LSTM MODELS TO SUPPORT THE SELECTIVE ANTIBIOTIC TREATMENT STRATEGY OF DAIRY COWS IN THE DRY PERIOD

Joana Baptista Rafael

Dissertation

presented as partial requirement for obtaining the Master Degree Program in Data Science and Advanced Analytics

NOVA Information Management School
Instituto Superior de Estatística e Gestão de Informação
Universidade Nova de Lisboa

NOVA Information Management School
Instituto Superior de Estatística e Gestão de Informação
Universidade Nova de Lisboa

LSTM MODELS TO SUPPORT THE SELECTIVE ANTIBIOTIC TREATMENT STRATEGY OF DAIRY COWS IN THE DRY PERIOD

by

Joana Baptista Rafael

Dissertation presented as partial requirement for obtaining the Master's degree in Advanced Analytics, with a Specialization in Data Science

Advisor: Prof. Dr. Roberto Henriques

Co-Advisor: Prof. Telmo Nunes

November 2022

STATEMENT OF INTEGRITY

I hereby declare having conducted this academic work with integrity. I confirm that I have not used plagiarism or any form of undue use of information or falsification of results along the process leading to its elaboration. I further declare that I have fully acknowledge the Rules of Conduct and Code of Honor from the NOVA Information Management School.

Lisbon, November 22th, 2022

ACKNOWLEDGMENTS

Reaching the end of one more step in my academic journey, I consider that this Master's degree has truly had a unique feeling. Firstly, for having acquired valuable knowledge at the cost of hard work and resilience. Secondly, in this dissertation, I could combine two areas I cherish: data science and veterinary medicine.

I would like to thank my advisor, Prof. Dr. Roberto Henriques, for his support and availability during the development of this dissertation.

I would like to thank my co-advisor, Prof. Telmo Nunes, for his motivation and always being able to count on his availability, dedication, and countless brainstorming sessions.

I would like to give a special acknowledgment to Dr. Luís Pinho, from SVA, for his kindness in providing the data, his precious support, and our weekly meetings to address and clarify my queries, and share ideas.

To my wonderful classmates, Ricardo Santos and Daniel Correia. They have proven that companionship and friendship at work are the best things we can have.

To Christopher Rosky, for being an outstanding teacher and for his extraordinary patience in reviewing the texts.

To my soulmate, Pedro for his care, support, and easy laugh that makes everything easier.

To my Father, who always encourages me to enjoy life as much as possible.

And finally, I would like to thank my Mother, who has always been there. Who supported me in the defeats and celebrated with me in the victories.

ABSTRACT

Udder inflammation, known as mastitis, is the most significant disease of dairy cows worldwide, invoking substantial economic losses. The current common strategy to reduce this problem is the prophylactic administration of antibiotics treatment of cows during their dry period. Paradoxically, the indiscriminate use of antibiotics in animals and humans has been the leading cause of antimicrobial resistance, a concern in several public health organizations. In light of these assumptions, at the beginning of 2022, the European Union made it illegal to routinely administer antibiotics on farms, with Regulation 2019/6 of 11 December 2018.

Considering this new scenario, the objective of this study was to produce a model that supports the decisions of veterinarians when administering antibiotics in the dry period of dairy cows. Deep learning models were used, namely LSTM layers that operate with dynamic features from milk recordings and a dense layer that uses static features. Two approaches were chosen to deal with this problem. The first is based on a binary classification model that considers the occurrence of mastitis within 60 days after calving. The second approach was a multiclass classification model based on veterinary expert judgment. In each approach, three models were implemented, a Vanilla LSTM, a Stacked LSTM, and a Stacked LSTM with a dense layer working in parallel.

The best performances from binary and multiclass approaches were 65% and 84% accuracy, respectively. It was possible to conclude that the models of the multiclass classification approach had better performance than the other classification.

The capture of long- and short-term dependencies in the LSTM models, especially with the combination of static features, obtained promising results, which will undoubtedly contribute to producing a machine learning system with a prompt and affordable response, allowing for a reduction in the administration of antibiotics in dairy cows to the strictly necessary.

KEYWORDS

Machine Learning; Deep Learning; LSTM; Dairy Cows; Antimicrobial Resistance; Mastitis

INDEX

LIST OF FIGURES

LIST OF TABLES

LIST OF ABBREVIATIONS AND ACRONYMS

AMR	Antimicrobial Resistance
AMS	Automatic Milking Systems
AUCROC	Area Under the Rock Curve
BCE	Binary Cross Entropy
BDCT	Blanket Dry Cow Therapy
CAP	Common Agricultural Policy
CCE	Categorical Cross Entropy
CNN	Convolutional Neural Networks
CRISP-DM	Cross Industry Standard Process for Data Mining
DCT	Blanket Dry Cow Therapy
DIM	Days in Milk
DLNN	Deep Learning Neural Networks
DrCIF	Interval-Based Diverse Representation Canonical Interval Forest
DTW	Dynamic Time Warping
FCN	Fully Convolutional Network
FFNN	Feed Forward Neural Networks
HIVE-COTE	Hierarchical Vote Collective of Transformation-Based Ensembles
IMI	Intramammary Infection
ITS	Internal Teat Sealant
LDH	Lactate Dehydrogenase
LSTM	Long Short-Term Memory
MFD	Milk Fat Depression
ML	Machine Learning
MLP	Multi-Layer Perceptron

MPD	Milk Protein Depression
MTSC	Multivariate Time Series Classification
NLP	Natural Language Processing
PPV	Proportion of Positive Values
RNN	Recurrent Neural Networks
ROCKET	Random Convolutional Kernel Transform
SARA	Sub-Acute Ruminal Acidosis
SCC	Somatic Cell Count
SDCT	Selective Dry Cow Therapy
STC	Shapelet Transform Classifier
SVA	Veterinary Services and Associates
TDE	Temporal Dictionary Ensemble
UE	European Union
WHO	World Health Organization

1. INTRODUCTION

The milk industry is a challenging business in our current times due to its complex system and reduced profit margins (Gołaś, 2017). For this market to continue circumventing inflationary pressures and political restrictions, herds have chosen to implement new technologies which enhance production and select dairy cows genetically more capable of producing milk. However, animals with greater productive ability are associated with a higher incidence of mastitis resulting from the metabolic, hormonal and immunity stress they are subject to in intensive production.

Mastitis, defined as the inflammatory response resulting from the infection of the udder tissue, is responsible for significant economic losses to dairy herds and the milk processing industry, provoking reduced milk yield, waste milk, treatments, and veterinary services costs (Neculai-Valeanu & Ariton, 2022) Some strategies to reduce the disease include implementing hygiene plans (such as installation cleanliness, milking materials, and post-milking teat disinfection), monitoring environmental factors (e.g., humidity/temperature) and evaluating milk parameters. Still, antibiotics have conventionally been seen as the primary line of treating and preventing mastitis, especially in the dry period.

The continued and indiscriminate use of antibiotics in animals and humans has allowed the rapid growth of antibiotic-resistant microorganisms. This problem has been scrutinized by different public health organizations worldwide. Research by the World Health Organization warned about the inability of antibiotics to eliminate bacterial infections that are expected to be one of the leading causes of death globally in the near future (Higgins et al., 2017).

On 28 January 2022, 'Regulation 2019/6 of 11 December 2018' was widely implemented in the European Union, which restricts the administration of antibiotics prophylactically. It was stressed that all animals must be evaluated individually, considering the risk of infection.

This new panorama has accelerated the need to create alternatives that support veterinarians' decisions. This study aims to produce a deep learning model, adapted to the reality of Portuguese herds, that supports veterinarians in making more conscious decisions to mitigate the indiscriminate use of antibiotics. The data were requested from consulting company SVA (Veterinary Services and Associates), which follows and monitors herds in mainland Portugal and the Azores. Dairy cows were selected based on pre-defined criteria from January 2018 to December 2019. After the data were prepared and transformed, they served as input to LSTM models, in which two different approaches were applied. The first uses a model with a binary mastitis detection output based on a threshold of >200.000 Somatic Cell Count during the first 60 days of milk production, and the second employing a model that uses multiclass output based on the decision of a veterinarian specialized in the field.

The results proved to be the beginning of reducing the administration of dairy cows to what is strictly necessary.

2. LITERATURE REVIEW

2.1. DAIRY COWS

2.1.1. The Milk Industry in the EU and the Portuguese Context

The European Union (EU) dairy industry is quite heterogenous, considering the diversity of environmental, political, and strategic factors into which herd falls (Philippidis & Waschik, 2019; Čechura et al., 2021). According to a report by the European Commission, the dairy industry is considered the second largest agricultural sector, about 12% of total output, which translates into an average production of 152 million tons of cow's milk per year (European Commission, 2022). The large capacity to produce milk allows the EU to be self-sufficient and one of the leading international market suppliers of consumer milk, cheese, and powdered milk (Eurostat, 2022).

Over the past few years, there has been a widespread trend toward a decline in small herds, mainly because of successive reforms of the Common Agricultural Policy (CAP), which caused the European quota system to fall in 2015. In other words, with this restructuring, there was the abolition of production limitations on herds and the elimination of fixed-price guarantees in milk, which had long been immune to the law of the supply and demand market (Bórawski et al., 2020; Martinho, 2020). This factor, coupled with a timid demand from the external market in the same year, caused devastating effects on small producers, and there was the necessity for an urgent reorganization of the dairy industry facing this new reality (Bórawski et al., 2020; Čechura et al., 2021). These adaptations resulted in more robust herds caused by new technological, structural, and nutritional implementations, which increased the cattle population's environmental and health quality. This naturally resulted in the presence of animals more capable of producing a more substantial quantity of milk (Nalon & Stevenson, 2019).

The Portuguese context follows the overall trend of the European dairy market. Portugal has been self-sufficient in producing milk and butter, products considered to have lower added value but deficient in yogurts and cheeses, products of greater added value (INE, 2015; Martinho, 2020). Even though there was a 90% reduction in the number of herds between 1989 and 2013, milk production in Portugal has shown a regular growth trend (Sottomayor et al., 2020). According to the data from the European Milk Observatory, although currently facing severe challenges related to the rampant inflation of essential raw materials such as food and energy, as well as indirectly, the labor force and products needed for production, the country systematically has the lowest milk sales price in the EU (European Commission, 2022). From a geographical point of view, two large milk-producing areas host about 80% of national production, the northern and central areas and the Autonomous Region of the Azores, with about 50% and 30% of production, respectively. This fact seems to be determined by the regions' good climatic and strategic conditions (Sottomayor et al., 2020).

2.1.2. The Lactation Cycle

The lactation curve of dairy cows is a model designed to provide information on the performance of herd lactation cycles. According to 2019 INE data, Portugal has a population of approximately 245,000 dairy cattle (PARCA, 2021). Each animal remains in a herd for about 5.5 years, averaging 2.6 lactation cycles (Stilwell, 2013; ANABLE, 2014).

The lactation cycle consists of two phases: the lactation period and the dry period, as shown in Figure 1. The lactation period begins with the calving and has an average duration of 305 days in cases when the insemination is within the established period (Ball et al., 2004; Dobson et al., 2007). The lactation peak is reached within this period, which occurs in the first 30 to 60 days postpartum (Castro et al., 2009; Rodrigues et al., 2012). At this stage, there is more outstanding milk production, between 38 and 45L per day, which can persist for up to 4 weeks (Webster, 2020). Once there is a more significant energy requirement, there is also a greater need for food intake, which is considered a critical period for mastitis and metabolic diseases. After the peak of lactation, a slow reduction in milk yield begins until the arrival of the dry period (Webster, 2020).

The dry period specifies when there is no milk yield. At this stage, there is induction of drying of the teats and, typically, by administrating intra-mammary antibiotics. Several studies indicate that this break in yield allows better performance in the subsequent lactation cycle, as the cow is in the last two months of gestation and needs the mammary system to be involute and prepared for the new cycle (Hoeij et al., 2017).

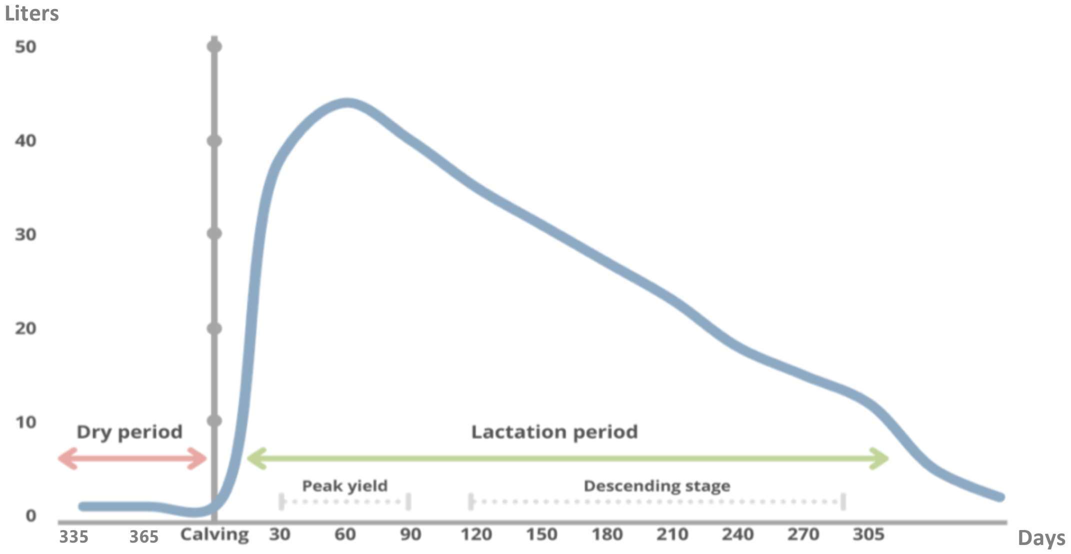


Figure 1: Lactation Cycle: Dry period - Starts 60 days before calving; Lactation period - between 0-305 days. This period includes Peak yield and the Descending stage.

2.1.3. Milk Recordings

Milk recording is an optional method practiced in herds with ambitious planning. The aim is to obtain information on the quantity and quality of milk production for each dairy cow in the herd.

Traditionally, in Portugal, the milk recording is carried out on a monthly basis by accredited technicians collecting individual samples of all milking over 24 hours. They are subsequently sent to an official laboratory for analysis (ICAR, 2017). However, other methods are also eligible, namely, the individual collection of a single daily sample by the technician or milking robots with built-in milk analysis systems. Whichever method is chosen, it must be approved by the International Committee for Animal Recording (ICAR, 2017) and comply with national standards in order to guarantee the uniformity of the outcomes obtained (Ordinance nº 1066/91, October 22th, 1991)

In each milk sample, several parameters are evaluated, among which stand out: somatic cell count (the set of epithelial cells and leukocytes from the mammary glands), protein content, fat content, urea, and lactose (ALIP, 2021). Other relevant data are also recorded: the total amount of milk produced, the number of days in lactation, and the number of lactation cycles performed (DGAV, 2017)

Incorporating these parameters into a database, in conjunction with their analysis, provides reliable information on the general and individual health of the cattle population. It can also play a significant role in culling decisions or even selecting animals with interesting phenotypic characteristics that allow genetic improvement in the herd (Balaine et al., 2019, 2019; Klepacki et al., 2020).

2.1.4. Mastitis Risk Factors

The incidence of clinical mastitis in European herds is estimated to vary from 20 to 35% per year (Nalon & Stevenson, 2019). This disease brings a considerable economic impact with losses in milk production, decreased fertility, and, consequently, premature culling of cows(Gröhn et al., 2004; Stilwell, 2013).

Mastitis is an inflammation of the mammary gland commonly associated with intramammary infection (IMI) and typically of bacterial etiology; it is one of the most debated problems in dairy herds (Erskine et al., 2003). These infectious agents may have an environmental origin (e.g., *Escherichia coli* and *Streptococcus uberis*) and naturally inhabit areas such as beds or stables. Alternatively, they may be of a contagious origin (e.g., *Staphylococcus aureus and Streptococcus agalactiae*), usually commensal agents of the cow (Pyörälä, 2008; Cheng & Han, 2020). IMI could present two clinical configurations: clinical mastitis, in which there is a visible inflammatory response to the infection (in which there are changings in milk properties (e.g., color, fibrin clots), the cow is prostrate and has a swollen and painful udder) or subclinical mastitis (in which the cow is asymptomatic, though the milk recording shows a slight decrease in milk production and an increase in somatic cells (≥200.000 cells/mL)) (Alhussien & Dang, 2018; Peters et al., 2015).

A complex set of environmental and host factors enhances the risk of mastitis in herds. At the environmental level, for example, poor hygiene conditions of the herds, poor ventilation, lack of humidity/temperature monitoring, and high population density allow for the favorable development of microorganisms (Nalon & Stevenson, 2019; Cheng & Han, 2020). At the host level, susceptibility to mastitis is mainly related to the immune system's competence to overcome the infection(Pyörälä, 2008).

