

Article

Not peer-reviewed version

Machine Learning Approach for Early Lactation Mastitis Diagnosis by Applying Total and Differential Somatic Cell Counts

[Alfonso Zecconi](#)*, [Francesca Zaghen](#), [Gabriele Meroni](#), [Flavio Sommariva](#), Silvio Ferrari, [Valerio Sora](#)

Posted Date: 26 March 2025

doi: 10.20944/preprints202503.1970.v1

Keywords: Mastitis; antimicrobial resistance; early diagnosis; differential cell count; machine learning



Preprints.org is a free multidisciplinary platform providing preprint service that is dedicated to making early versions of research outputs permanently available and citable. Preprints posted at Preprints.org appear in Web of Science, Crossref, Google Scholar, Scilit, Europe PMC.

Copyright: This open access article is published under a Creative Commons CC BY 4.0 license, which permit the free download, distribution, and reuse, provided that the author and preprint are cited in any reuse.

Article

Machine Learning Approach for Early Lactation Mastitis Diagnosis by Applying Total and Differential Somatic Cell Counts

Alfonso Zecconi ^{1,*} Francesca Zaghen ^{1,2}, Gabriele Meroni ¹, Flavio Sommariva ³, Silvio Ferrari ³ and Valerio Sora ^{1,3}

¹ Department of Biomedical, Surgical and Dental Sciences, School of Medicine, University of Milano, Via Pascal 36, 20133 Milano, Italy; alfonso.zecconi@unimi.it (AZ); francesca.zaghen@unimi.it (FZ); gabriele.meroni@unimi.it (GM); valerio.sora@unimi.it (VS).

² Department of Clinical and Community Sciences, School of Medicine, University of Milan, Via Celoria 22, 20133 Milan, Italy (francesca.zaghen@unimi.it).

³ Associazione Regionale Allevatori Lombardia, Via Kennedy 30, 26013 Crema, Italy; f.sommariva@aral-lom.it (FS) silviof58@gmail.com (SF).

* Correspondence: alfonso.zecconi@unimi.it

Simple Summary: The development of new technologies in many areas, including milk production, has significantly increased the amount of data available and the need to extract useful information from these data to improve production efficiency. This paper reports the results of applying machine learning to detect cow at risk of intramammary infections caused by major pathogens based on total and differential cell counts. The results confirm that these markers are useful in identifying infected cows and that the ML approach efficiently compares the accuracy of the diagnosis.

Abstract: Dairy herds around the world are undergoing several changes. Herd sizes are increasing, as are both milk yield and quality. The implementation of new technologies in various domains of dairy production is leading to an increase in the amount of data available. This, in turn, creates a need to extract useful information from these data to improve production efficiency. This paper presents the findings of a preliminary study that utilizes a machine learning (ML) approach to assess the accuracy of somatic cell count (SCC) and neutrophils + lymphocytes Count/mL (NLCC) in identifying cows at risk of developing intramammary infection (IMI) due to major pathogens (*S. aureus*, *S. agalactiae*, *S. uberis*, and *S. dysgalactiae*). The study identified these pathogens either by real-time PCR (qPCR) methods or by conventional bacteriology, following the cows' calving process. The study encompassed a total of 424 cows and 1,696 quarter milk samples. A comparison of the two methods revealed significant disparities in the prevalence of MajP, with the qPCR method demonstrating a higher prevalence than conventional bacteriology. However, the prevalence of negative results was comparable, with both methods yielding approximately 71.0% and 72.1%, respectively. The comprehensive results of the study substantiated that all the cellular markers exhibited the most precise values when MajP IMI was diagnosed using quarter milk samples. The cellular markers exhibited nearly equivalent performance, irrespective of the ML algorithm employed. The findings indicate that approaches based on SCC or PLCC may be useful for identifying healthy cows or quarters. However, it is essential to confirm all "non-negative" results through subsequent analysis within 7-15 days to ensure accuracy. However, further studies are necessary to enhance the diagnostic accuracy.

Keywords: Mastitis; antimicrobial resistance; early diagnosis; differential cell count; machine learning

1. Introduction

Dairy herds worldwide are currently experiencing a series of changes that present both challenges and opportunities. Herd sizes are increasing, as are milk yield and quality [1]. New technologies have been implemented in several areas of dairy production [2,3]. Concurrently new regulations have been implemented related to the control of antimicrobial resistance and the improvement of welfare [4].

The proliferation of novel technologies across diverse domains, including milk production, has led to a substantial augmentation in the availability of data. This, in turn, increased the need to extract useful information from these to improve production efficiency. In the context of dairy production, insights derived from data analysis have the potential to enhance the efficacy of early diagnosis, particularly for mastitis, thereby optimizing treatment protocols and reducing antimicrobial usage, ultimately enhancing cow welfare.

This analytical process was described as “knowledge discovery” defined as the ‘non-trivial extraction of implicit, previously unknown and potentially useful information from data’ [5]. One of the major components of the knowledge discovery process is data mining which may be defined as ‘finding the existing patterns in data, which is the base for the further analyses and statistical processes, including machine learning (ML)’ [6]. These techniques can analyze large amounts of data to predict different outcomes, including diseases, and dairy production has been applied to ketosis, lameness and heat stress [7,8]. Recently ML approach was also applied to mastitis diagnosis in cow, buffalo and sheep [9–15].

The availability of differential cell count (DSCC) on milk samples may be considered an additional source of information that could augment our capacity to diagnose subclinical mastitis [16,17]. Furthermore, DSCC data may be incorporated into the machine learning (ML) approach.

Within a project (FEASR 16.1 project MOOH) encompassing the development and application of contagious mastitis pathogens control and mastitis early diagnosis as a tool to reduce and rationalize antimicrobial usage in dairy herds, this paper presents the results of a preliminary study on the application of machine learning in detecting cows at risk to have intramammary infections (IMI) caused by major pathogens (MajP), *S. aureus*, *S. agalactiae*, *S. uberis*, and *S. dysgalactiae* after calving.

2. Materials and Methods

2.1. Herd and Cow Selection

This study considered 424 cows that calved from June 2023 to December 2023 from 12 dairy herds in the Lombardy Region and enrolled in the Italian Breeder Association (AIA) monthly individual milk test (DHI). The herd size ranged from 90 to 500 lactating cows and 95% of the cows were Italian Friesian.

