FACULTY OF ECONOMICS AND
BUSINESS ADMINISTRATION

USING DEEP LEARNING FOR ANIMAL MONITORING TO IMPROVE ANIMAL WELFARE IN DAIRY CATTLE

Arno Liseune

Supervisors: Prof. Dr. Dirk Van den Poel & Prof. Dr. Miel Hostens

A dissertation submitted to Ghent University in partial fulfilment of the requirements for the degree of Doctor of Applied Economics

Academic year: 2021 – 2022

GHENT
UNIVERSITY

Copyright © 2022 by Arno Liseune

All rights are reserved. No part of this publication may be reproduced or transmitted in any form or by any means electronic or mechanical, including photocopying, recording, or by any information storage and retrieval system, without permission in writing from the author.

Doctoral Advisory Committee

Prof. dr. Patrick Van Kenhove (Dean, Ghent University)

Prof. dr. Dirk Van den Poel (Promotor, Ghent University)

Prof. dr. Miel Hostens (Co-Promotor, Utrecht University)

Prof. dr. Dries F. Benoit (Secretary, Ghent University)

Prof. dr. Matthias Bogaert (Ghent University)

Dr. Ines Adriaens (Wageningen University & Research)

Dr. Jeroen Degroote (Ghent University)

Acknowledgements

Pursuing my PhD would not have been possible without the help of several people. First and foremost, I would like to thank my promotor Dirk. With no specific research project in mind, Dirk was the first to give me the opportunity to pursue a PhD and he gave me the time to let me discover which research field I wanted to explore. He always believed in me as a researcher and was always available when I was struggling with my work. I also would like to thank him for believing in me as an assistant and entrusting me with the ACRM and Big Data courses. I am also grateful to Miel, who entrusted me with his research projects. Without him, I would never have been able to pursue a PhD in animal healthcare. I would also like to thank him for all the valuable feedback and comments he gave me throughout my entire PhD. I would also like to thank the members of the jury for reading my PhD and formulating many helpful suggestions to this work. Furthermore, I would like to express my gratitude to all my colleagues. In particular, I would like to thank Waikit, my office buddy, who always tried his best to answer all my questions, even though I never really understood his answers. I would also like to thank him for the many epic ping pong battles, which were essential for my PhD. Kate, my other office buddy, who educated me in the fantasy genre. Nanouk and Ziad, for teaching me everything that had nothing to do with my PhD such as food, friendship and love. Bram and Lisa, for helping me with the courses, for reviewing my work and for giving me an amazing time in Seattle. Except for my colleagues, I would also like to acknowledge all the other people who contributed to my PhD by supporting me or by helping me relax and enjoy. My roommates, Thomas, William and Lars, who were always willing to clean up my dirty kitchen for the amazing food I made for us. My other friends with whom I shared amazing days, evenings, nights and holidays, which allowed me to find joy and happiness outside my work. Last but not least, I would like to thank my grandmother, my parents and my sister. They always believed in me and supported me throughout me entire life. They always welcomed me with open arms in their houses with warmth and love. Without them, finishing my PhD would not have been possible. Thank you!

Table of Contents

List of Figures

List of Tables

Nederlandse samenvatting

De zuivelindustrie is de op één na grootste landbouwsector in Europa en vertegenwoordigt meer dan 12% van de totale landbouwproductie. Toch staat de zuivelindustrie momenteel voor een aantal uitdagingen. Ten eerste is de consument steeds meer begaan met de veiligheid, de diervriendelijkheid en de duurzaamheid van de productie van zuivelproducten. Ten tweede leidt een toename van de globale vraag naar zuivelproducten tot een stijging in het gebruik van grondstoffen en een enorme uitstoot van broeikasgassen. Ten derde vormt de klimaatverandering een grote bedreiging voor de veehouderij omdat het de kwaliteit van voedergewassen, de beschikbaarheid van water, de productiviteit van dieren, dierziekten en de biodiversiteit in het gedrang brengt. Gedreven door de noodzaak om zowel te kunnen voldoen aan de toenemende vraag naar duurzame producten alsook de uitstoot van broeikasgassen te verminderen, schakelt de zuivelindustrie over naar een meer duurzame aanpak van landbouw. Precision Livestock Farming (PLF) is een reeks geavanceerde technologieën gericht op het automatisch en real-time monitoren van dierenwelzijn, gezondheid, milieu en productie. Bijgevolg wordt het gebruik van deze technologieën aanzien als één van de voornaamste oplossingen voor het realiseren van de overgang naar een meer robuuste en veerkrachtige landbouw. Tot op heden worden er reeds verscheidene monitoringsystemen die gebruik maken van PLF technologieën gecommercialiseerd en geïmplementeerd op melkveebedrijven. Door de voortdurende toename van de omvang van de veestapel en de snelle ontwikkeling van PLF technologieën worden melkveehouders echter geconfronteerd met een exponentiële groei van het volume, de verscheidenheid en de complexiteit van de verzamelde data. In deze situatie neemt de performantie van huidige monitoringsystemen snel af, waardoor ze onbruikbaar worden voor praktische toepassingen. Het gebruik van meer geavanceerde algoritmes die wel in staat zijn om te leren van complexe datasets wordt daarom steeds belangrijker in de zuivelindustrie. Het doel van dit proefschrift is om in deze behoefte te voorzien door dierenmonitoringsystemen te ontwikkelen op basis van deep learning algoritmen die geschikt zijn voor praktische doeleinden. In Hoofdstuk 2 is een deep learning raamwerk ontwikkeld om alle ontbrekende melkgiftes van een lactatiecurve van een koe te schatten. We onderzoeken of een melkgifte nauwkeurig bepaald kan worden aan de hand van alle informatie waargenomen in de lactatiecyclus, ongeacht de hoeveelheid waargenomen data, het tijdstip van de waargenomen data en de tijdsintervallen tussen de waargenomen data. De resultaten tonen aan dat dit raamwerk gebruikt kan worden voor zowel het interpoleren als het voorspellen van ontbrekende melkgiften. Bovendien tonen we aan dat het toevoegen van informatie rond de veestapel, de pariteit, de gezondheid en de vruchtbaarheid van de koe leidt tot meer nauwkeurige voorspellingen. De methodologie voorgesteld in deze studie kan het monitoren van dieren aanzienlijk verbeteren, aangezien het onverwachte dalingen in melkproductie en dus ziekten snel kan opsporen en de impact van de gezondheidstoestand op de productiviteit kan inschatten. In Hoofdstuk 3 stellen we een methode voor die de volledige melkcurve van een bepaalde lactatiecyclus voorspelt. De melkcurve wordt voorspeld aan de hand van alle waargenomen informatie in de voorgaande cyclus zoals melkgifte-, pariteit-, veestapel- en gezondheidsinformatie. Deze methodologie stelt veehouders in staat om de werkelijke melkgifte van een koe te vergelijken met haar verwachte melkgifte over het gehele verloop van de lactatiecyclus. Dit kan het monitoren van koeien in het eerste stadium van de lactatie aanzienlijk vergemakkelijken. In Hoofdstuk 4 wordt een deep learning model voorgesteld dat het moment van afkalven voorspelt. We onderzoeken of sensorgegevens over verschillende gedragingen zoals eten, herkauwen, liggen en staan gebruikt kunnen worden om het moment van afkalving automatisch te voorspellen. Verder gaan we na of intelligente imputatiestrategieën ontwikkeld kunnen worden om deze voorspellingsmodellen geschikt te maken voor praktische toepassingen. De methode voorgesteld in dit hoofdstuk stelt boeren in staat om tijdig hulp te verlenen bij het afkalven en hun afkalfmanagement te optimaliseren.

English summary

The dairy industry is the second biggest agricultural sector in Europe, representing more than 12% of the total agricultural output. Nevertheless, the dairy industry currently faces many challenges. First, consumers are increasingly more concerned about the safety, animal friendliness and sustainability of the production of dairy products. Second, the increasing global demand also leads to a high use of natural resources and large emission of greenhouse gasses. Third, global warming is a major threat to the livestock sector as it compromises the quality of feed crop, water availability, animal productivity, animal diseases and biodiversity. Driven by a need to meet the increasing demand in sustainable products and reduce emission gasses, the dairy sector is transitioning towards a more sustainable approach of agriculture. Precision Livestock Farming (PLF) is a set of advanced technologies aimed at automatic, real-time monitoring of animal welfare, health, environmental impact, and production. These technologies have, therefore, been proposed to transition towards a more robust and resilient agriculture. To date, many monitoring systems supported by PLF are already being commercialized and deployed on dairy farms. However, due to an ongoing increase in herd size as well as the rapid development of PLF technologies, dairy farmers are faced with an exponential growth of the volume, variety and complexity of the data they collect. In these situations, the performance of current monitoring systems tend to drop rapidly, making them unusable for practical applications. More advanced algorithms that are able to learn from such complex datasets are therefore becoming increasingly important in the dairy industry. The objective of this dissertation is to address this need by developing animal monitoring systems based on deep learning algorithms that are suitable for practical implementations. In Chapter 2, a deep learning framework is developed to infer all missing milk yields along the lactation curve of a dairy cow. In particular, we investigate whether milk yield can be accurately inferred by using all the observed information in the lactation cycle, regardless of the amount of data, the recording time of the data and the time interval between the data. Results show that this framework can be used to accurately interpolate as well as predict missing milk yields. In addition, we find that adding information on herd, parity and health and reproduction events improve the predictions. The framework facilitates animal monitoring as it allows to rapidly detect unexpected milk losses and therefore diseases and to assess the impact of health and reproduction events on the cow's productivity. In Chapter 3, we propose a framework that predicts the entire milk yield curve of the subsequent lactation cycle. The milk yield curve is generated by using all the observed information in the preceding cycle, including

milk, parity, herd and health information. This forecasting methodology allows farmers to compare a cow's actual and expected milk yield over the entire course of the lactation cycle which facilitates animal monitoring in early lactation. In Chapter 4, we present a deep learning model that predicts the moment of calving. In particular, we analyze whether sensor data on behavioral activities such as eating, ruminating, lying and standing can be used to automatically detect the moment of parturition. Furthermore, we investigate whether smart imputation strategies can be used to make calving prediction models suitable for practical applications. The proposed methodology allows farmers to provide timely assistance and optimize their calving management.

1
Introduction

1.1 The dairy industry

The importance of the dairy industry in the EU is undeniable, as it represents more than 12% of the total agricultural output (Laure and Granier, 2018). In fact, with a total milk production of 160.1 million tonnes in 2020, the dairy sector is the second biggest agricultural sector in the EU (Eurostat, 2022). This is unlikely to change in the near future, as a global increase in dairy consumption is expected in the coming years, which in turn is driven by growth in population and income (OECD et al., 2020). Along with this increase in demand, consumers are more concerned with the sustainability, safety and animal-friendliness of their food production (Werkheiser, 2018). Simultaneously, the livestock sector is facing increasing pressure as it contributes to approximately 14.5% of the global emission of greenhouse gasses, driving global warming (Cheng et al., 2022). Climate change in turn is a major threat for livestock production as it impacts the quality of feed crop and forage, water availability, animal and milk production, livestock diseases, animal reproduction, and biodiversity (Rojas-Downing et al., 2017). As a result, the dairy industry needs to meet an increase in demand for dairy products while reducing the number of milking cows. Hence, driven by a need to increase the sustainability of production systems and decrease livestock emissions, farming is transitioning towards a more resilient approach of agriculture (Lovarelli et al., 2020). Resilient animals can be considered as animals that avoid early culling by coping well with the farm's management conditions. These animals reproduce easily, produce consistently, and react well to imposed challenges and (physiological) stress (Adriaens et al., 2020). During this transition towards a more resilient livestock farming, it is fundamental that many aspects such as productivity, environmental impact and animal welfare are monitored vigorously.

Precision Livestock Farming (PLF) represents a set of advanced technologies that allow for automatic and real-time monitoring of production, reproduction, health and welfare of livestock and environmental impact (Norton and Berckmans, 2017). In previous research, it has been widely recognized that the development of PLF technologies is essential for more robust and resilient dairy production systems (Lindblom et al., 2017). In particular, animal monitoring systems based on PLF technologies enable farmers to improve animal welfare and produce food safely with a reduced environmental impact through the early warning of illness, higher reproductive performance, genetic improvement, more efficient use of nutrients and the reduction of emissions (Lindblom et al., 2017). As a result, PLF technologies have been proposed by the agricultural European Innovation Partnership (EIP-AGRI) to transition towards a more sustainable agriculture (Laure and Granier, 2018). To date, several countries are already investing in smart farming approaches (Rose and Chilvers, 2018). Indeed, PLF technologies such as cameras, microphones and advanced sensor technologies that alert farmers via connected devices (e.g. phones or computers) about production details and behavioral activities are now being used to monitor cattle (Berckmans, 2017).

Although these technologies are already commercialized and being deployed on farms, many challenges still remain. In particular, to improve the efficiency of production systems, it is essential for smart farming approaches to collect, process and analyze the data correctly (García et al., 2020). To date, most animal monitoring systems still rely on traditional machine learning models such as k-nearest neighbors, random forest, support vector machines and logistic regression models to analyze the PLF data (García et al., 2020). Due to the rapid development and adoption of these PLF technologies in the dairy industry, however, the volume, variety, velocity, and complexity of the data rapidly increases. Milk meters, for example, record milk yield on a daily basis. Accelerometers and pedometers continuously measure lying, standing, eating and ruminating behavior. Video cameras record live video footage from cows and temperature loggers continuously track the body temperature. As a result, dairy farmers now collect multidimensional datasets, involving complex interdependencies between interacting variables measured over time (Rosa, 2021). Furthermore, the collection of complete farm data may be challenging. Lactation cycles may be incomplete due to defective milkmeters, disease, animal treatment, discarded milk, culled or deceased cows and variable milk recording schemes across herds. Data on health and reproduction events may be incorrect due to human error. Sensor data on behavioral activities may contain missing values due to faulty transmission of data, malfunction of sensors and animal treatment. Therefore, the need for more advanced algorithms that are able to learn from such complex datasets increases in the dairy industry. The objective of this dissertation is to explore frameworks based on deep learning algorithms to improve current animal monitoring systems.

1.2 AI for animal monitoring

1.2.1 A short introduction to machine learning and deep learning

Before delving into the notion of using Artificial Intelligence (AI) to support animal monitoring in the dairy industry, let us start by introducing the fundamentals and motivating the use of AI, machine learning and deep learning. AI is a very broad field of computer science in which computers are learned to mimic human behavior. Machine learning is a branch of AI that aims to develop algorithms to automatically learn patterns from data without being explicitly programmed (Bishop and Nasrabadi, 2006). The procedure of learning from data is called model training, as the model is trained to improve its performance on extracting patterns from the data. Once a machine learning model has been trained with sufficient accuracy, it can leverage the uncovered patterns in the data to make predictions on future data and support decision-making under uncertainty (Murphy, 2012). In general, machine learning models can be broadly categorized in two different types of problems:

- **Supervised Learning**: machine learning task that learns to map a relationship between input data or features and output data or labels. An example of supervised learning is detecting crop disease based on the image of the crop.
- **Unsupervised Learning**: machine learning task that learns to discover interesting patterns from the input data without experiencing labels. Clustering dairy cows based on their milking characteristics for breeding purposes is an example of an unsupervised learning problem.

Many AI tasks can be solved by manually designing the right set of features for that task, and passing these features to a machine learning algorithm. For example, a useful set of features to predict the total milk yield of a cow comprises the breed, the parity, the birth weight and the season of calving. In the last decade, however, manual feature engineering became much more challenging, as datasets became much larger and more unstructured, and tasks became increasingly more complex, e.g. identifying and spraying weeds based on live video footage of drones to reduce the use of crop protection chemicals. Deep learning is a subset of machine learning that can solve more complex problems by automatically learning the features from the data. It does this by representing the task as a hierarchy of concepts, whereby each concept is expressed in terms of other, more simple concepts (Goodfellow et al., 2016). With recent advancements in computer technologies and algorithmic techniques, deep learning methods have led to major breakthroughs in various research domains such as speech recognition (Graves et al., 2013), machine translation (Bahdanau et al., 2016), object classification (He et al., 2015) object detection (Redmon et al., 2016) and image generation (Goodfellow et al., 2014).

In the dairy industry, most animal monitoring systems are still based on traditional machine learning models. Due to the ongoing modernization and digitalization of the dairy industry, however, enormous amounts of heterogeneous data are continuously collected by various PLF technologies. Therefore, the need for algorithms that can learn from more complex data increases in the dairy industry.

1.2.2 Predicting the milk yield

Accurately predicting the milk yield is one of the most valuable assets for a dairy farmer as it enables more efficient herd management. In particular, predicting milk yields allows for better financial planning as dairy farmers can obtain timely projections of their future production as well as costs such as required feed intake, proteins and nutrients as well as energy consumption and plant utilization (Murphy et al., 2014). Lactation models can be used for revenue optimization as they quantify the effect of various factors such as breeding, calving, feeding and culling decisions on the milk production (Ehrlich, 2010; Lormore and Galligan, 2001). Additionally, predicting a cow's productivity can support breeding and culling decisions. For breeding purposes, lactation models facilitate early identification of the most productive females as well as superior bulls based on the analysis of the total productivity of its offspring (Kliś et al., 2021; Lacroix et al., 1995). On the other hand, accurate forecasts of a cow's productivity contributes to improved culling decisions (Njubi et al., 2010). Furthermore, predicting the expected milk yield allows farmers to identify unforeseen milk losses and gain insights in health-related problems (Jensen et al., 2018). For example, monitoring systems can automatically alert farmers when a cow's realized milk yield deviates much from the expected lactation curve. This can help to rapidly detect diseases such as mastitis (Adriaens et al., 2018; Wilson et al., 2004), metritis (Giuliodori et al., 2013; Wittrock et al., 2011) and ketosis (McArt et al., 2012) and, therefore, reduce costs associated with these diseases which include lower production, discarded milk because of antibiotic therapy, labor, veterinary costs and treatments, and culling or death (Wilson et al., 2004). At the same time, deviations between the expected and observed milk yield also provide dairy farmers valuable feedback on the recovery and cure of the animal (Adriaens et al., 2018).

In early research, lactation curves were modeled by parametric functions fitted on empirical data. Probably the most popular parametric lactation model is the

Wood incomplete gamma function which is characterized by an increasing phase until a peak yield, followed by a more steady decline (Wood, 1967). Several modifications to the Wood model were proposed to fit a wider range of possible shapes, e.g. a combination of an exponential and linear model (Wilmink, 1987), a polynomial regression model (Ali and Schaeffer, 1987), a Legendre polynomial (Kirkpatrick et al., 1994) and an exponential model (Ehrlich et al., 2011). An overview of the available parametric lactation models are given in Bouallegue and M'Hamdi (2020). The main purpose of these empirical models was to describe the lactation traits of homogeneous groups of animals for management purposes such as breeding decisions (Macciotta et al., 2011). Later, due to developments in PLF technologies and an increasing need for individual animal monitoring, several models have been proposed to predict individual lactation (Macciotta et al., 2011). Vasconcelos et al. (2004) and Macciotta et al. (2002) employed autoregressive (AR) models to predict a daily yield based on a limited number of preceding yields. In the study presented by Grzesiak et al. (2006), a framework was developed to predict a milk yield on a given day by a multilayer perceptron (MLP) network. In addition to predicting daily yields, many models have also been developed to predict a cow's 305d cumulative milk yield (Grzesiak et al., 2003a,b; Njubi et al., 2010; Sharma et al., 2007).

Although many lactation models have already been suggested by different authors, some problems still remain. First, a fixed number of test-day milk yields measured at fixed time intervals is required by most of the previously proposed models. In reality, however, missing milk recordings and variable time intervals may exist due to animal treatment, system updates, defective recording machines and different recording schemes between herds. In these cases, previously proposed lactation models either fail to generate predictions or heavily depend on the quality of the missing value imputations to perform well. However, little information on the generalizability of these models on incomplete data is known, as most of these models were trained and evaluated on complete datasets without missing information. Second, individual curve-fitting models, such as those proposed by Macciotta et al. (2002) and Vasconcelos et al. (2004) only take into account historical milk yields to make predictions. Occasionally though, recorded milk yields exist before and after the missing milk yield that is to be inferred. In that case, modeling the correlations between the prediction and all its adjacent milk yields may yield more accurate estimates. Third, individual milk yields are either modeled by group averages or by a limited amount of preceding yields recorded in the same lactation cycle. This means that milk yield in early lactation is either estimated by only using herd statistics or not modeled at all.

This dissertation contributes to the literature on lactation modeling in two ways. First, we present a lactation model that infers all missing milk yields of a lactation curve based on all the observed information in the lactation cycle. More specifically, in Chapter 2, we present a method to predict a certain daily milk yield using all the milk yields observed before and after the moment of prediction, regardless of the number of observations and the recording interval between the different milk yields. We also investigate whether adding information on herd, parity and health and reproduction events improves the estimated milk yield curve. The presented framework can be used to more accurately infer missing milk yields along the entire lactation curve and therefore obtain more realistic estimates of the herd's productivity and hence revenues and costs. Moreover, the model is also able to predict future yields, which facilitates data-driven culling and breeding decision as well as improved animal monitoring. Second, we propose a predictive framework that generates the expected lactation curve of a certain lactation cycle based on all the observed information in the preceding cycle. In particular, in Chapter 3, we investigate whether a combination of data on milk yield, herd and parity statistics as well as health and reproduction events observed in the preceding cycle can generate a more accurate estimation of the milk yield curve than the lactation curve generated by a classic parametric model. This framework can be used to calculate the milk losses immediately after calving and, therefore, supports animal monitoring during the entire course of the lactation cycle. In addition, the framework enables farmers to increase their forecast horizon with respect to the farm's future profitability.

1.2.3 Predicting the moment of calving

Together with forecasting milk yields, monitoring the moment of calving has become a crucial aspect of dairy farm management as the moment of parturition is one of the most critical moments in the life of both the dam and the newborn (Barrier et al., 2013). Dystocia, i.e. difficulties or abnormalities experienced during calving, can severely compromise the animal's welfare and represents a problem worldwide for the dairy industry (Barrier et al., 2013; Mee et al., 2014). In fact, the incidence rate of dystocia in dairy cattle comprises 2% to 22% in Europe (Crociati et al., 2022). The physiological effects of dystocia on cows are well known. Dams that experience dystocia are at increased risk of damage to the uterus which may cause uterine diseases such as metritis and endometritis (Bruun et al., 2002; Ghanem et al., 2013). Moreover, it is believed that dystocia is a painful and stressful experience for dams (Huxley and Whay, 2006; Laven et al., 2009). On the other hand, dystocial calves can show signs of prolonged hypoxia and significant acidosis (Lombard et al., 2007) as well as internal bleedings and external lesions (Berglund et al., 2003). Moreover, it has been reported that more than 50% of stillbirths can be directly attributed to dystocia (Meyer et al., 2000). Hence, as difficulties with delivery adversely impact animal welfare and therefore farm economics, dystocia is seen as a major economic trait for a dairy farmer (Abdela and Ahmed, 2016). In fact, previous research has estimated the total cost associated with a difficult calving at

€500 (McGuirk et al., 2007). This cost can be attributed to a lower fertility, milk production and survival rate of the dam (Tenhagen et al., 2007) as well as a lower future productivity, growth rate and survival rate of the newborn (Crociati et al., 2022). In addition, difficult calvings also affect costs due to an increased need for veterinary assistance and the lost value of dead newborns (McGuirk et al., 2007). Prevention of dystocia should therefore be a top priority of dairy farms to improve animal welfare and farm economics.

Several risk factors of dystocia such as the biology of the dam (e.g. breed, parity, pelvic diameter), the weight, sex and position of the calve as well as seasonal effects have been identified (Johanson and Berger, 2003; Norman et al., 2010). Good farm management can significantly reduce some of these risks and increase the reproductive performance (Tenhagen et al., 2007). Timely delivery assistance, for example, may significantly reduce the risk of dystocia, reduce the pain and stress experienced during labor and improve the reproductive performance of the dam (Mainau and Manteca, 2011). Providing timely assistance, however, requires an accurate estimate of the moment of parturition. In the past, dairy farmers mainly relied on the day of the last insemination or on visual ispection to estimate the expected calving time. Yet, as herd sizes tend to increase yearly, manually inferring the calving time becomes challenging even for experienced personnel (Borchers et al., 2017; Lange et al., 2017). Automated monitoring systems that accurately predict the moment of calving have therefore become essential tools for dairy farmers (Ouellet et al., 2016).

Around the moment of calving, several behavioral changes take place in dairy cattle. Eating, ruminating and grooming behavior decrease, while lying, restlessness and tail raising behavior increase (Huzzey et al., 2005; Jensen, 2012; Miedema et al., 2011a,b; Speroni et al., 2018). Visual observation of these behavioral parameters, however, is time-consuming, subjective and prone to human error and requires experienced observers (Zehner et al., 2019). Various sensors such as pedometers, accelerometers, microphones and thermometers have therefore been developed to automatically detect these behaviors (an overview of commercially available sensors is given in Crociati et al. (2022)). As a result, most frameworks that have been proposed in literature to predict the moment of calving rely on data obtained from these sensors to generate calving alerts (Borchers et al., 2017; Fadul et al., 2017; Keceli et al., 2020; Ouellet et al., 2016; Rutten et al., 2017; Zehner et al., 2019).

Yet, some caution should be exercised in the interpretation of the results of the previously proposed frameworks. Most previous studies rely on a very limited sample of cows coming from one herd. The generalizability of these models on cows coming from different herds should therefore be investigated in more detail. Additionally, most previous studies use traditional machine learning algorithms to generate calving predictions. When dealing with multidimensional time series data containing complex interdependencies between interacting variables, however, these models tend to fall short in learning patterns compared to more advanced deep learning algorithms. Furthermore, previous studies rely on the assumption that the data in real everyday conditions is of adequate quality, as observations with missing values are removed from the analysis. In reality, however, the data obtained from sensors such as pedometers and accelerometers can contain many missing values due to animal treatment, defective recording machines, faulty transmission and software updates. As a result, the performance of the models developed in previous studies may considerably decrease when deployed in practice.

Hogeveen et al. (2010) defined three criteria that must be satisfied for a calving detection model to be implemented for commercial livestock production: a high performance, a relevant time window and a high degree of similarity between the study design and the real everyday conditions in commercial farms. In this dissertation, we contribute to the literature on calving prediction by providing a framework that satisfies these three conditions and therefore is suitable for practical implementations. In particular, we provide a framework that has been validated on a large dataset for different time intervals with respect to the moment of calving, i.e. 24h, 12h, 6h, 3h and 1h. This approach allows farmers to receive hourly alerts on the probability of calving starting within different time frames. Moreover, no assumption was made about the quality of the data, as we developed a framework that is able to generate reliable predictions, regardless of the amount of missing values in the sensor data. Hence, the framework proposed in this study can be used in practice to predict the moment of calving, and therefore facilitate timely supervision as well as improve animal welfare.

1.3 Data

Chapter 2 and 3. The data leveraged in these chapters were gathered from 104 different dairy farms. Each of these farms was equipped with a Herd Management System (HMS) that allowed dairy farmers to record animal lifetime events and collect milkmeter information. The lifetime events which include calving, heat, pregnancy, insemination, mastitis, disease, abort, culling and death were manually recorded by the farmer on connectable devices such as tablets, smartphones or computers. The HMS data sources were collected by a cloud-based dairy analysis application (www.mmmooogle.com). The platform applied basic cleaning procedures on incoming data. Duplicate milk yields were dropped and unrealistic event sequences were corrected (e.g. calving between pregnancy and dry-off). By using a proprietary data ontology model, the platform standardized and mapped the HMS data to a uniform data representation. Herd statistics as well as animal KPI's were automatically calculated based on the collected HMS data. The data were collected between 2013 and 2019 and comprised 12790342 milk yield recordings coming from 59122 lactations of 35133 unique cows. Furthermore, 304742 recordings of 13 unique lifetime events were registered during the time period of this research study.

Chapter 4. The data for this chapter was obtained from 8 commercial dairy farms with freestall barns in the Netherlands between August 2016 and November 2020. Each of these farms had implemented the Nedap Infrastructure (Nedap, Groenlo, The Netherlands) that consisted of a server, antenna and wearable sensors. From the moment the infrastructure was running, each cow was equipped with the Nedap Smarttag Leg and the Nedap Smarttag Neck sensor for the entire period of this study. The Nedap Smarttag sensors use G-sensors, which use acceleration as a measure of movement and the x-, y-, and z-axes (3-dimensional space) to determine the angle. Every second, acceleration is recorded by the Nedap sensors. A proprietary neural network was used to determine whether the cow was displaying the specified behavior per minute. Daily aggregates were automatically obtained by the Nedap software. The Leg sensor was attached to the front leg and measured the number of steps, standing time, walking time and lying time. The Neck sensor was attached to the neck and measured eating time, rumination time and inactive time, i.e., time not spend eating and ruminating. Apart from the sensor data, the moment of calving of each of these cows was manually recorded by the farmer. In total, the day of calving was registered for 3902 different calvings. For 572 of these calvings, the exact timestamp was registered by the farmer at the moment the farmer visually observed the parturition, i.e. the completed birth of the calve.

1.4 Outline and contributions of this dissertation

The frameworks presented in this dissertation allow dairy farmers to continuously monitor their cows as they transition through their lactation cycles. This helps dairy farmers to better manage the welfare of their animals, particularly during the transition period. This is the period between late pregnancy and early lactation and is one of the most critical moments of a dairy cow as most health disorders occur during this time (Drackley, 1999). This is demonstrated in more detail by **Fig** 1.1, in which a schematic overview of the frameworks developed in this dissertation is given. In Chapter 2, missing milk yields of a certain lactation cycle (dashed blue line) are predicted by herd statistics as well as the observed milk yields and health events in the same lactation cycle (solid blue line). At the end of the lactation cycle, the lactation curve of the next cycle (dashed yellow line) is predicted by all the observed milk yields, health events and herd statistics in the previous lactation cycle (solid yellow line) in Chapter 3. Then, before the start of the next lactation cycle, the moment of calving (dashed green line) is predicted in Chapter 4 by using behavioral sensor data (solid green line). In the remainder of this Section an outline will be provided for the remainder of the dissertation and the main contributions in

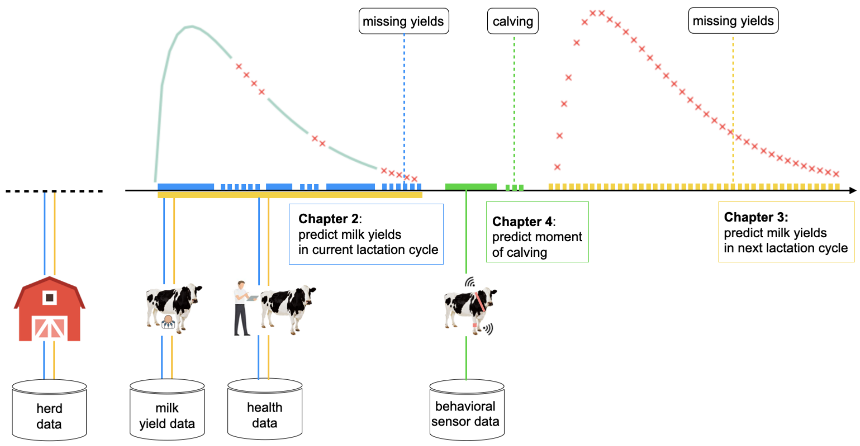


Figure 1.1: Schematic overview of framework of this dissertation. Solid horizontal line=time interval of features, dashed horizontal line=time interval of labels

each chapter are highlighted.

In **Chapter 2**, a deep learning framework to infer missing milk yields along the entire lactation cycle of a dairy cow is assessed. Past studies have primarily focused on predicting the expected yield at a specific point in the lactation cycle based on a fixed number of historical milk yields measured at constant time intervals. These lactation models, however, only partially take into account the complex interdependencies between all the interacting factors measured over different time intervals. Furthermore, these models fall short when missing milk yields or variable lengths of intervals between different measurements are present in the lactation curve. We extend previous research by proposing a lactation model in which missing milk yields along the lactation curve are dynamically updated as soon as new information in the corresponding lactation cycle becomes available. In particular, a sequential autoencoder (SAE) is trained to encode and decode all the available information on milk yields, herd statistics, parity, and health and reproduction events observed during the lactation cycle. It is shown how the framework can be used by dairy farmers to accurately predict a cow's milk yield, which in turn allows farmers to obtain more accurate forecasts on their future revenues and to enhance their animal monitoring systems.

In **Chapter 3**, a deep learning model is implemented that predicts the entire milk yield curve of a certain lactation cycle. In previous research, lactation models rely on a fixed amount of observed milk yield in early lactation to predict milk yield curves. This methodology, however, makes animal monitoring particularly difficult in early lactation, as there exists no information on expected milk yields in the period immediately after calving. In contrast, this study proposes a framework to predict the entire lactation curve of dairy cows by leveraging information on milk yields, health and reproduction events and herd statistics observed in the preceding cycle. This framework allows dairy farmers to increase the forecast horizon with respect to the herd's future productivity as well as to improve animal monitoring in early lactation.

In **Chapter 4**, a forecasting methodology is developed that aims to help dairy farmers to provide timely calving assistance by predicting the moment of calving. While machine learning frameworks were already developed to predict the onset of calving, their lack to accurately impute missing values and extract useful patterns from high-dimensional sequential data remains a barrier to the deployment of these models in practice. This study evaluates deep learning models to predict calving within 24h, 12h, 6h, 3h and 1h based on sensor data measuring a cow's eating, ruminating, walking, lying and standing behavior. Additionally, a novel methodology is analyzed to impute missing sensor values. This approach shows how deep learning algorithms can leverage all the available information to impute missing values in sensor data. The results show that deep learning algorithms outperform machine learning methods to predict the moment of calving based on sensor data. Moreover, it is shown that using the missing value imputations significantly improves the predictive performance for observations containing missing values. The framework proposed in this study can be used by farmers to optimize their calving management and hence improve animal monitoring.

References

Abdela, N. and Ahmed, W. M. Risk factors and economic impact of dystocia in dairy cows: a systematic review. *Journal of Reproduction and Infertility*, 7(2): 63–74, 2016.