The most critical period for a high-production cow to develop not only mastitis but also other infections and metabolic diseases (e.g., ketosis or subacute ruminal acidosis) is the transition period, which is between 3 weeks pre- and 3 weeks postpartum (Drackley, 1999; Cheng & Han, 2020). During this period, large reserves of energy and nutrients are mobilized for milk synthesis. However, the animal's feed intake during this period does not satisfy this requirement (Webster, 2020). The cow, in turn, enters a negative energy balance, concomitant with the physiological and endocrine changes that occur at this stage, promoting depression of the animal's immune system and, thus, making it susceptible to diseases. Therefore, in this phase, it is crucial to control the animal's body condition and to have adequate food management of dry matter, fat, and protein to minimize the occurrence of said issues (Butler, 2005).

2.1.5. The Problem with Antibiotics

Since the beginning of the 50s, Blanket Dry Cow Therapy (BDCT), or the administration of antibiotics at the end of the lactation period, has been an underlying practice in mastitis management and control plans on herds around the world (Berry & Hillerton, 2002; Biggs et al., 2016). This practice aims to eliminate latent infections and prevent the appearance of new IMIs in the dry period through the administration of long-acting antibiotics in each quarter (Scherpenzeel et al., 2014).

In BDCT, the administration of antibiotics is done without prior consideration of each animal’s context, i.e., regardless of the individual evaluation of the milk recordings or the risk of the cattle population developing mastitis, this is continuously performed in a prophylactic way.

The short-term benefits of this practice are indisputable, and several studies indicate that there is less clinical mastitis in treated than untreated cows. However, the selective pressure exerted on bacteria over the years by the indiscriminate use of antibiotics in humans and animals leads to prioritizing measures to control antimicrobial resistance (AMR) (Huijps & Hogeveen, 2007; Scherpenzeel et al., 2014).

AMR occurs when microorganisms develop biological mechanisms that allow them to resist the effects of pharmacological substances, which have, until recently, been efficient in eliminating them. These adaptations already existed long before the discovery of antibiotics, although their incorrect and excessive use accelerated this process (Berry & Hillerton, 2002; Regea, 2018).

Research by the World Health Organization (WHO) estimates that, by 2050, AMR could compromise 10 million lives if urgent measures are not implemented (Higgins et al., 2017). Given this evidence (and under Regulation 2019/6 of December 11th, 2018), the EU banned the routine administration of antibiotics on farms on January 28th, 2022. Regarding dairy cows, Article 107.3 of the regulation states that they can be used in exceptional cases where the risk of developing mastitis is very high. Nonetheless, a responsible analysis must be carried out on each animal (Nunan, 2022). Some European countries, such as Sweden, Denmark, and the Netherlands, have adopted Selective Dry Cow Therapy (SDCT) for decades. In some of them, there has even been a limitation in dosing and treatment length (Armstrong et al., 2018; Niemi et al., 2020).

In SDCT, based on the clinical history of the cow, the veterinarian assesses the need for medication administration. Generally, the decision is to administer Internal Teat Sealant (ITS), which provides a physical barrier between the quarters and external microorganisms. Antibiotics can also be included (Biggs et al., 2016; Nickerson & Ryman, 2019; Biscarini et al., 2020). Beyond just the treatment, raising producers' awareness to carry out rapid diagnoses and promote good hygiene on the herd seems to be the key to reducing the incidence of mastitis (Armstrong et al., 2018).

2.2. MULTIVARIATE TIME SERIES CLASSIFICATION

2.2.1. A Brief Introduction to Time Series

A time series is a temporal sequence of points with delimited temporal spacing. Also called a stochastic process, this can be described as a set of random variables $\{x1, x2, ..., xt, t \in T\}$ indexed in time, where T is an ordered set of indices. The stochastic process can be continuous or discrete (Chatfield & Xing, 2019). In a continuous time series, the observed variable is typically recorded continuously. An example is that of the heart's electrical activity data recorded on an ECG for 10 minutes, with the expression:

$$T = [0, \infty [$$

A discrete time series is presented by a set of numerable T values separated by time. An example of a discrete-time would be the recording of milk production between the years 2018 and 2019, given by the following expression:

$$T = \{0, 1, 2, . . . n\}$$

Time series can also be classified as univariate if the sequence of measurements comes from a single variable (a single sequence of real numbers) or multivariate, where the sequence of measurements from multiple variables is collected (a vector of real numbers).

2.2.2. Current Methods for Multivariate Time Series Classification

In recent years, Multivariate Time Series Classification (MTSC) has received significant interest since the data can be addressed to many real-life problems, such as disease diagnosis, human activity recognition, or anomaly detection (e.g., bank fraud, machine fault detection).

In general, multivariate time series can be defined as ordered observation series made sequentially through time.

Developing algorithms that work with MTSC problems has been a constant challenge, given the extraordinary complexity of analyzing numerous variables. These have the particularity of progressing in time and can be simultaneously, or not, dimension-dependent.

Traditional methods without machine learning algorithms only solve straightforward recognition patterns where it is possible to draw a threshold. Meanwhile, standard supervised machine learning classification algorithms are not adapted to be applied to problems with the data ordered sequentially (Faouzi, 2022; Ruiz et al., 2021). Considering these limitations, different approaches have been developed involving Machine Learning (ML) and distance-based algorithms, of which the following stand out:

2.2.2.1. Dynamic Time Warping

Dynamic Time Warping (DTW) is a distance-based algorithm used to measure the distance between two similar time series. DTW, combined with the 1-nearest neighborhood, is used to find the optimal global alignment between similar time series. The algorithm has been the gold standard for univariate time series classification for the past decade. For MTSC problems, Shokoohi-Yekta et al. carried out three possible modifications of the original model:

a) Independent DTW calculates each dimension independently and then sums up the result of the DTW distances.

b) Dependent DTW, instead of calculating the distance between two points (from the same dimension), calculates between two vectors representing all dimensions simultaneously.

c) Adaptive DTW to select the type of DWT that best adapt to the dataset.

2.2.2.2. Shapelets

Shapelets are designated as subseries of the time series with unique sequences capable of differentiating classes (Ye & Keogh, 2009). To assess the Shapelets, the first step is to extract those with the minimum euclidean distances between the time series of the same class and a large distance from the values of the other classes. Subsequently, the Shapelets are ordered using a variable ranking method (e.g., Information gain, F-Stat) (Wistuba et al., 2015). The selected Shapelets can be included as features in other standard classifiers (Faouzi, 2022). For MTSC, two possible approaches were described (Bostrom & Bagnall, 2017) :

a) Random Shapelet Forest considers each variable as an independent time series. Therefore, each time series is partitioned into random subseries, and, finally, a random forest is built into each dimension.

b) Learn Shapelets begin with a set of random subseries. They are clustered using k-means, and the centroids from these clusters are then refined using stochastic gradient descent.

2.2.2.3. ROCKET

ROCKET (Random Convolutional Kernel Transform) is an algorithm proposed by Dempster et al., 2020 that combines a large number of single 'layer' kernels (default value of 10,000 kernels) to capture relevant patterns in each time series class.

The logic behind Rocket is similar to that of Shapelets. However, convolution kernels are made up of a set of random parameters (length, weight, bias, dilation, and padding) with a fixed distribution. The result of the convolution between an input time series and random convolution kernel is a feature map (a dot product between two vectors) followed by the evaluation of the global max pooling and the Proportion of Positive Values (PPV). The resulting features of these rating methods allow the classifier to measure the prevalence of a given pattern in a time series.

The generated features can be subsequently used with a linear classifier. Usually, for small and large datasets, ridge regression and logistic regression are used, respectively. For multivariate time series, kernels are randomly assigned for each variable (dimension) (Ruiz et al., 2021). Convolution, in this case, can be interpreted as the dot product between two matrices as the kernel convolves longitudinally across the series.

2.2.2.4. HIVE-COTE

HIVE-COTE (The Hierarchical Vote Collective of Transformation-Based Ensembles) is an ensemble model consisting of a heterogeneous set of classifiers. According to the most recent update of this algorithm made by Middlehurst et al., 2021, the Hive-Cote 2.0 is composed of 4 independent classifiers: Temporal Dictionary Ensemble (TDE), the interval-based Diverse Representation Canonical Interval Forest (DrCIF), an ensemble of Rocket algorithm (Arsenal), and an updated version of the Shapelet Transform Classifier (STC). Each classifier aims to train the time series input, independently of the other classifiers, to produce a probability estimate for each class. The final result will be the combination of the weighted average of the probabilities returned by each classifier, considering that the weight of the final decision is estimated according to the performance of each classifier in the individual training (Faouzi, 2022).

For multivariate time series, each classifier produces a sub-classifier on every dimension. The combination of the predictions of each dimension will produce a single probability distribution, which will later be weighted with the performance of the remaining classifiers (Ruiz et al., 2021).

2.2.3. Deep Learning Approaches for MTSC

Deep Learning Neural Networks (DLNN) has recently been implemented in time series classification problems. This trend was motivated by the promising results of these models in a wide variety of classification problems, e.g., speech recognition, image classification, or Natural Language Processing (NLP) (Fawaz et al., 2019; Hsieh et al., 2021).

The DLNN models relied on the simple structures of neural networks to create highly complex versions constituted by multiple hidden layers of interconnected nodes.

The training starts with random initialization of the weights, and then the input flows along the network, passing from layer to layer. The next layer receives, as an input, the previous layer's output, and generates a new output through an activation function, also called a non-linearity function (e.g., Relu) (LeCun et al., 2015). At the end of the network, the last layer, also known as a descriptive layer, is applied to the last activation function (e.g., Softmax), which will distribute the estimated probabilities of an input belonging to each class. Based on the result of each training, the difference between the desired output and the predicted output is calculated by a loss function (e.g., categorical cross-entropy). With this estimation, the signal error flows from the last layer to the previous layer and updates the weights of the nodes by using an optimization function (e.g., Gradient Descent) and a learning rate. This process is called back-propagation and works iteratively to minimize loss and enable the active learning of the model.(LeCun et al., 2015; Vasilev et al., 2019).

Compared to the standard ML models, the most significant difficulty of DLNN models is their black box behavior. Derived from the number of parameters that interaction during the execution of the learning process makes, the tracking for relevant time series variables, or the time intervals, is unclear. However, deep learning models are very efficient when working with MTSC problems because of their ability to deal with the curse of dimensionality on a large scale (Ismail Fawaz et al., 2019; Ruiz et al., 2021). Moreover, it could find time patterns and sets of variables that can easily discriminate classes.

Different DLNN structures are being implemented. The ones with the most exciting results use Recurrent Neural Networks (RNN), Convolutional Neural Networks (CNN), and Multi-Layer Perceptron

(MLP). Some authors have published prototypes combining several of these structures, as is the case of Inception Time, an ensemble of CNNs developed by Karim et al. and Fawaz et al., proposed using the LSTM-FCN model that uses an LSTM (Long Short-Term Memory) layer and an FCN (Fully Convolutional Network) layer to extract features from the time series. The final layer is contemplated with a Softmax activation function to predict the class. More recently, Zhang et al. developed TapNet architecture for MTSC, with an even more complex architecture consisting of three key elements: the Random Dimension Permutation, a stage where a set of randomly-selected dimensions is created. The Multivariate Time Series Encoding, a stage with a set of LSTM, 1D convolutions, and FCNs, are designed to reduce the dimensional feature of the selected series. Furthermore, the last stage, Attentional Prototype Learning, evaluates if there is a similarity between the candidate representing the class and the remaining class elements through a distance-based loss function.

2.2.4. Long Short-Term Memory model

The RNNs are a variety of neural networks designed to work with sequential data due to their ability to retain information from past events.

While in Feed Forward Neural Networks (FFNN), the information flows in one direction only, the RNN contains an internal feedback loop (Le et al., 2019; Staudemeyer & Morris, 2019) corresponding to unfolding layers, over time, with equal weights. This particular structure allows each hidden node to learn based on its input and a second input from the hidden states' information. However, it is difficult for even a simpler RNN to capture long-term dependencies derived from the vanishing gradient problem. As the back-propagation proceeds, the error signal diffuses from layer to layer and from a hidden state to the previous ones. If there are many time steps, the gradient tends to diminish, making it impossible for the network to evolve with its learning in a reasonable amount of time (Staudemeyer & Morris, 2019).

To fill this gap, a new version of the RNN was redesigned by Hochreiter & Schmidhuber and improved by Gers et al. (Greff et al., 2017) as Long Short-Term Memory, as shown in Figure 2.

The LSTM, as RNN, contains memory blocks called cells (Le et al., 2019). Each cell in its structure has a mechanism similar to human memory due to its ability to retain information in two ways: long-term and short-term memory.

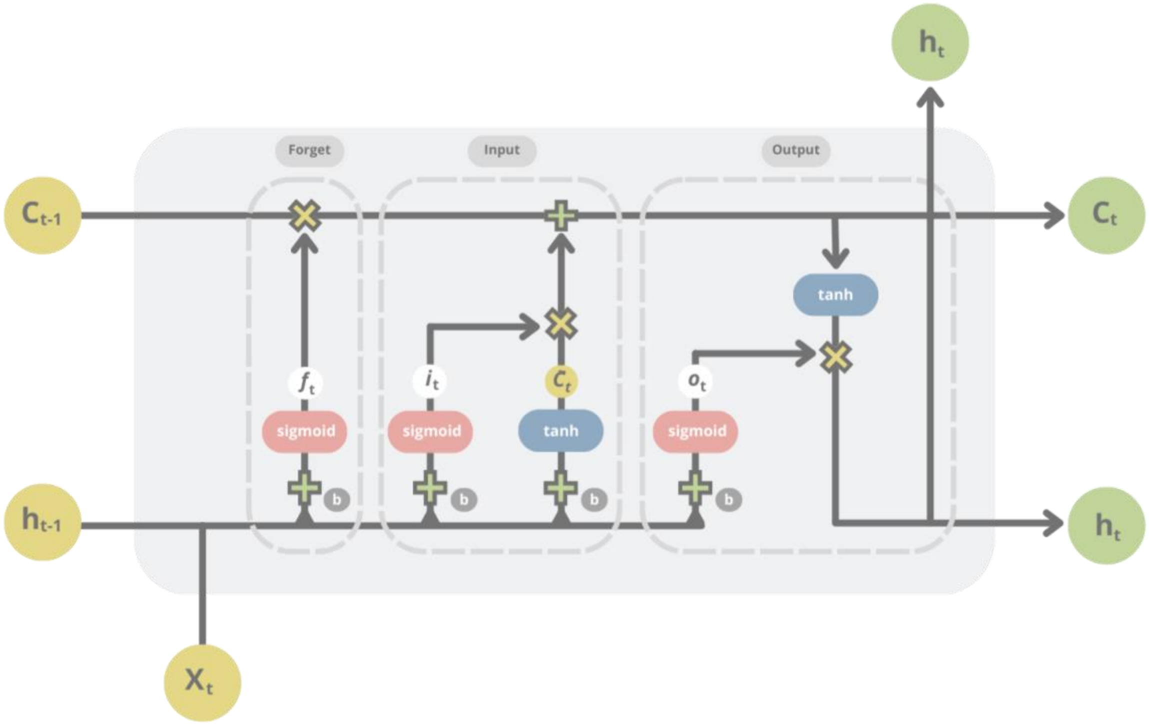


Figure 2: Structure of LSTM Cell: xt – current input; h_{t-1} - previous hidden state; h_t - current hidden state; C_{t-1} - previous cell state; C_t - current cell state; f_t - forget gate; i_t - input gate; o_t - output gate; $\tilde{c}$ - candidate vector; b - bias; + - addition; x - multiplication.

The cell state (C_t) represents the concept of long-term memory. The C_t is an essential mechanism that extracts features from said information to determine which time steps are important (high forget weights) and which are not (low forget weights) (Gupta et al., 2020). The following expression represents the cell state:

$$C_t = f_t \,.\, C_{t-1} + i_t . \tilde{C}$$

In this process, two stages aid in selecting relevant information: the forget stage and the input stage. The forget stage controls how much of the overall memory is kept. This stage starts with the forget gate, which uses the current input at time t (x_t) and the input from the previous hidden state (h_{t-1}), to pass this information through a sigmoid function that acts as a filter. The result is a vector that ranges between [0,1], which corresponds to the proportion of essential information to (1) retain or (0) forget (Liu et al., 2019; Staudemeyer & Morris, 2019).