2.2. Sample Collection

Individual cow samplings (MTR) were performed by certified methods currently applied by AIA at the laboratories of the Regional Breeders Association of Lombardy (ARAL) using Lactocorder™ (WMB AG, Balgach, CH), and delivered refrigerated to ARAL labs the same day, and analyzed within 30 h from sampling.

At the lab, an aliquot of 1 ml was taken from each sample and stored in a sterile tube to be analyzed to identify major pathogens by qPCR as described below.

Quarter milk samples (QMS) were taken within 1-3 days after previous individual sampling following the procedure described by N.M.C., 2017 [18], and delivered refrigerated to ARAL laboratories

2.3. Cellular Marker Analyses

Milk analyses on MTR samples included somatic cell count (SCC) and DSCC and were carried out on Fossomatic™ 7DC (Foss A/S, Hillerød, DK). The DSCC was assessed by the method described by Damm et al. [19]. This method allows to identify within a milk sample the macrophages (MAC) and the combination of polymorphonuclear leukocytes (PMN) and lymphocytes (LYM). DSCC is expressed as the combined proportion (%) of PMN and LYM in the overall count of milk cells.

2.4. Conventional Microbiological Analysis

Bacteriological analyses were performed by streaking 10 µl of QMS on Tryptic Soy Agar + 5% v/v defibrinated sheep blood, according to N.M.C., 2017 [18].

After incubation (18-24h at 37°C) the colonies recovered were identified by Vitek™ system (Biomérieux, Lion, F).

Based on the results of the bacteriological analysis, a quarter was classified similarly to a previous paper [20]. Briefly a quarter sample was classified as positive for an IMI due to major pathogens when 1 or more colonies of *S. agalactiae* or *S. aureus* were isolated, or when 5 or more colonies of *S. uberis* and *S. dysgalactiae* were isolated. It was considered positive for an IMI due to other bacteria when 5 or more colonies of the same species of Gram-negative pathogens (*E. coli*, *Klebsiella* spp., other coliforms) were isolated, or when 10 or more colonies of the same genus (coagulase negative *Staphylococcus* species, other environmental *Streptococcus* species, *Enterococcus* species) were isolated. The cases positive for bacteria other than major pathogens were classified as other bacteria in the statistical analysis.

2.5. Real-Time PCR Analysis

A commercial diagnostic kit was used (Mastitis 4A kit; DNA Diagnostic A/S), following producer's instruction. This kit allows bacterial DNA extraction, identification and quantification of *S. aureus*, *S. agalactiae*, *S. uberis* and *S. dysgalactiae* using qPCR. The reaction conditions of qPCR were as follows: 95°C for 1 min, 40 amplification cycles at 95°C for 5 s and 60°C for 25 s. Cycle threshold (Ct) values were considered positive when the value were ≤ 37 , as suggested by the manufacturer. The qPCR reactions were performed on Stratagene Mx3005P (Agilent Technologies Inc., Santa Clara, CA).

2.6. Statistical Analysis

Data were collected in a database including: herdID, cowID, days in milk (DIM), results of the bacteriological analysis by quarter, SCC and DSCC, and a variable (PLCC) calculated by multiplying $SCC \times DSCC$. This variable represents the total number of PMN+LYM/ml.

The model supplied to the ML software algorithms considered as response variables: MajP identified by qPCR, MajP identified by conventional bacteriology or all positive samples identified by conventional bacteriology, while explanatory variables were SCC or DSCC or PLCC, DIM (three classes A 5-15 DIM, B 6-45 DIM and C 45-90 DIM) and parturition (2 classes: primiparous and pluriparous cows).

The machine learning approach was based on Orange software [21], a Python-based tool for data mining and a machine learning suite. It consists of a set of widgets for data preprocessing, features compute modeling, model comparison, and exploration methods. Components are called widgets, ranging from simple data visualization, subset selection and preprocessing, to empirical evaluation of learning algorithms and predictive modeling [22]. Orange Data Mining has positive reviews when analyzing big data, is easy to work with, and is an open-source software tool [21,23].

Orange Data Mining provides a vast range of data mining algorithms (DMA). Naïve Bayes, Decision Tree, Random Forest, and Logistic Regression were used in this study [23,24] after a preliminary analysis led to exclude the algorithms with poorer performances. Figure 1 describes the workflow of ML process. The analysis was performed applying "leave-one-out" method which

randomly splits the data into the training and testing, holding out one instance at a time, inducing the model from all others and then classifying the instances. This method was selected because it is very stable and reliable.

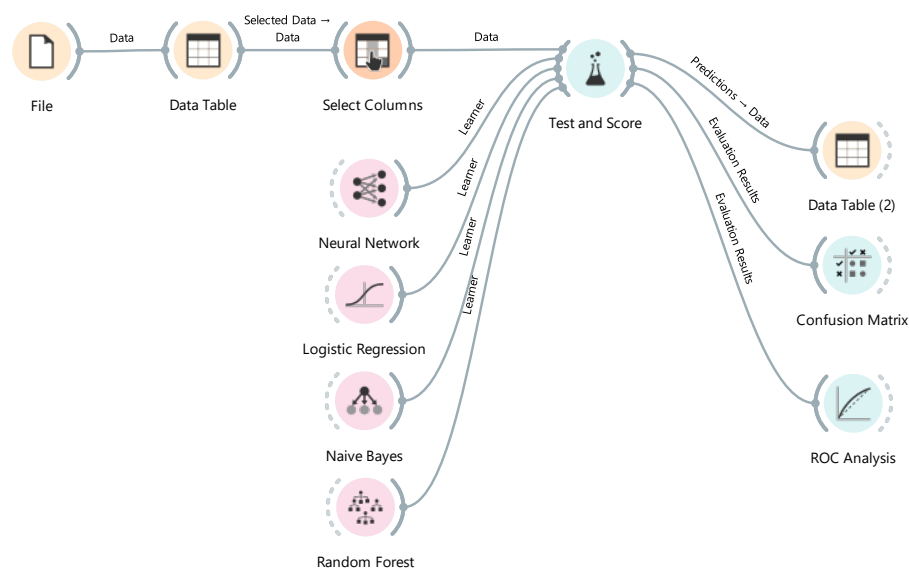


Figure 1. diagrams of the workflow of Orange software applied to the milk samples data.

