

**MATHEMATICAL MODELING AND MACHINE LEARNING PREDICTION FOR
PREVALENCE DYNAMICS OF CLINICAL MASTITIS IN DAIRY HERDS**

**MATEMATIČKO MODELIRANJE I PREDVIĐANJE MAŠINSKOG UČENJA ZA DINAMIKU
PREVALENCIJE KLINIČKOG MASTITISA U MLEČNIM STADAMA**

**Dimitar Nakov^{a*}, Biljana Zlatanovska^b, Mirjana Kocaleva Vitanova^b,
Marija Miteva^b, Slavča Hristov^c, Branislav Stanković^c**

^aFaculty of Agriculture, Goce Delcev University, Stip, North Macedonia,

^bFaculty of Computer Science, Goce Delcev University, Stip, North Macedonia

^cFaculty of Agriculture, University of Belgrade, Belgrade-Zemun, Serbia

*Corresponding author: dimitar.nakov@ugd.edu.mk



D.N. <https://orcid.org/0000-0001-6159-5595>, B.Z. <https://orcid.org/0000-0003-4300-2877>, M.K.V. <https://orcid.org/0000-0002-2444-2917>,
M.M. <https://orcid.org/0000-0001-5326-2301>, S.H. <https://orcid.org/0000-0002-1719-9971>, B.S. <https://orcid.org/0000-0003-3925-6102>

ABSTRACT

Mastitis remains one of the major diseases in dairy herds, causing profound economic losses to the entire milk production chain. The main aim of the study was an application of mathematical models and machine learning algorithms for the prediction of mastitis transmission in the dairy cow population. Data used for mathematical models and machine learning algorithms were obtained in a cross-sectional longitudinal survey lasting for one year by analyzing data for clinical mastitis occurrence in three dairy herds. For data prediction, simple SIR and SIRS mathematical models without vital dynamics and Weka software were applied. The annual prevalence rate of clinical mastitis for the entire population of cows was 34.13% on the cow level, 30.07% on the lactation level, while lactation incident risk was 45.86%. Most of the cows manifested one (68.24%) or two (18.63%) cases of clinical mastitis during lactation. The SIR model revealed that after a short time, the epidemic will disappear. From the explanation and the graphical presentations, it can be concluded that the stable point DFE attracts the trajectories of the system. The mastitis on the farms is calming down, and with these parameters of the model, an epidemic cannot occur. With the use of the decision table as one of the most used classification rules and cross-validation folds 10 we can best predict mastitis occurrence in dairy farms. Implementation of a good mastitis prevention program in dairy herds by increasing the rates of control parameters will reduce the mastitis pathogens transmission rates leading to a reduction of mastitis incidence.

Key words: mastitis, dairy cows, SIR, SIRS, machine learning

SAŽETAK

Mastitis je jedna od najznačajnijih bolesti u mlečnim stadama, koji izaziva velike ekonomske gubitke u celom lancu proizvodnje mleka. Osnovni cilj rada je bila primena matematičkih modela i algoritama mašinskog učenja za predviđanje prenošenja uzročnika mastitisa u populaciji muznih krava. Podaci, korišćeni za matematičke modele i algoritme mašinskog učenja, su dobijeni u longitudinalnom istraživanju poprečnog preseka u trajanju od godinu dana analizom podataka o kliničkoj pojavi mastitisa u tri mlečna stada. Za predviđanje podataka primenjeni su jednostavni SIR i SIRS matematički modeli bez vitalne dinamike i Veka softver. Godišnja stopa prevalencije kliničkog mastitisa za celokupnu populaciju krava iznosila je 34,13% na nivou krava, 30,07% na nivou laktacije, dok je rizik pojave u laktaciji iznosio 45,86%. Većina krava je u toku laktacije ispoljila jedan (68,24%) ili dva (18,63%) slučaja kliničkog mastitisa. SIR model je otkrio da će nakon kratkog vremena epidemija mastitisa nestati. Iz objašnjenja i grafičkih prikaza

može se zaključiti da stabilna tačka DFE privlači putanje sistema. Mastitis na farmama se smiruje, a sa ovim parametrima modela ne može doći do epidemije. Primenom tabele odluka kao jednog od najčešće korišćenih klasifikacionih pravila i preklopa za unakrsnu validaciju 10 može se najbolje predvideti pojava mastitisa na farmama mlečnih krava. Sprovođenje dobrog programa prevencije mastitisa u mlečnim stadima povećanjem stope kontrolnih parametara će smanjiti stope prenošenja patogenih mikroorganizama uzročnika mastitisa, što će dovesti do smanjenja incidencije mastitisa.

Ključne reči: mastitis, mlečne krave, SIR, SIRS, mašinsko učenje

1. INTRODUCTION

Mastitis continues to be a significant concern within dairy herds, leading to considerable economic ramifications throughout the milk production chain. The dairy industry experiences substantial losses primarily attributable to decreased milk production and compromised milk quality (1,2). National statistics from the leading milk-producing European countries indicate that each year, 30 to 40% of dairy cows exhibit clinical signs of mastitis during the lactation period (3,4).

