the most significant variables were the variance of electrical conductivity between quarters, quarter electrical conductivity, difference in electrical conductivity between quarters, and the maximum electrical conductivity of a quarter. The overall results also indicated that the explanatory variables closer to the CMSCC sampling event were more significant. Nonlinear relationships, estimated by the screening model, were found for several of the explanatory variables, visualized by smooth plots expressing the nonlinear relationship between the variable and the outcome (Figure 3).

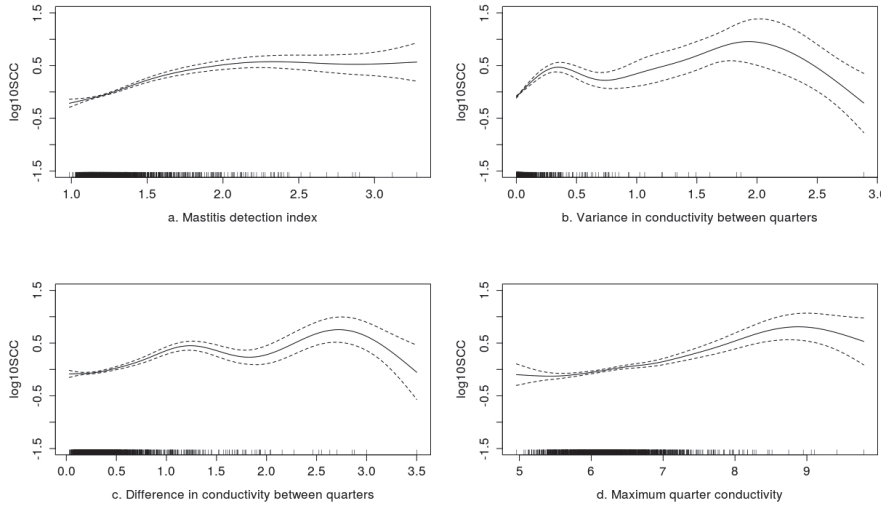


Figure 3. The partial effects of the: a) mastitis detection index, b) variance in conductivity between quarters, c) maximum quarter conductivity, and d) difference in conductivity between quarters estimated by the screening model. The pointwise 95% confidence interval is shown by the dashed lines. The vertical lines on the x-axis show the individual datapoints of each variable. The y-axis shows the composite milk somatic cell count.

4.1.2 Modeling CMSCC (I)

The model including all explanatory variables for seven days before the CMSCC sampling was the best model (lowest AIC and highest variance explained). However, compared with the results of models excluding the MDi variable, the differences in the AIC as well as in variance explained by each model were found to be small. The difference between the models restricted to data from three days before the CMSCC sampling event with or without the MDi was minimal. Models including milking data from the same milking as the CMSCC sampling had consistently lower AIC values, which implies that using data from closer to the CMSCC sampling resulted in better model performance.

4.1.3 Method comparison for CMSCC predictions (II)

In the method comparison, overall results indicated that the differences in MSe between the methods were quite small, larger for the RF than for the GAM and MLP. The differences in MSe were greater within method,

between explanatory variable setups for all methods except the RF, in which MSe was almost unaffected by any changes in variables. Removing the explanatory variable Cow from the models increased the MSe for all methods, for the GAM and MLP more than for the RF.

The lowest MSe (i.e., best model) was found for the GAM, where MSe was equally low in both evaluations and for both explanatory variable setups. However, an equally low MSe was found in the five-fold cross-validation for the MLP using the three-day-lag explanatory variable setup, implying that the performance of the methods is very similar. This is also illustrated by plots of the results of predictions on future data using the three-day-lag explanatory variable setup (Figure 4).

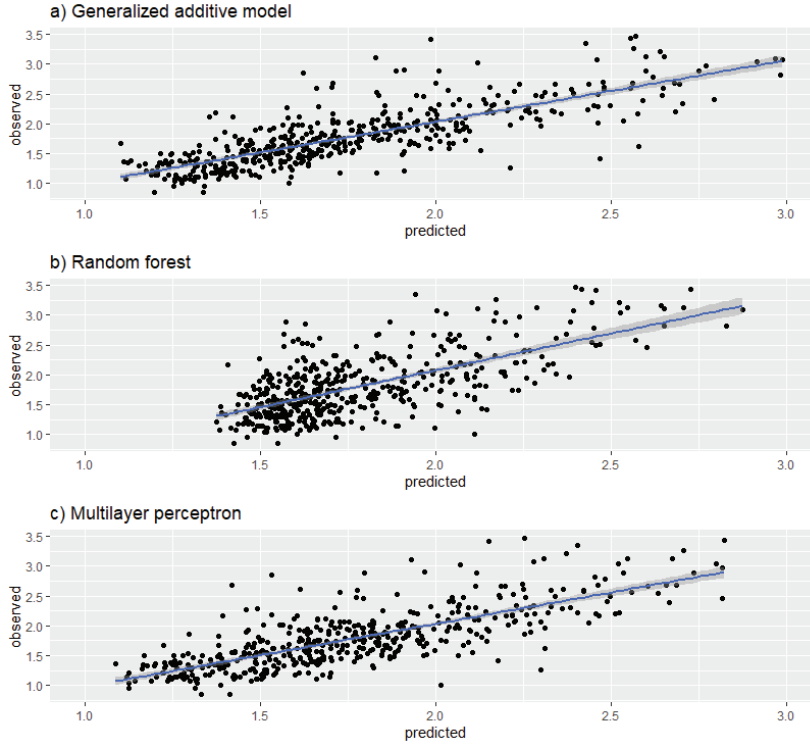


Figure 4. Observed versus predicted values of cow composite somatic cell count divided by 1000 on a log10 scale estimated by the (a) generalized model, (b) random forest, and (c) multilayer perceptron on future data, based on the explanatory variables for the three-day lags.

Both the GAM and MLP (Figure 4a and c) displayed a more balanced pattern across the predicted axis than did the RF. The log10CMSCC values predicted by the RF (Figure 4d) were clustered around 1.5 on the *x*-axis, which indicates that the method overestimated the low log10CMSCC values.

4.2 Milk homogeneity density scores in AMS (III & IV)

In total, 21,335 milk inspections were performed during 5424 milkings and 1656 periods of 624 unique cows. The number of samples with milk changed in homogeneity was 932 from 303 unique cows, and after discarding samples due to missing image, wrong cow number, etc., images of 892 quarters with scores >0 were available for the analysis, together with the 20,410 quarters with no identified clots (score = 0).

Substantial agreement was achieved between scorers (0.72), and the scoring results were as follows: 379 images received a score of 1, 303 images received a score of 2, 135 images received a score of 3, 67 images received a score of 4, and eight images received a score of 5. The total prevalence of clots and traces (QMS >0) was 4.2%, and prevalence of clots (QMS >1) was 2.4% in the total dataset from all periods. Of the collected samples, traces (QMS = 1) were found in 42%, mild and moderate cases in 49%, and heavy cases (QMS ≥4) in 9% (Figure 5).

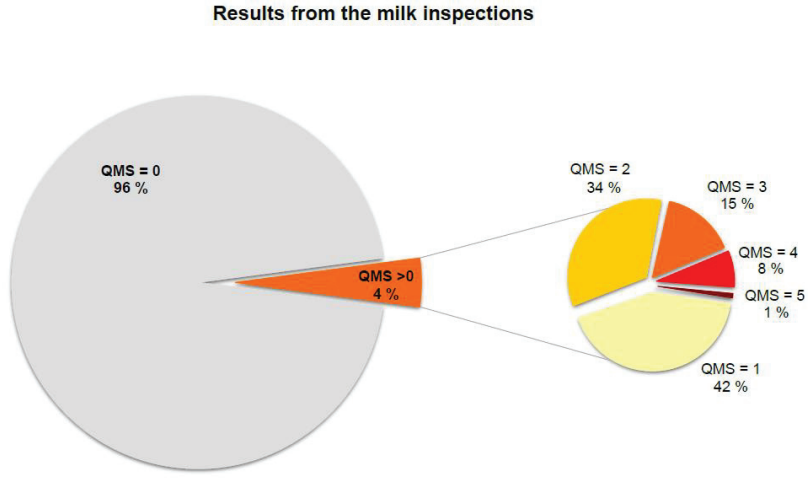


Figure 5. The results from the milk inspections showing the proportion of samples for each score.

4.2.1 Dynamics of density scores (III)

CPSS distribution and score dynamics between periods

The CPSS ranged from 0 to 50, with higher scores indicating more milkings with clots or heavier cases of clots in a period. A CPSS of 0 or 1 was found in a majority of periods, and in only 18% of the cow periods clots occurred (QMS ≥2). In 9% of the periods, CPSS values were ≥4, including 100% of the heavy, 88% of the moderate, and 69% of the mild cases. Hence, a threshold of CPSS ≥4 captured the periods and cows with the majority of high scores.

The 88 cows in the subset of cows with three sampling periods and at least one CPSS ≥4 were analyzed further, including all scores and also including traces, as they were found to be overrepresented in this group of cows. Seven of these cows had a CPSS ≥4 for three repeated periods, 20 cows for two repeated periods, and 61 cows for one period. The quarters with clots or traces, as well as the QMS scores, varied considerably between and within periods, as demonstrated in Figure 6. Of the 311 periods for cows with CPSS ≥4, 201 traces (QMS = 1) were found, which was 4.9 times higher per sampling period than in the remaining dataset; the corresponding numbers for QMS of 2 and 3 were 16.2 and 41.9 times higher, respectively.

		Period 1					**Period 2**						**Period 3**					
Cow	**Quarter**	MI 1	MI 2	MI 3	QPSS	CPSS	MI 1	MI 2	MI 3	MI 4	QPSS	CPSS	MI 1	MI 2	MI 3	MI 4	QPSS	CPSS
I	Left rear	2	0	2	4	**8**	2	1			2	**2**	3	2			5	**9**
	Left front	1	0	0	0		0	0			0		2	0			2	
	Right front	2	0	2	4		0	0			0		0	2			2	
	Right rear	0	0	0	0		0	1			0		0	0			0	
II	Left rear	2	3		5	**18**	0	0	0	0	0	**2**	0	0	0	0	0	**0**
	Left front	3	3		6		0	0	0	0	0		0	0	0	1	0	
	Right front	0	0		0		0	0	0	0	0		0	0	0	0	0	
	Right rear	3	4		7		1	0	0	2	2		0	0	0	0	0	
III	Left rear	0	0		0	**0**	0	0			0	**0**	0	0			0	**5**
	Left front	0	0		0		0	0			0		0	0			0	
	Right front	0	0		0		0	0			0		0	2			2	
	Right rear	0	0		0		0	0			0		3	0			3	
IV	Left rear	0	0	0	0	**5**	0	0	0		0	**10**	0	0			0	**6**
	Left front	0	1	2	2		2	1	2		4		3	3			6	
	Right front	0	0	1	0		0	0	0		0		0	0			0	
	Right rear	3	1	0	3		3	3	0		6		0	0			0	

Figure 6. Examples of scores within milk inspections (MI), quarter period sum score (QPSS), and cow period sum score (CPSS) distributions for four cows with CPSS ≥4 in one or more periods.

Score dynamics within periods

Of 338 quarters with scores of 2–5, 38% had QMS ≥2 at the following milking. A high QMS was related to a higher QMS at the following milk inspection. For QMS ≥4, 74% had at least QMS ≥2 at the following milk inspection, as compared with QMS = 2 being repeated in only 22% of cases.

The logistic regression analysis showed that the linear and quadratic terms for ΔCPSS, previous QMS, days in milk, milking interval, lactation number, and farm were significantly associated with the probability of observing a positive QMS. The probability of a cow having a positive QMS decreased with increasing number of days in milk. A longer milking interval corresponded to increased odds of a cow having a positive QMS. Furthermore, the analysis indicated that the odds of a random quarter in a period having clots (QMS ≥2) increased with higher ΔCPSS. In general, the probability of any random QMS being positive was very low (2.4%).

4.2.2 Detection and prediction of changes in milk homogeneity (IV)

The results of all models, i.e., the detection and prediction of clots for a single milking as well as detecting cow periods with clots showed an low sensitivity (i.e., 14–26%), while the specificity was high (i.e., 97–100%). The PPV results were intermediate to high (i.e., 42–72%). Adding information regarding activity and LDH did not improve the prediction performance much compared with detecting clots for a single milking. These results are summarized in Table 2 below.

Single milkings with clots

The misclassification rate of the model detecting clots at single milkings was lower among CMC having QMS ≥3 as the highest score within the combined cow score. For CMC having QMS ≥4 as the highest score, 63% were correctly identified while CMC incorporating QMS = 5 as the highest score were 100% correctly identified. The trends were very similar for the model predicting single milkings with clots.

Periods with clots in milk

Models detecting cow periods with CPC.3 as the outcome (i.e., no quarter in the period received a QMS ≥3 in the negative category) performed better (i.e., had higher sensitivity) than did models detecting cow periods with CPC.4 as the outcome (i.e., no quarter in the period received a QMS ≥4 in the negative category), Table 2. The results were opposite in terms of

specificity, i.e., periods with CPC.4 as the outcome were easier to distinguish as free of clots.