2.7. Diagnostic Parameters

For each DMA the following parameters were calculated:

- Area under the curve (AUC) of ROC curve: it represents the degree or measure of separability, the higher the AUC the better the model is at predicting the true status of the sample (positive / negative).
- Accuracy: expressed as a proportion of correctly classified subjects [true positive (TP) + true negative (TN)] among all subjects.
- Sensitivity (Se): the proportion of TP / [TP + false positive (FP)].
- Specificity (Sp): the proportion of TN / [false negative (FN) + TP].
- Positive predictive value (PPV): TP / (TP+FN).
- Negative predictive value (NPV): TN / (TN+FP).

3. Results

3.1. Data Description

The dataset containing all the information on cows, cell counts, and microbiological analysis (conventional and qPCR) was checked to identify missing data, such as the absence of SCC or DSCC, and for bacteriologically contaminated samples. All the records with these latter features were discarded and the final database included 424 valid records for cows and 1696 for quarter milk samples. The cellular markers data were summarized in Table 1.

The SCC, DSCC and PLCC mean values were very close among the three DIM periods, and the lower values were observed in period B. Primiparous cows, as expected, showed lower means for cellular markers when compared to pluriparous ones, even if the values between these two groups were relatively close.

Table 1. Description of sample characteristics: cellular markers mean values \pm standard deviation of cow classified by days in milk at sampling and by number of parturitions.

	N	SCC ¹ \pm Std.dev (Log ₁₀ /ml)	DSCC ² \pm Std.dev (%)	PLCC ³ \pm Std.dev (Log ₁₀ /ml)
Lactation period				
A (5-15 d)	88	4.98 \pm 0.22 ^{a,4}	63.1 \pm 17.7 ^a	4.76 \pm 0.90 ^a
B (16-45 d)	111	4.78 \pm 0.83 ^a	61.9 \pm 18.1 ^a	4.55 \pm 0.93 ^a
C (46-90 d)	225	5.00 \pm 0.82 ^b	64.4 \pm 18.3 ^a	4.79 \pm 0.92 ^b
Parturition				
Primiparous	133	4.81 \pm 0.66 ^a	63.8 \pm 19.5 ^a	4.60 \pm 0.73 ^a
Pluriparous	291	5.01 \pm 0.92 ^b	64.0 \pm 15.2 ^a	4.80 \pm 1.02 ^b

¹ SCC = somatic cell count. ² DSCC = differential cell count. ³ PLCC = SCC \times DSCC. ⁴ rows with different superscripts are statistically different ($\alpha=0.05$).

Table 2 reports the distribution of qPCR-positive results for *S. agalactiae*, *S. aureus*, *S. uberis* and *S. dysgalactiae* (MajP), classified in the three lactational periods. The frequency of negative samples was higher in period B, supporting the lower cellular mean values observed, and in the same period we had the lowest frequency for *S. agalactiae*, *S. uberis* and *S. dysgalactiae*. Whereas *S. aureus* frequency showed a marked increase as DIM increased. The *S. uberis* frequency was higher than the other pathogens, particularly in the period up to 45 DIM.

The comparison of cellular markers (Figure 2) based on the udder health status confirmed that MajP increased the mean values by nearly 1 log compared to negative samples.

Table 2. Udder health status: distribution of intramammary infections based on qPCR analysis targeting *S. agalactiae*, *S. aureus*, *S. uberis* and *S. dysgalactiae*, classified by days in milk.

Lactation period	<i>S.aureus</i>	<i>S.agalactiae</i>	<i>S.uberis</i>	<i>S.dysgalactiae</i>	Negative
A (5-15 d)	2.3	2.3	17.0	4.5	73.9
B (16-45 d)	5.4	0.0	8.9	0.9	84.8
C (46-90 d)	14.5	0.0	19.3	3.1	63.1
Total	9.6	0.5	16.1	2.8	71.0

When data were classified by the number of parturitions (Table 3), the results showed that nearly 80% of samples from primiparous were negative, while only 67% of pluriparous cows were free from MajP infections. In this latter group, *S. uberis* and *S. aureus* IMI frequencies nearly doubled compared to primiparous cows. The few *S. agalactiae* IMI observed were all related to pluriparous cows.

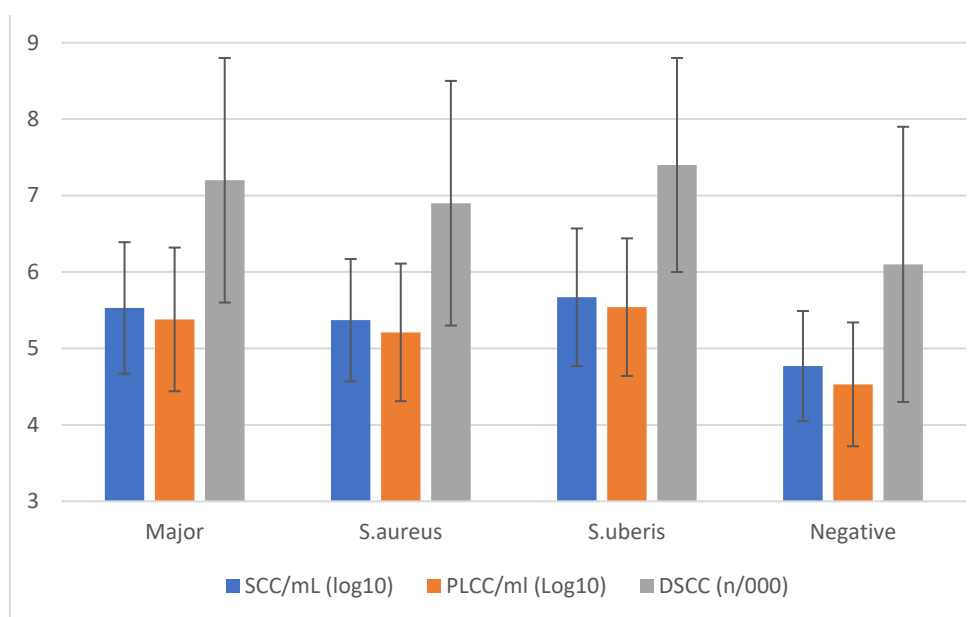


Figure 2. Cellular markers' means (\pm std.dev.) classified by the results of qPCR analysis on individual milk samples.