The expression for forget gate is represented below:

$$f_t = \sigma\big(W_{xf}\, x_t + W_{hf}\, h_{t-1} + b_f \big)$$

The vector generated by f_t will later be multiplied by a vector from the previous cell state (C_{t-1}).
The input stage decides what new information will be stored in the cell state. In this stage, the input gate behaves similarly to the gate described above. However, to regulate the network, a candidate value is calculated (R. Zhang et al., 2020), containing a $tanh$ operator that will squish values between 1 and -1 to generate a vector $\tilde{c}$ which gives the level of importance from x_t and h_{t-1} (Le et al., 2019). If the value is negative, the information is deducted from the cell state. Otherwise, the information is added. The expressions for i_t and $\tilde{c}$ are respectively:

$$i_t = \sigma(W_{xi}\, x_t + W_{hi}\, h_{t-1} + b_i)$$

$$\tilde{c} = tanh(W_{x\tilde{c}}\, x_t + W_{h\tilde{c}}\, h_{t-1} + b_{\tilde{c}})$$

The values obtained in these two stages are added and incorporated into the cell state, containing strictly relevant information from x_t ,h_{t-1} and C_{t-1} .This will be scrutinized again in the next time-step.
The short-term memory concept is advocated by the current hidden state (h_t). At the stage of producing output, the output gate is used (which uses pretty much the same behavior as the other gates), and a vector is generated by applying a tanh function to the newly generated cell state (Le et al., 2019; Liu et al., 2019), which can be understood as an attenuated version of the C_t.
The expressions of o_t and h_t are respectively:

$$o_t = \sigma(W_{xo}\, x_t + W_{ho}\, h_{t-1} + b_o)$$

$$h_{t=}\, o_t . tanh(C_t)$$

The result of both outputs will be multiplied point-by-point. The information will then be passed to the next hidden stage to be fed back into the next LSTM cell or used to make a prediction.

3. RELATED WORK

The usage of machine learning models in the dairy industry is gradually revolutionizing the sector due to its ability to obtain automated answers without significant economic effort.

As is well known, the demand of the dairy sector increasingly forces it to contain costs while improving both production and animal health. This particular aspect has allowed advances in machine learning and deep learning techniques application. These techniques include disease diagnosis, animal behavior recognition, animal body condition score, cattle identification, cattle counting, and monitoring health parameters (Mahmud et al., 2021).

It must be borne in mind that the specific objective of this study is the construction of a model that supports the veterinarians' decision regarding the administration of antibiotics in the dry period as a way to prevent the occurrence of mastitis. No studies were found that fit these parameters exactly. However, concerning studies involving mastitis detection in dairy cows, 20 papers published between 2007 and 2022 were pre-selected through a google scholar search, with the keywords "Machine Learning + Mastitis" and "Deep learning + Mastitis".

The final selection of articles was based on criteria that matched the study carried out: the dataset, the type of problem, and the output.

Regarding the dataset used, all studies included milk recordings from automated milking systems. Considering that the meaning of the work in question is a classification problem, studies with binary output related to the presence or absence of clinical or subclinical mastitis were selected.

Ten papers were selected, seven of which used traditional machine learning techniques while three used deep learning, as shown in Table 1.

From the papers that used machine learning models, it is noteworthy that in the studies in which several models were compared, the algorithm that obtained the best performance was Random Forest, as was the case in the studies carried out by Motohashi et al., 2020, Bobbo et al., 2021, Fadul-Pacheco et al., 2021 and Ghafoor & Sitkowska, 2021. Papers that used deep learning techniques, such as those by Cavero et al., 2008 and Hajaya et al., 2019, employed multi-layer perceptron with three layers. While in Naqvi et al., 2022 the approach was less traditional and used LSTM in conjunction with synthetic data, adding different noise scenarios and outcome misclassification to produce an output.

Table 1: Summary of the ten related papers selected. From essays that used ML models, the models used were: CBA- Classification based on association rules; RF- Random Forest; XGB-XGBoost; DT-Decision Tree; KNN-K-Nearest-Neighbors; NB- Naive Bayes; NN- Neural Network; SVM- Support Vector Machine; LDA - Linear Discriminant Analysis; LR-Logistic Regression; GBT- Gradient Boost Tree; GMT-Generalized Linear Model; LGBM- Light Gradient Boosting Model. Papers that used deep learning models were - MLP- Multi-Layer Perceptron; LSTM -Long Short-Term Memory. For the best model, the metrics used to infer the model performer were: ACC- accuracy; SEPC-Specificity; PREC- Precision; REC-Recall.

Authors	Field	Dataset Size	Models	Best Model
Ebrahimie et al., 2021	ML	150.000	CBA	CBA ACC 83.8%
Ghafoor & Sitkowska, 2021	ML	6.600	26 models (e.g., RF, XGB, DT, KNN, AdaBoost, SVM, Perceptron, LGBM, NB)	RF ACC 99.2%
Bobbo et al., 2021	ML	18.442	8 models (LDA, GLM, NB, CART, KNN, SVM, RF, NN)	LDA, RF ACC 79.7%
Motohashi et al., 2020	ML	11.935	2 models (SVM, RF)	RF SEPC 81% PREC 46%
Fadul-Pacheco et al., 2021	ML	NA	3 models (NB, RF, EGB)	RF ACC 85%
Ebrahimi et al., 2019	ML	364.249	6 models (NB, NN, RF, GLM, GBT, LR)	GBT ACC 84.9%
Mammadova & Keskin, 2013	ML	170	SVM	SVM SPEC 92% REC 89%
Naqvi et al., 2022	DL	4.600	LSTM	LSTM SPEC 90% REC 86%
Hajaya et al., 2019	DL	29.000	MLP	MLP SPEC 99% REC 97%
Cavero et al., 2008	DL	403.537	4 models MLP	MLP SPEC 78.3 %

As mentioned, the features used in the various papers come primarily from dairy records generated by the automatic milk system. However, some authors chose to include information regarding the animal (e.g., temperature, number of lactations, breed) or the herd (e.g., environmental temperature, number of cows per unit).

These features typically undergo a transformation process that could be normalized by placing them on a logarithmic scale or removing the outliers, for better model interpretation. Figure 3 presents the most common features used in the reviewed papers, where milk yield and electrical conductivity are the most used for input.

In order to create the output for binary mastitis, most papers chose to use the somatic cell count (SCC) as a threshold (e.g., > 200.000 SCC – Mastitis). Nevertheless, other authors chose to use other parameters, such as the measurement of LDH (biomarker lactate dehydrogenase) or adapted epidemiological formulas.

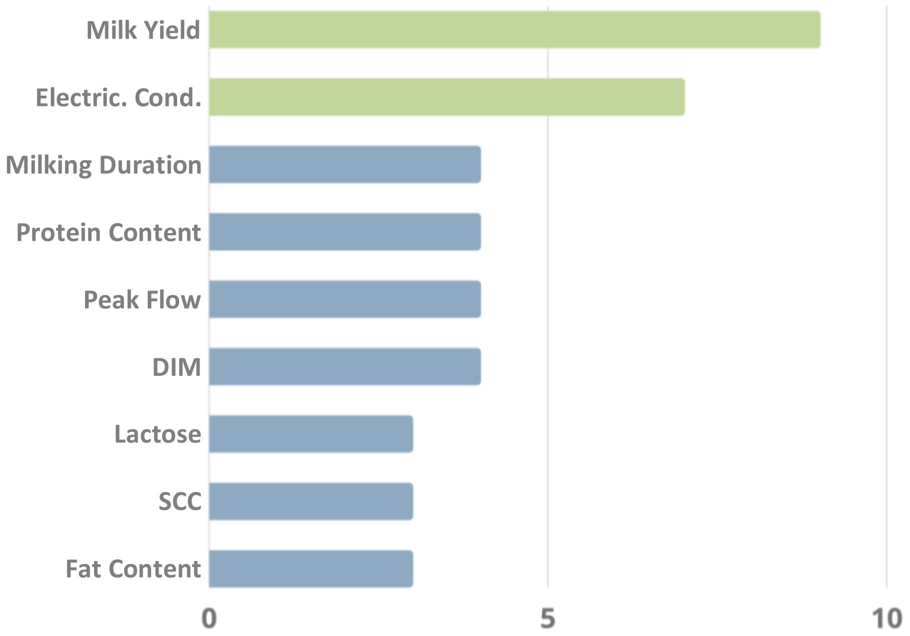


Figure 3: The most used features in selected essays. Highlighted in green, the Milk Yield was used in 9/10 papers, and the Electrical Conductivity in 7/10 papers. The other features are Milking Duration, Protein Content, Peak Flow, DIM - Days in Milk (appears in 4/10 papers). Lactose, SCC – Somatic Cells Count and Fat Content (appears in 3/10 papers).

Although these authors have not used milk records in mastitis detection, these approaches are still quite exciting and should be mentioned. The paper published by Xudong et al., 2020 used a CNN MobileNet and Yolo version to identify mastitis through thermal images of the eyes and udders of dairy cows. Another paper by (Dhoble et al., 2019) used Cytometric fingerprints to collect cellular parameters of milk samples to detect not only mastitis but also the type of pathogen and the cow lactation stage through standard ML models.

4. METHODOLOGY

This chapter aims to explain the model development process for this study. Given its specificity, the following methodology is proposed, inspired by the CRISP-DM (Cross Industry Standard Process for Data Mining) method and divided into five stages, as represented schematically in Figure 4.

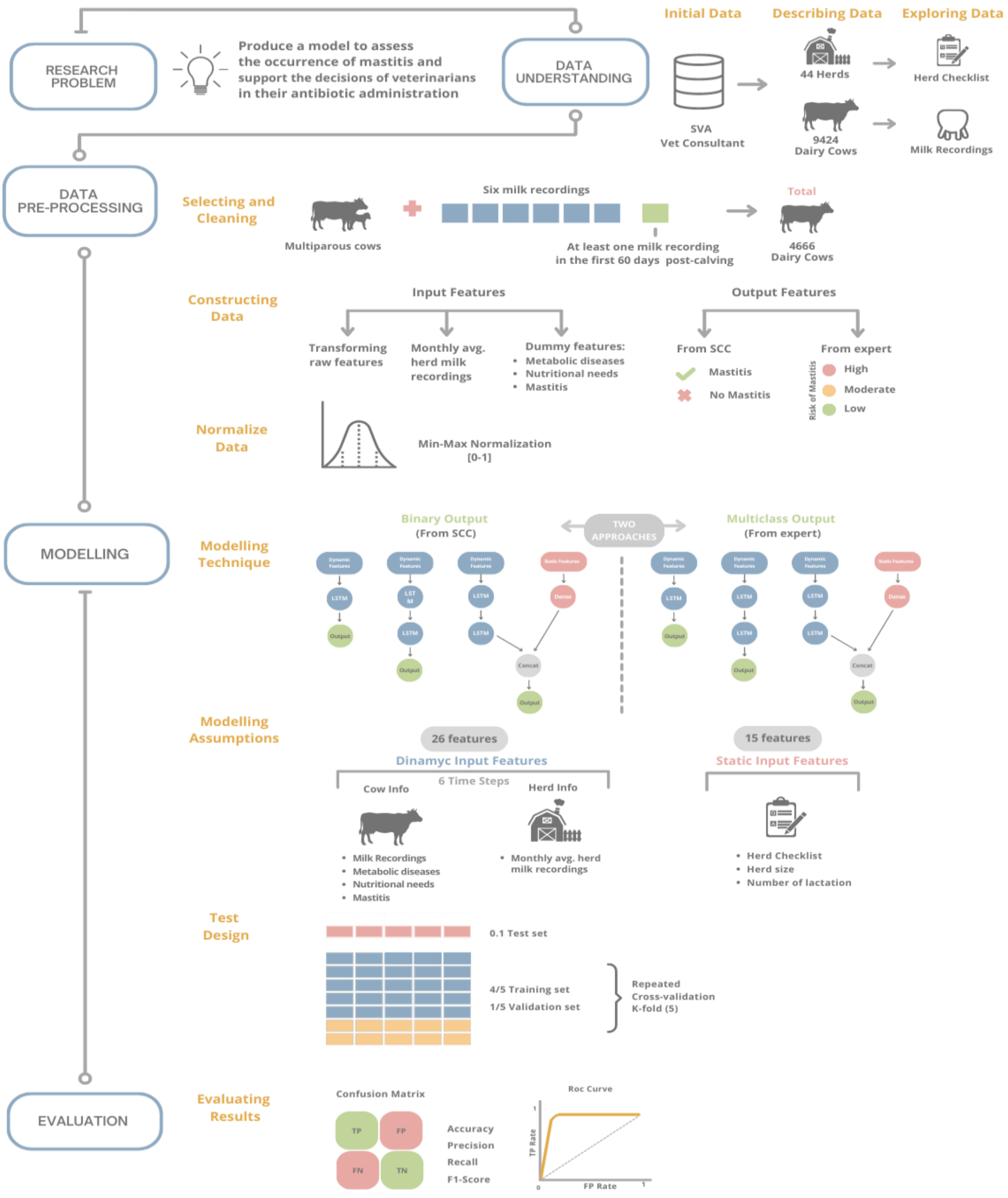


Figure 4: Workflow diagram of model development process. The process is divided into five steps: Research Problem, Data Understanding, Data Pre-Processing, Modelling, and Evaluation.

The first stage is Research Problem, which explains the motivations, the study objectives, and the plan to be executed. The following stage is Data Understanding, in which the initial data is accessed, and the features from the original datasets are explored. Data Pre-Processing is probably the most exhaustive stage of the study. Allowing to improve the model results, this stage, is dedicated to selecting, cleaning, standardizing, transforming, feature extraction, and data integration. During the next stage, Modelling, the model selection takes place, followed by the Evaluation stage, where the metrics are implemented to evaluate the model's performance. The results obtained from this stage are susceptible to being interpreted and confronted with previous results to review the entire process and, thus, enhance its performance. This last stage will be discussed later in the Results and Discussion chapter.

The model development process is iterative, and it is necessary to have a sense of judgment regarding the model's behavior to optimize and retain the best performance possible.

Throughout the development of this study, Python 3.9 was used as a programming tool in the IDE Jupyter Notebook. The most used libraries were Numpy, Pandas, Seaborn, and Matplotlib, used mainly in the data understanding and pre-processing stages, and Scikit-learn and Keras, used for the model production and Evaluation stages.

In the following subsections, each step of the methodology is explained in detail.

4.1. RESEARCH PROBLEM

As explained in the literature review, the new directives regarding the controlled use of antibiotics in animals, which came at the beginning 2022, have accelerated the necessity to develop tools that support the decisions of veterinarians in their AB administration.

Considering this challenge, this study aims to produce a model that helps veterinarians make more conscious decisions to mitigate the indiscriminate use of antibiotics on dairy herds.

4.2. DATA UNDERSTANDING

4.2.1. Initial Data

The primary data source was requested from consulting company SVA (Veterinary Services and Associates), which follows and monitors herds in mainland Portugal and the Azores. They granted access to milk records through the Bovinfor online platform and to the checklists of herd hygiene conditions resulting from the audits carried out by veterinarians.

These data will serve as input to LSTM models, which will interpret it sequentially, referring to the last six milk records (one for each month) before the cow enters the dry period.

The data collected refer to the period from January 2018, to December 2019. In this period, 44 herds located in mainland Portugal (22 herds) and Azores (22 herds) were covered, a total of 9.424 cows.

4.2.2. Describing Data

Describing the data sources used, the Bovinfor is an online platform where the monthly milk recordings for each herd are recorded. Each monthly collection contains the results of milk samples collected from each animal. The register includes the following features: Snirb (identification number of the dairy cow), number of lactations, milk yield (liters), fat content (%), protein content (%), urea (mg/Kg), SCC ($x10^3$), calving date, and the date on which the milk sample was collected.

The checklists carried out on the herds determine the quality of hygiene. Features that are related to the presence of mastitis were selected heuristically, which are: are the presence of contagious agents, the occurrence of drying therapy, type of drying mode, use of internal sealant, presence of hyperkeratosis in more than 20% of the herd, presence of udder inflammation in more than 20 % of the herd, dry cow separation, the type of bedding used, litter hygiene, mastitis vaccination and if the mastitis control program has been implemented.

4.2.3. Exploring Data

In order to understand the distribution of features in the 44 selected herds, data without previous processing was used.

On average, the herds comprise 216 cows each, with a standard deviation of 213 cows. As seen in the histogram of Figure 5, these herds are mostly small or medium.

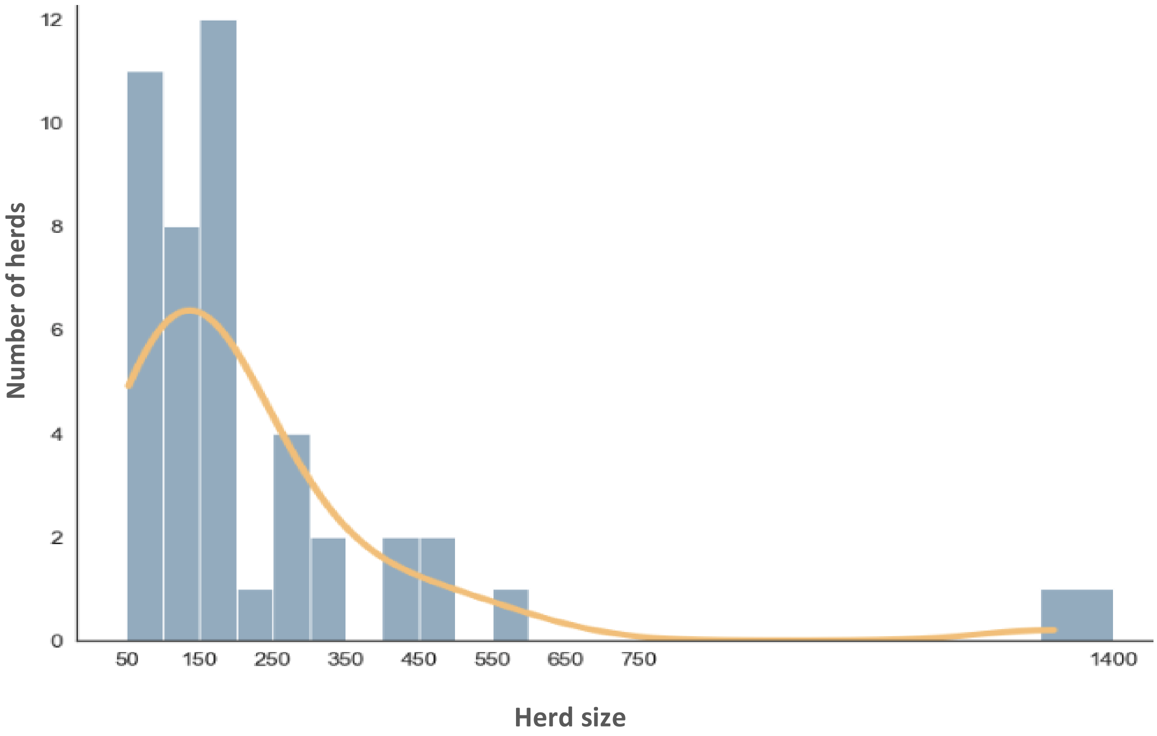


Figure 5: Histogram representing the number of herds by size. The smallest herd contains 53 animals, while the largest has 1319. The quartiles are as follows Q1: 101; Q2: 152; Q3: 258.