Table 2. *Sensitivity, specificity, positive predictive value (PPV), and negative predictive value (NPV) for the cow milk class (CMC) as well as for the cow period class (CPC) model with two different cow period class thresholds (i.e., CPS.3 and CPC.4)*

	Sensitivity	Specificity	PPV	NPV
CMC.D[1]	0.26	0.98	0.53	0.95
CMC.P[2]	0.25	0.98	0.47	0.95
CMC.D[1,3]	0.23	0.98	0.42	0.95
CMC.D[1,4]	0.22	0.97	0.38	0.95
CPC.3	0.23	0.98	0.72	0.87
CPC.4	0.14	1.00	0.71	0.94

[1]*Detection model using data from the milking of the milk inspection and three milkings before;* [2]*Prediction model excluding data from the milking of the current milk inspection;* [3]*Additional model using data from two farms with LDH and cow activity variable included;* [4]*Additional model using data from two farms with LDH and cow activity variable excluded.*

5. Discussion

5.1 Modeling and predicting CMSCC (I & II)

5.1.1 Explanatory variables considered for modeling

Quarter conductivity

Conductivity at the quarter level was found to be a strong explanatory variable for CMSCC, demonstrated by the scanning model as well as by both main models in paper I. The relationship between quarter conductivity and CMSCC has not previously been investigated using nonlinear modeling, which makes it difficult to compare the current results with previous findings. However, a positive relationship between SCC and conductivity has been reported (Hamann & Zecconi 1998), which was also found by the scanning model in paper I, as expected. The pattern (Figure 7) of the relationships between quarter conductivity and CMSCC differed greatly between the four separate quarters. The relationships were not always positive and differed the most for the left-rear quarter by being more linear and also, not significant in the scanning model. Using more data, i.e., conductivity data from more quarters with a wider distribution of values as well as a wider distribution of CMSCC values than we had in our data, might have equalized the difference between the patterns displayed by each quarter. Furthermore, it cannot be ignored that the SCC was collected at the cow composite level while conductivity was measured at the quarter level, which would affect the relationships found.

Conductivity values from separate quarters alone have been stated to be a poor predictor of clinical mastitis (Kamphuis *et al.* 2008b; Khatun *et al.* 2018). In contrast, the results in paper I suggest that maximum quarter

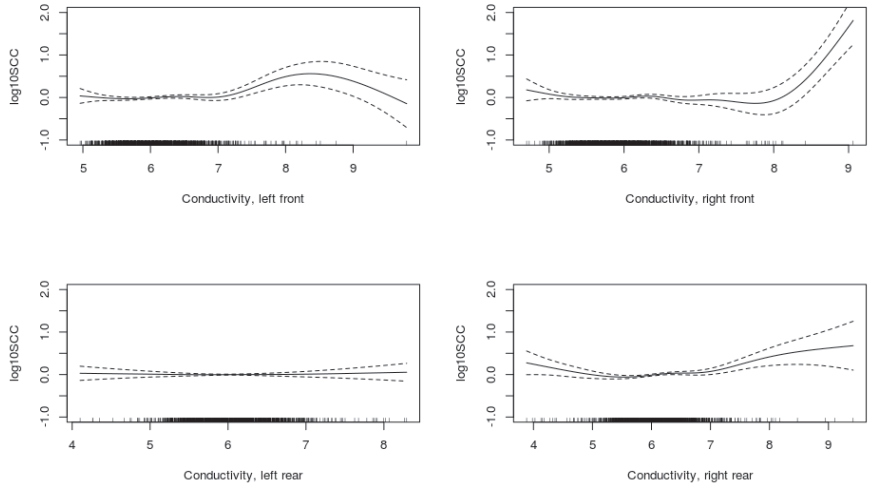


Figure 7. The effect of quarter conductivity on composite milk somatic cell count, estimated by the screening model. The pointwise 95% confidence interval is shown by the dashed lines. The vertical lines on the x-axis show the individual quarter conductivity datapoints.

conductivity, i.e., the highest conductivity value within a cow udder, was important in modeling the CMSCC. Furthermore, the relationship between maximum quarter conductivity and CMSCC was clearly positive. This demonstrates that information regarding conductivity from single quarters could also be important to consider if CMSCC is the prediction outcome. Also, including variables from single quarters, such as maximum conductivity, might strengthen the prediction performance of models in which CMSCC is part of the outcome.

Combined quarter conductivity

Strong significant relationships were found between CMSCC and several of the created variables for relative quarter conductivity, i.e., difference in conductivity between quarters, variance in conductivity between quarters, and, foremost, the MDi. Within-cow comparison of quarter conductivity is recommended (Kitchen 1981; Hamann & Zecconi 1998), and differences in conductivity between healthy and mastitis quarters have been observed (Nielen *et al.* 1995a; Bansal *et al.* 2005). Combining quarter conductivity variables was a successful method to model CMSCC. The input variables for the MDi are different phases of quarter conductivity values during milking

and threshold levels of blood at the quarter level. Because of the few observations of blood measured in milk, the MDi was probably mainly triggered by the conductivity variable

The relationship between the MDi and CMSCC was found to be positive and nonlinear. Poor correlations between SCC and the MDi have previously been reported (Lusis *et al.* 2017), but too few observations might have had an effect on the results as well as the nonlinear relationship that was observed in our study. The results of paper I clearly demonstrate that the MDi is a very important explanatory variable when modeling the CMSCC; notably, the models including the MDi generally performed better (paper I).

Other explanatory variables

Contrary to previous findings, neither short milking duration (Hammer *et al.* 2012) nor milking duration as such (Ebrahimie *et al.* 2018) was found to be significantly associated with CMSCC (paper I). The machine-on time can differ considerably between quarters when quarters are milked separately (Hogeveen *et al.* 2001). Due to a lack of more detailed data, milking duration was defined as the time required for all four quarters to be milked out, even though milking was conducted at the quarter level. This could explain why an association between milking duration and CMSCC could not be found.

Milking interval was not included as an explanatory variable in modeling or predicting CMSCC, since the milking interval was fixed due to batch milking at the farm where data were collected for papers I and II. However, if the study were to be replicated in an AMS system with irregular milking intervals, the milking interval might merit inclusion, since variation in milking interval as such has been suggested to have an impact on the SCC (Mollenhorst *et al.* 2011).

Milk yield was suggested to be one of the most important variables when predicting SCC using a decision tree model (Ebrahimie *et al.* 2018), and one of the least important variables in another study (Sitowska *et al.* 2017). Milk yield was excluded from our models due to the dilution effect of milk on SCC (Green *et al.* 2006), which could have affected the causal path between the other variables of interest and CMSCC. However, we included day-to-day variation in milk yield in the model, since it was significant in the scanning model. The variable might indicate an irregularity that could be associated with disorders in the udder and was created by calculating the difference between the current milking and the previous corresponding milking, i.e., morning milking versus previous morning milking. This type

of variable was not considered an intervening variable and was also significant in the main model where MDi was included.

We also added the variable incompletely milked quarters to the scanning model to investigate whether an incompletely milked quarter could be a potential predictor variable of CMSCC, since such quarters display higher SCC than do quarters that are milked out more completely (Penry *et al.* 2016). The number of incompletely milked quarters was significant in the variable scanning, and was therefore added to the models in paper I. However, as there were very few (i.e., 32) incompletely milked quarters, it is difficult to draw any strong conclusion from the results.

The occurrence of blood in milk was evaluated as an explanatory variable for CMSCC by adding two threshold levels for blood (ppm ≤1000 and ≤2000), but the results in paper I indicated that neither threshold was significantly associated with CMSCC. Similar results were found by Khatun *et al*. (2018), who scanned data with a univariable model to find explanatory variables for a clinical mastitis detection model. In contrast, Hammer *et al.* (2012) found blood in milk to be a significant risk factor for clinical mastitis, and consequently included it in a multivariable model. The association between blood in milk and clinical mastitis may depend on the type of pathogen causing the mastitis, because some pathogens are more commonly associated with blood in milk than others are (Pyörälä *et al.* 2011). Furthermore, there is more likely to be an association between blood in milk (or milk color, which is how blood in milk is commonly detected) and clinical mastitis than between blood in milk and elevated SCC as such. Also, the prevalence of milk samples with blood was very low in our material, which may also partly explain the non-significant associations with CMSCC.

5.1.2 Modeling CMSCC

Types of variables included in the modeling

Previous studies have shown that including several different types of conductivity variables, such as variance or maximum conductivity, increases the specificity of a mastitis prediction model (Norberg *et al.* 2004). The high degree of explanation of the models in paper I could probably partly be explained by the inclusion of the different types of conductivity variables, which describe different traits (e.g., difference between quarters), as well as the MDi, which combines conductivity results from different phases of

milking. Different variables will probably contribute in different ways to the model in describing the association with CMSCC. This was further demonstrated using models in which only three conductivity-based explanatory variables were fitted, i.e., the difference, variance, and maximum conductivity of a quarter (Anglart *et al.* 2019) For this smaller model, the variance explained was 0.78 and the AIC was 297; this is comparable to the best model presented in paper I, for which the variance explained was 0.80 and the AIC was 246. These very similar results indicate that including quarter-combined conductivity variables might play a more important role than including all possible variables.

Time-restricted data

Generally, the time needed to collect sufficient system data to make accurate predictions should be as short as possible. As new cows enter the system, the farmer could probably accept that the system cannot immediately gather all types of information, though this state should preferably be as short as possible. Thus, it is important to investigate how many days or milkings before an event are needed in order to predict the outcome.

Using data from three or seven days before the CMSCC sampling event did not affect the models' performance much, in either modeling (paper I) or predicting (paper II) CMSCC. In paper I, excluding information from the milking session when the sample was taken impaired the performance of the models as these models overall received higher AIC values. Thus, information from the milking for which CMSCC is to be predicted is important. From a practical point of view, this might be good enough, since this would be faster than dairy herd improvement program sampling results, or as fast as online sampling, in systems where such sampling is possible.

Including data on all variables from all seven days before the CMSCC sampling event gave the best model fit (paper I). However, the model fit was not noticeably affected by excluding the four days farthest from the CMSCC sampling, and we concluded that using all variables for three days before the CMSCC sampling should be sufficient to describe CMSCC. To investigate this further, predictions of CMSCC in paper II were evaluated using both three- and seven-day lags in the variable setup. The results in paper II confirmed our theory, since the difference in prediction error was very small between the two variations of day-lagged data used.

5.1.3 Predicting CMSCC

Reflections on data used

The main aim of paper II was to compare methods to predict CMSCC. The results of paper I gave insights into what variables to use and how much data should be sufficient. Hence, mainly conductivity variables were included in the models used in paper II and, additionally, the variable Cow was removed from both model variations. This was done to evaluate the importance of the model incorporating previous information regarding the cow that is not captured by the other variables. In a dairy farm, new cows are continuously presented at the milking, due to dry-off and calving but also due to recruitment. The comparison between including or not including the variable Cow as an effect in the model is important in order to determine how much and what kind of information the algorithm needs to accurately predict the CMSCC for newly introduced cows.

The results indicated that the GAM was the superior model, closely followed by the MLP, while the prediction performance of the RF was poorer overall. Removing Cow as an explanatory variable worsened the prediction performance of the GAM and MLP, while the RF was almost unaffected. The Cow variable captures information regarding the cows' overall level of CMSCC based on previous samplings, which implies that some information regarding the previous CMSCC is necessary to predict the next CMSCC. If information regarding cows' previous CMSCC is available, the model would gain in accuracy by including it.

One of the main findings in paper II was that the performance of the models differed more between the predictor variable setups than between methods. The differences in prediction performance using the three-day-lag variable setup compared with the seven-day-lag setup were very small for both the MLP and GAM. In fact, the MSe for the MLP was lower when using fewer days of data, which made the MLP perform as well as the GAM. Adding more data is generally a solution suggested to improve the accuracy of machine learning methods (Zhang & Ling 2018). Although more data were used in the seven-day-lag variable setup, the data were further in time from the CMSCC sampling event. This implies that, in this case, the MLP performed better with input data from days close to the outcome, i.e., the CMSCC sampling, even though it meant fewer data.

In line with our findings, Ankinakatte *et al*. (2013) found that the performance of a GAM was slightly better than that of an artificial neural

network, depending on the input variables used. The performance of the GAM in paper II, compared with the other methods, might have benefitted somewhat from the fact that the variable scanning was performed using the same method (paper I). However, the variables selected for model fit were similar to the explanatory variables suggested in other studies predicting CMSCC cutoffs using diverse methods (Panchal *et al*. 2016; Sitowska *et al*. 2017; Ebrahimi *et al*. 2019), so they were likely valid for predicting CMSCC independent of the method used for prediction.

The performance of the RF was mostly unaffected across different variable setups, which might also reflect a generally lower predictive accuracy of decision trees (James *et al.* 2013). The poor prediction performance of RF models in comparison with regression models has previously been reported (Miller & Franklin 2002; Sitowska *et al.* 2017).

5.2 Milk homogeneity density scores in AMS (III & IV)

5.2.1 Comments on data collection

Before data collection, a pilot trial was performed at farm A, where routines were developed for best performing the milk inspections and documenting the samples. The first concern was foam and milk residues on the filter, which made it difficult to acquire a clear image of the clots, as the milk and foam could also be interpreted as clots in the image. This was also reported by Rasmussen (2005), who found that samples of normal milk were scored as having clots, probably due to droplets of milk resembling clots in the images. To address this problem, after careful inspection of the filter, water was gently poured through the filter from above, as was also done by Rasmussen (2004). The risk of flushing away clots during this rinsing was likely very low, while the rinsing noticeably reduced the risk of misclassification when scoring images.

The second concern was to avoid subjective evaluation during data collection, i.e., to prevent those collecting the data from deciding what should or should not be considered clots (or flakes) if the case was unclear. Those involved were trained to acquire an image of everything that they suspected might be milk deposits on the filter. It was important that the data collection should not interfere with the assessment of density scores, which was avoided by applying this criterion. However, this might have had an

impact on the prevalence of traces or flakes found in the collected material compared with previous studies (see more below). Also, this explains why some images were scored as 0 (no clots) by the assessors setting the scores.

