Algorithm for detecting the latent mastitis state of animals in a dairy farms on the based of data fusion from different types sensors

L V Antonov[1], A A Orlov[1] and A V Astafiev[1]

[1]Murom Institute (branch) Vladimir State University, Orlovskaya street 23, Murom, Russia, 602264

Abstract. The problem of latent mastitis identification in livestock enterprises is analyzed. The necessity of automatizing the mastitis identification process is shown. Biological methods for determining the presence of the disease are considered in the article. Common methods of data fusion for the extraction of an informative trait is analyzed in the work. A new algorithm for identifying mastitis in animals based on data fusion from the livestock enterprise sensors is proposed. The developed algorithm as compared to the conventional method of determining mastitis increases the accuracy of the disease's identification by 6.5 percent.

1. Introduction

The most common non-communicable disease animals contract in dairy plants is mastitis. The disease has two main types: subclinical (hidden) and clinical (open visual symptoms of the disease). The most dangerous of them is subclinical, when the udder and milk produced by the cow both look absolutely normal. Hidden mastitis occurs 5-10 times more frequently than clinical. If subclinical mastitis is not detected in a timely manner, then it goes to the clinical stage after some time. This leads to a situation where the animal must be removed from production for treatment with antibiotics. Animal milk becomes unfit for further processing after medicines are used [1]. The risk of mastitis has a high probability that the animals can get sick at any time of the year and at any stage of the production cycle. For example, it may be in the early days of the postpartum period or in the period of maximum lactation and even in the dry period. The disease is more dangerous in the dry period because the animal ceases to give milk and the condition of the animal's udder is not checked [2]. Mastitis is one of the most common causes of the animals' mass culling. The statistics from [3] shows that 5-35% of culled cows are animals which have mastitis that includes atrophy of parts of the udder. Statistics [1] show that the overall incidence of dairy mastitis in animals is 24-45% in all kinds of farms. The procedure for identifying the disease requires visual veterinary inspection and chemical analysis of milk from different parts of the animal's udder, if possible. The consequences of mastitis are very serious. The next lactation of animals, who have been successfully cured, becomes less productive at 7-32% compared to the average level of the herds' milk productivity [4]. Thus, we meet the topical problem of identifying subclinical mastitis cases in the early stages of the disease.

2. Problem formulation.

The usual method for the detection of subclinical mastitis is an expensive and complex clinical analysis. The essence of this method is a survey of sample milk from the udder of the sick animal.

This helps determine the number of somatic cells. Complex chemical analysis is conducted using expensive reagents [6].

The standard procedure for detection of mastitis includes several different methods used together:
- Clinical study of the udder and milk ejection.
- Taste test of the milk.
- Chemical analysis and measurement of the electrical conductivity of the milk.

Another problem is discussed in [3-4]. Detection of latent mastitis is a very time-consuming task for dairy herds where the number of animals is a few thousand. As stated, the precise identification requires veterinary examination and chemical analysis. This process takes from a few hours to a few days if there is a free laboratory and doctors who are available. Most animals are unexamined in this case. Therefore, there are methods of clipping and reducing the number of test animals in [5]. Early examination of animals suspected of mastitis is formed on the basis of their secondary features. After that, the expert decision about the treatment of each animal is accepted. On the other hand, there is a large number of different sensors installed on the animals. They are used to solve specialized problems [7,12]. Some useful information from the sensors may be lost or not used to solve a number of problems, including the task of identifying mastitis. But the information from some types of sensors can be useful for automatizing solutions to the process of determining latent mastitis on dairy enterprises. The analysis of biological methods shows that visual veterinary inspection and chemical analysis of milk from different parts of the animal's udder is required for accurate identification of the disease [8-10]. High time complexity, the use of expensive equipment and the creation of laboratory facilities within the company are required to perform this analysis. Terms of risky agriculture including tough weather conditions, short duration of daylight hours and the location in the temperate latitudes do not allow for high yields for high profitability and the creation of a company's own laboratory stations. Mastitis disease is the most common disease on farms of any type. The problem of identification of mastitis is one of the main problems. The accurate diagnosis of latent mastitis is time consuming and requires significant financial costs [11]. The creation and application of algorithms using data collected from a variety of sensors installed in the enterprise is offered as a way to reduce costs for the identification of animals suffering from mastitis.