Table 3. Udder health status: distribution of intramammary infections based on qPCR analysis targeting *S. agalactiae*, *S. aureus*, *S. uberis* and *S. dysgalactiae*, classified by number of parturitions.

Parturition	<i>S.aureus</i>	<i>S.agalactiae</i>	<i>S.uberis</i>	<i>S.dysgalactiae</i>	Negative
Primiparous	6.8	0.0	10.5	3.0	79.7
Pluriparous	11.0	0.7	18.6	2.7	67.0
Total	9.6	0.5	16.1	2.8	71.0

When quarter milk samples were considered (Table 4), the results showed nearly identical overall frequencies of negative samples compared to qPCR results, even if there were differences in the different DIM classes. However, the frequency of MajP was lower than in individual samples.

Table 4. Udder health status: distribution of positive samples from conventional microbiological analysis of quarter milk samples, classified by days in milk.

Lactation period	Quarter (N)	MajP ¹	Other pathogens	Negative
A (5-15 d)	352	1.9%	18.8%	79.3%
B (16-45 d)	444	2.9%	22.3%	74.8%
C (46-90 d)	900	6.5%	25.6%	67.9%
Total	1696	4.6%	23.3%	72.1%

¹ MajP: major pathogens *S.agalactiae*, *S.uberis*, *S.dysgalactiae*, and *S.aureus*.

S. aureus was overall the most frequent pathogen recovered among the MajP with a prevalence of 2.6%, but this value was mainly due to the frequency observed in period C (4.2%). At the same time, *S. uberis* showed an overall prevalence of 2.0% and frequency of 1.7%, 2.2% and 2.1%, respectively in period A, B and C. These two pathogens represented >90% of the MajP isolated.

The results of SCC, DSCC and PLCC (Figure 3) based on the outcome of the microbiological analysis confirm that the presence of MajP is associated with an inflammatory status with values higher than in quarter positives for minor pathogens or bacteriologically negatives. The mean values for the three different cellular markers were very similar in these latter two cases.

When data were classified by the number of parturitions (Table 4), the frequency of IMI in primiparous cows was lower than in pluriparous cows, but the difference between the two groups was smaller than those observed for individual milk samples. To be noticed that the increase in overall IMI frequency was mainly due to the increase in MajP frequency.

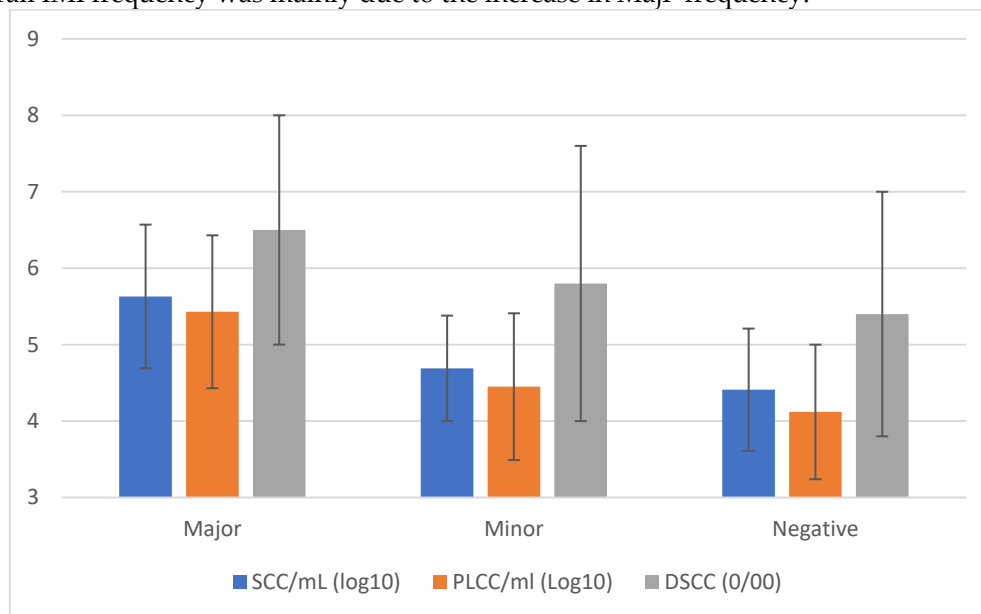


Figure 3. Cellular markers' means (\pm std.dev.) classified by the results of quarter milk samples analysis by conventional microbiological methods.

Table 4. Udder health status: distribution of positive samples from conventional microbiological analysis of quarter milk samples, classified by number of parturitions.

Parturition	Quarter (N)	MajP ¹	Other pathogens	Negative
Primiparous	648	2.6%	22.8%	74.6%
Pluriparous	1048	5.9%	23.6%	70.5%
Total	1696	4.6%	23.3%	72.1%

¹ MajP: major pathogens *S. agalactiae*, *S. uberis*, *S. dysgalactiae*, and *S. aureus*.

3.2. Machine Learning Analysis

The data were analyzed by data-mining algorithms, considering as response variables the positive results at qPCR analysis of individual milk samples, the positive results for MajP at the conventional bacteriological analysis of QMS or the positive results at the conventional bacteriological analysis of QMS.

Table 5 reports the results of the qPCR analysis, and an accuracy in the range 0.690-0.774 was observed among the different algorithms and cellular markers. Logistic regression showed the highest values among algorithms and SCC and PLCC among the cellular markers. Overall sensitivity was small, with values of 57.6% only for PLCC and DSCC when logistic regression was applied. The specificity was higher with values of about 79% for logistic regression and SCC and PLCC, while the highest values were observed for Random Forest with values around 81%. Subsequently, NPV was >95%, while PPV was <20% for most algorithms and cellular markers.

Table 5. Results of the four algorithms analysis in identifying major pathogens by qPCR in individual milk samples.