Adriaens, I., Friggens, N., Ouweltjes, W., Scott, H., Aernouts, B., and Statham, J. Productive life span and resilience rank can be predicted from on-farm first-parity sensor time series but not using a common equation across farms. *Journal of Dairy Science*, 103(8):7155–7171, 2020. ISSN 0022-0302. doi: https://doi.org/10.3168/jds.2019-17826.

Adriaens, I., Huybrechts, T., Aernouts, B., Geerinckx, K., Piepers, S., De Ketelaere, B., and Saeys, W. Method for short-term prediction of milk yield at the quarter level to improve udder health monitoring. *Journal of Dairy Science*, 101 (11):10327–10336, 2018. ISSN 0022-0302. doi: https://doi.org/10.3168/jds.2018-14696.

Ali, T. E. and Schaeffer, L. R. Accounting for covariances among test day milk yields in dairy cows. *Canadian Journal of Animal Science*, 67(3):637–644, 1987.

Bahdanau, D., Cho, K., and Bengio, Y. Neural machine translation by jointly learning to align and translate. *ArXiv e-prints*, 2016.

Barrier, A., Haskell, M., Birch, S., Bagnall, A., Bell, D., Dickinson, J., Macrae, A., and Dwyer, C. The impact of dystocia on dairy calf health, welfare, performance and survival. *The Veterinary Journal*, 195(1):86–90, 2013. ISSN 1090-0233. doi: https://doi.org/10.1016/j.tvjl.2012.07.031.

Berckmans, D. General introduction to precision livestock farming. *Animal Frontiers*, 7(1):6–11, 01 2017. ISSN 2160-6056. doi: 10.2527/af.2017.0102.

Berglund, B., Steinbock, L., and Elvander, M. Causes of stillbirth and time of death in Swedish Holstein calves examined post mortem. *Acta Veterinaria Scandinavica*, 44(3):111, 2003. doi: 10.1186/1751-0147-44-111.

Bishop, C. M. and Nasrabadi, N. M. *Pattern recognition and machine learning*, volume 4. Springer, 2006.

Borchers, M., Chang, Y., Proudfoot, K., Wadsworth, B., Stone, A., and Bewley, J. Machine-learning-based calving prediction from activity, lying, and ruminating behaviors in dairy cattle. *Journal of Dairy Science*, 100(7):5664–5674, 2017. ISSN 0022-0302. doi: https://doi.org/10.3168/jds.2016-11526.

Bouallegue, M. and M'Hamdi, N. Mathematical modeling of lactation curves: A review of parametric models. *Lactation in farm animals-biology, physiological basis, nutritional requirements, and modelization*, 1:1–20, 2020.

Bruun, J., Ersbøll, A., and Alban, L. Risk factors for metritis in danish dairy cows. *Preventive Veterinary Medicine*, 54(2):179–190, 2002. ISSN 0167-5877. doi: https://doi.org/10.1016/S0167-5877(02)00026-0.

Cheng, M., McCarl, B., and Fei, C. Climate change and livestock production: A literature review. *Atmosphere*, 13(1), 2022. ISSN 2073-4433. doi: 10.3390/atmos13010140.

Crociati, M., Sylla, L., De Vincenzi, A., Stradaioli, G., and Monaci, M. How to predict parturition in cattle? a literature review of automatic devices and technologies for remote monitoring and calving prediction. *Animals*, 12(3), 2022. ISSN 2076-2615. doi: 10.3390/ani12030405.

Drackley, J. K. Biology of dairy cows during the transition period: the final frontier? *Journal of Dairy Science*, 82(11):2259–2273, 1999. ISSN 0022-0302.

Ehrlich, J. L. Quantifying shape of lactation curves, and benchmark curves for common dairy breeds and parities. *The Bovine Practitioner*, 45(1):88–95, Dec. 2010. doi: 10.21423/bovine-vol45no1p88-95.

Ehrlich, J. et al. Quantifying shape of lactation curves, and benchmark curves for common dairy breeds and parities. *Bovine Practitioner*, 45(1):88, 2011.

Eurostat. Milk collection (all milks) and dairy products obtained - annual data, 2022.

Fadul, M., Bogdahn, C., Alsaaod, M., Hüsler, J., Starke, A., Steiner, A., and Hirsbrunner, G. Prediction of calving time in dairy cattle. *Animal Reproduction Science*, 187:37–46, 2017. ISSN 0378-4320. doi: https://doi.org/10.1016/j.anireprosci.2017.10.003.

García, R., Aguilar, J., Toro, M., Pinto, A., and Rodríguez, P. A systematic literature review on the use of machine learning in precision livestock farming. *Computers and Electronics in Agriculture*, 179:105826, 2020. ISSN 0168-1699. doi: https://doi.org/10.1016/j.compag.2020.105826.

Ghanem, M. E., Higuchi, H., Tezuka, E., Ito, H., Devkota, B., Izaike, Y., and Osawa, T. Mycoplasma infection in the uterus of early postpartum dairy cows and its relation to dystocia and endometritis. *Theriogenology*, 79(1):180–185, 2013. ISSN 0093-691X. doi: https://doi.org/10.1016/j.theriogenology.2012.09.027.

Giuliodori, M., Magnasco, R., Becu-Villalobos, D., Lacau-Mengido, I., Risco, C., and de la Sota, R. Metritis in dairy cows: Risk factors and reproductive performance. *Journal of Dairy Science*, 96(6):3621–3631, 2013. ISSN 0022-0302. doi: https://doi.org/10.3168/jds.2012-5922.

Goodfellow, I., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., Courville, A., and Bengio, Y. Generative adversarial nets. In Ghahramani, Z., Welling, M., Cortes, C., Lawrence, N., and Weinberger, K. Q., editors, *Advances in Neural Information Processing Systems*, volume 27. Curran Associates, Inc., 2014.

Goodfellow, I., Bengio, Y., and Courville, A. *Deep learning*. MIT press, 2016.

Graves, A., Mohamed, A.-r., and Hinton, G. Speech recognition with deep recurrent neural networks. In *2013 IEEE International Conference on Acoustics, Speech and Signal Processing*, pages 6645–6649, 2013. doi: 10.1109/ICASSP.2013.6638947.

Grzesiak, W., Lacroix, R., Wójcik, J., and Blaszczyk, P. A comparison of neural network and multiple regression predictions for 305-day lactation yield using partial lactation records. *Canadian Journal of Animal Science*, 83(2):307–310, 2003a.

Grzesiak, W., Wojcik, J., and Binerowska, B. Prediction of 305-day first lactation milk yield in cows with selected regression models. *Archives Animal Breeding*, 46(3):213–224, 2003b.

Grzesiak, W., Błaszczyk, P., and Lacroix, R. Methods of predicting milk yield in dairy cows—predictive capabilities of Wood's lactation curve and artificial neural networks (ANNs). *Computers and Electronics in Agriculture*, 54(2):69 – 83, 2006. ISSN 0168-1699.

He, K., Zhang, X., Ren, S., and Sun, J. Deep residual learning for image recognition. *ArXiv e-prints*, 2015.

Hogeveen, H., Kamphuis, C., Steeneveld, W., and Mollenhorst, H. Sensors and clinical mastitis—the quest for the perfect alert. *Sensors*, 10(9):7991–8009, 2010. ISSN 1424-8220.

Huxley, J. and Whay, H. Current attitudes of cattle practitioners to pain and the use of analgesics in cattle. *Veterinary Record*, 159(20):662–668, 2006.

Huzzey, J., von Keyserlingk, M., and Weary, D. Changes in feeding, drinking, and standing behavior of dairy cows during the transition period. *Journal of Dairy Science*, 88(7):2454–2461, 2005. ISSN 0022-0302. doi: https://doi.org/10.3168/jds.S0022-0302(05)72923-4.

Jensen, D. B., van der Voort, M., and Hogeveen, H. Dynamic forecasting of individual cow milk yield in automatic milking systems. *Journal of Dairy Science*, 101(11):10428–10439, 2018. ISSN 0022-0302. doi: https://doi.org/10.3168/jds.2017-14134.

Jensen, M. B. Behaviour around the time of calving in dairy cows. *Applied Animal Behaviour Science*, 139(3):195–202, 2012. ISSN 0168-1591. doi: https://doi.org/10.1016/j.applanim.2012.04.002.

Johanson, J. and Berger, P. Birth weight as a predictor of calving ease and perinatal mortality in holstein cattle. *Journal of dairy science*, 86(11):3745–3755, 2003.

Keceli, A. S., Catal, C., Kaya, A., and Tekinerdogan, B. Development of a recurrent neural networks-based calving prediction model using activity and behavioral data. *Computers and Electronics in Agriculture*, 170:105285, 2020. ISSN 0168-1699. doi: https://doi.org/10.1016/j.compag.2020.105285.

Kirkpatrick, M., Hill, W. G., and Thompson, R. Estimating the covariance structure of traits during growth and ageing, illustrated with lactation in dairy cattle. *Genetical Research*, 64(1):57–69, 1994.

Kliś, P., Piwczyński, D., Sawa, A., and Sitkowska, B. Prediction of lactational milk yield of cows based on data recorded by ams during the periparturient period. *Animals*, 11(2):383, 2021.

Lacroix, R., Wade, K., Kok, R., and Hayes, J. Prediction of cow performance with a connectionist model. *Transactions of the ASAE*, 38(5):1573–1579, 1995.

Lange, K., Fischer-Tenhagen, C., and Heuwieser, W. Predicting stage 2 of calving in Holstein-Friesian heifers. *Journal of Dairy Science*, 100(6):4847–4856, 2017. ISSN 0022-0302. doi: https://doi.org/10.3168/jds.2016-12024.

Laure, M. and Granier, A. The eu dairy sector: Main features, challenges and prospects, 2018.

Laven, R., Huxley, J., Whay, H., and Stafford, K. Results of a survey of attitudes of dairy veterinarians in New Zealand regarding painful procedures and conditions in cattle. *New Zealand Veterinary Journal*, 57(4):215–220, 2009. doi: 10.1080/00480169.2009.36904.

Lindblom, J., Lundström, C., Ljung, M., and Jonsson, A. Promoting sustainable intensification in precision agriculture: review of decision support systems development and strategies. *Precision Agriculture*, 18(3):309–331, 2017.

Lombard, J., Garry, F., Tomlinson, S., and Garber, L. Impacts of dystocia on health and survival of dairy calves. *Journal of Dairy Science*, 90(4):1751–1760, 2007. ISSN 0022-0302. doi: https://doi.org/10.3168/jds.2006-295.

Lormore, M. and Galligan, D. Economics of atypical lactation. *Journal of Dairy Science*, 84:E212–E215, 2001. ISSN 0022-0302. doi: https://doi.org/10.3168/jds.S0022-0302(01)70220-2.

Lovarelli, D., Bacenetti, J., and Guarino, M. A review on dairy cattle farming: Is precision livestock farming the compromise for an environmental, economic and social sustainable production? *Journal of Cleaner Production*, 262:121409, 2020. ISSN 0959-6526. doi: https://doi.org/10.1016/j.jclepro.2020.121409.

Macciotta, N. P., Dimauro, C., Rassu, S. P., Steri, R., and Pulina, G. The mathematical description of lactation curves in dairy cattle. *Italian Journal of Animal Science*, 10(4):e51, 2011.

Macciotta, N., Vicario, D., Pulina, G., and Cappio-Borlino, A. Test day and lactation yield predictions in Italian Simmental cows by ARMA methods. *Journal of Dairy Science*, 85(11):3107 – 3114, 2002. ISSN 0022-0302.

Mainau, E. and Manteca, X. Pain and discomfort caused by parturition in cows and sows. *Applied Animal Behaviour Science*, 135(3):241–251, 2011.

McArt, J., Nydam, D., and Oetzel, G. Epidemiology of subclinical ketosis in early lactation dairy cattle. *Journal of Dairy Science*, 95(9):5056–5066, 2012. ISSN 0022-0302. doi: https://doi.org/10.3168/jds.2012-5443.

McGuirk, B. J., Forsyth, R., and Dobson, H. Economic cost of difficult calvings in the United Kingdom dairy herd. *Veterinary Record*, 161(20):685–687, 2007. doi: https://doi.org/10.1136/vr.161.20.685.

Mee, J. F., Sánchez-Miguel, C., and Doherty, M. Influence of modifiable risk factors on the incidence of stillbirth/perinatal mortality in dairy cattle. *The Veterinary Journal*, 199(1):19–23, 2014. ISSN 1090-0233. doi: https://doi.org/10.1016/j.tvjl.2013.08.004.

Meyer, C., Berger, P., and Koehler, K. Interactions among factors affecting stillbirths in Holstein cattle in the United states. *Journal of Dairy Science*, 83(11):2657–2663, 2000. ISSN 0022-0302. doi: https://doi.org/10.3168/jds.S0022-0302(00)75159-9.

Miedema, H. M., Cockram, M. S., Dwyer, C. M., and Macrae, A. I. Behavioural predictors of the start of normal and dystocic calving in dairy cows and heifers. *Applied Animal Behaviour Science*, 132(1):14–19, 2011a. ISSN 0168-1591. doi: https://doi.org/10.1016/j.applanim.2011.03.003.

Miedema, H. M., Cockram, M. S., Dwyer, C. M., and Macrae, A. I. Changes in the behaviour of dairy cows during the 24h before normal calving compared with behaviour during late pregnancy. *Applied Animal Behaviour Science*, 131(1): 8–14, 2011b. ISSN 0168-1591. doi: https://doi.org/10.1016/j.applanim.2011.01.012.

Murphy, K. P. *Machine learning: a probabilistic perspective*. MIT press, 2012.

Murphy, M., O'Mahony, M., Shalloo, L., French, P., and Upton, J. Comparison of modelling techniques for milk-production forecasting. *Journal of Dairy Science*, 97(6):3352–3363, 2014. ISSN 0022-0302. doi: https://doi.org/10.3168/jds.2013-7451.

Njubi, D., Wakhungu, J., and Badamana, M. Use of test-day records to predict first lactation 305-day milk yield using artificial neural network in kenyan holstein–friesian dairy cows. *Tropical animal health and production*, 42(4):639–644, 2010.

Norman, H., Hutchison, J., and Miller, R. Use of sexed semen and its effect on conception rate, calf sex, dystocia, and stillbirth of holsteins in the united states. *Journal of Dairy Science*, 93(8):3880–3890, 2010. ISSN 0022-0302. doi: https://doi.org/10.3168/jds.2009-2781.

Norton, T. and Berckmans, D. Developing precision livestock farming tools for precision dairy farming. *Animal Frontiers*, 7(1):18–23, 01 2017. ISSN 2160-6056. doi: 10.2527/af.2017.0104.

OECD, Food, and of the United Nations, A. O. *OECD-FAO Agricultural Outlook 2020-2029*. 2020. doi: https://doi.org/https://doi.org/10.1787/1112c23b-en.

Ouellet, V., Vasseur, E., Heuwieser, W., Burfeind, O., Maldague, X., and Charbonneau, É. Evaluation of calving indicators measured by automated monitoring devices to predict the onset of calving in Holstein dairy cows. *Journal of Dairy Science*, 99(2):1539–1548, 2016. ISSN 0022-0302. doi: https://doi.org/10.3168/jds.2015-10057.

Redmon, J., Divvala, S., Girshick, R., and Farhadi, A. You only look once: Unified, real-time object detection. *ArXiv e-prints*, 2016.

Rojas-Downing, M. M., Nejadhashemi, A. P., Harrigan, T., and Woznicki, S. A. Climate change and livestock: Impacts, adaptation, and mitigation. *Climate Risk Management*, 16:145–163, 2017. ISSN 2212-0963. doi: https://doi.org/10.1016/j.crm.2017.02.001.

Rosa, G. J. M. Grand challenge in precision livestock farming. *Frontiers in Animal Science*, 2, 2021. ISSN 2673-6225. doi: 10.3389/fanim.2021.650324.

Rose, D. C. and Chilvers, J. Agriculture 4.0: Broadening responsible innovation in an era of smart farming. *Frontiers in Sustainable Food Systems*, 2, 2018. ISSN 2571-581X. doi: 10.3389/fsufs.2018.00087.

Rutten, C., Kamphuis, C., Hogeveen, H., Huijps, K., Nielen, M., and Steeneveld, W. Sensor data on cow activity, rumination, and ear temperature improve prediction of the start of calving in dairy cows. *Computers and Electronics in Agriculture*, 132:108–118, 2017. ISSN 0168-1699. doi: https://doi.org/10.1016/j.compag.2016.11.009.

Sharma, A. K., Sharma, R., and Kasana, H. Prediction of first lactation 305-day milk yield in Karan Fries dairy cattle using ANN modeling. *Applied Soft Computing*, 7(3):1112 – 1120, 2007. ISSN 1568-4946.

Speroni, M., Malacarne, M., Righi, F., Franceschi, P., and Summer, A. Increasing of posture changes as indicator of imminent calving in dairy cows. *Agriculture*, 8(11), 2018. ISSN 2077-0472.

Tenhagen, B.-A., Helmbold, A., and Heuwieser, W. Effect of various degrees of dystocia in dairy cattle on calf viability, milk production, fertility and culling. *Journal of Veterinary Medicine Series A*, 54(2):98–102, 2007. doi: https://doi.org/10.1111/j.1439-0442.2007.00850.x.

Vasconcelos, J., Martins, A., Petim-Batista, M., Colaço, J., Blake, R., and Carvalheira, J. Prediction of daily and lactation yields of milk, fat, and protein using an autoregressive repeatability test day model. *Journal of Dairy Science*, 87(8): 2591 – 2598, 2004. ISSN 0022-0302.

Werkheiser, I. Precision livestock farming and farmers' duties to livestock. *Journal of Agricultural and Environmental Ethics*, 31(2):181–195, 2018.

Wilmink, J. Adjustment of test-day milk, fat and protein yield for age, season and stage of lactation. *Livestock Production Science*, 16(4):335 – 348, 1987. ISSN 0301-6226.

Wilson, D., González, R., Hertl, J., Schulte, H., Bennett, G., Schukken, Y., and Gröhn, Y. Effect of clinical mastitis on the lactation curve: A mixed model estimation using daily milk weights. *Journal of Dairy Science*, 87(7):2073–2084, 2004. ISSN 0022-0302. doi: https://doi.org/10.3168/jds.S0022-0302(04)70025-9.

Wittrock, J., Proudfoot, K., Weary, D., and Von Keyserlingk, M. Metritis affects milk production and cull rate of holstein multiparous and primiparous dairy cows differently. *Journal of dairy science*, 94(5):2408–2412, 2011.

Wood, P. Algebraic model of the lactation curve in cattle. *Nature*, 216(5111): 164–165, 1967.

Zehner, N., Niederhauser, J. J., Schick, M., and Umstatter, C. Development and validation of a predictive model for calving time based on sensor measurements of ingestive behavior in dairy cows. *Computers and Electronics in Agriculture*, 161:62–71, 2019. ISSN 0168-1699. doi: https://doi.org/10.1016/j.compag.2018.08.037. BigData and DSS in Agriculture.

2

Leveraging latent representations for milk yield prediction and interpolation using deep learning [1]

2.1 Abstract

In this study, we propose a lactation model that estimates the daily milk yield by using autoencoders to generate a latent representation of all milk yields observed during the entire lactation cycle, irrespective of the length of the time interval between the different measurements. More specifically, we propose a sequential autoencoder (SAE) to process the sequential data, extract and decode the low-dimensional representations and generate the milk yield sequences. The SAE is compared with a more traditional multilayer perceptron model (MLP) which uses herd and parity information and lagged milk yields as input. Results show that incorporating the recorded daily milk yields, lactation number, herd statistics as well as reproduction and health events the cow encountered during the lactation cycle results in the most qualitative latent representations. Moreover, by leveraging these low-dimensional encodings, the SAE reconstructed the entire milk yield curve with a higher accuracy than the MLP. Hence, we present a framework that is able to infer missing milk yields along the entire lactation curve which facilitates

[1] Based on: Liseune, A., Salamone, M., Van den Poel, D., Van Ranst, B., & Hostens, M. (2020). Leveraging latent representations for milk yield prediction and interpolation using deep learning. *Journal of Computers and Electronics in Dairy Science.*

selection and culling decisions as well as the estimation of future earnings and costs. Furthermore, the model allows farmers to enhance their animal monitoring systems as it incorporates the sequence of health and reproduction events to forecast the cow's future productivity.

2.2 Introduction

Lactation models are one of the most important analytical tools in the dairy cattle industry as this enables the farmer to get a projection of the herd's total productive capacity and its future earnings (Ehrlich et al., 2011; Grzesiak et al., 2003; Sanzogni and Kerr, 2001). Moreover, by estimating each cow's expected milk yield, the required feed intake, energy and protein requirement and plant utilization can be assessed, allowing the farmer to forecast its costs (Grzesiak et al., 2006; Murphy et al., 2014). In addition, lactation models facilitate the selection process as it enables the timely identification of the most productive females as well as superior bulls based on the analysis of the total productivity of its offspring (Lacroix et al., 1995). Furthermore, early detection of unproductive animals supports more informed culling decisions (Njubi et al., 2010). Finally, by comparing a cow's predicted milk yields with the cow's actual lactation curve, diseases such as mastitis as well as other factors affecting the animal's health could be detected more accurately and hence facilitate more enhanced animal monitoring systems.

In early research, lactation curves were modeled by mathematical functions describing the general pattern of the lactation cycle. These were often characterized by an initial steep increase until a peak yield, followed by a longer and more gradual decline. Wood (1967) for example proposed a gamma function expressing the relationship between a specific moment in time in the lactation cycle and the associated expected milk yield by means of three parameters. Later, several extensions to this model were proposed in order to fit a wider range of possible shapes (Ali and Schaeffer, 1987; Ehrlich et al., 2011; Wilmink, 1987). The general purpose of these models was to describe the lactation curves of homogeneous groups of animals by its deterministic components as individual data on animals was often still lacking (Macciotta et al., 2011). As a result, the expected yield of an animal was entirely determined by the average curve of the group to which the animal could be assigned.

As time passed however, more data on individual animals became available and the need to monitor individual animals increased. As a result, more complex models such as polynomial and spline regressions were proposed that, in addition to the moment in time, also took into account individual features such as age, season of calving and lactation number (Grzesiak et al., 2003). Later, with the advent of Artificial Intelligence (AI), several other attempts were made to predict milk yields by making use of neural networks. Lacroix et al. (1995) were the first to successfully train a multilayer perceptron model (MLP) to predict the 305d cumulative milk yield of a cow. In later studies, this model was improved by using more sophisticated data preprocessing techniques (Lacroix et al., 1997) and by constructing multiple networks, each assigned to a specific task (Salehi et al., 1998). Furthermore, neural networks have been used to predict the 305d cumulative milk yield of the first lactation cycle (Njubi et al., 2010; Sharma et al., 2007) as well as to forecast a herd's total production (Sanzogni and Kerr, 2001). Other studies on the other hand investigated the use of neural networks in modeling the entire lactation curve rather than predicting the cumulative productivity (Grzesiak et al., 2006).

In most of these models however, the predicted milk yield does not take into account the sequence of previously produced milk yields and is mainly dependent on a couple of environmental factors as well as animal characteristics which remain constant during the entire lactation cycle. As a result, once these models are applied, the predictions remain constant during the entire lactation cycle, regardless of newly produced yields observed after the moment of prediction. Yet, common environmental factors such as weather, nutrition and herd management as well as an animal's repeatability of yields result in covariances between adjacent milk yields (Ali and Schaeffer, 1987). Therefore, some research presented frameworks in which correlations between consecutive milk yields were taken into account. Macciotta et al. (2002) and Vasconcelos et al. (2004) for example used Autoregressive (AR) models which predicted a test-day yield (TD) based on a fixed number of lagged TD records. Murphy et al. (2014) on the other hand applied a MLP which incorporated lagged features as well as yields to forecast the entire herd's production in the subsequent time step.

Nevertheless, these models were all based on the assumption of the prediction being dependent on a sequence of a fixed number of historical milk yields measured at constant time intervals. Yet, in reality, variable intervals between measured observations often occur due to various reasons, e.g. defective recording machines, animal treatment and variable recording schemes between different herds. Furthermore, as the main focus of these models was to forecast yields based on historical observations, they only partially took into account the existing correlations between the prediction and all the adjacent recordings. As a result, these models lacked the ability to exploit past as well as future data to interpolate gaps of missing information across the lactation curve.

In this research, we propose a lactation model in which missing milk yields along the lactation curve are dynamically updated as soon as new information in the corresponding lactation cycle becomes available. In addition to using all the recorded milk yields along the lactation curve, the lactation number, herd statistics and a cow's sequence of reproduction and health events it encountered during the lactation cycle are used to infer the missing parts of the lactation curve. More specifically, a sequential autoencoder (SAE) is trained to generate a latent representation of the entire quantity of information available in a certain lactation cycle, and uses this low-dimensional encoding to reconstruct the entire milk yield curve. As a result, the estimate of a milk yield at a specific point in time is generated by making use of all present information before and after the moment of prediction, regardless of the length of the time intervals between the different data observations. We believe that the presented framework can add value to dairy farmers in three different ways. First, the presented framework can be used to continuously monitor their cattle by comparing the true milk yield with the expected lactation curve. This helps farmers to detect unforeseen milk losses, which may indicate health-related problems. Moreover, by incorporating the impact of reproduction and health events, the framework can generate more realistic estimates of the cow's future productivity. Second, predicting milk yields allows dairy farmers to obtain forecasts of their herd's total production and hence of their future revenues as well as costs such as required feed intake as well as energy consumption. Third, the framework supports breeding and culling decisions by providing timely predictions of a cow's productivity.

2.3 Materials and Methods

2.3.1 Data

The data used in this study was collected from 104 different farms between 2013 and 2019. Farms were equipped with different Herd Management Systems (HMS) to record the animals lifetime records and collect milkmeter information. The HMS data sources were streamed and standardized using a cloud-based dairy analysis application (www.mmmooogle.com). In total, 59122 lactations of 35133 distinct cows have been collected, with an average of 216 recorded daily milk yields per lactation. In addition, 304742 recordings of 13 unique events were collected. Based on the milk and event recordings, the HMS also calculated several herd statistics per parity. In addition to metrics such as the average milk produced and the average lactation duration, the herd statistics also included a score to identify the average recording quality of the lactation cycles. This score was developed by the platform and indicates how much the sequence of events in certain lactation cycle deviates from the standard sequence of events during a lactation: Calving, Heat, Breeding, Pregnancy Positive, Calving. In the case of a Pregnancy Negative event, the Heat and Breeding events are repeated in the standard sequence. Each lactation cycle is initialized with a score of 1 and for each missing event in sequence the score is multiplied by 0.5. The resulting product per lactation is then subtracted from 1 and results in a score between 0 and 1, where 0 represents a standard sequence of events. Averaging this score over all the lactations generated in a specific herd produces the herd's average sequence quality score. The final dataset was obtained by extracting milk yields and events recorded during the first 305 days of each lactation cycle, with zero values representing missing milk yields and a special PAD symbol to indicate if no events occurred that day. Events that did occur in the validation or test set but not in the training set were labeled as UNKNOWN and were considered as a rare but unknown event. Subsequently, each record was augmented with the corresponding herd's statistics and lactation number, resulting in a two-dimensional sequence of yields and events and a static vector comprising the lactation number and herd statistics per observation. **Table** 3.1 gives an overview of all the independent variables used in this study.

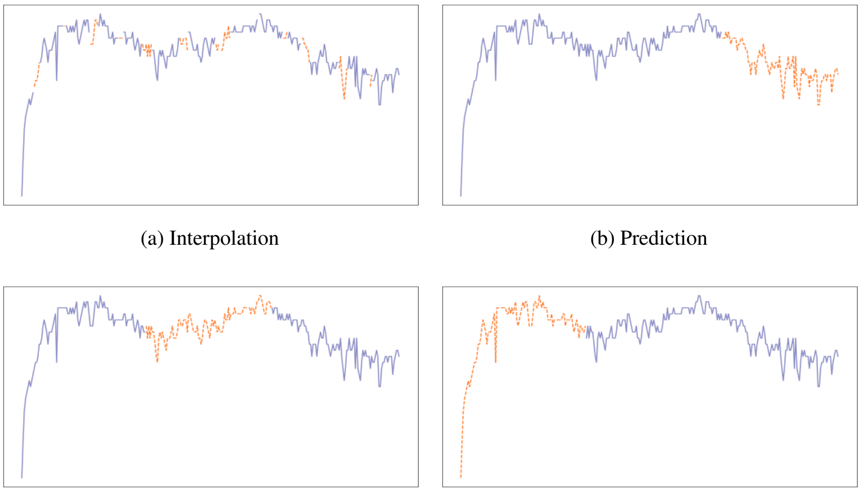
Variable Group	**Dimension**	**Variable Name**
Milk Yields	1 x 305	
		Milk Yield
Herd Statistics	10 x 1	
		Avg 21d Milk
		Avg 75d Milk
		Avg 305d Milk
		Avg Milk
		Avg Days Dry
		Avg Days Open
		Avg Days Pregnant
		Avg Days In Milk
		Avg Calving Interval
		Avg Sequence Quality
Events	1 x 305	
		Mastitis
		Abort
		Breeding
		Stop Breeding
		Pregnancy Negative
		Pregnancy Positive
		Calving
		Disease
		Died
		Heat
		Cull
		Dryoff
		PAD
		UNKNOWN
Parity	1 x 1	
		Lactation Number

Table 2.1: Independent variables used in this study

After normalizing the milk yields and herd statistics, the data was partitioned in three sets. A training set for training the models and a validation set for hyperparameter selection. A test set was used for obtaining an unbiased evaluation metric of the final model. This validation setup was preferred over cross-validation, which commonly requires a repeated model training procedure (Li et al., 2020). Therefore, using cross-validation for deep learning models often becomes practically infeasible, as the computational burden linearly grows with the number of folds. Since the validation and test set were both used for model evaluation, a stratified sampling procedure was used such that only records with complete milk yield curves were assigned to these sets (Parsons, 2017). From the entire collection of observations containing no missing milk yields, 1000 observations were randomly assigned to the validation as well as the test set. The remaining observations with complete information were assigned to the training set together with the collection of observations containing incomplete milk yield curves. However, in order to train a model that is able to reconstruct an entire milk yield curve given a random subset of data gathered in the corresponding lactation cycle, the training set should contain a representative sample of all possible configurations of information available in a lactation cycle. Hence, instead of feeding the training data directly into the model, missing values were randomly inserted in the sequential data of each observation every time it was passed to a training iteration. More specifically, in addition to removing each daily observation with a certain probability, a complete window of records across the lactation cycle determined by a randomly chosen start and end day could be set to missing. Hence, as depicted by **Fig.** 2.1, one of four possible missing data imputation patterns was applied to each observation every training iteration: interpolation, prediction, missing window or backtracking. For the prediction, backtracking and missing window imputation patterns, the start and end days are randomly sampled from the lactation cycle, with each day having the same probability of being sampled. For the interpolation imputation pattern, a milk yield is randomly dropped according to a probability that is determined for each lactation curve separately, such that the milk yields of lactation curves with few missing values have a higher probability of being removed, while the milk yields of lactation curves with many missing values have a lower probability of being removed. In particular, the probability of each milk yield being removed from a certain lactation cycle is determined by the following formula:

$$P_{drop} = 0.7 * (s/305)$$

with P_{drop} being the probability of a milk yield being set to missing and s being the number of non-missing milk yields in the lactation curve. As a result, the drop probability ranges between 0 and 0.7, with 0 for lactation curves with no recorded milk yields and 0.7 for complete lactation curves. Hence, since each observation is randomly injected with missing values every training iteration, the model is constantly fed with new observations. This reduces the chance of overfitting and hence improves the model's generalization capacity. In addition, the model is forced to learn multiple tasks such as prediction and interpolation simultaneously, resulting in a more versatile lactation model.



(a) Interpolation (b) Prediction (c) Missing Window (d) Backtracking

Figure 2.1: Missing data imputation patterns. Purple solid line = observed milk yields, orange dotted line = injected missing milk yields

2.3.2 Sequential Autoencoder

2.3.2.1 Convolutional Neural Network

Convolutional neural networks (CNN) are a type of deep learning algorithms specifically designed to process grid-like data and have been responsible for major breakthroughs in object detection and classification (He et al., 2016; Szegedy et al., 2015). Yet, while CNNs were initially developed for computer vision applications, recent research has shown that these type of architectures can also be of great value for time series analysis given their ability to represent a chronological sequence by a set of automatically extracted features (Zhao et al., 2017). In general, a CNN consists out of a sequence of blocks with each block typically comprising a convolutional layer for feature extraction and a pooling stage for downsampling and obtaining the most salient elements. More specifically, **Fig.** 3.1 shows how a CNN architecture is constructed for a time series of length S and width N. A kernel of size $K < S$ and width N is slided from the beginning to the end of the sequence. Each time the kernel is shifted one position, the kernel weights are multiplied with the elements of the sequence that are covered by the kernel at that point. Subsequently, a non-linear activation function, such as rectified linear units (ReLu), is applied to

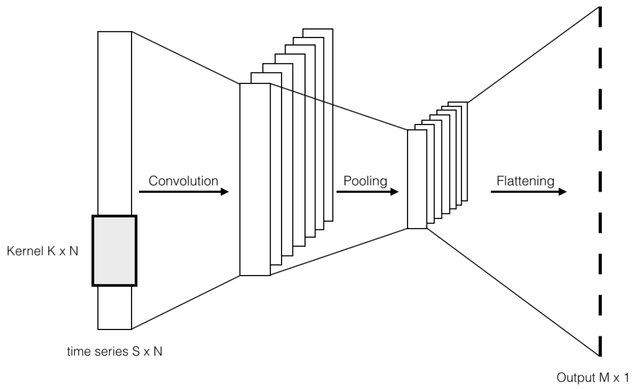


Figure 2.2: Convolutional neural netowork for time series

the sum of the outputs of the multiplication and hence results in a new time series of the features extracted by that kernel. The feature map is then downsampled by a pooling layer, which summarizes the presence of the feature in every specific time window. Finally, the pooling layer's output is flattened to obtain a feature vector summarizing the entire time series and can be used for upstream tasks. By altering the number of kernels or the kernel size, different properties of the time series can be extracted, while adding convolutional blocks allows to learn more complex patterns. In this research, a CNN was applied on the sequential data comprising the milk yields and events encountered by a cow during a specific lactation cycle. However, as the events were represented by a sequence of symbols, an embedding matrix was used to convert every event into a numerical representation, defined in a continuous vector space. More specifically, by initializing a random matrix with values from a normal distribution of size 14 x k, each of the 14 events occurring in the sequence was replaced by the corresponding matrix row to produce a 305 x k sequence. Yet, instead of keeping the embedding vectors fixed, the matrix values were considered as network parameters and were updated during training in order to obtain the most optimal embeddings. Additionally, instead of applying a ReLu activation function which is expressed as follows:

$$f = max(0, \boldsymbol{x})$$

with $\boldsymbol{x}$ being the feature map produced by a specific kernel, a Leaky ReLu function was used to obtain the non-linear activations and is defined as follows:

$$f = max(\alpha\boldsymbol{x}, \boldsymbol{x})$$

with α being a constant and often set to small values. While the output of the ReLu function equals 0 for every negative value of the input $\boldsymbol{x}$, the Leaky ReLu function contains a positive slope for the entire range of x which can facilitate gradient-based optimization. The feature maps were downsampled by making use of max pooling layers which extract the maximum response value of a specific feature in a certain time interval. The output of the last convolutional block was flattened and summarized the sequential data gathered during an entire lactation cycle.