Mastitis is inflammation of the mammary gland that more often strikes cows with high levels of milk production, significantly impacting their milk yield (1). This condition not only affects the immediate productivity of dairy cows but also limits their genetic potential for milk production. The severity of milk yield reduction due to clinical mastitis (CM) varies based on the cow's number of lactations and the timing of the disease's onset during the lactation period (5). The most substantial decreases in milk production are typically seen when CM occurs early in the lactation period (6), with affected cows generally unable to achieve their expected milk yield for the remainder of that lactation period.

However, cases of CM represent just a tip of the total mastitis incidents. In fact, for every instance of clinical mastitis, there are 20 to 40 times more cases of subclinical mastitis (7). These subclinical cases may either evolve into CM or linger unnoticed for an extended duration. The emergence of mastitis within dairy herds is the result of a complex interaction between the cows, their environment, and various pathogens. Risk factors for CM in dairy operations are typically categorised into two main groups: those related to individual cows and those associated with environmental conditions (8,9,10,11).

The prevalence of udder pathogens responsible for both subclinical and clinical mastitis exhibits significant variation (12). Historically, primary

udder pathogens in dairy herds included *Streptococcus agalactiae* and *Staphylococcus aureus*. However, with the adoption of modern milking practices and the implementation of mastitis control programs, the prevalence of these organisms has notably declined across many contemporary dairy farms. Presently, the common environmental pathogens encompass Coagulase Negative *Staphylococci* (CNS), *Streptococcus uberis*, *Streptococcus dysgalactiae*, *Klebsiella spp.*, and *Escherichia coli* (13,14).

Research efforts focused on developing decision-support tools for mastitis detection and management in dairy herds are ongoing. Mathematical modelling serves as a critical tool in comprehending and addressing mastitis within these herds (15). Additionally, understanding the pathways through which mastitis pathogens are transmitted is essential for making proper management decisions.

This paper discusses mathematical models framed within the broader concept of the epidemic structural equation models, focusing on SIR (Susceptible, Infected, Recovered) and SIRS (Susceptible, Infected, Recovered, Susceptible) frameworks to analyze the binary value of mastitis in dairy cows, making distinguish between healthy cows and those afflicted with clinical mastitis. These models intend to outline the causal, whether simultaneous or sequential, relationships among phenotypes, a common occurrence in various biological systems. For instance, in dairy cattle, a correlation exists where high milk production may lead to an increased risk of mastitis, which, in turn, negatively impacts milk yield.

In the realm of mathematical modelling, distinctions are made between SIR and SIRS models that do not incorporate vital dynamics (not considering rates of newly involved heifers and culling dairy cows) and those that do. This paper focuses on the application of SIR and SIRS models that exclude vital dynamics. These models categorize the population into three primary groups: Susceptible (S), Infected

(I), and Recovered (R), aiming to predict the transmission dynamics of clinical mastitis within dairy herds. Through the analysis of these models, the paper also seeks to recommend strategies for a mastitis control program aimed at dairy herds, contributing to both theoretical and practical advancements in managing mastitis.

Furthermore, the paper utilizes machine learning (ML), a subfield of artificial intelligence (AI) dedicated to crafting algorithms and models that empower computers to learn from data autonomously and make predictions or decisions without explicit programming. ML finds application across diverse domains, including epidemiology, image and speech recognition, natural language processing, recommendation systems, healthcare, finance, and beyond. Within epidemiology, ML techniques are applied for disease surveillance, diagnostic and prognostic modelling, genomic epidemiology, risk factor identification, and various other scenarios.

2. MATERIALS AND METHODS

2.1. Epidemiological data

This study was conducted across three dairy farms situated in North Macedonia. Data were collected through a cross-sectional longitudinal survey spanning one year, focusing on the occurrence of clinical mastitis among milking cows. Each milking cow's yearly observation period began at a specific point in time and continued for one year, or from the start of the cow's lactation period within the observed timeframe until the finishing of lactation or the time point when the cow was removed from the herd. Consequently, a total of 1031 cows of the black-white breed, encompassing 1267 lactations, were monitored throughout the study duration.

The incidence of new cases of clinical mastitis was recorded daily throughout the research trial, using standard clinical observation under typical field conditions. Diagnosis of clinical mastitis involves a clinical examination of the udder (assessing for swelling, firmness, colour, pain, and impaired function) and evaluation of milk abnormalities (including the presence of watery milk, flakes, clots, blood, pus, or discolouration). The occurrence of clinical mastitis was quantified as prevalence per cow and lactation.

$$\text{prevalence rate per cow (\%)} = \frac{\text{number of cows with clinical mastitis}}{\text{total number of observed dairy cows}} \times 100$$

$$\text{prevalence rate per lactation (\%)} = \frac{\text{number of lactation with case of clinical mastitis}}{\text{total number of observed lactation}} \times 100$$

Lactation incidence risk (LIR) was calculated using the density method:

$$\text{LIR (\%)} = \frac{\text{total number of clinical mastitis cases}}{\text{total number of observed lactation}} \times 100$$

During the same lactation, a nine-day interval was used to make a distinction between two successive cases of clinical mastitis. This interval comprised four days of antibiotic treatment administered to the affected quarters of the mammary gland, followed by an additional four days during which the antibiotics remained present in the milk. Throughout this period, the milk was waved aside. Finally, on the ninth day, any abnormal changes in the milk were assessed to confirm the presence of mastitis symptoms (16).

Related to the mastitis status, dairy cows were allocated into three classes: the cows at risk of developing a clinical form of mastitis, cows that suffer from clinical mastitis and the class of dairy cows that have recovered from the mastitis and got temporary immunity.