To obtain the density scores used in papers III and IV, three assessors independently scored each sample and the agreement was substantial (Landis & Koch 1977). Using several scores strengthens the accuracy of the assessment and gives a higher degree of confidence in the results (Boyer & Verma 2000). Three assessors were chosen to allow the score to be determined even if only two out of three assessors were in agreement, while excluding scores when all three assessors disagreed. Furthermore, by consciously including in the scoring several images in which no clots were found during the milk inspections, the results might have been strengthened further. Another method to assess the images could be image analysis algorithms, which also were considered. However, the images were not consistent enough, since some images had reflections (due to lighting conditions on the farm or at the time of the day), which was considered a problem for the image analysis.

A sample was considered failed if the cow could not be sampled for some reason. This was noted in the list of sampled cows by the assessors, as was whether the cow number was unknown or the image was missing. The total number of samples (i.e., milk inspections) that needed to be discarded due to failed sampling was very low (30), which implies that the collected data accurately reflect the occurrence of clots and traces at each farm.

5.2.2 Prevalence and spread of changes in milk homogeneity

Quarter-level sampling and assessment of scores

The quarter-level density scoring of clots collected by inline filters has not previously been investigated. Thus, comparing the prevalence of clots at the quarter level with previous findings is somewhat difficult, mainly due to different scales used for scoring, but also due to the sampling procedure and farm specific factors such as udder health. The main difference between collecting clots using inline filters and using foremilk samples is that, with the former, sampling is performed throughout the milking and all milk from the quarter will be investigated. Thus, a different distribution of clot occurrence might be expected, since clinical signs can also appear later in the milking (Rasmussen 2004). In the data collected for papers III and IV,

one squirt of the foremilk per quarter was discarded into the teat-cleaning cup of the AMS when each cow was pre-milked, which probably had a negligible effect when investigating the occurrence of clots or flakes.

A rough comparison of visually inspected foremilk samples for changes in milk homogeneity at the quarter level further illustrates the findings in papers III and IV (Table 3). Rasmussen *et al*. (2005) reported prevalence of 2.1% for clots and 1.4% for flakes. This is somewhat in agreement with our findings of prevalence of 2.4% for clots (scores 2–5) and 1.8% for traces or flakes (score 1). A higher prevalence of clots was found by Kamphuis *et al.* (2008a), who noted a joint 2.9% prevalence of three categories of clots (Table 3). The prevalence of cases categorized as "watery milk with small flakes" was noted as 2.0% by Kamphuis *et al*. (2008a); this category could probably be compared to our "trace" category or to the "flakes" category used by Rasmussen *et al*. (2005) and thus represents a slightly higher level. The number of cases decreased with increased clot category (Kamphuis *et al.* 2008a), i.e., fewer cases with more clots or higher density scores, which was also found in the data collected for papers III and IV. Rasmussen (2005) found the prevalence of clots to be 3.7%, which is somewhat in agreement with the finding of Kamphuis *et al*. (2008a), although the higher prevalence might also be because only one category was assessed or reflect the udder health on these farms.

Table 3. *Rough comparison of the prevalence of the different categories in each study; sampling was inline for papers III and IV, while other samplings were of foremilk*

Categories of flakes and clots and their prevalence			
Paper III, and IV	Rasmussen *et al*. (2005)	Kamphuis *et al*. (2008a)	Rasmussen (2005)
Trace density score 1 (1.8%)	Flakes yes/no (1.4%)	Watery and small flakes (2.0%)	
Clot density scores 2–5 (2.4%)	Clots yes/no (2.1%)	Few clots, mainly clots, mainly clots and milk appearance disappeared (2.9%)	Clots yes/no (3.7%)

Since the variation in clot prevalence was larger between the farms in the data collected for papers III and IV (data not shown) than between the studies compared in Table 3, the amount of clots found in these studies is probably still comparable.

In the data collected for papers III and IV, 42% of samples were classified as having a QMS of 1 (i.e., trace), a score that would likely not be noticed evaluating milk samples on cow composite level. The traces were 4.9 times more common in all milk inspections from cows who also held the higher scores (i.e., one or more CPSS ≥4), which might indicate that traces could be interpreted as an early stage of clinical mastitis, or a cow in the recovery stage of a clinical mastitis. In chronic cows, QMS of 1 could indicate a transition between a non-clinical and a clinical stage. The score dynamics between periods in paper III, showed that for cows with CPSS ≥4 in at least one period, the scores varied considerably within and between quarters, which could be an indicator of cows with chronic mastitis. Hence, collecting clots on quarter level is valuable, since important information is missed out as scores or grade of severity for each quarter are hidden in the whole milk. It also appears that a cow having clots in one quarter, is more likely to display changes in other quarters, hence dynamics of elevated QMS between the quarters of a cow needs further investigation.

Inline sampling and assessment of density scores

The density scoring of clots sampled by inline filters has previously been treated as the gold standard (Claycomb *et al.* 2009; Kamphuis *et al.* 2013, 2016). As the samples in previous studies were collected at the cow composite level, the 0–3 scale provided by the filter supplier could be utilized for density scoring. For papers III and IV, the scale needed to be adjusted, since we expected milk from single quarters to generate a different density distribution on the filters (i.e., some of the clots collected at the cow composite level might originate from several quarters), so a scale of 0–5 was developed. It is therefore difficult to compare the numbers of cases with different density scores between the scoring systems, although a rough comparison is attempted between the studies discussed.

As expected, the proportions of clots with different density scores differed somewhat between the studies compared (Table 4), i.e., papers III and IV, Claycomb *et al*. (2009), and Kamphuis *et al*. (2013, 2016). Notably, the proportion of cases increased with increasing density score in *Claycomb et al*. (2009), while the opposite relationship was found in the other studies,

with the proportion of cases decreasing with increasing density score. To make the scoring from papers III and IV more comparable, the proportions of scores are also presented with scores of 1 and 2 considered negative. However, the proportions of clots scored in the different density categories still differed, mainly for the density score of 3 (or graded 1 for composite level sampling), even though the scores of 1 and 2 in our data were removed. Simultaneously, the prevalence of clots sampled at the quarter level (Table 3) was quite similar between the studies, which emphasizes that single-quarter samples generate a different density distribution on the filters compared with cow composite samples. Hence, a scoring scale with a wider range was valid for use in papers III and IV.

Table 4. *Proportions in % of different categories of clots scored for density using inline samples in three studies. Results from papers III and IV are also presented with scores of 1 and 2 considered negative*

	Proportions of positive scores (%) at different clot density levels					
	Quarter level			Cow composite level		
Score[1]	Papers III, & IV[2], ≥1	≥2	≥3	Claycomb et al. (2009)[2]	Kamphuis et al. (2013)[3]	Kamphuis et al. (2016)[3]
1	42	-	-	-	-	-
2	34	59	-	-	-	-
3/1	15	26	64	20	11	40
4/2	8	13	32	28	50	38
5/3	1	2	4	52	39	22

[1]*Not scored using an equal scale;* [2]*Quarter scale (1-5);* [3]*Cow composite scale (1-3)*

5.2.3 Predicting and detecting changes in milk homogeneity

Prediction performance

The results of the prediction models in paper IV indicated a high ability to distinguish cow milkings and cow milking periods free of clots, as demonstrated by a high specificity. However, the low sensitivity indicates a poor ability to detect cow milkings or cow milking periods when clots occurred. Models were also evaluated in terms of PPV and NPV, statistical measures that depend on the prevalence (which was very low), which showed that the farmers would experience a low false-positive rate in practice, as the overall PPV was moderate to high (47–72%) for the main models.

Improving prediction performance

One of the most common suggestions for improving the predictions of machine learning algorithms is to use more data (Zhang & Ling 2018). The data collection was quite labor intensive, and of the 21,335 inspections at the quarter level, we observed 513 quarters with clots (scores 2–5). The option to collect as much data as needed to possibly improve predictions was thus not considered. It has been proposed that true clinical mastitis cases should include repetitive occurrence of clots in consecutive milkings (Rasmussen *et al*. 2005; Claycomb *et al*. 2009; Kamphuis *et al*. 2016). However, records of treated clinical mastitis cases are often used as 'true case', probably also since this type of data will not demand a massive or expensive data collection. When evaluating models for mastitis prediction, it is important that the gold standard definition is independent from the sensors in the evaluation, as treated cows are most likely were detected using information from the system sensors. Collecting clots inline is a reference method that is entirely independent from the system sensors. The low accuracy in prediction results in paper IV might also reflect this difference in comparison to other studies, where the reference for true cases were not completely independent.

The dataset was also unbalanced, meaning that most of the observations belonged to one class. More observations in one class will make the model learn this class better, affecting the prediction performance regarding the minority class (He & Ma 2013), which was also demonstrated by the model's greater ability to correctly classify cows without clots (i.e., high specificity). Furthermore, the positive class (i.e., having clots) was not only rare, but also comprised different types of severity cases (or scores). This may also influence the prediction ability of the model, as different types of severity cases may be due to various underlying causes. Although the prevalence was low, it was not extremely low, likely reflecting the "real-life" situation. It is therefore not obvious that collecting more data would have improved the prediction performance of the model, especially not the sensitivity.

Adding additional explanatory variables

Adding the explanatory variables LDH and daily cow activity did not improve the predictive performance of the models. The cow activity variable as such, measuring steps per hour, captures how much the cow is moving in the barn, but not whether the cow is standing or lying, which is suggested to be an udder inflammation indicator (Siivonen *et al.* 2011; Medrano-Galarza

et al. 2012; Fogsgaard *et al.* 2015). Furthermore, since the cow activity variable was a rough simplification, i.e., the daily mean and coefficient of daily variation, more advanced modeling could possibly have contributed more to the prediction performance. Daily cow activity could be an interesting variable to evaluate as an indicator of severe disease; however, adding the variable to the model did not improve the predictive performance regarding clots in milk.

Even though LDH is suggested as a marker of clinical mastitis (Bogin & Ziv 1973; Chagunda *et al*. 2006b), it might not be as useful for changes in milk homogeneity, which is also what our results imply. Increased LDH activity has been reported in udders infused with bacteria (Bogin & Ziv 1973), and high correlations have been observed between LDH and SCC (Bogin & Ziv 1973; Chagunda *et al.* 2006b). Since the reasons for the occurrence of clots in milk are not entirely understood (Rasmussen & Larsen 2003), clots do not necessarily indicate the presence of bacteria in the udder, while high SCC does not necessarily mean that the milk has changed in homogeneity (Rasmussen *et al.* 2005). Furthermore, cow factors such as parity or days in milk affect the activity levels of LDH (Nyman *et al.* 2014), which might impair the predictive performance of the variable.

Since LDH values were not available on a daily basis and were sampled rarely in some cows, the value was reused from the previous sampling occasion when missing; this might also have affected the results, as information regarding the actual LDH activity between the samplings was missing. Additionally, LDH was only available at two farms but the prediction results with and without LDH in the model for these two farms were similar, which further implies that LDH did not improve prediction performance.

The modeling approach

Initially, we investigated other methods along with MLP for predicting clots, i.e., GAM and a gradient boost classifier. The sensitivity and specificity results of these other models were poor and did not improve over the results of the MLP reported in paper IV. Another approach would be to evaluate ensemble learning, i.e., using multiple algorithms that could be tuned differently on different parts of the data (Pujari & Gupta 2012), which might improve prediction performance. This was not evaluated, however, due to time limitations.

Modeling the dependencies (i.e., correlations) between milkings within quarters might add some information and reduce noise in the model. This was successfully done by Franzén *et al.* (2012) who genetically evaluated mastitis by modeling transition probabilities between mastitis and non-mastitis cases. This should be investigated further, although the prediction performance of the approach would probably be only slightly improved, due to the variation of occurrence of clots in milk found in paper III, which did not follow a specific pattern.

The outcome variable

The quarters scored as 1 were excluded from the positive class for the models predicting or detecting clots in single milkings, since they were considered to represent traces of flakes. Single cases like these are likely a non-concern for farmers from a practical and milk hygiene point of view (e.g., see Figure 2), and would also be captured by the milk filter before the milk is delivered to the bulk tank. Furthermore, the presence of small flakes is reportedly a weak indication of bacteriological infection (Giesecke & van den Heever 1974). However, the findings in paper III indicated that "traces" accumulate to a larger extent among cows with heavy cases (QMS ≥4) and were therefore included in the models detecting periods of milkings with clots. Hence, it should be investigated whether excluding the "traces," even from the models detecting clots over a longer period, could have improved the prediction performance regarding periods with clots, as it might reduce some noise. This was not possible, however, due to time limitations.

5.2.4 Changes in milk homogenity as a prediction outcome

Usefulness of detecting clots during periods

The presence of clots over time is a generally considered important indicator of clinical mastitis. The presence of clots in two out of three consecutive milkings has been suggested to be used as the gold standard definition in clinical mastitis detection models (Mein & Rasmussen 2008; Kamphuis *et al.* 2013, 2016), and the International Organization for Standardization (2007) suggests testing for 36 consecutive hours. Predicting the accumulated presence of clots, as indicated by, for example, the CPSS or QPSS, could be a useful tool, since it would also provide estimations over time. The 30-hour period is not in accordance with International Organization for Standardization (2007) but might be sufficient in AMS, since the average

number of milk inspections per cow within a period, in the data collected for papers II and IV, was between 3.0 and 3.8, and thus covering several consecutive milkings.

Usefulness of detecting cow milkings with clots

Mollenhorst *et al*. (2012) concluded that farmers find it important to detect severe mastitis cases in a timely manner. The more severe cases, i.e., cow milk scores that included QMS ≥4, were correctly classified to a large extent as milkings with clots, compared to QMS <4, by the detection and prediction models in paper IV. This indicates that the models were better at categorizing the severe than the mild cases. From an udder health perspective, the severe cases are probably more important to find, and not the least to enable prediction. However, all types of single clot occurrences (QMS ≥2) might alert the farmer to take precautions. Furthermore, information regarding single occurrences of clots in milk detected by prediction models could be valuable input to clinical mastitis detection models.