2.3.2.2 Autoencoder

While MLPs and CNNs are mainly used for supervised learning tasks, autoencoders (AE) are a type of neural networks specifically designed to compress the input data in an unsupervised manner (Hinton and Zemel, 1994). In particular, the AE takes an input $\boldsymbol{x} \in \mathbb{R}^d$ and maps it into an encoded representation $\boldsymbol{z} \in \mathbb{R}^l$ with $l < d$ by means of an encoder parameterized by $\mathbf{W}_e$. A decoder defined by $\mathbf{W}_d$ then converts the latent representation $\boldsymbol{z}$ back into a reconstruction of the input $\widetilde{\boldsymbol{x}} \in \mathbb{R}^d$. The entire model is trained by minimizing the reconstruction loss for all observations:

$$\begin{aligned} loss &= \sum_{i=1}^{n} L(\boldsymbol{x}^{(i)}, \widetilde{\boldsymbol{x}}^{(i)}) \\ &= \sum_{i=1}^{n} L(\boldsymbol{x}^{(i)}, f(\mathbf{W}_d; \boldsymbol{z}^{(i)}) \\ &= \sum_{i=1}^{n} L(\boldsymbol{x}^{(i)}, f(\mathbf{W}_d; f(\mathbf{W}_e; \boldsymbol{x}^{(i)}))) \end{aligned}$$

In addition to extracting the most salient features by applying dimensionality reduction, an AE can also learn to impute missing values by corrupting the input before it is passed to the encoder (Bengio et al., 2013). More specifically, instead of minimizing the original reconstruction loss, the AE is now trained to undo the corruption of the original input:

$$loss = \sum_{i=1}^{n} L(\boldsymbol{x}^{(i)}, f(\mathbf{W}_d; f(\mathbf{W}_e; \hat{\boldsymbol{x}}^{(i)})))$$

with $\hat{\boldsymbol{x}}$ being a corrupted version of the original input $\boldsymbol{x}$. Hence, by randomly injecting missing values into the original observations and forcing the dimensionality of the latent representation to be smaller than that of the input, the model learns to impute the missing values by extracting the most salient features from the incomplete information. Hence, in this research, an AE was applied on the features extracted by the CNN as discussed in the previous section. Since these features represent a summary of the sequential information still present after applying one of the four possible imputation schemes as discussed in Section 2.3.1, the AE learns to extract the most prominent traits of the lactation curve from the available information and uses this latent representation to reconstruct the complete feature set. In order to get a more informative encoding however, other features such as the herd statistics and lactation number were added to the encoder's input, as these could partially explain the shape of the lactation curve. The AE's encoder consisted out of a sequence of layers decreasing in size, with each hidden layer comprising a linear transformation followed by a Leaky ReLu activation. The output layer was provided with a Sigmoid function in order to obtain latent features distributed between 0 and 1. The decoder was the exact image mirror of the encoder except for the output layer. The latter had the same dimensionality as the CNN's output, since we were only interested in the reconstruction of the lactation curve rather than the entire feature set gathered during the lactation cycle.

2.3.2.3 Deconvolutional Neural Network

As discussed in the previous section, the AE's output comprises a feature vector of the same dimensionality as that of the CNN's output. In other words, it represents the reconstruction of several abstract features extracted from the lactation curve rather than the sequence of milk yield itself. Hence, in order to convert this static feature representation back into its corresponding lactation curve, a deconvolutional neural network (DNN) is used to reverse the CNN's operations (Zeiler et al., 2010). However, since the initial input of the CNN comprises the sequence of milk yields as well as embedded events, the DNN's final layer was forced to return a one-dimensional sequence instead of producing multivariate time series. The other DNN's layers on the other hand were the exact inversions of the corresponding CNN's layers. The combination of the CNN, AE and DNN results in the sequential autoencoder (SAE).

2.3.2.4 Training

The SAE was trained by applying the backpropagation algorithm (Rumelhart et al., 1986). In this algorithm, the inputs are first propagated forward through the network to produce an output. A loss is then calculated by comparing the outputs with the true labels. Next, the gradients with respect to weights are computed by propagating the errors backward through the network. Finally, a gradient-based optimization algorithm such as stochastic gradient descent or Adam (Kingma and Ba, 2014) is used to update the parameters. However, as the training observations included complete as well as incomplete lactation curves, a loss calculated between the reconstruction and the true input over the entire lactation cycle would be biased. The reconstruction loss therefore only took into account the errors between the milk yields which were actually recorded and the corresponding reconstructions.

Several model hyperparameter settings were assessed by making use of an early stopping procedure in which a model was trained as long as its performance on the validation set increased. The next configuration of hyperparameters was determined by a Bayesian optimization procedure. In this algorithm, the exploration of a parameter setting with uncertain results is traded off against the exploitation of a point in parameter space with high model performance. In addition to the parameters defining the model's architecture, the inclusion of the event, herd and parity information was also considered as hyperparameter in order to obtain the most discriminative group of features. Furthermore, the dimensionality of the encoder's output was also set as a hyperparameter in order to find the optimal encodings for the lactation curves. The best performing model was retrained on the combination of the training and validation set and was subsequently evaluated on the test set. In order to ensure a wide variety of evaluation observations and a fair model selection procedure, four fixed and significantly different variations of each imputation scheme (see Section 2.3.1) were applied to each validation and test record. As a result, the validation and test set both comprised 16000 records. An overview of the entire methodology of this study is illustrated by **Fig.** 4.4.

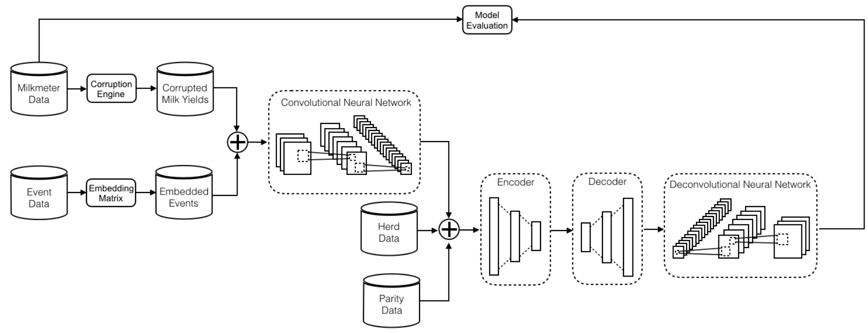


Figure 2.3: Schematic overview of the methodology used in this study

2.3.3 Multilayer Perceptron Model

In order to benchmark the SAE's performance, a standard MLP was trained to predict milk yields. In contrast to the SAE, an MLP is not designed to process sequential data. It is a supervised learning model constructed to infer a relationship between an input and a corresponding output. Hence, in order to construct the MLP's observations, the sequences of milk yields presented in **Table** 3.1 were transformed. In particular, for a certain day t in a lactation cycle l with a non-missing milk yield, one observation was generated with the output y_{l_t} being the milk yield MY_{l_t} and the input x_{l_t} consisting of the days in milk t and a number of k

preceding milk yields $MY_{l_{t-1}}$, $MY_{l_{t-2}}$, ..., $MY_{l_{t-k}}$. In addition, the corresponding parity and herd information were added to the input. This resulted in a training set comprising 9086759 observations and a validation as well as test set consisting out of 305000 observations. In correspondence to the SAE's encoder, the MLP was constructed out of a sequence of layers with each layer comprising a linear transformation followed by a Leaky ReLu activation function. In addition to the model's architecture, the number of lags was also considered as hyperparameter. A random search in hyperparameter space determined the MLP configuration in each training iteration. The MLP weights were updated by applying the Adam optimization and backpropagation algorithms (Kingma and Ba, 2014; Rumelhart et al., 1986). In order to avoid overfitting, the early stopping procedure was applied in which the MLP was trained until its validation performance started to degrade.

2.3.4 Model Evaluation

The model performance was assessed by four metrics commonly used for evaluating lactation models.

Coefficient of Determination (R^2). The R^2 value is defined as the proportion of the variance of the dependent variable explained by the independent variables and indicates the model's goodness of fit. More specifically, if y_i denotes the true value for the ith observation and $\hat{y}_i$ the corresponding prediction, then the R^2 is given by

$$1 - \frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{\sum_{i=1}^{n}(y_i - \bar{y})^2}$$

where $\bar{y}$ represents the mean of the dependent variable.

Pearson Correlation Coefficient (ρ). The ρ measures the linear correlation between two variables. It ranges between -1 and 1, with 1 meaning that both variables are perfectly positively correlated, -1 meaning that the variables are perfectly negatively correlated and 0 meaning that both variables are not linearly related.

$$\frac{\mathrm{cov}(y, \hat{y})}{\sigma(y)\sigma(\hat{y})}$$

Root Mean Squared Error (RMSE). The RMSE is a widespread measure for evaluating regression models and is equal to the the square root of the average of squared differences between the predictions and actual observations.

$$\frac{1}{n}\sqrt{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}$$

Since the errors are squared, the model will be penalized more for making predictions that differ greatly from the corresponding true value.

Mean Absolute Percentage Error (MAPE). While the analysis of the RMSE greatly depends on the scale of the dependent variable, the MAPE statistic measures how much the model's predictions deviate from the corresponding true value on average and hence allows for a more easy interpretation.

$$\frac{1}{n}\sum_{i=1}^{n}\frac{|y_i - \hat{y}_i|}{y_i} \times 100\%$$

The model's general performance was assessed by calculating the metrics on all the test observations. In addition, the model was evaluated on every imputation task by obtaining the metrics on each subset of samples that were generated by a specific imputation pattern. For the model's predictive performance, the metrics were obtained for different windows of available data. Furthermore, milk yields in the window of available data were randomly set to missing according to a pre-specified sampling rate in order to inspect the model's robustness towards variable lengths of time intervals between different recordings.

2.3.5 Variable Importance

In order to assess the importance of each feature towards inferring missing milk yields, the permutation variable importance (VI) score was calculated (Breiman, 2001). This score measures how much the model's error increases on average when a specific feature is randomly permutated:

$$VI_k = \frac{1}{n}\sum_{i=1}^{n}\frac{\widetilde{e}_{k,i}}{e}$$

with e being the model's error with complete information and $\widetilde{e}_{k,i}$ being the model's error obtained for the ith random permutation of the feature k. By randomly permutating multiple features simultaneously, the VI score of an entire feature group can be calculated. The VI score, expressed in terms of the relative increase in RMSE, was calculated for each entire group of features as well as for its individual members.

2.3.6 Programming Tools

All data processing and analyses were done in Python 2.7 (Python Software Foundation, https://www.python.org/) with the add-on packages Pandas (pandas development team, 2020) and NumPy (Harris et al., 2020) for data preprocessing, scikit-learn (Pedregosa et al., 2011) for machine learning modeling and model evaluation, TensorFlow (Abadi et al., 2015) and Keras (Chollet et al., 2015) for deep learning modeling and Matplotlib (Hunter, 2007) and seaborn (Waskom, 2021) for data visualization.

2.4 Results

2.4.1 Model Selection

For the SAE's early stopping procedure, each model configuration was evaluated in terms of the RMSE on the validation set every 5000 training iterations with a batch size of 32. The model's updated weights were saved every time the validation performance increased. Model training was terminated once the evaluation metric obtained on the validation set did not improve for 5 consecutive times. The SAE's next configuration of hyperparameters was determined by a Bayesian optimization algorithm which was initialized by evaluating 5 randomly sampled model configurations. The best performing model included a CNN that was trained on the milk yields as well as events, with the events being embedded into a 5-dimensional vector space. The CNN's architecture consisted out of 6 convolutional blocks, each one obtained by means of kernels with size 3, same padding and a Leaky ReLu function for non-linear activation with α being 0.2. The first two convolutional blocks applied 16 different kernels, while 32 kernels were used in block 3 and 4. The last two convolutional blocks both used 64 kernels. In order to extract the most salient elements, a max pooling layer of size 4 was applied to the output of the first block, while a max pooling layer of size 2 was used for block 3 and block 5. The best results were obtained by including the lactation number and herd statistics in the encoder's input. The MLP encoder contained 4 hidden layers of sizes 200, 100, 100 and 50 respectively and an output layer of size 20. Except for the output layer, the MLP decoder was the exact mirror image of the encoder. A Leaky ReLu activation function with α being 0.2 was used for each hidden layer in both the MLP encoder and decoder. The DNN consisted out of the exact inversions of every CNN's hidden layer. The DNN's output layer comprised one kernel of size 3 and a Sigmoid activation function. In order to match the input shape, a valid padding was used and the last observation of the output sequence was discarded. As a result, the DNN then returned a sequence of 305 values ranging between 0 and 1. The best performing MLP configuration was obtained by a random search over the hyperparameter space and by applying an early stopping procedure in which model training was completed when the validation RMSE did not improve after 3 entire training epochs. Milk yields were most accurately predicted by taking into account 15 lags and with the MLP's architecture consisting out of 3 layers with 100, 50 and 50 neurons respectively. Each layer consisted out of a linear transformation followed by a Leaky ReLu activation function with α being 0.1. Training was completed after 6 training epochs. The SAE and MLP were both trained by making use of the Adam optimization algorithm and an initial learning rate of 0.0001.

2.4.2 Model Performance

The goodness of fit of the MLP and the SAE on the training set is presented in **Table** 2.2. The R^2 and the ρ of the SAE on the entire training set were 0.84 and 0.91 respectively, while the MLP achieved an R^2 and a ρ of 0.83 and 0.91. The SAE's and MLP's goodness of fit however, was particularly lower for the curves of the first lactation than for the other parities.

	$\mathbf{R^2}$		ρ	
Parity	MLP	SAE	MLP	SAE
1	0.75	0.76	0.87	0.87
2	0.83	0.83	0.91	0.91
3	0.83	0.84	0.91	0.92
4	0.84	0.85	0.92	0.92
5	0.84	0.85	0.92	0.92
6	0.84	0.84	0.91	0.92
All	0.83	0.84	0.91	0.91

Table 2.2: Goodness of fit of the MLP and SAE on the training set

The performance of the models on the test set in general as well as for each particular imputation task as discussed in Section 2.3.1 is presented in **Table** 4.3. Results show that the SAE performed better than the MLP in general as well as on every task in terms of every evaluation metric. In particular, while the SAE's maximum MAPE was 17%, the MLP's MAPE ranged from 12% to 21%. As expected, the MLP performed particularly worse on the backtracking task than the SAE, while the interpolation and missing window tasks were best accomplished by both models. While the RMSE of the SAE for the prediction and backtracking tasks comprised 5.68 and 5.97, the RMSE for the missing window and interpolation tasks comprised 4.76 and 4.06 respectively. In terms of all the other metrics, the SAE also performed better on the missing window and interpolation tasks with the latter obtaining the overall highest scores. For the prediction and backtracking tasks, all the evaluation metrics achieved by the SAE did not differ considerably, except for the MAPE. This can be explained due to a higher average value of the missing milk yields for the backtracking task than for the prediction task. The SAE's reconstruction of the lactation curve for each fixed variation of every imputation pattern applied on a test observation is depicted by **Fig.** 2.4.

Table 2.4 presents the predictive performance of the SAE for daily as well

	MLP				SAE			
Imputation method	**RMSE**	**MAPE**	$\mathbf{R^2}$	ρ	**RMSE**	**MAPE**	$\mathbf{R^2}$	ρ
Prediction	6.77	0.18	0.48	0.76	5.68	0.17	0.63	0.80
Backtracking	8.16	0.21	0.35	0.62	5.97	0.15	0.65	0.81
Missing window	6.29	0.15	0.56	0.78	4.76	0.11	0.75	0.87
Interpolation	4.62	0.12	0.79	0.89	4.06	0.10	0.84	0.92
All	6.52	0.17	0.57	0.77	5.15	0.13	0.73	0.86

Table 2.3: Performance of the MLP and SAE on the test set in general as well as for the different missing data imputation patterns

as 305d yields given different windows of available data. The metrics for the daily milk yields were obtained by solely considering the model's errors on the predictions of missing milk yields. The metrics for the 305d yield were calculated by comparing the true 305d milk yield with the predicted 305d yield, i.e. the summation of the cumulative observed yield and the cumulative predicted yield. In order to inspect the SAE's robustness towards variable lengths of time intervals between different recordings, 0%, 30% and 60% of the milk yields in the window of available data were randomly set to missing. The results show that the SAE's performance with respect to predicting daily milk yields consistently increased when more information became available. While the SAE's RMSE was 6.31 for the prediction of 275 remaining milk yields, it decreased to 5.10 when half of the lactation curve was known and to 3.93 for the prediction of the last 30 missing milk yields. Similarly, the MAPE initially equalled 18% for 30 available observations and decreased with one percentage point each time 60 additional observations became available. Furthermore, the MAPE increased by maximum 1 percentage point when the sampling rate was increased to 30% and 2 percentage points when the sampling rate was increased to 60% for every possible forecast horizon. The other metrics also didn't change substantially when the sampling rate was increased, with a maximum deviation of 0.26, 0.03 and 0.02 for the RMSE, R^2 and ρ respectively. These results demonstrate the robustness of the SAE towards varying samples of available recordings.

For the predictions of the cumulative milk yield, the SAE's performance increased by a higher extent when more data became available. With only 30 recordings, the SAE was already capable of predicting the 305d yield with a MAPE of 10%, which decreased by a factor of 2 when 90 more recordings became available. Halfway through the lactation cycle, the SAE was able to predict the 305d yield with a 4% error rate. Furthermore, the MAPE for the 305d yield remained mostly constant for the different sampling rates for every window of available data, with a maximum increase of 1 percentage point.

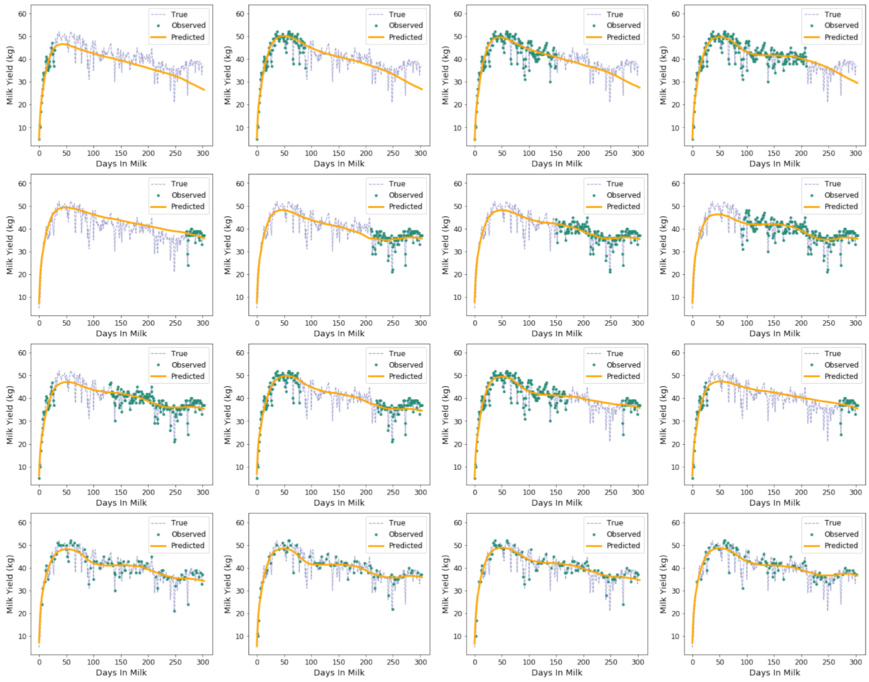


Figure 2.4: Visualization of the SAE's inference of missing milk yields for one observation in the test set for 4 fixed variations of each imputation pattern. Purple dashed line = true milk yield, green dots = observed milk yields, orange solid line = predicted milk yields

An example of how the SAE dynamically adapts its predictions for different windows of available data for a given test observation is visualized in **Fig.** 2.5. First, given a set of 30 available observations, the SAE generates a lactation curve that slightly underestimates the peak yield. Next, when 30 more observations are recorded and the SAE becomes aware of the actual peak yield, predictions are positively adjusted and the remaining milk yields are now slightly overestimated by the model. However, as soon as information of the declining phase after the peak becomes available, the SAE readjusts the estimated lactation curve and predicts reasonable values for remaining milk yields. The SAE keeps improving its predictions as more observations become available and adjusts the estimations for the known observations to improve its entire fit of the lactation curve.

Finally, **Table** 2.5 shows the correlation between the SAE's predictions and true values of a milk yield that is generated a certain number of days after the end of a given a window of available data. As expected, the correlations were largest

Days with observed data	Sampling rate	Daily yield				305d yield			
		RMSE	MAPE	R^2	ρ	RMSE	MAPE	R^2	ρ
30	0.0	6.31	0.18	0.60	0.77	1171.17	0.10	0.67	0.82
30	0.3	6.40	0.18	0.58	0.77	1196.12	0.10	0.66	0.81
30	0.6	6.48	0.19	0.57	0.76	1231.51	0.10	0.64	0.80
60	0.0	5.90	0.17	0.61	0.79	929.50	0.08	0.79	0.90
60	0.3	5.92	0.18	0.61	0.79	928.90	0.08	0.79	0.90
60	0.6	6.16	0.19	0.58	0.78	1015.65	0.09	0.75	0.88
90	0.0	5.64	0.17	0.62	0.79	763.90	0.06	0.86	0.93
90	0.3	5.64	0.17	0.62	0.78	768.75	0.06	0.86	0.93
90	0.6	5.80	0.18	0.59	0.78	821.78	0.07	0.84	0.92
120	0.0	5.43	0.16	0.62	0.79	634.57	0.05	0.90	0.95
120	0.3	5.47	0.17	0.61	0.78	650.50	0.05	0.90	0.95
120	0.6	5.62	0.18	0.59	0.77	690.96	0.06	0.89	0.94
150	0.0	5.10	0.16	0.64	0.80	487.96	0.04	0.94	0.97
150	0.3	5.16	0.16	0.63	0.79	499.61	0.04	0.94	0.97
150	0.6	5.32	0.17	0.61	0.78	536.80	0.04	0.93	0.97
180	0.0	4.87	0.15	0.65	0.81	372.44	0.03	0.97	0.98
180	0.3	4.94	0.16	0.64	0.80	383.65	0.03	0.96	0.98
180	0.6	5.07	0.16	0.62	0.79	410.52	0.03	0.96	0.98
210	0.0	4.63	0.15	0.67	0.82	267.39	0.02	0.98	0.99
210	0.3	4.73	0.16	0.65	0.81	281.75	0.02	0.98	0.99
210	0.6	4.83	0.16	0.64	0.80	295.07	0.02	0.98	0.99
240	0.0	4.39	0.15	0.68	0.83	176.39	0.01	0.99	1.00
240	0.3	4.52	0.15	0.67	0.82	188.62	0.02	0.99	1.00
240	0.6	4.54	0.15	0.66	0.81	189.64	0.01	0.99	1.00
270	0.0	3.93	0.14	0.74	0.86	79.43	0.01	1.00	1.00
270	0.3	4.11	0.14	0.72	0.86	88.81	0.01	1.00	1.00
270	0.6	4.13	0.15	0.71	0.85	89.67	0.01	1.00	1.00

Table 2.4: SAE's predictive performance for daily and 305d milk yield given different windows of observed data

for yields closest to the end of the window of available data and decreased as the forecast horizon increased. An average correlation of 0.83 was obtained for milk yields 30 days in the future, while the correlation between true yield and predictions made 60 days in advance was on average 0.78. For the largest possible forecast horizon of 270 days, the SAE achieved 0.38 correlation between its predictions and the true values.

2.4.3 Variable Importance

The variable importance scores for each group of features as well as each feature individually are presented in **Fig.** 2.6. **Fig.** 2.6a shows that the parity was the most discriminative feature with an average VI of 1.14, meaning that the SAE's total RMSE of 4.65 increased to 5.3 on average when the lactation number was randomly permutated. With a VI close to 1.12, the herd statistics was the second most informative feature group, with the herd's average milk produced in the first

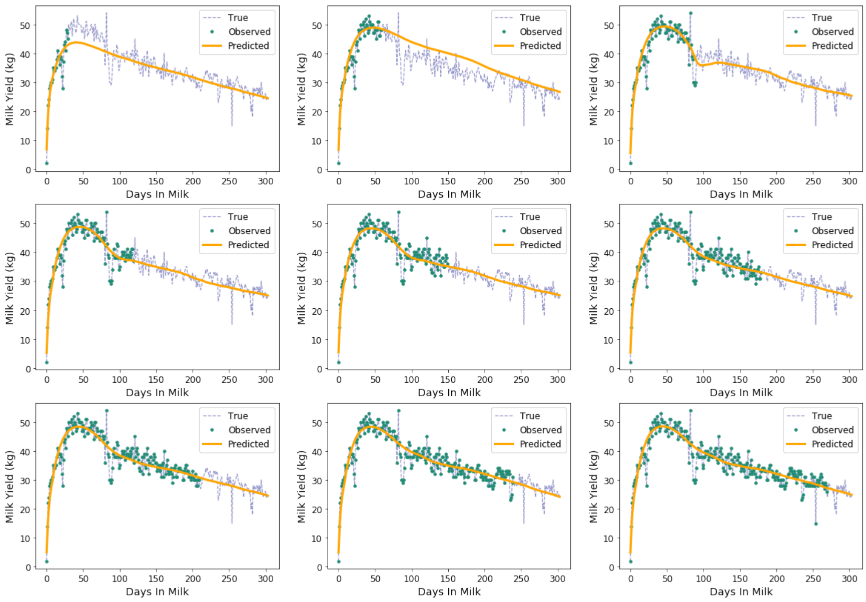
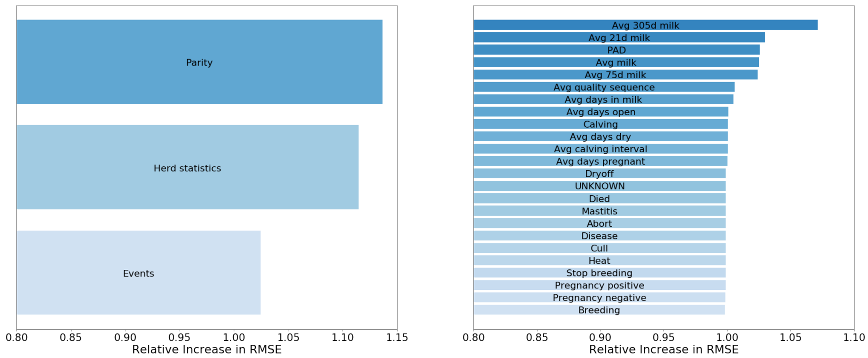


Figure 2.5: Visualization of the SAE's predictions for one observation in the test set for different windows of available data. Purple dashed line = true milk yields, green dots = observed milk yields, orange solid line = predicted milk yields

	Number of days used for prediction								
Forecast horizon	30	60	90	120	150	180	210	240	270
30	0.80	0.83	0.83	0.84	0.85	0.82	0.84	0.83	0.84
60	0.75	0.79	0.79	0.80	0.79	0.79	0.78	0.77	
90	0.71	0.75	0.76	0.75	0.75	0.74	0.72		
120	0.69	0.73	0.71	0.69	0.69	0.66			
150	0.64	0.67	0.65	0.64	0.62				
180	0.58	0.61	0.59	0.56					
210	0.53	0.55	0.51						
240	0.46	0.48							
270	0.38								

Table 2.5: SAE's performance in terms of Pearson Correlation between predicted and observed milk yields for different forecast horizons given different windows of observed data

305 days and 21 days as well as the herd's total average produced milk being the most discriminative variables as is shown by **Fig.** 2.6b. The average days in milk, average days dry and average calving interval contributed the least to the SAE's reconstruction capacity within the herd feature group which can also be seen from **Fig.** 2.6b. The events were the least informative group of variables with a VI score of around 1.02, though this could be expected since this feature group was very sparse.



(a) VI feature groups

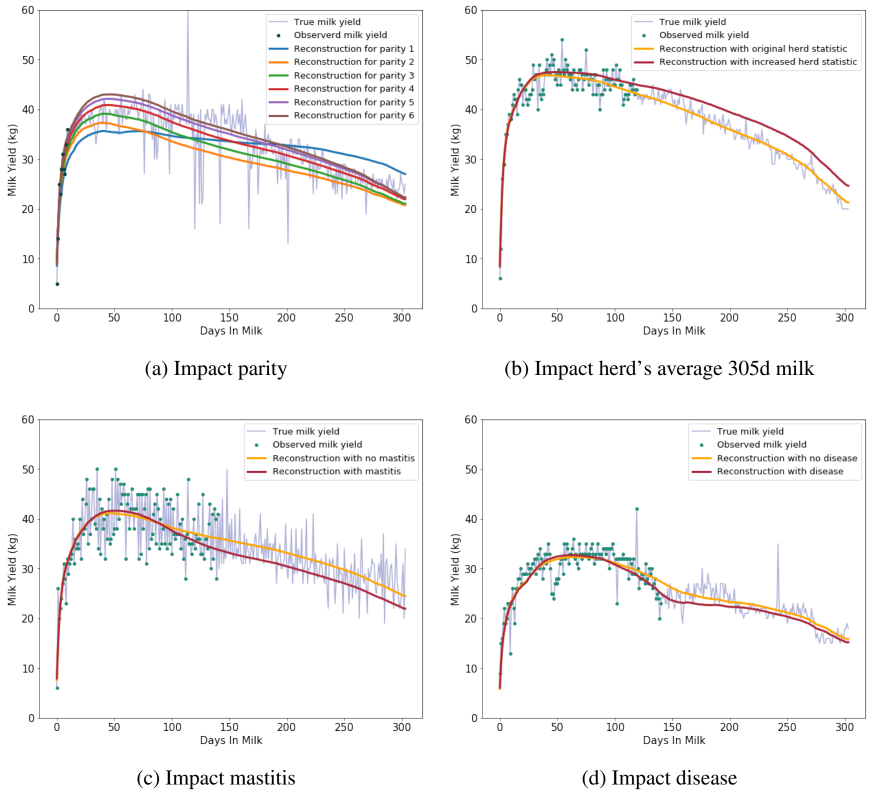
(b) VI individual features

Figure 2.6: Variable importance

The impact of some features on the SAE's predictions for a random test observation and a random window of available data is visualized by **Fig.** 2.7a, 2.7b, 2.7c and 2.7d. For a small number of recordings available in the beginning of the lactation cycle, the SAE positively adjusted the lactation curve for every consecutive parity. When mastitis was manually injected at the end of the window of 140 available recordings, the SAE predicted lower future yields than if the cow were to be healthy. Similarly, the SAE adjusted its predictions downwards when the disease event was injected at the end of the window of 132 available recordings. Finally, for a window of 120 available data points, the SAE adjusted the lactation curve upwards when the observation's initial normalized value of 0.7 for the herd's average 305d milk yield statistic was increased to its maximum value of 1.

2.5 Discussion

The R^2 obtained by the SAE was 0.84 and was slightly higher than the R^2 of 0.83 obtained by the MLP. An ANN similar to the MLP trained in this study was proposed by Grzesiak et al. (2006) who reported an R^2 of 0.77. Furthermore, Olori et al. (1999) stated that an $\mathrm{R}^2 > 0.70$ implies a good model fit, while a model with an $\mathrm{R}^2 < 0.40$ should not use for prediction. Hence, these results indicate a good fit of the SAE and MLP models on the data. However, the SAE performed better than the MLP on every individual imputation task. The MLP's large errors for the backtracking task can be explained by the fact that it only takes into account lagged milk yields to infer a missing milk yield. The SAE on the other hand is able to leverage every available observation recorded during the entire lactation cycle for inference. The SAE was therefore also better able to interpolate missing values than to predict or backtrack entire parts of the lactation curve since both past as well as future information could be exploited for interpolating missing values. Nevertheless, the SAE still achieved reasonable results on the prediction and backtracking tasks with correlations of 0.80 and 0.81 respectively. Hence these results show the versatility of the SAE to infer missing yields, irrespective of the pattern of missing observations along the lactation curve.