2.2. Mathematical models

The SIR and SIRS frameworks were applied as predictive models for understanding the spread of mastitis within dairy cow populations. These models operate under the assumption of stable population size, denoted as N, thereby excluding vital dynamics such as the introduction of new heifers, culling, or deaths within the dairy cow population. Such models are referred to as closed epidemic models. Due to the infectious nature of mastitis, it is presumed that the entire milking cow population is at risk of infection. Cows exhibiting symptoms of clinical mastitis act as the infection's source, facilitating its spread within the herd. Typically, around nine days after showing signs of clinical mastitis, cows are again at risk of developing the disease.

Thus, the SIR model assumed that cows once recovered from mastitis cannot be reinfected, while the SIRS model posits that recovered cows remain

susceptible and can indeed contract mastitis once more. The graphical visualizations of mathematical models were made in the mathematical program Wolfram Mathematica 7.0. 18.43403. (17).

2.2.1. The SIR model without vital dynamics

The SIR model without vital dynamics (18,19) for mastitis transmission in the dairy cow population is a simple mathematical model. The observed population at risk was divided into three groups: the susceptible or cows under risk (S), the cows with clinical mastitis (I), and the recovered cows (R). Related to the nature of the practical problem that should be solved, the symbols S , I and R , which are used to mark the groups according to the SIR model, will be changed. The susceptible group will be denoted with S_c , the cows that suffer from clinical mastitis will be denoted with CM_c , and the recovered cows will be denoted with H_c . Figure 1 shows the SIR model.

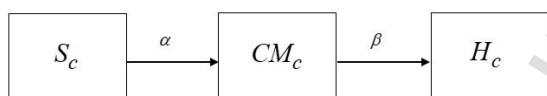


Figure 1. The SIR model without vital dynamics

The fixed size of the total population N is $N = S_c(t) + CM_c(t) + H_c(t)$ for all time t . The functions $S_c(t)$, $CM_c(t)$ and $H_c(t)$ depend on the time t . The rate of infection with which infect the susceptible' cows is denoted by α and the rate of recovery is denoted by β . The time of recovery is $\tau = \frac{1}{\beta}$. This model is described by the following

ordinary differential equations:

$$\frac{dS_c}{dt} = -\frac{\alpha S_c CM_c}{N} \quad (1)$$

$$\frac{dCM_c}{dt} = \frac{\alpha S_c CM_c}{N} - \beta CM_c$$

$$\frac{dH_c}{dt} = \beta CM_c$$

with the initial values

$$S_c(0) = S_{c_0} > 0, CM_c(0) = CM_{c_0} > 0, H_c(0) = H_{c_0} \geq 0$$

By replacement

$$S(t) = \frac{S_c}{N}, CM(t) = \frac{CM_c(t)}{N}, H(t) = \frac{H_c(t)}{N},$$

this system

$$\frac{dS}{dt} = -\alpha S CM \quad (2)$$

$$\frac{dCM}{dt} = \alpha S CM - \beta CM$$

$$\frac{dH}{dt} = \beta CM$$

with the initial values $S(0) = S_0 > 0$, $CM(0) = CM_0 > 0$, $H(0) = H_0 \geq 0$ and $S(t) + CM(t) + H(t) = 1$ for all time t is transformed.

The initial behavior of the system is that we have a small number of infections and no recovered individuals, while most of the population is susceptible. Therefore, we can take $S_{c_0} \approx N$ i.e. $S_0 \approx 1$ then the first and second equations of the system (1) have transformed them to

$$\frac{dS}{dt} \approx -\alpha CM$$

$$\frac{dCM}{dt} \approx \alpha CM - \beta CM = (\alpha - \beta) CM$$

Clear that, if $\alpha > \beta$ i.e. $\frac{\alpha}{\beta} > 1$ then $\frac{dCM}{dt} > 0$ and

thus that I is increasing. If $\alpha < \beta$ i.e. $\frac{\alpha}{\beta} < 1$ than

$\frac{dCM}{dt} < 0$ thus that CM is lowering i.e. we have a system within the outbreak's initial epidemical stage.

Therefore, the quotient $\frac{\alpha}{\beta}$ is an important number

for this model and represents the basic reproductive

number $\mathcal{R}_0 = \frac{\alpha}{\beta}$. By definition, the basic

reproduction number $\mathcal{R}_0$ is the average number of secondary infections that a single infectious individual will give rise to for his infection, in the susceptible population.

Clear that, the epidemic will occur if the proportion of infectives increases with time i.e. $\frac{dCM}{dt} > 0$.

But we won't have an epidemic for $\frac{dCM}{dt} < 0$.

We can say:

If $\mathcal{R}_0 > 1$ then $\frac{dCM}{dt} > 0$ and we will have an epidemic. If $\mathcal{R}_0 < 1$ then $\frac{dCM}{dt} < 0$ and we won't have an epidemic.
 The second equation of the model (2)

$$\frac{dCM}{dt} = \alpha S CM - \beta CM = CM(\alpha S - \beta)$$

 gave us new information that $\frac{dCM}{dt} > 0$, if it is satisfied $\alpha S - \beta > 0 \Rightarrow S > \frac{\beta}{\alpha}$. This relation means that the infection will invade in the populations if the condition $S_0 > \frac{\beta}{\alpha}$ i.e. $S_0 > \frac{1}{\mathcal{R}_0}$ is satisfied. Biologically, this means that the infection will enter the population if the initial number of susceptible individuals $S_0 > \frac{1}{\mathcal{R}_0}$.