Ideas on quarter level predictions

The prediction target in paper IV was to find cows, rather than specific quarters, with changes in milk homogeneity, so combined cow scores were created as the outcome variable. Predictions of quarter milkings with clots were initially investigated, as well as predictions of the density score as such, but the performance was poor (data not shown) and would probably not be better than cow composite predictions.

Predictions of clots in single quarters might be interesting. By setting the QPSS variable (investigated in paper III) as the outcome, a specific quarter could be tracked over time. The findings in paper III indicated that, of quarter milkings with clots, the heavy cases and very heavy cases were found in the same quarter in 74% of the subsequent milkings; of the quarters scored as mild, the corresponding figure was 22%. By predicting QPSS, quarters free of clots would not be included in the combined quarter period score to the same extent as they would in a combined cow period score. This might reduce noise from those quarters, which also might improve the prediction performance. However, since the repeatability was different in quarters scored as mild, outcomes such as cow score or cow period score are still valid and useful for some types of cases, in which the cow as such needs to be highlighted.

Furthermore, clot prediction in quarter milkings or periods might be interesting to combine with clot prediction at the cow level, since it might

increase the ability to detect different types of cases by customizing each algorithm for each task. Unfortunately, neither the QMS nor QPSS was evaluated as the outcome in paper IV due to time limitations, so it remains to be investigated how these variables could be combined.

Watery milk

Watery milk with small flakes is suggested to be an indicator of clinical mastitis caused by coliform bacteria (Eberhart *et al.* 1979; Pyörälä & Syväjärvi 1987). Watery milk cannot be detected using inline filters (Rasmussen 2005), while flakes or traces in milk, collected by inline filters, may originate from this type of milk. Thus, if such cases were present during our data collection, they were overlooked. Kamphuis *et al*. (2008a) suggested that "watery milk" should be treated as a homogeneity category between "few clots" and "mainly clots" in terms of severity, based on the increased mean electrical conductivity values in each investigated category. Furthermore, Montgomery *et al.* (1987) found watery milk to be significantly associated with coliform mastitis, while clots were not. How cases of watery milk should be handled when collecting samples using online filters merits further review.

The prediction target

Given this background, the question remains as to what level of density score is considered severe and why. We chose to set scores of 1 as negative, but kept them as positive for predicting clots during periods in paper IV. Claycomb *et al*. (2009) investigated several different combinations of gold standard definitions, excluding clots scored as 1 (on a cow composite scoring scale) or excluding cows with only one observation of clots, as also suggested by Kamphuis *et al*. (2013). In paper III, cows with only one milk inspection in a period were removed, independently of degree of score, since the dynamics between milkings were of main interest. Excluding milkings with single observations of clots was not considered for the predictions in paper IV, since we wanted to investigate the models ability to predict clots of all types. However, the inclusion of quarters scored as 2 as predictions of single-milking clots might be considered. The findings in paper III indicated that lower density scores also have lower repeatability. Cases scored as ≥3 should probably be set as positive (i.e., targeting mastitis) since their repeatability was over 50%. Thus, the prediction target in paper IV for cow milkings as well as cow periods might have been misguided. Kamphuis *et*

al. (2016) suggested that analyses excluding low-density scores should be presented in addition to analyses including all density scores. Prediction models intended to detect and predict higher density scores as targets should therefore be evaluated further.

5.3 Comments on methods, variables and evaluation

5.3.1 Choice of methods

In the framework of this thesis, three different types of methods were investigated, GAM representing regression models, RF representing decision trees and MLP representing artificial neural networks. In paper I, a GAM was used to investigate the relationship between potential explanatory variables and the outcome, CMSCC, since GAM can provide information regarding the relationships independently of the parametric function form (Hastie & Tibshirani 1990). This made it possible to reveal the underlying patterns between the variables of interest and CMSCC, which could be applied in the prediction models for CMSCC in paper II. In paper II, three methods were compared for their ability to predict CMSCC, with the GAM being found as the superior method. However, the MLP performed equally well using the three-day variable setup and, most importantly, the convergence speed of the MLP was much higher than that of GAM (data not shown). This is a major advantage if the method is to be used for applications operating on real-time data. Furthermore, a GAM cannot make predictions for an observation in a test set if a class in one of the factor variables is not represented by any observations in the training set. For balanced data with few factor variables this is not a major problem, but for large models with unbalanced data it is, which was the case when modeling CMSCC. For paper IV, some initial tests with the GAM were conducted; however, the MLP was chosen as the method for predicting clots due to its better performance and because artificial neutral networks have previously been suggested as a method for clinical mastitis detection (Nielen *et al.* 1995a; Sun *et al.* 2010; Ankinakatte *et al.* 2013), which clots are an indicator of.

One difference between the methods evaluated in this thesis framework is that the RF and MLP required parameter tuning, which the GAM did not. Tuning the hyperparameters creates the structure and configuration of the algorithm, and is an important step in model development that affects model

performance (Larochelle *et al*. 2007; Smith, 2018). The hyperparameters were tuned manually when constructing the RF and MLP for papers II and IV, based on default settings and established suggestions for starting points. Manual tuning is a common approach for optimizing hyperparameters (Bergstra & Bengio 2012).

By using an ensample of MLP as approach, the hyperparameters of each model in the ensemble could be tuned on different parts of the data. Hence, this could improve the overall performance, compared to the performance of one model tuned on all data (Sollich & Krogh 1995) and should be investigated further. We combined quarter scores of different grade into one cow score, which probably masked information in several ways. Not only the density (or severity) as such, but also the as appearance of clots differed between the samples. The most diverse type of appearance was found for score 3 (data not shown) which could indicate different stages of inflammation (upcoming mastitis, self-cure or chronic case). Hence, the indicators in the milk might be different depending on the stage of inflammation or type of case, and by tuning several different models, targeting different cases, overall prediction might improve.

The cow milkings and cow milking periods free of clots were correctly categorized to a very high degree by the MLP in paper IV. Hence, it could been interesting to combine this prediction model, that accurately predicts the 'none cases' with some other method, that could better predict the milkings where clots occur. Let us say that we let the MPL first decide on which cows we do not have to consider having clots in their milk, and thereafter we can use some other prediction method for the remaining cases. Each model would then be specialized on a different task. This type of approach would be an opposite of ensemble learning, where several models are trained for the same task.

For predictions of CMSCC, the RF was the method most unaffected by the presence or absence of the variable Cow. Predictions of cases where information regarding the cows previous CMSCC was missing, could possibly be improved by using an ensemble of RF, or maybe by combining the RF with one of the other two methods investigated in paper IV. Using different types of methods as an ensemble, might improve the chance of more accurate predictions, since different methods were affected by changes in data in different ways, as observed in the model comparison in paper II.

5.3.2 Choice of variables

The first selection of potential explanatory variables in paper I was based on variable scanning using GAM, corrected for multiple comparisons and correlations. Thus, the variables selected for paper II were based on the findings in paper I. For paper IV, we selected the variables available from the herd management system that were considered to be associated with changes in milk homogeneity (for instance, the number of attachments was removed while the variable indicating failed attachment was kept). As correlation affects the coefficients and *P*-values (Kutner *et al.* 1997), variance of inflation tests, measuring the amount of multicollinearity (Fox & Monette 1992), were performed after variable scanning in paper I. However, correlation between explanatory variables will generally not affect the prediction performance of a model (Kutner *et al.* 1997) and was therefore not applied in paper IV, since the aim was not to evaluate the importance of different predictor variables. In paper IV, we did not combine quarter variables as we did in papers I and II, since MLP unlike GAM, capture the interactions between the explanatory variables (Haykin 2009).

The MLP is a "black box" algorithm and little can be known regarding how the input affects the outcome. For instance, the relative importance of adding LDH or any other variables could not be determined. We therefore evaluated the effects on prediction performance of adding LDH and daily cow activity by creating separate models for this purpose. This is opposite to the RF situation, in which variable importance can easily be determined. The RF has been suggested to be an algorithm suitable for variable selection (Genuer *et al.* 2010), but can also result misleading diagnostics (Hooker & Mentch 2019).

Although variable selection was not considered in paper IV, some initial checks were done. Variable importance was estimated by the RF, but also by a gradient boost classifier (results not shown). Both methods showed that OCC, days in milk, and MDi were important variables for predicting clots. Accordingly, the SCC as well as quarter-combined conductivity seem to be related to the occurrence of clots. Days in milk was also one of the variables found associated with positive QMS in the logistic regression in paper III, together with milking interval, lactation number and farm, also included as explanatory variables in paper IV. The outcome of the logistic regression model could probably be expected, since cows in early lactation have an increased risk of mastitis (e.g., Suriyasathaporn *et al.* 2000; Svensson *et al.*

2006; Steeneveld *et al.* 2008) as well as older cows tend to have elevated SCC and more cases of clinical mastitis (e.g., Barkema *et al.* 1998; Suriyasathaporn *et al.* 2000; Steeneveld *et al.* 2008). The effect of farm could partly reflect type of udder health related problems at the specific farm i.e., type of bacteria present. Prolonged milking interval (up to 24 hours) seems to increase the number of polymorphonuclear (PMN) cells in milk from previously healthy udders (Lakic *et al.* 2009). The proteolytic activity of proteases released from somatic cells, such as PMN cells, during inflammation, inducing the generation of para-κ-casein and leading to the precipitation of caseins in mastitis milk, has been described as one of the mechanisms (Rasmussen & Larsen 2003). Depending on factors such as the possible presence and type of bacteria and stage of inflammation in the udder, a prolonged milking interval might increase the risk of clots in milk. Hence, the presence of different type of proteases and the effect of protease specific inhibitors on clotting could give valuable information, together with information from differential cell count and also bacteriological findings. That is, a more complete biological explanation of why milk clot would make it easier to find the type data or variables and variable combinations needed for accurate predictions.

Cow behavior, such as lying time and rumination might be two interesting variables to evaluate for prediction of clots, although they would probably only have importance on the most severe cases. However, if clotting of milk is an indication of an upcoming inflammation or if clotting of milk actually is a way of handling an already ongoing inflammation, it would be interesting to investigate what comes first; shorter lying duration or clots in the milk? For CMSCC predictions, variables related to cow behavior are probably not as interesting, since subclinical mastitis is the type of mastitis free from general sickness signs. Hence, variables aiming at deviations in the udder, and above all, between the quarters should be valuable to investigate further. Comparison of curves of conductivity and flow for individual quarters during milking as well as deviations or variation in milk yield and quarter milking duration are some suggestions for this purpose.

5.3.3 Combined outcome variables

Combining predictions of the two dual-purpose indicators, i.e., clots in milk and SCC, a detection model based on the suggestions of the International Organization for Standardization (2007) can be targeted. Furthermore, this

approach could increase the possibility of pinpointing important cases, as Claycomb *et al*. (2009) found the highest sensitivity for a gold standard definition by combining SCC with the occurrence of clots in milk. Depending on how the predictions are combined, the possibility of identifying a range of mastitis cases that affect milk quality, milk homogeneity, and udder health in different ways also increases. Either OCC or samples of SCC from dairy herd improvement programs could be used for this purpose. Eventually, predictions of CMSCC could also be combined with prediction models for the occurrence of clots.

Other variables that could be interesting to evaluate in combination with clots as outcome are MDi, days in milk, milking interval or LDH. Combining clots and LDH as outcome might improve the prediction of more severe cases i.e., as LDH indicates tissue damage (Bogin *et al.* 1976) while clots most likely indicate a first stage, since clotting has been proposed as a way of isolating the bacteria. The combined outcome would thus be defined more precisely, by targeting cows indicated by different variables, which could reduce noise and improve prediction performance. Adding one of the variables MDi, days in milk or milking interval to the outcome, could help to target the cows at risk i.e., selecting the cows with higher MDi, fresh cows or cows that have longer milking intervals.

5.3.4 Model evaluation

In paper I, the corrected AIC was used as the method for model comparison that also corrects for sample size, as described by Wood *et al.* (2016). The AIC gives an estimate of how well one model in comparison with another will make predictions on new data by using maximum likelihood. The adjusted R^2 was also estimated, but only to evaluate how much variance is explained by each model. In paper II, cross-validation, an alternative approach to AIC, was one of the methods used for method comparison. Cross-validation is a useful method for comparing and evaluating the performance of different methods based on the test error. The difference between the AIC and cross-validation is that the latter will not make assumptions about any underlying distribution (e.g., normal distribution; James *et al.* 2013) and also that the prediction error for each method was obtained.

To determine the model performance on unseen data, separate datasets for testing and training could be created. This is a crucial step in model

evaluation, since predictions on the same data on which the model was trained will overestimate performance (Hastie *et al.* 2009). This approach was applied in the predictions made for paper IV, in which 70% of the data were used for training and 30% for testing, but also in one of the evaluations for paper II, in which the model was trained on data from six out of eight CMSCC sampling occasions (i.e., 75/25 split). These proportions, number of CMSCC samplings, were used to get sufficient enough data for training and prediction, however predicting CMSCC, the preferable scenario would be to predict the CMSCC just including one previous sampling (or none), together with historical from the AMR (day lags). From a practical perspective, it would be beneficial to find the cutoff for how much information regarding previous CMSCC is needed to make accurate predictions of the next CMSCC.

Another evaluation approach could be to train the model on data from one farm, and to test the model on data from another farm. This strategy was applied by Khatun *et al*. (2018), who trained their mastitis detection model at one farm and used it to predict mastitis at another farm. This approach could merit investigation if the model becomes specific to the farm where it was trained. Although the focus of paper II was comparison of prediction methods, making predictions on other farms would have made the results more generalizable, and might also have decreased the prediction performance achieved, since there was a small number of samples with high cell count in our data. This could of course also be the opposite case, if the model was trained on data with very high cell counts, it might be harder to predict the low numbers. Furthermore, due to sensor drifting, the explanatory variables measured might differ some between farms, which also could have an effect on prediction, and is something that might need consideration evaluating models on different farms.