(a) Impact parity

(b) Impact herd's average 305d milk

(c) Impact mastitis

(d) Impact disease

Figure 2.7: Visualization of variable impact on reconstruction of lactation curve

As expected, the predictive performance of the SAE increased as the forecast horizon became shorter, which is showed in **Table** 2.5. The SAE's correlation between predictions and yields measured 30, 60 and 90 days after the first 30 recorded observations comprised 0.80, 0.75 and 0.71 respectively. These correlations are higher than those obtained by Njubi et al. (2010) who reported correlations of 0.75, 0.70 and 0.67 for the second, third and fourth predicted TD record by making use of an MLP with the first TD as a predictor. Macciotta et al. (2002)

and Vasconcelos et al. (2004) on the other hand achieved an average correlation of 0.85 for the prediction of the next TD based on the previous TDs, while an average correlation of 0.83 was found for the SAE's predictions of milk yields 30 days in the future. These slightly smaller correlations obtained by the SAE can be explained by the fact that the SAE is trained to perform well on multiple tasks simultaneously, while the aforementioned two studies constructed models specifically designed to predict the next TD yield based on historical data. Furthermore, instead of predicting TD yields which are measured every 30 days on average, the SAE is able to generate predictions for every daily milk yield of the curve. This facilitates improved monitoring systems since deviations from the expected milk yield curve can be detected more rapidly. In addition, the time of prediction of models which take lagged records into account increases linearly with the amount of missing points in the lactation curve since each prediction is dependent on the previously estimated values of the model. Instead of predicting the daily milk yields sequentially however, the SAE generates a reconstruction of the entire lactation curve at once. Likewise, curve fitting models such as proposed by Ali and Schaeffer (1987); Ehrlich et al. (2011); Wilmink (1987); Wood (1967) also generate entire lactation curves fitted to the data. However, Silvestre et al. (2006) found that these lactation models heavily depend on the sampling properties of the input data, with the accuracy quickly deteriorating as the amount of data decreased and the moment of the initiation of data collection was delayed. More specifically, the correlations between predicted and true yields obtained by Wood model decreased from 0.88 to 0.41 when the interval between calving and the first TD record was increased from 30 days to 60 days, while the Ali model's correlation decreased to 0.08 for the same sampling property. When the frequency of sampling was decreased by a factor of two on the other hand, the correlations decreased by a maximum of 0.04 for the Wood model and by a maximum of 0.09 for the Ali model. On the contrary, the SAE's correlation for every forecast horizon decreased by a maximum of 0.01 when 30% of the observations was randomly set to missing and by a maximum of 0.02 when a sampling rate of 0.60 was applied. This demonstrates the SAE's robustness towards varying sampling properties. In addition, curve fitting models are either fitted on lactation curves of homogeneous groups of animals or on each lactation curve individually. In the former case, the estimated parameters describe the group's lactation cycle and do not take into account individual variations in the lactation curve. In the latter case, one model is trained for every individual lactation curve which makes it impossible to learn from patterns occurring across multiple observations. The SAE on the other hand is a model trained on the entire dataset and learned to extract a latent representation from a specific lactation curve and to make a reconstruction by leveraging patterns learned from the entire population.

The results presented in **Table** 2.4 show that the SAE was able to predict the

305d milk yields with high precision for different windows of available data. Many previous studies also developed models to predict the 305d milk yield of a lactation curve based on a couple of observed TD records. Grzesiak et al. (2003) for example reported a prediction error of 6% by making use of a spline model with 4 TDs recorded before the 118th day of lactation. Similar results were obtained by the SAE with a MAPE of 5% for a window of 120 available observations. For a comparable window of available data, Dongre et al. (2012) achieved an R^2 of 0.86 by making use of an MLP, while an R^2 of 0.90 was achieved by the SAE for 120 days of data. The ANN proposed by Lacroix et al. (1995) achieved correlations of 0.897 and 0.963 when 115 and 210 observations were available, while correlations of 0.95 and 0.99 were obtained by the SAE for windows of equal size. By making use of autoregressive models, Vasconcelos et al. (2004) reported correlations between the predicted and true 305d milk yields of 0.85 and 0.94 for 2 and 4 available TDs, while Macciotta et al. (2002) obtained correlations of 0.88 and 0.96 by applying an autoregressive moving average model. The SAE on the other hand achieved correlations of 0.90 when 60 days of data were available and 0.97 when 150 days of data were available, regardless of the sampling rate. Hence, these results suggest that the SAE is capable of achieving state-of-the art results with respect to predicting 305d milk yields for different windows of available data. This can help farmers to more accurately forecast a cow's entire productivity which facilitates better selection and culling decisions as well as more accurate estimates of future costs and revenues. In addition, while most previous studies relied on a fixed number of observations for predicting the cumulative milk yield, the results demonstrated the robustness of the SAE's 305d predictions with respect to varying samples of available recordings. This can be of great value when variable lengths of time intervals exist between different yield measurements or when missing observations are present.

Finally, as shown by **Fig.** 2.6, the parity feature contributed the most to the SAE's predictive capacity with a VI of 1.14. This is not surprising as dairy cows generally produce more in each subsequent lactation cycle (Ehrlich et al., 2011; Macciotta et al., 2011). This is also demonstrated by **Fig.** 2.7a which shows how the SAE positively adjusts its predictions for higher parities. Furthermore, the results presented in this research are also in line with those found by other studies in that the quality of herd management expressed by statistics such as average milk yield production has a positive impact on a cow's milk production (Jeretina et al., 2015; Lacroix et al., 1995). A loss in milk on the other hand can be expected in case of illness such as mastitis (Adriaens et al., 2018). As shown by **Fig.** 2.7c and 2.7d, the SAE learned to detect these patterns and lowered its forecasted milk yields when the cow was exposed to mastitis or another disease. Hence, by incorporating a cow's sequence of reproduction and health events, more realistic milk yield predictions are made by the SAE and allows for more accurate forecasts of the farm's total production. Additionally, the model presented in this study facilitates improved animal monitoring systems as it enables a more accurate comparison between an animal's expected and true lactation curve.

2.6 Conclusion

Many lactation models currently exist to forecast the missing milk yields along the lactation curve of a cow (Macciotta et al., 2011, 2005; Zhang et al., 2016). Most of these models however, assume that the estimation of a specific point in the lactation curve depends on a fixed number of previously measured milk yields with constant time intervals between each recording. Yet, common environmental factors such as herd management, weather and nutrition as well as the animal's repeatability result in covariances between a specific milk yield and its preceding as well as subsequent recordings. Furthermore, due to animal treatment and defective recording machines, missing observations can occur and variable lengths of intervals may exists between the different measurements. In this study, we propose a model that infers a missing milk yield along the lactation curve by leveraging a latent representation of all information available in the lactation cycle. As a result, a missing milk yield is inferred by all the observations recorded before and after the moment of prediction, irrespective of the length of the time interval between these recordings. Results showed that the quality of the model's encodings increased when the recorded milk yields were augmented with the parity and herd statistics as well as the health and reproduction events encountered by the cow during the lactation cycle. Furthermore, the model was able to accurately predict missing milk yields for different windows of available data, regardless of the sampling properties. Hence, the model presented in this study can be used to predict and interpolate missing milk yields along the lactation curve. In addition, the model is able to assess the impact of herd management and events such as mastitis on the cow's productivity. As a result, this framework allows the farmer to obtain more accurate forecasts on its production as well as costs and facilitates more enhanced animal monitoring systems.

References

Abadi, M., Agarwal, A., Barham, P., Brevdo, E., Chen, Z., Citro, C., Corrado, G. S., Davis, A., Dean, J., Devin, M., Ghemawat, S., Goodfellow, I., Harp, A., Irving, G., Isard, M., Jia, Y., Jozefowicz, R., Kaiser, L., Kudlur, M., Levenberg, J., Mané, D., Monga, R., Moore, S., Murray, D., Olah, C., Schuster, M., Shlens, J., Steiner, B., Sutskever, I., Talwar, K., Tucker, P., Vanhoucke, V., Vasudevan, V., Viégas, F., Vinyals, O., Warden, P., Wattenberg, M., Wicke, M., Yu, Y., and Zheng, X. TensorFlow: Large-scale machine learning on heterogeneous systems, 2015. URL `https://www.tensorflow.org/`. Software available from tensorflow.org.

Adriaens, I., Huybrechts, T., Aernouts, B., Geerinckx, K., Piepers, S., Ketelaere, B. D., and Saeys, W. Method for short-term prediction of milk yield at the quarter level to improve udder health monitoring. *Journal of Dairy Science*, 101(11): 10327 – 10336, 2018. ISSN 0022-0302.

Ali, T. E. and Schaeffer, L. R. Accounting for covariances among test day milk yields in dairy cows. *Canadian Journal of Animal Science*, 67(3):637–644, 1987.

Bengio, Y., Yao, L., Alain, G., and Vincent, P. Generalized denoising auto-encoders as generative models. In *Advances in Neural Information Processing Systems 26*, pages 899–907. 2013.

Breiman, L. Random forests. *Machine Learning*, 45(1):5–32, 2001.

Chollet, F. et al. Keras. `https://github.com/fchollet/keras`, 2015.

Dongre, V., Gandhi, R., Singh, A., and Ruhil, A. Comparative efficiency of artificial neural networks and multiple linear regression analysis for prediction of first lactation 305-day milk yield in Sahiwal cattle. *Livestock Science*, 147(1):192 – 197, 2012. ISSN 1871-1413.

Ehrlich, J. et al. Quantifying shape of lactation curves, and benchmark curves for common dairy breeds and parities. *Bovine Practitioner*, 45(1):88, 2011.

Grzesiak, W., Wojcik, J., and Binerowska, B. Prediction of 305-day first lactation milk yield in cows with selected regression models. *Archives Animal Breeding*, 46(3):213–224, 2003.

Grzesiak, W., Błaszczyk, P., and Lacroix, R. Methods of predicting milk yield in dairy cows—predictive capabilities of Wood's lactation curve and artificial neural networks (ANNs). *Computers and Electronics in Agriculture*, 54(2):69 – 83, 2006. ISSN 0168-1699.

Harris, C. R., Millman, K. J., van der Walt, S. J., Gommers, R., Virtanen, P., Cournapeau, D., Wieser, E., Taylor, J., Berg, S., Smith, N. J., Kern, R., Picus, M., Hoyer, S., van Kerkwijk, M. H., Brett, M., Haldane, A., del Río, J. F., Wiebe, M., Peterson, P., Gérard-Marchant, P., Sheppard, K., Reddy, T., Weckesser, W., Abbasi, H., Gohlke, C., and Oliphant, T. E. Array programming with NumPy. *Nature*, 585(7825):357–362, September 2020. doi: 10.1038/s41586-020-2649-2. URL `https://doi.org/10.1038/s41586-020-2649-2`.

He, K., Zhang, X., Ren, S., and Sun, J. Deep residual learning for image recognition. In *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 770–778, 2016.

Hinton, G. E. and Zemel, R. S. Autoencoders, minimum description length and Helmholtz free energy. In *Advances in neural information processing systems*, pages 3–10, 1994.

Hunter, J. D. Matplotlib: A 2d graphics environment. *Computing in Science & Engineering*, 9(3):90–95, 2007. doi: 10.1109/MCSE.2007.55.

Jeretina, J., Babnik, D., and Škorjanc, D. Prediction of standard lactation curves for primiparous Holstein cows by using corrected regression models. *Italian Journal of Animal Science*, 14(4):3776, 2015.

Kingma, D. and Ba, J. Adam: A method for stochastic optimization. *ArXiv e-prints*, 2014.

Lacroix, R., Wade, K., Kok, R., and Hayes, J. Prediction of cow performance with a connectionist model. *Transactions of the ASAE*, 38(5):1573–1579, 1995.

Lacroix, R., Salehi, F., Yang, X., and Wade, K. Effects of data preprocessing on the performance of artificial neural networks for dairy yield prediction and cow culling classification. *Transactions of the ASAE*, 40(3):839–846, 1997.

Li, W., Geng, C., and Chen, S. Leave zero out: Towards a no-cross-validation approach for model selection. *arXiv*, abs/2012.13309, 2020.

Macciotta, N. P., Dimauro, C., Rassu, S. P., Steri, R., and Pulina, G. The mathematical description of lactation curves in dairy cattle. *Italian Journal of Animal Science*, 10(4):e51, 2011.

Macciotta, N., Vicario, D., Pulina, G., and Cappio-Borlino, A. Test day and lactation yield predictions in Italian Simmental cows by ARMA methods. *Journal of Dairy Science*, 85(11):3107 – 3114, 2002. ISSN 0022-0302.

Macciotta, N., Vicario, D., and Cappio-Borlino, A. Detection of different shapes of lactation curve for milk yield in dairy cattle by empirical mathematical models. *Journal of Dairy Science*, 88(3):1178 – 1191, 2005. ISSN 0022-0302.

Murphy, M., O'Mahony, M., Shalloo, L., French, P., and Upton, J. Comparison of modelling techniques for milk-production forecasting. *Journal of Dairy Science*, 97(6):3352 – 3363, 2014. ISSN 0022-0302.

Njubi, D., Wakhungu, J., and Badamana, M. Use of test-day records to predict first lactation 305-day milk yield using artificial neural network in Kenyan Holstein–Friesian dairy cows. *Tropical animal health and production*, 42(4):639–644, 2010.

Olori, V., Brotherstone, S., Hill, W., and McGuirk, B. Fit of standard models of the lactation curve to weekly records of milk production of cows in a single herd. *Livestock Production Science*, 58(1):55 – 63, 1999. ISSN 0301-6226.

pandas development team, T. pandas-dev/pandas: Pandas, February 2020. URL `https://doi.org/10.5281/zenodo.3509134`.

Parsons, V. L. *Stratified Sampling*, pages 1–11. John Wiley & Sons, Ltd, 2017. ISBN 9781118445112. doi: https://doi.org/10.1002/9781118445112.stat05999.pub2.

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., and Duchesnay, E. Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research*, 12:2825–2830, 2011.

Rumelhart, D. E., Hinton, G. E., and Williams, R. J. Learning representations by back-propagating errors. *Nature*, 323(6088):533–536, 1986.

Salehi, F., Lacroix, R., and Wade, K. Improving dairy yield predictions through combined record classifiers and specialized artificial neural networks. *Computers and Electronics in Agriculture*, 20(3):199 – 213, 1998. ISSN 0168-1699.

Sanzogni, L. and Kerr, D. Milk production estimates using feed forward artificial neural networks. *Computers and Electronics in Agriculture*, 32(1):21 – 30, 2001. ISSN 0168-1699.

Sharma, A. K., Sharma, R., and Kasana, H. Prediction of first lactation 305-day milk yield in Karan Fries dairy cattle using ANN modeling. *Applied Soft Computing*, 7(3):1112 – 1120, 2007. ISSN 1568-4946.

Silvestre, A., Petim-Batista, F., and Colaço, J. The accuracy of seven mathematical functions in modeling dairy cattle lactation curves based on test-day records from varying sample schemes. *Journal of Dairy Science*, 89(5):1813 – 1821, 2006. ISSN 0022-0302.

Szegedy, C., Liu, W., Jia, Y., Sermanet, P., Reed, S., Anguelov, D., Erhan, D., Vanhoucke, V., and Rabinovich, A. Going deeper with convolutions. In *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 1–9, 2015.

Vasconcelos, J., Martins, A., Petim-Batista, M., Colaço, J., Blake, R., and Carvalheira, J. Prediction of daily and lactation yields of milk, fat, and protein using an autoregressive repeatability test day model. *Journal of Dairy Science*, 87(8): 2591 – 2598, 2004. ISSN 0022-0302.

Waskom, M. L. seaborn: statistical data visualization. *Journal of Open Source Software*, 6(60):3021, 2021. doi: 10.21105/joss.03021. URL `https://doi.org/10.21105/joss.03021`.

Wilmink, J. Adjustment of test-day milk, fat and protein yield for age, season and stage of lactation. *Livestock Production Science*, 16(4):335 – 348, 1987. ISSN 0301-6226.

Wood, P. Algebraic model of the lactation curve in cattle. *Nature*, 216(5111): 164–165, 1967.

Zeiler, M. D., Krishnan, D., Taylor, G. W., and Fergus, R. Deconvolutional networks. In *2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, pages 2528–2535, June 2010.

Zhang, F., Murphy, M. D., Shalloo, L., Ruelle, E., and Upton, J. An automatic model configuration and optimization system for milk production forecasting. *Computers and Electronics in Agriculture*, 128:100 – 111, 2016. ISSN 0168-1699.

Zhao, B., Lu, H., Chen, S., Liu, J., and Wu, D. Convolutional neural networks for time series classification. *Journal of Systems Engineering and Electronics*, 28(1): 162–169, Feb 2017. ISSN 1004-4132.

3

Predicting the milk yield curve of dairy cows in the subsequent lactation period using deep learning[1]

3.1 Abstract

Existing lactation models predict milk yields based on a fixed amount of observed milk production in early lactation. In contrast, this study proposes a model to predict the entire lactation curve of dairy cows by leveraging historical milk yield information observed in the preceding cycle. More specifically, we present a deep learning framework to encode the model inputs, predict the latent representation of the milk yield sequences and generate the corresponding lactation curves. Results show that the proposed framework outperforms the baseline models and that during the first 26 days of lactation, the model's predictions are more accurate than those of a state-of-the-art lactation model which is able to leverage the observed milk yields. As a result, the framework presented in this study allows farmers to increase their forecast horizon with respect to predicting its herd's total production and hence facilitates optimal herd management. Additionally, the model can be used to compare a cow's actual and expected milk yield over the entire course of the lactation cycle. This in turn can help to accelerate disease detection and

[1]Based on: Liseune, A., Salamone, M., Van den Poel, D., Van Ranst, B., & Hostens, M. (2021). Predicting the milk yield curve of dairy cows in the subsequent lactation period using deep learning. *Journal of Computers and Electronics in Dairy Science.*

enhance current animal monitoring systems. Finally, as the model incorporates the impact of health and reproduction events as well as herd management on the cow's productivity, future earnings and costs can be estimated more accurately.

3.2 Introduction

Forecasting milk yield is an important asset for dairy farmers as it can lead to improved decision making for optimal herd management (Dematawewa et al., 2007; Grzesiak et al., 2006). In particular, lactation models help to forecast the dairy farm's income (Ehrlich et al., 2011; Grzesiak et al., 2003b), determine the required nutrition and energy consumption (Murphy et al., 2014), optimize selection and culling decisions (Njubi et al., 2010; Sharma et al., 2006) and enhance animal monitoring systems (Adriaens et al., 2018; Silvestre et al., 2006).

Early lactation models were determined by mathematical functions describing the general milk yield pattern of homogeneous groups of animals (Brody et al., 1923). Lactation was modeled as a function of time with an increasing phase until a peak yield, followed by a more steady decline (Olori et al., 1999). Several mathematical functions have been widely used for predicting dairy milk yields, including an incomplete gamma, (Wood, 1967), a polynomial (Ali and Schaeffer, 1987), an exponential (Wilmink, 1987) and a Legendre polynomial (Kirkpatrick et al., 1994). Over time, the need to model individual variations from the mean lactation curve increased as more animal records were collected and farm management software improved (Macciotta et al., 2011). This led to several authors developing new models in order to fit more complex shapes and to include more input features (Murphy et al., 2018). For example, Grzesiak et al. (2003a) presented a multivariate regression model (MLR) that in addition to days in milk, also used test-day (TD) records, month of calving and the percentage of Holstein-Friesian (HF) genes as features to predict the 305d milk yield. Vasconcelos et al. (2004) and Macciotta et al. (2002) proposed autoregressive (AR) models in order to predict a milk yield based on a sequence of preceding TD records. Græsbøll et al. (2016) on the other hand presented a robust prediction model for cow level milk yield using lactation curves with reduced number of parameters, which is useful in case of sparse data.

Later, several artificial neural networks (ANN) have been proposed to predict milk yield. Lacroix et al. (1995) trained the first successful multilayer perceptron model (MLP) to predict the 305d yield based on 16 variables. In subsequent studies, this model was improved by applying more sophisticated data preprocessing techniques (Lacroix et al., 1997) and by training multiple networks each assigned to make specific predictions (Salehi et al., 1998). Furthermore, MLPs have been used to predict the 305d milk yield (Gorgulu, 2012; Grzesiak et al., 2003a), the 305d milk yield of the first lactation (Njubi et al., 2010; Sharma et al., 2006, 2007), the daily milk yield (Grzesiak et al., 2006; Torres et al., 2005) and dairy herd's total production (Murphy et al., 2014; Sanzogni and Kerr, 2001). In contrast to MLPs, recent research has shown that convolutional neural networks (CNN), most commonly applied for processing image data, can also be of great value for time series analysis (Zhao et al., 2017; Zheng et al., 2014). This was also shown in **Chapter** 2 that presented a sequential autoencoder (SAE) to interpolate as well as predict missing milk yields along the entire lactation cycle by leveraging a latent representation of all the information available in the lactation cycle.

With the advent of individual curve fitting models, animal monitoring systems improved significantly. More specifically, by comparing a cow's expected and actual milk yield, diseases such as mastitis and ketosis could be detected more accurately (Adriaens et al., 2018; Grzesiak et al., 2003a). Such early detection systems are very valuable for the farmer since there can be a lot of costs associated with diseases of dairy cattle, e.g. lower production, discarded milk, treatment and culling or death (Gröhn et al., 2004; Wilson et al., 2004). Furthermore, the knowledge of the expected lactation curve made it easier to assess the impact of different treatments (Tekerli et al., 2000). However, in order to infer the expected lactation curve to which the actual milk yield can be compared, an initial number of milk yields recorded in early lactation is generally required. As a result, a reliable reference in early lactation is often unavailable which makes herd management as well as health monitoring of dairy cows particularly difficult in the period immediately after calving.

In this research, we propose a novel methodology to predict the entire lactation curve of a cow. More specifically, this paper contributes to previous research in multiple ways. Firstly, we present a model that predicts a lactation curve by using the sequence of milk yields generated in the preceding cycle. Moreover, instead of using the raw sequence of milk yields, the corresponding latent representation is used in order to disentangle the sequential information and to reduce the feature dimensionality. Secondly, we formulate a framework that models the impact of animal and herd Key Performance Indicators (KPI), lactation number and the sequence of health and reproduction events the cow encountered during the preceding cycle on the milk production. Finally, a new prediction approach is presented that generates the entire lactation curve non-sequentially. In particular, an MLP is used to generate the curve's latent encoding which is subsequently converted back into its corresponding milk yield sequence. The predictions obtained by the proposed model can be used to calculate the milk losses immediately after calving and hence support animal monitoring systems. In addition, the framework enables farmers to increase their forecast horizon with respect to the farm's future profitability.

3.3 Materials and Methods

3.3.1 Data

The same data as described in **Chapter 2** 2.3.1 was used for this study. In addition to the features defined in **Chapter 2** 2.3.1, two more sets of features were constructed for this study. For each lactation, several animal KPIs were calculated and by averaging the milk yields of the entire herd per parity, the average lactation curve per herd and per parity was obtained. The final dataset was constructed by extracting every available pair of consecutive lactation cycles and therefore exclusively consisted of data on cows with at least 2 lactation cycles. This resulted in a total of 23745 observations, with each observation's dependent variable comprising the sequence of milk yields generated in the predicted lactation cycle. The features included the sequence of milk yields, animal KPIs and events generated in the preceding cycle together with the lactation number, herd KPIs and the herd's average lactation curve corresponding to the predicted lactation cycle. Milk yields and events recorded after 305 days in lactation were removed and days on which no events occurred were represented by a special `PAD` token. Events that did occur in the validation or test set but not in the training set were labeled as `UNKNOWN` and were considered as a rare but unknown event. The animal and herd KPIs were normalized between 0 and 1 and missing values were imputed with the mean value of the variable. From the entire collection of observations containing no missing milk yields in the predicted period, 2000 randomly sampled observations were assigned to the validation set and 2000 randomly sampled observations were assigned to the test set. The remaining observations with complete information in the predicted period were assigned to the training set together with the collection of observations containing missing milk yields in the predicted lactation cycle. An overview of all the variables used in this study is given in **Table** 3.1.

3.3.2 Feature Extraction

In general, MLP's are used to generate predictions for static input data. When the input data contains sequential or spatial data, however, other specialized architectures such as LSTMs or CNNs are first used to extract one-dimensional representations of the non-static data. The obtained static feature vectors are then passed to an MLP which generates the final predictions. This study uses a combination of sequential and static features. Therefore, different feature extraction techniques are used before passing the features to the final classification model. In the following sections, we discuss the different techniques used to preprocess the input data.

Variable Group	Lactation Cycle	Dimension	Variable Name
Milk Yields	Preceding	1 x 305	Preceding Milk Yield
Milk Yields	Predicted	1 x 305	Predicted Milk Yield
Herd Yields	Predicted	1 x 305	Avg Milk Yield Per Herd Per Parity
Herd KPIs	Predicted	10 x 1	Avg 21d Milk Avg 75d Milk Avg 305d Milk Avg Milk Avg Days Dry Avg Days Open Avg Days Pregnant Avg Days In Milk Avg Calving Interval Avg Quality Sequence
Animal KPIs	Preceding	13 x 1	Age At First Calving Age At First Insemination Days Pregnant Days In Milk Minimum Milk Yield Maximum Milk Yield Total Milk Yield 305d Milk Yield 75d Milk Yield 21d Milk Yield Avg Milk Std Milk Quality Sequence

Events	Preceding	1 x 305	Mastitis
			Abort
			Breeding
			Stop Breeding
			Pregnancy Negative
			Pregnancy Positive
			Calving
			Disease
			Died
			Heat
			Cull
			Dryoff
			`PAD`
			`UNKNOWN`

Parity	Predicted	1 x 1	Lactation Number

Table 3.1: Variables used in this study. Lactation Cycle = the cycle from which the data was obtained. Dimension = the number of features belonging to the feature group and the number of time steps at which the features were measured.

3.3.2.1 Convolutional Neural Network

Generally, a CNN's architecture exists out of a sequence of blocks, with each block typically comprising a convolutional layer, followed by a non-linear activation function for feature extraction. Pooling layers on the other hand are specifically designed for reducing the dimensionality of the hidden representation and making the network invariant to small translations in the input by obtaining the most prominent features. **Fig.** 3.1 gives an example of how a CNN with one convolutional block is applied on a sequence with S time steps and N features at each time step. A kernel of length $K < S$ and width N is slided over the entire input sequence and the dot products between the entries of the kernel and the input at any position are calculated. By sliding more than one kernel over the input sequence, multiple one-dimensional sequences are produced, with the elements of the sequences containing the response values of the corresponding kernels at every time step. Subsequently, an element-wise non-linear activation function is applied, followed by a pooling layer which summarizes the feature response in a certain time window. Finally, the output of the pooling layer is flattened and results in a vector that represents all the features extracted by every kernel at every time step and hence can be used for upstream tasks.

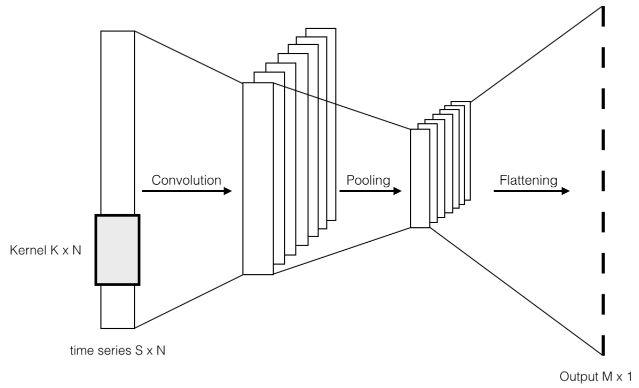


Figure 3.1: Convolutional neural netowork for time series (Liseune et al., 2020)

In this study, a similar architecture was applied on the sequence comprising the M last reproduction and health events encountered by the cow in the preceding lactation cycle. Before feeding the events directly to the CNN however, an embedding matrix of size 14 x k was used to convert each possible event occuring in the sequence into its corresponding numeric vector. Hence, the sequence of M events was converted to a M x k sequence with each time step of the sequence containing the corresponding event's embedding. In order to find the optimal event representations, the values of the embedding matrix were considered as network parameters and were updated during training. The sequences of embedding vectors were passed to a CNN of which each block consisted out of a linear transformation followed by a batch normalization layer. This layer normalizes the hidden activations and supports faster training as well as regularization (Ioffe and Szegedy, 2015). More specifically, if $\mathbf{X}$ represents a batch of hidden activations of a certain layer, then the normalization can be obtained as follows:

$$\hat{\mathbf{X}} = \frac{\mathbf{X} - \mu}{\sigma}$$

with μ and σ being the activation's means and standard deviations respectively. The output of the batch normalization layer can then be calculated as follows:

$$y = \gamma \hat{\mathbf{X}} + \beta$$

with γ and β being trainable parameters. Subsequently, a Leaky ReLU activation function was applied and is defined as follows:

$$f = max(\alpha \boldsymbol{x}, \boldsymbol{x})$$

with α often being set at a small constant value. As a result, the Leaky ReLU function will produce positive gradients for its entire range and hence facilitates gradient-based optimization. In addition, a dropout layer was applied to the output of the non-linear activation. With this mechanism, each neuron is dropped from the network with a certain probability and hence enforces the neurons to perform well, regardless of which other neurons are present in the network (Hinton et al., 2012). Dimensionality reduction of the hidden representation was obtained by applying max pooling layers which extract the most activated presence of a feature in a specific time interval. The last convolutional block's output was flattened and resulted in the static feature representation of the sequence of events.

3.3.2.2 SAE Encoder

The sequential autoencoder (SAE), as presented in **Chapter** 2, is an artificial neural network (ANN) specifically designed to infer missing milk yields along the lactation curve. More specifically, the SAE comprises a CNN which is used to transform the lactation cycle's sequential information into an extensive set of time-dependent features. A neural autoencoder then extracts a latent representation from the lactation curve and uses this encoding to get a reconstruction of the input features. Finally, a deconvolutional neural network (DNN) converts the reconstructed features back into the corresponding sequence of milk yields. As a result, the SAE is able to infer missing milk yields by making use of all the information available in a lactation cycle, irrespective of the length of the time intervals between the different observations. In **Fig.** 3.2, a schematic overview of the SAE is given. The SAE's encoder, which comprises the CNN and the autoencoder's compression side, takes an incomplete milk yield curve as input and extracts a latent representation from this curve. Subsequently, the autoencoder's reconstruction side and DNN (i.e., the

SAE's decoder) convert the latent representation back into the reconstruction of the milk yield curve. In this study, the SAE was used to impute missing values along the milk yield curves of the preceding lactation cycles. The reconstructed lactation curve however is large in dimensionality and needs to be processed sequentially. Hence, instead of applying the entire SAE's architecture, the output of the SAE's encoder was used. This output comprises the latent representation of the lactation curve and hence entails its most prominent traits in a low-dimensional vector. Likewise, the SAE's encoder was applied to convert the herd's average lactation curve per parity into its low-dimensional representation.

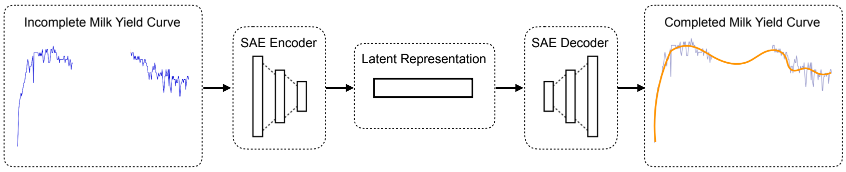


Figure 3.2: Schematic overview of the SAE

3.3.3 Milk Yield Prediction

In order to predict the lactation curve of the predicted lactation cycle, a standard MLP was used. In this study, each hidden layer comprised a linear transformation followed by a batch normalization layer, a Leaky ReLU function and a dropout layer. The MLP's inputs comprised the latent representation of the preceding and the herd's average milk yield sequence as well as the processed sequence of health and reproduction events, the animal and herd KPIs and the lactation number. Instead of predicting the entire sequence of milk yields however, the MLP's output was constrained to be of the same dimensionality as that of the latent encoding generated by the SAE's encoder. This vector was then fed to the SAE's decoder, which generated the reconstruction of the corresponding lactation curve. As a result, the MLP was trained to predict the lactation curve's latent encoding, rather than the entire sequence which would be time-consuming and prone to overfitting. The combination of all the different model components to forecast the lactation curve resulted in the Subsequent Lactation Milk Yield Predictor (SLMYP). A schematic overview of the SLMYP is depicted by **Fig.** 3.3.

3.3.4 Training

The entire SLMYP model was trained by making use of the backpropagation algorithm, firstly introduced by Rumelhart et al. (1986). More specifically, the inputs were first propagated through the entire network to produce a lactation

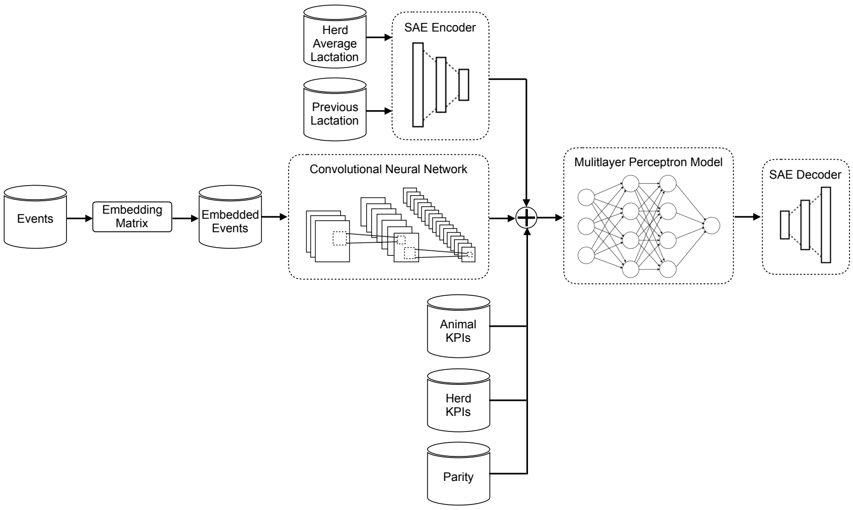
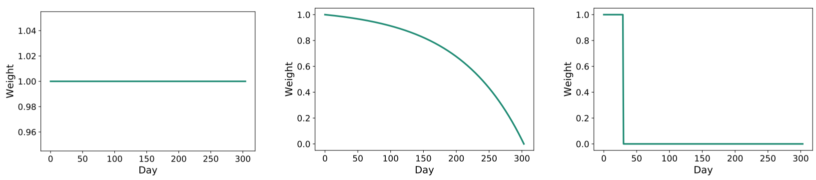


Figure 3.3: Schematic overview of the SLMYP

curve. A loss between the model's predicted and true lactation curve was then calculated and the gradient of the loss function was propagated backward through the network. The weights were then updated by applying the Adam gradient-based optimization algorithm (Kingma and Ba, 2014). In order to avoid overfitting, an early stopping procedure was applied in which the model was trained until its performance on the validation set started to degrade. Furthermore, as neural networks are typically characterized by a large range of possible configurations, a Bayesian optimization procedure was applied to find the optimal hyperparameter setting. More specifically, once the early stopping procedure was terminated, the next configuration was determined by a trade-off between the exploration of a parameter setting with uncertain results against the exploitation of a point in parameter space with high model performance. In addition to the hyperparameters defining the CNN's and MLP's model architecture, the inclusion of all the model inputs except for the preceding lactation curve were also set as hyperparameter in order to obtain only those features with a significant predictive power. Furthermore, a boolean hyperparameter was included that determined whether the data should be balanced with respect to the lactation number. In the case of data balancing, observations with rare lactation numbers corresponding to the predicted lactation cycles were upsampled during training such that each training batch consisted out of an equal number of observations per lactation number. Finally, each model was trained with one out of three possible loss functions which weighted each daily prediction in the predicted period differently. A uniform loss assigned the same weight to each daily prediction such that the resulting model tried to fit each milk yield of the lactation curve equally well. In order to solely focus on generating predictions during the period for which current lactation models generally lack predictions, a step loss was used that assigned a uniform weight to the first 30 predictions of the predicted period while disregarding later days. This resulted in a model that was trained to approximate milk yields in early lactation as best as possible, while ignoring the predictive accuracy in the period after early lactation. Finally, a logarithmic loss assigned weights to every milk yield in the lactation curve, but with a lower weight for each subsequent milk yield. As a result, the model trained with the logarithmic loss function was particularly focused on generating accurate predictions in early lactation, yet without ignoring the predictions made for the remaining part of the lactation cycle. The weights used by each loss function are depicted by **Fig.** 3.4.