We are not concerned about what will happen to the infection over time. For that purpose, we divide the first equation by the third equation of the model (2) and the following differential equation

$$\frac{dS}{dH} = \frac{-\alpha S}{\beta} \Rightarrow \frac{dS}{S} = -\mathcal{R}_0 dH.$$

By its solving, the solution

$$S(t) = S_0 e^{-\mathcal{R}_0(H(t)-H_0)} \quad (3)$$

is obtained.

Clear that $H(t) \leq 1$ for all time t and the solution

$S(t) > 0$ for all time t . Because $\frac{dS}{dt} < 0$, the number of individuals in the susceptible group decrement over the time t . The function $S(t)$ is positive and monotonous function means $\lim_{t \rightarrow \infty} S(t) = S_\infty$ (S_∞ is the final size of individuals in

the susceptible group). Because $\frac{dH}{dt} > 0$, the number of individuals in the recovery group increment over the time t . The function $R(t)$ is monotonous and bounded by 1 means $\lim_{t \rightarrow \infty} H(t) = H_\infty$ (H_∞ is the final size of individuals in the recovery group). The epidemic will be

stopped, if the final size of individuals in the infected group $H_\infty = 0$ ($\lim_{t \rightarrow \infty} H(t) = H_\infty$).

Because $S(t) + CM(t) + H(t) = 1$, the final sizes and the initial values $H_\infty = 1 - S_\infty, H_0 = 1 - S_0$ are obtained. The solution (3) will have the following form

$$S_\infty = S_0 e^{\mathcal{R}_0(S_\infty - S_0)}.$$

This equation can be solved only numerically, and its solution will give us the final size of individuals in the susceptible group.

2.2.2. SIRS-model without vital dynamics

The SIRS model without vital dynamics (20) for the spreading of infectious diseases is a simple mathematical model obtained from the SIR model without vital dynamics. In this model, all recovered individuals return unprotected to the susceptible group, and they can be infected again. In Figure 2, the SIRS model is shown.

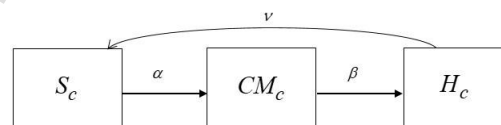


Figure 2. The model SIRS without vital dynamics

The rate by which recovered cows return to the susceptible group is denoted by v . The time of recovery (in the case of mastitis unspecific immune protection against mastitis pathogens) is $\tau = \frac{1}{v}$.

This model is described by the following ordinary differential equations:

$$\begin{aligned} \frac{dS_c}{dt} &= -\frac{\alpha S_c CM_c}{N} + v H_c \\ \frac{dCM_c}{dt} &= \frac{\alpha S_c CM_c}{N} - \beta CM_c \\ \frac{dH_c}{dt} &= \beta CM_c - v H_c \end{aligned}$$

with the initial values

$$S_c(0) = S_{c_0} > 0, CM_c(0) = CM_{c_0} > 0, H_c(0) = H_{c_0} \geq 0.$$

We need to observe the long-term behavior of the system, i.e., we are interested in finding equilibrium points where none of the variables change. For the variables that are not changed, their first derivatives must be zero:

$$\begin{aligned} -\frac{\alpha S_c CM_c}{N} + \nu H_c &= 0 \\ \frac{\alpha S_c CM_c}{N} - \beta CM_c &= 0 \\ \beta CM_c - \nu H_c &= 0 \end{aligned}$$

By solving this system, two solutions are obtained. This means that model has two equilibrium points: the disease-free equilibrium point (DFE) $(S_c^0, CM_c^0, H_c^0) = (N, 0, 0)$ and the endemic equilibrium point EE $(S_c^*, CM_c^*, H_c^*) = (\frac{N}{\mathfrak{R}_0}, \frac{N}{\mathfrak{R}_0} \cdot \frac{\nu}{\nu + \beta} (\mathfrak{R}_0 - 1), \frac{N}{\mathfrak{R}_0} \cdot \frac{\beta}{\nu + \beta} (\mathfrak{R}_0 - 1))$,

where the basic reproductive number is $\mathfrak{R}_0 = \frac{\alpha}{\beta}$. If

$\mathfrak{R}_0 > 1$ then we will have an epidemic. If $\mathfrak{R}_0 < 1$ then we won't have an epidemic. If $\mathfrak{R}_0 < 1$ then the DFE will be a stable node. If $\mathfrak{R}_0 > 1$ then the DFE will be a saddle point (unstable).

If $[\nu(\alpha + \nu)]^2 - \nu(\alpha - \beta)[2(\beta + \nu)]^2 \geq 0$ and $\mathfrak{R}_0 < 1$ then the EE will be a saddle point (unstable).

If $[\nu(\alpha + \nu)]^2 - \nu(\alpha - \beta)[2(\beta + \nu)]^2 \geq 0$ and $\mathfrak{R}_0 > 1$ then the EE will be a stable node. If $[\nu(\alpha + \nu)]^2 - \nu(\alpha - \beta)[2(\beta + \nu)]^2 < 0$ and $\mathfrak{R}_0 > 1$ then the EE will be a stable focus and the

following condition $\mathfrak{R}_0 > \frac{[\nu(\alpha + \nu)]^2}{4\beta\nu(\beta + \nu)^2} + 1$.

