A Thesis Submitted for the Degree of Doctor of Philosophy at

Harper Adams University

Copyright and moral rights for this thesis and, where applicable, any accompanying data are retained by the author and/or other copyright owners. A copy can be downloaded for personal, non-commercial research or study without prior permission or charge. This thesis and the accompanying data cannot be reproduced or quoted extensively without first obtaining permission in writing from the copyright holder/s. The content of the thesis and accompanying research data (where applicable) must not be changed in any way or sold commercially in any format or medium without the formal permission of the copyright holder/s.

When referring to this thesis and any accompanying data, full bibliographic details, including the author, title, awarding institution and date of the thesis, must be given.

Understanding and Predicting Pen Fouling, Tail Biting, and Diarrhoea in Farmed Pigs.

Thesis submitted to the Harper Adams University for the degree of Doctor of Philosophy

Engineering Department.

Yuvraj Domun

BEng. MSc.

2023

Harper Adams University

Declaration

The thesis has been written by me and describes the work carried out by myself unless otherwise stated. Information from other sources has been fully acknowledged and referenced in the text.

Yuvraj Domun

April 2023

Acknowledgements

This work was possible thanks to the funding provided by John the Oldacre Foundation and the Douglas Bomford Trust. Trust. The data source used for this research was funded by the Green Development and Demonstration Programme under the Ministry of Food, Agriculture and Fisheries, Denmark, Project IntactTails j.nr. 34009-13-074.

Table of Contents

Abstract

This PhD research investigated precision animal farming, specifically emphasising commercially reared pigs and their welfare, addressing concerns like pen fouling, tail-biting, and diarrhoea. While animal welfare in pig farming is critical, there is a lack of comprehensive predictive models that integrate various factors affecting pig behaviours. The primary objective was to create advanced algorithms and predictive models that combine mechanistic modelling and machine learning to better understand and predict pig behavioural dynamics related to welfare issues. Various methods were employed, including transfer function models to link water consumption with temperature differences, analysing spatial positioning in relation to fouling events, and employing neural network architectures for time series data. Bayesian networks were utilised for simulating intervention scenarios.

Several significant discoveries were made during the research. Anomalies in pigs' water consumption that were linked to temperature variations were effectively identified by the transfer function model, giving valuable insights into pen fouling and tail-biting incidents. It was also discovered that a crucial role in influencing fouling events in pigs is played by spatial positioning and temperature differences between different activity areas within pig pens. Superior predictive capabilities for events such as fouling, tail-biting, and diarrhoea were demonstrated by the innovative application of a neural network approach to predict these events. Furthermore, an early warning system that utilised hierarchical clustering and principal component analysis was introduced, which showed strong predictive potential. Finally, this research also demonstrated that Bayesian Network simulations can be used as a non-invasive method to test for potential strategies to mitigate welfare issues in farmed pigs while also providing practical insights for better farm management.

This research offers vital tools and insights for advancing precision pig farming, fostering a more sustainable and ethical approach. The developed algorithms not only contribute to better pig welfare but also enhance monitoring, potentially leading to increased farm profitability. While the models are promising, further refinement and research into the various factors affecting pig behaviour are recommended.

Chapter 1: General Introduction

1.1 Introduction

The influence of the thermal environment on pigs' physiological, nutritional, and psychological responses has been well documented in numerous studies (Huynh et al., 2005; Huynh et al., 2006; Jones & Nicol, 1998; Patience et al., 2005). These studies have highlighted the effects of temperature on pigs' behaviours, including lying and excretion (Banhazi et al., 2008; Huynh et al., 2005). Furthermore, it is worth noting that intensive pig production contributes significantly to the emission of ammonia and greenhouse gases, with potential health risks to humans and animals and adverse impacts on the environment and climate (Banhazi et al., 2008; Philippe et al., 2011; Philippe & Nicks, 2015; Phillips et al., 1998)

Historically, attempts to control these environmental conditions have been made to minimise the pigs' adverse behaviours (Geers et al., 1986; Geers & Bosschaerts, 1989). A consistent argument among researchers is the importance of maintaining a stable and homogeneous airflow distribution to reduce unwanted behaviour (Huynh & Aarnink, 2004; Wiegand et al., 1994). Studies suggest that pen fouling in growing pigs can be mitigated by adequately controlling floor surface temperature, air temperature, and air velocity.

Indeed, research by Huynh & Aarnink (2004), like earlier work by Geers et al. (1986) and Geers & Bosschaerts (1989), demonstrated the impact of floor cooling in partially solid floor systems on the behaviour and performance of growing-finishing pigs. It was found that pigs preferred lying on the cooled floor in high ambient temperature conditions, improving their welfare and promoting growth. However, while floor temperature control offered some behavioural stabilisation, it fell short in managing airborne air quality factors, such as moisture and toxicity levels.

Although there are helpful design recommendations to foster homogeneous airflow within pig buildings (Geers & Bosschaerts (1989)), these suggestions are over 25 years old, and control engineering and ventilation system designs have since advanced significantly. The indoor climate of livestock buildings undeniably plays a crucial role in the animals' well-being, health, and production performance (e.g., daily weight gain). The Literature (Aarnink et al., 2001; Statham et al., 2009) indicates that animal behaviour can be regulated through an effective micro-climatic controller. However, the complexity of the interaction between behavioural activity and the micro-climate is a challenge.

Current control systems are insufficiently robust and primarily focused on controlling set temperatures and humidity levels. There is a conspicuous research gap in designing control systems considering floor usage and activity within housing areas. Preliminary research on poultry by Youssef et al. (2015) has shown promise in designing a control system that adjusts the inlet temperature and ventilation rate based on a dynamic activity index.

1.2 Environmental Control Systems

The central role of an environmental control system is to optimise the various atmospheric variables for both animals and farmers through the modulation of heating, ventilation, and air humidifiers. Within the scope of this research, the focus is on leveraging the environmental control system to stabilise animal behaviour.

The agriculture industry primarily utilises conventional staged ventilation to maintain favourable conditions within livestock farming environments. Several works on behavioural research, including work by Fraser (2003), Gallo et al. (2022) and Mendl et al. (2010), have provided insights into the ideal environmental parameters that can foster a high welfare environment for livestock. These studies underscore the necessity of understanding animal emotions, using behaviour for assessing animal welfare, and the pivotal role of behaviour in understanding the needs of animals. Moreover, they highlight the pressing need for ongoing research into animal welfare, particularly in identifying the right environmental conditions to promote high welfare.

Today's prevalent control systems operate through discrete ventilation and heating stages to counter deviations from the desired set point. The widely accepted and most effective practice is an automatic feedback control based on the internal air temperature gauged at a singular position within the building's volume (Timmons et al., 1995). However, despite modern technology, maintaining constant vigilance and control over these fixed set points proves challenging for micro-climatic controllers (Geers et al., 1986). Existing control laws for indoor-farmed animals often rely on the premise that indoor micro-climate can be solely regulated by maintaining specific temperature or humidity levels. However, such approaches have proven to be unreliable (Roque et al., 2016). The primary reason for this unreliability is that these simplistic control methods aren't robust enough to handle the complexities and uncertainties of indoor farming systems. Research suggests that indoor livestock environments can contain a multitude of contaminants, including harmful microorganisms. These contaminants pose risks to the health of both the animals and the farm workers (Roque et al., 2016). The quality of air, especially its microbial content, in these settings is influenced by factors like stocking density, barn cleanliness, microclimatic conditions (like temperature, humidity, and gases), and the efficiency of the ventilation system (Roque et al., 2016; Shao et al., 2019).

The omission of the spatial distribution of animals in the current climatic control systems for indoor farming contributes to the inability to stabilise behaviour within housing units (Roque et al., 2016). The lack of consideration for the spatial distribution of animals makes it problematic to accurately control and regulate the microclimate conditions in different areas of the facility. This underlines the need for robust control schemes considering animals' spatial dynamics and behaviour. Hence, it is necessary to develop more reliable control schemes that consider the complex relationship between the spatial dynamics of animals and microclimate parameters to manage indoor farming environments and stabilise animal behaviour.

Existing environmental controllers depend on indoor temperature and apply a constant minimum ventilation rate to govern humidity levels. This approach arises from the impracticality of developing a thermal index to control the micro-environment (Lemay et al., 2001; Soldatos et al., 2005). The corrosive conditions within livestock housing facilities can degrade the humidity sensor, compromising the humidity data's accuracy. This results in some sensors failing after short-term barn exposure, which limits the integration of relative humidity data into the environmental control strategy for livestock buildings. Hence, most current animal production facilities adopt temperature-only control strategies as humidity control is not extensively implemented. This imprecise control over ventilation rates can lead to substantial production losses and ventilation-related health issues in modern livestock buildings (Taylor et al., 2004).

Behavioural scientists emphasise that the quality of micro-climatic conditions inside livestock buildings emerges from intricate interactions among multiple climatic parameters. A control system emphasising temperature overlooks the complex interrelationships between temperature, airspeed, and humidity. In addition, actuators such as windows and regulation valves in these controlled environments are susceptible to varying external factors like wind velocity and outside temperature. Consequently, there is a significant requirement for a system that can monitor animal behaviour and adjust and control key environmental conditions in response. The outputs of this system would primarily involve regulating heating and ventilation within livestock buildings, ensuring optimal conditions are maintained for animal welfare and productivity. Such a comprehensive approach allows for a more effective and robust control scheme that adapts to the complexity of the livestock environment.

In conclusion, a climate control system should ideally be devised harmoniously with animals' behavioural responses to environmental stimuli. The main hurdle is in accurately and consistently capturing these behavioural dynamics. The following section will delve into potential solutions to overcome this challenge.

1.2.1 Bio-Energetic Models

Bio-energetic models are valuable tools in designing control systems that optimally maintain a comfortable environment for animals. These models typically estimate the heat production of animals by considering changes in their physiology, various environmental factors, or a combination of both. They aim to capture the complex energy exchange processes between the animal and its environment. These bio-energetic models can be broadly categorised into three types:

1. **Empirical Models:** These models are developed based on observations and data collection, relying heavily on statistical techniques to form relationships.
2. **Mechanistic Models:** These models rely on the underlying biological mechanisms and processes. They incorporate known physiological and biophysical principles to predict outcomes.
3. **Dynamic Data-Based Models**: These models are founded on real-time data, capturing the dynamic relationship between variables over time.

These models aim to provide a comprehensive understanding of animal bio-energy systems, leading to the development of more efficient and effective environmental control systems.

1.2.1.1 Empirical models

Empirical heat production models are used to understand the cause and effect of specific parameters on the bio-energetic variables (e.g., heat and moisture) (Aerts et al., 2000; Bridges & Gates, 2009; D. M. Green & Parsons, 2006). Early efforts (Bond et al., 1952; 1959, 1965; Heitman et al., 1958; Morrison et al., 1967) have seen the development of heat and moisture models for pigs in response to the live animal weight and room temperature. Other empirical models can be found in the literature that describe the bio-energetic system of dairy cows (Brody, 1945), beef cattle (Yeck et al., 1960; Yeck & Stewart, 1959) and broilers (Deaton et al., 1969; Longhouse, 1967; Reece & Lott, 1982a, 1982b).

One drawback of empirical models is that they are case-specific, i.e., they can only predict the deterministic responses in scenarios that match the experimental conditions that prevailed during the data-collection phase of the model development (Black, 2014; Bridges & Gates, 2009; A. R. Green & Xin, 2009). From the control systems design perspective, these models have limited use when used alone (Aerts et al., 2003; Wathes et al., 2008) as they are static models, meaning that they do not change or adapt over time. Thus, they fall short in capturing the evolving and dynamic nature of bio-energetic systems, especially in response to different input variables (Aerts et al., 2003; Wathes et al., 2008).

1.2.1.2 Mechanistic Models

Mechanistic models represent the behavioural and biological systems of animals, such as thermal regulation, metabolism, and feeding performance, in relation to changes in their microenvironment. Grounded in the laws of physics (Norton et al., 2007, 2009, 2010) and chemistry, these models often exhibit better reliability than empirical models. This is because the physical and chemical laws that describe biological responses are usually consistent across various experimental conditions.

There are multiple models (Black et al., 1986; Bruce & Clark, 1979; Close & Mount, 1978; Jacobson, 1983; Teter et al., 1973b, 1973a; Usry et al., 1992; Watt et al., 1987) that used a mechanistic modelling approach to predict the behavioural and physiological responses of pigs. For instance, Teter et al. (1973a, 1937b) developed several models for pigs, beef cattle, and broilers to estimate physiological responses, such as feed intake, weight gain, and feed efficiency, in relation to changes in air temperature. Other literature also offers mechanistic models that predict heat production in growing pigs concerning their live weight and metabolic energy intake (Bruce & Clark, 1979; Close & Mount, 1978; Holmes & Close, 1977; Jacobson, 1983).

Mechanistic models hold significant potential for modelling and predicting physiological and biological responses due to their in-depth representations of the systems under study (Tsamandouras et al., 2015). However, their inherent complexity and the many undefined variables they incorporate often make them challenging to manage (Tsamandouras et al., 2015). As such, these models frequently use estimates derived from empirical models for initialization, an approach necessitated by the difficulties associated with quantifying every variable (Tsamandouras et al., 2015). In response to these challenges, researchers are increasingly exploring minimal or semi-mechanistic models, which maintain the physiological mechanistic nature where it is most relevant, offering enhanced flexibility (Tsamandouras et al., 2015). Furthermore, a promising strategy involves bridging mechanistic and phenomenological models using tools like the Manifold Boundary Approximation Method (MBAM), which can simplify complex mechanistic models into more digestible phenomenological counterparts (Transtrum & Qiu, 2016). Despite the intricacies, mechanistic models remain indispensable tools in physiological research, yielding invaluable insights into the processes they describe.

Mechanistic models, however, require regular updates due to their dependence on empirical estimates. Continuous research is needed to ensure the validity of these estimates. Studies (Brown-Brandl et al., 2004; Chepete & Xin, 2001) have shown that the physiological responses in pigs to their micro-environment have evolved due to the change in animal genetics. For example, fasting heat production changed by 18% from 1984 to 2002 because of increased lean tissue accretion rates (Brown-Brandl et al., 2004). With time, it has also been found that pigs' thermal response and sensitivity change (ASABE, 2012). Hence, any mechanistic models will need to be updated regularly as their validity can be questioned with time.

1.2.1.3 Dynamic Data-Based Models

Dynamic data-based models provide a solution to some of the limitations of empirical and mechanistic models as they recursively update the model parameters by continuously extracting new features from the newly available data set within the bio-system of interest. Dynamic data-based modelling can deal with the non-stationary response of biological systems (heat production, behaviour, and feed intake) to disturbing factors (climate). However, the main limitation of this class of model is its reliability on data; hence, in some cases, they may require an expensive set of sensors, which is not always possible at the farm level. For example, sensors for monitoring heat production as required by existing dynamic data-based models (Aerts et al., 2000; Madsen et al., 2005) are not usually available on farms.

There is very little research (Aerts et al., 2000; Madsen et al., 2005) on dynamic data-based models to quantify the dynamic biological response of pigs. However, some research has been carried out for broilers, where predictive models were developed to control broilers’ growth trajectory by recursively adjusting the food supply.

1.2.2 Thermal Comfort Indices

To design a control system to optimise animal comfort, it is essential to have adequate functions that can be used to translate how the animals perceive the combinations of climate control variables. Consequently, much research has focused on developing animal comfort indices (ACIs) (Brown-Brandl, 2013; Hahn et al., 2009; Nääs et al., 2006). ACIs are empirical and are identified by comparing selected performance criteria (e.g., core body temperature, respiration rate, growth, feeding behaviour) to environmental variables.