(a) Uniform loss (b) Logarithmic loss (c) Step loss

Figure 3.4: Loss functions

3.3.5 Benchmark Models

The SLMYP was compared with four models used to predict the lactation curves of future cycles with no recordings. The first benchmark model uses the milk yields generated in the preceding cycle as forecast. A second model approximates the lactation curve by the herd's average milk yield curve corresponding to the predicted period. Third, a Wood's curve was fitted on the lactation data for each distinct parity. The predictions were then generated by the Wood's curve corresponding to the parity of the predicted lactation cycle. Furthermore, the SLMYP's predictions were also compared with the predictions generated by the SAE for increasing windows of observed milk yields in the predicted period.

3.3.6 Model Evaluation

The performance of the models were evaluated by four metrics frequently used in similar research: the Pearson correlation coefficient (ρ), the root mean squared error (RMSE), the mean absolute error (MAE) and the mean absolute percentage error (MAPE) (Grzesiak et al., 2003b; Lacroix et al., 1995; Liseune et al., 2020). In contrast to the MAE and RMSE, which are absolute measures of fit, the MAPE is not scale dependent and indicates how much the predictions deviate from the true values on average. It is defined by the following formula:

$$MAPE = \frac{1}{n}\sum_{i=1}^{n}\frac{|y_i - \hat{y}_i|}{y_i} \times 100\%$$

3.3.7 Variable Importance

The variable importance (VI) score measures the relative increase of the model's error for a random perturbation of a specific feature (Breiman, 2001). More specifically, if e is the error of the model using all the features and $\widetilde{e}_{k,i}$ is the error of the model for the ith random perturbation of feature k, then the VI of feature k can be calculated as follows:

$$VI_k = \frac{1}{n}\sum_{i=1}^{n}\frac{\widetilde{e}_{k,i}}{e}$$

Likewise, the relative increase of the model's error and hence the VI score for an entire feature group can be calculated by randomizing all the features belonging to that feature group.

3.3.8 Programming Tools

All data processing and analyses were done in the same environment as described in **Chapter 2** 2.3.6.

3.4 Results

3.4.1 Model Selection

The Bayesian optimization procedure was initialized by evaluating 50 randomly sampled model configurations. Each parameter setting was evaluated on the validation set after every 1000 training iterations with a batch size of 32. Every time the validation RMSE decreased, the model's weights were saved and training was terminated when the performance did not improve for 10 consecutive times. The best performing model was retrained on the training and validation set and evaluated on the test set. This model included the lactation number, the latent representation of the herd's average milk yield curve, the animal and herd KPIs as well as the sequence of the last 300 health and reproductions events encountered by the cow in the preceding lactation cycle. The events were embedded into 5-dimensional vectors and were passed to the CNN which consisted out of 4 blocks with 16 and 32 kernels of size 3 in the first two and last two blocks respectively. The output of each block's non-linear activation was dropped with a probability of 0.5 and a max pooling layer of size 4 was applied in block 1 and 4. The MLP contained 2 layers with 100 and 50 neurons and a dropout probability of 0.2 in each layer. The α of the Leaky ReLU layers in both the CNN and MLP was 0.5 and the initial learning rate of the Adam optimization algorithm was set at 0.001. Furthermore, the best results were obtained when the training data was balanced with respect to the lactation number.

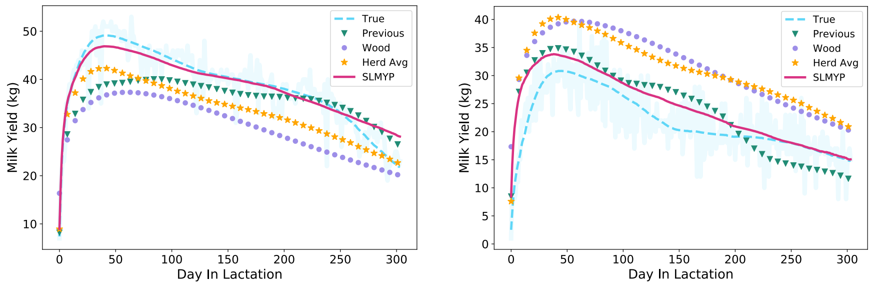
3.4.2 Model Performance

The predictive performance on the daily as well as 305d yield of the SLMYP trained with the uniform loss function and the baseline models on the test set is depicted in **Table** 3.2. For the daily yields, the metrics were obtained by taking into account the errors between the non-missing milk yields in the predicted period and the corresponding model predictions. The performance scores for the 305d yields were obtained by comparing the true 305d milk yield with the predicted 305d yield (i.e., the summation of all the predicted yields in a certain cycle). The worst performing baseline model was the Wood's model, with an average MAE and a MAPE of 9.58 kg and 30% for the daily milk yield and 2338.56 kg and 18% for the 305d yield, respectively. Using the average curves per herd and per parity as predictions on the other hand resulted in a RMSE of 7.97 kg and a MAPE of 25% for the daily predictions. The SLMYP performed best on the daily as well as 305d yield predictions with respect to every metric. On average, the SLMYP's RMSE between the daily milk yield predictions and the true values comprised 7.38 kg. For the 305d milk yield, the prediction error obtained by the SLMYP was 11% and was 2 percentage points lower than the best performing baseline model. **Fig.** 3.5 visualizes the predictions made by the baseline models as well as by the SLMYP for two random examples from the test set. **Fig.** 3.5a shows how the SLMYP is better able to model the peak of the lactation curve compared to the predictions made by the baseline models. **Fig.** 3.5b on the other hand shows that in contrast to the baseline models, the SLMYP is able to predict lower milk yield returns than expected.

The impact of the different loss functions on the SLMYP's performance for different forecasting windows is displayed in **Table** 3.3. As expected, the SLMYP trained with the step loss function achieved the lowest MAE for the first recorded milk yields in the predicted lactation cycle. In particular, by applying the step loss function, the MAE was 5.57 kg for the first week of lactation and 5.67 kg for the first month of lactation. For larger windows however, the performance obtained by the step loss started to decrease rapidly, with the MAE being 6.81 kg for the first 60 days and 10.05 kg for the entire lactation cycle. In contrast, the MAE of the SLMYP trained with the uniform and logarithmic weights never exceeded 6.0 kg

Model	Daily yield				305d yield			
	RMSE	MAE	MAPE	ρ	RMSE	MAE	MAPE	ρ
Baseline 1	9.22	7.13	0.26	0.62	2071.33	1081.01	0.16	0.61
Baseline 2	7.97	6.21	0.25	0.70	1705.09	1346.82	0.13	0.61
Baseline 3	9.58	7.67	0.30	0.53	2338.56	1887.68	0.18	0.18
SLMYP	7.38	5.58	0.23	0.75	1448.95	1093.30	0.11	0.73

Table 3.2: Performance on daily as well as 305d yield of the SLMYP and baseline models. Baseline 1 = lactation curve of preceding cycle, Baseline 2 = average lactation per herd per parity, Baseline 3 = Wood's curve



(a) (b)

Figure 3.5: Visualization of SLMYP and baseline predictions for two random examples of test set

for every possible window. Furthermore, the model trained with the logarithmic loss function performed better than the model trained with the uniform loss weights for the first 180 days of lactation. For larger windows, both models performed equally well except for the largest possible forecasting window in which the model with the uniform loss achieved the best results overall.

Furthermore, the performance of the SLMYP trained with the logarithmic loss function as well as the best performing baseline model was compared with the SAE presented in **Chapter** 2 during the predicted period. More specifically, **Fig.** 3.6 compares the SAE's predictions for increasing windows of observed milk yields in the predicted period with the SLMYP's predictions made in the preceding cycle. For each possible window of observed yields, the performance of both models was calculated for the remaining unknown part of the lactation curve. As expected, the SAE's daily as well as 305d predictions deviated a lot from the true values in the beginning of the lactation cycle but became more accurate as more milk yields were

	Forecast horizon												
Loss	7	14	21	30	60	90	120	150	180	210	240	270	300
Uniform	5.67	5.58	5.63	5.72	5.82	5.78	5.74	5.71	5.68	5.65	5.62	5.59	5.58
Logarithmic	5.58	5.50	5.56	5.67	5.79	5.76	5.73	5.70	5.67	5.65	5.62	5.59	5.59
Step	5.57	5.49	5.56	5.67	6.81	9.75	10.02	10.07	9.99	9.89	9.87	9.91	10.05

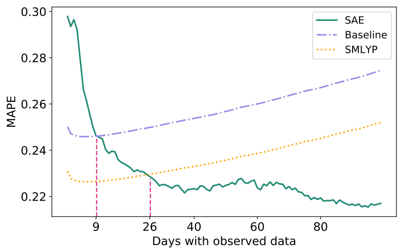
Table 3.3: Performance of SLMYP trained with different loss functions for multiple forecast horizons in terms of MAE (kg)

observed. More specifically, the SAE's MAPE for the daily predictions decreased from 30% when no observations were available to 23% when 20 milk yields were recorded. On the contrary, the MAPE of the SLMYP remained below 23% for the first 20 days of lactation. However, as the beginning of the forecasting window was shifted towards the end of the lactation period, the MAPE of the SLMYP for the daily yields increased gradually as the MAE remains more or less constant while the average milk yield becomes smaller towards the end of the lactation cycle. From 26 days onwards, the SAE outperformed the SLMYP in terms of MAPE by leveraging the observed data. On the other hand, the MAPE of the best performing baseline model that predicted a subsequent lactation curve as the herd's average milk yield per parity was already surpassed by the SAE after 9 days of observed data. For the 305d predictions, the performance of the SAE, SLMYP and baseline model increased for larger windows of recorded milk yields as the predicted 305d yield comprises the cumulative predicted yield as well as the cumulative observed yield. Yet, in this case, the SAE's performance surpassed the baseline model already at the 4th day of lactation while it surpassed SLMYP's performance at the 20th day of lactation, with the MAPE of both models being around 11% at that day.

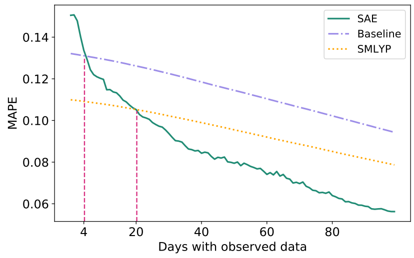
3.4.3 Variable Importance

The variable importance scores of each group of features are visualized in **Fig.** 3.7. The latent representation of the average milk yield curve per herd and per parity had the highest VI of 1.47, meaning that the SLMYP's total RMSE of 7.38 kg increased to 10.84 kg when the latent features were randomly permutated. The animal KPIs and latent representation of the previous lactation curve were the second and third most discriminative feature groups with VI scores of 1.08. The least discriminative feature group were the herd KPIs with a VI score of 1.02.

The impact of several features on the SLMYP's predictions for a random test



(a) Daily performance



(b) 305d performance

Figure 3.6: Daily and 305d milk yield performance of the SAE, SLMYP and baseline models for different windows of observed data in the predicted period

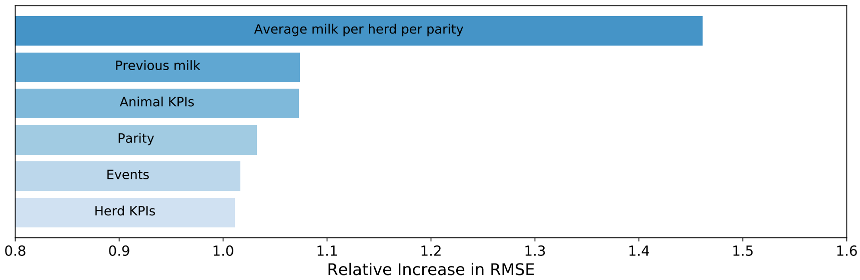
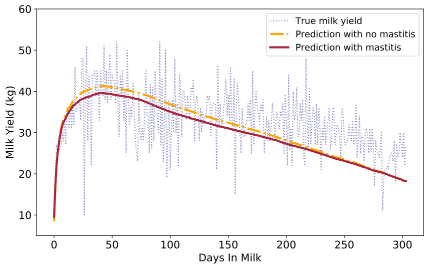
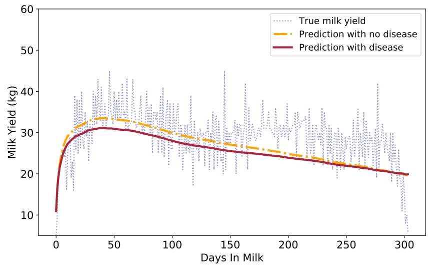


Figure 3.7: Variable importance scores of feature groups

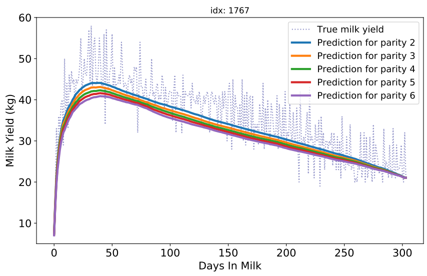
observation are depicted by **Fig.** 3.8. When mastitis was manually injected at the 170th day of the previous lactation cycle, the SLMYP predicted a slightly lower milk production than when the cow would be healthy. Likewise, the SLMYP adjusted the milk yield curve downwards when the disease event was injected at the 174th day of the previous lactation cycle. For each consecutive parity, the SLMYP predicted a slightly lower milk yield curve for a fixed lactation curve in the preceding cycle. When one of the latent variables of the low-dimensional representation of the previous lactation curve that was related to the cow's persistency was increased to its maximum value of 1, a more gradual decline of the subsequent curve was predicted as well. Finally, when the animal KPI corresponding to the 305d milk yield was decreased to its minimum value of 0, the SLMYP predicted lower returns for the entire lactation cycle. The milk yield curve was shifted upwards when the herd KPI related to the average days open was set to its minimum value of 0.



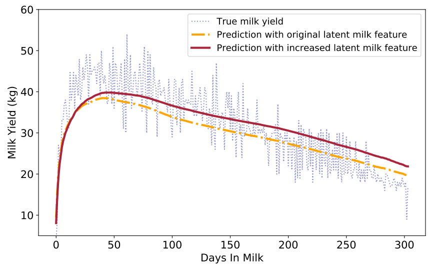
(a) Impact mastitis



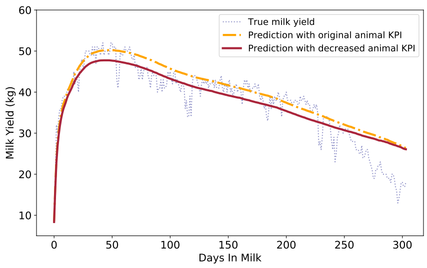
(b) Impact disease



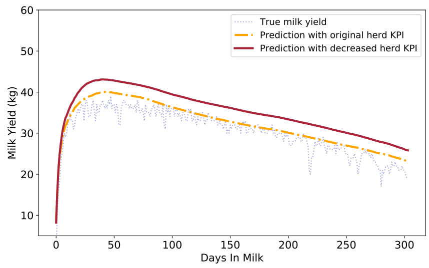
(c) Impact parity



(d) Impact latent variable previous milk yield



(e) Impact animal KPI 305d milk yield



(f) Impact herd KPI avg days open

Figure 3.8: Variable Importance

3.5 Discussion

Overall, the SLMYP performed better than all the baseline models with an average correlation of 0.75 between the predictions and the true values for the daily milk yields. The SAE presented in **Chapter** 2 on the other hand, obtained a correlation of 0.77 when 30 days of data were observed. Furthermore, the SLMYP made the most accurate 305d milk yield forecasts with an average prediction error of 11%. This error is slightly higher compared to the results found by Grzesiak et al. (2003a) who reported a prediction error of 9% by making use of a spline model with one test-day (TD) record observed during the first 28 days of lactation. By applying autoregressive models on 2 TD records observed in the first 2 months of lactation, Macciotta et al. (2002); Vasconcelos et al. (2004) reported correlations between the predicted and true 305d milk yields of 0.85 and 0.88 respectively. For a comparable window of available data, the SAE proposed in **Chapter** 2 obtained an even higher correlation of 0.90. On the contrary, the SLMYP achieved a correlation of 0.73 between the predicted and true 305d values. Yet, while the previously mentioned studies leveraged data observed during early lactation, the SLMYP generates its milk yield predictions before the start of lactation. As shown by **Fig.** 3.6, the SLMYP was still able to produce better predictions for the daily milk yields than the SAE until the 26th day of the lactation cycle. In terms of 305d milk yields, the SLMYP generated better predictions than the SAE until the 20th day of the lactation cycle. Hence, during the first 26 days of lactation, the SLMYP enables farmers to obtain more accurate estimates of milk losses in early lactation and hence facilitate animal monitoring systems. In addition, the model presented in this study allows farmers to increase their forecast horizon with respect to their herd's total productivity with 20 days on average. After that period, predictions of lactation models such as the SAE become more accurate as these model are able to leverage the observed milk yields.

While no studies currently exist that predict the entire lactation curve by using data obtained in the previous cycle, curve fitting models such as those proposed by Ali and Schaeffer (1987); Wilmink (1987); Wood (1967) are able to make forecasts for future cycles by fitting curves on lactation data of homogeneous groups of animals. The resulting parameters of the fitted curves thus describe the group's average production and hence do not incorporate historical information from individual animals. The SLMYP however generates its predictions by taking into account both group statistics as well as individual information regarding historical milk production and reproduction as well as health events. In addition, Silvestre et al. (2006) showed that the accuracy of curve fitting models heavily depends on the sampling properties of the recorded milk yields. This is less a problem for SLMYP as it uses the latent representation of the historical as well as the herd's average milk yield extracted by the SAE. More specifically, as was shown in **Chapter** 2, the SAE's MAPE for reconstructing the entire lactation curve decreased by a maximum of 2 percentage points when 60% of the input milk yields were randomly dropped. Using the latent representation instead of the raw milk yield sequences thus makes the SLMYP particularly robust for missing data in the features corresponding to the cow's and herd's lactation curves.

Finally, as shown by **Fig.** 3.7, the latent representation of the herd's average milk yield curve corresponding to the predicted lactation cycle contributed the most to the predictions. This could be expected since cows from the same herd usually share the same breed, feeding systems, herd management and climate, which have been shown to significantly affect milk production (Rekik and Gara, 2004). In addition to variables summarizing the average herd production, other features related to herd management also had an impact on the milk production. In **Fig.** 3.8f for example, the SLMYP increased its predictions for a lower value of the average number of days open. This is not surprising as a delay in pregnancy will inversely influence milk production (Cattaneo et al., 2015). In contrast to previous studies which found that cows generally produced more in each subsequent lactation cycle (Ehrlich et al., 2011; Macciotta et al., 2011), the SLMYP predicted lower milk yield returns for higher parities. Yet, this could be explained by the fact that the curves corresponding to the previous lactation cycle as well as to the herd's average lactation remained fixed and hence became abnormally low for higher parities. Furthermore, the results presented in this research are in line with previous studies in that a cow's historical milk production is positively related to its future production (Ali and Schaeffer, 1987). As shown by **Fig.** 3.8e and **Fig.** 3.8d, the SLMYP lifted the predicted curve upwards as the total milk production of the preceding cycle was increased or when the persistency of the preceding cycle was positively adjusted. On the contrary, a loss in milk yield can be expected in case of mastitis or disease (Adriaens et al., 2018). This was shown by **Fig** 3.8a and **Fig** 3.8b in which the SLMYP slightly adjusted its predictions downwards when the cow was sick during the preceding cycle. Hence, the SLMYP is able to generate more realistic milk yield predictions by taking into account the sequence of reproduction and health events. As a result, differences between expected and produced milk yields can be calculated more accurately which improves disease detection.

3.6 Conclusion

Current lactation models rely on a fixed number of milk yields recorded in early lactation to forecast individual milk yield curves. As a result, animal monitoring becomes particularly difficult in early lactation as expected milk yields are often missing in the period immediately after calving. In addition, forecasts of a cow's total productivity can only be obtained from the moment the model's last required milk yield input is observed. Curve fitting models on the other hand are able to generate entire lactation curves. These curves however represent group averages and hence remain constant irrespective of the animal's individual variation. In this study, we present a framework that, in addition to herd statistics, uses the cow's historical sequence of milk yields as well as reproduction and health events in the preceding cycle to predict the cow's entire lactation curve in the subsequent cycle. Results show that by leveraging individual data, the model is able to generate more accurate predictions than by solely using group averages. As a result, the framework presented in this research can be used to assess the impact of herd management, health and reproduction events as well as a cow's historical milk yield on the cow's future productivity. In addition, the model allows to increase the farmer's forecast horizon with respect to the herd's future productivity as well as to improve animal monitoring systems in early lactation.

References

Adriaens, I., Huybrechts, T., Aernouts, B., Geerinckx, K., Piepers, S., Ketelaere, B. D., and Saeys, W. Method for short-term prediction of milk yield at the quarter level to improve udder health monitoring. *Journal of Dairy Science*, 101(11): 10327 – 10336, 2018. ISSN 0022-0302.

Ali, T. E. and Schaeffer, L. R. Accounting for covariances among test day milk yields in dairy cows. *Canadian Journal of Animal Science*, 67(3):637–644, 1987.

Breiman, L. Random forests. *Machine Learning*, 45(1):5–32, 2001.

Brody, S., Ragsdale, A. C., and Turner, C. W. The rate of decline of milk secretion with the advance of the period of lactation. *The Journal of General Physiology*, 5(4):441–444, 1923.

Cattaneo, L., Baudracco, J., Lazzarini, B., and Ortega, H. Methodology to estimate the cost of delayed pregnancy for dairy cows. An example for Argentina. *Revista Brasileira de Zootecnia*, 44(6):226–229, 2015.

Dematawewa, C., Pearson, R., and VanRaden, P. Modeling extended lactations of Holsteins. *Journal of Dairy Science*, 90(8):3924 – 3936, 2007. ISSN 0022-0302.

Ehrlich, J. et al. Quantifying shape of lactation curves, and benchmark curves for common dairy breeds and parities. *Bovine Practitioner*, 45(1):88, 2011.

Gorgulu, O. Prediction of 305-day milk yield in Brown Swiss cattle using artificial neural networks. *South African Journal of Animal Science*, 42(3):280–287, 2012.

Grzesiak, W., Lacroix, R., Wójcik, J., and Blaszczyk, P. A comparison of neural network and multiple regression predictions for 305-day lactation yield using partial lactation records. *Canadian Journal of Animal Science*, 83(2):307–310, 2003a.

Grzesiak, W., Wojcik, J., and Binerowska, B. Prediction of 305-day first lactation milk yield in cows with selected regression models. *Archives Animal Breeding*, 46(3):213–224, 2003b.

Grzesiak, W., Błaszczyk, P., and Lacroix, R. Methods of predicting milk yield in dairy cows—predictive capabilities of Wood's lactation curve and artificial neural networks (ANNs). *Computers and Electronics in Agriculture*, 54(2):69 – 83, 2006. ISSN 0168-1699.

Græsbøll, K., Kirkeby, C., Nielsen, S. S., Halasa, T., Toft, N., and Christiansen, L. E. Models to estimate lactation curves of milk yield and somatic cell count in dairy cows at the herd level for the use in simulations and predictive models. *Frontiers in Veterinary Science*, 3:115, 2016. ISSN 2297-1769.

Gröhn, Y., Wilson, D., González, R., Hertl, J., Schulte, H., Bennett, G., and Schukken, Y. Effect of pathogen-specific clinical mastitis on milk yield in dairy cows. *Journal of Dairy Science*, 87(10):3358 – 3374, 2004. ISSN 0022-0302.

Hinton, G., Srivastava, N., Krizhevsky, A., Sutskever, I., and Salakhutdinov, R. Improving neural networks by preventing co-adaptation of feature detectors. *ArXiv e-prints*, 2012.

Ioffe, S. and Szegedy, C. Batch normalization: Accelerating deep network training by reducing internal covariate shift. *ArXiv e-prints*, 2015.

Kingma, D. and Ba, J. Adam: A method for stochastic optimization. *ArXiv e-prints*, 2014.

Kirkpatrick, M., Hill, W. G., and Thompson, R. Estimating the covariance structure of traits during growth and ageing, illustrated with lactation in dairy cattle. *Genetical Research*, 64(1):57–69, 1994.

Lacroix, R., Wade, K., Kok, R., and Hayes, J. Prediction of cow performance with a connectionist model. *Transactions of the ASAE*, 38(5):1573–1579, 1995.

Lacroix, R., Salehi, F., Yang, X., and Wade, K. Effects of data preprocessing on the performance of artificial neural networks for dairy yield prediction and cow culling classification. *Transactions of the ASAE*, 40(3):839–846, 1997.

Liseune, A., Salamone, M., Van den Poel, D., Van Ranst, B., and Hostens, M. Leveraging latent representations for milk yield prediction and interpolation using deep learning. *Computers and Electronics in Agriculture*, 175:105600, 2020. ISSN 0168-1699.

Macciotta, N. P., Dimauro, C., Rassu, S. P., Steri, R., and Pulina, G. The mathematical description of lactation curves in dairy cattle. *Italian Journal of Animal Science*, 10(4):e51, 2011.

Macciotta, N., Vicario, D., Pulina, G., and Cappio-Borlino, A. Test day and lactation yield predictions in Italian Simmental cows by ARMA methods. *Journal of Dairy Science*, 85(11):3107 – 3114, 2002. ISSN 0022-0302.

Murphy, M., O'Mahony, M., Shalloo, L., French, P., and Upton, J. Comparison of modelling techniques for milk-production forecasting. *Journal of Dairy Science*, 97(6):3352 – 3363, 2014. ISSN 0022-0302.

Murphy, M., Zhang, F., Upton, J., Shine, P., and Shalloo, L. *A review of milk production forecasting models*, pages 14 – 61. 04 2018.

Njubi, D., Wakhungu, J., and Badamana, M. Use of test-day records to predict first lactation 305-day milk yield using artificial neural network in Kenyan Holstein–Friesian dairy cows. *Tropical animal health and production*, 42(4):639–644, 2010.

Olori, V., Brotherstone, S., Hill, W., and McGuirk, B. Fit of standard models of the lactation curve to weekly records of milk production of cows in a single herd. *Livestock Production Science*, 58(1):55 – 63, 1999. ISSN 0301-6226.

Rekik, B. and Gara, A. Factors affecting the occurrence of atypical lactations for Holstein–Friesian cows. *Livestock Production Science*, 87(2):245 – 250, 2004. ISSN 0301-6226.

Rumelhart, D. E., Hinton, G. E., and Williams, R. J. Learning representations by back-propagating errors. *Nature*, 323(6088):533–536, 1986.

Salehi, F., Lacroix, R., and Wade, K. Improving dairy yield predictions through combined record classifiers and specialized artificial neural networks. *Computers and Electronics in Agriculture*, 20(3):199 – 213, 1998. ISSN 0168-1699.

Sanzogni, L. and Kerr, D. Milk production estimates using feed forward artificial neural networks. *Computers and Electronics in Agriculture*, 32(1):21 – 30, 2001. ISSN 0168-1699.

Sharma, A. K., Sharma, R., and Kasana, H. Empirical comparisons of feed-forward connectionist and conventional regression models for prediction of first lactation 305-day milk yield in Karan Fries dairy cows. *Neural Computing & Applications*, 15(3-4):359–365, 2006.

Sharma, A. K., Sharma, R., and Kasana, H. Prediction of first lactation 305-day milk yield in Karan Fries dairy cattle using ANN modeling. *Applied Soft Computing*, 7(3):1112 – 1120, 2007. ISSN 1568-4946.

Silvestre, A., Petim-Batista, F., and Colaço, J. The accuracy of seven mathematical functions in modeling dairy cattle lactation curves based on test-day records from varying sample schemes. *Journal of Dairy Science*, 89(5):1813 – 1821, 2006. ISSN 0022-0302.

Tekerli, M., Akinci, Z., Dogan, I., and Akcan, A. Factors affecting the shape of lactation curves of Holstein cows from the Balikesir province of Turkey. *Journal of Dairy Science*, 83(6):1381 – 1386, 2000. ISSN 0022-0302.

Torres, M., Hervás, C., and Amador, F. Approximating the sheep milk production curve through the use of artificial neural networks and genetic algorithms. *Computers & Operations Research*, 32(10):2653 – 2670, 2005. ISSN 0305-0548.

Vasconcelos, J., Martins, A., Petim-Batista, M., Colaço, J., Blake, R., and Carvalheira, J. Prediction of daily and lactation yields of milk, fat, and protein using an autoregressive repeatability test day model. *Journal of Dairy Science*, 87(8): 2591 – 2598, 2004. ISSN 0022-0302.

Wilmink, J. Adjustment of test-day milk, fat and protein yield for age, season and stage of lactation. *Livestock Production Science*, 16(4):335 – 348, 1987. ISSN 0301-6226.

Wilson, D., González, R., Hertl, J., Schulte, H., Bennett, G., Schukken, Y., and Gröhn, Y. Effect of clinical mastitis on the lactation curve: A mixed model estimation using daily milk weights. *Journal of Dairy Science*, 87(7):2073 – 2084, 2004. ISSN 0022-0302.

Wood, P. Algebraic model of the lactation curve in cattle. *Nature*, 216(5111): 164–165, 1967.

Zhao, B., Lu, H., Chen, S., Liu, J., and Wu, D. Convolutional neural networks for time series classification. *Journal of Systems Engineering and Electronics*, 28(1): 162–169, Feb 2017. ISSN 1004-4132.

Zheng, Y., Liu, Q., Chen, E., Ge, Y., and Zhao, J. L. Time series classification using multi-channels deep convolutional neural networks. In *Web-Age Information Management*, pages 298–310, 2014. ISBN 978-3-319-08010-9.

4

Leveraging sequential information from multivariate behavioral sensor data to predict the moment of calving in dairy cattle using deep learning[1]

4.1 Abstract

Calving is one of the most critical moments during the life of a cow and their calves. Timely supervision is therefore crucial for animal welfare as well as the farm economics. In this study, we propose a framework to predict calving within 24h, 12h, 6h, 3h and 1h of dairy cows using sequential sensor data. In particular, data were extracted from 2363 cows coming from 8 commercial farms between August 2016 and November 2020. Two sensors attached to the neck and leg of each cow measured rumination, eating, lying, standup, walking and inactive behavior on a minute basis. A novel methodology was used to impute the missing values in the sensor sequences by leveraging the observed values of all the behavioral activities recorded by the sensors. A deep learning model was then used to predict the moment of calving on an hourly basis using the imputed sensor sequences.

[1]Based on: Liseune, A., Van den Poel, D., Hut, P. R., Hostens, M., & van Eerdenburg, F. J.C.M. (2021). Leveraging sequential information from multivariate behavioral sensor data to predict the moment of calving in dairy cattle using deep learning. *Journal of Computers and Electronics in Dairy Science.*

Results show that for the 24h interval, the model achieves a Sensitivity of 65% with a Precision of 77%, while for the 3h interval, the model achieves a Sensitivity of 57% with a Precision of 49%. Moreover, we find that using the missing value imputations significantly improves the predictive performance for observations containing up to 60% of missing values. The framework proposed in this study can be used by farmers to optimize their calving management and hence improve animal monitoring.

4.2 Introduction

Calving is one of the most critical moments of both the cow's and the calf's life (Barrier et al., 2013; Mee, 2013). Dystocia, i.e. difficulties or abnormalities encountered during calving, can severely affect health and welfare of dairy cattle (Barrier and Haskell, 2011). In particular, dams that experience dystocia can be at increased risk of injury as well as contracting uterine diseases such as metritis and endometritis (Rutten et al., 2017). Moreover, it is reported that dystocia is one of the most painful conditions for dairy cows (Laven et al., 2009). Dystocial calves on the other hand can experience many physiological problems such as prolonged hypoxia and significant acidosis (Lombard et al., 2007) as well as physiological stress and internal injuries (Berglund et al., 2003). This in turn can reduce the calf's long-term survival or even result in stillbirth (Lombard et al., 2007). In fact, 7% of all the calves born in the United States die directly within 48 hours and 50% of the stillbirths can be directly attributed to dystocia (Meyer et al., 2000). Difficulties with calving can therefore negatively affect animal welfare as well as farm economics (Mee, 2004). Specifically, dystocia can be very costly to dairy farmers as it is associated with a lower fertility, milk production and survival rate of the dam (Tenhagen et al., 2007). Additionally, the need for veterinary assistance contributes to the economic cost of dystocia. In particular, the total cost associated with a difficult calving has been estimated at €500 (McGuirk et al., 2007). The financial losses related to stillbirth even average $938 per case (Mahnani et al., 2018). Reducing difficulties with calving is therefore crucial to the dairy producer.