Figure 6 presents the boxplots of features from the milk recordings: Number of Lactations, Fat Content, Protein Content, Urea and SCC. The points represent the average values from each herd.

In an overview of the study population, it is notable that most lactating cows are in their second lactation, the protein content does not show significant variation, and the SCC feature has an average and median of around 250.000 (with some variation to these values). It should be stressed that this data still needs to be subject to selection and filtering, these values being merely indicative.

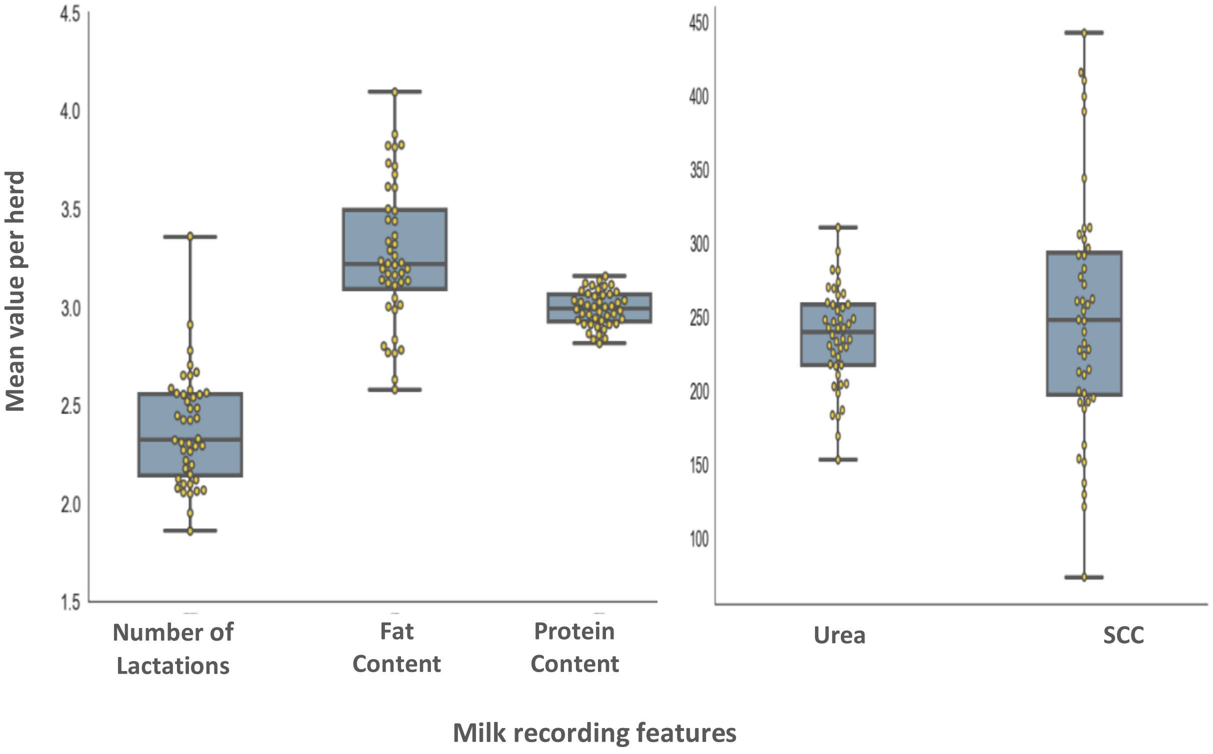


Figure 6: Boxplots representing the distribution of milk recording features. Number of lactations; Fat Content (%); Protein Content (%); Urea (mg/kg) and SCC (x10^3)

From the checklist, the surveys carried out for each herd were collected. Figures 7 and 8 present bar graphs with each question's positive and negative answers. In the graph of Figure 7, it is possible to observe that most herds opt for separating dry cows from lactating and that using internal sealants is also common. Conversely, it appears not a common practice to proceed with gradual drying since practically all opt for the abrupt practice.

Regarding the type of bedding used, in the graph of Figure 8, the vast majority is in the pasture. This aspect occurs because half of the herds are placed in the Azores, where it is common for dairy cows to live in extensive breeding.

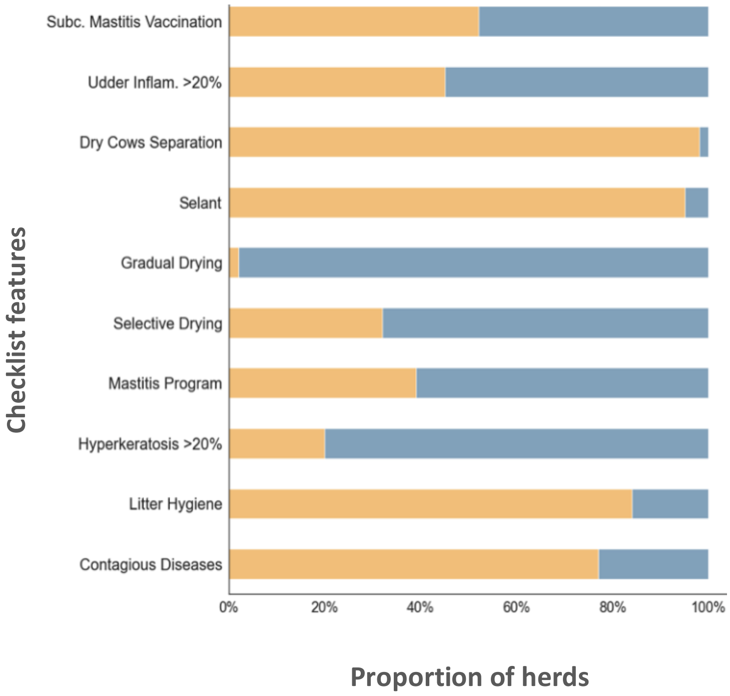


Figure 7: Graph bars indicating the proportion of herds per checklist feature. Orange for positive answers and blue for negative.

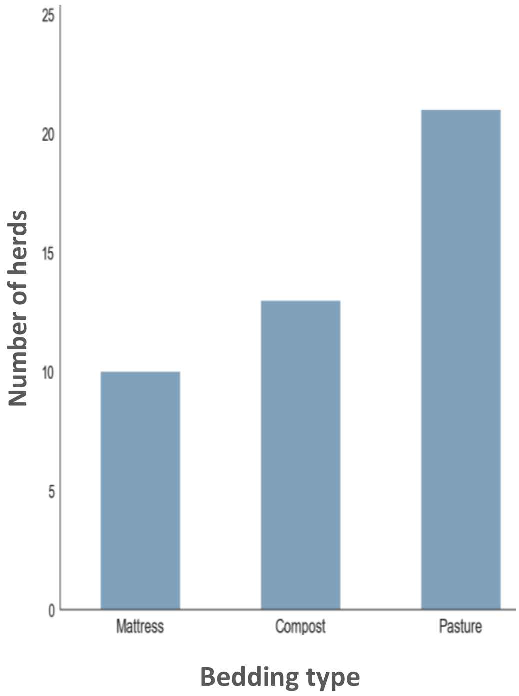


Figure 8: Graph bars demonstrating the number of herds per bedding type.

4.3. DATA PRE-PROCESSING

The first steps to compose data for model ingestion are cleaning and selecting data based on specific criteria defined for the LSTM model, namely using sequential data.

The following stage is constructing the input data. This process was begun by reformulating some features and later creating new features that better identify the herd's general panorama and understand the dairy cows' health information.

The following step is to prepare the output features, which will be based on two separate outputs, the first coming from the SCC and the other from an expert veterinarian.

This data preparation stage is concluded with the normalization of input data performed by the min-max scaler.

4.3.1. Selecting and Cleaning

Selecting a number of dairy cows that were compatible with the study, it was necessary to validate the following assumptions:

- All cows included in this study are multiparous, which indicates that they have had at least one calving.
- In terms of selection, the cows' last six milk recordings before entering the dry period, and at least one milk recording in the first 60 days post-calving, will be included.

Figure 9 exemplifies the feature behaviors in a set of milk recordings from a given dairy cow throughout 2018 and 2019. The graph illustrates a period window that includes the six selected milk recordings, the dry period window, and the window comprising the first 60 DIM (which will be used for models with binary output).

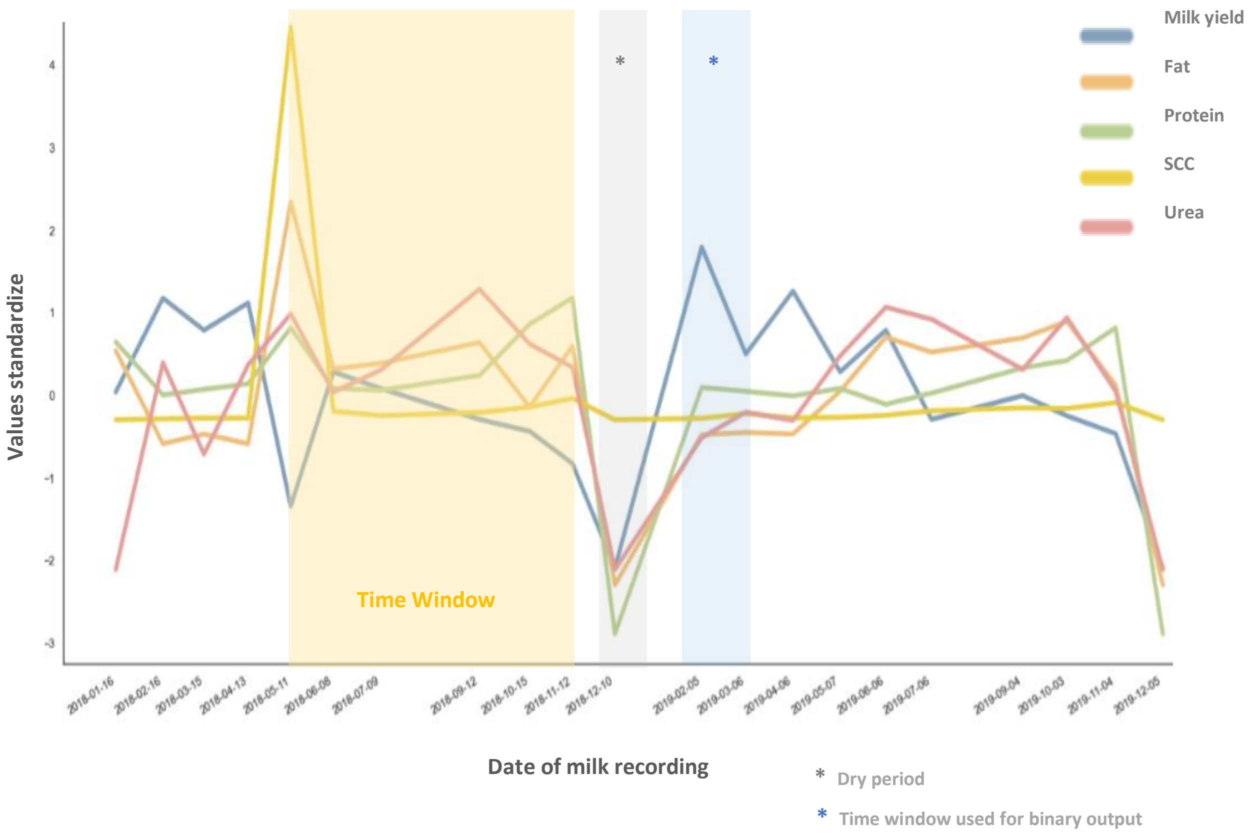


Figure 9: Example of changes in the milk recording features of a dairy cow after taking their logarithms, from 2018 to 2019.

Of the 9424 cows, 4666 were selected for the study. The main eliminatory factor was that they did not have a dry period within the time window from 2018 to 2019.

4.3.2. Constructing Data

Constructing the data for learning the models is a somewhat complex process, requiring some input features to be reformulated or created from scratch. This was the case with some of the milk recording features, the dummy features that reflect the risk of metabolic diseases, nutritional needs, or mastitis, as well as the aggregated milk recording features that report the general condition of the herds.
Considering that different approaches were used for the same study, the SCC variable will be used for the binary classification output, and the output for multiclassification will be based on the score provided by a veterinarian expert.

4.3.2.1. New Input Features

a) Transforming raw features

Regarding the DIM (Days in Milk), one of the most used features in previous papers, which mirrors the lactation stage of the cow, this feature was included in this study.

- DIM - The difference between the milk recording date and the calving date

b) Monthly average values per herd

Milk record values were added per herd, and the monthly average was obtained (in order not to compromise the results, all 0 values were transformed into NaN values).
The new features served to project the general condition of the herd each month. They are the following:

- Avg. number of lactations
- Avg. milk yield
- Avg. fat content
- Avg. protein content
- Avg. urea
- Avg. SCC

c) Metabolic diseases, nutritional needs, and mastitis dummy features

The method used to create dummy features that reproduce the presence or absence of metabolic problems, mastitis, or nutritional needs was made from milk recordings features.

The reason for including dummy features based on metabolic diseases is that they are directly linked to the occurrence of mastitis. For example, when the fat/protein ratio is below 1, it is assumed that there is a risk that the cow could develop Sub-Acute Ruminal Acidosis (SARA). This occurrence may affect reproductive performance and cause mineral deficiency. On the other hand, if the fat/protein ratio is above 1.5, it indicates a risk of subclinical ketosis that may be associated with nutritional diseases such as metritis, mastitis, milk fever, lameness, and displaced abomasum, which are commonly observed after calving (Cabezas-Garcia et al., 2021).

Typically, in milk, the percentage of protein follows the variations of fat. When this assumption does not occur, two types of changes could take place. The first, Milk Fat Depression (MFD), is a phenomenon consisting of a reduction of milk fat content as expected, and it is usually associated with subtle changes in the cow's rumen (Koch & Lascano, 2018; Stokes et al., 2009). The second, Milk Protein Depression (MPD), may be attributed to a limitation of amino acids for protein synthesis derived from mammary gland issues (Casper & Schingoethe, 1989; Stokes et al., 2009).

Two continuous variables were also included, the values from Protein/Fat and Fat/Protein ratios, to better observe these changes.

Dummy Features

- MFD - Protein /Fat ratio > 0.95
- MPD - Protein /Fat ratio < 0.8
- Subclinical Ketosis - Fat / Protein ratio >= 1.5
- SARA - Fat/Protein ratio < 1
- Subclinical Mastitis - SCC > 200.000

Continue Features

- Protein /Fat ratio
- Fat / Protein ratio

The relationship between urea and protein is based on conditions, as represented in Table 2. These reflect some nutritional deficiencies that may be occurring, namely, if the protein or carbohydrate intake is being done correctly. These nutritional needs were added to the study for the reason they are coadjutants of metabolic diseases and mastitis (Miranda, 2014).

Table 2: Features describing the carbohydrate and protein needs through the relationship with protein content and urea. Adapted from (Miranda, 2014).

Features expressed protein and carbohydrates needs	Conditions	
	Urea (mg/kg)	Protein (%)
• Prot ↓ and CH ↓	< 256.0	< 3.0
• Prot ↓ and CH ↑	< 256.0	> 3.2
• Prot ↓	< 256.0	3 - 3.2
• Prot ↑ and CH ↓	> 342.0	<=3.2
• Prot ↑	> 342.0	> 3.2
• Prot ↑ and CH ↑	256.0 - 342.0	> 3.2
• CH↓	256.0 - 342.0	< 3.0

4.3.2.2. The Output Features

As mentioned before, two types of outputs were used in this study. The first output is designed for binary classification models based on a threshold, and the second is designed to be used in multi-classification models based on three classes. The section below will better explain the process of obtaining these outputs.

a) From the SCC

The output was produced based on the presence or absence of subclinical mastitis in the first 60 days post-calving. This period may include one or two milk recordings. In cases where the SCC exceeds the threshold of 200.000 during this period, the output is 1; otherwise, it is 0.

b) From the expert

Since it would be humanly impossible to produce a case-by-case judgment, it was agreed to produce an output based on a score that accounts for metabolic diseases, nutritional needs, and the SCC animal and herd values from the last six milk recordings before the calving. The particularity of this score is that it applies a weight of importance depending on if the milk recording is closer to the dry period, as shown in Table 3.