2.3. Machine learning

For data prediction, Weka software (21) was used. Weka is a collection of machine learning (ML) algorithms for data mining tasks. Weka is widely used in academia, research, and industry for educational purposes, prototyping machine learning solutions, and exploring data mining techniques. For prediction, we used all available data for 1031 cows from the database and classification rules, such as DecisionTable, M5Rules and ZeroR rules. We used cross-validation folds 10.

3. RESULTS

The annual prevalence of clinical mastitis, calculated as the rate per 100 cow/years at risk, per 100 lactation and as lactation incidence risk is shown in Table 1.

Table 1. The annual prevalence rate of clinical mastitis in the observed population of dairy cows

Farms	Prevalence rate per 100 cows/years at risk	Prevalence rate per 100 lactations	Lactation incidence risk
A	24.69%	20.10%	25.00%
B	50.70%	51.35%	95.58%
C	25.59%	19.97%	21.49%
Total	34.14%	30.07%	45.86%

The total prevalence of clinical mastitis for the entire observed population of dairy cows, calculated per 100 cows/years at risk, was 34.14% and 30.07% calculated per 100 lactations. The lactation incidence risk for clinical mastitis was 45.86%.

Table 2 shows the number of cases of clinical mastitis during lactation and the occurrence of recurrent cases during the same lactation.

Table 2. Reoccurrence of clinical mastitis cases within a single lactation

Cows in lactation that experienced						
Farm s	1 case of CM	2 cases of CM	3 cases of CM	4 cases of CM	5 cases of CM	6 cases of CM
A	82.92 %	9.75 %	7.31 %			
B	50.23 %	27.27 %	12.44 %	7.17 %	1.91 %	0.95 %
C	92.36 %	7.63 %				
Total	68.24 %	18.63 %	7.61 %	3.93 %	1.04 %	0.52 %

From the analysis of showed results in Table 2 it might be noticed that most of the cows suffered from one case of clinical mastitis during the same lactation, and the number of cows suffered from recurrent consecutive cases (two, three, four or more) was rare.

The SIR model

The total population of the cows in the trial is $N = 1031$ with the initial values $S_{c_0} = 439$, $CM_{c_0} = 352$, $H_{c_0} = 240$. The rate of infection is $\alpha = 0,3414$ and the rate of recovery is $\beta = 0,6824$. The time of recovery is $\tau = \frac{1}{\beta} = \frac{1}{0,6824} = 1,46542$ days plus a nine-day

interval needed for the pathogenesis of processes in the mammary gland related to mastitis. The basic reproduction number

$$\mathfrak{R}_0 = \frac{\alpha}{\beta} = \frac{0,3414}{0,6824} \approx 0,5003 < 1 \text{ i.e. the epidemic}$$

will be avoided. In addition to this the fact that i.e.

$$S_0 = \frac{S_{c_0}}{N} = \frac{439}{1031} \approx 0,4258 > \frac{1}{\mathfrak{R}_0} = \frac{1}{0,5003} = 1,9988 \cdot$$

In Figure 3 is shown the functions $S_c(t)$, $CM_c(t)$ and $H_c(t)$:

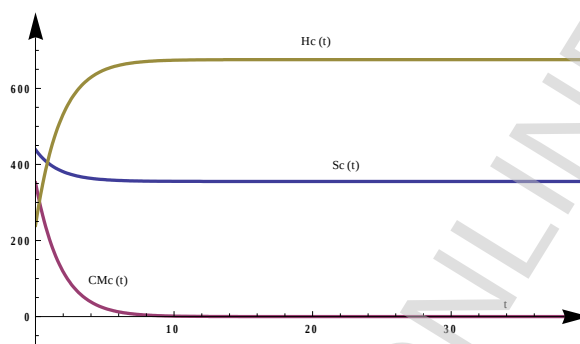


Figure 3. The functions $S_c(t)$, $CM_c(t)$ and $H_c(t)$

The SIRS model

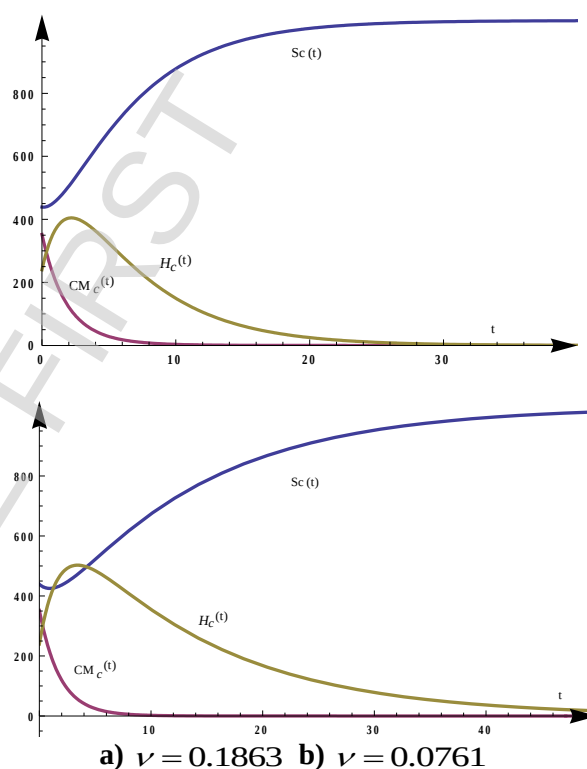
The total population of the cows included in the trial is $N = 1031$ with the initial values $S_{c_0} = 439$, $I_{c_0} = 352$, $R_{c_0} = 240$. The rate of infection is $\alpha = 0,3414$ and the rate of recovery is $\beta = 0,6824$. The basic reproduction number is