Several risk factors causing dystocia include parity, calf weight, sex, body size and pelvic diameters of the dam as well as seasonal effect and environmental stress (Tenhagen et al., 2007). Yet, farm management such as breeding decisions and human supervision can strongly influence calving difficulties as well (Rutten et al., 2017; Van Pelt and de Jong, 2011). More specifically, it has been shown that providing timely human intervention reduces the risk of dystocia, the pain experienced during labor and the reproductive decline of the dam (Borchers et al., 2017). Individual animal monitoring, however, becomes increasingly more difficult as the number of cattle per farm globally increases over time (Raussi, 2003). In fact, even with intensive monitoring, it remains difficult to forecast the moment of calving correctly (Lange et al., 2017). One way to organize human supervision more efficiently is by employing models that are able to accurately predict the moment of parturition. Such models can automatically alert farmers of an imminent calving and hence facilitate timely calving supervision (Ouellet et al., 2016). Physical and behavioral changes may provide clues to detect when cows are about to calve (Huzzey et al., 2005). More specifically, it has been shown that behaviors such as eating, rumination and grooming decrease, while restlessness and lying bouts increase during the period around calving (Jensen, 2012; Miedema et al., 2011; Schirmann et al., 2013). Visually assessing these behavioral changes, however, is subjective, time consuming and prone to human error (Ouellet et al., 2016). Several frameworks were, therefore, presented to predict the onset of calving by automatically processing these changes in behavioral patterns. Ouellet et al. (2016), for example, constructed three different models to predict the moment of calving based on four calving indicators, i.e., vaginal temperature, rumination time, lying time and lying bouts. More specifically, three logistic regression models were built that predicted the start of parturition within 24h, 12h and 6h based on the optimal combination of the four aforementioned indicators. Similarly, Fadul et al. (2017) trained a logistic regression model with a stepwise selection procedure to predict the onset of calving within the next 3 hours based on rumination time and chews, lying bouts, boluses as well as other activities not related to ruminating, feed intake or drinking. Whereas the two previously mentioned studies removed missing data, all observations with missing values were assigned to the training set in the study presented by Zehner et al. (2019). A Naive bayes model was then trained and evaluated on a validation set, which exclusively consisted of observations with complete information. Rutten et al. (2017) on the other hand, presented a methodology to impute the missing values by a weighted average of sensor data recorded during the previous three days at the same time period. A logistic regression model was then trained on the imputed data to generate the calving predictions. Borchers et al. (2017) applied more complicated machine learning techniques such as random forests, linear discriminant analysis and neural networks to predict the start of calving. The same dataset was used in a subsequent study conducted by Keceli et al. (2020) who applied a Bidirectional Long Short-Term Memory (Bi-LSTM) to process the data sequentially.

Yet, in most of the previously mentioned studies, the proposed frameworks disentangle the temporal information in the sensor sequences. As result, these models are not able to leverage the sequential patterns in the the behavioral changes, which can negatively affect model performance. Additionally, the previously presented frameworks are difficult to generalize and may not be suitable for practical applications. In particular, in most studies, observations with missing values are removed (Borchers et al., 2017; Fadul et al., 2017; Keceli et al., 2020; Ouellet et al., 2016; Zehner et al., 2019). As a result, these models won't be able to generate reliable calving predictions when missing values are present in the observed sensor sequences. In one study, however, missing values were imputed by the moving averages of sensor recordings observed at previous time steps (Rutten et al., 2017). Yet, in case of large periods with missing data, this approach will also not be able to impute the missing values in a reliable way. Furthermore, all of the previously mentioned studies were conducted on datasets with a limited number of recorded calvings, mostly coming from one herd. Hence, the reported performance scores of these models were obtained on very limited test observations and are thus difficult to generalize towards calving events not observed in the data.

In order to fill this gap in literature, we present a framework that is generalizable and suitable for practical implementations. More specifically, this study was conducted on a large dataset containing sensor data coming from 2363 animals from 8 different herds (Hut et al., 2021). Additionally, we propose a novel methodology to infer missing values by leveraging the values recorded by all the sensors. Finally, we present a model that accurately predicts the moment of calving by sequentially processing the multivariate sensor sequences. There are several reasons why we believe that the proposed framework can be valuable for calving management. First, human supervision for calving can be organized better as farmers are automatically alerted when a cow goes into labor. This way, stock personnel does not need to permanently supervise their cattle. Second, cow welfare can be drastically increased as timely supervision significantly reduces the negative consequences of dystocia (Borchers et al., 2017; Schuenemann et al., 2011, 2013; Szenci et al., 2012). Finally, we propose a model that generates reliable predictions, irrespective of the data quality of the recorded sensor data. This is a valuable tool as missing values and outliers frequently occur in sensor sequences due to faulty data transmission or malfunction of the sensors.

4.3 Materials and Methods

4.3.1 Data

For this study, data was collected from 2363 cows coming from 8 commercial dairy farms with freestall barns in the Netherlands between August 2016 and November 2020. No external personnel was employed by the farms. From the 8 farms, 6 farms were Holstein Friesian, 1 were Fleckvieh and 1 farm were crossbreeding Holstein Friesian, Fleckvieh and Scandinavian Red. From the moment the Nedap infrasture (Nedap, Groenlo, The Netherlands) was completely implemented at a farm, each cow was equipped with the Nedap Smarttag Leg and Nedap Smarttag Neck sensor for the entire period of this study. The sensors were attached to the front legs and the neck of the cow with the former recording the number of steps, standing time, walking time and lying time and the latter recording the eating time, rumination time and inactive time, i.e., time not spend eating and ruminating. Sensor data was recorded every minute. Hourly as well as daily measurements were obtained by summing all the values of each activity recorded during each hour and day respectively. For the data aggregated on a daily basis, the data supplier provided some additional features, e.g., the number of bouts, the average bout length as well as the average length between different bouts for several activities. Additionally, the parity, i.e. the number of different times a dam has had an offspring, and the season of calving (summer, spring, autumn, winter) were provided for each calving event. Table 4.1 shows the raw sensor recordings as well as the derived features on an hourly and daily basis obtained from the data provider.

Activity	**1h features**	**24h features**
Walking	minutes	minutes
Standing	minutes	minutes
		number of bouts
Eating	minutes	minutes
		number of bouts
		avg bout minutes
		avg inter bout minutes
Rumination	minutes	minutes
		number of bouts
		avg bout minutes
		avg inter bout minutes
Lying	minutes	minutes
		number of bouts
		avg bout minutes
Inactivity	minutes	minutes
		number of bouts
		avg bout minutes
		avg inter bout minutes
Leg activity	number of steps	number of steps

Table 4.1: The sensor activities and their corresponding features recorded on an hourly and daily basis

The moment of calving was manually recorded by the farmer. In total, the day of calving was registered for 3902 different calvings. For 572 of these calvings, the exact timestamp was registered by the farmer at the moment the farmer visually observed the parturition. In total, 159 calvings were registered in the morning (from

6am to 12pm), 170 in the afternoon (from 12pm to 6pm), 178 in the evening (from 6pm to 12am) and 65 at night (from 12am to 6am). For each calving event, the daily features observed during the 21 days before calving were extracted. For calving events with an exact time stamp, the hourly sensor values were extracted from the day before calving until the moment of parturition. The training, validation and test set for the daily and hourly data were constructed by randomly sampling 60%, 20% and 20% of the daily and hourly calvings respectively. In order to extract the features and labels of the daily calvings, a sliding window of 14 days was shifted over the daily sequences by one day. This resulted in 8 observations for each calving event, with one observation containing the sensor sequences observed the day before calving, and 7 observations with sensor sequences observed 2 or more days before calving. For calving events with an exact time stamp, a sliding window of 24 hours was then shifted over the hourly sequences by one hour. A graphical depiction of the validation setup and the sliding window procedure is given by **Fig.** 4.1. In total, 31216 sequences of daily data X^d and 8275 sequences of hourly data X^h were extracted. For the sensor data aggregated on a daily basis, each observation X_i^d contained 14 recordings $X_{i_t}^d$, with each recording comprising 19 sensor values. For the sensor data recorded on an hourly basis, every observation X_i^h consisted of 24 recordings $X_{i_t}^h$ of 7 sensor values. Outliers were removed by the median absolute deviation method (Leys et al., 2013). This method consists of removing observations according to the absolute difference between the observation and the median value. Hence, it is more robust for extreme outliers. In total, 4783 outlying sensor recordings were replaced by missing values, which comprises 0.3% of the data. After the removal of the outliers, sensor values were normalized between 0 and 1. Table 4.2 gives an overview of the data used in this study.

	Hourly Prediction	**Daily Prediction**
Calving events	572	3902
Farms	8	8
Parity 1	110	782
Parity 2	148	927
Parity 3+	314	2193
Recording interval	1h	24h
Sliding window size	24h	14d
Number of sequences	8275	31216
Features	7	19
Training size	4896	18728
Validation size	1721	6240
Test size	1658	6248

Table 4.2: Overview of data used in this study

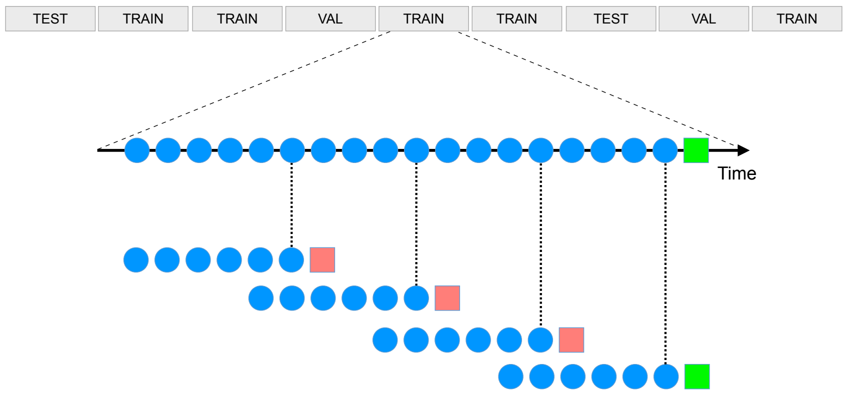


Figure 4.1: Schematic overview of validation setup and the sliding window procedure. Each sequence containing the sensor data observed before a calving of a specific cow (gray square) is randomly assigned to the training, validation or test set. A sliding window is then applied to each sequence to produce observations with positive (green square) and negative (red square) calving events.

4.3.2 Deep Learning Models

Mutlilayer Perceptron Models (MLP) are a type of neural networks and consist of an input layer, one ore more hidden layers and an output layer. In each hidden layer, every neuron is a linear combination of all the neurons from the previous layer, followed by a non-linear activation function. In particular, if h_j represents the outputs of layer j, then the output of layer $j+1$ can be calculated as follows:

$$h_{j+1} = f(h_j \cdot W_{j+1} + b_{j+1})$$

with W_{j+1} and b_{j+1} being the weights matrix and biases corresponding to layer $j+1$, and f being a non-linear activation function, commonly a ReLU function. The activation function of the final layer is generally a sigmoid, softmax or identity function, depending on whether the label is binary, multiclass or continuous respectively. In general, MLPs are suitable for any supervised learning task. In practice, however, MLPs are rarely used when temporal or spatial dependencies exist among the features of the input data. Long Short-Term Memory Models (LSTM) on the other hand, have been specifically designed to process time-series data as they have recurrent connections between the different inputs (Hochreiter and Schmidhuber, 1997). In particular, information from each time step t is fed to an LSTM unit, which is composed out of four units: a memory cell c_t, an input gate i_t with the corresponding weight matrices W_R^i, W_I^i and b^i, an output gate o_t with the corresponding weight matrices W_R^o, W_I^o and b^o and a forget gate f_t with the corresponding weight matrices W_R^f, W_I^f and b^f, as shown by the following equations:

$$i_t = \sigma(W_R^i h_{t-1} + W_I^i x_t + b^i)$$

$$f_t = \sigma(W_R^f h_{t-1} + W_I^f x_t + b^f)$$

$$c_t = c_{t-1} \odot f_t + i_t \odot tanh(W_R^c h_{t-1} + W_I^c x_t + b^c)$$

$$o_t = \sigma(W_R^o h_{t-1} + W_I^o x_t + b^o)$$

$$h_t = o_t \odot tanh(c_t)$$

where x_t represents the observed features at the current time step, h_{t-1} represents the output of the previous time step, c_{t-1} represents the cell state of the previous time step and σ and $tanh$ represent the sigmoid and hyperbolic tangent function respectively. The memory cell c_t stores information extracted from the previous time steps and the gates determine the information flow between the cells. The output at the last time step represents a compact summary of the entire observed sequence. Sometimes, however, one LSTM layer is insufficient to compress all the observed data into one single feature vector. In such cases, more informative vectors can be obtained by stacking multiple LSTM layers on top of each other. In contrast to LSTMs, Convolutional Neural Networks (CNN) were originally developed for computer vision applications (Krizhevsky et al., 2012; LeCun et al., 1998; Szegedy et al., 2015). Lately, however, they have also shown great performance on time series data as they can extract time-dependent features in parallel (Zhao et al., 2017). In general, a CNN exists of multiple convolutional blocks, with each block typically comprising a linear transformation and a non-linear activation stage for feature extraction. In particular, for a time series with K features and T time steps, a filter of size KxS with $S < T$ is slided over the sequential data along the time dimension. Each time the filter is shifted one position, the filter weights are multiplied with the elements of the data that are covered by the filter at that point. Subsequently, a non-linear activation function, such as ReLU, is applied to the sum of the outputs of the multiplication and results in a new time series of the features extracted by that filter. In order to downsample the output and to make the model invariant to small translations in the input, a pooling stage or strided convolution is used at some of the layers to summarize the presence of the feature in every specific time window. By applying multiple convolutional blocks and flattening the output of the last layer, a vector is obtained representing all the features extracted from the input data. By altering the number of filters of the filter size, or by adding convolutional blocks, the model can learn more complex patterns. Finally, hybrid approaches are now also used to leverage the unique capabilities of different models. C-LSTM models that combine CNNs with LSTMs for example, have been successfully used to process time-series data (Alhussein et al., 2020; Pak et al., 2018). In these architectures, a CNN first extracts a set of time-dependent features from the input data, while an LSTM then sequentially processes these features and encodes them into a one-dimensional feature vector. The motivation to use this kind of architecture is that the CNN is able to extract meaningful features in parallel from the multivariate timeseries, while the LSTM can extract temporal patterns from long-term sequences.

4.3.3 Missing Value Imputation

A major concern regarding the data quality of sensor data is the frequent occurrence of missing values. Missing gaps in sensor sequences can occur due to several reasons such as malfunction of the sensors and faulty transmission of data. One way of dealing with missing values present in the data sequences is by imputation by the mean, whereby the missing values of a certain feature are replaced by the feature mean. For time series data, however, this often results in unrealistic realizations of sequences, as the imputed value does not take into account the values observed before or after the missing value (Liseune et al., 2020). In contrast, linear and spline interpolation impute missing values by interpolating between known data points. While the linear interpolant equals a straight line between two known points, the spline interpolant is a piecewice polynomial fitted to a small subset of known values. However, in case of a multivariate time series, correlation may exist among the different sequential features, which can not be leveraged by linear or spline interpolation. Hence, in addition to the three previously mentioned imputation methods, a model was also built to impute the missing values for each of the behavioral sequences (e.g. eating) based on the values observed in that behavioral sequence as well as the values recorded in the other behavioral sequences, hereinafter referred to as the dependent and independent sequential features. More specifically, a CNN was used to obtain a one-dimensional vector from the independent sequential features. In order to leverage the values that were observed in the dependent sequential feature, the sequence was used as input as well. During the training stage, observed values of the dependent sequential feature were randomly set to missing to obtain a set of missing and true values. This vector was then concatenated with the CNNs output and was subsequently fed to an MLP which predicted the entire dependent sequential feature. Finally, the mean squared error loss between the values of the dependent sequential feature that were set to missing and the corresponding predictions was calculated and backpropagated through the entire network. For each feature of the daily and hourly data, a missing imputation model was trained and was used to impute all the missing values. An example of how one particular sequential feature is imputed by a missing value imputation model is given in **Fig.** 4.2.

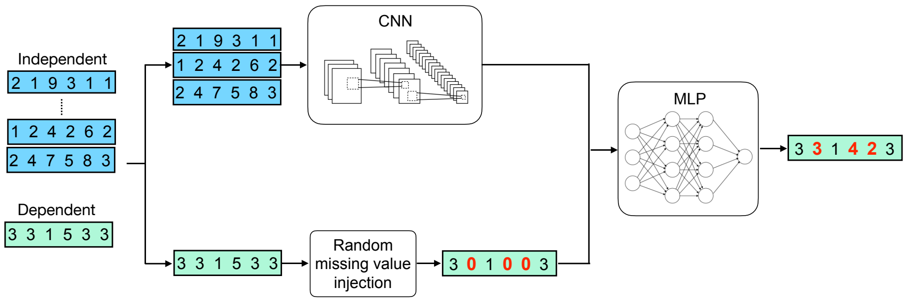


Figure 4.2: Schematic overview of the missing value imputation model

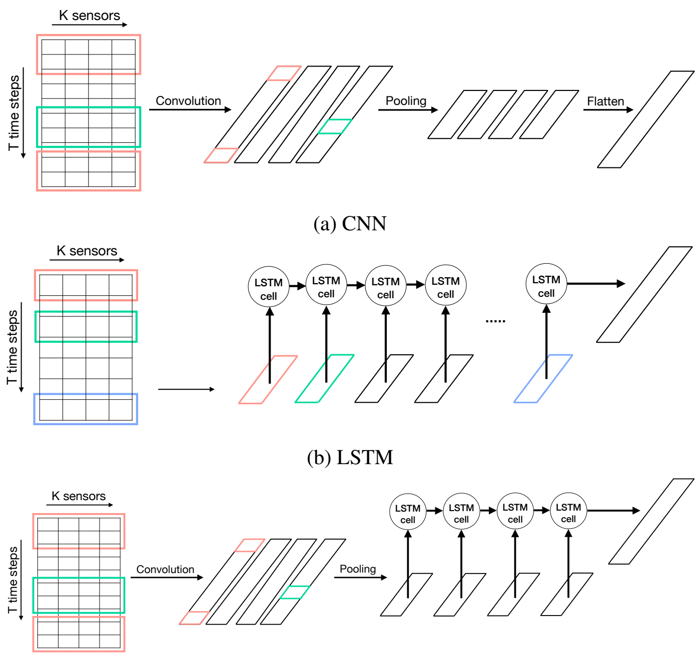
4.3.4 Predictive Models

In order to predict the moment of calving, two machine learning models and three deep learning models were trained on the sensor data. For predicting the moment of parturition within 24h, the sensor data aggregated on a daily basis X^d was used as input. The hourly sensor data X^h was used to predict calving within 12h, 6h, 3h and 1h. For the machine learning models, the data was flattened to obtain non-sequential observations. In particular, for the daily data X^d, each observation X_i^d was flattened by concatenating each of the 14 recordings $X_{i_t}^d$ of 19 sensor values into a one-dimensional vector: $X_{i_{1_1}}^d$, $X_{i_{1_2}}^d$, ..., $X_{i_{14_{19}}}^d$. Likewise, the hourly data was flattened by concatenating each observation's recording into a one-dimensional vector: $X_{i_{1_1}}^h$, $X_{i_{1_2}}^h$, ..., $X_{i_{24_7}}^h$. For every prediction window, each model was trained on the imputed data as well as the raw data with the missing values. Like most of the previous studies, a logistic regression model was trained on the flattened daily and hourly sensor data as this model is not able to sequentially process the input features. For the daily and hourly predictions, the logistic regression model can be expressed as follows:

$$y_i^d = \frac{1}{1 + \exp^{-(\beta_0 + \beta_1 * X_{i_{1_1}}^d + \beta_2 * X_{i_{1_2}}^d + ... + \beta_{266} * X_{i_{14_{19}}}^d)}}$$

$$y_i^h = \frac{1}{1 + \exp^{-(\beta_0 + \beta_1 * X_{i_{1_1}}^h + \beta_2 * X_{i_{1_2}}^h + ... + \beta_{168} * X_{i_{24_7}}^h)}}$$

with y_i^d being the predicted probability of calving the next day for observation i of the daily data and y_i^h being the predicted probability of calving the next 1h, 3h, 6h or 12h for observation i of the hourly data. Additionally, a random forest model was trained on the flattened data as this model does not assume a linear decision boundary, unlike the logistic regression model. In contrast to the machine learning models, three deep learning models that are able to sequentially process the sensor data were used to predict calving. A CNN model was implemented by applying multiple convolutional layers on the time series data. In each layer, several filters were shifted along the time dimension to extract different sets of time-dependent features. Pooling stages were used to downsample the feature space. The output of the last layer was flattened to obtain a vector comprising all the extracted features. For the LSTM model, the sensor values observed at each time step were processed sequentially by one or two LSTM layers. The output of the last LSTM cell was used as a compact summary of all the observed sensor sequences. Finally, the C-LSTM model comprised a CNN and LSTM unit, with the CNN extracting several time-dependent feature vectors by applying multiple convolutional blocks, and the LSTM processing these features sequentially and obtaining a compact feature representation. The feature representations obtained by the LSTM, CNN and C-LSTM models were passed to an MLP with a sigmoid activation function in the final layer to predict the probability of calving. An overview of the three deep learning models applied in this study are shown by **Fig.** 4.3



(a) CNN

(b) LSTM

(c) C-LSTM

Figure 4.3: Deep learning architectures used in this study

4.3.5 Model Training

All the deep learning models were trained by using the backpropagation algorithm (Rumelhart et al., 1986). In this algorithm, the gradient with respect to loss is calculated and propagated through the network by using the chain rule. The Adam gradient-based optimization algorithm was then used to update the weights (Kingma and Ba, 2014). For the missing value imputation model, the mean squared error between the values of the dependent sensor that were randomly set to missing and the corresponding predictions was calculated and backpropagated through the entire network. The negative log-likelihood between the predicted probabilities and the true calving observations was used to train the prediction models. All the models were trained on the training set by using the early stopping procedure in which model training continues as long as the performance on the validation set improves in order to avoid overfitting. The hyperparameters of all the prediction models were tuned using a random search. More specifically, for each training cycle of a model, a hyperparameter setting was determined by randomly sampling values from the model's predefined hyperparameter space. After a predefined number of training cycles, the optimal hyperparameter setting was determined by obtaining the model with the highest validation performance. For the logistic regression model, the regularization method and strength as well as the number of training iterations were optimized. The number and depth of trees as well as the the number of samples required to split an internal node and to be at a leaf node were set as hyperparameters for the random forest model. For the deep learning models, the number of layers, the number of neurons in each layer, the activation function, the dropout rate as well as the inclusion of batch normalization were all considered as tunable parameters. Finally, for every prediction model, the inclusion of the static data features, i.e. the parity and season of calving, as well as the balancing scheme was considered as a hyperparameter as well. In particular, the data could be upsampled or downsampled, the loss function could be weighted with respect to the class proportions or no adjustment could be made to the data. An overview of all the hyperparameters that were assessed for each of the different models is provided in 4.6. In **Fig.** 4.4, a schematic overview of the methodology for each prediction model is given.

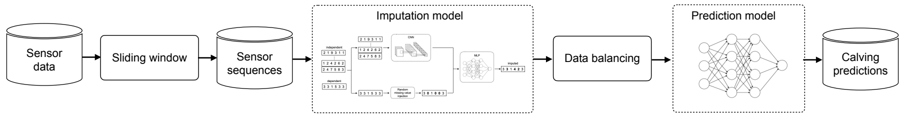


Figure 4.4: Overview of methodology

4.3.6 Model Evaluation

The performance of the prediction models was evaluated by five metrics that are widely accepted as appropriate evaluation metrics for binary classification algorithms, namely the AUC (Area Under ROC Curve), the Sensitivity (Se), the Precision or Positive Predicted Value (PPV), the Specificity (Sp) and the Average Precision (AP). The Se equals the proportion of correctly classified positive examples. The PPV measures how much positive examples were retrieved from the positive predictions. The Sp calculates the proportion of negative examples that were identified by the model. In contrast to the aforementioned evaluation metrics, the AUC and AP are not dependent on a specific threshold, i.e. the cutoff point above which a predicted probability is considered as a positive prediction and a negative prediction otherwise. Hence, these metrics allow to compare how well models are ordering the predictions, without considering any specific decision threshold. The AUC can be interpreted as the probability that a random positive observation gets a higher score than a random negative observation (Bradley, 1997). An AUC score of 0.5 represents a model that does not perform any better than random, while an AUC score of 1 is obtained by a perfect model. In case of imbalanced data, however, it has been shown that the AP is more informative than the AUC when evaluating binary classification models (Saito and Rehmsmeier, 2015). The AP is the area under the precision recall curve (PRC) and indicates how well the model can correctly identify all the positive examples without predicting too much negative examples as positive. A random classifier has an AP equal to the proportion of positive examples while a perfect model has an AP equal to 1.

4.3.7 Model Selection

Each missing value imputation model comprised 5 convolutional layers, with 32, 64, 64, 128 and 128 filters respectively. In each layer, a filter size of 3 and a ReLU activation function was applied. The output of the second and fourth layer was downsampled by applying a stride of 2. The output of the last layer was flattened and passed to an MLP with one hidden layer of size 100 and a ReLU activation function. Each imputation model was evaluated in terms of the RMSE on the validation set every 5000 training iterations with a batch size of 32. Every time the validation RMSE decreased, the model's weights were saved. Training was terminated when the performance did not improve for 5 consecutive times. For each of the predictive models, 50 random hyperparameter configurations were assessed. After convergence on the training set, the AP of the machine learning models on the validation set was calculated. Every deep learning model was evaluated on the validation set in terms of the AP after 1 training epoch. Model weights were saved when the AP on the validation set increased. Training was terminated when the validation AP did not increase for 5 consecutive times. For each predictive model, the parameter configuration that rendered the highest validation performance was retrained on the combination of the training and validation set and was evaluated on the test set. The learning rate applied for the Adam optimization algorithm was 0.001 for both the missing value imputation as well as the deep learning predictive models.

4.3.8 Programming Tools

All data processing and analyses were done in Python 3.9 (Python Software Foundation, https://www.python.org/). The same add-on packages as described in **Chapter 2** 2.3.6 were used.

4.4 Results

4.4.1 Model Performance

The performance of the models in terms of the AP on the test set for each prediction window is presented in **Table** 4.3. For the daily predictions, the deep learning models clearly outperformed the machine learning models on the data with missing values. While the logistic regression and random forest model achieved an AP of 0.32 and 0.65, the LSTM, CNN and C-LSTM obtained AP scores of 0.72, 0.75 and 0.73 respectively. The highest scores, however, were obtained on the imputed data. While the performance of the CNN and C-LSTM increased by 3% and 7% respectively, the AP of the LSTM increased by 12.5% to 0.79, resulting in the best performance on the daily predictions. Likewise, the performance of the LSTM and C-LSTM improved considerably when trained on the imputed data for the smallest prediction window. More specifically, the LSTM's performance increased by 0.02 when trained on the imputed data. The C-LSTM on the contrary, improved its performance from 0.19 to 0.29, which resulted in the highest score for the 1h prediction interval. The added value of the imputations with respect to the predictive performance is also visible for the other prediction intervals. The CNN trained on the imputed data obtained the highest performance scores for the 3h prediction window with an AP equal to 0.49, thereby outscoring the best performing model trained on the missing data with 0.02. For the 6h prediction interval, the best performance was obtained by the C-LSTM and random forest model trained on the imputed data as well as the random forest model trained on the data with missing values.

In contrast to the AP, the Se, Sp and PPV are dependent on the chosen threshold. For each prediction interval, the values of these metrics are therefore shown for different thresholds in **Table** 4.4 for the best performing model, i.e., the C-LSTM model trained on the imputed data for the 6h and 1h prediction interval, the CNN model trained on the imputed data for the 12h and 3h prediction interval and the LSTM model trained on the imputed data for the 24h prediction interval. As expected, the Se increases for lower thresholds, as more observations are classified as positive. Yet, as more observations are predicted as being positive, the number of false positives will increase as well, hence resulting in lower levels of PPV. For a threshold equal to 0.8, the daily prediction model is able to detect 65% of calvings that will occur in 24 hours with approximately 77% of the positive predictions being correct. When the threshold is lowered to 0.3, almost 90% of all calving events are identified but with a lower PPV being equal to 0.4. For the same threshold, the model predicting a calving event within 1 hour is comparable to the model predicting the moment of calving within 24 hours in terms of the Se. The PPV of the 1h model, however, is 0.13 and, therefore, considerably lower than that of the 24h model.

Prediction window	Non-Imputed					Imputed				
	LR	RF	LSTM	CNN	C-LSTM	LR	RF	LSTM	CNN	C-LSTM
24h	0.32	0.65	0.72	0.75	0.73	0.52	0.76	0.79	0.77	0.78
12h	0.89	0.89	0.89	0.90	0.88	0.86	0.89	0.90	0.90	0.89
6h	0.65	0.68	0.64	0.66	0.66	0.62	0.68	0.64	0.65	0.68
3h	0.41	0.46	0.41	0.47	0.44	0.41	0.46	0.44	0.49	0.44
1h	0.18	0.21	0.21	0.24	0.19	0.19	0.21	0.23	0.24	0.29

Table 4.3: Performance in terms of the AP of the models on imputed and non-imputed test set for the different prediction windows. The imputations on the test data were made by the imputation model

Furthermore, **Table** 4.5 shows the performance of the C-LSTM model for the different imputation strategies in terms of the AP. Regarding the traditional imputation methods, the spline interpolation renders the highest performance for the 24h and 12h prediction interval, with an AP equal to 0.36 and 0.86 respectively. Imputations made by linear interpolation on the contrary, achieve the highest results for the 6h and 1h interval, with AP scores equal to 0.59 and 0.11 respectively. Yet, for every prediction window, using the imputations inferred by the deep learning model clearly results in better performance with respect to predicting the moment of calving than using the imputations made by the more traditional imputation methods. In particular, for the 1h interval, the C-LSTM model trained on the model imputations outperforms the C-LSTM model trained on the imputations made by linear interpolations by 0.18. For the 24h interval, the C-LSTM model leveraging the model imputations even outperforms the best performing model using a traditional imputation method by 0.42. Additionally, the performance of the models trained on the data with missing values as well as the missing values imputed by the imputation model is visualized in more detail in **Fig.** 4.5. More specifically,

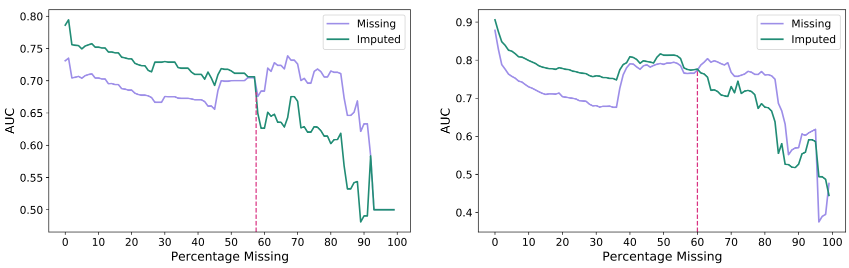
Prediction Window	**Threshold**	**Se**	**PPV**	**Sp**
24h	0.8	0.65	0.77	0.97
	0.5	0.79	0.53	0.90
	0.3	0.87	0.40	0.81
	0.1	0.93	0.28	0.67
12h	0.8	0.57	0.89	0.79
	0.5	0.89	0.81	0.39
	0.3	0.98	0.77	0.15
	0.1	1.0	0.75	0.01
6h	0.8	0.43	0.66	0.85
	0.5	0.77	0.58	0.63
	0.3	0.91	0.52	0.43
	0.1	1.00	0.42	0.09
3h	0.8	0.12	0.67	0.99
	0.5	0.57	0.49	0.85
	0.3	0.80	0.37	0.65
	0.1	0.95	0.26	0.32
1h	0.8	0.30	0.31	0.95
	0.5	0.66	0.16	0.75
	0.3	0.88	0.13	0.55
	0.1	0.99	0.09	0.21

Table 4.4: Performance of the C-LSTM in terms of Se, Sp and PPV for different thresholds

the AUC of the C-LSTM model trained on the missing and imputed data for the smallest and largest prediction interval are compared for different subsets of test observations comprising a minimum percentage of missing values. As expected, the AUC of the models decreased when more missing values were present in the observations. For observations with at least 20% of the sensor values missing, the AUC of both models for the 1h prediction interval decreased by 0.05 compared to the AUC obtained on observations with no missing values. For the 24h prediction interval, the AUC of the model trained with missing data decreased by 0.17 while the AUC of the model trained on the data with imputations only decreased by 0.13. Yet, while the performance of all the models steadily decreased for increasing amounts of missing values, the models trained on the imputed data clearly outperformed the models trained on data with missing values for observations with a tolerable number of missing values. In particular, for sensor sequences with at least 30% of the values missing, the model leveraging the imputations scored an AUC of 0.73 and 0.76 on the 1h and 24h prediction interval respectively. In contrast, the models that didn't have access to the imputations obtained an AUC score of 0.68 for the same subset on both prediction intervals. However, when approximately 60% or more of the sensor values were missing, the AUC of the models trained on the imputed data started to decrease rapidly, resulting in higher performance scores obtained by the models trained on the missing data. This could be explained by the fact that for these observations, the imputations only rely on a small subset of recorded values, hence resulting in less qualitative estimations. Yet, for more reasonable amounts of missing data, imputations are far more precise and therefore the resulting calving predictions as well.