Model	Parameter	AUC ⁴	Accuracy	Sensitivity	PPV ⁵	Specificity	NPV ⁶
Logistic regression	PLCC ¹	0.740	0.774	57.6%	17.6%	78.9%	96.0%
	SCC ²	0.743	0.774	57.6%	17.6%	78.9%	96.0%
	DSCC ³	0.665	0.763	0.0%	0.0%	100%	76.3%
Neural network	PLCC	0.733	0.760	42.3%	20.4%	78.7%	91.4%
	SCC	0.739	0.765	51.2%	19.4%	79.0%	93.7%
	DSCC	0.651	0.760	33.3%	0.9%	76.3%	99.4%
Naïve Bayes	PLCC	0.711	0.758	48.1%	24.1%	79.6%	91.9%
	SCC	0.717	0.758	48.1%	24.1%	79.6%	91.9%
	DSCC	0.662	0.763	0.0%	0.0%	76.3%	100.0%
Random forest	PLCC	0.684	0.745	45.6%	38.0%	81.6%	85.9%
	SCC	0.656	0.727	40.9%	33.3%	80.4%	85.0%
	DSCC	0.630	0.690	30.6%	24.1%	77.8%	83.0%

¹ PLCC = SCC×DSCC/ml; ² SCC = somatic cell count/ml; ³ DSCC = differential cell count/ml; ⁴ AUC = area under the curve of the ROC curve; ⁵ PPV = positive predictive value; ⁶NPV = negative predictive value.

When conventional bacteriological analysis on quarter milk samples was considered (Table 6) in identifying MajP, the results differed from the analogous analysis based on qPCR. Indeed, the accuracy was always close to 95%, independently from the algorithm and the cellular markers. However, this result is only due to a very high specificity, while sensibility was very poor.

When a positive bacteriological result was considered instead of only MajP (Table 7) the results showed an overall accuracy in the range 0.657-0.733, with the higher values observed for logistic regression and neural network algorithms and PLCC and SCC.

Logistic regression on PLCC and SCC as cellular markers showed the highest sensitivity values, respectively 61.4% and 64.3%, while specificity was around 74% for both markers. These values led to a low PPV, but a NPV >95% for both PLCC and SCC.

Finally, DSCC showed the poorest performance for all four algorithms and the three outcomes considered.

Table 6. Results of the four algorithms analysis in identifying major pathogens by conventional microbiology on quarter milk samples.

Model	Parameter	AUC ⁴	Accuracy	Sensitivity	PPV ⁵	Specificity	NPV ⁶
Logistic regression	PLCC ¹	0.816	0.952	0.0%	0.0%	95.3%	99.9%
	SCC ²	0.821	0.952	0.0%	0.0%	95.3%	99.8%
	DSCC ³	0.686	0.953	n.a. ⁷	0.0%	95.3%	100.0%
Neural network	PLCC	0.806	0.952	0.0%	0.0%	95.3%	99.9%
	SCC	0.811	0.951	16.7%	1.2%	95.4%	99.7%
	DSCC	0.658	0.953	n.a.	0.0%	95.3%	100.0%
Naïve Bayes	PLCC	0.770	0.953	n.a.	0.0%	95.3%	100.0%
	SCC	0.780	0.953	n.a.	0.0%	95.3%	100.0%
	DSCC	0.639	0.953	n.a.	0.0%	95.3%	100.0%
Random forest	PLCC	0.723	0.933	21.5%	16.5%	96.0%	97.1%
	SCC	0.711	0.943	32.7%	20.0%	96.2%	98.0%
	DSCC	0.602	0.951	0.0%	0.0%	95.3%	99.7%

¹PLCC = SCC×DSCC/ml; ² SCC = somatic cell count/ml; ³ DSCC = differential cell count/ml; ⁴AUC = area under the curve of the ROC curve; ⁵ PPV = positive predictive value; ⁶ NPV = negative predictive value. ⁷n.a. = not available.

Table 7. Results of the four algorithms analysis in identifying all pathogens by conventional microbiology on quarter milk samples.

Model	Parameter	AUC ⁴	Accuracy	Sensitivity	PPV ⁵	Specificity	NPV ⁶
Logistic regression	PLCC ¹	0.638	0.728	61.4%	16.9%	73.8%	95.7%
	SCC ²	0.637	0.732	64.3%	17.0%	73.9%	96.1%
	DSCC ³	0.595	0.710	n.a. ⁶	0.0%	71.0%	100.0%
Neural network	PLCC	0.657	0.733	60.5%	22.9%	74.9%	93.9%
	SCC	0.653	0.733	60.7%	22.5%	74.8%	94.0%
	DSCC	0.621	0.716	55.8%	10.0%	72.5%	96.7%
Naïve Bayes	PLCC	0.618	0.713	51.6%	18.6%	73.6%	92.9%
	SCC	0.616	0.710	n.a.	0.0%	71.0%	100.0%
	DSCC	0.587	0.710	n.a.	0.0%	71.0%	100.0%
Random forest	PLCC	0.586	0.657	39.6%	34.8%	74.6%	78.3%
	SCC	0.595	0.683	43.1%	29.2%	74.4%	84.3%
	DSCC	0.543	0.660	35.1%	20.1%	72.2%	84.8%

¹PLCC = SCC×DSCC/ml; ² SCC = somatic cell count/ml; ³ DSCC = differential cell count/ml; ⁴AUC = area under the curve of the ROC curve; ⁵ PPV = positive predictive value; ⁶ NPV = negative predictive value. ⁷n.a. = not available.

4. Discussion

4.1. Intramammary Infections

The early assessment of udder health after calving has become increasingly important as selective dry-cow therapy is widely applied in many countries. Indeed, it facilitates the identification of infected animals, thereby enabling the judicious and rational administration of antimicrobial therapy, which in turn reduces the risk of antimicrobial resistance (AMR) [28]. This approach is relatively labor intensive (milk sampling), may be expensive (microbiological analysis). Additionally, its implementation in dairy herds with >200 cows can pose significant challenges.