$$\mathfrak{R}_0 = \frac{\alpha}{\beta} = \frac{0,3414}{0,6824} \approx 0,5003 < 1. \text{ The equilibrium}$$

point DFE is $DFE = (1031, 0, 0)$.

In our case, we will compare the epidemic behaviour to the recurrence of the disease one time, two, three, four, and five times. In all cases, the

disease-free equilibrium points $DFE = (1031, 0, 0)$ and the basic reproduction number $\mathfrak{R}_0 \approx 0,5003 < 1$ are the same. The graphical presentations of the functions $S_c(t)$, $CM_c(t)$ and $H_c(t)$ for the different values of the rate by which recovered cows return to the susceptible group at times of recurrence of the disease ν are obtained in Figure 4:



a) $\nu = 0.1863$ b) $\nu = 0.0761$

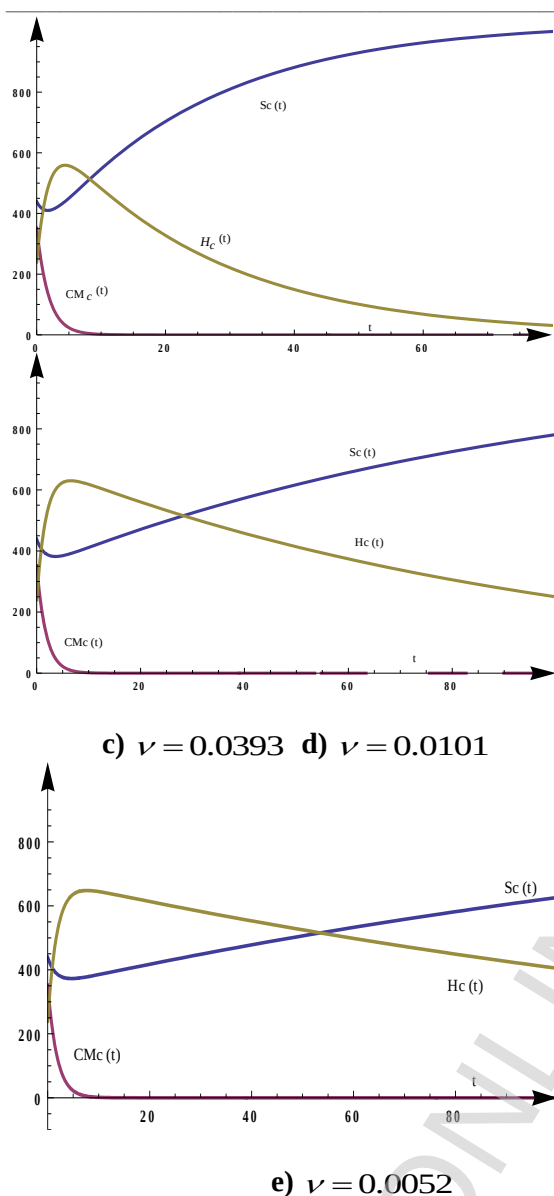


Figure 4. The graphical presentations of the functions $S_c(t)$, $CM_c(t)$ and $H_c(t)$ for the different values of ν

The time of recovery (in case of mastitis unspecific immune protection against mastitis pathogens)

$\tau = \frac{1}{\nu}$ for different values of ν (days) are:

- $\tau \approx 5.37$ for $\nu = 0.1863$;
- $\tau \approx 13.14$ for $\nu = 0.0761$;
- $\tau \approx 25.44$ for $\nu = 0.0393$;
- $\tau \approx 99.01$ for $\nu = 0.0101$;
- $\tau \approx 192.31$ for $\nu = 0.0052$.

The disease-free equilibrium point $DFE = (1031, 0, 0)$ is a stable node for all cases of ν .

Results from ML are given in Table 3 below. According to the results, we can see that the best correlation coefficient we have while using decision table rules. That means the strength of a linear relationship between variables is good. With ZeroR we have a negative coefficient. Root mean squared error is small in all cases, but the relative absolute error is smaller while using the decision table and M5Rules. The best results are obtained using a decision table, and the worst are obtained using ZeroR classification rules. So, that means that with the decision table, we can best predict mastitis in dairy herds and with that can help prevent disease.

Table 3. Machine learning classification rules

	Classification rules		
	Decision Table:	M5Rules:	ZeroR:
Correlation coefficient	0.9988	0.8385	-0.0586
Mean absolute error	0.001	0.0109	0.406
Root mean squared error	0.0223	0.3023	0.4506
Relative absolute error	0.2418%	2.6937%	100 %
Root relative squared error	4.9541%	67.0884%	100 %

4. DISCUSSION

The widespread occurrence of clinical mastitis significantly impacts the economic stability of dairy farms due to its high prevalence and associated risks. Mastitis exemplifies a multifactorial disease, suggesting that variations in its prevalence across farms can be attributed to differences in farm management, environmental conditions, breeding systems, hygiene practices, health management, milk production, and the genetic diversity among cows, particularly in their resistance to mastitis.