Prediction Window	**Mean Imputation**	**Linear Interpolation**	**Spline Interpolation**	**Model Imputation**
24h	0.16	0.27	0.36	0.78
12h	0.81	0.82	0.86	0.89
6h	0.56	0.59	0.50	0.68
3h	0.31	0.29	0.27	0.44
1h	0.09	0.11	0.10	0.29

Table 4.5: Performance of the C-LSTM model for the different imputation techniques in terms of the AP

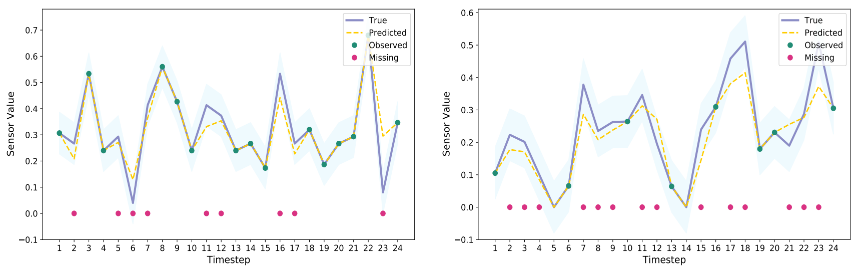


(a) AUC C-LSTM 1h

(b) AUC C-LSTM 24h

Figure 4.5: AUC of the C-LSTM models trained on the data with missing values as well as imputations on subsets of test observations with increasing amounts of missing values. Purple solid line = AUC of model trained on data with missing values, Green solid line = AUC of model trained on data with imputed missing values.

An example of how the imputation model infers the missing values of two different sensor sequences recorded on an hourly basis is visualized in **Fig.** 4.6. For the sensor sequence measuring Inactivity, 10 of the 24 values were randomly set to missing. By observing the remaining Inactivity values as well as the sequences representing the 6 other behavioral activities, the imputation model is capable of accurately approximating the true sensor values. For time step 6 for example, the model correctly infers a strong decrease in Inactivity, before increasing back to a local maximum. For time step 11 and 23, the model also correctly identifies the true direction of the sensor activity, yet slightly underestimates the true increase and decline of the sensor values. For the sequence representing Leg Activity behavior, 14 values were randomly set to missing. Again, the imputation model is able to correctly infer the direction of the sensor values for most of the time steps. From time step 2 to 5, the model rightly predicts a slight increase followed by strong decrease. Likewise, the model is able to detect an increase in sensor values for time steps 7, 11, 18 and 23. For time step 21 however, the model assumes an increase in Leg Activity behavior while a decrease in sensor activity was truly observed.



(a) Inactivity (b) Leg Activity

Figure 4.6: Visualization of the imputation model's predictions of missing sensor values for two random sensor sequences of the test set. Blue solid line = true sensor sequence, Pink dots = sensor values randomly set to missing, Green dots = observed sensor values, Orange dashed line = predicted sensor values

Finally, an example of how the hourly calving models change their predicted probabilities according to the time until calving is shown by **Fig.** 4.7. For one animal, the probabilities are generated by the models by observing the sequence of sliding windows of sensor data before calving. As the moment of calving approaches, the predicted probabilities of the 4 models increase. For the model trained to predict calving within 12 hours, the probabilities become considerably larger than 0.5 when calving starts in 9 hours. The 6h model on the other hand, only starts generating probabilities larger than 0.5 when the moment of parturition is in 6 hours or less. The predictions made by the 3h model start to increase rapidly when calving approaches within 4 hours, while the 1h model only predicts probabilities larger than 0.5 when calving starts within 2 hours.

4.5 Discussion

The results depicted in **Table** 4.3 clearly indicate that the deep learning models, which are able to leverage the sequential patterns in the sensor data, perform better than the more traditional machine learning models, which used the flattened sensor data as input. Except for the 6h prediction interval, the highest AP was always obtained by one of the deep learning algorithms, irrespective of the data

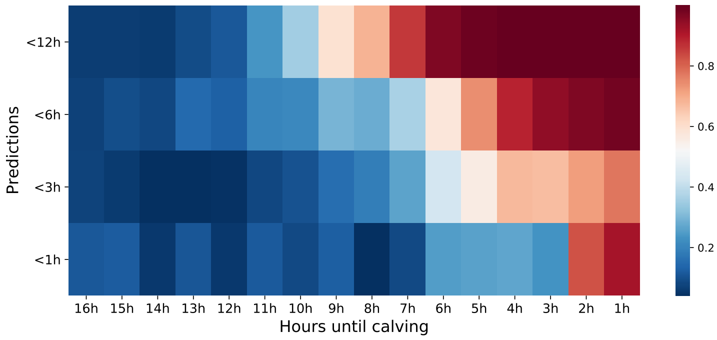


Figure 4.7: The predicted probabilities of the 12h, 6h, 3h and 1h calving models for different time periods until calving

preprocessing method. This indicates that the temporal patterns in the sequences of sensor data contain valuable clues regarding the moment of calving. Traditional machine learning models are not able to leverage sequential information as they do not process the time series in a sequential fashion.

Furthermore, it is also clear from **Table** 4.3 that the models trained on the data imputed by the imputation model predict calving more accurately than when the data with missing values was used. For every prediction window, the best performing model trained on the data imputed by the deep learning model performed as well or better than the best performing model trained on the missing data. For predicting calving within 24 hours and 1 hour, the missing value imputations had the largest impact, with an increase of 0.04 and 0.05 in terms of AP respectively. Additionally, **Table** 4.5 shows how the predictions with respect to the moment of calving were considerably more accurate for every prediction window by using the imputations made by the deep learning model than by using the imputations made by the more traditional imputation methods. Moreover, the results from **Table** 4.3 and **Table** 4.5 indicate that imputation by the mean, linear interpolation or spline interpolation even harm performance, as the C-LSTM model trained on the non-imputed data obtains higher AP scores for every prediction window. This can be attributed to the fact that entire gaps of missing values are more present in the sequences than single missing data points, which in turn may be the result of sensors not transmitting data for a certain period, rather than a single moment. For such large gaps of missing data, the imputations made by the more traditional imputation methods will likely be unrealistic. In particular, imputations generated by the mean and linear interpolation will lie on a horizontal and linear line respectively, while the imputations generated by spline interpolation will lie on a parabolic line. The deep learning imputation model, however, is able to leverage all the information available in the data, including the observed values of other features, and is able to generate more complex patterns for the gaps of missing data. The added value of the imputations was also visualized by **Fig.** 4.5. In particular, it was demonstrated that for observations with reasonable amounts of missing values, the models trained on the imputed data perform consistently better than the models trained on the missing data. For test observations with 1% to approximately 60% of missing values, the 1h as well as the 24h predictions were more accurate when the missing values were imputed. However, for more missing values, the accuracy of the imputations starts to decrease rapidly and hence results in even worse performance than using the raw data as input. These results suggest that as long as no more than half of the data is missing, using intelligent imputation methods can considerably increase the predictive performance to predict the moment of calving.

In order to investigate the added value of using sequential deep learning models for imputation as well as prediction in further detail, the results from this study are compared with the results obtained by similar studies. In the study presented by Rutten et al. (2017), a logistic regression model that used the relative differences in sensor values to predict calving within 1 hour obtained a Se of 0.21 with a PPV of 0.05. For approximately the same level of Precision, the Naive Bayes model presented by Zehner et al. (2019) obtained a much higher Se of 0.82. In this study, the C-LSTM trained on the imputed data was able to detect more positive calving events at a higher precision. More specifically, for a threshold of 0.3, 88% of the true calving events were detected with a PPV of 0.13, while for a threshold of 0.1, the model was able to detect 99% of positive cases with a PPV of 0.09. The logistic regression model was also used by Rutten et al. (2017) to predict the start of calving within 3 hours. For this prediction interval, they reported a Se and a PPV of 0.42 and 0.09 respectively. In this study, a Se of 0.95 with a PPV equal to 0.26 was achieved by the CNN trained on the imputed data, given a threshold of 0.1. A logistic regression model using the relative changes in sensor values was also proposed by Fadul et al. (2017) to predict calving for a 3h interval. For multiparous cows, they reported a Se of 0.85 in correspondence to a Sp of 0.74. In this research, a similar Se of 0.8 was obtained for a slightly lower level of Sp equal to 0.65, given a threshold of 0.3. Yet, the results presented by Fadul et al. (2017) were obtained on the same 9 observations which were used to fit the model parameters and therefore could be biased. For predicting the start of calving within 6 and 12 hours, a logistic regression model was also used by Rutten et al. (2017) and Ouellet et al. (2016). For the 6h predictions interval, Ouellet et al. (2016) reported a Se of 0.71 and a PPV of 0.17, while for the 12h interval, a Se and PPV of 0.7 and 0.3 were obtained. These results, however, should be interpreted with caution as they were also obtained on the same 33 calving events used to train the model. A better comparison can therefore be made with the 6h and 12h models presented by Rutten et al. (2017), as they did use a separate test set to evaluate the predictive performance. For the 6h prediction interval, the model proposed by Rutten et al. (2017) obtained a Se and PPV of 0.49 and 0.11 respectively, while for a window of 12 hours, a Se of 0.51 and a PPV of 0.13 was reported. For a threshold of 0.8, the 6h model proposed in this study achieved a similar level of Se of 0.43, but at a much higher PPV, i.e. 0.66. Likewise, the CNN model trained on the imputed data was much more accurate in predicting the start calving within 12 hours. In particular, 89% of the positive cases could be detected at a predictive accuracy of 81% for a threshold of 0.5. Finally, the model proposed in this study that predicted calving within 24 hours obtained a Se of 0.65 with a PPV of 0.77 given a threshold of 0.8. The model proposed by Rutten et al. (2017) obtained lower values for both the Se as well as PPV, namely a Se of 0.36 and a PPV of 0.6. Borchers et al. (2017) on the other hand, presented an MLP that was able to detect every single positive calving event with a PPV of 0.4. Better performance scores were even reported by Keceli et al. (2020) who used an LSTM architecture on the same dataset. In particular, they reported a Se and PPV of 1.0. However, while the results in this study are obtained on a test set comprising 115 calvings coming from 8 different herds, the results reported by by Borchers et al. (2017) and Keceli et al. (2020) were obtained on only 10 calving events coming from the same herd. Additionally, while in this research the results are obtained on test observations containing missing values, observations with missing values were removed from the analysis conducted by the two aforementioned studies. This is also true for the frameworks proposed by Fadul et al. (2017) and Ouellet et al. (2016). In practice, however, sensor sequences often contain missing values. The prediction errors reported by these studies will therefore be underestimates of the true errors obtained on new observations containing missing values. Finally, the models presented by Borchers et al. (2017) and Keceli et al. (2020) are categorical classification algorithms that predict the number of days until calving during the two weeks preceding calving. In order to predict the number of days until calving, a fixed window of 14 days of observed data was used as feature by Keceli et al. (2020). As a result, the model can correctly predict the number of days until calving by solely counting the number of available features, regardless of the values of these features. For unseen calving events, however, the days until calving are unknown and therefore also the number of features. As a result, it is much more difficult to generalize these results towards other calving events than the results obtained by the present study.

4.6 Conclusion

Dystocia is a major problem for the dairy cattle industry as it significantly affects the animal welfare as well as the farm economics. Accurately predicting the moment of calving is, therefore, a valuable tool for dairy farmers as it allows them to provide timely supervision. In this study, we propose a framework to predict the moment of calving by using sensor data measuring behavioral activities such as eating, ruminating, walking and lying. The present study shows that leveraging the sequential patterns from the sensor data increases the performance of calving prediction models. More specifically, we show how deep learning models are able to accurately infer missing values by using all the behavioral activities observed by the sensors. In addition to increasing the overall predictive performance, using the missing value imputations also significantly improves the performance on observations containing up to 60% of missing values. Additionally, we show how using sequential deep learning algorithms are better able to predict the moment of parturition than more traditional machine learning algorithms, which are not able to exploit the sequential patterns hidden in the sensor data. In particular, the presented models could detect 65% of the calvings within 24 hours with a precision of 77%, while 57% of calvings occurring within 3 hours could be identified with a precision equal to 49%. Hence, the framework proposed in this study can be used to enhance calving predictions, and therefore facilitate timely supervision as well as improve animal welfare.

Acknowledgements

This research was supported by the Dutch research program SenseOfSensors, which is a collaboration between Utrecht University (Utrecht, the Netherlands), Wageningen University & Research (Wageningen, the Netherlands), Nedap Livestock Management (Groenlo, the Netherlands), Vetvice (Bergen op Zoom, the Netherlands) and Elsbeth Stassen (Adaptation Physiology Group, Wageningen University, De Elst 1, 6708 WD Wageningen).

Appendix A. Hyperparameters

Predictive Model	Hyperparameter	Settings
Logistic Regression	Number of iterations	1000, 2000, ..., 5000
	Regularization method	None, L1, L2, Elastic Net
	Regularization strength	0.001, 0.01, 0.1, 0.2, 0.5, 1, 10, 100
	Balancing method	None, Downsample, Upsample, Weighted
Random Forest	Number of trees	100, 200, ..., 1000
	Maximum depth of a tree	10, 20, ..., 100
	Minimum number of samples per split	2, 5, 10
	Minimum number of samples per leave	1, 2, 4
	Maximum features per split	sqrt(number of features)
	Use static features	True, False
	Balancing method	None, Downsample, Upsample, Weighted
LSTM	Number of LSTM layers	1, 2
	Size of hidden state	50, 100, 200
	Activation function	ReLU, Leaky ReLU
	Dropout Rate	0.0, 0.1, ..., 0.5
	Use batch normalization	True, False
	Number of MLP layers	0, 1, 2
	Size of MLP layers	50, 100
	Use static features	True, False
	Balancing method	None, Downsample, Upsample, Weighted
CNN	Number of CNN layers	2, 4, 6, 8
	Number of filters	16, 32, 64, 128
	Size of filter	3
	Downsample layer	None, Stride, MaxPool
	Stride or MaxPool size	2
	Activation function	ReLU, Leaky ReLU
	Dropout Rate	0.0, 0.1, ..., 0.5
	Use batch normalization	True, False
	Number of MLP layers	0, 1, 2
	Size of MLP layers	50, 100
	Use static features	True, False
	Balancing method	None, Downsample, Upsample, Weighted
C-LSTM	Number of CNN layers	1, 2
	Number of filters	16, 32, 64
	Size of filter	3
	Downsample layer	None, Stride, MaxPool
	Stride or MaxPool size	2
	Number of LSTM layers	1, 2
	Size of hidden state	50, 100
	Activation function	ReLU, Leaky ReLU
	Dropout Rate	0.0, 0.1, ..., 0.5
	Use batch normalization	True, False
	Number of MLP layers	0, 1, 2
	Size of MLP layers	50, 100
	Use static features	True, False
	Balancing method	None, Downsample, Upsample, Weighted

Table 4.6: Hyperparameters of the models

References

Alhussein, M., Aurangzeb, K., and Haider, S. I. Hybrid CNN-LSTM model for short-term individual household load forecasting. *IEEE Access*, 8:180544–180557, 2020. doi: 10.1109/ACCESS.2020.3028281.

Barrier, A. and Haskell, M. Calving difficulty in dairy cows has a longer effect on saleable milk yield than on estimated milk production. *Journal of Dairy Science*, 94(4):1804–1812, 2011. ISSN 0022-0302. doi: https://doi.org/10.3168/jds.2010-3641.

Barrier, A., Haskell, M., Birch, S., Bagnall, A., Bell, D., Dickinson, J., Macrae, A., and Dwyer, C. The impact of dystocia on dairy calf health, welfare, performance and survival. *The Veterinary Journal*, 195(1):86–90, 2013. ISSN 1090-0233. doi: https://doi.org/10.1016/j.tvjl.2012.07.031.

Berglund, B., Steinbock, L., and Elvander, M. Causes of stillbirth and time of death in Swedish Holstein calves examined post mortem. *Acta Veterinaria Scandinavica*, 44(3):111, 2003. doi: 10.1186/1751-0147-44-111.

Borchers, M., Chang, Y., Proudfoot, K., Wadsworth, B., Stone, A., and Bewley, J. Machine-learning-based calving prediction from activity, lying, and ruminating behaviors in dairy cattle. *Journal of Dairy Science*, 100(7):5664–5674, 2017. ISSN 0022-0302. doi: https://doi.org/10.3168/jds.2016-11526.

Bradley, A. P. The use of the area under the roc curve in the evaluation of machine learning algorithms. *Pattern Recognition*, 30(7):1145–1159, 1997. ISSN 0031-3203. doi: https://doi.org/10.1016/S0031-3203(96)00142-2.

Fadul, M., Bogdahn, C., Alsaaod, M., Hüsler, J., Starke, A., Steiner, A., and Hirsbrunner, G. Prediction of calving time in dairy cattle. *Animal Reproduction Science*, 187:37–46, 2017. ISSN 0378-4320. doi: https://doi.org/10.1016/j.anireprosci.2017.10.003.

Hochreiter, S. and Schmidhuber, J. Long short-term memory. *Neural computation*, 9(8):1735–1780, 1997.

Hut, P., Hostens, M., Beijaard, M., van Eerdenburg, F., Hulsen, J., Hooijer, G., Stassen, E., and Nielen, M. Associations between body condition score, locomotion score, and sensor-based time budgets of dairy cattle during the dry period and early lactation. *Journal of Dairy Science*, 104(4):4746–4763, 2021. ISSN 0022-0302. doi: https://doi.org/10.3168/jds.2020-19200.

Huzzey, J., von Keyserlingk, M., and Weary, D. Changes in feeding, drinking, and standing behavior of dairy cows during the transition period. *Journal of Dairy*

Science, 88(7):2454–2461, 2005. ISSN 0022-0302. doi: https://doi.org/10.3168/jds.S0022-0302(05)72923-4.

Jensen, M. B. Behaviour around the time of calving in dairy cows. *Applied Animal Behaviour Science*, 139(3):195–202, 2012. ISSN 0168-1591. doi: https://doi.org/10.1016/j.applanim.2012.04.002.

Keceli, A. S., Catal, C., Kaya, A., and Tekinerdogan, B. Development of a recurrent neural networks-based calving prediction model using activity and behavioral data. *Computers and Electronics in Agriculture*, 170:105285, 2020. ISSN 0168-1699. doi: https://doi.org/10.1016/j.compag.2020.105285.

Kingma, D. and Ba, J. Adam: A method for stochastic optimization. *ArXiv e-prints*, 2014.

Krizhevsky, A., Sutskever, I., and Hinton, G. E. Imagenet classification with deep convolutional neural networks. *Advances in neural information processing systems*, 25:1097–1105, 2012.

Lange, K., Fischer-Tenhagen, C., and Heuwieser, W. Predicting stage 2 of calving in Holstein-Friesian heifers. *Journal of Dairy Science*, 100(6):4847–4856, 2017. ISSN 0022-0302. doi: https://doi.org/10.3168/jds.2016-12024.

Laven, R., Huxley, J., Whay, H., and Stafford, K. Results of a survey of attitudes of dairy veterinarians in New Zealand regarding painful procedures and conditions in cattle. *New Zealand Veterinary Journal*, 57(4):215–220, 2009. doi: 10.1080/00480169.2009.36904.

LeCun, Y., Bottou, L., Bengio, Y., and Haffner, P. Gradient-based learning applied to document recognition. *Proceedings of the IEEE*, 86(11):2278–2324, 1998.

Leys, C., Ley, C., Klein, O., Bernard, P., and Licata, L. Detecting outliers: Do not use standard deviation around the mean, use absolute deviation around the median. *Journal of Experimental Social Psychology*, 49(4):764–766, 2013. ISSN 0022-1031. doi: https://doi.org/10.1016/j.jesp.2013.03.013.

Liseune, A., Salamone, M., Van den Poel, D., Van Ranst, B., and Hostens, M. Leveraging latent representations for milk yield prediction and interpolation using deep learning. *Computers and Electronics in Agriculture*, 175:105600, 2020. ISSN 0168-1699.

Lombard, J., Garry, F., Tomlinson, S., and Garber, L. Impacts of dystocia on health and survival of dairy calves. *Journal of Dairy Science*, 90(4):1751–1760, 2007. ISSN 0022-0302. doi: https://doi.org/10.3168/jds.2006-295.

Mahnani, A., Sadeghi-Sefidmazgi, A., and Keshavarzi, H. Performance and financial consequences of stillbirth in holstein dairy cattle. *Animal*, 12(3):617–623, 2018. ISSN 1751-7311. doi: https://doi.org/10.1017/S1751731117002026. URL https://www.sciencedirect.com/science/article/pii/S1751731117002026.

McGuirk, B. J., Forsyth, R., and Dobson, H. Economic cost of difficult calvings in the United Kingdom dairy herd. *Veterinary Record*, 161(20):685–687, 2007. doi: https://doi.org/10.1136/vr.161.20.685.

Mee, J. F. Managing the dairy cow at calving time. *Veterinary Clinics of North America: Food Animal Practice*, 20(3):521–546, 2004. ISSN 0749-0720. doi: https://doi.org/10.1016/j.cvfa.2004.06.001.

Mee, J. F. Why do so many calves die on modern dairy farms and what can we do about calf welfare in the future? *Animals*, 3(4):1036–1057, 2013. ISSN 2076-2615. doi: 10.3390/ani3041036. URL https://www.mdpi.com/2076-2615/3/4/1036.

Meyer, C., Berger, P., and Koehler, K. Interactions among factors affecting stillbirths in Holstein cattle in the United states. *Journal of Dairy Science*, 83(11):2657–2663, 2000. ISSN 0022-0302. doi: https://doi.org/10.3168/jds.S0022-0302(00)75159-9.

Miedema, H. M., Cockram, M. S., Dwyer, C. M., and Macrae, A. I. Behavioural predictors of the start of normal and dystocic calving in dairy cows and heifers. *Applied Animal Behaviour Science*, 132(1):14–19, 2011. ISSN 0168-1591. doi: https://doi.org/10.1016/j.applanim.2011.03.003.

Ouellet, V., Vasseur, E., Heuwieser, W., Burfeind, O., Maldague, X., and Charbonneau, É. Evaluation of calving indicators measured by automated monitoring devices to predict the onset of calving in Holstein dairy cows. *Journal of Dairy Science*, 99(2):1539–1548, 2016. ISSN 0022-0302. doi: https://doi.org/10.3168/jds.2015-10057.

Pak, U., Kim, C., Ryu, U., Sok, K., and Pak, S. A hybrid model based on convolutional neural networks and long short-term memory for ozone concentration prediction. *Air Quality, Atmosphere & Health*, 11(8):883–895, 2018.

Raussi, S. Human–cattle interactions in group housing. *Applied Animal Behaviour Science*, 80(3):245–262, 2003. ISSN 0168-1591. doi: https://doi.org/10.1016/S0168-1591(02)00213-7. Behavior and welfare of cattle housed in large groups.

Rumelhart, D. E., Hinton, G. E., and Williams, R. J. Learning representations by back-propagating errors. *Nature*, 323(6088):533–536, 1986.

Rutten, C., Kamphuis, C., Hogeveen, H., Huijps, K., Nielen, M., and Steeneveld, W. Sensor data on cow activity, rumination, and ear temperature improve prediction of the start of calving in dairy cows. *Computers and Electronics in Agriculture*, 132:108–118, 2017. ISSN 0168-1699. doi: https://doi.org/10.1016/j.compag.2016.11.009.

Saito, T. and Rehmsmeier, M. The precision-recall plot is more informative than the ROC plot when evaluating binary classifiers on imbalanced datasets. *PloS one*, 10(3):e0118432, 2015.

Schirmann, K., Chapinal, N., Weary, D., Vickers, L., and von Keyserlingk, M. Short communication: Rumination and feeding behavior before and after calving in dairy cows. *Journal of Dairy Science*, 96(11):7088–7092, 2013. ISSN 0022-0302. doi: https://doi.org/10.3168/jds.2013-7023.

Schuenemann, G., Nieto, I., Bas, S., Galvão, K., and Workman, J. Assessment of calving progress and reference times for obstetric intervention during dystocia in Holstein dairy cows. *Journal of Dairy Science*, 94(11):5494–5501, 2011. ISSN 0022-0302. doi: https://doi.org/10.3168/jds.2011-4436.

Schuenemann, G., Bas, S., Gordon, E., and Workman, J. Dairy calving management: Description and assessment of a training program for dairy personnel. *Journal of Dairy Science*, 96(4):2671–2680, 2013. ISSN 0022-0302. doi: https://doi.org/10.3168/jds.2012-5976.

Szegedy, C., Liu, W., Jia, Y., Sermanet, P., Reed, S., Anguelov, D., Erhan, D., Vanhoucke, V., and Rabinovich, A. Going deeper with convolutions. In *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 1–9, 2015.

Szenci, O., Nagy, K., Takács, L., Mádl, I., Bajcsy, Á. C., et al. Farm personnel management as a risk factor for stillbirth in hungarian Holstein-Friesian dairy farms. *Magyar Állatorvosok Lapja*, 134(7):387–393, 2012.

Tenhagen, B.-A., Helmbold, A., and Heuwieser, W. Effect of various degrees of dystocia in dairy cattle on calf viability, milk production, fertility and culling. *Journal of Veterinary Medicine Series A*, 54(2):98–102, 2007. doi: https://doi.org/10.1111/j.1439-0442.2007.00850.x.

Van Pelt, M. and de Jong, G. Genetic evaluation for direct and maternal livability in The Netherlands. *Interbull Bulletin*, (44), 2011.

Zehner, N., Niederhauser, J. J., Schick, M., and Umstatter, C. Development and validation of a predictive model for calving time based on sensor measurements of ingestive behavior in dairy cows. *Computers and Electronics in Agriculture*, 161:62–71, 2019. ISSN 0168-1699. doi: https://doi.org/10.1016/j.compag.2018.08.037. BigData and DSS in Agriculture.

Zhao, B., Lu, H., Chen, S., Liu, J., and Wu, D. Convolutional neural networks for time series classification. *Journal of Systems Engineering and Electronics*, 28(1):162–169, Feb 2017. ISSN 1004-4132.

5 Discussion and conclusion

5.1 Discussion

This dissertation discussed several techniques to improve current animal monitoring systems in dairy cattle and make them suitable for practical applications. The notion of a system that can effectively be used for practical applications is often an underexposed topic in literature. According to Hogeveen et al. (2010), three criteria must be fulfilled for a prediction model to be applied in practice for commercial dairy farming: 1) a high performance, 2) a relevant prediction time window and 3) a high degree of similarity between the research data and the real everyday data in commercial farms. Since our goal was to develop monitoring systems suitable for practical implementations, we discuss the frameworks proposed in this dissertation with regard to every of these three criteria.

In **Chapter 2**, we present an autoencoder that infers all missing milk yields along the lactation curve. To do this, the encoder compresses all available information in a latent representation that comprises the most important traits of the lactation curve. The decoder then uses this latent representation to generate the entire milk yield curve. This approach allows the model to leverage all available information, regardless of the number and recording time of the observations. Results show that the model presented in this chapter is capable of achieving state-of-the art performance with respect to predicting individual as well as 305d milk yields for different windows of available data. Moreover, by deliberately training the model on data with varying sampling properties, the model's Mean Absolute Percentage

Error (MAPE) increased by a maximum of 2 percentage points when 60% of the observed milk yields were set to missing for every possible forecast horizon. This is in contrast to previously developed lactation models whose performance quickly deteriorates when observations are lacking (Silvestre et al., 2006). All of the above properties make it clear that the proposed approach fulfills the three predefined criteria. First, the prediction window is relevant as the model is capable to infer any missing milk yield along the lactation curve. Moreover, the model dynamically changes its predictions when new information is observed. This enables farmers to predict any milk yield at any given point in the lactation cycle. Second, there is a high similarity between the research design and the reality as the model was explicitly trained to perform well under varying conditions of data completeness. Third, results show that we have developed a robust lactation model that can compete with the state-of-the-art in terms of individual as well as 305d milk yield prediction.

In **Chapter 3**, we propose a novel methodology to predict the milk yield curve of a lactation cycle based on all the observed information from the preceding cycle. In particular, we first extract the latent representation of the observed milk yield curve using the encoder from Chapter 2. Next, we predict the low-dimensional latent encoding from the lactation curve in the subsequent cycle. Finally, we use the decoder from Chapter 2 to transform the latent encoding into the corresponding curve. This approach has some interesting properties. First, it allows to forecast a cow's entire milk yield curve before the cow even has calved. This in turn allows dairy farmers to monitor unexpected milk losses in early lactation, increase the forecast horizon with respect to production and costs and facilitate timely culling and breeding decisions. Second, the curve can be generated regardless of the number of milk yields observed in the preceding cycle, as the encoder from Chapter 2 was explicitly trained to extract a curve's most important traits by leveraging all available information. Hence, as long as the latent representation is of high enough quality, the model generates accurate predictions. As a result, these two properties make that the proposed methodology meet the last two requirements, i.e. relevant prediction window and correspondence between research design and reality. Furthermore, it was shown that the model presented in this chapter obtained a smaller prediction error for every milk yield in the predicted cycle compared to the group's average lactation curve. Moreover, it was shown that for the first 26 days of lactation, the predictions made by this model using data from the preceding cycle are more accurate than those of a lactation model that uses data observed in the predicted cycle. Therefore, we can conclude that with regards to the early lactation, we have developed a model that, in addition to the last two criteria (i.e. a relevant prediction window and a high similarity between research design and reality) also meets the first criterion (i.e. a high performance).

In **Chapter 4**, we specifically designed a calving prediction framework to fulfill the last two criteria (i.e. a relevant prediction window and a high similarity between research design and reality), as such frameworks have not been covered adequately in previous literature. To meet the last criterion, we developed a deep learning methodology to infer missing values in the sensor recordings of behavioral activities by leveraging the observed sensor data. This approach ensured that the model could still generate reliable calving alerts when facing several data quality issues (e.g. defective sensors and faulty transmission of data). In fact, results showed that this strategy outperformed more traditional imputation methods and significantly improved calving detection for cows with sensor data containing up to 60% of missing values. To fulfill the second criterion, we analyzed the ability of deep learning algorithms to accurately predict the moment of calving within different time frames. In particular, we investigated whether 14 days of sensor data could be used to predict calving within 24h and whether 24 hours of sensor data could be used to predict calving within 12h, 6h, 3h and 1h. This way, dairy farmers can continuously obtain the likelihood a cow is about to calve within several time frames. Moreover, we validated the performance of each of these models on a previously unseen large sample of cows coming from different herds in terms of Average Precision (AP) to analyze the framework's generalizability. A random classifier has an AP equal to the proportion of positive examples while a perfect model has an AP equal to 1. For each of the prediction intervals, the models obtained an AP considerably higher than the average proportion of positives. In particular, for the 24h, 12h, 6h, 3h and 1h prediction intervals, the AP scores were 0.79, 0.90, 0.68, 0.49 and 0.29 respectively, while the proportion of positives equalled 0.125, 0.26, 0.40, 0.20 and 0.07 respectively. Moreover, we found that for each of the predictive time frames, our framework outperformed the only model previously developed that could largely satisfy the three predefined criteria (Rutten et al., 2017). Therefore, we believe that, especially for the larger prediction windows, we have developed an accurate calving prediction model that can compete with current state-of-the-art models and hence fulfills all three criteria.

5.1.1 Additional recommendations

Given the fact that the presented prediction models largely satisfy each of the three predefined criteria, i.e. a high performance, a relevant prediction time window and a high degree of similarity between the research data and the real everyday data in commercial farms, we believe that this dissertation provides useful insights on how dairy farmers can use their data to develop state-of-the-art monitoring systems that are applicable in practice. Before actually implementing the frameworks, however, dairy farmers should satisfy a minimal amount of requirements, which are visualized by **Figure** 5.1. In particular, dairy farmers should record a minimum amount of daily milk yields as well as health and reproduction events of their cows to implement the lactation models from Chapter 2 and 3. Most modern dairy farms are equipped with milkmeters that automatically register milk yields with each milking. Since the SAE from Chapter 2 was explicitly trained on data with variable recording intervals, monthly test-day records can also be used, albeit with a potential loss in prediction accuracy. The option to register health and reproduction related events is included as standard in most HMS. Farms with no HMS can register these events manually on their computers or mobile devices. The herd KPIs, animal KPIs and average milk yield per parity that were used for the lactation models presented in Chapter 2 and 3 were automatically calculated by the cloud-based dairy analysis application (www.mmmooogle.com). These statistics can also be derived directly from the produced milk yields and recorded health and reproduction events. In addition, mmmooogle cleans the collected HMS data by removing duplicate milk yields and correcting unrealistic event sequences. For example, a calving event that occurs between a pregnancy and dry-off event is removed from the event sequence. These cleaning procedures can also be done internally without relying on third parties. For predicting the moment of calving, behavioral sensor data were used. Therefore, another minimal requirement is that dairy farms should equip their cows with devices measuring the same behavioral activities before the moment of calving. In addition to the Nedap neck and leg accelerometers, many other devices exist that measure the behavioral activities that were used in this study. For example, pressure sensing devices, microphones and boluses have been used to measure eating and ruminating behavior (Crociati et al., 2022). From the moment the dairy farms are equipped with the necessary PLF technologies to record the required data, the models can be deployed continuously while cows transition through their lactation cycles, as is demonstrated by **Figure** 1.1. As the presented models were completely developed and trained in TensorFlow, they can be deployed by TensorFlow Serving (Olston et al., 2017). This is a flexible, language-neutral and high-performance serving system for machine learning models, designed for production environments.