In pigs, a combination of temperature and humidity has been used in literature to quantify the amount of heat energy (enthalpy) to measure animal comfort (Moura et al., 1997; Rodrigues et al., 2011). Other indices used radiation and airspeed as a measure of comfort (Gaughan et al., 2012; Hahn et al., 2009). An in-depth review of ACIs is given by (Fournel et al., 2017). However, many indices are empirical and developed 20-50 years ago (Baker, 2004; Buffington et al., 1981). Moreover, as previously discussed, empirical model validity can be questioned and is not always reliable due to time dependent variations in animal behaviour. Furthermore, some developed indices (Da Silva et al., 2007; Gebremedhin & Wu, 2005) rely on impractical farm-level measurements (respiration rate, skin temperature, sweating rate).

1.2.3 Behavioural Responses

The feeding behaviour of animals serves as a valuable indicator of their overall well-being and health status (Banhazi et al., 2007; Brown-Brandl, 2013; Kashiha et al., 2013). In many cases, alterations in feeding behaviour have been employed for early detection of changes in the animals' physical conditions (Madsen et al., 2005; Nienaber & Hahn, 2000), such as the onset of diarrhoea, which might indicate digestive issues or adjustment to dietary changes.

However, it is essential to note that other phenomena, such as pen fouling or tail-biting, while potentially indicative of discomfort or behavioural disruptions in the animals, may not directly correspond to a decline in health. Pen fouling, for example, could be a response to thermal stress rather than a sign of poor health. Similarly, tail-biting might primarily be a behavioural issue, potentially triggered by environmental stressors or inadequate enrichment, although health problems could also contribute.

Other behavioural measures, such as animal activity within its enclosure, can also contain helpful information to indicate the animal’s physiological state (Frost et al., 1997). Very sophisticated sensors, such as ultrasonic proximity sensors

(Hillman et al., 2000)], pedometers (Walker et al., 1985), and accelerometers (Darr & Epperson, 2009; Müller & Schrader, 2003; Ouellet et al., 2016; Robert et al., 2009) have been used to monitor animal activity.

The list of sensors used in precision livestock farming is not exhaustive. However, the main problem remains the availability of cheap, dependable, and robust sensors to implement model-based control algorithms in agricultural buildings (Fournel et al., 2017).

1.3 Research Objective and Hypotheses Formulation

Predicting animal welfare issues is a challenging task that necessitates accurate algorithms based on dependable data. The use of traditional statistics and machine learning techniques is used in this research to understand and to forecast animal welfare issues is proposed in this research, with data derived from multiple sources including water consumption, pig positions in pens, temperature sensors, and information from the farm climate system.

1.3.1 Objectives

The goal of this study is to have a set of targeted algorithms created that can accurately predict and help to better understand animal welfare problems. Although issues such as pen fouling, tail biting, and diarrhoea in pigs raised for farming are the focus of this research, it is understood that welfare challenges are multifaceted and varied. To address this, a range of algorithms has been used to tailor each model to the specific characteristics of each welfare problem. This approach will enable greater precision and usefulness to be achieved, potentially allowing other welfare issues beyond the scope of this research to be tackled by the algorithms.

1.3.1.1 Research objectives

To achieve the general objective, three specific objectives were set:

Objective 1: Investigate the usefulness of using farm-level information to predict tail-biting, pen-fouling, and diarrhoea.

This Objective was achieved through chapters 2 and 3. In these two chapters, mechanistic modelling and statistical analysis were performed to investigate potential information that can be used to predict specific welfare issues. Chapter 2 proposed using a mechanistic behavioural model based on the water consumption of pigs and their pen's activity to predict tail-biting. Chapter 3 uses statistical analysis to investigate the relationship between environmental factors and pen fouling.

Objective 2: Develop a general approach that can be applied and adapted to predict specific welfare issues within farmed animals.

This objective was achieved through Chapter 4. In Chapter 4, it was demonstrated that a novel deep neural network could learn to perform discriminate prediction of tail-biting, pen-fouling, and diarrhoea.

Objective 3: Develop approaches that can be used to interpret the learning from black-boxed artificial intelligence (AI) models.

Our objective was reached through chapters 5 and 6. A hybrid method that combines feature selection techniques to identify crucial indicators in time series data for the early detection of tail-biting, pen-fouling, and diarrhoea in pigs was introduced in Chapter 5. This hybrid approach integrates similarity-based features and other data processing methods to improve the accuracy and effectiveness of predictive models. In Chapter 6, scenario models were created using Bayesian Networks and do-Calculus that can determine the best scenarios to decrease the frequency of tail-biting, diarrhoea, and pen-fouling outbreaks.

1.3.2 Hypotheses Formulation

Drawing from the objectives, the study postulates several hypotheses:

Mechanistic Behavioural Model: It is hypothesized that an innovative mechanistic behavioural model, utilizing water consumption and pen activity parameters, will be effective in predicting tail-biting behaviour in commercially farmed pigs. It is expected that by filtering the frequency to reduce noise in temperature data, an improvement in the model's ability to detect tail-biting and pen fouling events with acceptable sensitivity will be achieved.

Spatial Positioning and Temperature Differences: It is suggested that the positioning of pigs and the temperature differences between their resting and excreting areas are important factors in determining fouling events. Furthermore, a negative relationship between the likelihood of fouling and the amount of water consumed by pigs is hypothesized.

Machine Learning and Pattern Recognition: The study asserts that a specifically designed machine learning structure, incorporating stacked bidirectional long short-term memory and feedforward neural network, can efficiently learn from and categorize patterns in time-series data, thereby enabling accurate predictions of tail-biting, fouling, and diarrhoea events.

Early Warning System: It is postulated that a meticulously designed early warning system, aiming to monitor and predict pig behaviours related to fouling, tail-biting, and diarrhoea and employing hierarchical clustering and principal component analysis for data pre-processing and feature extraction, will show satisfactory predictive capabilities for these behaviour issues.

Intervention Scenarios: It is proposed that the identification of effective strategies to prevent pen fouling can be achieved using Bayesian Networks to simulate and assess different intervention scenarios. Specifically, it is anticipated that the likelihood of fouling can be significantly decreased by measures such as reducing extreme ventilation output and keeping the lying area cooler.

These hypotheses lay the foundation for this doctoral research, steering investigations and algorithm development towards the early prediction and comprehension of animal welfare issues in commercially farmed pigs. This study is consequently dedicated to elevating the surveillance and management of pigs.

Chapter 2: Investigating the feasibility of using mechanistic modelling to detect tail biting and pen fouling.

2.1 Abstract

This study investigates the dynamics of water consumption and zonal temperature difference in a pig population to establish a transfer function model that captures the relationship between these aforementioned variables. The zonal temperature difference is a proxy for animal spatial positioning within the pens. System identification techniques are used to identify the optimal model structure and accurately characterise the underlying behavioural patterns. Two distinct behavioural modes are discovered by analysing the poles and employing k-means clustering, each representing unique aspects of pig behaviour. The event detection system implemented in this study offers valuable insights into abnormal behaviour dynamics, although further refinements are necessary to enhance the overall predictive performance. In conclusion, this study makes a valuable contribution to the progress of precision animal farming practices, highlighting the significance of interdisciplinary approaches in enhancing animal welfare within controlled environments.

2.2 Introduction

The growing need for precision in livestock farming has highlighted the importance of advanced monitoring technologies for diagnosing and managing animal health and behaviour. This doctoral research focuses on early detection and a thorough understanding of animal welfare issues in commercial pig farming. While the broader thesis investigates three critical welfare issues, pen fouling and tail-biting, this chapter focuses on detecting tail-biting and fouling. Chapter 1 sets the groundwork by introducing the topics under discussion and outlining the overall scope of the research. Building on this base, this chapter presents an innovative behavioural model based on detailed water consumption and pen activity analysis. The primary goal is to predict early tail-biting and pen-fouling in pigs.

In this chapter, the proposed model uses cost-effective sensor technology to observe and interpret pig behaviour. Frequency filtering is employed to overcome the challenges posed by the high noise level in sensor data collected from farms. In this chapter, ‘frequency’ refers to the rate at which specific behaviours occur within a defined period. Recurring behavioural patterns can be isolated by identifying these fundamental frequencies, serving as critical indicators to detect deviations. Such deviations from the main frequencies may signal signs of stress or poor welfare conditions in pigs. The model employs frequency analysis to understand pig behaviour, aiming to promote better animal welfare.

The mechanistic modelling approach presented in this chapter advocates a proactive method for detecting tail-biting and pen-fouling incidents in pigs, with the potential to improve pig behaviour monitoring. In this research, ‘mechanistic modelling’ means modelling the system using the relationship between water consumption and spatial positioning with the pens. This study aims to develop a robust and responsive system by understanding recurring behavioural patterns under normal conditions, enabling early detection of signs of distress or discomfort. This chapter’s primary goal is to assess the proposed model’s feasibility. This chapter is a step towards achieving the broader objectives of this PhD thesis as more innovative methods for enhancing animal welfare are explored in subsequent chapters.

2.3 Material and Methods

The research study was carried out over two years, from 2015 to 2016, strictly adhering to a protocol approved by the Danish Animal Experiments Inspectorate (Journal no. 2015-15-0201-00593). During this period, four cohorts of pigs were sequentially introduced into each pen, marking distinct rounds of the experiment. The respective timelines for each round were as follows:

- Round 1: June 16th, 2015, to September 3rd, 2015.
- Round 2: September 14th, 2015, to December 3rd, 2015.
- Round 3: January 12th, 2016, to March 31st, 2016.
- Round 4: September 7th, 2016, to November 26th, 2016.

All the data collected for this thesis were part of a broader study (Larsen et al., 2016, 2017), and all welfare event observations followed a predefined observation protocol (Lyderik et al., 2016).

2.3.1 Animal Selection, Housing, and Management

The research included 1624 finisher pigs from the same herd, divided among 112 pens. The pigs, bred from Danavl Yorkshire x Danavl Landrace dams and inseminated with Danavl Duroc semen, adhered to Danish production standards. The experiment was executed in four batches at the Department of Animal Science, Aarhus University, Denmark. The research comprised one weaner section (7 to 30 kg) and two finisher sections, each with 16 pens. All the work done in this thesis uses the data from the finisher sections. Figure 2.1 shows an illustration of all the pens used.

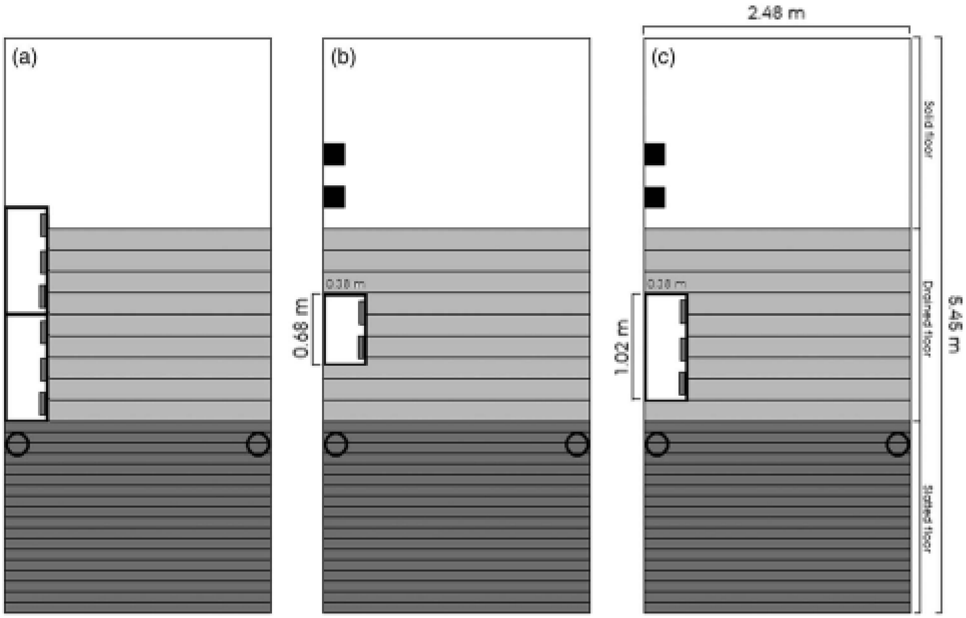

Figure 2.1: The various pen designs used in the study. Three distinct pen designs were used: one for weaners (a), another for finisher pens accommodating 1.21 m^2 per pig (b), and a third for finisher pens providing 0.73 m^2 per pig (c). Despite having identical outer dimensions, the pens differed in their internal features. Water cups' positions are marked with circles, while solid black squares represent wooden sticks placed in separate racks.

2.3.2 Weaner Section

Upon arrival, the pigs had an average weight of 9.1 ± 1.7 kg and were distributed among 14 weaner pens. Half the pens hosted pigs with docked tails, while the other half accommodated pigs with undocked tails. Staff were also trained to promptly identify and address early signs of tail damage, which resulted in the identification and removal of only five obsessive biters throughout the four batches.

2.3.3 Finisher Section

Once the pigs reached an average weight of 31.6 ± 6.6 kg, they were transferred to the finisher section. The pens were 5.45 m by 2.48 m, with the floor equally divided into solid concrete (rest area), drained (activity Area) and slatted floor (excreting area). The gap between the slats in the slatted floor and the drained floor was 2 cm, while the respective slats were 180 mm and 80m wide. An overview of a pen with sensor locations is given in Figure 2.2.

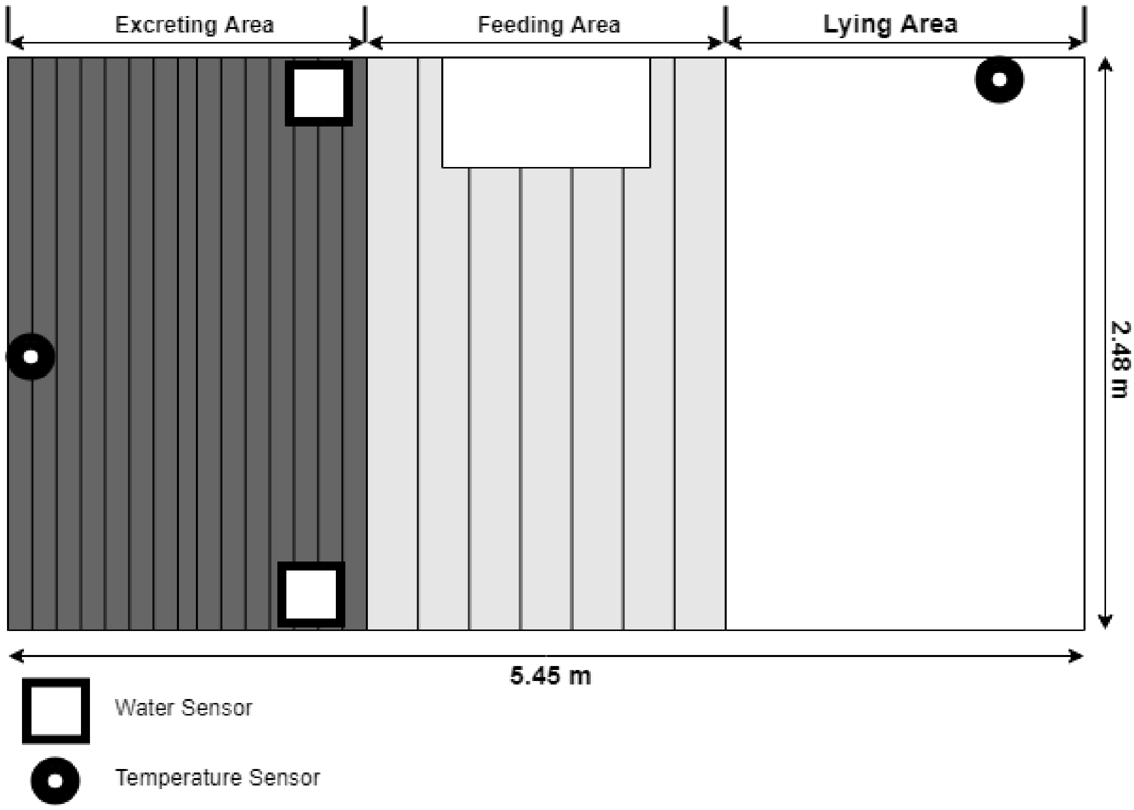

Figure 2.2: A top view of a pen with the location of different sensors and the designated rest, feeding, and lying area.