3. Development of algorithm to identify the latent mastitis based on allocation of complex feature.

Data from 800 animals with real livestock enterprises are used as a source in the research. The connection between the traits of mastitis, including milk yield, conductivity and the emergence of the animal's mastitis, has been previously established. These parameters are the input. The presence of mastitis in animals is an output parameter of the mathematical model. It is represented in the database as 0 and 1, and it was determined by the expert (herd manager) at the company. Thus, we have a set of three parameters.

Let us introduce the following notation. We shall accept t as the average number of days in lactation (about 305 days), n as the number of the company's animals. Let us assume $\mu_i(t)$ is the value of milk production index in every day observations, $\eta_i(t)$ is the conductivity value of the index for each day of observation, $v_i(t) = \{0,1\}$ - expert opinion on the existence or lack of mastitis in the animal. A general view of the mathematical regression model looks as follows:

$$v_i(t) = k_1 \mu_i(t) + k_2 \eta_i(t) + b, \tag{1}$$

where $i = 1..n$.

Thus, the relationship between input and output parameters must be found to assess the regression coefficients k_1 and k_2. Regression analysis allows us to determine the influence of individual independent features on the result (dependent feature). If we shall accept the input parameters $\mu(t)$ and $\eta(t)$ as matrix X with two columns, output parameter $v(t)$ as a vector Y, then the vector of the regression coefficients r calculated by the ordinary least squares is:

$$r = (X^T X)^{-1} X^T Y \tag{2}$$

But the use of the regression model and the least squares method requires that the data were the same size. Therefore, the difference $|\mu_i(t) - \eta_i(t)|$ (where $i = 0,1...n$) should not be large. In our case, the index of milk production is measured in thousands of mL per day, but the electrical conductivity of

milk is measured in several mS/cm. According to the definition of the Euclidean norm it makes one parameter much more significant than the other, when in fact it is not. It is noteworthy that the inputs have different distribution functions. Features must be made independent of time for the mathematical model to produce correct results. Thus, pre-processing of the necessary data must be carried out. Normalization and centering of random functions must be performed.

Let us perform the centering of the function $\mu_i(t)$ according to (3), (4).

$$T(t) = \frac{1}{n}\sum_{i=1}^{n}\mu_i(t) \tag{3}$$

where $i = 1..n$.

$$R_i(t) = \mu_i(t) - T(t) \tag{4}$$

where $i = 1..n$.

The expected value of the function $R(t)$ is 0, and the data are uniformly distributed relative to the x-axis after the conversions (3), (4). Thus, the parameter's values were converted. The distribution of random function $R(t)$ does not depend on the time after the conversion.

Let us perform the normalization of a centered function $R(t)$. Calculation of the standard deviation (SD) for the data of the herd's milk productivity produced by the (5), (6).

$$\sigma = \sqrt{\frac{1}{n}\sum_{i=1}^{n}R(t)^2} \tag{5}$$

$$\hat{\mu}_i(t) = \frac{R_i(t)}{\sigma} \tag{6}$$

where $i = 1\ldots n$.

Then we can calculate the value of the regression coefficients and the free terms of the equation by using (2). Values that were calculated are show in Table 1.

Table 1. The regression coefficients for equation (1).