In paper IV, data from all farms were used for both testing and training, so the effect of farm was incorporated in the model. Still, this was a limited number of farms with similar farm conditions. The model performance would probably have been different (i.e., lower) in conditions such as grazing or other type of AMS such as the automatic milking rotary. As the reasons for the occurrence of clots is not totally known, farm specific factors such as pathogens present on the farm or farm layout (i.e., cows have longer milking intervals due to long waiting times) could affect the prediction performance more than expected.

6. Conclusions

The work performed here has achieved a better understanding of how to use and combine data generated from AMS for predicting two important milk hygiene and udder health indicators: CMSCC and milk changed in homogeneity. Furthermore, relationships between CMSCC and data regularly recorded in AMS have been illustrated. In addition, the dynamics of milk homogeneity changes in cows milked in AMS have been reported for the first time. The main conclusions are:

- The most important variables in modeling and predicting CMSCC were combinations of quarter conductivity variables. The MDi, incorporating the quarter conductivity phases during milking, had the strongest association with CMSCC, followed by quarter conductivity and the difference in conductivity between quarters.
- Information regarding the cows' overall CMSCC, based on previous sampling, improved prediction performance.
- Using only three days of data before predicting CMSCC had no general disadvantages compared with using seven prior days of data.
- The variables milking interval, days in milk, lactation number, farm, and previous QMS were significantly associated with the occurrence of clots in milk.
- Elevated clot density scores were found in a limited group of cows. Higher density scores recurred more often in the subsequent milking of the same quarter than did lower density scores. The amount of flakes or traces recurred more often over time in cows having higher density scores.

- GAM are suitable for modeling relationships and predicting CMSCC. The performance of the MLP to predict CMSCC was equally good using fewer data, and had a faster conversion speed than did GAM.
- Cow milkings and cow periods free of clots were correctly categorized to a very high degree by the MLP. The prediction performance in detecting and predicting cow milkings and cow periods with clots was weak, however. The model was equally good in predicting the occurrence of clots as in detecting the occurrence of clots.
- The repeatability of low density scores was below 50%, which likely also meant that the prediction target was misguided. Clots scored as 1 and 2 should be treated as negative, targeting mastitis prediction.

7. Future perspectives

Insights into statistical methods for improving milk quality and mastitis detection have been evaluated in the framework of this thesis. However, some additional questions regarding the findings have been raised, as well as thoughts on how these findings could be applied and developed further.

Sources of SCC to include for improved model performance

The cow's previous CMSCC levels were an important variable improving prediction performance. Accordingly, applications in which the previous CMSCC levels could be included in prediction would be very interesting to evaluate further. The possibility of including CMSCC measurements from other available sources, such as sampling data from dairy herd improvement programs or OCC values, should be investigated. To predict the CMSCC of new cows entering milking, other sources of CMSCC data, such as the CMT, monthly farm SCC baseline, bulk tank SCC, or CMSCC measurements from other groups of cows, should be investigated further to improve model performance. For cows in second lactation, in addition, the cow's baseline or deviations from the baseline CMSCC in previous lactations could also be considered.

Possible applications for predicting CMSCC

Applications for predicting CMSCC between routine samplings could potentially save time and money by reducing the number of samples needed and yielding additional information on the udder heath status of individual cows. Alternatively, a group of cows could be sampled for CMSCC while the remaining cows' CMSCC values are predicted. The OCC measures the SCC for the current milking; to obtain the future CMSCC levels, OCC data could be included in the prediction model. Predictions of CMSCC could also

be included in mastitis detection models as an input variable or as part of the outcome to improve accuracy.

Improving the prediction performance of models detecting clots

The clot prediction performance was rather poor. Variations of the combined cow variables explored within the present framework should be further investigated, as how the information from single traces should be used in prediction models. Variable scanning of data available in AMS could give sources of information with which to improve models for predicting clots, i.e., insights into relationships that are important for better understanding the outcome, which is likely an important key. In addition, how to combine clot occurrence with LDH, SCC, OCC, MDi, days in milk or milking interval should be investigated further, since these variables likely indicate different types of disturbances and severity levels.

Why does milk clot?

The relationship between different types of clot cases and other indicators of inflammation deserve further investigation. Exploring the relationship between the SCC and clot occurrence could yield additional information regarding the type of case, as would bacteriological findings or the type of cells found in corresponding quarters. Furthermore, differential cell count data could give useful information regarding the type of cells involved in the milk clotting as well as proteases involved. Knowledge of the milk fraction that usually clots could give additional information regarding the types of cases that should be considered important to detect. In other words, do clots in the foremilk and clots occurring throughout the milking indicate different things? In addition, more knowledge of the severity of the recurrence of flakes and traces within the udders of some cows would add valuable input about the subject.

Do prediction models need to be general or should they be farm specific?

It is often said that results need to be generalized over several farms to be valid or reliable. However, a farm-specific model could be optimized and trained on data from the farm where it is supposed to operate. The effects of adding corrections for farm-specific factors would be interesting to incorporate into such models. For instance, information such as the SCC baseline of the farm, latest bulk tank SCC, type of pathogens from historical treatment data and previous cases of mastitis, and farm location and type

(e.g., grassland or intensive) could be utilized. Furthermore, by letting the farmer register the true cases into categories, the model could become even more specialized by training on these data.

The future is now

For predicting clots in milk, we targeted clots larger than 2 mm in size. Neither this nor previous projects using clots of this size as the gold standard in detection models have achieved impressive prediction performance or concluded that clots of all sizes above 2 mm are important to detect. Clots occur on a continuous size scale, and the size that should be considered important to monitor should be based on science. Is the suggested standard for the testing of detection systems still valid or does it need to be updated? The standard was formulated at the beginning of the AMS era, and much has happened since. Farmers, advisers, veterinarians, and researchers now have much more experience and knowledge of the challenges farmers are facing every day, searching for cows that impair overall milk quality or need action due to udder health disturbances. Hence, a revision of the standard formulated in 2007 might be in order.

> Perhaps the most important news of our day is that datasets – not algorithms – might be the key limiting factor to development of human-level artificial intelligence. (Alexander Wissner-Gross, edge.org)

References

Agabriel, C., Cornu, A., Journal, C., Sibra, C., Grolier, P. & Martin, B. (2007). Tanker milk variability according to farm feeding practices: Vitamins A and E, carotenoids, color, and terpenoids. *Journal of Dairy Science*, vol. 90 (10), pp. 4884–4896 Elsevier. DOI: https://doi.org/10.3168/jds.2007-0171

Agatonovic-Kustrin, S. & Beresford, R. (2000). Basic concepts of artificial neural network (ANN) modeling and its application in pharmaceutical research. *Journal of Pharmaceutical and Biomedical Analysis*, vol. 22 (5), pp. 717–727

Åkerstedt, M., Forsbck, L., Larsen, T. & Svennersten-Sjaunja, K. (2011). Natural variation in biomarkers indicating mastitis in healthy cows. *Journal of Dairy Research*, vol. 78 (1), pp. 88–96

Anglart, D., Hallén-Sandgren, C. & Emanuelson, U. (2019). Predicting somatic cell counts in an automatic milking rotary using conductivity and generalized additive models. In: Precision livestock farming '19. O'Brien, B., Hennessy, D., & Shalloo, L. (eds.), Cork, 2019. pp. 241–245. Cork: Teagasc Research

Ankinakatte, S., Norberg, E., Løvendahl, P., Edwards, D. & Højsgaard, S. (2013). Predicting mastitis in dairy cows using neural networks and generalized additive models: A comparison. *Computers and Electronics in Agriculture*, vol. 99, pp. 1–6

Bansal, B.K., Hamann, J., Grabowski, N.T. & Singh, K.B. (2005). Variation in the composition of selected milk fraction samples from healthy and mastitic quarters, and its significance for mastitis diagnosis. *Journal of Dairy Research*, vol. 72 (2), pp. 144–152 Teagasc Research. DOI: https://doi.org/10.1017/S0022029905000798

Barbano, D.M., Ma, Y. & Santos, M. V. (2006). Influence of raw milk quality on fluid milk shelf life. *Journal of dairy science*, vol. 89 Suppl 1 (April), pp. E15–E19. DOI: https://doi.org/10.3168/jds.s0022-0302(06)72360-8

Barkema, H.W., Schukken, Y.H., Lam, T.J.G.M., Beiboer, M.L., Wilmink, H., Benedictus, G. & Brand, A. (1998). Incidence of Clinical Mastitis in Dairy Herds Grouped in Three Categories by Bulk Milk Somatic Cell Counts. *Journal of Dairy Science*, vol. 81 (2), pp. 411–419 Elsevier. DOI: https://doi.org/10.3168/jds.S0022-0302(98)75591-2

Bergstra, J. & Bengio, Y. (2012). Random search for hyper-parameter optimization. *Journal of Machine Learning Research*, vol. 13, pp. 281–305

Bogin, E. & Ziv, G. (1973). Enzymes and minerals in normal and mastitic milk. *Cornell Vet.*, vol. 63 (4), pp. 666–676

Bogin, E., Ziv, G. & Avidar, J. (1976). Enzyme Activities in Normal and Inflamed Bovine Udder Tissues. *Zbl. Vet. Med. A*, vol. 23, pp. 460–466

Boyer, K.K. & Verma, R. (2000). Multiple raters in survey-based operations management research: A review and tutorial. *Production and Operations Management*, vol. 9 (2), pp. 128–140

Breiman, L. (2001). Random Forests. *Machine Learning*, vol. 45 (1), pp. 5–32. DOI: https://doi.org/10.1023/A:1010933404324

Cavero, D., Tölle, K.H., Buxadé, C. & Krieter, J. (2006). Mastitis detection in dairy cows by application of fuzzy logic. *Livestock Science*, vol. 105 (1–3), pp. 207–213

Cavero, D., Tölle, K.H., Henze, C., Buxadé, C. & Krieter, J. (2008). Mastitis detection in dairy cows by application of neural networks. *Livestock Science*, vol. 114 (2–3), pp. 280–286

Chagunda, M.G.G., Friggens, N.C., Rasmussen, M.D. & Larsen, T. (2006a). A model for detection of individual cow mastitis based on an indicator measured in milk. *Journal of Dairy Science*, vol. 89 (8), pp. 2980–2998

Chagunda, M.G.G., Larsen, T., Bjerring, M. & Ingvartsen, K.L. (2006b). L-lactate dehydrogenase and N-acetyl-β-D-glucosaminidase activities in bovine milk as indicators of non-specific mastitis. *Journal of Dairy Research*, vol. 73 (4), pp. 431–440

Chollet, F. (2017). *Keras (2015)*. *URL http://keras. io* Accessed Seprember 21, 2020

Claycomb, R.W., Johnstone, P.T., Mein, G.A. & Sherlock, R.A. (2009). An automated in-line clinical mastitis detection system using measurement of conductivity from foremilk of individual udder quarters. *New Zealand Veterinary Journal*, vol. 57 (4), pp. 208–214

Eberhart, R.J., Natzke, R.P., Newbould, F.H.S., Nonnecke, B. & Thompson. P (1979). Coliform Mastitis – A Review. *Journal of Dairy Science*, vol. 62 (1), pp. 1–22. DOI: https://doi.org/10.3168/jds.S0022-0302(79)83196-3

Ebrahimi, M., Mohammadi-Dehcheshmeh, M., Ebrahimie, E. & Petrovski, K.R. (2019). Comprehensive analysis of machine learning models for prediction of sub-clinical mastitis: Deep Learning and Gradient-Boosted Trees outperform other models. *Computers in Biology and Medicine*, vol. 114 (May), p. 103456 Elsevier Ltd. DOI: https://doi.org/10.1016/j.compbiomed.2019.103456

Ebrahimie, E., Ebrahimi, F., Ebrahimi, M., Tomlinson, S. & Petrovski, K.R. (2018). Hierarchical pattern recognition in milking parameters predicts mastitis prevalence. *Computers and Electronics in Agriculture*, vol. 147, pp. 6–11 Elsevier. DOI: https://doi.org/10.1016/j.compag.2018.02.003

Emanuelson, U., Olsson, T., Holmberg, O., Hageltorn, T., Mattila, T., Nelson, L. & Åström, G. (1987). Comparison of some screening tests for fetecting mastitis. *Journal of Dairy Science*, vol. 70, pp. 880–887

European Commission (2004). Regulation (EC) N° 853/2004 of the European Parlamient and of the Council of 29 April 2004 laying down specific hygiene rules for on the hygiene of foodstuffs. *Official Journal of the European Union*, vol. L 139 (853), p. 55

Fleiss, J.L. (1971). Measuring nominal scale agreement among many raters. *Psychological Bulletin*, vol. 76 (5), pp. 378–382

Fogsgaard, K.K., Bennedsgaard, T.W. & Herskin, M.S. (2015). Behavioral changes in freestall-housed dairy cows with naturally occurring clinical mastitis. *Journal of Dairy Science*, vol. 98 (3), pp. 1730–1738 Elsevier.