Strict monitoring ensured that no pig with tail damage entered this section. As per EU legislation, each pen was provided with a dry feeder and two wooden sticks for enrichment. Artificial light was on from 0530 to 1830 h (182 lux). The finisher units' climate (temperature (°C) and humidity (%)) were controlled by activating the heating, cooling or ventilation system via the climate control systems (SKOV A/S, Roslev, DK). Each pen included an automatically controlled shower system (SKOV A/S, Roslev, Denmark) above the slatted floor. The pigs were fed ad libitum with a commercial dry feed, and the feeders were filled three times a day at 0300, 1000 and 1830h. The educated farm staff performed the general farm management. Between 1000h and 1200h, the general routine in the stables was performed, including cleaning, straw provision, and a general health check of each pen.

2.3.4 Experimental Design

The finisher pens were randomly assigned various treatments concerning tail docking, provision of straw, and stocking density. The tail docking procedure, performed within the first four days post-birth using a hot-iron cutter, conformed to Danish legislation. During the research, both docked and undocked pigs were placed in pens with or without straw. The pigs were allocated different space allowances, precisely 1.21 m²/pig and 0.73 m²/pig. Each combination of conditions had four replicates for the first, third, and fourth batches. However, the second batch had three replicates for most conditions, except undocked pigs in straw and non-straw pens with a space allowance of 1.21 m²/pig, where only one replicate was observed (Larsen et al., 2016, 2018).

2.3.5 Data Collection and Usage

The scope of this study required the collection and use of various variables. These data points were integral in analysing the pigs' environmental conditions, physiological states, and behaviours that could potentially influence fouling events within the pig pen environment. The following sections present the variables used, their sampling rate, and, where relevant, the units of measurement.

2.3.5.1 Environmental and Water Variables

A set of environmental and water variables was captured to assess the conditions in the pig pens. These variables include:

- Temperature (Solid Floor): The pig pens' solid floor temperature was measured in degrees every minute.
- Temperature (Slatted Floor): Similarly, the temperature of the slatted floor was recorded every minute in degrees.
- Water Consumption (Drinker 1 and Drinker 2): The water consumed by two different drinkers was recorded every 10 seconds, measured in litres.
- Relative Humidity (Finisher Unit): The relative humidity in the finisher unit was measured every minute and expressed as a percentage.
- Ventilation Output (Finisher Unit): The output of the ventilation system in the finisher unit was recorded every minute and expressed as a percentage.
- Heating Output (Finisher Unit): The heating output in the finisher unit was measured every minute and expressed as a percentage.
- Cooling Output (Finisher Unit): The cooling output in the finisher unit was also captured every minute and expressed as a percentage.
- Temperature (Finisher Unit): The ambient temperature in the finisher unit was recorded every minute..

2.3.5.2 Physiological Variables

Physiological variables related to the pigs were collected to understand the correlation between their states and the fouling events:

- Age of Pigs: The age of the pigs was recorded daily in terms of days after their insertion into the pens.
- Number of Pigs: The number of pigs in each pen was recorded daily.

2.3.5.3 Behavioural Variables

To further understand the behavioural aspects leading to fouling events, additional data was collected:

- Straw: This binary variable indicates whether straw was present in the pig pens.
- Tail Type: These binary variable records whether pigs had curly or straight tails.

2.3.5.4 Behavioural Video Observations

Behavioural patterns were recorded from view observation for five days before each fouling event occurred. Additionally, each pen identified with a fouling event was paired with a corresponding control pen for comparative analysis. The variables include:

- % Lying (Rest): The percentage of pigs lying in the resting area was recorded daily.
- % Lying (Activity): The percentage of pigs lying in the activity area was recorded daily.
- % Lying (Excreting): The percentage of pigs lying in the excreting area was recorded daily.

Observations were made during morning (06:00-08:00) and evening (12:00-14:00) times with an image sampled every 10 minutes.

2.3.5.5 Additional Category Variables

A corresponding control pen was identified for every pen that experienced a fouling event. This control pen had the same treatment and physiological conditions as the fouling pen but did not register any fouling event, thereby serving as a benchmark for comparative study. The variables used are:

- Foul Pen (FOUL): Pens with a recorded fouling event.
- Control Pen (NO FOUL): Pens with no recorded fouling event chosen as a control.

2.3.5.6 Output Variables

The outcome variables were binary and recorded daily to identify welfare events in the pig pens. All the observations were made according to a set protocol by

Lyderik et al. (2016). The observation protocol involves daily monitoring of Diarrhoea, Tail Biting, and Fouling in the pens. The output variables are defined as:

- Fouling: This variable denotes whether a fouling event occured on any day. A fouling event is characterised by over half of the lying area (solid floor) being covered with excreta and/or urine.
- Diarrhoea: This variable denotes whether any pig showed signs of diarrhoea on any given day. The presence of diarrhoea was recorded through daily visual inspections of the pen from the outside, with a positive record being made when at least one instance of faeces with a liquid or runny consistency was spotted.
- Tail Biting: This variable denotes whether any incidence of tail biting was observed on any given day. An occurrence of tail biting was recorded when at least one pig was visible in the pen with fresh blood on its tail.

2.3.6 Variable Usage Across Different Chapters

Table 2.1: Variable Utilisation Across Research Chapters.

Variable Name	Ch. 2	Ch. 3	Ch. 4	Ch. 5	Ch. 6
Temperature (Solid Floor)	✓	✓	✓	✓	✓
Temperature (Slatted Floor)	✓	✓	✓	✓	✓
Water Consumption (Drinker 1)	✓	✓	✓	✓	✓
Water Consumption (Drinker 2)	✓	✓	✓	✓	✓
Relative Humidity (Finisher Unit)			✓	✓	✓
Ventilation Output (Finisher Unit)			✓	✓	✓
Heating Output (Finisher Unit)			✓		
Cooling Output (Finisher Unit)			✓		
Temperature (Finisher Unit)		✓	✓	✓	✓
Age of Pigs		✓	✓		✓
Number of Pigs		✓	✓		✓
Straw		✓	✓		✓
Tail Type		✓	✓		✓
% Lying (Rest)		✓			✓
% Lying (Activity)		✓			✓
% Lying (Excreting)		✓			✓
Foul Pen (FOUL)		✓			
Control Pen (NO FOUL)		✓			
Fouling	✓	✓	✓	✓	✓
Diarrhoea			✓	✓	
Tail Biting	✓		✓	✓	

Table 2.1 provides an overview of the collected variables used across different chapters. Some variables were employed in multiple chapters, while others were used for analyses.

In addition to the variables, Table 2.2 summarises the data recorded for each batch in the study, providing an account of the number of pens, samples, fouling instances, tail-biting incidents, and diarrhoea cases.

Table 2.2: Summary of Data recorded for each batch.

Batch	Section	No. of Pens	Total Days	Fouling	Tail Biting	Diarrhoea
1	1	16	944	288	6	52
1	2	16	944	182	16	80
2	1	16	1072	158	7	35
3	1	16	1040	76	18	59
3	2	16	1040	84	14	72
4	1	16	1296	175	9	8
4	2	16	1296	163	8	10

In the context of this chapter, the availability and utility of data were unfortunately hampered by technical challenges. Specifically, only the data from one batch (Round 3) was used in this chapter. This limitation was due to missing data in the data of the other batches, which was primarily the result of recurrent sensor failures. Consequently, the model development in this study will rely solely on data procured from Round 3. Round four data was not included in this study as the data was not available at the time the analysis was performed. The algorithm used in this chapter does not handle corrupt/missing data.

2.4 Theory and Equations

In this chapter outlined in Figure 2.3, the study employs the concept of diurnal frequencies - patterns recurring over 24 hours - a phenomenon prevalent in the behaviour of pigs (Bigelow & Houpt, 1988; Bornett et al., 2000; Jarissa et al., 2017). To remove noise in the data, bandpass filters were used. These bandpass filters selectively allow frequencies that align within a specified range - in this scenario, the diurnal frequencies - to pass through while obstructing the rest. Following the filtration process, dynamic models are constructed from the resulting data. These models effectively serve as mathematical representations of the ongoing activities within the pig pens. They represent the evolving patterns of behaviour that are continually observed.

It is important to note that these dynamic models are not static; they are designed to update and refine themselves as new data is received continually. This continuous process of adjusting and enhancing the models with new inputs is known as recursive identification. This ensures that the models remain updated, accurate and effectively represent the real-time behaviour observed within the pens.

Finally, these models are consistently monitored for any deviations or abnormal behaviour. Within the context of this study, ‘abnormal’ is defined as any significant departure from the patterns that the dynamic models represent.

In summary, by integrating the principles of diurnal frequencies, bandpass filters, recursive identification, and consistent monitoring, this methodology provides a robust platform for analysing and comprehending the behavioural patterns within the pig pens. Furthermore, it equips the researchers to swiftly detect abnormal behaviour, facilitating proactive responses to changes within the pens.

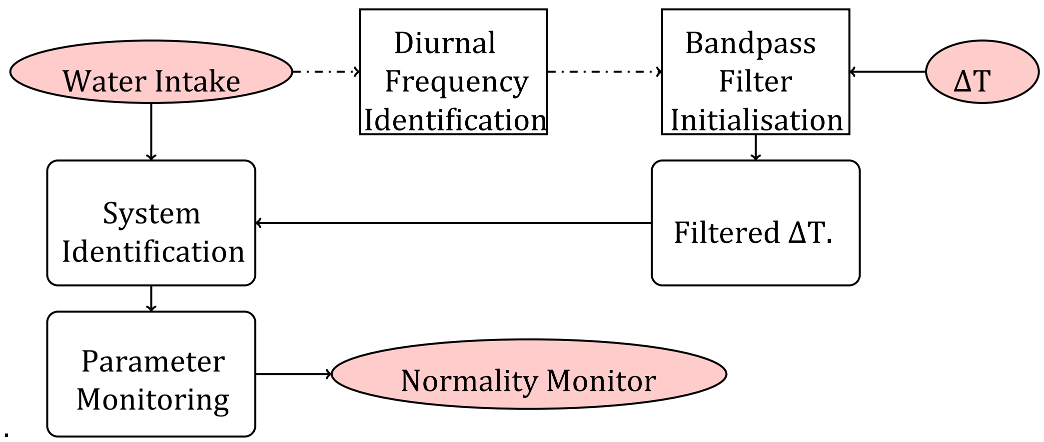

Figure 2.3: Schematic representation of the modelling approach utilised in this research. The process begins with collecting raw temperature data from two probes within each pen, which are T1 and T2. This data is used to calculate the difference in temperature between the two points (ΔT), which is used as a proxy for the pigs’ distribution within the pen. Concurrently, water intake data, denoted as w(t), is also recorded, and used to build the dynamic model through system identification and parameter monitoring.

2.4.1 Diurnal Frequency Identification using Discrete Fourier Transform (DFT)

The study uses the observation of water intake, labelled as w(t), and the data from two temperature probes (T1 and T2) situated in each pen. To simplify data analysis, these recorded measurements were averaged on an hourly basis. This approach enabled the streamlining of the data, making pattern recognition over time significantly more manageable. The averaging of the data each hour ensured the necessary detail to track daily behavioural and environmental cycles without losing valuable information. Additionally, averaging in time series reduces noise in the data, making the system identification process to converge when finding an optimal solution.

In this experiment, the difference in readings between the two temperature probes was used, designated as ΔT. Instead of the direct temperature readings, the rationale for utilising ΔT lies in its capacity to nullify potential influences from the pens' climate control systems. This is because any consistent temperature changes caused by the climate control system would impact both probes equally, meaning the difference, or ΔT, would remain unaffected. As a result, this provided a more accurate behavioural analysis of the pigs, detached from environmental temperature variations.

Moreover, ΔT was critical in revealing the animals' locations within the pen. The reason is that the temperature within each region where the probes were placed would shift based on the pigs' presence. The area inhabited by the pigs would naturally record a higher temperature due to their body heat. Thus, tracking ΔT made it possible to determine the pigs' current location in the pen.

Each data set in this study includes N samples representing a single pen. Data was sampled every hour. The primary frequencies present in the water intake, denoted as w(t), can be identified using the Fast Fourier Transform (FFT) technique.

The FFT is a computational tool that dissects a complex signal into constituent frequencies or harmonics. In other words, it breaks down the w(t) signal into a series of simpler sinusoidal signals, each characterised by a unique frequency Madsen et al. (2005). The water consumption signal, w(t), is not a singular entity but a composite of infinite sinusoidal signals. Each of these signals has its unique frequency and phase. The Discrete Fourier Transform, a form of FFT, provides a means to analyse the frequency spectrum of the w(t) signal. Through this analysis, it is possible to determine the key frequencies that make up the water consumption signal, w(t), thereby revealing the primary rhythms of water intake. The Fourier transform of the w(t) is defined as:

$$\mathcal{FT}\{w(t)\} = W_{ft}(\omega) = \int_{-\infty}^{\infty} w(t)e^{-2\pi ft}dt \qquad \text{Equation 2.1}$$

The outcome of this transform is a function in terms of frequency, represented by ω. Wft(ω) quantifies the power or intensity of the water intake function, w(t), at a specific frequency ω. In this context, the notion of ‘frequency’ can be thought of as the rate at which a specific pattern in the water intake data repeats itself over time. In other words, it refers to the regularity of specific behaviours in the water consumption of the pigs.

However, because the water intake signal is both discrete (i.e., it consists of distinct, separate values) and of finite duration (encompassing a specific number of samples, N), a version of the Fourier Transform tailored for such data, known as the Discrete Fourier Transform (DFT) was employed. The DFT is computed over a finite range of frequencies called the sampled frequencies ωk’s where:

$$\omega_k = \frac{2\pi}{N}\mathrm{k}, \quad \text{where} \quad k = 0,1,2,\ldots,\mathrm{N}-1. \qquad \text{Equation 2.2}$$

The Discrete Fourier Transform (DFT) of a sequence of N complex numbers $w(t), t = 0,\ldots,N-1$ is another sequence of N complex numbers $W(\omega_k), k = 0,\ldots,N-1$, defined by the formula:

$$W(\omega_k) = \sum_{\{N=0\}}^{\{N-1\}} w(t) * e^{-i\omega_k t} \qquad \text{Equation 2.3}$$

The summation is over t from 0 to N-1 (inclusive), and i is the imaginary unit, which satisfies the equation $i^2 = -1$. The exponential function $e^{\{-i\omega_k t\}}$ represents a complex exponential, a mathematical construct used to facilitate calculations involving both amplitude and phase. In this context, ‘n’ signifies the nth instance within a data set that consists of ‘N’ observations or samples. These observations are collected over a given period - in this case, each hourly interval within each pen. Therefore, each ‘n’ represents an hourly water intake and temperature observation within a single pen. The full dataset ‘N’ incorporates all these hourly observations made over the study period. It can be shown (Sundararajan, 2001) that the DFT of a signal at a frequency ω can be written in the complex plane as:

$$WDft(\omega k) = a + bj \qquad \text{Equation 2.4}$$

In this equation, ‘a’ and ‘b’ represent the real and imaginary components of the DFT at the specific frequency ωk. This enables the calculation of the magnitude and phase of the water intake signal, w(t), at the different sampled frequencies, ωk.