The availability of new technologies such as DSCC and qPCR on the one hand, and new approaches to data analysis such as ML on the other hand, paved the way for the development of new protocols to identify cows at risk. A previous study supported this approach by showing that early diagnosis based on DSCC is cost effective [20].

The present study aimed to improve this approach by evaluating whether qPCR and ML analytical approaches can increase the accuracy and efficiency of early mastitis diagnosis.

The analysis of the results of mastitis diagnosis on MTR samples and QMS allowed us to identify some critical points. Indeed, the analysis of the cellular and microbiological data showed that the different time windows influenced the result. The mean values observed for both SCC and PLCC markers in the first two periods considered (5-15, 16-45 DIM) were significantly different for the third period (46-90 DIM). In addition, the values in period B (16-45 DIM) were the lowest. Similarly, IMI prevalence was lowest in the first two DIM intervals for both MTR and QMS.

The results of the comparison between MTR samples (qPCR) and QMS (conventional bacteriology) showed, as expected, large differences in the prevalence of MajP as a whole and when considering the individual species. In fact, the prevalence of MajP in MTR was 16.1% compared to 4.6% in QMS. Despite these differences, the prevalence of negative MTR or QMS was very close at 71.0% and 72.1%, respectively. The differences in MajP pathogens may be explained, at least in part,

by the different diagnostic methods, and it can be argued that qPCR identifies DNA and not live bacteria, and that some of the positive results may be due to environmental contamination (in the case of *S. uberis* and *S. dysgalactiae*). However, this argument does not apply to infectious bacteria. The qPCR was advantageous to identify the presence of IMI when the bacterial concentrations were below the detection limits of conventional bacteriology or the samples were contaminated [29,30]. Therefore, the analysis of the association of cellular markers with infection status may be useful to identify the presence of an evolving infection.

4.1. Machine Learning Approach

The availability of automated instruments to measure DSCC in milk samples has greatly improved our ability to diagnose mastitis [20,31–33]. When applied to MTR samples, this analysis generates a large amount of data that may be analyzed using the ML approach. In fact, ML has been used to diagnose clinical and subclinical mastitis [9,12,34], but DSCC were not included in these studies.

The different algorithms applied to the same set of data showed, as expected, some differences. Indeed, overall logistic regression and neural network performed better than naïve Bayes and random forest. Different outcomes related to the algorithm applied were expected and already observed [12], and they may also be explained by the characteristics of the database considered. The capability of Orange tool and other ML software to compare different algorithms is pivotal to identify the best algorithm and model to be applied on a routine basis, when the ML approach is used to diagnose diseases.

Our study included a relatively small number of data, but the sample size is comparable or greater than other studies on the diagnostic of dairy cows [7,35] or human diseases [36,37].

The overall results confirmed that all cellular markers showed the highest accuracy values when MajP IMI was diagnosed using quarter milk samples. In fact, regardless of the algorithm used, the accuracy was always >90%. This result was due to the very high specificity observed (>95%), while the sensitivity was very low and absent in most cases. This result was expected due to the low frequency of MajP observed in QMS. When MTR and qPCR were considered, the accuracy ranged from 63-74%, depending on the algorithm, with sensitivity in the range of 30-57% and specificity in the range of 78-82%. In both cases, MTR and QMS, the NPV was very high, reaching 100% in some cases. The observed sensitivity and specificity values were similar to the results of other studies [9,13], but in contrast to those observed by Ebrahimi *et al.* [38].

DSCC had the lowest performance among the cellular markers, while SCC and PLCC had nearly equal performance, regardless of the algorithm used. This result has useful implications in practice. Indeed, DSCC and PLCC are not available without the specific counting instrument (Fossomatic™ 7DC). However, SCCs are always available when a DHI program is applied, as is the case in most countries with large milk production. Therefore, in the latter case, the identification of cows at risk for MajP IMI is possible, while when PLCC can be calculated, it is an additional information to be applied at cow level to better define the cow's udder health status [39]. Overall, these results suggest that approaches based on SCC or PLCC may help to identify healthy cows or quarters, while all "non-negative" should be confirmed by a subsequent analysis in the following 7-15 days.

5. Conclusions

Early and accurate diagnosis of disease is important in any field of medicine. In the case of dairy cow IMI, such diagnosis has several positive implications: higher cure rate, rational and reduced use of antimicrobials, improved cow welfare, higher milk yield and quality. Early diagnosis is particularly important after calving, when cows are exposed to pathogens due to their compromised immune status as a result of calving. Changes in dairy herd management include new approaches and technologies that could help identify cows at risk for MajP IMI. The results of this study under field conditions confirm that cellular markers related to total and differential cell counts may be

useful in identifying these cows and that the ML approach is efficient in comparing the accuracy of diagnosis. However, further studies are needed to increase the diagnostic accuracy.

Author Contributions: Conceptualization, A.Z.; methodology, G.M., F.Z., F.S., V.S.; validation A.Z., F.Z.; formal analysis, F.Z., G.M., S.F., V.S.; statistical analysis, A.Z., V.S.; investigation, A.Z., F.Z., G.M., S.F., V.S.; data curation, F.Z., G.M., F.S., V.S.; writing—original draft preparation, A.Z., F.Z.; writing—review and editing, A.Z., F.Z., G.M., F.S., V.S.; funding acquisition, A.Z., F.S.. All authors have read and agreed to the published version of the manuscript.

Funding: The study was funded by FEASR—Programma di Sviluppo Rurale 2014–2020 Misura 16.1 project MOOH

Institutional Review Board Statement: According to national legislation, the Ethics Committee approval was not required. Furthermore, this study was conducted in accordance with EU Directive 2010/63/EU on the protection of animals used for scientific purposes.

Informed Consent Statement: Not applicable since the dataset was supplied anonymized, and sampling and record storage are regulated by agreement between the farmers and Italian Breeder Association which includes the availability of data for research purposes.