Forsbäck, L., Lindmark-Månsson, H., Andrén, A., Åkerstedt, M., Andrée, L. & Svennersten-Sjaunja, K. (2010). Day-to-day variation in milk yield and milk composition at the udder-quarter level. *Journal of Dairy Science*, vol. 93 (8), pp. 3569–3577 Elsevier. DOI: https://doi.org/10.3168/jds.2009-3015

Forsbäck, L., Lindmark-Månsson, H., Åndrén, A., Kerstedt, M. & Svennersten-Sjaunja, K. (2009). Udder quarter milk composition at different levels of somatic cell count in cow composite milk. *Animal*, vol. 3 (5), pp. 710–717

Fox, J. & Monette, G. (1992). Generalized collinearity diagnostics. *Journal of the American Statistical Association*, vol. 87 (417), pp. 178–183

Franzén, J., Thorburn, D., Urioste, J.I. & Strandberg, E. (2012). Genetic evaluation of mastitis liability and recovery through longitudinal analysis of transition probabilities. *Genetics Selection Evolution*, vol. 44 (1), pp. 1–10

Genuer, R., Poggi, J.M. & Tuleau-Malot, C. (2010). Variable selection using random forests. *Pattern Recognition Letters*, vol. 31 (14), pp. 2225–2236

Giesecke, W.H. & van den Heever, L.W. (1974). The diagnosis of bovine mastitis with particular reference to subclinical mastitis: a critical review of relevant literature. *Onderstepoort Journal of Veterinary Research*, vol. 41 (4), pp. 169–211

Green, L.E., Schukken, Y.H. & Green, M.J. (2006). On distinguishing cause and consequence: Do high somatic cell counts lead to lower milk yield or does high milk yield lead to lower somatic cell count? *Preventive Veterinary Medicine*, vol. 76 (1–2), pp. 74–89. DOI: https://doi.org/10.1016/j.prevetmed.2006.04.012

Haidet, K.K., Tate, J., Divirgilio-Thomas, D., Kolanowski, A. & Happ, M.B. (2009). Methods to improve reliability of video-recorded behavioral data. *Research in Nursing and Health*, vol. 32 (4), pp. 465–474

Halasa, T., Huijps, K., Østerås, O. & Hogeveen, H. (2007). Economic effects of bovine mastitis and mastitis management: A review. *Veterinary Quarterly*, vol. 29 (1), pp. 18–31

Hamann, J., & Krömker, V. (1997). Potential of specific milk composition variables for cow health management. *Livestock Production Science*, vol 48(3), pp. 201–208.

Hamann, J. & Zecconi, A. (1998). Evaluation of the electrical conductivity of milk as a mastitis indicator. *Bulletin-FIL-IDF (Belgium)*, vol. 334 (January), pp. 5–22.

Hammer, J.F., Morton, J.M. & Kerrisk, K.L. (2012). Quarter-milking-, quarter-, udder- and lactation-level risk factors and indicators for clinical mastitis during lactation in pasture-fed dairy cows managed in an automatic milking system. *Australian Veterinary Journal*, vol. 90 (5), pp. 167–174

Hassan, K.J., Samarasinghe, S. & Lopez-Benavides, M.G. (2009). Use of neural networks to detect minor and major pathogens that cause bovine mastitis. *Journal of Dairy Science*, vol. 92 (4), pp. 1493–1499

Hastie, T., Tibshirani, R. & Friedman, J. (2009). *The Elements of Statistical Learning*. New York, NY: Springer New York. DOI: https://doi.org/10.1007/978-0-387-84858-7

Hastie, T.J. & Tibshirani, R.J. (1990). *Generalized additive models*. New York: Chapman and Hall, New York.

Haykin, S. (2009). *Neural networks and learning machines*. 3rd. ed. Upper Saddle River: Pearson Education.

He, H. & Ma, Y. (2013). *Imbalanced learning: foundations, algoritms and applications*. New Jersey: John Wiley & sonc, Inc.

Heald, C.W., Kim, T., Sischo, W.M., Cooper, J.B. & Wolfgang, D.R. (2000). A computerized mastitis decision aid using farm-based records: An artificial neural network approach. *Journal of Dairy Science*, vol. 83 (4), pp. 711–720 Elsevier. DOI: https://doi.org/10.3168/jds.S0022-0302(00)74933-2

Hillerton, J.E. (2000). Detecting Mastitis Cow-Side., 2000. pp. 48–53

Hogeveen, H., Huijps, K. & Lam, T.J.G.M. (2011). Economic aspects of mastitis: New developments. *New Zealand Veterinary Journal*, vol. 59 (1), pp. 16–23

Hogeveen, H., Kamphuis, C., Steeneveld, W. & Mollenhorst, H. (2010). Sensors and clinical mastitis-the quest for the perfect alert. *Sensors*, vol. 10 (9), pp. 7991–8009

Hogeveen, H. & Ouweltjes, W. (2011). MASTITIS THERAPY AND CONTROL | Automatic On-Line Detection of Abnormal Milk. *Encyclopedia of Dairy Sciences*, vol. 3, pp. 1735–1740

Hogeveen, H., Ouweltjes, W., Koning, C.J.A.M. De & Stelwagen, K. (2001). Milking interval , milk production and milk flow-rate in an automatic milking system. vol. 72, pp. 157–167

Hooker, G. & Mentch, L. (2019). Please stop permuting features an explanation and alternatives. *arXiv*, pp. 1–15

Hovinen, M. & Pyörälä, S. (2011). Invited review: Udder health of dairy cows in automatic milking. *Journal of Dairy Science*, vol. 94 (2), pp. 547–562

International Dairy Federation (IDF) (2011). *Suggested interpretation of mastitis terminology*. Vol.448, Bulletin IDF. Brussels, Belgium.

International Dairy Federation (IDF) (2013). *Guidelines for the use and interpretation of bovine milk somatic cell count.* Vol.466, Bulletin IDF. Brussels, Belgium.

International Organization for Standardization (2005). *ISO/IEC 17025 — General requirements for the competence of testing and calibration laboratories* Available at: https://www.iso.org/publication/PUB100424.html

International Organization for Standardization (2006). *ISO 13366-2:2006 [IDF 148-2:2006] Milk — Enumeration of somatic cells — Part 2: Guidance on the operation of fluoro-opto-electronic counters*

International Organization for Standardization (2007). *ISO 20966:2007— Automatic milking installations-Requirements and testing Installations de traite automatique-Exigences et essais. 1sr edition. Geneva, Switzerland.*

International Organization for Standardization (2008). *ISO 707:2008 [IDF 50:2008] Milk and milk products — Guidance on sampling* Available at: https://www.iso.org/standard/37882.html

James, G., Witten, D., Hastie, T. & Tibishirani, R. (2013). *An Introduction to Statistical Learning with Applications in R*. *Springer Texts in Statistics* 6th. ed. New York: Springer Science & Business Media.

Jensen, D.B., Hogeveen, H. & De Vries, A. (2016). Bayesian integration of sensor information and a multivariate dynamic linear model for prediction of dairy cow mastitis. *Journal of Dairy Science*, vol. 99 (9), pp. 7344–7361 American Dairy Science Association. DOI: https://doi.org/10.3168/jds.2015-10060

Kamphuis, C., Pietersma, D., van der Tol, R., Wiedemann, M. & Hogeveen, H. (2008a). 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*, vol. 62 (2), pp. 169–181

Kamphuis, C., Sherlock, R., Jago, J., Mein, G. & Hogeveen, H. (2008b). Automatic detection of clinical mastitis is improved by in-line monitoring of somatic cell count. *Journal of dairy science*, vol. 91 (12), pp. 4560–70. DOI: https://doi.org/10.3168/jds.2008-1160

Kamphuis, C. (2010). *Making sense of sensor data: detecting clinical mastitis in automatic milking systems*. Diss. Utrecht University.

Kamphuis, C., Mollenhorst, H., Heesterbeek, J.A.P. & Hogeveen, H. (2010). Detection of clinical mastitis with sensor data from automatic milking systems is improved by using decision-tree induction. *Journal of Dairy Science*, vol. 93 (8), pp. 3616–3627

Kamphuis, C., Dela Rue, B., Mein, G. & Jago, J. (2013). Development of protocols to evaluate in-line mastitis-detection systems. *Journal of Dairy Science*, vol. 96 (6), pp. 4047–4058. DOI: https://doi.org/10.3168/jds.2012-6190

Kamphuis, C., Dela Rue, B.T. & Eastwood, C.R. (2016). Field validation of protocols developed to evaluate in-line mastitis detection systems. *Journal*

of Dairy Science, vol. 99 (2), pp. 1619–1631. DOI: https://doi.org/10.3168/jds.2015-10253

Khatun, M., Clark, C.E.F., Lyons, N.A., Thomson, P.C., Kerrisk, K.L. & García, S.C. (2017). Early detection of clinical mastitis from electrical conductivity data in an automatic milking system. *Animal Production Science*, vol. 57 (7), p. 1226. DOI: https://doi.org/10.1071/AN16707

Khatun, M., Thomson, P.C., Kerrisk, K.L., Lyons, N.A., Clark, C.E.F., Molfino, J. & García, S.C. (2018). Development of a new clinical mastitis detection method for automatic milking systems. *Journal of Dairy Science*, vol. 101, pp. 9385–9295

Kingma, D.P. & Ba, J.L. (2015). Adam: A method for stochastic optimization. *3rd International Conference on Learning Representations, ICLR 2015 - Conference Track Proceedings*, pp. 1–15

Kitchen, B.J. (1981). Review of the progress of dairy science: Bovine mastitis: Milk compositional changes and related diagnostic tests. *Journal of Dairy Research*, vol. 48 (1), pp. 167–188

Koldeweij, E., Emanuelson, U. & Janson, L. (1999). Relation of Milk Production Loss to Milk Somatic Cell Count. *Acta Veterinaria Scandinavica*, vol. 40 (1), pp. 47–56

de Koning, C.J.A.M. (Kees) (2010). Automatic Milking – Common Practice on Dairy Farms., Lelystad, Netherlands, 2010. p. 16. Lelystad, Netherlands: Wageningen UR Livestock Research,

Krawczel, P., Ferneborg, S., Wiking, L., Dalsgaard, T.K., Gregersen, S., Black, R., Larsen, T., Agenäs, S., Svennersten-Sjaunja, K. & Ternman, E. (2017). Milking time and risk of over-milking can be decreased with early teat cup removal based on udder quarter milk flow without loss in milk yield. *Journal of Dairy Science*, vol. 100 (8), pp. 6640–6647

Kutner, M., Nachtsheim, C. & Neter, J. (1997). *Applied Linear Statistical Models*. (T, R. & Jr, H., eds.) *Journal of Quality Technology* 4th. ed. New York: McGraw-Hill/Irwin.

Lakic, B., Wredle, E., Svennersten-Sjaunja, K. & Stensson, K. (2009). Is there a special mechanism behind the changes in somatic cell and polymorphonuclear leukocyte counts, and composition of milk after a single prolonged milking interval in cows? *Acta Veterinaria Scandinavica*, vol. 51 (1)

Landis, J. & Koch, G. (1977). The measurement of observer agreement for categorical data. *Biometrics*, vol. 33 (1), pp. 159–174

Larochelle, H., Erhan, D., Courville, A., Bergstra, J. & Bengio, Y. (2007). An empirical evaluation of deep architectures on problems with many factors of variation. *ACM International Conference Proceeding Series*, vol. 227 (2006), pp. 473–480

Larsen, T., Rontved, C.M., Ingvartsen, K.L., Vels, L. & Bjerring, M. (2010). Enzyme activity and acute phase proteins in milk utilized as indicators of acute clinical E. coli LPS-induced mastitis. *Animal*, vol. 4 (10), pp. 1672–1679

Leitner, G., Merin, U., Krifucks, O., Blum, S., Rivas, A.L. & Silanikove, N. (2012). Effects of intra-mammary bacterial infection with coagulase negative staphylococci and stage of lactation on shedding of epithelial cells and infiltration of leukocytes into milk: Comparison among cows, goats and sheep. *Veterinary Immunology and Immunopathology*, vol. 147 (3–4), pp. 202–210 Elsevier B.V.

Lind, O., Ipema, A.H., de Koning, C.J.A.M. (Kees) & Mottram, T.. (2000). Automatic milking. *Bulletin of the IDF*, (348), pp. 3–14

Linzell, J.L. & Peaker, M. (1971). Intracellular concentrations of sodium, potassium and chloride in the lactating mammary gland and their relation to the secretory mechanism. *The Journal of Physiology*, vol. 216 (3), pp. 683–700

Lohuis, J.A.C.M., Schukken, Y.H., Verheijden, J.H.M., Brand, A. & Van Miert, A.S.J.P.A.M. (1990). Effect of Severity of Systemic Signs During the Acute Phase of Experimentally Induced Escherichia coli Mastitis on Milk Production Losses. *Journal of Dairy Science*, vol. 73 (2), pp. 333–341 Elsevier. DOI: https://doi.org/10.3168/jds.S0022-0302(90)78678-X

Lusis, I., Antane, V. & Laurs, A. (2017). Effectiveness of mastitis detection index for cow monitoring and abnormal milk detection in milking robots., May 24 2017. pp. 1383–1387

Mammadova, N.M. & Keskin, I. (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 (OF), pp. 671–679

Medrano-Galarza, C., Gibbons, J., Wagner, S., de Passillé, A.M. & Rushen, J. (2012). Behavioral changes in dairy cows with mastitis. *Journal of Dairy Science*, vol. 95 (12), pp. 6994–7002 Elsevier.