DFTs are valuable tools as they reveal patterns (or ‘periodicities’) within the input signal and quantify the relative influence of individual components at these specific frequencies, ωk. Numerous resources are available in the scientific literature (Oppenheim & Schafer, 2014; Sundararajan, 2001) for a more comprehensive understanding of DFT.

Using the Fourier Transform, the water intake signals were deconstructed into their various frequency components and dynamically identify the primary harmonics that define the pigs' drinking patterns. This study used the Fast Fourier Transform (FFT), a more efficient algorithm than the DFT.Compared to the standard DFT computation, FFT reduces the necessary calculations for a dataset of N points from $2N^2$ to $2NlogN$ (Cooley & Tukey, 1964). This enhanced efficiency makes FFT a preferred choice, especially for larger datasets like the one used in this study.

2.4.2 Recursive detection of the main diurnal frequencies

Two bandpass filters were recursively updated only to allow a range of frequencies corresponding to the main diurnal frequencies in the water data. A 20th Butterworth bandpass filter was used for this study; the filter was initialised in MATLAB using MATLAB's designfilt function. The filters were recursively updated daily using data from a backward window of 7 days (as shown in Figure 2.4). Constantly updating the filter with the new frequency range minimises information lost in the filtered data. An illustration of two bandpass filters implemented on the water signal is shown in Figure 2.5a and Figure 2.5b.

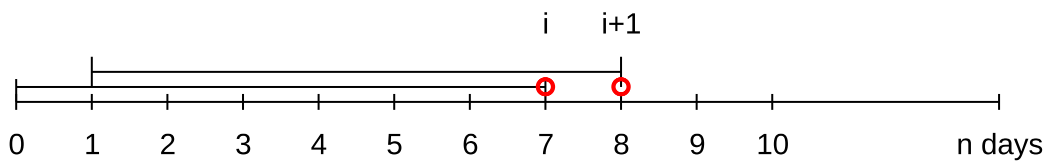

Figure 2.4: This picture shows the timeline used to estimate new frequency ranges recursively. On the i[th] day, two samples from seven days before are used to estimate the new diurnal frequencies and update the bandpass filter with the newly identified frequency range.

In this research, a "window" refers to a specific frame for which the data is analysed: seven days. A "sample" denotes a single observation within the data. Since this study deals with time-series data, each sample corresponds to a specific moment when water consumption data was recorded. Instead of analysing this whole data simultaneously, it is broken down into smaller, more manageable segments or "windows". Each window in the analysis includes seven days, equating to 168 samples because data is recorded every hour (24 hours/day * 7 days = 168 samples).

For each window, the bandpass filters were initialised by minimising the variance in error between the filtered and the original data. The range of frequencies is defined as having centre frequency ωBPc with a lower frequency and upper frequency bound of ωBPR1 and ωBPR2 respectively, where:

$$\omega BPR1 = \omega BPc - R1 \qquad \text{Equation 2.5}$$

$$\omega BPR2 = \omega BPc + R2 \qquad \text{Equation 2.6}$$

R1 and R2 are constants. A range of bandpass filters are simulated with varying ωBPc. The band pass filter with the least variance of the error signal is chosen. Once the bandpass filters are identified, the ΔT signal is passed through the filter. ΔT is inherently a very noisy signal. Filtering out ΔT at the diurnal frequencies only allows meaningful information to be retained from the ΔT signal.

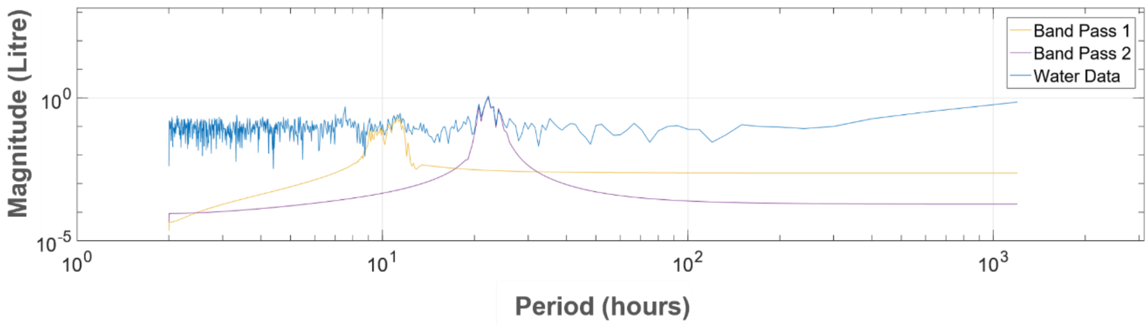

Figure 2.5a: Frequency decomposition for one week of water data. The effect of two bandpass filters is also shown.

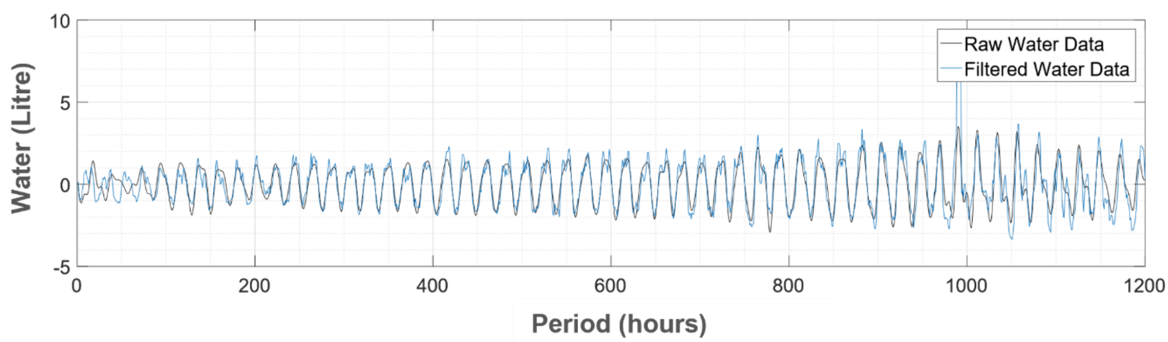

Figure 2.5b: Water Signal with Bandpass filter applied.

2.4.3 Identifying System Relationships

A transfer function model was used to identify the relationship between the filtered water wf(t) and temperature difference time signal $\Delta T_f\ (s)$. Transfer functions $(G(S))$ models are helpful tools to describe the dynamic relationship between a pair of input ($W_{f(s)}$) and output ($\Delta T_f\ (s)$):

$$\mathrm{G}(s) = \frac{\Delta T_f(s)}{w_f(s)}$$ Equation 2.7

$W_{f(s)}$ and $\Delta T_f\ (s)$ are the filtered water and temperature difference signals transformed in the Laplace domain. The polynomial form of the transfer function can be represented as:

$$G(s) = \frac{s^{nz} + a_1 s^{nz-1} + a_2 s^{nz-2} + \ldots k1}{s^{np} + b_1 s^{np-1} + b_2 s^{np-2} + \ldots k2}$$ Equation 2.8

Where np and nz are the numbers of poles and zeros of the transfer functions. a, b and k constants, which are identified using MATLAB. These parameters are identified using the default MATLAB setting: the instrument variable (IV) method as described by Young & Jakeman (1979).

The numerator and denominator of the transfer function in equation 2.8 can be factorised. The denominator of the transfer functions is called the characteristic equation and contains the underlying dynamics of the relationship between the input and output of the system. Poles of the transfer functions are obtained by factorising the characteristic equation of the transfer function. Based on the chosen nz^{th} order of the transfer function, several modes can be extracted from the characteristic's equation.

A mode of a transfer function is defined as a complex conjugate pair of poles, i.e.:

$$\mathrm{M_k} = (\mathrm{a} + \mathrm{bj})(\mathrm{a} - \mathrm{bj})$$ Equation 2.9

where a is the real and b is the imaginary part of the kth mode of the system. A graphical explanation of the pole placement is given by Figure 2.6. The properties of the mode of the transfer function are as follows:

$$Natural\ Frequency\ (\omega_n) = \sqrt{a^2 + b^2}$$ Equation 2.10

$$Damping\ ratio\ (\zeta) = -\frac{a}{\sqrt{a^2 + b^2}}$$ Equation 2.11

$$Resonant\ Frequency\ (\omega_r) = \omega_n * \sqrt{1 - 2\zeta^2}$$ Equation 2.12

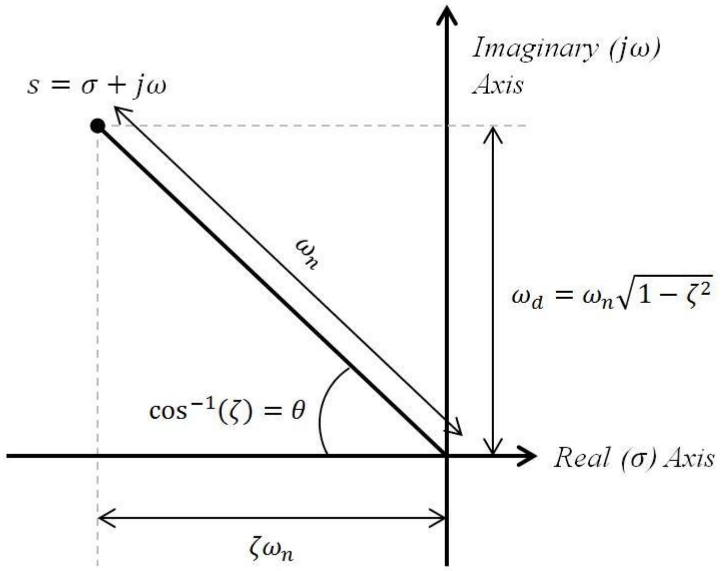

Figure 2.6: Significance of pole placement.

2.5 Results and Discussion

2.5.1 Frequency Decomposition of water data

To gain insight into the water data, spectral plots for experimental rounds 1, 2 and 3 were obtained using the FFT, see Figures 2.7, 2.8 and 2.9. A summary of the observed peaks in the spectral plots is given in Table 2.2. Four dominant peaks can be observed in the FFT of the water signal. Table 2.2 shows the frequencies at which the primary harmonics in the water signals occur. Furthermore, similar results were obtained by Madsen and Kristensen (2005).

The four harmonics in the water data have periods of about 6, 8, 12 and 23 hours. From the frequency responses (Figures 2.7, 2.8 and 2.9), it can be observed that there is a fifth harmonic, consisting of a large period (>100 hours). These harmonics can be interpreted as a quasi-linear increase in water consumption in the pigs as they grow. This slow increase in water consumption is not of interest in this study, as it provides no information on the diurnal pattern. However, it could be used in other studies to model growth in pigs.

The harmonics observed using the FFT confirm the finding by Madsen and Kristensen (2005). It was reported that a dynamic linear model composed of four harmonics gave the optimum performance in simulating water intake in pigs. The harmonics can be regrouped into two explanatory sets. The first set is the 23-hour harmonic, which could explain the pigs' main daily feeding pattern. The second set, composed of the other three harmonics, could be a result of daily adjustments in drinking and feeding.

Table 2.2: Periods and magnitudes of the four primary harmonics constituting the water signal in two sections for each of the three experimental rounds. Each trial corresponds to an experimental round, and each peak corresponds to one of the primary harmonics. The periods are measured in hours, and the magnitudes are measured in litres. For each section and peak, the period and magnitude values indicate when and how strongly the harmonic appears in the water signal.

		Section 1		**Section 2**	
	Peak	**Period (hrs)**	**Magnitude (L)**	**Period (hrs)**	**Magnitude (L)**
Round one	Peak one	5.9	0.22	5.7	0.10
	Peak two	7.7	0.43	7.5	0.31
	Peak three	11.6	0.21	11.5	0.20
	Peak four	22.9	1.49	21.9	1.09
Round two	Peak one	n/a	n/a	5.9	0.28
	Peak two	n/a	n/a	7.8	0.56
	Peak three	n/a	n/a	11.7	0.32
	Peak four	n/a	n/a	22.9	1.52
Round three	Peak one	5.9	0.17	5.9	0.19
	Peak two	7.5	0.27	7.8	0.36
	Peak three	11.3	1.9	11.6	0.22
	Peak four	22.4	0.70	22.4	0.98

Although the frequencies of the harmonics reported in this study are coherent with what was reported in the literature (Madsen & Kristensen, 2005), both studies were conducted on a conventional Danish farm setting, hence a very similar housing structure. Hence when modelling the water intake, adjustments in the four harmonic frequencies might be needed, and the FFT has proven beneficial.

The water signal in a normal state should show a stable diurnal pattern centred around the four harmonic frequencies identified in this study. Hence by monitoring for this expected dynamic in the water data, it can be expected that the chance of an outbreak of welfare issues would increase should the pig's drinking behaviour deviates from normality, that is, the expected pig behaviour.

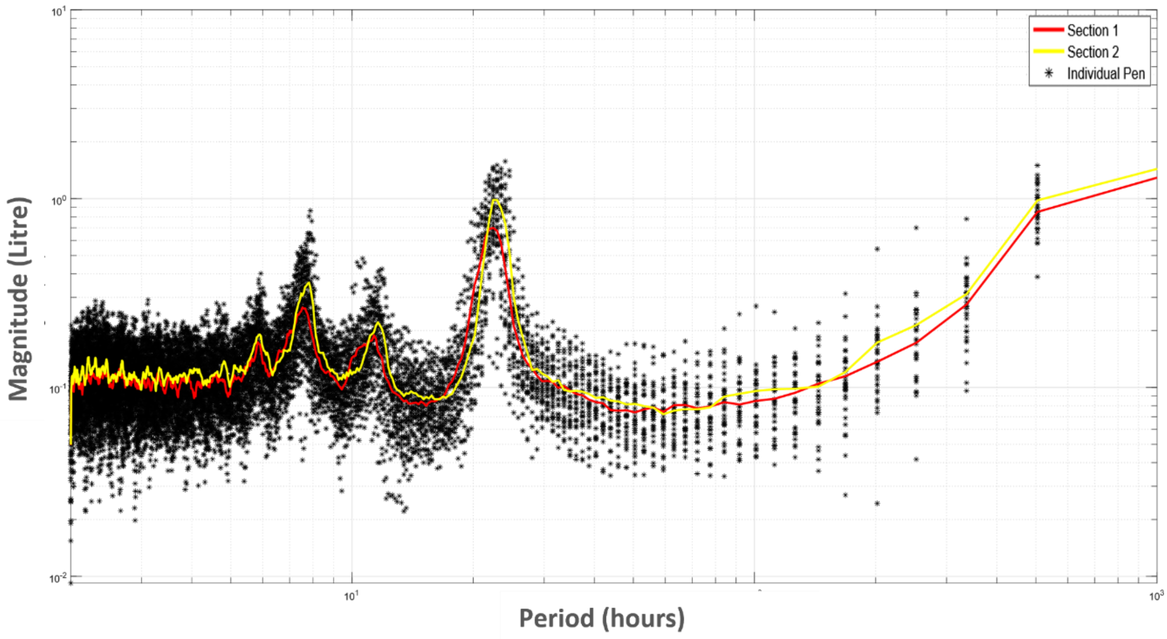

Figure 2.7: Spectral analysis of water intake data for each pen from Round 1. The averaged frequency response for each section 1 (red) and section 2 (yellow) are also shown in the figure.