Regression Coefficient	Value
k_1 (Coefficient of milk production)	-0.0493
k_2 (Conductivity coefficient)	0.1933
b (free term)	0.1022

Thus, the plane that approximates the initial space of the mastitis traits is shown as a set of points in the graph constructed by the input coordinates is found (Fig. 1). The regression model gives the required complex feature, allowing for the estimation of the probability of mastitis in the animals. The regression model to determine the presence of mastitis in the animals is shown in (7). Coefficients are calculated using the least squares method.

$$v_i(t) = -0.0493 \cdot \hat{\mu}_i(t) + 0.1933 \cdot \hat{\eta}_i(t) + 0.1022 \tag{7}$$

where $i = 1..n$.

The practical purpose of the research is the detection of mastitis using the threshold value of the complex feature. The recognition threshold (P) is calculated according to the range of confidence probability, which determines the livestock expert enterprise. Entering the threshold manually allows us to the identify of 100 animals with mastitis because the value of confidence interval is low. Nevertheless, among the sampled animals suspected to disease there are also many healthy cows. Therefore, manual inspection of all the suspected animals is not possible because of the large count of cows in the sample. There is type II error. Thus, the confidence interval may be changed depending on how many animals need to be checked by the workers per day, and what level of identification accuracy is required.

The first step is building a histogram of the probability density distribution for calculating complex mastitis traits. The next stage is obtaining the threshold value P in accordance with the installed confidence interval. Suspicion of disease in the animal is determined when the value of the complex

mastitis trait, based on milk yield and conductivity exceeds the threshold P. The animal is entered into a special sample. These methods are based on a single parameter—the electrical conductivity of milk. Thresholds of electrical conductivity to determine the animal's health conditions are given in [6] (Table 2).
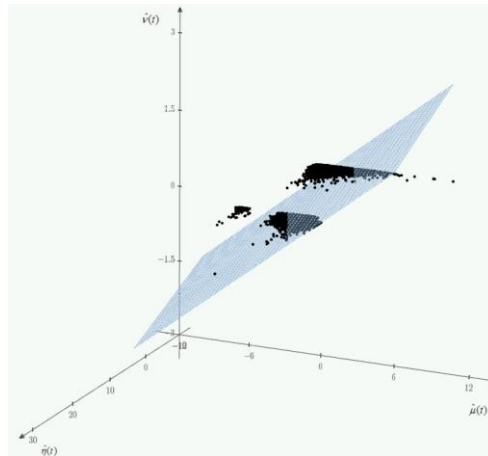

Figure 1. The plane which approximates the feature space.

Table 2. Milk conductivity in healthy and ill quarters (operating instructions).

The Health Status of the Udder	Milk conductivity (mS/cm)
Healthy udder	< 5.5
Suspicion of mastitis udder	$5.5 - 6.6$
Subclinical mastitis	> 6.5
Clinical mastitis (Milk must not be delivered)	> 8

4. Experimental results.

Currently there are several approaches for preliminary identification of mastitis in animals without using chemical analysis. Generally, mastitis in animals is identified when the milk conductivity exceeds the threshold value of 6 mS/cm. This approach shows good results when you need to create a list of animals suspected of having mastitis disease without expensive equipment [6].

The algorithm based on the complex feature is compared with the algorithm for identifying mastitis in animals using the threshold 6 mS/cm of milk conductivity.

Data obtained for more than 800 cows, for about 305 days from the two sensors' measuring parameters were treated in the research. The objective quantitative results of the algorithms' experimental research for data about daily milk yield and conductivity are shown in Tables 3 and 4. The results of the algorithm were compared with an expert's estimates. The expert estimates that the results about the animal disease are not 100% accurate, because it is carried out at an early stage of the disease. But the application of this approach without long diagnostic procedures increases the efficiency of decisions. Therefore, the expert assessments are the reference data. The percentage of sick animals by the algorithm among expert data about healthy cows is type II error. The percentage of healthy animals by the algorithm among expert data about sick cows is type I error. Type I error is more important than type II error because if a sick animal is not identified, then the disease will become clinical and the animal will be removed from production. Tables 3 and 4 show that the developed algorithm based on the complex feature gives better operation results in comparison with the algorithm of fixed threshold conductivity.