Acknowledgments: The authors are grateful for the technical support offered by the field and laboratory technicians of Associazione Regionale Allevatori della Lombardia.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. FAO. *Dairy Market Review: Overview of global market developments in 2024*; FAO: Rome, 2024.
2. Martin, T.; Gasselin, P.; Hostiou, N.; Feron, G.; Laurens, L.; Purseigle, F.; Ollivier, G. Robots and transformations of work in farm: a systematic review of the literature and a research agenda. *Agron. Sustain. Dev.* **2022**, *42*, 20. <https://doi.org/10.1007/s13593-022-00796-2>.
3. Cogato, A.; Brscic, M.; Guo, H.; Marinello, F.; Pezzuolo, A. Challenges and Tendencies of Automatic Milking Systems (AMS): A 20-Years Systematic Review of Literature and Patents. *Animals* **2021**, *11*, 21. <https://doi.org/10.3390/ani11020356>.
4. Garcia, S.N.; Osburn, B.I.; Cullor, J.S. A one health perspective on dairy production and dairy food safety. *One Health* **2019**, *7*, 9. <https://doi.org/10.1016/j.onehlt.2019.100086>.
5. Bramer, M. *Principles of Data Mining*, IV ed.; Srpinger: London 2020; p. 576.
6. Qamar, U.; Raza, M.S. *Data Science Concepts and Techniques with Applications*; Springer Singapore: 2020; p. 196.
7. Zhou, X.J.; Xu, C.; Wang, H.; Xu, W.; Zhao, Z.X.; Chen, M.X.; Jia, B.; Huang, B.Y. The Early Prediction of Common Disorders in Dairy Cows Monitored by Automatic Systems with Machine Learning Algorithms. *Animals* **2022**, *12*, 14. <https://doi.org/10.3390/ani12101251>.
8. Lasser, J.; Matzhold, C.; Egger-Danner, C.; Fuerst-Waltl, B.; Steininger, F.; Wittek, T.; Klimek, P. Integrating diverse data sources to predict disease risk in dairy cattle—a machine learning approach. *J. Anim. Sci.* **2021**, *99*, 14. <https://doi.org/10.1093/jas/skab294>.
9. Bobbo, T.; Biffani, S.; Taccioli, C.; Penasa, M.; Cassandro, M. Comparison of machine learning methods to predict udder health status based on somatic cell counts in dairy cows. *Sci Rep* **2021**, *11*, 10. <https://doi.org/10.1038/s41598-021-93056-4>.
10. Bobbo, T.; Matera, R.; Pedota, G.; Manunza, A.; Cotticelli, A.; Neglia, G.; Biffani, S. Exploiting machine learning methods with monthly routine milk recording data and climatic information to predict subclinical mastitis in Italian Mediterranean buffaloes. *J. Dairy Sci.* **2023**, *106*, 1942–1952. <https://doi.org/10.3168/jds.2022-22292>.

11. Tian, H.; Zhou, X.J.; Wang, H.; Xu, C.; Zhao, Z.X.; Xu, W.; Deng, Z.J. The Prediction of Clinical Mastitis in Dairy Cows Based on Milk Yield, Rumination Time, and Milk Electrical Conductivity Using Machine Learning Algorithms. *Animals* **2024**, *14*, 14. <https://doi.org/10.3390/ani14030427>.
12. Fadul-Pacheco, L.; Delgado, H.; Cabrera, V.E. Exploring machine learning algorithms for early prediction of clinical mastitis. *Int. Dairy J.* **2021**, *119*, 9. <https://doi.org/10.1016/j.idairyj.2021.105051>.
13. Satola, A.; Satola, K. Performance comparison of machine learning models used for predicting subclinical mastitis in dairy cows: Bagging, boosting, stacking, and super-learner ensembles versus single machine learning models. *J. Dairy Sci.* **2024**, *107*, 3959-3972. <https://doi.org/10.3168/jds.2023-24243>.
14. Mitsunaga, T.M.; Garcia, B.L.N.; Pereira, L.B.R.; Costa, Y.C.B.; da Silva, R.F.; Delbem, A.C.B.; dos Santos, M.V. Current Trends in Artificial Intelligence and Bovine Mastitis Research: A Bibliometric Review Approach. *Animals* **2024**, *14*, 19. <https://doi.org/10.3390/ani14142023>.
15. Kiouvrekis, Y.; Vasileiou, N.G.C.; Katsarou, E.I.; Lianou, D.T.; Michael, C.K.; Zikas, S.; Katsafadou, A.I.; Bourganou, M.V.; Liagka, D.V.; Chatzopoulos, D.C.; et al. The Use of Machine Learning to Predict Prevalence of Subclinical Mastitis in Dairy Sheep Farms. *Animals* **2024**, *14*, 21. <https://doi.org/10.3390/ani14162295>.
16. Zecconi, A.; Vairani, D.; Cipolla, M.; Rizzi, N.; Zanini, L. Assessment of Subclinical Mastitis Diagnostic Accuracy by Differential Cell Count in Individual Cow Milk. *Ital J Anim Sci* **2018**, *18*, 435-440. <https://doi.org/10.1080/1828051X.2018.1533391>.
17. Halasa, T.; Kirkeby, C. Differential Somatic Cell Count: Value for Udder Health Management. *Front. Vet. Sci.* **2020**, *7*, 7. <https://doi.org/10.3389/fvets.2020.609055>.
18. N.M.C. *Laboratory Handbook on Bovine Mastitis*; N.M.C.: New Prague Wi, USA, 2017.
19. Damm, M.; Holm, C.; Blaabjerg, M.; Bro, M.N.; Schwarz, D. Differential somatic cell count-A novel method for routine mastitis screening in the frame of Dairy Herd Improvement testing programs. *J. Dairy Sci.* **2017**, *100*, 4926-4940. <https://doi.org/10.3168/jds.2016-12409>.
20. Zecconi, A.; Meroni, G.; Sora, V.; Mattina, R.; Cipolla, M.; Zanini, L. Total and differential cell counts as a tool to identify intramammary infections in cows after calving. *Animals* **2021**, *11*, 14.
21. Demsar, J.; Curk, T.; Erjavec, A.; Gorup, C.; Hocevar, T.; Milutinovic, M.; Mozina, M.; Polajnar, M.; Toplak, M.; Staric, A.; et al. Orange data mining toolbox in Python. *Journal of Machine Learning Research* **2013**, *14*, 2349-2353.
22. Hosseini, S.; Sardo, S.R. Data mining tools -a case study for network intrusion detection. *Multimedia Tools and Applications* **2021**, *80*, 4999-5019. <https://doi.org/10.1007/s11042-020-09916-0>.
23. Antas, J.; Rocha Silva, R.; Bernardino, J. Assessment of SQL and NoSQL Systems to Store and Mine COVID-19 Data. *Computers* **2022**, *11*, 29.
24. Ozella, L.; Rebuli, K.B.; Forte, C.; Giacobini, M. A Literature Review of Modeling Approaches Applied to Data Collected in Automatic Milking Systems. *Animals* **2023**, *13*, 23. <https://doi.org/10.3390/ani13121916>.
25. Zecconi, A.; Gusmara, C.; Di Giusto, T.; Cipolla, M.; Marconi, P.; Zanini, L. Observational study on application of a selective dry-cow therapy protocol based on individual somatic cell count thresholds. *Ital J Anim Sci* **2020**, *19*, 1341-1348. <https://doi.org/10.1080/1828051X.2020.1842812>.
26. Dziuba, M.; Caixeta, L.S.; Boyum, B.; Godden, S.; Royster, E.; Rowe, S. Negatively controlled trial investigating the effects of dry cow therapy on clinical mastitis and culling in multiparous cows. *J. Dairy Sci.* **2023**, *106*, 5687-5695. <https://doi.org/10.3168/jds.2022-22845>.
27. Rowe, S.M.; Godden, S.M.; Nydam, D.V.; Gorden, P.J.; Lago, A.; Vasquez, A.K.; Royster, E.; Timmerman, J.; Thomas, M. Randomized controlled non-inferiority trial investigating the effect of 2 selective dry-cow