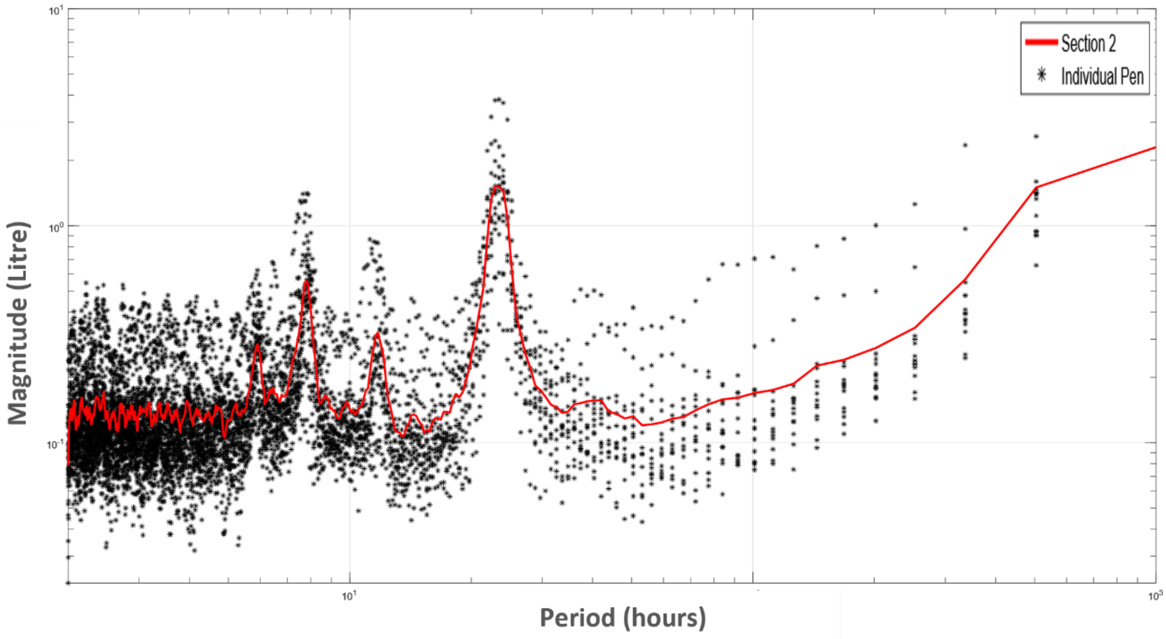

Figure 2.8: Spectral analysis of water intake data for each pen from Round 2. The averaged frequency response for each section 1 (red) and section 2 (yellow) are also shown in the figure.

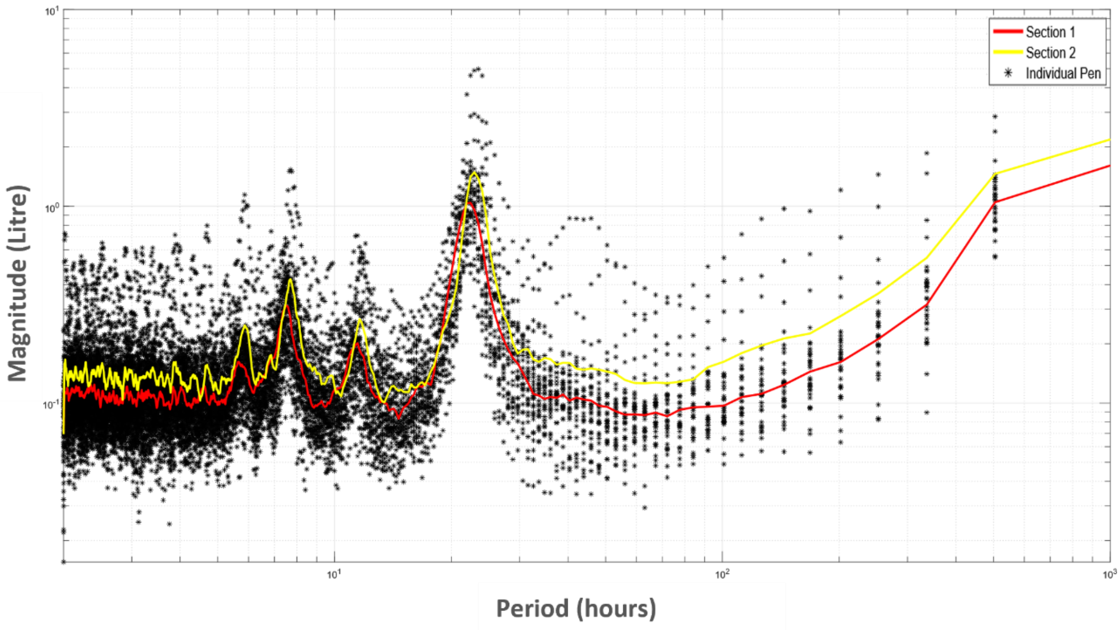

Figure 2.9: Spectral analysis of water intake data for each pen from Round 3. The averaged frequency response for each section 1 (red) and section 2 (yellow) are also shown in the figure.

2.5.2 Frequency filtering

It has been shown by Bigelow and Houpt (1988) that 75% of the water bouts were closely related to eating bouts. Hence, it is reasonable to assume that the frequencies driving the water signal are related to the overall diurnal feeding frequencies. Using the diurnal frequencies to initialise a bandpass filter provides a robust filtering approach to remove all noise (dynamics not related to diurnal behaviour). The filtering approach proposed in this study can also be used to reduce noise in the water signal further.

To obtain signals with frequencies around the driving harmonics in the water signals, two bandpass filters were used: the first one to allow the dynamics to have a period of around 23 hours and the second one to allow the three smaller harmonics through, i.e., ranging from a period of 5 to 13 hours. An illustration of the two bandpass filters is given in Figure 2.10. Each of the filters was designed to allow signals within a range of an eight-hour window. The optimal window was obtained by minimising the error signal variance between the filtered and unfiltered water signals. The new bandpass filters were initialised daily based on seven days of prior data. A simulation was performed to obtain the percentage error between the original and filtered data (from the optimal filter). The error was computed for all the pens for three rounds. The mean percentage error of the filtering algorithm was calculated to be 0.030% with a standard deviation of

1.51%. Figure 2.14 shows the distribution of the mean percentage error. The filtering algorithm successfully removes noise in the water data by effectively retaining only the four primary harmonics. This robust performance in noise reduction validates the efficacy of the filtering approach in enhancing the accuracy of the water data analysis.

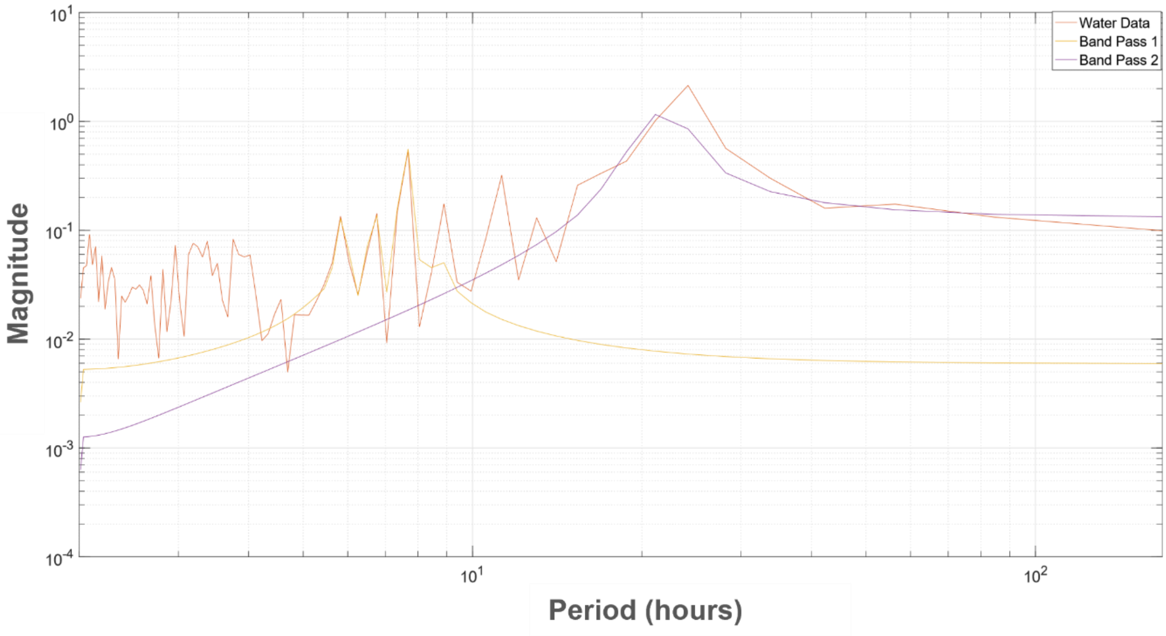

Figure 2.10: Frequency decomposition for one week of water data. The Effects of two bandpass filters are also shown.

2.5.3 Relationship Modelling

Two main components are needed to build a transfer function model: an input signal and an output signal. In this context, the filtered water signal wf(t) is the input, while the filtered zonal temperature difference at the diurnal frequencies, ∆Tf(t), acts as the output. These filtered signals are used to capture the key frequencies that constitute the diurnal rhythms in water consumption and temperature difference, eliminating unnecessary noise and facilitating a more precise system identification.

System identification is a method in control engineering for building mathematical models of dynamic systems from measured data. In this scenario, the system identification algorithm in MATLAB is employed to generate the transfer function model, which is a mathematical representation of the relationship between the input and output of the system. The transfer function model maps how a change in water consumption (input) influences this pig population’s zonal temperature difference (output).

However, the structure of the transfer function model is not predetermined; it must be identified from the data. Different structures or orders of transfer function models can be tested to find the best-fitting model. The “order” here refers to the degree of the differential equation in the model, with higher-order models having more parameters and potentially capturing more complex dynamics.

To determine the optimal order for the transfer function model, various system identification criteria are used. By comparing these criteria across different orders of transfer function models, the most suitable model order can be determined, and the structure of the transfer function model can be finalised. This optimal model will best estimate the system’s dynamics, capturing the underlying relationship between the pigs’ water consumption and the zonal temperature difference. The different system identification criteria used to compare different order of transfer function models are:

1. Final Prediction Error (FPE), and
2. Mean Square Error (MSE).

The research results are shown in Table 2.3. For the optimal transfer function structure, the aim was to have the smallest FPE and MSE values while considering stability constraints. The FPE and MSE are measures of the model’s accuracy, with lower values indicating better accuracy. FPE reflects the error in the model’s predictions, with a smaller value suggesting that the model’s predictions are closer to the actual data. MSE is a measure of the average squared difference between the model’s predictions and the actual data; again, a smaller MSE indicates a more accurate model.

Only consider models with an even number of poles were considered to ensure that the model is identifiable. This is done so that conjugate pairs of poles (two complex poles that are reflections of each other across the real axis in the complex plane) can be identified, which is essential in understanding the oscillatory behaviour of the system. Furthermore, the model must have fewer zeros than poles for dynamic stability - an important system characteristic that ensures it will not produce unexpected or extreme responses. A system with more poles than zeros naturally tend to dampen out oscillations, ensuring stable behaviour.

Upon examination of Table 2.3, it becomes evident that the optimal structure for the transfer function is a sixth-order model with two zeros. Higher-order models were not pursued as they could result in excessively complex dynamics. Although the sixth-order model presented optimal results, a fourth-order model with two zeros was chosen to reduce the complexity of the dynamic model identified. The minor differences observed between the 4th and 6th-order models supported the choice. Therefore, the selected model structure is as follows:

$$\frac{\Delta T(s)}{w(s)} = \frac{(s+z_1)(s+z2)}{(s\pm p_1)(s\pm p2)} \qquad \text{Equation 2.13}$$

Table 2.3: System Identification Criteria for Transfer Function models of different orders. The table shows the number of zeros and poles for each order and corresponding values, Final Prediction Error (FPE), and Mean Squared Error (MSE). These metrics provide a measure of the quality of fit of each model.

Model	Zeros	Poles	FPE	MSE
1	1	2	0.04850	0.04795
2	1	4	0.04859	0.04792
3	1	6	0.04879	0.0479
4	1	8	0.04964	0.04861
5	2	2	0.04815	0.04734
6	2	4	0.03882	0.03809
7	2	6	0.03605	0.03513
8	2	8	0.03792	0.03669
9	4	6	0.03571	0.03467

Where $\Delta(T(s))$ is the zonal temperature difference and w(s) is the water consumption in the Laplace domain.

The pole locations for each day were calculated, and an average was taken across all the pens. These averages are graphically represented in Figure 2.12. Upon visual analysis of this figure, it was observed that the two modes of the transfer function could be separated into two distinct regions within the complex plane. The characteristics of these modes are discernible based on their positions on the complex plane, as represented in Figure 2.12. A closer look at Figure 2.12 reveals that the identified modes can indeed be categorised into two distinct regions.

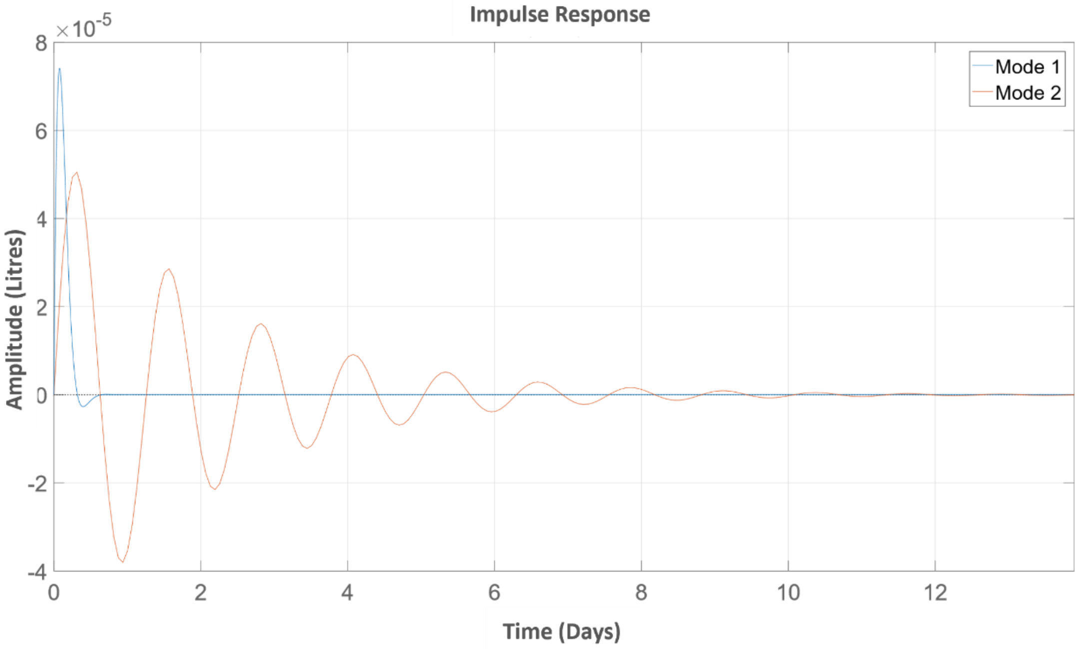

Figure 2.11: Illustration of the impulse response of the modes.