Mein, G.A. & Rasmussen, M.D. (2008). Perfromance of system for automated monitoring of udder health: would the real gold standard please stand up? In: Lam, T.J.G.M. (ed.), Wageningen, 2008. pp. 259–266. Wageningen: Wagening Academic

Miller, J. & Franklin, J. (2002). Modeling the distribution of four vegetation alliances using generalized linear models and classification trees with spatial dependence. *Ecological Modelling*, vol. 157 (2–3), pp. 227–247

De Mol, R.M. & Ouweltjes, W. (2001). Detection model for mastitis in cows milked in an automatic milking system. *Preventive Veterinary Medicine*, vol. 49 (1–2), pp. 71–82

Mollenhorst, H., Hidayat, M.M., van den Broek, J., Neijenhuis, F. & Hogeveen, H. (2011). The relationship between milking interval and somatic cell count in

automatic milking systems. *Journal of Dairy Science*, vol. 94 (9), pp. 4531–4537 Elsevier. DOI: https://doi.org/10.3168/jds.2011-4244

Mollenhorst, H., Rijkaart, L.J. & Hogeveen, H. (2012). Mastitis alert preferences of farmers milking with automatic milking systems. *Journal of Dairy Science*, vol. 95 (5), pp. 2523–2530

Mollenhorst, H., van der Tol, P.P. & Hogeveen, H. (2010). Somatic cell count assessment at the quarter or cow milking level. *J Dairy Sci*, vol. 93 (7), pp. 3358–3364. DOI: https://doi.org/10.3168/jds.2009-2842

Montgomery, M.E., White, M.E. & Martin, S.W. (1987). A comparison of discriminant analysis and logistic regression for the prediction of coliform mastitis in dairy cows. *Canadian journal of veterinary research, Revue canadienne de recherche vétérinaire*, vol. 51 (4), pp. 495–498

Muhammad, P.R.G., Expert, S. & Rashid, I. (2015). Causes , Diagnosis and Treatment of Blood in Milk (Hemolactia) in Cows and Buffaloes. *World Wide Web*, vol. 2 (1), pp. 1–6

Nielen, M., Schukken, Y., Brand, A., Haring, S. & Ferwerda-Van Zonneveld, R. (1995a). Comparison of Analysis Techniques for On-Line Detection of Clinical Mastitis. *Journal of Dairy Science*, vol. 78, pp. 1050–1061

Nielen, M., Spigt, M.H., Schukken, Y.H., Deluyker, H.A., Maatje, K. & Brand, A. (1995b). Application of a neural network to analyse on-line milking parlour data for the detection of clinical mastitis in dairy cows. *Preventive Veterinary Medicine*, vol. 22 (1–2), pp. 15–28

Nielsen, C. & Emanuelson, U. (2013). Mastitis control in Swedish dairy herds. *Journal of Dairy Science*, vol. 96 (11), pp. 6883–6893

National Mastitis Council. 2013. Recommended Milking Procedures, National Mastitis Council. .http://www.nmconline.org/wp-content/uploads/2016/09/Recommended-Milking-Procedures.pdf. Accessed: December, 13, 2020

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

Nørstebø, H., Dalen, G., Rachah, A., Heringstad, B., Whist, A.C., Nødtvedt, A. & Reksen, O. (2019). Factors associated with milking-to-milking variability in somatic cell counts from healthy cows in an automatic milking system. *Preventive Veterinary Medicine*, vol. 172 (May), p. 104786 Elsevier. DOI: https://doi.org/10.1016/j.prevetmed.2019.104786

Nyman, A.-K., Persson Waller, K., Bennedsgaard, T.W., Larsen, T. & Emanuelson, U. (2014). Associations of udder-health indicators with cow factors and with intramammary infection in dairy cows. *Journal of Dairy Science*, vol. 97 (9), pp. 5459–5473 Elsevier. DOI: https://doi.org/10.3168/jds.2013-7885

Panchal, I., Sawhney, I.K., Sharma, A.K. & Dang, A.K. (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*, vol. 127, pp. 242–248 Elsevier B.V. DOI: https://doi.org/10.1016/j.compag.2016.06.015

Pastell, M., Tiusanen, J., Hakojärvi, M. & Hänninen, L. (2009). A wireless accelerometer system with wavelet analysis for assessing lameness in cattle. *Biosystems Engineering*, vol. 104 (4), pp. 545–551

Penry, J.F., Endres, E.L., de Bruijn, B., Kleinhans, A., Crump, P.M., Reinemann, D.J. & Hernandez, L.L. (2016). Effect of incomplete milking on milk production rate and composition with 2 daily milkings. *Journal of Dairy Science*, vol. 100 (2), pp. 1535–1540 American Dairy Science Association. DOI: https://doi.org/10.3168/jds.2016-11935

Pinzón-Sánchez, C. & Ruegg, P. (2011). Risk factors associated with short-term post-treatment outcomes of clinical mastitis. *Journal of Dairy Science*, vol. 94, pp. 3397–3410

Politis, I. & Ng-Kwai-Hang, K.F. (1988). Effects of Somatic Cell Count and Milk Composition on Cheese Composition and Cheese Making Efficiency. *Journal of Dairy Science*, vol. 71 (7), pp. 1711–1719 Elsevier. DOI: https://doi.org/10.3168/jds.S0022-0302(88)79737-4

Pujari, P. & Gupta, J.B. (2012). Improving Classification Accuracy by Using Feature Selection and Ensemble Model. *International Journal of Soft Computing and Engineering (IJSCE)*, vol. 2 (2), pp. 380–386

Pyörälä, S. (2003). Indicators of inflammation in the diagnosis of mastitis. *Veterinary Research*, vol. 34 (5), pp. 565–578

Pyörälä, S., Hovinen, M., Simojoki, H., Fitzpatrick, J., Eckersall, P.D. & Orro, T. (2011). Acute phase proteins in milk in naturally acquired bovine mastitis caused by different pathogens. *Veterinary Record*, vol. 168 (20), p. 535

Pyörälä, S. & Pyörälä, E. (1997). Accuracy of Methods Using Somatic Cell Count and N-Acetyl-β-D-Glucosaminidase Activity in Milk to Assess the Bacteriological Cure of Bovine Clinical Mastitis. *Journal of Dairy Science*, vol. 80 (11), pp. 2820–2825

Pyörälä, S. & Syväjärvi, J. (1987). Bovine Acute Mastitis Part I. Clinical Aspects and Parameters of Inflammation in Mastitis Caused by Different Pathogens. *Journal of Veterinary Medicine, Series B*, vol. 34 (1–10), pp. 573–584. DOI: https://doi.org/10.1111/j.1439-0450.1987.tb00438.x

Quist, M.A., LeBlanc, S.J., Hand, K.J., Lazenby, D., Miglior, F. & Kelton, D.F. (2008). Milking-to-Milking Variability for Milk Yield, Fat and Protein Percentage, and Somatic Cell Count. *Journal of Dairy Science*, vol. 91 (9), pp. 3412–3423 Elsevier. DOI: https://doi.org/10.3168/jds.2007-0184

Rasmussen, M.D. (2004). Detection and separation of abnormal milk in automatic milking systems. *Automatic Milking - a better understanding*, pp. 189–197

Rasmussen, M.D. (2005). Visual scoring of clots in foremilk. *Journal of Dairy Research*, vol. 72 (4), pp. 406–414

Rasmussen, M.D. & Bjerring, M. (2005). Visual scoring of milk mixed with blood. *Journal of Dairy Research*, vol. 72 (3), pp. 257–263

Rasmussen, M.D., Bjerring, M. & Skjøth, F. (2005). Visual appearance and CMT score of foremilk of individual quarters in relation to cell count of cows milked automatically. *Journal of Dairy Research*, vol. 72 (1), pp. 49–56

Rasmussen, M.D. & Larsen, B.L. (2003). Milking hygiene: new issues and opportunities from automatic milking. *Italian Journal of Animal Science*, vol. 2 (4), pp. 283–289

Rosenblatt, F. (1958). The perceptron: A probabilistic model for information storage and organization in the brain. *Psychological Review*, vol. 65 (6), pp. 386–408. DOI: https://doi.org/10.1037/h0042519

Rossing, W. & Hogewerf, P.H. (1997). State of the art of automatic milking systems. *Computers and Electronics in Agriculture*, vol. 17 (1), pp. 1–17

Ruegg, P.L. (2003). Investigation of mastitis problems on farms. *Veterinary Clinics of North America - Food Animal Practice*, vol. 19 (1), pp. 47–73

Ruegg, P.L. (2012). New Perspectives in Udder Health Management. *Veterinary Clinics of North America - Food Animal Practice*, vol. 28 (2), pp. 149–163 Elsevier.

Rutten, C.J., Velthuis, A.G.J., Steeneveld, W. & Hogeveen, H. (2013). Invited review: Sensors to support health management on dairy farms. *Journal of Dairy Science*, vol. 96 (4), pp. 1928–1952. DOI: https://doi.org/10.3168/jds.2012-6107

Schmidt, G.H. & Smith, T.R. (1986). Use of Dairy Herd Improvement Testing Programs by Dairy Farmers. *Journal of Dairy Science*, vol. 69 (12), pp. 3156–3164

Schmidt Madsen, P. (1975). Fluoro-opto-electronic cell-counting on milke. *Journal of Dairy Research*, vol. 42 (2), pp. 227–239

Schukken, Y.H., Wilson, D.J., Welcome, F., Garrison-Tikofsky, L. & Gonzalez, R.N. (2003). Monitoring udder health and milk quality using somatic cell counts. *Veterinary Research*, vol. 34 (5), pp. 579–596

Siivonen, J., Taponen, S., Hovinen, M., Pastell, M., Lensink, B.J., Pyörälä, S. & Hänninen, L. (2011). Impact of acute clinical mastitis on cow behaviour. *Applied Animal Behaviour Science*, vol. 132 (3–4), pp. 101–106 Elsevier B.V.

Sitowska, B., Piwczyński, D., Aerts, J., Kolenda, M. & Özkaya, S. (2017). Detection of high levels of somatic cells in milk on farms equippedwith an automatic milking system by decision trees technique. *Turkish Journal of Veterinary and Animal Sciences*, vol. 41, pp. 532–540

Smith, L.N. (2018). a Disciplined Approach To Neural Network Hyper-Parameters: Part 1 – Learning Rate, Batch Size, Momentum, and Weight Decay. *arXiv*, pp. 1–21

Sollich, P. & Krogh, A. (1995). Learning with ensembles : how over-fitting can be useful., Denver, Colorado, 1995. pp. 4–10. Denver, Colorado

Sørensen, L.P., Bjerring, M. & Løvendahl, P. (2016). Monitoring individual cow udder health in automated milking systems using online somatic cell counts. *Journal of Dairy Science*, vol. 99 (1), pp. 608–620 Elsevier. DOI: https://doi.org/10.3168/jds.2014-8823

Steeneveld, W., Hogeveen, H., Barkema, H.W., Van Den Broek, J. & Huirne, R.B.M. (2008). The influence of cow factors on the incidence of clinical mastitis in dairy cows. *Journal of Dairy Science*, vol. 91 (4), pp. 1391–1402

Stekhoven, D.J. (2013). missForest: Nonparametic Missing Value Imputation using Random Forest. R Package version 1.4. Available at: http://cran.r-project.org/web/packages/missForest/index.html Assessed June 26, 2020

Sun, Z., Samarasinghe, S. & Jago, J. (2010). Detection of mastitis and its stage of progression by automatic milking systems using artificial neural networks. *Journal of Dairy Research*, vol. 77 (2), pp. 168–175

Suriyasathaporn, W., Schukken, Y.H., Nielen, M. & Brand, A. (2000). Low somatic cell count: A risk factor for subsequent clinical mastitis in a dairy herd. *Journal of Dairy Science*, vol. 83 (6), pp. 1248–1255 Elsevier. DOI: https://doi.org/10.3168/jds.S0022-0302(00)74991-5

Svennersten-Sjaunja, K.M. & Pettersson, G. (2008). Pros and cons of automatic milking in Europe. *Journal of animal science*, vol. 86 (13 Suppl), pp. 37–46

Svensson, C., Nyman, A.K., Waller, K.P. & Emanuelson, U. (2006). Effects of housing, management, and health of dairy heifers on first-lactation udder health in southwest Sweden. *Journal of Dairy Science*, vol. 89 (6), pp. 1990–1999 Elsevier. DOI: https://doi.org/10.3168/jds.S0022-0302(06)72266-4

Tyler, J.W., Thurmond, M.C. & Lasslo, L. (1989). Relationship between test-day measures of somatic cell count and milk production in California dairy cows. *Canadian journal of veterinary research, Revue canadienne de recherche vétérinaire*, vol. 53 (2), pp. 182–187

Viguier, C., Arora, S., Gilmartin, N., Welbeck, K. & O'Kennedy, R. (2009). Mastitis detection: current trends and future perspectives. *Trends in Biotechnology*, vol. 27 (8), pp. 486–493

Wenz, J.R., Barrington, G.M., Garry, F.B., Dinsmore, R.P. & Callan, R.J. (2001). Use of systemic disease signs to assess disease severity in dairy cows with acute coliform mastitis. *Journal of the American Veterinary Medical Association*, vol. 218 (4), pp. 567–572

Wenz, J.R., Jensen, S.M., Lombard, J.E., Wagner, B.A. & Dinsmore, R.P. (2007). Herd management practices and their association with Bulk tank somatic cell count on United States Dairy perations. *Journal of Dairy Science*, vol. 90 (8), pp. 3652–3659.

Wiener, A. & Liaw, M. (2002). Classification and Regression by randomForesttle. *R News*, vol. 2 (3), pp. 18–22. Available at: https://cran.r-project.org/doc/Rnews/

Wong, N.P., Jenness, R., Keeney, M. & Marth, E.H. (1988). *Fundamentals of dairy chemistry*. Springer Science & Business Media.

Wood, S. (2011). Fast stable restricted maximum likelihood and marginal likelihood estimation of semiparametric generalized linear models. *Journal of the Royal Statistical Society*, vol. 73 (1), pp. 3–36

Wood, S.N., Pya, N. & Säfken, B. (2016). Smoothing Parameter and Model Selection for General Smooth Models. *Journal of the American Statistical Association*, vol. 111 (516), pp. 1548–1563

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*, vol. 65 (2), pp. 187–198

Zhang, Y. & Ling, C. (2018). A strategy to apply machine learning to small datasets in materials science. *npj Computational Materials*, vol. 4 (1), pp. 28–33 Springer US. DOI: https://doi.org/10.1038/s41524-018-0081-z

Popular science summary

Delivering high-quality milk from healthy dairy cows is not only important for dairy farmers but is also regulated by the law in many countries, which also ensures safe and high quality dairy products for the consumer. Traditionally, the milker pre-strips the quarters and inspects the stripped milk before attaching the milking unit. The milker is looking for deviations such as clots, color changes, or other abnormalities in the milk in order to identify sick cows and prevent abnormal milk from ending up in the bulk tank. In addition, regular samplings measuring the levels of somatic cells (i.e., white blood cells) in the milk are performed to monitor the udder health as well as milk quality.