Table 3: Score provided by a veterinary expert. Dummy features included: Subclinical Ketosis, SARA, MFD, MPD, and all of the features from nutritional needs. Continuous features SCC and Avg SCC herd. 1M corresponds to the closest month before the dry period, and 6M is the farthest.

Milk Record	SCC		Sub. Ketosis /SARA	MFD/ MPD	Urea and Protein	Avg SCC herd		Total score
1M	> 400 200 - 400	10 2	1	0.5	0.25	> 400 200 - 400	10 2	22.25
2M	> 400 200 - 400	2 1	0.5	0.25	0.125	> 400 200 - 400	2 1	5.125
3M	> 400 200 - 400	2 1	0.5	0.25	0.125	> 400 200 - 400	2 1	5.125
4M	> 400 200 - 400	1 0.5	0.25	0.125	0.0625	> 400 200 - 400	1 0.5	2.5625
5M	> 400 200 - 400	1 0.5	0.25	0.125	0.0625	> 400 200 - 400	1 0.5	2.5625
6M	> 400 200 - 400	1 0.5	0.25	0.125	0.0625	> 400 200 - 400	1 0.5	2.5625
Total score	17		2.75	1.375	0.6875	17		**40.18**

From the score distribution, the outputs were cataloged. These were delimited in:

- **High risk of mastitis** if the value is equal to or greater than 10.
- **Moderate risk of mastitis** if the value is between 7 and 10.
- **Low risk of mastitis** if the value is equal to or less than 7.

4.3.3. Normalize Data

All input features were normalized with a min-max scaler estimator from the Scikit-learn library that scales and translates each value to the range 0 and 1, meaning the minimum and maximum value of a feature, respectively. The formula is given below:

$$x_{scaled} = \frac{x - min\ (x)}{max\ (x) - min\ (x)}$$

4.4. MODELLING

4.4.1. Selecting Modelling Technique

LSTM models applied to classification problems are designed to deal with sequential data that reproduce the variation of features as they are returned in time. The incorporation of static features in RNN models causes duplication of information at each time step, impairs the model's performance, and implies excessive memory and time consumption (Miebs et al., 2020).

The first study that relates dynamic and static features to predict clinical events based on patient history was developed by Esteban et al. Later, other authors, such as Hyland et al., Lin et al. and Yang et al., used and improved this method to predict diseases or for making medical decisions. Inspired by previous work, in this study, two types of methods were tested (Leontjeva & Kuzovkin, 2016). The first is the unimodal method, which uses only the dynamic features as input. In this case, two models with this configuration were used: a baseline model Vanilla LSTM (fig. 10A) and a two-layer Stacked LSTM (fig. 10B).

In the bimodal method, dynamic and static features were used. For this composition, a Stacked LSTM was employed for the dynamic features and a dense layer for the static features, fig. 10C.

The structure of the models presented are many-to-one models since they are injected with input features, and the result is precisely one output.

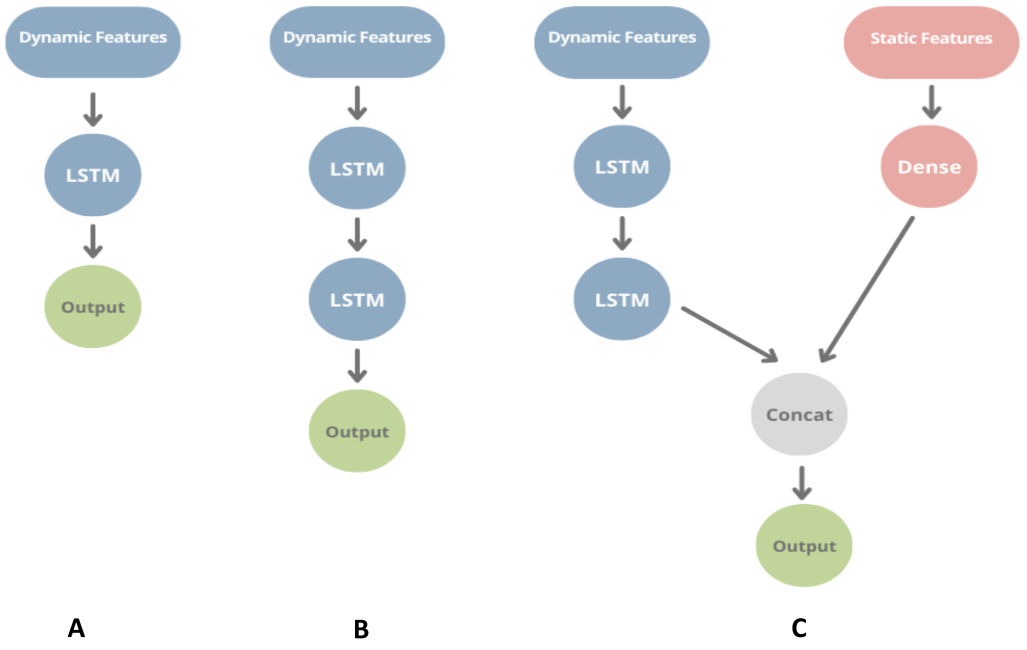


Figure 10: LSTM models used in the study. A- Vanilla LSTM with dynamic features; B- Stacked LSTM model with dynamic features; C- Stacked LSTM and Dense layer with dynamic and static features

As represented in Figures 10B and 10C, Stacked LSTMs are used. One or more layers are added with the same configuration, allowing the model to expand the capacity to capture more complex sequences.

The relation of the output/input connections is the only relation between two adjacent layers, i.e., on the depth dimension level, hidden state t on level L-1 will be fed as the input into the following hidden state on level L with the same time t, as represented in Figure 11. On the time dimension level, the behavior is the default, and it transfers both the cell and hidden states along the time dimension, respecting the temporal order. Moreover, this approach makes it possible to better manage the relationship to the great dimensionality of the times series during the learning process, enhancing the power of neural networks.

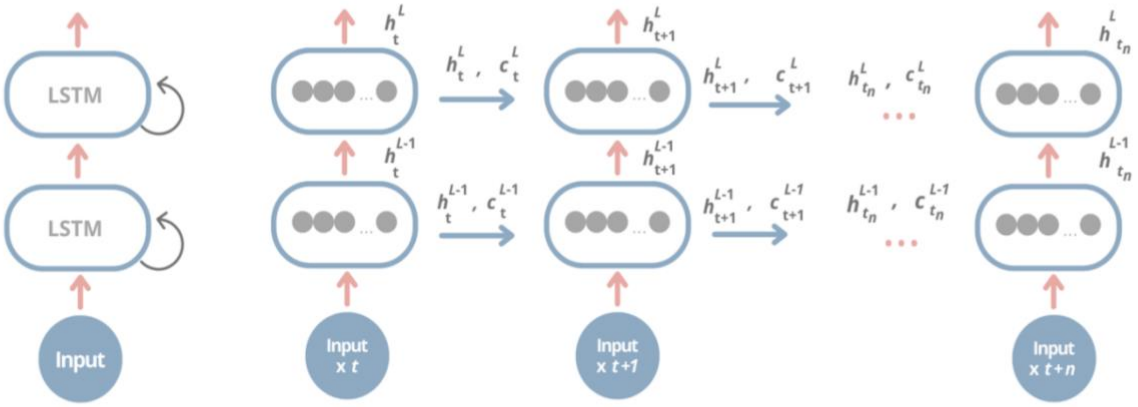


Figure 11: Representation of unrolled stacked LSTM model with two layers. *h* - hidden state; *c* – cell state; *L* – depth level, t-time dimension level; *x(*t) – input at time t.

As model 10C shows, in addition to the stacked LSTM that receives dynamic features as input, a dense layer composed of fully connected neurons is used in parallel, which exclusively receives the input vector of N static features. The vector coming from the last time step of stacked LSTM and the vector from the dense layer are joined through concatenation. The resulting vector is passed through a dense layer and finally passed into an activation function to make the final prediction.

4.4.2. Modelling Assumptions

In models optimized to perform with sequential data, the data must have pre-defined particularities. Thus, input data has been vectorized and, according to behavior over time, assigned as dynamic or static features, as shown in Figure 12.

Represented in tabular form, the variables from different time steps were collected and aggregated into a single 6-dimensional vector. A total of 26 dynamic features represented the information related to dairy cows and herds. In the dairy cow information, the features are subdivided into continuous features, such as milk recordings (6 features), and dummy features, such as metabolic diseases (5 features), nutritional needs (7 features), and mastitis (1 feature). Regarding the information related to the herds, continuous features are related to the average monthly value of milk recordings in each herd (5 features).

In the static features that have been converted into a one-dimensional input vector, a total of 15 features are represented. The continuous features (2 features) refer to the number of dairy cow lactations and the herd size, and the dummy features (13 features) include features from the herd checklist related to hygiene behaviors to prevent mastitis.

A comprehensive explanation of the dynamic and static features is summarized in the appendix.

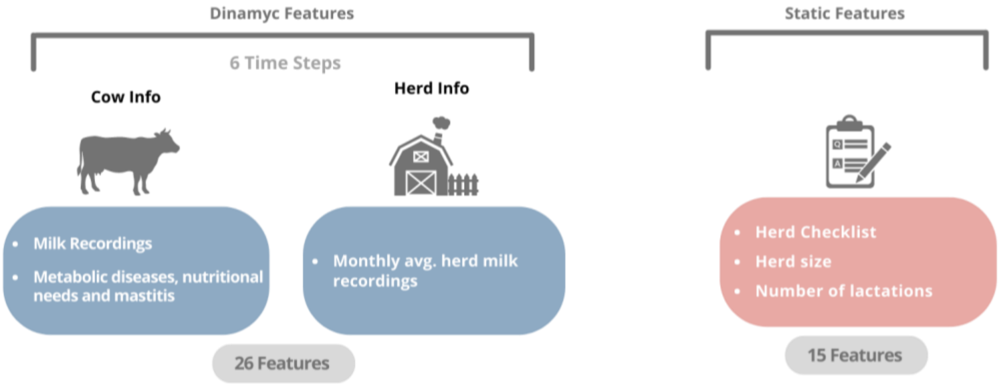


Figure 12: Illustration of dynamic and static features

Considering the reasonable size of the dataset (4666 dairy cows) and the reduced number of features, this study only used the qualities inherent in the LSTM mechanisms to capture long and short-term memory, which act as feature selection by capturing the attributes that stand out over time from dynamic data.

4.4.3. Test Design

From selected data, 10% was reserved for testing using the GroupShuffleSplit function and random state 21. This resampling technique aims to randomly divide a portion of the data for the training model and another for testing. At the same time, it adds the condition of keeping a third-party group aggregated, i.e., it preserves the instances with information on the milk contrasts allocated to the ID of each animal.

With the remaining data for training the model, the Repeated Cross Validation function was employed, which uses traditional cross-validation repeated several times. Cross-validation aims to divide the data into non-overlapping folds of similar size to allow all folds to be equally utilized for training and validation, providing realistic results(Berrar, 2019).

In this study, a fold of five (four for training and one for validation) and a random state of 21 were used for faithful replication of the results.

This procedure was repeated 20 times in order to bring the model to its exhaustion and validate the results. A total of 100 versions of the same model were produced.

4.4.4. Building the Models

Building deep learning models from scratch is necessary to observe the type of data, the output, and the particularities of each algorithm, to consider the parameters that best fit each model.

The GridSearchCv function was performed, allowing the models' best parameters to be obtained. This process was time-consuming due to the number of combinations employed and cross-validation with five folds.

This method was utilized in both intermediate models (Stacked LSTM) of the different approaches as guides for the hyperparameters to be used in the remaining models. The number of nodes tested in the LSTM layers (32, 64, 128), the batch size (15, 30, 50), the value of the regularizers (0.01, 0.001), the dropout value (0.1, 0.2, 0.3), the optimizer (Adam, Rmsprop) and weight initialization type (normal, glorot uniform).

From the parameters obtained, and after analyzing the models' behavior individually, the number of 200 epochs and batch size of 15 were defined as parameters shared among all models.

Table 4 shows the structures of the binary and multiclassification models where all the parameters mentioned are included and the number of nodes established.

Table 4: Structures of the models for the binary and multiclass approaches.

Vanilla LSTM	Stacked LSTM	Stacked LSTM + Dense
Binary Classification		
LSTM layer (64) Activation func: Tahn Output layer: Dense layer (2) Activation func: Sigmoid Loss func: BCE Optimizer: Adam	1º layer: LSTM (128) Activation func: Tahn Dropout (0.1) 2º layer: LSTM (64) Activation func: Tahn Dropout (0.1) Output layer: Dense (2) Activation func: Sigmoid Loss func: BCE Optimizer: Adam L2 Regularization: 0.01	1º layer: LSTM (128) Activation func.: Than Dropout (0.1) 2º layer: LSTM (64) Activation func: Tahn Dropout (0.1) Static layer: Dense (64) Activation func.: Relu Concat layer: Dense (32) Output layer: Dense (2) Activation func: Sigmoid Loss func: BCE Optimizer: Adam L2 Regularization: 0.01
Multi Classification		
LSTM layer (64) Activation func: Tahn Output layer: Dense layer (2) Activation func: SoftMax Loss func: CCE Optimizer: RmsProp L2 Regularization: 0.01	1º layer: LSTM (128) Activation func: Tahn Dropout (0.1) 2º layer: LSTM (64) Activation func: Tahn Dropout (0.1) Output layer: Dense (3) Activation func: SoftMax Loss func: CCE Optimizer: RmsProp L2 Regularization: 0.01	1º layer: LSTM (128) Activation func: Tahn Dropout (0.1) 2º layer: LSTM (64) Activation func: Tahn Dropout (0.1) Static layer: Dense (64) Activation func: Relu Concat layer: Dense (32) Output layer: Dense (3) Activation func: SoftMax Loss func: CCE L2 Regularization: 0.01 Optimizer: RmsProp

Choosing the proper parameters for the model is a daunting task, especially for deep learning models, which generally consist of thousands of combinations. It is necessary to understand what suitable parameters are needed for the given task before experimenting with all of them (Balaji et al., 2021).
In the models with binary and multi-class classification, it was necessary to make some adaptations for them to function correctly. One of the adaptations to be highlighted is the loss function, which is responsible for adjusting the weights through the calculation of the penalty score, based on the distance between the probability of belonging to a given label and the target label. In the case of binary classifiers, Binary Cross Entropy (BCE) was used, while for multi-class classification models Categorical Cross Entropy (CCE) was employed, in which true labels are one-hot encoded.

The activation function also plays a significant role in the decision of whether a given node should be activated by its relevance in the predicting process. Each activation function must be adapted to the type of layer. For example, in LSTM layers, because of their internal mechanisms to memorize or forget features, the Tahn function is often used for its advantage of squeezing action, producing values that vary between [-1, 1]. For the Static dense layers, the Relu function was employed for its linear rectified form, obtaining sparsity representations and avoiding easy saturation (Glorot et al., 2011). A Sigmoid function was used in the binary output layers. This function naturally coexists with BCE due to its need to compute the logarithm of the predicted output, which must have a range between 0 and 1. Meanwhile, the Softmax function coexists with the CCE that calculates the probabilities distribution of each target class over all possible target classes (Ho & Wookey, 2020).

Other parameters used in both approaches were the Optimizer function and Regularizers, crucial to improving the performance of the models.
Optimizer functions can modify attributes such as weights and learning rate, helping to reduce loss and consequently improving learning. In this study, two Optimization functions were used, RmsProp and Adam. Both functions are adaptive, ensuring the convergence of the minimum error function for a well-chosen learning rate. The Rmsprop optimization is based on keeping an exponential moving average of the squared gradients, dividing by the root of this average for each weight. Adam is similar to Rmsprop. It calculates the exponential moving average of the squared gradient but also uses the gradients' second momentum, consisting of the parameters beta1 and beta2 that control the decay rates of these moving averages (Kingma & Ba, 2014).
Regularizers are other components integrated into NNs that perform slight modifications to the learning algorithm such that the model generalizes better and prevents overfitting.
The main idea of the dropout technique is that nodes are randomly dropped out in each epoch (Labach et al., 2019). This approach intends to artificially corrupt the structure of the NN as a way of producing

different architectures, allowing the model to capture new patterns and, thus, generalize better (Srivastava et al., 2014).

Another regularizer used is the L2 parameter, composed of a Ridge Regression function, which penalizes the loss function by adding the “squared magnitude” of the weights coefficient as a regularization term. This adds a more significant penalty as the model complexity increases, forcing the weights to be very small without reaching 0, avoiding sparse results.