When a cow has been dried off and has been equipped with sensors to measure its behavioral activities, the calving models can be applied continuously on incoming batches of sensor data. On their mobile devices, dairy farmers can get a heatmap for each cow that is about to calve similar to the heatmap presented in **Figure** 4.7. These heatmaps visualize the predicted probabilities of each calving model at each time step and indicate how likely it is that the moment of calving will occur within each predictive window. Automatic alerts can be generated when the predicted probability of one of the calving models exceeds a specific threshold. This can help farmers to decide which cows to inspect for possible delivery assistance. Typically, automated detection models are evaluated in terms of Sensitivity, Specificity and Precision. Concretely, for a calving prediction model, a high Sensitivity means that the model is capable of generating alerts for most of the true calving events. A high Specificity means that the model is capable of not generating alerts for most of the

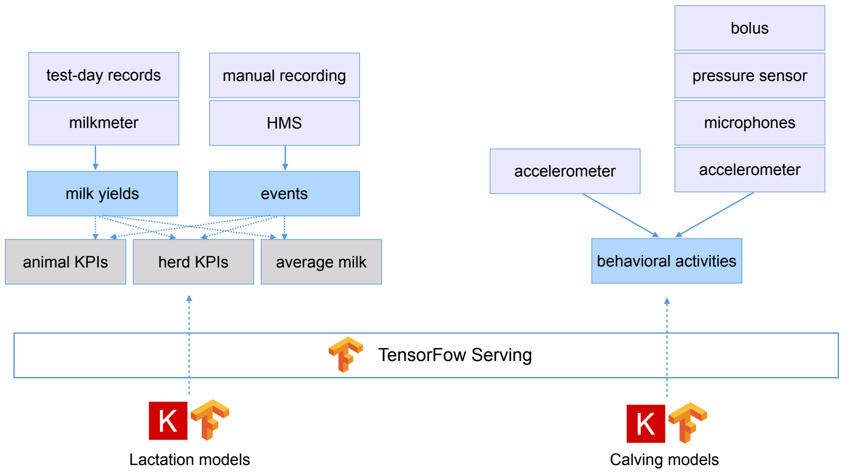


Figure 5.1: Overview of minimal requirements to implement presented frameworks. Purple = required PLF technologies, blue = required data

false calving events. Similarly, a high Precision means that little alerts that were generated related to false calvings events. Therefore, the Specificity and Precision provide more information on the reliability of a detection system and are extremely important for the farmer's trust in the prediction model. For example, when the Sensitivity of the calving model is 90%, but the Precision is 50%, this means that almost all calvings are detected successfully, but that 50% of the alerts are in fact erroneous. Inspecting cows that are not about to calve can be a large waste of time and hence be very costly. Hence, when the system generates too much false alerts, the farmer will lose its trust in the system and stop using it. According to the ISO 20966:2007 guideline, the Sensitivity and Specificity of automated detection systems should be as high as 80% and 99%. In practice, however, these performance scores are seldom achieved (Rutten et al., 2013). These performance scores were neither obtained by the calving models presented in this dissertation. In particular, for a Sensitivity of approximately 80%, the 24h model, 6h model and 3h model achieved a Specificity of 90%, 63% and 65% respectively. Nevertheless, as we provided calving models operating on different predictive time windows, farmers can choose to apply a customized rule system for each calving model, such that the framework still facilitates calving management. For example, farmers can decide to apply high thresholds for the larger prediction windows, while applying lower thresholds for the smaller prediction windows. This strategy will only generate alerts when cows are very likely to calve in the larger windows, while for the smaller windows alerts will be generated more quickly to avoid False Negatives. Also, for the cows that have been identified as being very likely to calve in the larger prediction windows, the heatmap presented in **Figure** 4.7 can be used to monitor the probabilities of the models of the smaller prediction windows. An example of how this strategy can be used in practice is given by **Figure** 5.2. Hogeveen et al. (2017) found that using a

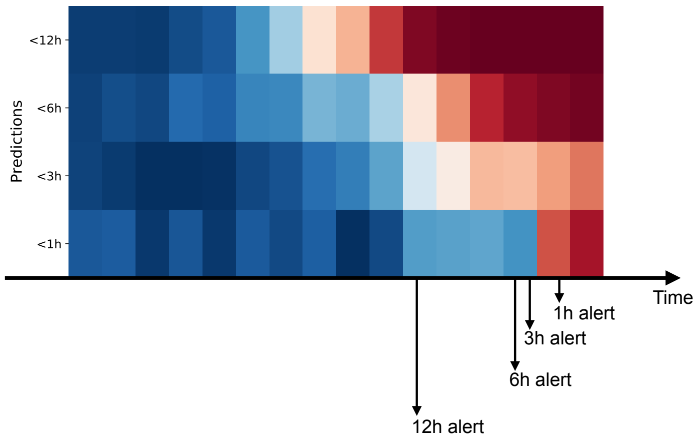


Figure 5.2: Example of calving models in practice. At every time step, each calving model generates predicted probabilities given the observed behavioral data. Blue = low predicted probability, red = high predicted probability

calving alarm system based on accelerometer data presented by Rutten et al. (2017) yielded an average net return of €5.71 per cow and per calving for the 1h model. Crociati et al. (2020) on the other hand, found that using a calving alert system based on an intravaginal device measuring light and temperature yields an average net return of €63.5 per cow per calving. The net returns of the two aforementioned calving systems were estimated by taking into account the costs and revenues that relate to the implementation and usage of a calving alert systems. The revenues corresponded to a decreased rate of stillbirths, a decreased number of days open and an increased amount of milk yield. The cost mainly included the installment costs as well as the cost of the central unit and sensor devices. The sensors used in this dissertation were accelerometers measuring behavioral activities and were very similar to the sensors used in Rutten et al. (2017). In Section 4.5 of Chapter 4, we showed that the calving models presented in this dissertation outperformed the model developed by Rutten et al. (2017) in terms of Sensitivity and Precision for every prediction window. The calving alert system used by Crociati et al. (2020) is based on an intravaginal sensor measuring light and temperature and generates an alert when it is expelled from the vagina at the beginning of stage 2 of labor. The cost of this system was estimated at €10 per calving and is slightly lower than the cost of the sensor system used in this dissertation, which is approximately €50 per cow and hence ranges between €12,5 and €16,67 per calving, assuming that a cow gives birth to 3 to 4 calves during its lifespan. The calving system evaluated in Crociati et al. (2020) achieved a Sensitivity of 86%, yet no information on the Specificity or Precision of the alarm system was provided. In case that the calving alarm system obtains similar levels of Precision and Specificity for a Sensitivity of 86% as the calving model presented in this dissertation, we can expect that using the calving models presented in this dissertation will yield a net revenue that ranges between €5.71 and €63.5. Future research should determine the net value of this system by performing a cost-profit simulation similar to that conducted by Crociati et al. (2020); Hogeveen et al. (2017)

Once a cow has delivered its calf and begins its next lactation cycle, the SLMYP can be deployed to forecast the entire milk yield curve that the cow is expected to produce. From 26 day onwards in the lactation cycle, the daily milk yield predictions are generated by the SAE, as **Figure** 3.6 showed that after 26 days, the SAE starts to generate more accurate predictions than the SLMYP. From that moment, the expected lactation curve is continuously updated as soon as new information on milk yields or health and reproduction events becomes available. Early in the lactation cycle, the expected lactation curve can be used to support breeding, culling and feeding decisions. Additionally, the milk yield predictions generated by the SLMYP and the SAE can be compared with the true yields on a daily basis to support animal monitoring. The daily milk losses can be displayed on mobile devices, and farmers can get alerts in case of considerable milk losses as these can indicate health-related problems. This is demonstrated by **Figure** 5.3. From the moment the farmer receives an alert, the cow can be submitted for inspection. As with the calving models, this detection system should also be evaluated in terms of Sensitivity and Precision. In particular, how many cows with health-related problems are detected by the monitoring system, and how many alerts were generated for cows with no problems. In addition to overestimates of the expected milk yields predicted by the lactation models, low Precision scores may also be caused by other factors. First, the alert can be generated too quickly if a too low boundary is applied for the milk loss that triggers the alarm. A higher boundary for the milk loss or longer periods with consecutive observed milk losses should then be considered to trigger the alerts. Second, due to varying milking intervals, daily milk yields can fluctuate much, which may trigger false alerts. In that case, it is advisable to standardize the daily milk yields with respect to the milking intervals (Adriaens et al., 2018). The net value of the lactation models can be estimated by using a similar procedure as used in Crociati et al. (2020); Hogeveen et al. (2017). This includes the estimation of the costs and revenues that correspond to using the lactation models. In contrast to the calving models, using the lactation models does not require additional investments in specialized equipment. Milk yields are by default recorded in almost every dairy farm, either by

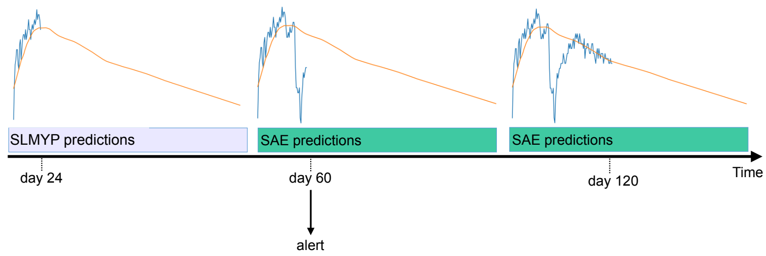


Figure 5.3: Example of lactation models in practice. Blue line = true milk yield, orange line = predicted milk yield

milkmeters or by test-day records. The health and reproduction events are recorded manually on a computer. Hence, as the installment costs are negligible, the total cost related to using the lactation models mainly exist out of the amount of time the farmer spend on investigating false detected cows, which is estimated to be €22.05 per hour for Belgian dairy farmers (emb, 2017). The revenues of using the lactation model mainly relate to the decreased rate of disease due to timely detection and inspection. This results in increased milk production and decreased veterinary costs. Other revenues of using the lactation model correspond to more selective breeding and culling and should also be taken into account. Hence, only when we can foresee the actual revenues generated by the lactation models, we will be able to look at the cost-efficiency of using the lactation models in practice. This will require empirical experiments in which the revenues of the lactation model generated for an experimental group is compared with a control group which were not submitted to the detection system.

In the short term, the models should be retrained on a regular basis to safeguard their performance. Due to TensorFlow Serving's version control management, TensorFlow models that are being deployed in production can be retrained continuously. The training of the lactation models can be continued when completed lactation cycles are being observed, while incremental training of the calving models can be conducted on new windows of observed sensor data. Furthermore, the models should be retrained regularly to cope with potential data drift. For example, it has been showed that milk production has been steadily increasing over the last years due to selective breeding, increased milking frequency, improved monitoring, better feeding and improved reproductive performance (Bórawski et al., 2020). In the long term, the models can be adapted when better processing techniques, new data and novel algorithms become available. In Section 5.3, we discuss some of these future avenues for research.

5.2 Conclusion and implications

In this dissertation, we set out to enhance animal monitoring systems in the dairy industry and make them more suitable for practical applications. Our goal was to provide evidence that advanced deep learning algorithms could improve the predictive performance of current monitoring systems and could cope with the real everyday data conditions in commercial farms. To do so, we developed and assessed three different monitoring systems based on data coming from various commercial dairy farms: a framework to predict the milk yield in the current lactation cycle, a framework to predict the moment of calving after the current lactation cycle and a framework to predict the milk yield in the subsequent lactation cycle. Combining these frameworks allows dairy farmers to better manage the welfare of their animals during the transition period, i.e. the period between late pregnancy and early lactation. This period is a critical moment for dairy cows, as most health disorders occur during this time (Drackley, 1999). Here, we summarize each chapter according to their methodological and theoretical contributions.

Chapter 2 focuses on inferring milk yields in a certain lactation cycle. This study evaluates whether a latent representation of all information observed in the lactation cycle can be used to accurately generate the entire milk yield curve. At the theoretical sight, our findings extend the current theories on the effect of mastitis (Fernandes et al., 2021), disease (Bareille et al., 2003), parity (Macciotta et al., 2011) and herd management (Jeretina et al., 2015) on the milk production. Furthermore, we find evidence that most important aspects of a lactation curve such as peak time, peak yield and persistence can be used to accurately summarize the entire lactation curve, which is in correspondence with the existing theory on lactation curve modeling (Bouallegue and M'Hamdi, 2020). From a methodological perspective, we contribute to the literature by presenting a deep learning model that can automatically derive the lactation curve parameters by using all available information in a lactation cycle. This approach allows missing milk yields to be inferred along the entire lactation curve, regardless of the number of observations and the recording interval between the different observations. Furthermore, the presented methodology dynamically updates its predictions when new observations are observed. This is in contrast with previous individual curve fitting models that generate fixed predictions once the required set of data points are observed. In addition, this study is the first to show how the sequence of the health and reproduction events can be included to improve milk yield predictions. This approach makes it possible to update the milk yield predictions once a critical event such as mastitis is observed. Our findings show that the presented framework achieves state-of-the-art results in terms of individual and 305d milk yield predictions.

Chapter 3 investigates whether a lactation curve can be entirely predicted by using information observed in the preceding cycle. On the methodological side, this study is the first to conduct such an analysis in terms of algorithms, variables as well as predictions. Our methodological insights are that an entire lactation curve can be generated non-sequentially by predicting and decoding its parameters using deep learning algorithms and that the lactation curve is estimated by a higher accuracy when parity, herd, health and reproduction information is used in addition to the milk yield observed in the preceding cycle. Moreover, we show that by using this methodology, the milk yield before the 26th day of lactation can be predicted more accurately than by using current lactation models. Therefore, this methodology contributes to the literature by enhancing animal monitoring in early lactation, which is a critical stage in the cow's life (Caixeta and Omontese, 2021). On the theoretical side, we find evidence that a cow's historical milk production is related to its future production, which is in line with results of previous studies (Ali and Schaeffer, 1987). Furthermore, we find that health events such as mastitis and disease are correlated with future milk production.

Chapter 4 assesses a deep learning model that predicts the moment of calving. The goal in this study is to determine whether sensor data on behavioral activities such as eating, ruminating, lying, walking and standing time can be used to predict the moment of calving within 24h, 12h, 6h, 3h and 1h. From a methodological perspective, this study contributes to the literature by using a deep learning framework to impute the missing values in the sensor data. The study finds that missing observations of a certain behavioral activity can be accurately imputed by leveraging the observed values of all the behavioral activities. Moreover, the study shows that this approach yields better predictions than more traditional imputation schemes such as mean imputation, linear and spline interpolation. Additionally, we find that that deep learning architectures specifically designed to process sequential data outperform machine learning models used in previous studies. From a theoretical side, our findings extend current research on calving prediction (Borchers et al., 2017; Fadul et al., 2017; Rutten et al., 2017; Zehner et al., 2019) in that behavioral changes around the time of calving may provide clues to detect when cows are about to calve.

5.3 Limitations and future research

5.3.1 Data limitations

A first limitation concerning the data used in this dissertation is that some input data were manually recorded by the dairy farmers and therefore are prone to human error. In Chapter 2 and 3 the health and reproduction events such as mastitis, insemination, pregnancy and disease were entered manually by the farmers on wearable devices such as tablets or smartphones. Hence, we are aware that it is very likely that some events that took place were not recorded and that some events that did not took place were recorded, though we estimate the probability of the latter case low since recording an event requires a manual action on the device. An interesting avenue for future research could, therefore, be to automatically register these events. Several authors have, for example, suggested many automated techniques to detect mastitis by using automated milking systems data (Khatun et al., 2018), infrared thermography data (Machado et al., 2021) or sensor data (Post et al., 2020). Another limitation in this regard is that the moment of calving was manually registered by the farmers in Chapter 4 and corresponded to the timestamp of the completed birth of the calve. The data, however, did not include which calvings were assisted. Therefore, it is likely that the timestamp of some of the calvings are underestimates of the true expected timestamps as delivery assistance speeds up the process of parturition. Together with the limited data on calvings with an exact timestamp, this could explain the lower performance of the models predicting calving within the shorter time frames. Furthermore, we were limited to predict the moment of parturition visually observed by the farmer and could not discriminate between the different stages of calving. This could be valuable as larger farms would benefit more from receiving alerts during the prodromal stage of labor (stage 1), while receiving alerts at the beginning of labor (stage 2) would be preferred by smaller farms (Crociati et al., 2022). Again, an interesting solution here could be to rely on automated approaches to register the different stages of parturition such as video cameras (Fadul et al., 2017) or pressure sensing devices (Scheurwater et al., 2021).

A second limitation is that we did not use the complete set of features that relate to milk production and calving. In Chapter 2 and 3, we exclusively used information on milk production, herd statistics, parity and health and reproduction events to predict milk yield. With regard to the health and reproduction events, we were limited by the set of events that were provided by the analytical cloud platform to the dairy farmers. Yet, other health related events such as lameness and infectious diseases also have an impact on the future milk production (Green et al., 2002; Statham et al., 2015). Moreover, it is known that environmental factors such as weather, climate and geography, feed intake and management practices also affect milk production (Collier et al., 2017; Rekik and Gara, 2004). For Chapter 4 on the other hand, we focused on using behavioral data measured by pedometers and accelerometers to predict the moment of calving. Yet, many alternative devices exist that measure other parameters that correlate with calving. For example, research found that vaginal temperature (Ouellet et al., 2016), tail base temperature (Cooper-Prado et al., 2011) and tail movements (Giaretta et al., 2021) tend to change around the moment of calving. Therefore, we suggest future studies to investigate the impact of the inclusion of some of these proposed features on the results reported in this dissertation.

A third limitation is that milking intervals were not taken into account in Chapter 2 and 3. The daily milk yield was equal to the accumulation of the yields of all the milking sessions during that day. Variable milking intervals, however, can introduce some variability in the data. Therefore, future research could normalize the daily milk yields according the milking intervals in order to reduce the variability of the daily yields. This approach has already been applied in previous research (Adriaens et al., 2018).

5.3.2 Algorithmic limitations

A first limitation related to the algorithmic choices made in this dissertation is that we did not asses the entire range of available deep learning architectures for our frameworks. In Chapter 2 and 3, we opted for Convolutional Neural Networks (CNN) to process the milk yield sequences as research has shown that these types of architectures can be of great value for time series analysis (Zhao et al., 2017). Moreover, in contrast to Long Short-Term Memory (LSTM) models, CNNs process the data in parallel, which can significantly decrease training and testing time. This model architecture was therefore also chosen in Chapter 4 to impute the missing values in the sensor data. However, as stated by the no free lunch theorem, there exists no single best algorithm, as the error averaged over all tasks is the same for any solution method (Wolpert, 1996). Hence, we suggest future research to benchmark different deep learning architectures such as CNNs, LSTMs, GRUs, CNN-LSTMS and Bi-LSTMS with respect to their performance on the different tasks tackled in this dissertation. Moreover, future research could asses the performance of transformers, which were originally developed for Natural Language Processing (NLP) tasks, but have recently also been used for time series forecasting and classification problems (Liu et al., 2021; Zerveas et al., 2021).

A second limitation is that we did not assess all the possible hyperparameter configurations for the predictive models. In Chapter 2, 3 and 4, we decided to fix the batch size and the optimization algorithm for the training process of the models. We also limited the possible number of layers and neurons per layer for all the models. Moreover, we did not explore the entire range of possible values for parameters such as kernel size, dropout, leaky ReLu and learning rate. In addition, we did not implement grid search as this would waste enormous amounts of computation (Goodfellow et al., 2016). Therefore, we opted for a Bayesian optimization strategy in Chapter 2 and 3, while a random search was applied in Chapter 4. It would be interesting to see whether the performance of the models would further increase by extending the hyperparameter tuning process.

5.3.3 Analytical limitations

A first limitation of the presented studies is that we did not use statistical tests to compare different models. We mainly drew conclusions on feature and model selection based on the predictive performance of the features and models on the test set. Performing a significance test requires the distribution of each of the models' performance score. One common way to obtain this distribution is by using cross-validation, as in this procedure a model is evaluated on multiple test samples. Training deep learning models, however, is computationally expensive. Hence, using cross-validation for deep learning models often becomes practically infeasible, as the computation time exponentially grows with the number of folds (Li et al., 2020). Therefore, a simpler strategy that involves one training, validation and test set is often preferred when training deep learning models. This validation procedure, however, only results in one performance score on the test set and, therefore, does not allow to draw statistical conclusions. A perspective for future research is to evaluate the trained models on different bootstraps drawn from the test set to obtain a distribution of the evaluation metric. The downside of this approach, however, is the potential presence of duplicates in and across the different test samples. Another approach would be to estimate the mean and standard error of the prediction errors, though this approach requires that the errors are normally distributed (Goodfellow et al., 2016). Another interesting avenue would be to implement Bayesian deep learning models, as this approach defines a posterior distribution on the model's parameters, which in turn results in a predictive distribution. These models, however, are significantly more complex than standard neural networks which makes them hard to implement in practice. Nevertheless, by using one of these approaches, future research could extend the findings of this dissertation by statistically determining which modeling techniques are most suited for milk and calving prediction. Moreover, by using significance tests, future research could establish which features significantly contribute to the model's performance.

A second limitation is that we didn't assess the direction of the relationship between the independent features and the labels. Deep learning models are black box models, providing almost no information on how the features affect the predictions. Interpretable modeling techniques, such as Shapley values and Partial Dependence Plots allow to infer the marginal effect of a feature on the outcome. However, we didn't implement Shapley values or Partial Dependence Plots in this dissertation. Again, this limitation is mainly due to the computational costs associated with these techniques. Inferring the marginal effect of a feature on the outcome generally requires the calculation of the predictions for all possible feature combinations. Therefore, the computation time increases exponentially with the number of features (Jia et al., 2019). Moreover, as deep learning models often comprise more than millions of parameters, generating predictions for large datasets also requires some computational effort. Also, at the time of writing this dissertation, no practical implementations of statistical packages to calculate these Shapley values or Partial Dependence Plots were available for deep learning models trained in Tensorflow. Therefore, in Chapter 2 and 3, we opted for calculating the

Variable Importance scores. This approach is a lot less computationally expensive, as calculating these scores does not require the need to compare the contribution of each feature to the entire feature space. The downside of this technique is that it only reports the ranking of each feature based on their contributions to the predictive accuracy (Livingston, 2005). As a result, in Chapter 2 and 3, we were limited by reporting which features were most important with regard to milk prediction. To improve the interpretability of the prediction models in Chapter 2 and 3, future research could include Shapley values for the input features. This could provide valuable information to farmers on how exactly different herd statistics, animal KPI's and health events affect the milk production, which in turn could lead to more effective herd management. In contrast to Chapter 2 and 3, we did not use any interpretable modeling technique in Chapter 4. The reason is that in this study we apply deep learning models on multivariate timeseries data, which makes calculating the Variable Importance scores particularly hard as the features are temporal sequences. Moreover, the computational effort related to calculating Shapley values would now not only depend on the number of features, but also on the sequence length. Nevertheless, a perspective for future research would be to provide some interpretable insights into the data of this study by implementing techniques such as those proposed by Guo et al. (2019).

5.4 Final note

Moving towards a more sustainable agriculture is becoming quintessential in today's world as it is characterized by an increasing population growth, global warming and an immense use of natural resources. PLF technologies have been proposed to catalyze this transition as it allows to vigorously monitor every aspect of the food chain. Yet, as these technologies rapidly advance, the data collected on dairy farms becomes increasingly more complex. This dissertation bridges the gap between novel research and practical applications of animal monitoring systems based on complex PLF data. The study provides the dairy industry with new insights on how to develop accurate monitoring systems in their complex data landscape. More specifically, we uncovered how deep learning algorithms can be applied to build milk prediction and calving prediction systems suitable for practical applications. We suggest future research to extend these findings by applying the proposed frameworks in real-life dairy farms and validate their value in practice.

References

Adriaens, I., Huybrechts, T., Aernouts, B., Geerinckx, K., Piepers, S., De Ketelaere, B., and Saeys, W. Method for short-term prediction of milk yield at the quarter level to improve udder health monitoring. *Journal of Dairy Science*, 101 (11):10327–10336, 2018. ISSN 0022-0302. doi: https://doi.org/10.3168/jds.2018-14696.

Ali, T. E. and Schaeffer, L. R. Accounting for covariances among test day milk yields in dairy cows. *Canadian Journal of Animal Science*, 67(3):637–644, 1987.

Bareille, N., Beaudeau, F., Billon, S., Robert, A., and Faverdin, P. Effects of health disorders on feed intake and milk production in dairy cows. *Livestock Production Science*, 83(1):53–62, 2003. ISSN 0301-6226. doi: https://doi.org/10.1016/S0301-6226(03)00040-X.

Borchers, M., Chang, Y., Proudfoot, K., Wadsworth, B., Stone, A., and Bewley, J. Machine-learning-based calving prediction from activity, lying, and ruminating behaviors in dairy cattle. *Journal of Dairy Science*, 100(7):5664–5674, 2017. ISSN 0022-0302. doi: https://doi.org/10.3168/jds.2016-11526.

Bouallegue, M. and M'Hamdi, N. Mathematical modeling of lactation curves: A review of parametric models. *Lactation in farm animals-biology, physiological basis, nutritional requirements, and modelization*, 1:1–20, 2020.

Bórawski, P., Pawlewicz, A., Parzonko, A., Harper, K., Jayson, and Holden, L. Factors shaping cow's milk production in the eu. *Sustainability*, 12(1), 2020. ISSN 2071-1050. doi: 10.3390/su12010420.

Caixeta, L. S. and Omontese, B. O. Monitoring and improving the metabolic health of dairy cows during the transition period. *Animals*, 11(2), 2021. ISSN 2076-2615. doi: 10.3390/ani11020352.

Collier, R. J., Xiao, Y., and Bauman, D. E. Chapter 1 - regulation of factors affecting milk yield. In Watson, R. R., Collier, R. J., and Preedy, V. R., editors, *Nutrients in Dairy and their Implications on Health and Disease*, pages 3–17. Academic Press, 2017. ISBN 978-0-12-809762-5. doi: https://doi.org/10.1016/B978-0-12-809762-5.00001-2.

Cooper-Prado, M. J., Long, N. M., Wright, E. C., Goad, C. L., and Wettemann, R. P. Relationship of ruminal temperature with parturition and estrus of beef cows1. *Journal of Animal Science*, 89(4):1020–1027, 04 2011. ISSN 0021-8812. doi: 10.2527/jas.2010-3434.

Crociati, M., Sylla, L., van Straten, M., Stradaioli, G., and Monaci, M. Estimating the net return of a remote calving alarm system in a dairy farm. *Journal of Dairy Science*, 103(10):9646–9655, 2020. ISSN 0022-0302. doi: https://doi.org/10.3168/jds.2020-18253.

Crociati, M., Sylla, L., De Vincenzi, A., Stradaioli, G., and Monaci, M. How to predict parturition in cattle? a literature review of automatic devices and technologies for remote monitoring and calving prediction. *Animals*, 12(3), 2022. ISSN 2076-2615. doi: 10.3390/ani12030405.

Drackley, J. K. Biology of dairy cows during the transition period: the final frontier? *Journal of Dairy Science*, 82(11):2259–2273, 1999. ISSN 0022-0302.

emb. What is the cost of producing milk?, 2017.

Fadul, M., Bogdahn, C., Alsaaod, M., Hüsler, J., Starke, A., Steiner, A., and Hirsbrunner, G. Prediction of calving time in dairy cattle. *Animal Reproduction Science*, 187:37–46, 2017. ISSN 0378-4320. doi: https://doi.org/10.1016/j.anireprosci.2017.10.003.

Fernandes, L., Guimaraes, I., Noyes, N., Caixeta, L., and Machado, V. Effect of subclinical mastitis detected in the first month of lactation on somatic cell count linear scores, milk yield, fertility, and culling of dairy cows in certified organic herds. *Journal of Dairy Science*, 104(2):2140–2150, 2021. ISSN 0022-0302. doi: https://doi.org/10.3168/jds.2020-19153.

Giaretta, E., Marliani, G., Postiglione, G., Magazzù, G., Pantò, F., Mari, G., Formigoni, A., Accorsi, P., and Mordenti, A. Calving time identified by the automatic detection of tail movements and rumination time, and observation of cow behavioural changes. *Animal*, 15(1):100071, 2021. ISSN 1751-7311. doi: https://doi.org/10.1016/j.animal.2020.100071.

Goodfellow, I., Bengio, Y., and Courville, A. *Deep learning*. MIT press, 2016.

Green, L., Hedges, V., Schukken, Y., Blowey, R., and Packington, A. The impact of clinical lameness on the milk yield of dairy cows. *Journal of Dairy Science*, 85(9):2250–2256, 2002. ISSN 0022-0302. doi: https://doi.org/10.3168/jds.S0022-0302(02)74304-X.

Guo, T., Lin, T., and Antulov-Fantulin, N. Exploring interpretable LSTM neural networks over multi-variable data. In Chaudhuri, K. and Salakhutdinov, R., editors, *Proceedings of the 36th International Conference on Machine Learning*, volume 97 of *Proceedings of Machine Learning Research*, pages 2494–2504. PMLR, 09–15 Jun 2019.

Hogeveen, H., Kamphuis, C., Steeneveld, W., and Mollenhorst, H. Sensors and clinical mastitis—the quest for the perfect alert. *Sensors*, 10(9):7991–8009, 2010. ISSN 1424-8220.

Hogeveen, H., Rutten, N., Kamphuis, C., and van der Voort, M. Economics of precision dairy monitoring techniques1. In *Conference on Precision Dairy Farming*, page 87, 2017.

Jeretina, J., Babnik, D., and Škorjanc, D. Prediction of standard lactation curves for primiparous Holstein cows by using corrected regression models. *Italian Journal of Animal Science*, 14(4):3776, 2015.

Jia, R., Dao, D., Wang, B., Hubis, F. A., Hynes, N., Gürel, N. M., Li, B., Zhang, C., Song, D., and Spanos, C. J. Towards efficient data valuation based on the shapley value. In *The 22nd International Conference on Artificial Intelligence and Statistics*, pages 1167–1176. PMLR, 2019.

Khatun, M., Thomson, P., Kerrisk, K., Lyons, N., Clark, C., Molfino, J., and García, S. Development of a new clinical mastitis detection method for automatic milking systems. *Journal of Dairy Science*, 101(10):9385–9395, 2018. ISSN 0022-0302. doi: https://doi.org/10.3168/jds.2017-14310.

Li, W., Geng, C., and Chen, S. Leave zero out: Towards a no-cross-validation approach for model selection. *arXiv*, abs/2012.13309, 2020.

Liu, M., Ren, S., Ma, S., Jiao, J., Chen, Y., Wang, Z., and Song, W. Gated transformer networks for multivariate time series classification. *CoRR*, abs/2103.14438, 2021.

Livingston, F. Implementation of breiman's random forest machine learning algorithm. *ECE591Q Machine Learning Journal Paper*, pages 1–13, 2005.

Macciotta, N. P., Dimauro, C., Rassu, S. P., Steri, R., and Pulina, G. The mathematical description of lactation curves in dairy cattle. *Italian Journal of Animal Science*, 10(4):e51, 2011.

Machado, N. A., Da Costa, L. B., Barbosa-Filho, J. A., De Oliveira, K. P., De Sampaio, L. C., Peixoto, M. S., and Damasceno, F. A. Using infrared thermography to detect subclinical mastitis in dairy cows in compost barn systems. *Journal of Thermal Biology*, 97:102881, 2021. ISSN 0306-4565. doi: https://doi.org/10.1016/j.jtherbio.2021.102881.

Olston, C., Fiedel, N., Gorovoy, K., Harmsen, J., Lao, L., Li, F., Rajashekhar, V., Ramesh, S., and Soyke, J. Tensorflow-serving: Flexible, high-performance ML serving. *ArXiv e-prints*, abs/1712.06139, 2017.

Ouellet, V., Vasseur, E., Heuwieser, W., Burfeind, O., Maldague, X., and Charbonneau, É. Evaluation of calving indicators measured by automated monitoring devices to predict the onset of calving in Holstein dairy cows. *Journal of Dairy Science*, 99(2):1539–1548, 2016. ISSN 0022-0302. doi: https://doi.org/10.3168/jds.2015-10057.

Post, C., Rietz, C., Büscher, W., and Müller, U. Using sensor data to detect lameness and mastitis treatment events in dairy cows: A comparison of classification models. *Sensors*, 20(14), 2020. ISSN 1424-8220.

Rekik, B. and Gara, A. Factors affecting the occurrence of atypical lactations for Holstein–Friesian cows. *Livestock Production Science*, 87(2):245 – 250, 2004. ISSN 0301-6226.

Rutten, C., Velthuis, A., Steeneveld, W., and Hogeveen, H. Invited review: Sensors to support health management on dairy farms. *Journal of Dairy Science*, 96(4): 1928–1952, 2013. ISSN 0022-0302. doi: https://doi.org/10.3168/jds.2012-6107.

Rutten, C., Kamphuis, C., Hogeveen, H., Huijps, K., Nielen, M., and Steeneveld, W. Sensor data on cow activity, rumination, and ear temperature improve prediction of the start of calving in dairy cows. *Computers and Electronics in Agriculture*, 132:108–118, 2017. ISSN 0168-1699. doi: https://doi.org/10.1016/j.compag.2016.11.009.

Scheurwater, J., Hostens, M., Nielen, M., Heesterbeek, H., Schot, A., van Hoeij, R., and Aardema, H. Pressure measurement in the reticulum to detect different behaviors of healthy cows. *PLOS ONE*, 16(7):1–13, 07 2021. doi: 10.1371/journal.pone.0254410.

Silvestre, A., Petim-Batista, F., and Colaço, J. The accuracy of seven mathematical functions in modeling dairy cattle lactation curves based on test-day records from varying sample schemes. *Journal of Dairy Science*, 89(5):1813 – 1821, 2006. ISSN 0022-0302.

Statham, J. M. E., Randall, L. V., and Archer, S. C. Reduction in daily milk yield associated with subclinical bovine herpesvirus 1 infection. *Veterinary Record*, 177(13):339–339, 2015. doi: https://doi.org/10.1136/vr.103105.

Wolpert, D. H. The Lack of A Priori Distinctions Between Learning Algorithms. *Neural Computation*, 8(7):1341–1390, 10 1996. ISSN 0899-7667. doi: 10.1162/neco.1996.8.7.1341.

Zehner, N., Niederhauser, J. J., Schick, M., and Umstatter, C. Development and validation of a predictive model for calving time based on sensor measurements of ingestive behavior in dairy cows. *Computers and Electronics in Agriculture*, 161:62–71, 2019. ISSN 0168-1699. doi: https://doi.org/10.1016/j.compag.2018.08.037.

Zerveas, G., Jayaraman, S., Patel, D., Bhamidipaty, A., and Eickhoff, C. A transformer-based framework for multivariate time series representation learning. In *Proceedings of the 27th ACM SIGKDD Conference on Knowledge Discovery & Data Mining*, KDD '21, page 2114–2124, New York, NY, USA, 2021. Association for Computing Machinery. ISBN 9781450383325. doi: 10.1145/3447548.3467401.

Zhao, B., Lu, H., Chen, S., Liu, J., and Wu, D. Convolutional neural networks for time series classification. *Journal of Systems Engineering and Electronics*, 28(1): 162–169, Feb 2017. ISSN 1004-4132.