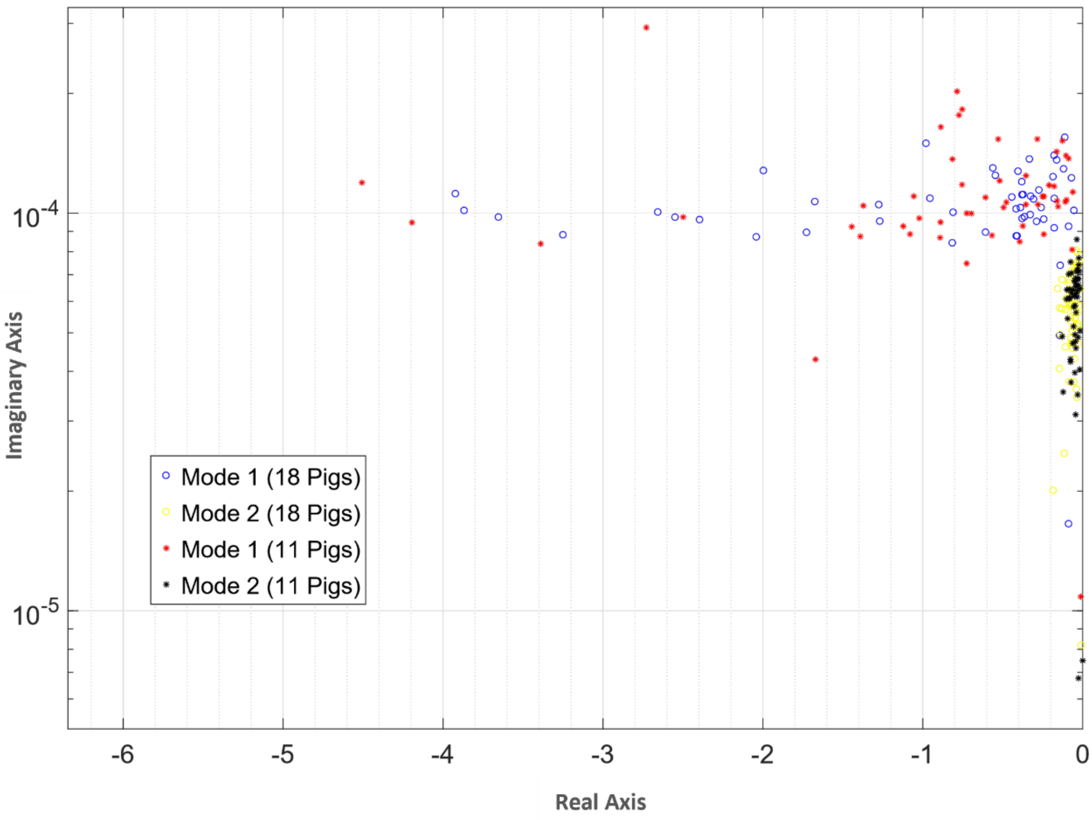

Figure 2.12: Poles and Zero maps of identified modes.

To substantiate the observed groupings, a k-means clustering algorithm was used. This unsupervised machine learning technique partitions a set of n observations into k clusters, such that each observation is associated with the cluster with the nearest mean, also known as the centroid. In this context, k-means clustering was leveraged to segregate the pole locations into distinct clusters, guided by their locations in the complex plane. The optimal number of clusters, k, was determined through the Elbow Method, as shown in Figure 2.13.

The Elbow plot clearly indicated a distinct “elbow” point at k-2, suggesting that the pole locations can be most effectively categorised into two distinct clusters. This data-driven approach provided a robust basis for the observations, reinforcing the notion that the transfer function’s two modes could be differentiated into two distinct regions within the complex plane.

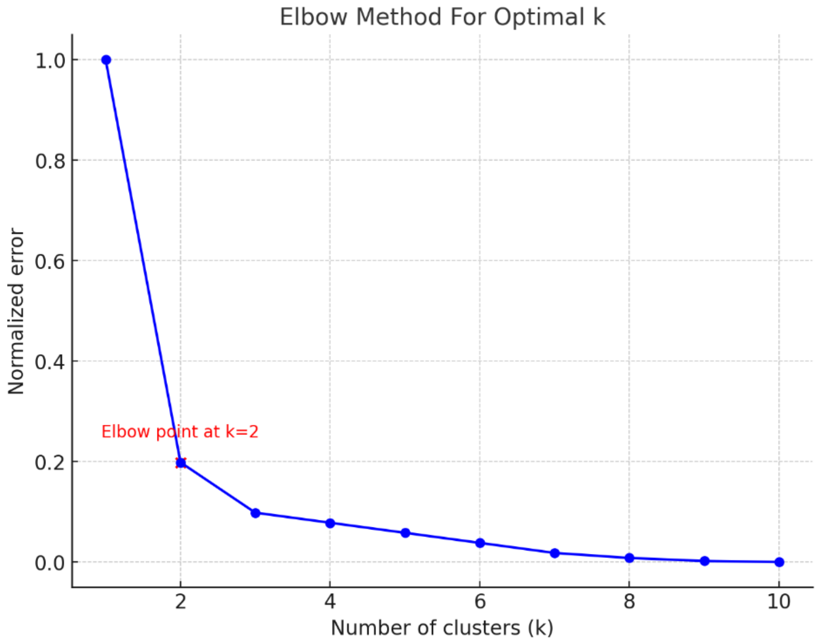

Figure 2.13: Elbow Method plot for determining the optimal number of clusters in k-means clustering. The plot shows the normalised error (within-cluster sum of squares) against the number of clusters (k). A clear ‘elbow’ point is observed at k=2 (highlighted in red), indicating that this dataset’s optimal number of clusters.

By employing this method, the two distinct regions were confirmed, and the mean locations for the two characteristic modes of the systems were precisely determined. This data-driven approach, utilising k-means clustering, provides a more robust confirmation of the observations made through visual inspection, ensuring a sound basis for the findings. The mean location for the two characteristic modes of the systems is given by:

$$\text{Mode 1: } (s \pm p1) \qquad \text{Equation 2.14}$$

$$\text{Mode 2: } (s \pm p1) \qquad \text{Equation 2.15}$$

Figure 2.12 shows the average pole placement for each throughout the whole experiment. The two modes can be used to describe two different dynamics:

1. Mode 1: A fast-moving mode (0-3 hours)
2. Mode 2: A slow mowing mode (4-7 hours)

The average location and the standard deviations of each of the modes for experimental round three are given in Table 2.4. The features of these poles are listed in Table 2.3. An illustration of the impulse responses of the two modes is given in Figure 2.11.

The first mode (Mode 1) can be used to characterise periods that are not related to their feeding pattern. That is, the animals move to the drinking station driven by their need to hydrate. This mode can be caused by animal socialising behaviour that triggers thirst. The pens with eleven pigs had faster periods than those with eighteen pigs. This difference could be the result of reduced accessibility of water stations. Hence, pigs tend to spend more time drinking in larger groups. Mode 1 is also characterised by a high damping ratio (between 0.6 to 1). A high damping ratio signifies that these modes are not oscillatory. Mode 1 represents an animal moving towards the drinking area for water and then returning to the resting area to rest. The faster mode can be used to characterise individual activities. The small undershoot in the fast response can be due to the increase in movement because of drinking.

The second and slower mode (Mode 2) can result from the animals' diurnal feeding pattern. Water and feed intake have a clear relationship (Bigelow & Houpt, 1988; Yang et al., 1981). Pigs tend to consume a large amount of water before feeding. Hence, these slower modes describe the daily needs for feeding. The mode with a period of 4-6 hours (as seen in Table 2.3) can be interpreted as follows: the pigs move towards the feeding and dunging area during their feeding period; this thus increases the temperature in the feeding and dunging area. Before feeding, the animals return to their resting area with an elevated body temperature, which temporarily elevates the resting area's temperature. Mode 2 is also observed to occur with a low damping ratio, which signifies that the mode is oscillatory; hence it would be the primary driving mode for a diurnal response.

Table 2.4: Characteristics of the mean modes for all experiments. The table shows the resonant frequency (ω_r), damping ratio (ζ), and natural frequency (ω_n) for two distinct modes in experiments with 18 and 11 pigs, respectively.

No of Pigs	Modes	ω_r (hr)	ζ	ω_n (hr)
18	Mode 1	1.68	2.44	0.73
18	Mode 2	4.84	4.46	0.091
11	Mode 1	0.51	2.69	0.98
11	Mode 2	4.97	5.01	0.12

Table 2.5: Mean values and standard deviations of the real and imaginary parts for the modes across all experiments. The table details the mean and standard deviation of the real and imaginary components of two different modes observed in the experiments.

	Real Part		Imaginary Part	
	Mean	Std	Mean	Std
Mode 1	-0.00012	0.00028	0.00011	0.000041
Mode 2	-0.0000052	0.0000028	0.000057	0.000016
Mode 1	-0.00053	0.0022	0.00010	0.000023
Mode 2	-0.0000068	0.0000038	0.0000055	0.000014

2.5.4 Event Detection

The present study introduces an event detection system to monitor the dynamics between poles to establish a frequency range defining 'normal' behaviour in pigs. This system predicts and flags potential fouling and tail-biting events, enhancing data accuracy. The section outlines the methodology, evaluates the alarm system's performance, and discusses potential improvement areas.

The methodology involved the identification of two modes, which demonstrated mean periods of 2.5 and 4.9 hours for their natural frequencies. For a preliminary trial, varying cut-off thresholds were set at 0.7, 0.8, and 0.9 σ (standard deviation). An alarm was set to trigger upon the occurrence of two consecutive pole movements that exceeded the set threshold within five days.

The event detection approach employed in this study is non-discriminatory, treating multiple instances of tail-biting and fouling collectively as a singular event, focusing on capturing general patterns of irregular behaviour rather than specific occurrences.

Throughout the 50-day study duration, data associated with these combined events were aggregated at the farm level. The recorded observations included fifteen events in pens housing eleven pigs and fourteen events in pens with eighteen pigs. The performance of the alarm system was evaluated using the AUROC (Area Under the Receiver Operating Characteristic) values, which are commonly used to assess the performance of binary classification models. The AUROC values obtained in this study for both datasets were approximately 0.145 for pens housing eleven pigs and 0.142 for pens housing eighteen pigs. These values are relatively low, suggesting that the model’s classification performance is not much better than random guessing.

Analysing the alarm system’s performance in Table 2.5 and Table 2.6, it is evident that setting a tighter threshold and quantifying a smaller percentile of extreme data as outliers leads to a reduction in sensitivity. Conversely, using a more relaxed threshold resulted in a higher number of false positives. The current detection method may not be optimal for event detection, as the literature offers better methodologies for setting alarms to detect events (Jensen, 2016). Previous research in the literature utilises a probabilistic approach to predict events, which is a valuable tool for observational data. However, this methodology lacks insights into the dynamics of behavioural models.

In contrast, the methodology presented in this paper for alarm detection adopts a mechanistic approach. Understanding the mechanics leading to abnormal behaviour makes it possible to develop methodologies to control these events. The current model incorporates water consumption as an input, but this may not be a controllable variable in an environment where water is distributed ad libitum. However, in more controlled environments with stricter control of feeding times, understanding the feeding and activity dynamics would significantly minimise animal stress, thereby increasing the system's welfare.

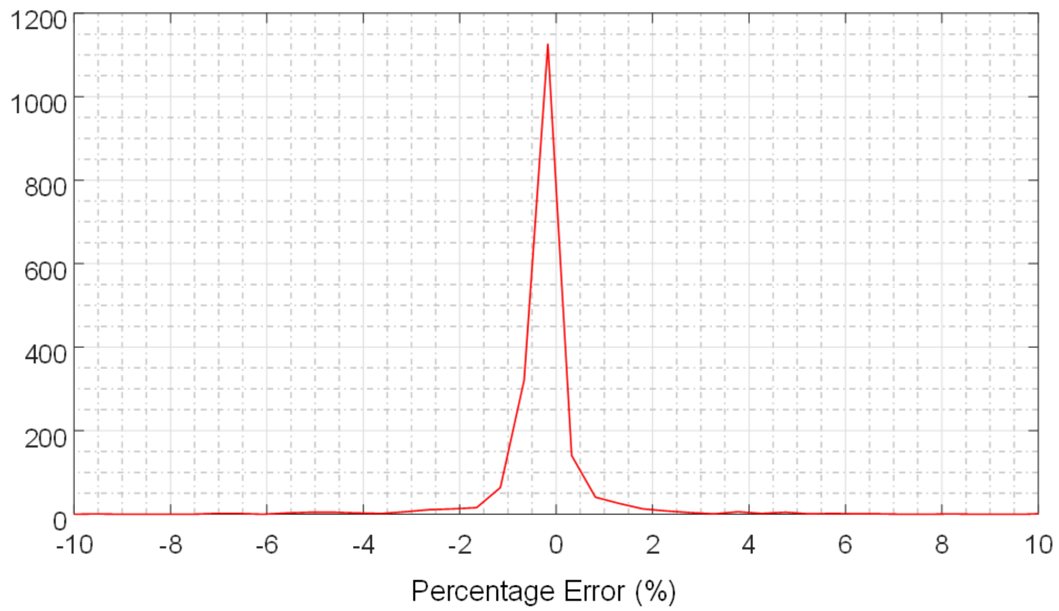

Figure 2.14: Gaussian distribution of mean percentage error between filtered and unfiltered water signal.

The relatively low AUROC values obtained could be attributed to several factors. Firstly, the non-discriminatory event detection approach might mask the distinctions between tail-biting and fouling events, reducing sensitivity in identifying specific behaviours. Secondly, the reliance on aggregated data at the farm level may have overlooked subtle variations in behaviour within individual pens, impacting the overall performance of the alarm system.

In conclusion, while the mechanistic approach to detecting events provides valuable insights into the dynamics of abnormal behaviour, further refinements may be necessary to improve the sensitivity and accuracy of the alarm system. Exploring alternative methodologies and incorporating more granular data at the pen level could enhance the system's ability to discriminate between different event types, leading to more effective event detection and management in pig farming practices.

Table 2.6: Event detection sensitivity for pens with eleven pigs.

Threshold (%)	False Positive	True Positive	False Negative	True Negative	Sensitivity
48 (0.7σ)	22	9	6	13	60%
54 (0.8σ)	14	8	7	21	53%
61 (0.9σ)	11	5	10	24	33%

Table 2.7: Event detection sensitivity for pens with eighteen pigs

Threshold (%)	False Positive	True Positive	False Negative	True Negative	Sensitivity
48 (0.7σ)	26	10	4	10	71%
54 (0.8σ)	23	9	5	13	64%
61 (0.9σ)	18	8	6	18	57%

2.6 Conclusions

This study developed a transfer function model to establish the relationship between water consumption and zonal temperature difference in a pig population. The optimal model structure was identified using system identification techniques, comparing various system identification criteria. The selected fourth-order transfer function model with two zeros exhibited an excellent fit to the data while maintaining model simplicity.

By conducting pole analysis and employing k-means clustering, two distinct modes in the system dynamics were found. Mode 1, characterised by a fast response period, captured the animals' movement towards drinking stations, likely influenced by socialising behaviour. Mode 2, with a slower response period, corresponded to the diurnal feeding pattern of the pigs.

The implemented event detection system used a non-discriminatory approach, capturing general patterns of irregular behaviour without distinguishing between tail-biting and fouling events. Despite obtaining relatively low AUROC values, the mechanistic approach provided valuable insights into abnormal behaviour dynamics. To enhance the sensitivity and accuracy of the alarm system, further research should explore alternative methodologies, including incorporating more granular data at the pen level, and consider adopting probabilistic approaches from the literature.