This especially applies to LSTM models, which have complex mechanisms, are exposed to high dimensional input, or, because they need a greater volume of training data to perform satisfactorily, are prone to overfitting. For this reason, regularizers are essential for the proper functioning of the models (Balaji et al., 2021).

4.4.4.1. Resampling Technique

In binary output models, the classes presented were initially unbalanced, which translates into having a majority class to the detriment of the other. The model’s performance from unbalanced data will produce imposter results, inducing better accuracy due to a very discrepant recall and precision between classes. In other words, during the training phase, the model tends to ignore the minority class(Swana et al., 2022), identifying it as noise and consecutively predicting the ruling class.

The dataset was distributed as 70% with “non-mastitis” output vs. 30% where the output is “mastitis.”

In order to solve this problem, class equalizing techniques were used. The most common approaches to resampling techniques manage to adjust the number of instances between classes through Undersampling (balances the classes by removing instances of the majority class) or Oversampling (balances the class by producing synthetic instances of the minority class) (Zeng et al., 2016).

SMOTE Tomek, a hybrid technique, was used to combine both resampling techniques. The main advantage is conciliating the properties from classic SMOTE, the Oversampling technique that generates random synthetic samples based on a k-nearest neighbor for the minority class, with the Undersampling technique Tomek Links, which removes the overlapping data to increase the class separation near the decision boundaries (Sasada et al., 2020; Swana et al., 2022).

4.5. EVALUATION

In machine learning problems, the Evaluation stage is used to assess the ability of training models to generalize. This iterative learning process aims to optimize the model to obtain the best possible result.

For classification problems, predictions generated at the end of model training are compared with actual predictions. These results are listed in a confusion matrix which is an N x N matrix, where N is the number of classes being predicted.

Each row of the matrix represents the predicted class, while the column represents the actual class, i.e., TP (True Positives) and TN (True Negatives) denote the number of positive and negative samples that are correctly classified (Hossin & Sulaiman, 2015). Meanwhile, FP (False Positive) and FN (False Negative) represent the number of misclassified negative and positive samples, respectively. These four outcomes are the basis for performing evaluation metrics used in this study, as shown in Table 5.

Table 5: Metrics applied to assess the models performance.

Metric	Formula
Accuracy	$\frac{TP + TN}{TP + FN + TN + FP}$
Recall / True Positive Rate	$\frac{TP}{TP + FN}$
Precision	$\frac{TP}{TP + FP}$
F1-Score	$2 * \frac{Precision * Recall}{Precision + Recall}$
False Positive Rate	$\frac{FP}{FP + TN}$

Accuracy is the most used evaluation metric in practice, either for binary or multi-classification problems. Due to its simplicity in understanding, this metric represents the proportion of the total number of correct predictions. Other metrics that allow checking whether accuracy is balanced (Grandini et al., 2020) are Recall and Precision .

Recall reflects the model's ability to detect positive samples, and is used to measure the ratio of truly positive cases that are correctly classified. Precision mirrors how reliable the model is in classifying samples as positive, measuring the ratio of positive samples that are correctly identified.

F1-Score is often used as a single value that provides high-level information about each model's output quality. This metric combines Precision and Recall using the harmonic mean, which gives equal weight to both metrics.

The AUC (Area Under the Curve) is considered one of the robust metrics for classification problems. It is composed of the ROC (Receiver Operating Characteristics) curve plot, which shows the relationship between the True Positive Rate (TPR - probability that a positive sample is correctly predicted in the positive class) and the False Positive Rate (FPR - probability that a negative sample is incorrectly predicted in the positive class) across all possible classification thresholds (Hajian-Tilaki, 2013). For better comprehension, the ROC is a probability curve where the AUC represents the measure of separability, which measures how well a model can distinguish between classes. To illustrate, the higher the AUC, the better the chance the model can correctly segregate the mastitis classes with no mastitis.

5. RESULTS AND DISCUSSION

In the process of ascertaining the experimental results, a total of 100 versions of the same model were produced. The evaluation of binary and multi-classification models performed by Accuracy, Recall, Precision, AUC, and F1-Score was applied to measure the classification quality, as represented in tables 6 and 7.

Table 6: Metrics of performance evaluation of binary models

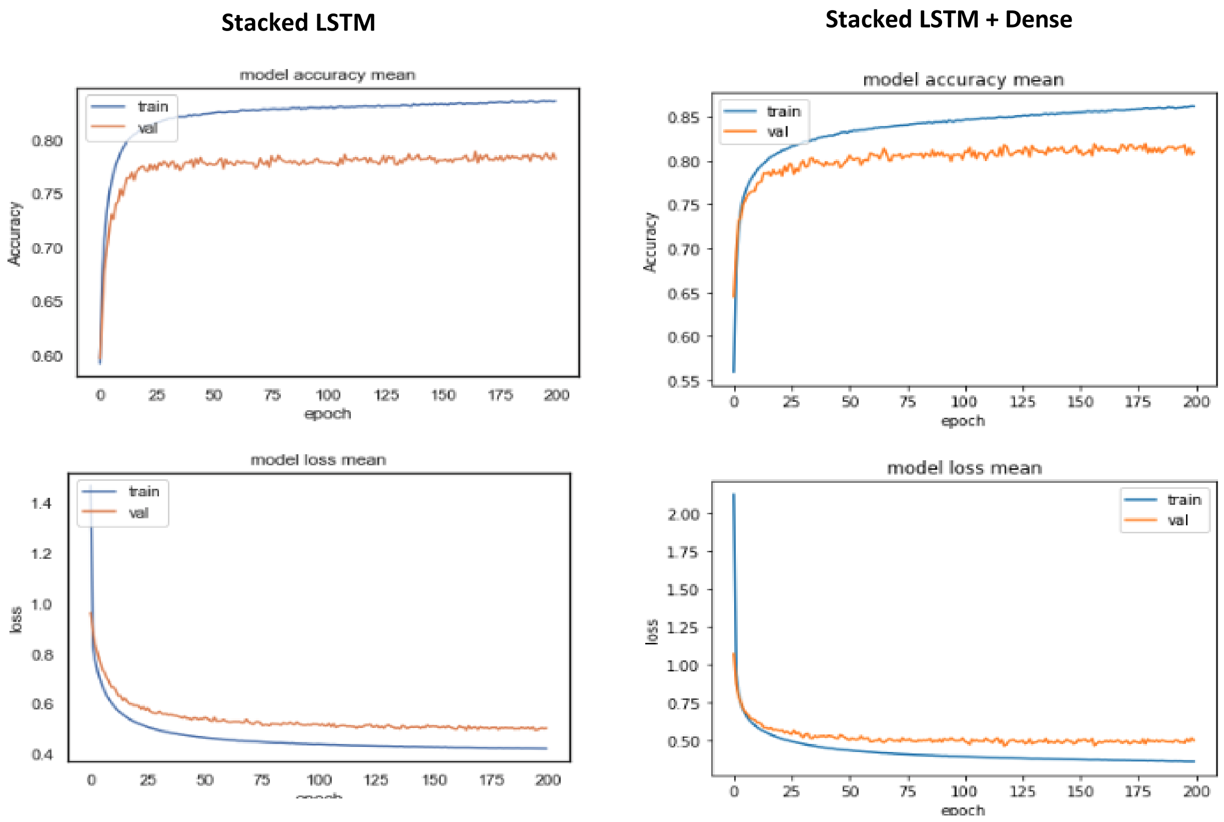
Binary Classification			
Metrics	Vanilla LSTM	Stacked LSTM	Stacked LSTM + Dense
Accuracy	0.61± 0.02	0.63± 0.04	**0.65± 0.03**
Recall	0.54± 0.02	**0.60± 0.02**	0.57± 0.02
Precision	0.54± 0.02	**0.59± 0.02**	0.58± 0.02
AUROC	0.54± 0.02	**0.61± 0.02**	0.56± 0.02
F1-Score			
Mastitis	0.35± 0.03	**0.45± 0.03**	0.38± 0.04
No Mastitis	0.72± 0.02	0.72± 0.05	**0.75± 0.04**

Table 7: Metrics of performance evaluation of multi classification models.

Multi Classification			
Metrics	Vanilla LSTM	Stacked LSTM	Stacked LSTM + Dense
Accuracy	0.81± 0.02	0.83± 0.02	**0.84**± 0.1
Recall	0.80± 0.01	0.82± 0.02	0.82± 0.06
Precision	0.80± 0.01	0.83± 0.02	0.83± 0.04
AUROC	0.86± 0.01	**0.87**± 0.01	0.79± 0.04
F1-Score			
Low Risk	0.84± 0.02	**0.87**±0.01	0.86±0.07
Moderate Risk	0.69± 0.03	**0.72**± 0.03	0.70± 0.09
High Risk	0.86± 0.02	**0.89**± 0.02	0.87± 0.05

From the two variants tested, it was noticeable that the models that generated the best performances were the multi-classification models, which successfully mimicked the choices of the veterinarian expert. Multiclass models outperformed all evaluative metrics compared to models of the binary classification approach.

The Stacked LSTM + Dense model from multi-classification obtained better accuracy than the remaining multi-classification models. Nonetheless, the Stacked LSTM model stands out for equating the Recall and Precision results and obtaining better AUCROC and F1-Score values than its competitor. The behavior of the best models is represented along 200 epochs, expressed in the graph's loss and accuracy in Figure 13.


Stacked LSTM

Stacked LSTM + Dense

Figure 13: Graphic representation of accuracy and loss behavior over 200 epochs in the Stacked LSTM and Stacked LSTM + Dense multi-classification models

As mentioned, a Grid Search was performed on the intermediate models of each approach. Nevertheless, when the Stacked LSTM parameters were replicated in the Stacked LSTM + Dense models, there was a performance loss. It was necessary to add 64 nodes in the dense layer and concatenation layer to reduce the overfitting, possibly because the models in which a dense layer was added in parallel increased the complexity of the models.

The performance of the binary models could have been more extraordinary, with the model having difficulty predicting the occurrence of mastitis even though the dataset was balanced with the SMOTE Tomek technique, which achieved more interesting Recall and Precision, as well as an increase in the F1-score for the minority class (mastitis) in the order of 20%.

The binary output could be the reason for less promising results. It was obtaining information for the binary output originating from whether or not the animal developed mastitis in the first 60 days after calving is an acceptable solution. However, it is known that most animals were administered antibiotics in the dry period, which could mask the proper behavior of the output to be analyzed, so it is anticipated that this assumption could be a significant bias in this type of approach.

Comparing all binary output models is possible to verify that the Stacked LSTM + Dense model presented better Accuracy and F1-Score, while the Stacked LSTM model had better Recall, Precision, and AUCROC values compared to the further models.

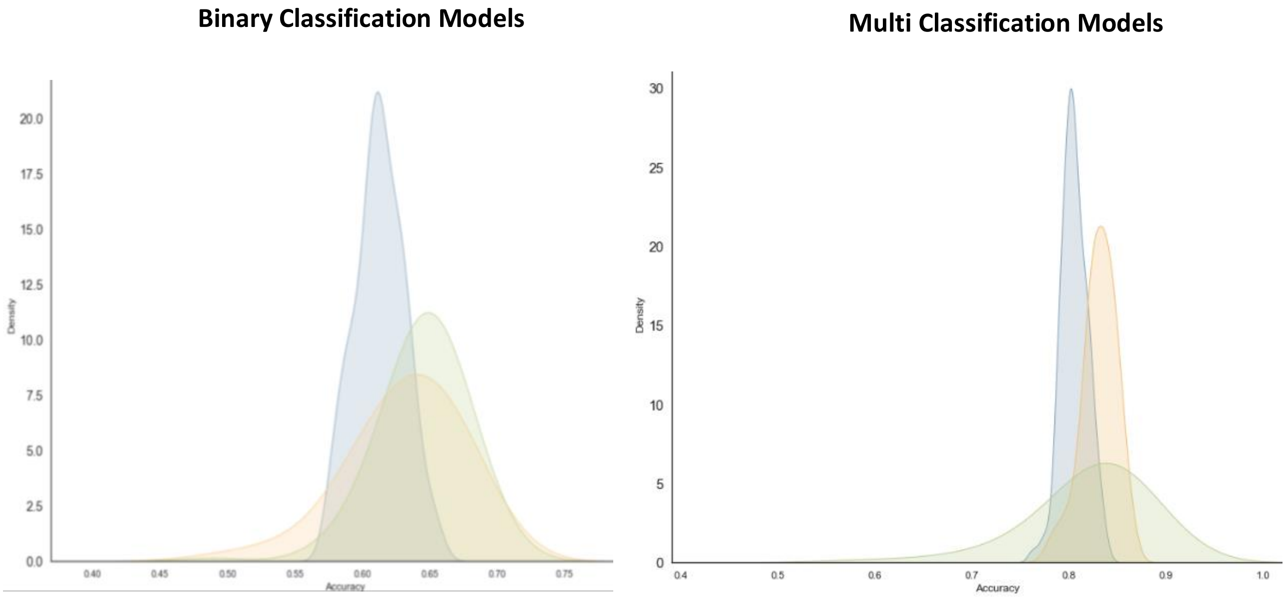
It was found that in the Stacked LSTM + Dense model, the addition of static features could have brought more improvement, as expected. This factor could be explained by the fact that the static features do not add value, possibly because the vast majority of the features are dummy data about the hygienic conditions of the herds. Regardless, the models with the addition of a dense layer were more consistent in their loss and accuracy behavior and easier to reduce overfitting than the rest of the participants.

A Paired t-student was used to verify the robustness of the results between the models of the two approaches.

The Paired Student's t-test is the most common statistical hypothesis test used for comparing the performance of ML models combined via random subsamples of the training dataset. Two hypotheses were produced inferring whether there are significant differences between the models' performances: H0 (null hypothesis), which assumes no differences between mean accuracy, or H1 (alternative hypothesis), which rejects H0. In other words, the alternative hypothesis assumes that two applied models perform differently.

In order to acquire benefit from this test, the variables must assume certain assumptions: the dependent variables are continuous, the observations made are independent, and the distribution must be approximately normal, as shown in Figure 14.

This test is merely indicative, considering that the results generated from a K-fold Cross-validation violate the assumptions of independence of the observations. Nevertheless, it is preferable to promote the replicability of the tests (Bouckaert & Frank, 2004) to the detriment of avoiding type-I errors (Dietterich, 1998).



Binary Classification Models

Multi Classification Models

Figure 14: Distribution of mean accuracy values over 100 repetitions. The models are represented in blue: Vanilla LSTM, orange: Stacked LSTM, and green: Stacked LSTM + Dense Layer.

The p-value was calculated to assess the improbability of the occurrence of a value in a determined sample space. The standard significance level was set to 0.05, which provides convincing evidence that the models perform differently.

The Paired Student t-test was performed among three models (I-Vanilla and Stacked LSTM models; II-Vanilla and Stacked LSTM + Dense layer; III- Stacked LSTM and Stacked LSTM + Dense layer) following both approaches. All results rejected H0 (p-value < 0.05), which indicates that the models presented significant differences in performance, corroborated by their different distributions.

A topic that is interesting to address in the discussion is the comparison of the outputs of the different approaches. Figure 15 shows a stacked bar graph representing the relationship between both outputs. The colors represent the multi-classification output referring to the choice of veterinarian, and the bars represent the binary output referring to the occurrence of mastitis.

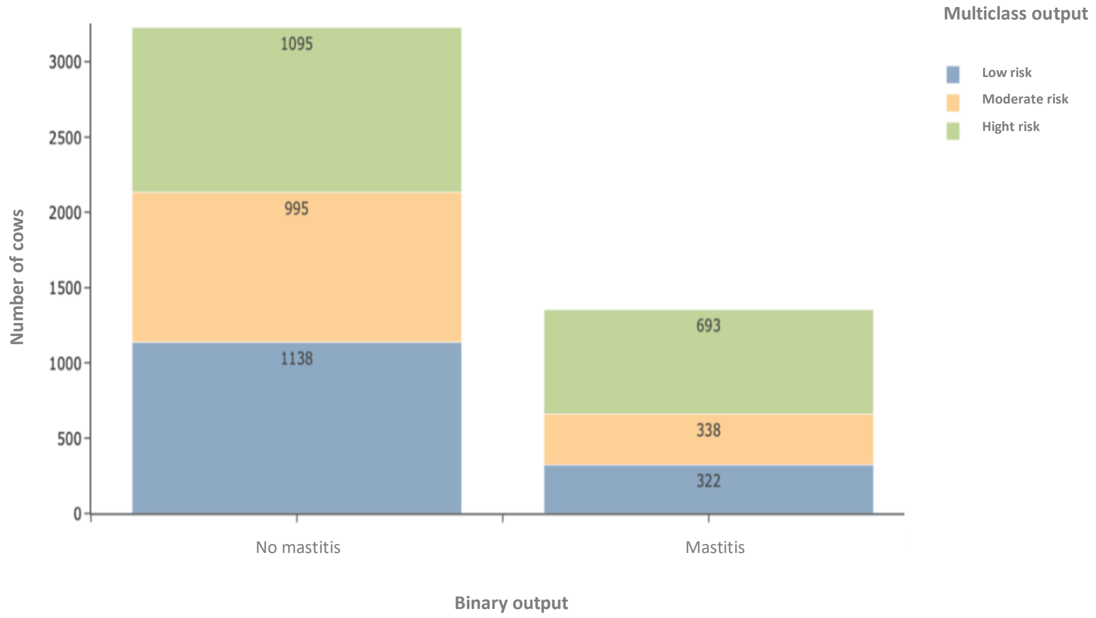


Figure 15: Stacked graph comparing the relation between binary outputs and multiclass outputs.