- therapy protocols on antibiotic use at dry -off and dry period intramammary infection dynamics. *J. Dairy Sci.* **2020**, *103*, 6473-6492. <https://doi.org/10.3168/jds.2019-17728>.
28. Afifi, M.; Kabera, F.; Stryhn, H.; Roy, J.P.; Heider, L.C.; Godden, S.; Montelpare, W.; Sanchez, J.; Dufour, S. Antimicrobial-based dry cow therapy approaches for cure and prevention of intramammary infections: a protocol for a systematic review and meta-analysis. *Anim. Health Res. Rev.* **2018**, *19*, 74-78. <https://doi.org/10.1017/s1466252318000051>.
 29. Koskinen, M.T.; Wellenberg, G.J.; Sampimon, O.C.; Holopainen, J.; Rothkamp, A.; Salmikivi, L.; van Haeringen, W.A.; Lam, T.; Pyorala, S. Field comparison of real-time polymerase chain reaction and bacterial culture for identification of bovine mastitis bacteria. *Journal of Dairy Science* **2010**, *93*, 5707-5715. <https://doi.org/10.3168/jds.2010-3167>.
 30. Taponen, S.; Salmikivi, L.; Simojoki, H.; Koskinen, M.T.; Pyorala, S. Real-time polymerase chain reaction-based identification of bacteria in milk samples from bovine clinical mastitis with no growth in conventional culturing. *Journal of Dairy Science* **2009**, *92*, 2610-2617. <https://doi.org/10.3168/jds.2008-1729>.
 31. Schwarz, D.; Lipkens, Z.; Piepers, S.; De Vliegher, S. Investigation of differential somatic cell count as a potential new supplementary indicator to somatic cell count for identification of intramammary infection in dairy cows at the end of the lactation period. *Preventive Veterinary Medicine* **2019**, *172*, 7. <https://doi.org/10.1016/j.prevetmed.2019.104803>.
 32. Kirkeby, C.; Toft, N.; Schwarz, D.; Farre, M.; Nielsen, S.S.; Zervens, L.; Hechinger, S.; Halasa, T. Differential somatic cell count as an additional indicator for intramammary infections in dairy cows. *J. Dairy Sci.* **2020**, *103*, 1759-1775. <https://doi.org/10.3168/jds.2019-16523>.
 33. Zecconi, A.; Zaghen, F.; Meroni, G.; Sora, V.; Martino, P.A.; Laterza, G.; Zanini, L. Early Milk Total and Differential Cell Counts as a Diagnostic Tool to Improve Antimicrobial Therapy Protocols. *Animals* **2023**, *13*, 1143.
 34. Luo, W.K.; Dong, Q.; Feng, Y. Risk prediction model of clinical mastitis in lactating dairy cows based on machine learning algorithms. *Prev. Vet. Med.* **2023**, *221*, 8. <https://doi.org/10.1016/j.prevetmed.2023.106059>.
 35. Goyache, F.; Díez, J.; López, S.; Pajares, G.; Santos, B.; Fernández, I.; Prieto, M. Machine Learning as an aid to management decisions on high somatic cell counts in dairy farms. *Arch. Tierz.-Arch. Anim. Breed.* **2005**, *48*, 138-148. <https://doi.org/10.5194/aab-48-138-2005>.
 36. DE CAMPOS, M.A.; OPPERMANN, M.L.R.; GENRO, V.K.; LEITAO, C.B.; HIRAKATA, V.N.; REICHEL, A.J. 170-LB: Predictors of Diabetes First Diagnosed in Pregnancy: A Machine-Learning Model. *Diabetes* **2020**, *69*. <https://doi.org/10.2337/db20-170-LB>.
 37. Tougui, I.; Jilbab, A.; El Mhamdi, J. Heart disease classification using data mining tools and machine learning techniques. *Health and Technology* **2020**, *10*, 1137-1144. <https://doi.org/10.1007/s12553-020-00438-1>.
 38. Ebrahimi, M.; Mohammadi-Dehcheshmeh, M.; Ebrahimie, E.; Petrovski, K.R. Comprehensive analysis of machine learning models for prediction of sub-clinical mastitis: Deep Learning and Gradient-Boosted Trees outperform other models. *Comput. Biol. Med.* **2019**, *114*, 9. <https://doi.org/10.1016/j.compbimed.2019.103456>.
 39. Zecconi, A.; Zanini, L.; Cipolla, M.; Stefanon, B. Factors Affecting the Patterns of Total Amount and Proportions of Leukocytes in Bovine Milk *Animals* **2020**, *10*, 1-12. <https://doi.org/10.3390/ani10060992>.

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.