Overall, this study contributes to understanding pig behaviour dynamics and establishes a foundation for more effective event detection and management in pig farming practices. By gaining insights into the complex interactions between water consumption and zonal temperature difference, opportunities to optimise pig welfare and potentially reduce stress in controlled environments with stricter feeding schedules emerge.

In conclusion, this research serves as a steppingstone towards better animal welfare practices, emphasising the importance of employing data-driven approaches to unravel the intricacies of animal behaviour. Future studies stand to benefit from exploring more sophisticated models and integrating advanced machine-learning techniques for improved event detection and monitoring in pig farming systems. The findings underscore the significance of interdisciplinary approaches that bridge engineering, biology, and data analysis to advance the field of precision animal farming and enhance animal welfare.

Chapter 3: Can early changes in behaviour be used to predict fouling?

3.1 Abstract

This study aims to explore the feasibility of using early changes in behaviour as indicators of fouling in conventionally housed slaughter pigs. The research investigates whether specific factors, including temperature variations between resting and excreting areas, water consumption, and spatial positioning of the pigs, can reliably predict fouling events. The spatial positioning of pigs was monitored using video recordings, allowing for accurate observations of their behaviour. The pen was divided into three areas: the Activity area, the Excreting area, and the Rest area. The Activity, Excreting and Rest areas are covered with slatted, drained, and solid areas, respectively.

The study found that a reduction in the number of pigs lying in the Activity area occurred 5 and 4 days before fouling events, resulting in 8.1% ($p=0.02$) and 9.0% ($p=0.03$) fewer pigs in that area, respectively. In contrast, pens experiencing fouling had, on average, 14.0% ($p=<0.001$) more pigs in the Activity area, 2.8% ($p=0.003$) fewer pigs in the Excreting area, and 11.0% ($p<0.001$) fewer pigs in the Rest area.

Moreover, the temperature difference between the Rest and Excreting areas emerged as a more reliable fouling indicator. Fouled pens exhibited a smaller temperature difference of 0.52°C ($p<0.007$). This finding suggests that the spatial positioning of pigs significantly influenced the temperature in both areas. For every 10% increase in pigs in the activity area, there was a decrease of -0.2°C ($p=0.01$) and an increase of 0.62°C ($p<0.001$) in temperature in the rest and slatted areas, respectively.

In conclusion, this study highlights the potential of early behaviour changes, particularly spatial positioning, as valuable predictors of fouling in conventionally housed slaughter pigs. The results underscore the importance of monitoring these indicators to improve the management of pig fouling events, leading to better animal welfare and enhanced productivity. Further research could explore temperature variation as a proxy for assessing pig activity, streamlining the monitoring process for farmers and researchers alike.

Keywords: Fouling, Linear Regression, Pigs Behaviour

3.2 Introduction

Pen fouling is a complex issue influenced by multiple factors, including space allowance, floor design, thermal climate, and pig behaviour (Bertelsen et al., 2017). The space allowance impacts pigs' ability to access designated excreting areas and distinguish between rest and excreting spaces. In contrast, excessive space may reduce pig movement and increase excretion in designated areas. Additionally, pigs are susceptible to overheating as they are homeothermic animals. In response to high ambient temperatures, pigs seek colder spots to lie down, resulting in the disappearance of distinct activity areas at lower temperatures (Fraser, 1985). While factors like floor design and pigs' earlier experience remain constant during pigs' housing, they are not the focus of this study.

Although prior research has explored the impact of temperature (Spoolder et al., 2012), humidity (Huynh et al., 2005), and space allowance (Bench et al., 2013) on pen fouling and overall animal behaviour, it remains unclear why some pens experience fouling while others do not, despite similar physiological status and environmental conditions. A comparative analysis between the factors presents prior to in contrast to non-fouling pens will be carried out in this chapter. The previous chapter's findings, where water consumption and zonal temperature difference were investigated as indicators of pig spatial positioning, will be built upon.

In Chapter 2, system identification techniques and event detection systems were used to identify unique behavioural patterns associated with pig behaviours. Valuable insights into abnormal behaviour dynamics were provided. In this chapter, a deeper understanding of pig behaviours will be achieved by examining specific factors during prior to the occurrence of fouling events.

The primary objectives of this chapter are to:

- Investigate the influence of temperature variations between resting and excreting areas prior to fouling events.
- Examine the changes in water consumption patterns prior to a fouling event.
- Analyse the impact of pigs' spatial positioning prior to fouling events.

Valuable insights will be gained in this chapter, which will form the basis for future chapters where machine learning models will be explored. By key factors that affect fouling behaviour being identified, a reliable predictive framework will be created that uses advanced machine learning techniques to accurately forecast fouling events in traditionally housed slaughter pigs. The integration of machine learning models will enhance the ability to manage and optimize pig welfare in controlled environments, representing a significant advance in precision animal farming.

3.3 Materials and Methods

The experiment was conducted at the Finisher unit at Aarhus University, Denmark, from 2015 to 2016. A more detailed description of the experimental setup can be found in Chapter 2. The study included 112 pens housing 1624 slaughter pigs, divided into four batches (batch 1, 3, and 4: n = 32 pens each; batch 2: n = 16 pens). The pigs were inserted randomly into the pens with an average weight of 31.6 ± 6.6 kg. Each pen measured 5.45 m by 2.48 m and had the floor divided into solid concrete (rest area), drained (activity area), and slatted floor (excreting area). Artificial light was provided from 0530 to 1830 h (182 lx). During the daily check-up performed by the herd staff from 1000 to 1200 h, a protocol for scoring diarrhoea, fouling and tail-biting events was followed. As mentioned in chapter 2, fouling was defined as when more than half of the solid floor was wet with excreta and/or urine. These zones are arranged from left to right, as indicated by the blue markings that demarcate the borders between the different floor types. A schematic of the pen is given in Figure 2.1 and Figure 3.1.

Figure 3.1: Screenshot from the video observation made via a recorder video, capturing the arrangement of the three distinct zones within the pig pens. The zones are designated as the "Rest Area," "Activity Area," and "Excreting Area."

3.3.1 Behavioural Video Observations

For each recorded fouling event, video data was collected for five consecutive days leading up to the event, including the day of fouling (day (0)). The average percentage of pigs lying in each of the three distinct areas (% Lying (Rest), % Lying (Activity), and % Lying (Excreting)), as shown in Figure 3.1, was calculated for each of these days (day (-5), day (-4), day (-3), day (-2), day (-1), day (0)).

To ensure comprehensive data collection, video recordings were taken twice daily during the morning (06:00-08:00h) and evening (12:00-14:00h), with images sampled every 10 minutes. For every pen with a fouling event, a control pen with the same treatment and physiological condition but without any fouling event was selected. The video footage of the control pen was also observed for the same five days. The study labelled fouled pens as "FOUL," while the control pens were labelled as "NO FOUL." This meticulous approach allowed for a thorough assessment of the factors influencing fouling behaviour.

3.3.2 Sensor Observations.

This chapter used environmental sensor recordings to assess the pig pens' environmental conditions and physiological states. These sensor data points were integral to understanding the factors that might influence fouling events and pig behaviour. The key sensor recordings used in this chapter are summarised in Table 3.1. For a more comprehensive view of the sensors used throughout this research, refer to Table 2.1 in Chapter 2.

Table 3.1: Summary of Key Sensor Recordings

Sensor Recording	Description
Temperature (Rest Area)	VE10-A temperature probes by Veng System (VengSystem, Denmark) were placed at pig level in the resting area.
Temperature (Slatted Area)	VE10-A temperature probes by Veng System (VengSystem, Denmark) were placed at pig level in the excreting area.
Temperature (Section)	The climate control system records the average temperature of all pens within the finisher unit.
Water Intake	Water flow was measured using flow meters in each pen, providing insights into drinking behaviour.

3.4 Statistical Analysis

To analyse the influence of different environmental factors on the occurrence of fouling events, a combination of logistic and linear regression models was employed. All computations were performed using R version 3.2.4.

The presence of fouling was used as the binary result variable in logistic regression, and environmental factors such as temperature and water consumed per pig were considered as predictors. The regression models were constructed with a stepwise selection process only to retain noteworthy variables. Linear regression models were used to further explore the relationship between the significant predictors and fouling. In these models, DAY and FOUL were considered as factors, re-levelled around DAY (0) and NO FOUL, respectively. The stepAIC function was also used here for dimensionality reduction.

3.4.1 Logistic Regression Analysis (Fouling)

In this study, twenty-one distinct fouling outbreaks were identified. An outbreak is defined as the first observation of fouling within a specific batch or section, marking the beginning of a fouling issue. The 21 outbreaks analysed in this section were selected from a broader set of pens evaluated for fouling. Any data compromised by corrupted sensors or video were excluded. It's important to differentiate between these outbreaks and the total of 1,126 fouling incidents, which accounts for every individual fouling occurrence throughout the research.

For each of these twenty-one fouling outbreaks, corresponding observations were made in twenty-one other control pens. This pairing resulted in a temporal dataset for each outbreak event. This dataset encompassed six-time steps (days) leading up to the onset of the fouling outbreak, allowing for a detailed understanding of the conditions and factors that may have contributed to the initiation of fouling. The dataset incorporated several features: average temperature (Section, Rest, Excreting), water consumed per pig, and the percentage of pigs lying in different areas (Rest, Activity, Excreting). The water consumed per pig was calculated by dividing the total amount of water consumed in each pen by the number of pigs in that pen.

A mixed logistic regression model was employed to investigate these independent variables' impact on fouling. The mixed logistic regression model can be expressed as:

$$log\big(P(Y=1)\big) = \beta^0 + \beta^1 X^1 + \beta^2 X^2 + \ldots + \beta_{\mathrm{n}} * X_{\mathrm{n}} + \varepsilon \qquad \text{Equation 3.1}$$

Here:

- $P(Y=1)$ represents the probability of a fouling event happening.
- β_0 is the intercept, which captures the log odds of fouling when all predictors are zero.
- $\beta_1, \beta_2, \ldots, \beta_{\mathrm{n}}$ are the coefficients of the predictors X_1, X_2, X_{n} respectively,

indicating the change in log odds of fouling for a unit increase in these predictors.

- $X_1, X_2, \ldots, X_\mathrm{n}$ represent the predictors.
- b represents the random effects, i.e., variability at the pen level.
- ε is the error term, capturing the variability not explained by the predictors or random effects.

The stepAIC function in R was used, which implements a hybrid feature selection strategy backwards and forward feature elimination) to create an optimal model. This model included all significant predictors, disregarding the insignificant interaction terms.

Since the R-Square statistic cannot be computed for regression models with a categorical dependent variable, an approximation known as McFadden's R-Square was used. This measure offers a ratio of the log-likelihood of the full model to the log-likelihood of a model with no predictors, thereby estimating the proportion of variance in the dependent variable that is explained by the independent variables.

3.4.2 Linear Regression Analysis (Dependent Variables)

Linear regression models were used to analyse the significant predictors of fouling, with the relationships between dependent variables and predictors being examined. The mathematical representation of the linear regression models used in this study is as follows:

$$Y = \beta_0 + \beta_1 \boldsymbol{X}_1 + \boldsymbol{\beta}_2 X_2 + \ldots + \beta_\mathrm{n} X_\mathrm{n} + \varepsilon \qquad \text{Equation 3.2}$$

In this formulation:

- Y represents the dependent variable, which in this context could be temperature or percentage of pigs lying in a specific area.
- β_0 is the intercept, indicating the expected value of Y when all predictors are zero.
- $\beta_1, \beta_2, \ldots, \beta_\mathrm{n}$ are the coefficients of the predictors $X_1, X_2, \ldots, X_\mathrm{n}$, respectively. They measure the change in the dependent variable for a unit increase in the corresponding predictor, all other predictors being held constant.
- $X_1, X_2, \ldots, X_\mathrm{n}$ are the predictors or independent variables. In the models in this chapter, these represent the environmental factors and pig behaviour variables.

- ε is the error term, which accounts for the variability in the dependent variable that the predictors do not explain.

In these linear regression models, DAY and FOUL were incorporated as factors. The factors were adjusted to use DAY (0) and NO FOUL as the reference levels. The stepAIC method, as with the logistic regression model, was used here for dimensionality reduction, resulting in a model that only includes the predictors that significantly contribute to the explanation of the dependent variable's variance.

3.5 Results

In Table 3.1, the dependent variable's intercept was determined to be 9.418, with a 95% confidence interval (CI) ranging from 3.078 to 16.612. This result was found to be statistically significant (P-Value = 0.021). When examining the full model for Foul, the influence of different variables such as % Lying (Rest Area), % Lying (Activity Area), % Lying (Excreting Area), Temperature (Section), Temperature (Rest Area), Temperature (Excreting Area), and water consumed per pig (Litres) were found to have various levels of significance, with varying degrees of statistical significance. The model's fitness was indicated by McFadden's Pseudo R Square, which was 0.370. In the reduced model for Foul, significant variables included the intercept, % Lying (Activity Area), % Lying (Excreting Area), Temperature (Rest Area), Temperature (Excreting Area), and Water consumed per pig (Litres). McFadden's Pseudo R Square for the reduced model remained the same as the full model (0.370), indicating the reduced model's effective representation of the data.

Table 3.2 focused on the effect of temperature in different areas. Different variables such as % Lying (Rest Area), % Lying (Activity Area), % Lying (Excreting Area), and Temperature (Section) demonstrated varying degrees of influence on the temperature in the rest area, excreting area, and temperature difference. The Adjusted R Square, used to show the goodness of fit of the models in this study was 0.845, 0.822, and 0.688 for Temperature (Rest Area), Temperature (Excreting Area), and Temperature Difference, respectively. Figure 3.3 shows the relationship between the temperature difference and the % of pigs in the excreting area.

Table 3.3 shows the dependent variables (Excreting Area, Activity Area, and Rest Area) and their relationship with various variables such as Day and FOUL were explored. The adjusted R-Squares for the Excreting Area, Activity Area, and Rest Area models were 0.150, 0.210, and 0.130, respectively, suggesting how well these variables could explain the percentage of pigs lying in the respective areas. Figure 3.2 summarises the percentage of pigs lying in each area before fouling.

Table 3.1: Full and reduced order model for predicting the logarithmic odds of fouling ($\hat{\beta}$), 5% and 95% confidence correspond to logistic regression coefficients and p-values. McFadden's Pseudo R Square is used for an indication of model fit.

	Dependent Variable	$\hat{\boldsymbol{\beta}}$	**CI (5%)**	**CI (95%)**	**P-Value**
	Intercept	9.418	3.078	16.612	0.021
Full Model (Foul)	% Lying (Rest Area)	0.076	-0.166	0.007	0.148
	% Lying (Activity Area)	0.075	-0.007	0.163	0.013
	% Lying (Excreting Area)	0.010	-0.069	0.087	0.016
	Temperature (Section)	0.075	-0.279	0.426	0.723
	Temperature (Rest Area)	-0.947	-1.502	-0.452	0.003
	Temperature (Excreting Area)	0.547	0.150	0.959	0.025
	Water consumed per pig (Litres)	-0.254	-0.438	-0.084	0.017
	Pseudo R Square (McFadden)	0.370			
Reduced Model (Foul)	Intercept	9.827	5.241	15.064	**<0.001**
	% Lying (Activity Area)	0.063	0.024	0.109	**<0.001**
	% Lying (Excreting Area)	0.084	-0.135	-0.046	**<0.001**
	Temperature (Rest Area)	0.253	0.092	0.423	**<0.001**
	Temperature (Excreting Area)	-0.571	-0.730	-0.355	0.056
	Water consumed per pig (Litres)	-0.085	-0.146	-0.028	0.030
	Pseudo R Square (McFadden)	0.370			

Table 3.2: Reduced-order linear regression model for predicting Temperature in Rest, Excreting Area, and Temperature Difference. The coefficients ($\hat{\beta}$), indicates the change in °C for a unit increase in the dependent variable. The 5% and 95% confidence intervals (CI) for each of the coefficients and p-values are also shown.