The SIR model revealed that after a short time, the epidemic will disappear. From the explanation and the graphical presentations, it can be concluded that the stable point DFE attracts the trajectories of the system. Biologically speaking, the disease on the

farms is calming down, and with these parameters of the model, an epidemic cannot occur.

The SIR and SIRS models are types of compartmental models used in epidemiology to describe how diseases spread through populations. These models can also be adapted to predict the occurrence of clinical mastitis in dairy herds by considering the unique aspects of the disease and the way it spreads among cattle. The rate at which the disease spreads depends on the contact rate between susceptible and infected cows and the effectiveness of transmission per contact. The rate of recovery represents the time needed for mastitic cows to return to health (either through natural recovery or treatment). To effectively use SIR or SIRS models for predicting clinical mastitis, data for the number of susceptible, infected, and recovered cows are needed, along with transmission rates of mastitis in the herd, recovery rates and rates at which recovered cows become susceptible again. In the early stages, the number of infected cows increases rapidly due to the high transmission rate. As more cows become infected, the number of susceptible cows declines. The model predicts a peak in infections after a certain period, depending on the transmission and recovery rates. This peak represents the maximum number of cows infected during the outbreak. As infected cows recover or are removed from the population (due to culling or isolation), the number of infected individuals decreases. This decline is influenced by the recovery rate. After a certain period, the SIR model reaches an endemic equilibrium where the number of susceptible, infected, and recovered cows stabilizes. This equilibrium level depends on the balance between transmission and recovery rates. Unlike the SIR model, the SIRS model accounts for the possibility of recovered cows becoming susceptible again. This introduces cyclic behaviour into the model, where individuals move between susceptible, infected, recovered, and susceptible states over time. Due to the cyclic nature of the SIRS model, there can be observed oscillations in the number of infected cows over time. These oscillations represent periodic outbreaks of clinical mastitis as recovered cows become susceptible again, leading to new infections. The SIRS model predicts long-term fluctuations in the prevalence of clinical mastitis within the dairy herd. These fluctuations are influenced by the balance between transmission, recovery, and the rate of susceptibility. Understanding the cyclic behaviour of the SIRS model can inform control

strategies. For instance, good management practices aimed at reducing transmission can help dampen the oscillations and stabilize the prevalence of clinical mastitis in the long term.

Therefore, these models can be made more complex and precise by including additional factors like varying susceptibility due to genetic factors, age, or stage of lactation, and by considering external factors like farm management practices.

With the use of the decision table as one of the most used classification rules and cross-validation folds 10 we can best predict mastitis. The decision table classifier scans through the decision table to find precise matches, focusing solely on the features specified in the schema. The cross-validation folds 10 approach involves partitioning a single dataset into ten randomly selected subsets. Nine of these subsets are utilized for training purposes, while one subset is reserved exclusively for testing. This process is iterated ten times, with each iteration selecting a different subset for testing while the remaining nine are used for training.

Single or deep ML models have traditionally been employed to classify cows into healthy and at risk of mastitis, whether clinical or subclinical. Common approaches include decision trees, distance-based methods like support vector machines, clustering models such as k-nearest neighbours and linear discriminant analysis, neural networks, and generalized linear models like logistic regression (20,21,22,23). 22,23,24,25

Recently, there has been a surge in interest in utilizing data collected from both automatic milking recording systems and routine milk recording procedures. These data, easily accessible to farmers and abundant in quantity, can be leveraged for training ML models geared towards classification or regression, aiding in the assessment of traits that may be challenging to measure directly. ML models tailored for dairy cattle have proven effective in early mastitis risk identification (26,23,27,25,28).

5.CONCLUSSIONS

Both the SIR and SIRS models provide valuable insights into the dynamics of clinical mastitis in dairy herds. While the SIR model offers a simplified representation of disease transmission and recovery, the SIRS model accounts for the loss of immunity over time, leading to cyclic behaviour in the prevalence of mastitis. By analyzing the results from these models, veterinarians and farm managers can

develop effective control strategies to minimize the impact of clinical mastitis on dairy production. With the use of the decision table as one of the most used classification rules and cross-validation folds 10 we can best predict clinical mastitis occurrence in dairy herds. Implementation of a good mastitis prevention program in dairy herds by increasing the rates of

control parameters will reduce the mastitis pathogens transmission rates leading to a reduction of mastitis incidence.

CONFLICTS OF INTEREST

The authors state that there are no conflicts of interest concerning the publication of this article.