As seen in the graph, 322 animals were judged to be at low risk of mastitis though they suffered a mastitis event simultaneously. As most of these animals were given antibiotics and evaluated by the veterinarian's criteria, this 7% of cases suggests other environmental or individual variables outside the scope of the study, which may impact the prediction. This fact makes the prediction of mastitis somewhat unpredictable and should require more detailed analysis for future predictions.

6. CONCLUSION

This study has aimed to create a model that would be an additional tool for veterinarians when appraising the administration of antibiotics in the dry period of dairy cows. These models were designed to be used in herds with traditional implementations of milk evaluation, which assumes that evaluations are carried out monthly.

During this study, a distinct-feature engineering approach using raw data was employed to increase the number of features and enhance the input data information into the models. This step involved integrated dummy features that reflected the metabolic problems or nutritional needs that were occurring in the animal.

More than simply achieving perfect performances, this study focused on applying a very recent and innovative concept, which makes it possible to use deep learning hybrid models that combine LSTM layers of dynamic features with a Dense layer of static features.

With these challenging starting points, two types of classification methods were created: one with binary classes based on the occurrence of mastitis after the dry period and the other with multi classes based on the opinion of a veterinary expert.

The multi-classification model that incorporates a Stacked LSTM and, in parallel, a Dense layer managed to obtain the best accuracy of the study, while the model that presents only Stacked LSTM in its structure obtained better performance in AUROC and F1-Score.

The performances of the binary models were inferior to their competitors, probably justified by the administration of antibiotics in the dry period that influenced the construction of the binary output, altering the results. Even so, it will be worthy of a new study when new data with antibiotic administration criteria are obtained.

Given these results, opting for models that reproduce the veterinarian's decision in the face of various scenarios seems to be the most sensible decision at the moment.

7. LIMITATIONS AND RECOMMENDATIONS

In carrying out this study, some limitations were found, especially in models with binary output, which directly affect performance and should be mentioned to ensure progress in this field.

One of the biases that have a crucial impact on the results is that most cows that participated in this study were still administered antibiotics in the dry period. In other words, the animals obtained additional protection contributing to the naturally lower prevalence of mastitis in herds. This fact masks attributes that could help model training, causing misinterpretation of the data. The possible method to overcome this problem would be, for example, implementing a control group of cows to which no antibiotic would be administered. This method would bring clear discriminatory advantages to the models since it would better evidence the features that cause mastitis. Nevertheless, the economic and logistical constraints when using a group containing many animals covering these particularities will be complex.

Another important fact that may impact the model's generalization is the inclusion of animals exposed to different realities such as geographic location, climatic conditions, environmental conditions, and extensive or intensive regimes into which the cows are inserted.

Other recommendable approaches to consider are building models that include, for example, CNN layers for feature selection. Alternatively, experiments with other machine learning models, such as Radom Forest, which uses tabular instead of vector data, have been employed and seem to have promising results. Another suggestion is to include more discriminative features, a possible hypothesis to be included in future studies. All these recommendations intend to support new lines of research, to reduce the administration of antibiotics to what is strictly necessary.

8. REFERENCES

Alhussien, M. N., & Dang, A. K. (2018). Milk somatic cells, factors influencing their release, future prospects, and practical utility in dairy animals: An overview. *Veterinary World*, *11*(5), 562–577. https://doi.org/10.14202/vetworld.2018.562-577

ALIP. (2021). *Anexo Técnico de Acreditação L0512-1*. http://www .ipac.pt/docsig/?61JC-R76C-AF94-4J5P

ANABLE. (2014). *Publicação de Resultados* (p. 164). ANABLE- Associação Nacional para o Melhoramento de Bovinos de Leite.

Armstrong, D., Koops, W., & Ekström, J. (2018). *EuroDairy: Practical strategies to reduce antimicrobial use in dairy farming* (p. 39). EuroDairy.

Balaine, L., Dillon, E., & Läpple, D. (2019). Milk recording – a winning formula for dairy farmers. *TRESEARCH*, *14*(1), 32–33.

Balaji, B., D., B., Elumalai, V. K., & R., V. (2021). Automatic and non-invasive Parkinson's disease diagnosis and severity rating using LSTM network. *Applied Soft Computing*, *108*, 107463. https://doi.org/10.1016/j.asoc.2021.107463

Ball, P. J. H., Peters, A. R., & Peters, A. R. (2004). *Reproduction in cattle* (3rd ed). Blackwell Pub.

Berrar, D. (2019). Cross-Validation. Em *Encyclopedia of Bioinformatics and Computational Biology* (pp. 542–545). Elsevier. https://doi.org/10.1016/B978-0-12-809633-8.20349-X

Berry, E. A., & Hillerton, J. E. (2002). The Effect of Selective Dry Cow Treatment on New Intramammary Infections. *Journal of Dairy Science*, *85*(1), 112–121. https://doi.org/10.3168/jds.S0022-0302(02)74059-9

Biggs, A., Barrett, D., Bradley, A., Green, M., Reyher, K., & Zadoks, R. (2016). Antibiotic dry cow therapy: Where next? *Veterinary Record*, *178*(4), 93–94. https://doi.org/10.1136/vr.i338

Biscarini, F., Cremonesi, P., Castiglioni, B., Stella, A., Bronzo, V., Locatelli, C., & Moroni, P. (2020). A Randomized Controlled Trial of Teat-Sealant and Antibiotic Dry-Cow Treatments for Mastitis Prevention Shows Similar Effect on the Healthy Milk Microbiome. *Frontiers in Veterinary Science*, *7*, 581. https://doi.org/10.3389/fvets.2020.00581

Bobbo, T., Biffani, S., Taccioli, C., Penasa, M., & Cassandro, M. (2021). Comparison of machine learning methods to predict udder health status based on somatic cell counts in dairy cows. *Scientific Reports*, *11*(1), 13642. https://doi.org/10.1038/s41598-021-93056-4

Bórawski, P., Pawlewicz, A., Parzonko, A., Harper, J., K., & Holden, L. (2020). Factors Shaping Cow's Milk Production in the EU. *Sustainability*, *12*(1), 420. https://doi.org/10.3390/su12010420

Bostrom, A., & Bagnall, A. (2017). *A Shapelet Transform for Multivariate Time Series Classification*. https://doi.org/10.48550/ARXIV.1712.06428

Bouckaert, R. R., & Frank, E. (2004). Evaluating the Replicability of Significance Tests for Comparing

Learning Algorithms. Em H. Dai, R. Srikant, & C. Zhang (Eds.), *Advances in Knowledge Discovery and Data Mining* (Vol. 3056, pp. 3–12). Springer Berlin Heidelberg. https://doi.org/10.1007/978-3-540-24775-3_3

Butler, W. (2005). Nutrition, negative energy balance and fertility in the postpartum dairy cow. *Cattle Practice*, *13*, 13–18.

Cabezas-Garcia, E. H., Gordon, A. W., Mulligan, F. J., & Ferris, C. P. (2021). Revisiting the Relationships between Fat-to-Protein Ratio in Milk and Energy Balance in Dairy Cows of Different Parities, and at Different Stages of Lactation. *Animals*, *11*(11), 3256. https://doi.org/10.3390/ani11113256

Casper, D. P., & Schingoethe, D. J. (1989). Model to Describe and Alleviate Milk Protein Depression in Early Lactation Dairy Cows Fed a High Fat Diet. *Journal of Dairy Science*, *72*(12), 3327–3335. https://doi.org/10.3168/jds.S0022-0302(89)79494-7

Castro, D., Ribeiro, C., & Simões, J. (2009). Medicina da produção: Monitorização do balanço energético negativo (BEN) em vacas leiteiras—Herd health management: Negative energy balance (NEB) evaluation in dairy cattle. *REDVET*, *X*.

Cavero, D., Tölle, K.-H., Henze, C., Buxadé, C., & Krieter, J. (2008). Mastitis detection in dairy cows by application of neural networks. *Livestock Science*, *114*(2–3), 280–286. https://doi.org/10.1016/j.livsci.2007.05.012

Čechura, L., Žáková Kroupová, Z., & Benešová, I. (2021). Productivity and Efficiency in European Milk Production: Can We Observe the Effects of Abolishing Milk Quotas? *Agriculture*, *11*(9), 835. https://doi.org/10.3390/agriculture11090835

Chatfield, C., & Xing, H. (2019). Chapter 3: Some Linear Time Series Models. Em *The analysis of time series: An introduction with R* (Seventh edition, pp. 41–53). CRC Press, Taylor & Francis Group.

Cheng, W. N., & Han, S. G. (2020). Bovine mastitis: Risk factors, therapeutic strategies, and alternative treatments — A review. *Asian-Australasian Journal of Animal Sciences*, *33*(11), 1699–1713. https://doi.org/10.5713/ajas.20.0156

Dempster, A., Petitjean, F., & Webb, G. I. (2020). ROCKET: Exceptionally fast and accurate time series classification using random convolutional kernels. *Data Mining and Knowledge Discovery*, *34*(5), 1454–1495. https://doi.org/10.1007/s10618-020-00701-z

DGAV. (2017). *Normas para Aplicapao e Validagao dos Programas de Conservagao Genetica Animal e Programas de Melhoramento Genetic° Animal* (7.2 - Contraste leiteiro). DGAV.

Dhoble, A. S., Ryan, K. T., Lahiri, P., Chen, M., Pang, X., Cardoso, F. C., & Bhalerao, K. D. (2019). Cytometric fingerprinting and machine learning (CFML): A novel label-free, objective method for routine mastitis screening. *Computers and Electronics in Agriculture*, *162*, 505–513. https://doi.org/10.1016/j.compag.2019.04.029

Dietterich, T. G. (1998). Approximate Statistical Tests for Comparing Supervised Classification Learning Algorithms. *Neural Computation*, *10*(7), 1895–1923. https://doi.org/10.1162/089976698300017197

Dobson, H., Smith, R., Royal, M., Knight, C., & Sheldon, I. (2007). The High-producing Dairy Cow and its Reproductive Performance. *Reproduction in Domestic Animals*, *42*(s2), 17–23. https://doi.org/10.1111/j.1439-0531.2007.00906.x

Drackley, J. K. (1999). Biology of Dairy Cows During the Transition Period: The Final Frontier? *Journal of Dairy Science*, *82*(11), 2259–2273. https://doi.org/10.3168/jds.S0022-0302(99)75474-3

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

Ebrahimie, E., Mohammadi-Dehcheshmeh, M., Laven, R., & Petrovski, K. (2021). Rule Discovery in Milk Content towards Mastitis Diagnosis: Dealing with Farm Heterogeneity over Multiple Years through Classification Based on Associations. *Animals*, *11*(6), 1638. https://doi.org/10.3390/ani11061638

Erskine, R. J., Wagner, S., & DeGraves, F. J. (2003). Mastitis therapy and pharmacology. *Veterinary Clinics of North America: Food Animal Practice*, *19*(1), 109–138. https://doi.org/10.1016/S0749-0720(02)00067-1

Esteban, C., Staeck, O., Yang, Y., & Tresp, V. (2016). *Predicting Clinical Events by Combining Static and Dynamic Information Using Recurrent Neural Networks*. https://doi.org/10.48550/ARXIV.1602.02685

European Comission. (2022). *EU PRICES of COW's RAW MILK*. Milk Market Observatory. https://ec.europa.eu/info/sites/default/files/food-farming-fisheries/farming/documents/eu-raw-milk-prices_en.pdf

Eurostat. (2022). *EU Dairy Exports to Third countries*. https://ec.europa.eu/info/sites/default/files/food-farming-fisheries/farming/documents/eu-dairy-extra-trade_en.pdf

Fadul-Pacheco, L., Delgado, H., & Cabrera, V. E. (2021). Exploring machine learning algorithms for early prediction of clinical mastitis. *International Dairy Journal*, *119*, 105051. https://doi.org/10.1016/j.idairyj.2021.105051

Faouzi, J. (2022). Time Series Classification: A review of Algorithms and Implementations, Classification de séries temporelles: Une revue d'algorithmes et d'implémentations. Em K. Kotecha (Ed.), *Machine Learning (Emerging Trends and Applications)*. Proud Pen. https://hal.inria.fr/hal-03558165

Gers, F. A., Schmidhuber, J., & Cummins, F. (2000). Learning to Forget: Continual Prediction with LSTM. *Neural Computation*, *12*(10), 2451–2471. https://doi.org/10.1162/089976600300015015

Ghafoor, N., & Sitkowska, B. (2021). MasPA: A Machine Learning Application to Predict Risk of Mastitis in Cattle from AMS Sensor Data. *AgriEngineering*, *3*(3), 575–584. https://doi.org/10.3390/agriengineering3030037

Glorot, X., Bordes, A., & Bengio, Y. (2011). Deep Sparse Rectifier Neural Networks. Em G. Gordon, D.

Dunson, & M. Dudík (Eds.), *Proceedings of the Fourteenth International Conference on Artificial Intelligence and Statistics* (Vol. 15, pp. 315–323). PMLR. https://proceedings.mlr.press/v15/glorot11a.html

Gołaś, Z. (2017). DETERMINANTS OF MILK PRODUCTION PROFITABILITY OF DAIRY FARMS IN THE EU MEMBER STATES. *Problems of Agricultural Economics*, *352*(3), 19–40. https://doi.org/10.30858/zer/83030

Grandini, M., Bagli, E., & Visani, G. (2020). *Metrics for Multi-Class Classification: An Overview*. https://doi.org/10.48550/ARXIV.2008.05756

Greff, K., Srivastava, R. K., Koutnik, J., Steunebrink, B. R., & Schmidhuber, J. (2017). LSTM: A Search Space Odyssey. *IEEE Transactions on Neural Networks and Learning Systems*, *28*(10), 2222–2232. https://doi.org/10.1109/TNNLS.2016.2582924

Gröhn, Y. T., Wilson, D. J., González, R. N., Hertl, J. A., Schulte, H., Bennett, G., & Schukken, Y. H. (2004). Effect of Pathogen-Specific Clinical Mastitis on Milk Yield in Dairy Cows. *Journal of Dairy Science*, *87*(10), 3358–3374. https://doi.org/10.3168/jds.S0022-0302(04)73472-4

Gupta, S., Shukla, R., Shukla, R. K., & Verma, R. (2020). Deep Learning Bidirectional LSTM based Detection of Prolongation and Repetition in Stuttered Speech using Weighted MFCC. *International Journal of Advanced Computer Science and Applications*, *11*, 1–12.

Hajaya, Samarasinghe, Kulasiri, & Benavides. (2019, dezembro 1). Detection of dairy cattle Mastitis: Modelling of milking features using deep neural networks. *El Sawah, S. (ed.) MODSIM2019, 23rd International Congress on Modelling and Simulation.* 23rd International Congress on Modelling and Simulation (MODSIM2019). https://doi.org/10.36334/modsim.2019.A1.khamaysahajaya

Hajian-Tilaki, K. (2013). Receiver Operating Characteristic (ROC) Curve Analysis for Medical Diagnostic Test Evaluation. *Caspian Journal of Internal Medicine*, *4*(2), 627–635.

Higgins, H. M., Golding, S. E., Mouncey, J., Nanjiani, I., & Cook, A. J. C. (2017). Understanding veterinarians' prescribing decisions on antibiotic dry cow therapy. *Journal of Dairy Science*, *100*(4), 2909–2916. https://doi.org/10.3168/jds.2016-11923

Ho, Y., & Wookey, S. (2020). The Real-World-Weight Cross-Entropy Loss Function: Modeling the Costs of Mislabeling. *IEEE Access*, *8*, 4806–4813. https://doi.org/10.1109/ACCESS.2019.2962617

Hochreiter, S., & Schmidhuber, J. (1997). Long Short-Term Memory. *Neural Computation*, *9*(8), 1735–1780. https://doi.org/10.1162/neco.1997.9.8.1735

Hoeij, R. J., Dijkstra, J., Bruckmaier, R. M., Gross, J. J., Lam, T. J. G. M., Remmelink, G. J., Kemp, B., & van Knegsel, A. T. M. (2017). The effect of dry period length and postpartum level of concentrate on milk production, energy balance, and plasma metabolites of dairy cows across the dry period and in early lactation. *Journal of Dairy Science*, *100*(7), 5863–5879. https://doi.org/10.3168/jds.2016-11703

Hossin, H., & Sulaiman, S. (2015). A Review on Evaluation Metrics for Data Classification Evaluations. *International Journal of Data Mining & Knowledge Management Process*, *5*(2), 01–11. https://doi.org/10.5121/ijdkp.2015.5201

Hsieh, T.-Y., Wang, S., Sun, Y., & Honavar, V. (2021). Explainable Multivariate Time Series Classification: A Deep Neural Network Which Learns to Attend to Important Variables As Well As Time Intervals. *Proceedings of the 14th ACM International Conference on Web Search and Data Mining*, 607–615. https://doi.org/10.1145/3437963.3441815

Huijps, K., & Hogeveen, H. (2007). Stochastic Modeling to Determine the Economic Effects of Blanket, Selective, and No Dry Cow Therapy. *Journal of Dairy Science*, *90*(3), 1225–1234. https://doi.org/10.3168/jds.S0022-0302(07)71611-9

Hyland, S. L., Faltys, M., Hüser, M., Lyu, X., Gumbsch, T., Esteban, C., Bock, C., Horn, M., Moor, M., Rieck, B., Zimmermann, M., Bodenham, D., Borgwardt, K., Rätsch, G., & Merz, T. M. (2020). Early prediction of circulatory failure in the intensive care unit using machine learning. *Nature Medicine*, *26*(3), 364–373. https://doi.org/10.1038/s41591-020-0789-4

ICAR. (2017). *Section 2—Guidelines for Dairy Cattle Milk Recording* (p. 27). International Committee for Animal Recording (ICAR). https://www.icar.org/Guidelines/12-Milk-Analysis.pdf

INE. (2015). *Estatísticas da Produção e Consumo de Leite*.