Among cows milked by milking robots, the milk is inspected by sensors measuring milk characteristics such as conductivity or color, alerting the farmer if there are any deviations. However, no detection system for detecting and predicting milk clots has yet been investigated. Additionally, knowledge of how often clots occur and why in the milk of cows milked by milking robots is scarce. The level of somatic cells can be measured automatically in the milking robot; however, this involves additional costs and is not possible in all types of robot systems. In this thesis, the possibility of predicting somatic cell counts and clots in dairy cow milk using sensor data from the milking robot was investigated. We also wanted to investigate the relationship between the milk characteristics measured by sensors and the somatic cell count as well as investigate the dynamics of clots in cows milked by milking robots.

The first part of this project focused on the relationship between the sensor data generated by the milking robot and the somatic cell count in the cow's milk. Quarter conductivity was found to be a key feature explaining the levels of somatic cells. The relationship between the quarters in terms of

conductivity was the most important explanatory variable. We also investigated for how many days the data should be recorded in order to make accurate predictions of the level of somatic cells. With knowledge of how the sensor data should best be used, three statistical methods for predicting somatic cell counts for individual cows were compared. The methods had different abilities and operated on the data in different ways. The results indicated that using three days of data was as good as seven days of data and that information about the cows' previous somatic cell level was important in order to predict the next. Two of the compared methods were slightly better than the third for predicting the somatic cell count.

In the second part, the dynamics of the occurrence of clots in milk were mapped by installing filters in the milk line and assessing them for clots according to a scale of how much of the filter was covered with clots. We found clots in the milk from a limited number of cows and that the larger the filter area covered with clots, the greater the risk of having clots again in the same quarter. We also found that cows that go a long time between milkings in the robot tend to have clots in their milk more often. Finally, we built several models to detect and predict the occurrence of clots in a single milking as well as during consecutive periods. The results indicated that the model was very good in classifying which cows had no clots in the milk during a single milking or consecutive periods. However, the model was generally not good in identifying cows that did have clots in their milk.

In conclusion, it is possible to use sensor data from the milking robot to predict the level of somatic cells in the cow's milk, but this finding needs to be investigated in other conditions, such as other farms. How these findings could be used to support farmers by supplying additional information about the cows' somatic cell levels between sampling periods should also be investigated. The models for predicting clots in milk were inadequate to use in practice. Based on the findings regarding clot dynamics, we suggest that the severity grade of clots that need to be detected merits further study. By having our models find all the clots, we may have misjudged the goal of the detection. More knowledge is also required regarding the biological explanations of why milk clots could help improve the performance of detection models.

Populärvetenskaplig sammanfattning

Mjölkkornas hälsa är viktigt för mjölkbonden och att leverera mjölk av hög kvalité till mejeriet är även reglerat i lag i många länder, vilket säkerhetsställer att konsumenterna får säkra och högkvalitativa mejeriprodukter produkter. Traditionellt inspekterar den som utför mjölkningen mjölken innan mjölkningskopparna kopplas på. Mjölkaren undersöker då mjölken genom att titta efter flockor, färgförändrningar eller andra avvikelser i syfte att hitta sjuka kor men också för att hindra dålig mjölk från att nå mjölktanken. Celltalsnivån i kornas mjölk (d.v.s. mängden vita blodkroppar) mäts också regelbundet, för att övervaka mjölkkvalitén och likaså kornas juverhälsa.

I anläggningar där korna mjölkas med mjölkningsrobotar övervakas mjölken av sensorer som mäter olika egenskaper i mjölken som konduktivitet och färg. Sensorerna är kopplade till larm som kan varna bonden om något ser konstigt ut. Trots detta har inget system för att upptäcka och förutse flockor i mjölken undersökts. Dessutom är kunskapen om hur ofta och varför flockor uppkommer begränsad, i synnerhet bland kor som mjölkas i robotsystem. Celltalsnivån i mjölken kan mätas, men detta är dyrt och är dessutom inte möjligt i att göra automatiskt i alla typer utav system. I denna avhandling undersöktes modeller för att förutse celltalsnivå och flockor i mjölk genom att använda sensordata från mjölkroboten. För att bättre förstå hur modellerna skulle byggas, undersöktes även sambanden mellan celltalsnivån och det sensordata som kommer ifrån roboten, samt dynamiken och flockornas förekomst hos kor som mjölkas i robotsystem.

Projektets första del fokuserade på sambandet mellan data från sensorer, variabler, och celltalsnivå. Undersökningen visade att ledningsförmåga på fjärdelsnivå var en nyckelvariabel, och att relationen mellan olika fjärdedelars ledningsförmåga var de viktigaste variablerna för att förklara

celltalsnivån. Vi undersökte också hur många dagars sensordata som behövde samlas in för att göra korrekta förutsägelser av celltalsnivå. Med kunskapen om hur sensordata bäst skulle användas jämfördes tre statistiska modeller för att förutse celltalsnivå för enskilda kor. Modellerna hade olika egenskaper och använde data på olika sätt. Resultatet visade att tre dagars sensordata gav lika bra förutsägelse om celltalet som sju dagars sensordata, och att information om kornas tidigare celltalsnivå var viktigt för att förutsägelsen skulle bli så korrekt som möjligt. Två av modellerna var något bättre på att förutse celltalsnivå än den tredje.

I projektets andra del undersöktes flockförekomsten och dynamik genom att installera filter i mjölkledingen som går till mjölkkopparna. Förekomsten av flockor bedömdes genom att filtren graderades utifrån hur stor del av dem som var täckt av flockor. Undersökningen visade att ett begränsat antal kor hade flockor i sin mjölk, men att för de kor vars prover hade flocker på en större del av filtret var risken även större att de skulle få flocker igen i samma juverdel. Vi upptäckte också att kor som inte gick till mjölkningen lika ofta tenderade att ha flockor i mjölken oftare. Slutligen byggde vi flera modeller för att upptäcka och förutse flockförekomst både i enskilda mjölkningar och över längre sammanhängande perioder. Resultaten visade att modellen var bra på att peka ut vilka kor som inte hade flockor i mjölken, men att den däremot var sämre på att hitta de kor som faktiskt hade flockor i sin mjölk.

Sammanfattningsvis har vi sett att det är möjligt att använda sensordata från mjölkningsroboten för att förutse celltalsnivå, men detta behöver undersökas vidare till exempel på andra gårdar. Hur dessa resultat kan användas för att hjälpa bönder genom att tillhandahålla ytterligare information om kornas celltalsnivå mellan provperioder behöver också undersökas vidare. Modellerna för att förutse flockor i mjölken visade sig inte vara tillräckligt bra för att kunna använda praktiskt. Baserat på resultaten gällande dynamiken i flockförekomst föreslår vi att den nivå av flockor som behöver upptäckas bör undersökas vidare. Genom att låta vår modell hitta alla flockor kan vi ha missriktat målet för detektionen. Mer kunskap om den biologiska förklaringen till uppkomsten av flockor kan också förbättra hur bra modellerna blir på att hitta dem.

Acknowledgements

The work of this thesis was carried out at the Department of Clinical Sciences at the Swedish University of Agricultural Sciences and DeLaval International AB. Founding was provided by the Swedish Foundation for Strategic Research (SSF).

I feel like a very lucky person! Not only did my dream about doing a PhD come true, I could also combine expertise from **DeLaval** and **SLU** into my work. There are so many people involved in the making of this thesis, it had not been possible without you! I would like to thank you all, with special thanks to:

My main supervisor, **Ulf Emanuelson**. Your patience is golden and your accuracy is outstanding. I could not have had better guidance throughout this project! It felt safe to know that you were always an e-mail (or several) away, giving good advice on great and small things of all kind. If there is something you don't know, it is probably not worth knowing! Thank you for everything and sorry for all the times you had to change an i.e. to an e.g. – or was it the other way around?!

Charlotte Hallén Sandgren, I am so impressed by your expertise and your quick mind, but above all that, just your way of being you! Thank you for the never ending enthusiasm and engagement in this project, always pushing for the best, always ready to answer my questions regardless of the time of the day. You should know that I am so truly thankful that you called me that day and asked if I wanted to join this project!

Lars Rönnegård, I am so thankful that you decided to just jump in to this project! No question was never too stupid or to simple. I really appreciate

your way of encouraging to learn more and making me think by asking the right questions. It has been a privilege to have you in the group!

It has not always been easy having three supervisors going towards three different directions (and me to a fourth), but I think the three of you formed a perfect group thanks to your differences. It has been a really fun group to work in!

Patrik Waldmann, thank you for giving me the opportunity to learn the basics in machine learning and introducing me to the book that became my bible during these years. **Claudia von Brömssen** and **Staffan Betnér,** thank you for all the statistical support and patience with all questions!

Collecting data is a challenge, and it had not been possible without help! **Johan Östlund**, jag är så otroligt tacksam att just du ställde upp och hjälpte mig med datainsamling, bidrog med goda råd och idéer kring det praktiska! **Harry Tuinier**, I am happy that we met and forever thankful for you participating in the crazy data collection by nights!

Henrik Veldman, I could not thank you enough for your help, support and hospitality from you and your family! Also big thanks to, **Hugo Lamot** and **Han Eijkelkamp** for all support. **Sara Phil,** for helping out with the scoring.

Martin Wiedemann and **Miriam Weber**, priceless to get support and access to the data collected!

And of course, thanks to all **farmers** who participated and helped out during and after the data collection. It could have not been possible without you!

To the people at **DeLaval** that just **makes a PhDs life easier**, thank you! **Thomas**, tack för att jag fått ringa i tid och otid, att du lyssnat, stöttat, och plockat fram data ur möjliga och omöjliga filer! **Conny**, tack för att exakt allt går att lösa och ingenting är någonsin ett problem! **Staffan,** du är en helt ovärderlig! Fixar allt man ber om och lite till. **Jöran**, **Anders**, **Hasse**, **Foppa** och **Desirée** tack för att ni svarat på alla mina frågor och för hjälpen med stort som smått!

Till hela **MEV-** gänget, tack för att jag får ta del av er gemenskap. Jag ska såklart bjuda på tårta! **Gunvor, Karin** och **Sofia**, tack för att för att jag fått koppla bort jobbet ibland och prata om det viktigaste i livet, hästar och hundar!

Bohao Liao, thank you for your endless support and interest in my project, **Ilka Klaas** for valuable input whenever needed and **Per Hasselqvist,** for helping with the administrative stuff just trusting me doing my job! Last but not least, **Jonas Hellman**, thank you for making this project possible!

You guys at **SLU**, **Isabel**, **Gabriella**, **Haseeb, Ann**, **Agneta, Marie** and **Jonas**, thank you for letting me be that person dropping in and out, and always feel very welcome and as a part of the group! **Karin A**, thanks for all dog walks, lunches and nice fikas! **Karin S**, I will never forget Scotland!! **Ylva** for all the help with the PhD education. **Nils**, for making me feel welcome at KV from the first day.

And off course, the **PhD students**, well, most of you actually former PhD students nowadays, with special mentioning of a few. **Ida**, **Karolina**, **Daniel**, **Emeli**, **Gunilla**, **Dinah**, **Niclas**, **Torun**, **Josef** and **Sara**. I am so glad that I met you all, but sad that I did not meet you as much as I had wished this last year. Hopefully we can celebrate altogether soon!

John, thanks for all the chats, good ideas and support! It has been nice to have a fellow PhD pal at DeLaval.

Lisa, you have been such a good friend from day one! I am so thankfully that you helped out with the scoring and the system as such, a person like you was really needed to get it right.

Lena-Marie, thank you for always being such a true friend! Also for always being one step in front, so I just can ask you anything, at any time, and you will always have the answer.

Additionally, I just don't know how to thank the both of you enough for the support during the last weeks!

Mamma och **pappa,** ni är mina förebilder i livet! Längtar tills vi ska sitta vid middagsbordet tillsammans och kalla varandra för "doktor Anglart", och ingen vet vem om pratar med vem ☺

Nisse, tack för att du håller balansen i båten när jag bestämmer mig för att spänna mina segel. Utan dig hade jag kapsejsat för länge sen!

Till mina hundar, vad vore livet utan er? **Sune** och **Bert**, för att ni snällt hängt med land och rike runt. **Kott**, för att du håller mig i topptrim med din oändliga energi och dina påhitt! Och **Lisen**, för att du är precis som du är ♥

Acta Universitatis Agriculturae Sueciae

Doctoral Thesis No. 2021:5

This thesis investigates two important indicators of mastitis and milk quality in automatic milking systems. Knowledge of the relationships between data generated by automatic milking systems and somatic cells was gained, and the dynamics of changes in milk homogeneity were outlined. The results indicate that somatic cell counts could be predicted with low prediction error, while the prediction of milk changed in homogeneity was unsatisfactory. These results can provide a basis for future applications used on farms.

Dorota Anglart received her postgraduate education at the Department of Clinical Sciences, Swedish University of Agricultural Sciences, Uppsala. She obtained her Master of Science degree in Agriculture (Animal Sciences) in 2010 at the same university.

Acta Universitatis Agriculturae Sueciae presents doctoral theses from the Swedish University of Agricultural Sciences (SLU).

SLU generates knowledge for the sustainable use of biological natural resources. Research, education, extension, as well as environmental monitoring and assessment are used to achieve this goal.

Online publication of thesis summary: http://pub.epsilon.slu.se/

ISSN 1652-6880
ISBN (print version) 978-91-7760-688-8
ISBN (electronic version) 978-91-7760-689-5