The algorithm developed reduces type I error to 6.5%. It is possible to increase the number of sick animals, which are found at exactly 10 units. Type II error has been decreased to 3%. Thus the algorithm helps achieve a smaller number of false positives for identification of bovine mastitis.

Table 3. Results comparison of algorithm based on the threshold of milk conductivity 6 ms/cm with expert opinion.

Percentage of Animals That are Found by the Algorithm	Expert Opinion	
	Sick Animals, %	Healthy Animals, %
Sick Animals	89.5	9
Healthy Animals	10.5	91

Table 4. Results comparison of developed algorithm based on the complex feature with expert opinion.

Percentage of Animals That are Found by the Algorithm	Expert Opinion	
	Sick Animals, %	Healthy Animals, %
Sick Animals	96	6
Healthy Animals	4	94

Errors of identification are decreased by the help of the additional trait, low level milk yield, for the calculation of complex mastitis traits. The scalable fragment of changes of milk yield values, milk conductivity and the calculated complex mastitis traits during lactation are shown in Fig. 2.
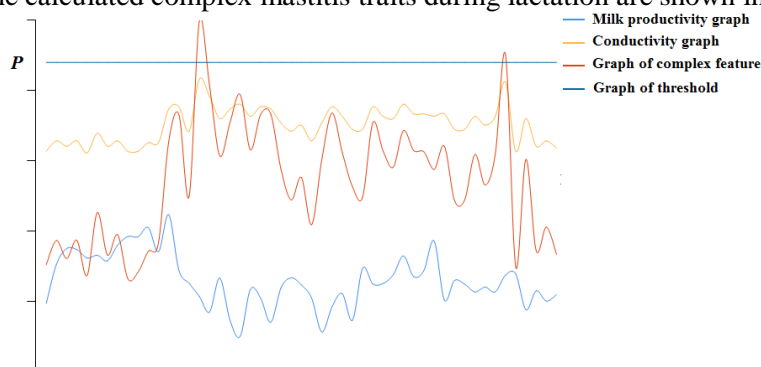

Figure 2. Comparison of the results of algorithms identification latent mastitis in animal.

The graphic fragment of lactation for one of the animals, including changing values measured by the sensors, is shown in Fig. 3. The dynamics of changes in the rate of electrical conductivity shows that there is mastitis in the animal. This is not the same as the expert evaluation. This is a mistake. The graph of the complex trait is stable and does not exceed the threshold P, because milk productivity is normal. Thus, type II error is eliminated.

The estimated algorithm for more than 800 cows was produced daily. The value of complex mastitis traits were calculated for each animal daily. Characteristic values exceeding the threshold P were considered abnormal. The confidence interval was 90%. The results of evaluation of the algorithm research are shown in Table 5.

Table 5. A Quantitative comparison of algorithms results of the mastitis identification.

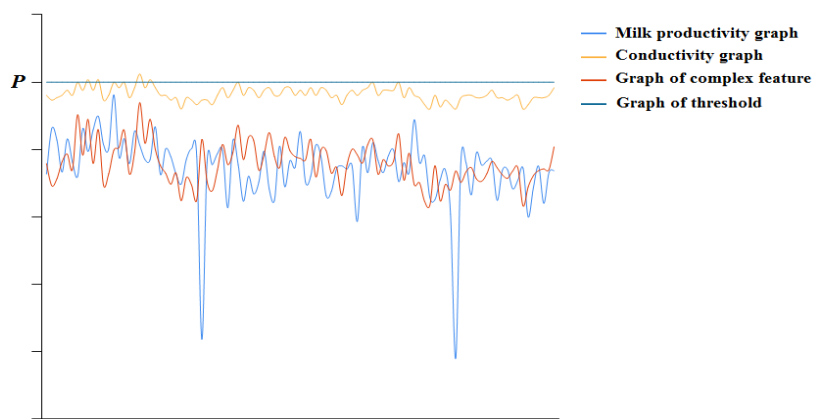| Applied Algorithm	All days	Critical Deviations Found
Identification using a fixed threshold of electrical conductivity 6 mS/cm	131274	14239 (10.8%)
Identification Based on the Complex Feature	131274	13804 (10.5%)

Figure 3. Comparison of the results of algorithms identification latent mastitis in animal.