Ismail Fawaz, H., Forestier, G., Weber, J., Idoumghar, L., & Muller, P.-A. (2019). Deep learning for time series classification: A review. *Data Mining and Knowledge Discovery*, *33*(4), 917–963. https://doi.org/10.1007/s10618-019-00619-1

Karim, F., Majumdar, S., Darabi, H., & Harford, S. (2018). *Multivariate LSTM-FCNs for Time Series Classification*. https://doi.org/10.48550/ARXIV.1801.04503

Kingma, D. P., & Ba, J. (2014). *Adam: A Method for Stochastic Optimization*. https://doi.org/10.48550/ARXIV.1412.6980

Klepacki, B., Olkowska, O., & Borawski, P. (2020). Milk Recording as a Factor Influencing the Performance of the Dairy Sector in Poland. *EUROPEAN RESEARCH STUDIES JOURNAL*, *XXIII*(Special Issue 1), 66–78. https://doi.org/10.35808/ersj/1745

Koch, L., & Lascano, G. (2018). Milk Fat Depression: Etiology, Theories, and Soluble Carbohydrate Interactions. *Journal of Animal Research and Nutrition*, *03*(02). https://doi.org/10.21767/2572-5459.100046

Labach, A., Salehinejad, H., & Valaee, S. (2019). *Survey of Dropout Methods for Deep Neural Networks*. https://doi.org/10.48550/ARXIV.1904.13310

Le, Ho, Lee, & Jung. (2019). Application of Long Short-Term Memory (LSTM) Neural Network for Flood Forecasting. *Water*, *11*(7), 1387. https://doi.org/10.3390/w11071387

LeCun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. *Nature*, *521*(7553), 436–444. https://doi.org/10.1038/nature14539

Leontjeva, A., & Kuzovkin, I. (2016). Combining Static and Dynamic Features for Multivariate Sequence Classification. *2016 IEEE International Conference on Data Science and Advanced Analytics (DSAA)*, 21–30. https://doi.org/10.1109/DSAA.2016.10

Lin, C., Zhang, Y., Ivy, J., Capan, M., Arnold, R., Huddleston, J. M., & Chi, M. (2018). Early Diagnosis and Prediction of Sepsis Shock by Combining Static and Dynamic Information Using Convolutional-LSTM. *2018 IEEE International Conference on Healthcare Informatics (ICHI)*, 219–228. https://doi.org/10.1109/ICHI.2018.00032

Liu, Y., Su, Z., Li, H., & Zhang, Y. (2019). An LSTM based classification method for time series trend forecasting. *2019 14th IEEE Conference on Industrial Electronics and Applications (ICIEA)*, 402–406. https://doi.org/10.1109/ICIEA.2019.8833725

Mahmud, M. S., Zahid, A., Das, A. K., Muzammil, M., & Khan, M. U. (2021). A systematic literature review on deep learning applications for precision cattle farming. *Computers and Electronics in Agriculture*, *187*, 106313. https://doi.org/10.1016/j.compag.2021.106313

Mammadova, N., & Keskin, İ. (2013). Application of the Support Vector Machine to Predict Subclinical Mastitis in Dairy Cattle. *The Scientific World Journal*, *2013*, 1–9. https://doi.org/10.1155/2013/603897

Martinho, V. J. P. D. (2020). The evolution of the milk sector in Portugal: Implications from the Common Agricultural Policy. *Open Agriculture*, *5*(1), 582–592. https://doi.org/10.1515/opag-2020-0061

Middlehurst, M., Large, J., Flynn, M., Lines, J., Bostrom, A., & Bagnall, A. (2021). HIVE-COTE 2.0: A new meta ensemble for time series classification. *Machine Learning*, *110*(11–12), 3211–3243. https://doi.org/10.1007/s10994-021-06057-9

Miebs, G., Mochol-Grzelak, M., Karaszewski, A., & Bachorz, R. A. (2020). Efficient Strategies of Static Features Incorporation into the Recurrent Neural Network. *Neural Processing Letters*, *51*(3), 2301–2316. https://doi.org/10.1007/s11063-020-10195-x

Miranda, P. (2014). *Análise Das Concentrações De Ureia No Leite De Vacas Frísias Holstein Da Ilha De São Miguel (Açores): Estudo Da Sua Relação Com O Arraçoamento E Dos Seus Efeitos Nos Índices Reprodutivos* [Faculdade de Medicina Veterinária da Universidade de Lisboa]. http://hdl.handle.net/10400.5/7579

Motohashi, H., Ohwada, H., & Kubota, C. (2020). Early Detection Method for Subclinical Mastitis in Auto Milking Systems Using Machine Learning. *2020 IEEE 19th International Conference on Cognitive Informatics & Cognitive Computing (ICCI*CC)*, 76–83. https://doi.org/10.1109/ICCICC50026.2020.9450258

Nalon, E., & Stevenson, P. (2019). Protection of Dairy Cattle in the EU: State of Play and Directions for Policymaking from a Legal and Animal Advocacy Perspective. *Animals*, *9*(12), 1066. https://doi.org/10.3390/ani9121066

Naqvi, S. A., King, M. T. M., DeVries, T. J., Barkema, H. W., & Deardon, R. (2022). Data considerations for developing deep learning models for dairy applications: A simulation study on mastitis detection. *Computers and Electronics in Agriculture*, *196*, 106895. https://doi.org/10.1016/j.compag.2022.106895

Neculai-Valeanu, A.-S., & Ariton, A.-M. (2022). Udder Health Monitoring for Prevention of Bovine

Mastitis and Improvement of Milk Quality. *Bioengineering*, *9*(11), 608. https://doi.org/10.3390/bioengineering9110608

Nickerson, S., & Ryman, V. E. (2019). Role of Antibiotic Therapy in Mastitis Control for Lactating and Dry Cows. *Bulletin 1516*.

Niemi, R. E., Vilar, M. J., Dohoo, I. R., Hovinen, M., Simojoki, H., & Rajala-Schultz, P. J. (2020). Antibiotic dry cow therapy, somatic cell count, and milk production: Retrospective analysis of the associations in dairy herd recording data using multilevel growth models. *Preventive Veterinary Medicine*, *180*, 105028. https://doi.org/10.1016/j.prevetmed.2020.105028

Nunan, C. (2022). Ending routine farm antibiotic use in Europe. *European Public Health Alliance (EPHA).*

PARCA. (2021). *Subcomissão Específica Dedicada ao Leite e Produtos Lácteios*. PARCA- Plataforma de Acompanhamento das Relações na Cadeia Agroalimentar. https://www.gpp.pt/images/Producao_e_Mercados/OrgProducao_CadeiaAlimentar/PARCA/PARCA_SubcomissoLeiteLcteos_RelDiagnostico__22122021.pdf

Peters, M. D. P., Silveira, I. D. B., & Fischer, V. (2015). Impact of subclinical and clinical mastitis on sensitivity to pain of dairy cows. *Animal*, *9*(12), 2024–2028. https://doi.org/10.1017/S1751731115001391

Philippidis, G., & Waschik, R. (2019). Melitz Meets Milk: The Impact of Quota Abolition on EU Dairy Export Competitiveness. *Journal of Agricultural Economics*, *70*(1), 44–61. https://doi.org/10.1111/1477-9552.12276

Ordinance nª 1066/91 de 22 de Outubro, 243, Ministério da Agricultura, Pescas e Alimentação, Artigo 1 5464 (1991).

Pyörälä, S. (2008). Mastitis in Post-Partum Dairy Cows. *Reproduction in Domestic Animals*, *43*, 252–259. https://doi.org/10.1111/j.1439-0531.2008.01170.x

Regea, G. (2018). *Review on Antibiotics Resistance and its Economic Impacts*.

Rodrigues, A. M., Guimarães, J., & Oliveira, C. (2012). Rentabilidade das explorações leiteiras em Portugal—Dados técnicos e económicos. *Livro de Resumos*, 109–129.

Ruiz, A. P., Flynn, M., Large, J., Middlehurst, M., & Bagnall, A. (2021). The great multivariate time series classification bake off: A review and experimental evaluation of recent algorithmic advances. *Data Mining and Knowledge Discovery*, *35*(2), 401–449. https://doi.org/10.1007/s10618-020-00727-3

Sasada, T., Liu, Z., Baba, T., Hatano, K., & Kimura, Y. (2020). A Resampling Method for Imbalanced Datasets Considering Noise and Overlap. *Procedia Computer Science*, *176*, 420–429. https://doi.org/10.1016/j.procs.2020.08.043

Scherpenzeel, C. G. M., den Uijl, I. E. M., van Schaik, G., Olde Riekerink, R. G. M., Keurentjes, J. M., & Lam, T. J. G. M. (2014). Evaluation of the use of dry cow antibiotics in low somatic cell count cows. *Journal of Dairy Science*, *97*(6), 3606–3614. https://doi.org/10.3168/jds.2013-7655

Shokoohi-Yekta, M., Hu, B., Jin, H., Wang, J., & Keogh, E. (2017). Generalizing DTW to the multi-dimensional case requires an adaptive approach. *Data Mining and Knowledge Discovery*, *31*(1), 1–31. https://doi.org/10.1007/s10618-016-0455-0

Sottomayor, M., Costa, L., & Ferreira, M. P. (2020). *Avaliação do impacto da reforma da PAC Pós 2020 sobre o setor do leite no continente português*. https://www.confagri.pt/content/uploads/2021/01/Estudo_Impacto_PAC_2023_Produtores_Leite_Fenalac.pdf

Srivastava, N., Hinton, G., Krizhevsky, A., Sutskever, I., & Salakhutdinov, R. (2014). Dropout: A Simple Way to Prevent Neural Networks from Overfitting. *Journal of Machine Learning Research*, *15*, 1929–1958.

Staudemeyer, R. C., & Morris, E. R. (2019). *Understanding LSTM -- a tutorial into Long Short-Term Memory Recurrent Neural Networks*. https://doi.org/10.48550/ARXIV.1909.09586

Stilwell, G. (2013). O Refugo nas Vacarias de Leite. *Agrotec- Revista técnico-científica agrícola*, *7*.

Stokes, S. R., Waldner, D. N., Jordan, E. R., & Looper, M. L. (2009). *Managing Milk Composition: Evaluating Herd Potential*. The Texas A&M University System.

Swana, E. F., Doorsamy, W., & Bokoro, P. (2022). Tomek Link and SMOTE Approaches for Machine Fault Classification with an Imbalanced Dataset. *Sensors*, *22*(9), 3246. https://doi.org/10.3390/s22093246

Vasilev, I., Slater, D., Spacagna, G., Roelants, P., & Zocca, V. (2019). Different types of activation function. Em *Python deep learning: Exploring deep learning techniques and neural network architectures with PyTorch, Keras, and TensorFlow* (Second edition, pp. 43–44). Packt Publishing Limited.

Webster, J. (2020). *Understanding the dairy cow* (Third edition). Wiley-Blackwell.

Wistuba, M., Grabocka, J., & Schmidt-Thieme, L. (2015). *Ultra-Fast Shapelets for Time Series Classification*.

Xudong, Z., Xi, K., Ningning, F., & Gang, L. (2020). Automatic recognition of dairy cow mastitis from thermal images by a deep learning detector. *Computers and Electronics in Agriculture*, *178*, 105754. https://doi.org/10.1016/j.compag.2020.105754

Yang, Y., Fasching, P. A., & Tresp, V. (2017). Predictive Modeling of Therapy Decisions in Metastatic Breast Cancer with Recurrent Neural Network Encoder and Multinomial Hierarchical Regression Decoder. *2017 IEEE International Conference on Healthcare Informatics (ICHI)*, 46–55. https://doi.org/10.1109/ICHI.2017.51

Ye, L., & Keogh, E. (2009). Time series shapelets: A new primitive for data mining. *Proceedings of the 15th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining - KDD ’09*, 947. https://doi.org/10.1145/1557019.1557122

Zeng, M., Zou, B., Wei, F., Liu, X., & Wang, L. (2016). Effective prediction of three common diseases by combining SMOTE with Tomek links technique for imbalanced medical data. *2016 IEEE*

International Conference of Online Analysis and Computing Science (ICOACS), 225–228. https://doi.org/10.1109/ICOACS.2016.7563084

Zhang, R., Liu, Y., & Sun, H. (2020). Physics-informed multi-LSTM networks for metamodeling of nonlinear structures. *Computer Methods in Applied Mechanics and Engineering*, *369*, 113226. https://doi.org/10.1016/j.cma.2020.113226

Zhang, X., Gao, Y., Lin, J., & Lu, C.-T. (2020). TapNet: Multivariate Time Series Classification with Attentional Prototypical Network. *Proceedings of the AAAI Conference on Artificial Intelligence*, *34*(04), 6845–6852. https://doi.org/10.1609/aaai.v34i04.6165

9. APPENDIX

Table I: Dynamic and static features descriptions

Dynamic Features	Description	Static features	Description
Dummy Features			
Metabolic and Infection Diseases		**Herd Checklist**	
SARA	Sub-Acute Ruminal Acidosis could occur when Fat/Protein ratio < 1	Contagious agents	Contagious mastitis pathogens are present in the herd
Subclinical Ketosis	Subclinical ketosis could occur when Fat / Protein ratio >= 1.5	Litter hygiene	Bed hygiene is carried out regularly
MFD	Milk Fat Depression could occur when Protein /Fat ratio > 0.95	Hyperkeratosis > 20%	More than 20% of the herd have teats hyperkeratosis
MPD	Milk Protein Depression could occur when Protein /Fat ratio < 0.8	Mastitis program	There is a mastitis prevention program in place for the herd
Subclinical Mastitis	Somatic Cells Count > 200.000	Selective drying	Antibiotic treatment is given based on an assessment of the individual cow.
Nutritional Needs: Urea (mg/kg) , Protein (%)		Gradual drying	Drying is carried out gradually
Prot ↓ and CH ↓	< 256.0 Urea and < 3.0 Prot.	Sealant	Use of internal teat sealant
Prot ↓ and CH ↑	< 256.0 Urea and > 3.2 Prot.	Dry cows separation	There is a physical separation between dry and lactating cows
Prot ↓	< 256.0 Urea and 3 - 3.2 Prot.	Udder inflammation > 20%	More than 20% of the herd have udder inflammation
Prot ↑ and CH ↓	> 342.0 Urea and < 3.0 Prot.	Subclinical mastitis vaccination	There is a vaccination plan against bovine mastitis
Prot ↑	> 342.0 Urea and < 3.0 Prot.	Mattress bedding	The animals rest on a mattress
Prot ↑ and CH ↑	256.0 - 342.0 Urea and < 3.0 Prot.	Compost bedding	The animals rest in a compost mixture (e.g., shavings, sawdust, straw, sawdust)
CH↓	256.0 - 342.0 Urea and > 3.2 Prot.	Pasture bedding	The animals rest in the pasture
Continuous Features			
From Milk Recordings			
Fat content	Milk fat (%)	Number of lactations	Lactation cycles of animal

Protein content	Milk protein (%)	Herd size	Number of animals per herd
Urea	Milk urea concentration (mg/kg)		
Milk Yield	Milk yield per day (L)		
DIM	Days In Milk		
SCC	Somatic Cells Count (x10^3)		
Ratio prot/fat	Protein and fat ratio		
Ratio fat/prot	Fat and protein ratio		
Avg. number of lactations	Average number of monthly lactations perherd		
Avg. milk yield	The average monthly milk yield per herd (L)		
Avg. fat content	The average monthly fat content per herd (%)		
Avg. protein content	Average monthly protein content per herd (%)		
Avg. urea	Average monthly urea per herd (mg/Kg)		
Avg. SCC	Average monthly SCC per herd		

NOVA Information Management School
Instituto Superior de Estatística e Gestão de Informação

Universidade Nova de Lisboa