REFERENCES

1. Nakov D., Trajchev M. Udder quarter risk factors associated with prevalence of bovine clinical mastitis. *Mac Vet Rev.* 2012, 35(2):55-64.
2. Fetai V., Nakov D., Jashari B., Trajchev M. Economics of milk yield losses in one dairy farm in Macedonia associated with clinical mastitis. *Universi-International Journal of Education, Science, Technology, Innovation, Health and Environment* 2015, 1(4):104-13.
3. Bradley A.J., Leach K.A., Breen J.E., Green L.E., Green M.J. Survey of the incidence and aetiology of mastitis on dairy farms in England and Wales. *Veterinary Record* 2007, 160(8):253-8.
4. Santman-Berends I.M., Lam T.J., Keurentjes J., Van Schaik G. An estimation of the clinical mastitis incidence per 100 cows per year based on routinely collected herd data. *Journal of dairy science* 2015, 98(10):6965-77.
5. Tommasoni C., Fiore E., Lisuzzo A., Giancesella M. Mastitis in dairy cattle: On-farm diagnostics and future perspectives. *Animals* 2023, 13(15):2538.
6. Hagnestam C., Emanuelson U., Berglund B. Yield losses associated with clinical mastitis occurring in different weeks of lactation. *Journal of dairy science* 2007, 90(5):2260-70.
7. Nakov D., Trajchev M. Monitoring udder health and milk hygiene on-farm using quick screening methods. *Journal of Agriculture and Plant Sciences* 2018, 16(2):57-63.
8. Nyman A.K., Ekman T., Emanuelson U., Gustafsson A.H., Holtenius K., Waller K.P., Sandgren C.H. Risk factors associated with the incidence of veterinary-treated clinical mastitis in Swedish dairy herds with a high milk yield and a low prevalence of subclinical mastitis. *Preventive veterinary medicine* 2007, 78(2):142-60.
9. Valde J.P., Lystad M.L., Simensen E., Østerås O. Comparison of feeding management and body condition of dairy cows in herds with low and high mastitis rates. *Journal of dairy science* 2007, 90(9):4317-24.
10. Nakov D., Hristov S., Andonov S., Trajchev M. Udder-related risk factors for clinical mastitis in dairy cows. *Veterinarski arhiv* 2014, 84(2):111-227.
11. Rowe S., Tranter W., Laven R. Longitudinal study of herd udder hygiene and its association with clinical mastitis in pasture-based dairy cows. *Journal of Dairy Science* 2021, 104(5):6051-60.
12. Levison L.J., Miller-Cushon E.K., Tucker A.L., Bergeron R., Leslie K.E., Barkema H.W., DeVries T.J. Incidence rate of pathogen-specific clinical mastitis on conventional and organic Canadian dairy farms. *Journal of dairy science* 2016, 99(2):1341-50.
13. Trajchev M., Nakov D., Petrovska M., Jankoska G. Mastitis pathogens and their antimicrobial susceptibility in early lactating dairy cows. *The Journal Agriculture and Forestry* 2017, 63(1).
14. Abdi R.D., Gillespie B.E., Ivey S., Pighetti G.M., Almeida R.A., Kerro Dego O. Antimicrobial resistance of major bacterial pathogens from dairy cows with high somatic cell count and clinical mastitis. *Animals* 2021, 11(1):131.
15. Izneid B.A., Gurmu E.D., Obsu L.L., Melese A.S., Kanan M., Al-Qerem A. Optimal Control Strategy on Mathematical Model for the Dynamics of Mastitis. *Appl. Math.* 2023, 17(4):699-711.
16. Sviland S., Waage S. Clinical bovine mastitis in Norway. *Preventive Veterinary Medicine* 2002, 54(1):65-78.
17. Wolfram Mathematica 7.0. 18.43403 Informer Technologies, Inc. <https://wolfram-mathematica.informer.com/7.0/> accessed 29/4/2024
18. Philipps S, Rossi D, Von Arb R, Capaldi A. Mathematical models of infectious diseases: Two-strain infections in metapopulations. 2011.
19. Doeschl-Wilson A. Modelling Epidemics. Lecture 5: Deterministic compartmental epidemiological models in homogeneous populations. https://jvanderw.une.edu.au/L5_ModellingEpidemics1.pdf accessed 29/4/2024.
20. Dadlani A., Afolabi R.O., Jung H., Sohraby K., Kim K. Deterministic models in epidemiology: from modelling to implementation. *arXiv preprint arXiv:2004.04675*. 2020.
21. WEKA Unveils Powerful AI-Native Data Platform Appliance for NVIDIA DGX SuperPOD with NVIDIA DGX H100 Systems. <https://www.weka.io/> accessed 29/4/2024.

22. Mammadova N.M., Keskin I. Application of neural network and adaptive neuro-fuzzy inference system to predict subclinical mastitis in dairy cattle. *Indian Journal of Animal Research* 2015, 49(5):671-9.
23. Sitkowska B., Piwczynski D., Aerts J., Kolenda M., Özkaya S. Detection of high levels of somatic cells in milk on farms equipped with an automatic milking system by decision trees technique. *Turkish Journal of Veterinary & Animal Sciences* 2017, 41(4):532-40.
24. 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. *Computers in biology and medicine* 2019, 114, 103456.
25. 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. *Scientific Reports* 2021, 11(1):13642.
26. Fadul-Pacheco L., Delgado H., Cabrera V.E. Exploring machine learning algorithms for early prediction of clinical mastitis. *International Dairy Journal* 2021, 119:105051.
27. Ebrahimie E., Ebrahimi F., Ebrahimi M., Tomlinson S., Petrovski K.R. Hierarchical pattern recognition in milking parameters predicts mastitis prevalence. *Computers and electronics in agriculture* 2018, 147:6-11.
28. Tian H., Zhou X., Wang H., Xu C., Zhao Z., Xu W., Deng Z. 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(3):427.

ONLINE FIRST