5. Conclusion

The relevance of timely identification of subclinical mastitis is shown in the work. The process of accurately detecting bovine mastitis is expensive. Furthermore, accurate identification requires the manual inspection of large numbers of animals. There are not enough workers in livestock to solve this problem. Therefore, automation of the disease identification process is necessary. Methods of forming a preliminary list of animals suspected to have mastitis are considered in the article. These methods are based on data analysis from sensors. The approach is based on the dedicated complex traits which allow for identifying the disease proposed in the work. The experimental results show that the developed algorithm for the mastitis detection is more accurate than the traditional approach, which is described in [6]. Use of complex mastitis traits allows for an increase in the accuracy of identifying mastitis by 6.5% and a 3% decrease in false positives. Thus, the proposed algorithm shows promising results. It can be used in program systems for monitoring the production of livestock farms.

6. References

[1] Gianneechini R E, Concha C and Franklin A 2002 Antimicrobial susceptibility of udder pathogens isolated from dairy herds in the west littoral region of Uruguay *Acta Vet Scand.* **43** 31-41

[2] Hillerton J E and Semmens J E 1999 Comparison of treatment of mastitis by oxytocin or antibiotics following detection according to changes in milk electrical conductivity prior to visible signs *J. Dairy Sci.* **82** 82-93

[3] Waage S, Jonsson P and Franklin A 1994 Evaluation of a cow-side test for detection of Gramnegative bacteria in milk from cows with mastitis *Acta Vet. Scand.* **35** 200-207

[4] Antonov L V, Makarov K V and Orlov A A 2015 Development and experimental research on production data analysis algorithm in livestock enterprises *Procedia Engineering* **129** 664-669

[5] Norberg E 2005 Electrical conductivity of milk as a phenotypic and genetic indicator of bovine mastitis *Livestock Production Science* **96** 129-139

[6] Janzekovic M, Brus M, Mursec B, Vinis B, Stajnko D and Cus F 2009 Mastitis detection based on electric conductivity of milk *Journal of Achievements in Materials and Manufacturing Engineering* **34** 39-46

[7] Kate M 2000 Millar respect for animal autonomy in bioethical analysis: the case of automatic milking systems (AMS) *Journal of Agricultural and Environmental Ethics* **12** 41-50

[8] Lovendahl P and Chagunda M 2006 Assessment of fertility in dairy cows based on electronic monitoring of their physical activity *Proceedings of the 8th World Congress on Genetics Applied to Livestock Pro-duction* 496-500

[9] Steeneveld W, van der Gaag L C, Ouweltjes W, Mollenhorst H and Hogeveen H 2010 Discriminating between true-positive and false-positive clinical mastitis alerts from automatic milking systems *Journal of Dairy Science* **93(6)** 2559-2568

[10] Sun D, Li C, Gu C, Chen J, Qu Y, Wang X, Gao J, Wei S, Wang J, Wu R and Guo D 2015 Analysis of Mineral Elements, Metabolism, and Inflammation Indexes in the Plasma of Dairy Cows Suffering from Different Degrees of Lameness *Biological Trace Element Research* **168** 372-379

[11] Zarchi H and Ingi R 2009 Improving Oestrus Detection in Dairy Cows by Combining Statistical Detection with Fuzzy Logic Classification *Proceedings Workshop on Advanced Control and Diagnosis*

[12] Yakovleva T V 2017 Determining the Phase Shift of Quasiharmonic Signals through Envelope Analysis *Computer Optics* **41(6)** 950-956 DOI: 10.18287/2412-6179-2017-41-6-950-956