Independent Variable	Dependent Variable	$\hat{\beta}$	CI (5%)	CI (95%)	P-Value
Temperature (Rest Area), °C	Intercept	***5.678***	**4.662**	**6.695**	**<0.001**
	% Lying (Rest Area)	***0.040***	**0.019**	**0.061**	**<0.001**
	% Lying (Activity Area)	*-0.020*	-0.029	-0.011	**<0.001**
	% Lying (Excreting Area)	*0.004*	-0.023	0.032	0.760
	Temperature (Section), °C	***0.816***	**0.754**	**0.878**	**<0.001**
	Adjusted R Square	0.845			
Temperature (Excreting Area), °C	Intercept	***-0.676***	**-2.911**	**1.560**	**0.552**
	% Lying (Rest Area)	***0.106***	**0.073**	**0.139**	**<0.001**
	% Lying (Activity Area)	*0.062*	0.032	0.092	**<0.001**
	% Lying (Excreting Area)	*0.025*	-0.008	0.057	0.132
	Temperature (Section), °C	***0.820***	**0.743**	**0.898**	**<0.001**
	Adjusted R Square	0.822			
Temperature Difference, °C	Intercept	***4.539***	**4.001**	**5.077**	**<0.001**
	% Lying (Activity Area)	*-0.516*	-0.895	-0.137	0.008
	% Lying (Excreting Area)	*-0.056*	-0.077	-0.035	**<0.001**
	No of Pigs	*-0.116*	-0.163	-0.069	**<0.001**
	Foul	*-0.051*	-0.062	-0.041	**<0.001**
	Adjusted R Square	0.688			

Table 3.3: Reduced order linear regression model for predicting the percentage of pigs lying in the Excreting, Activity and Rest Area. Temperatures were omitted from the reduced order mondel as they strongly correlate with activity within each zona. The model was level on the day (0), and no foul.

Dependent Variable	Dependent Variable	$\widehat{\beta}$	5% CI	95% CI	P-Value
% Lying (Excreting Area)	Intercept	19.060	16.524	22.517	**<0.001**
	Day (-5)	1.521	-2.548	5.788	0.504
	Day (-4)	3.077	-2.401	6.045	0.177
	Day (-3)	2.173	-2.473	6.377	0.358
	Day (-2)	1.022	-3.177	4.956	0.643
	Day (-1)	2.377	-1.723	6.507	0.286
	FOUL	-2.828	-7.384	-2.709	0.004
	Adjusted R-Square	0.150			
% Lying (Activity Area)	Intercept	16.395	10.747	22.811	**<0.001**
	Day (-5)	-8.102	-16.720	0.060	0.026
	Day (-4)	-8.992	-16.873	0.127	0.029
	Day (-3)	-5.868	-15.411	2.401	0.115
	Day (-2)	-5.411	-14.252	2.117	0.120
	Day (-1)	-0.404	-7.963	8.602	0.928
	FOUL	13.976	10.890	20.301	**<0.001**
	Adjusted R-Square	0.210			
Lying (Rest Area)	Intercept	47.957	43.060	54.023	**<0.001**
	Day (-5)	3.852	-4.117	11.132	0.340
	Day (-4)	3.595	-4.016	11.433	0.372
	Day (-3)	0.613	-6.585	9.602	0.883
	Day (-2)	3.043	-3.767	11.109	0.436
	Day (-1)	-2.467	-11.099	3.954	0.531
	FOUL	-10.952	-15.714	-7.161	**<0.001**
	Adjusted R-Square	0.130			

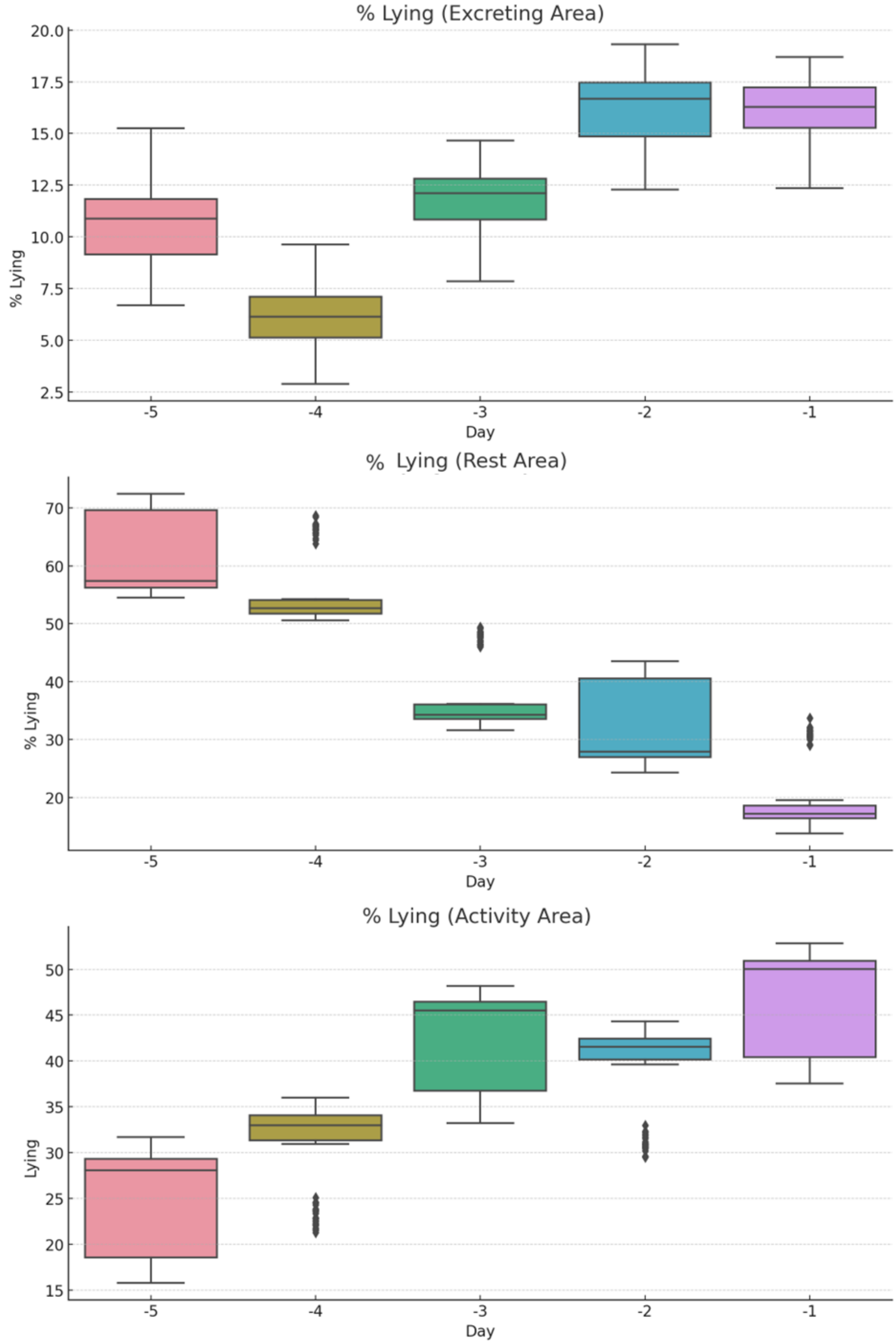

Figure 3.2: Box plots representing the distributions of the "% Lying" variables across five days for Pen where Fouling was observed. The variables are "% Lying (Excreting Area)", "% Lying (Rest Area)", and "Lying (Activity Area)", depicted from top to bottom respectively. Each box plot demonstrates the median (the horizontal line inside the box), interquartile range (the box itself), and minimum and maximum value.

Table 3.4 models the relationship between variables such as % Lying (Rest Area), % Lying (Activity Area), % Lying (Excreting Area), Temperature (Section), and Age, and the amount of water consumed per pig was analysed. Further details are to be presented in the following section.

Table 3.4: Reduced order linear regression model for predicting water consumed per pig (Litres). The coefficients ($\hat{\beta}$) indicate the change in Litres for a unit increase in the dependent variable. The 5% and 95% confidence intervals (CI) for each of the coefficients and p-values are also shown.

Dependent Variable	$\hat{\boldsymbol{\beta}}$	**5% CI**	**95% CI**	**P-Value**
Intercept	9.008	7.440	10.491	**<0.001**
% Lying (Rest Area)	0.025	0.063	0.134	0.018
% Lying (Activity Area)	0.028	0.022	0.032	**<0.001**
% Lying (Excreting Area)	-0.019	-0.168	0.020	**<0.001**
Temperature (Section)	-0.407	-0.380	-0.424	0.016
Age	0.160	0.124	0.207	**<0.001**
Adjusted R-Square	0.5455			

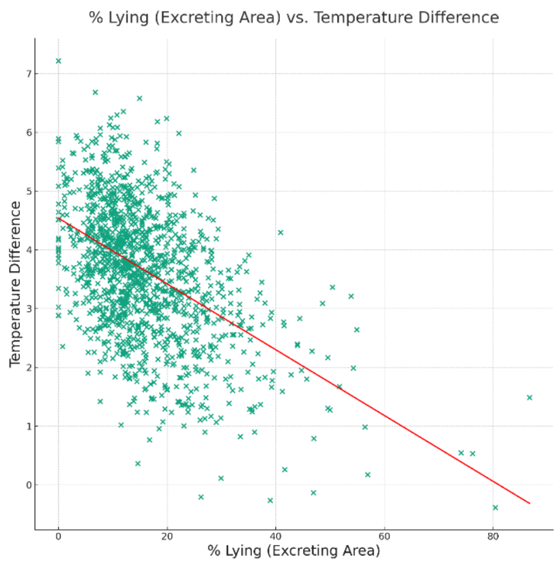

Figure 3.3: Plot Showing the Relationship between Temperature Difference and % Lying (Excreting Area).

3.6 Discussion

Factors increasing the odds of fouling were obtained using logistic regression (see Table 3.1). The probability of fouling can seem to be significantly influenced by the pig's preferred lying area, the temperature in the rest and excreting area and water consumption.

3.6.1 Pigs Activity Prior to Fouling

The connection between thermal distress and changes in lying posture in pigs highlights a behavioural adaptation that favours increased heat dissipation. When faced with heat stress, pigs tend to assume a lateral lying position, exposing more of their skin surface to the cooler ground. This pattern is potentially a precursor to severe heat stress reactions such as fouling.

Our statistical analysis shows a significant correlation ($P<0.01$) between an increased presence of pigs in the activity and excreting areas and the rising risk of fouling. However, the activity within the rest area had no noticeable impact on fouling. A multi-level linear regression model was employed to ascertain factors influencing pig lying behaviour across three distinct areas: rest, activity, and excreting areas. The model factored in variables including DAY and FOUL and their interactions, with DAY (0) and NO FOUL as the starting points.

No substantial difference was detected in lying activity and days leading to fouling within the Excreting and Rest areas. However, there was a noticeable decline in the number of pigs (8.1%, $p=0.02$ and 9.0%, $P=0.03$ respectively) lying in the Activity area five and four days before the detection of fouling. This trend was observed in both FOUL and NO FOUL pens. No significant difference was found when FOUL and DAYS interacted with pigs lying in the Activity area.

Despite similarities in pen design, climate control, and physiological characteristics between the fouled and non-fouling pens, variations in the activity area can be attributed to increased lateral lying by pigs, which necessitates more space for heat dissipation. This behaviour typically initiates when the temperature crosses a certain threshold (Aarnink et al., 2006). Before reaching this point, pigs employ other coping mechanisms to maximise heat dissipation, thereby increasing activity in both FOUL and NO FOUL pens.

As temperatures rise, the space requirements of pig's surge significantly due to the adoption of lateral lying positions aimed at maximising heat transfer by exposing more skin surface area. The occurrence of lateral lying increased by 0.8% per ^{o}C (Huynh et al., 2005), 1.9% per ^{o}C (Aarnink et al., 2006), and 1.40 %

per ^{o}C (Brehme et al., 2004).

Fouling pens showed an average of 14.0% ($p<0.001$) more pigs in the Activity area and 2.8% ($p=0.003$) and 11.0% ($p<0.001$) fewer pigs in the Excreting and Rest areas, respectively. This implies that a key distinction between fouling and non-fouling pens is the significant surge in the number of pigs in the Activity area in fouling pens. Although the increased presence of pigs in the activity area does not directly result in a fouling event, it is a potent indicator of thermal discomfort in pigs.

Notably, despite pigs' well-documented aversion to lying in excreting areas (Huynh et al., 2006; Stolba & Wood-Gush, 1989), this study suggests that extreme thermal distress might alter this behaviour. Under severe heat stress, pigs might lie in their excrement as an alternative form of wallowing without mud or water wallows. This behaviour can be seen as a step in behavioural thermoregulation and might be regarded as a desperate move by the pigs to find relief from heat stress.

This displacement from the activity area to the excreting area might suggest that a lack of space for adopting a lateral lying position could lead to this behavioural change. As the pigs' space requirements increase due to lateral lying, the excreting area could become an additional space for heat dissipation. This could further explain the significant correlation between the increased presence of pigs in the activity and excreting areas and the higher risk of fouling.

Drawing on the review by Bracke (2011), wallowing behaviour in pigs is primarily driven by the need for thermoregulation. Therefore, confined feeding operations should consider providing ample space for pigs to adopt lateral lying positions or alternatives such as mud wallows or water troughs to accommodate pigs' natural thermoregulatory behaviours.

In conclusion, thermal conditions heavily influence pigs' behaviour, such as lateral lying and potentially wallowing in faeces. These behaviours could serve as early indicators of thermal distress and an increased risk of fouling, providing valuable insights for farmers to optimise pig-rearing practices for improved animal welfare and productivity.

3.6.2 Water Consumption and Fouling

The present study found that the risk of fouling decreases with increasing water consumption ($p=0.03$). This is an unexpected observation, as, under heat stress, pigs have a higher daily water intake than those in a cold environment (Huynh et al., 2005; Mount, 1975; Patience et al., 2005). This study shows that a 1% increase of pigs in the resting and activity area is associated with a change of 0.025 ($p=0.02$) and 0.027 ($p<0.001$) Litres in water consumption per pig, respectively. Hence, the strong positive interaction between percentages of pigs in the activity area shows that pigs moving from the rest to the activity area will drink more water. A hypothetical explanation of why the risk of fouling decreases when pigs consume more water may be because pigs that are unable to thermoregulate through water consumption are more prone to fouling. No significant difference was observed in water consumption between the fouling and non-fouling groups over the five observed days.

Water consumption in pigs is found to be strongly influenced by their age ($p < 0.001$). This finding supports the logical hypothesis that as pigs grow, they tend to consume more water, which has been previously demonstrated in studies by Meiszberg et al. (2009), Patience et al. (2005), and Yang et al. (1981).

An intriguing aspect of the study is the adjustment of the stable temperature based on the age of the pigs. As the pigs mature, the temperature in the stable is gradually reduced to provide them with maximum comfort. It can be expected that the pigs would drink more water as the temperature increases to thermoregulate. However, the results showed an unexpected negative interaction between stable temperature and water consumption. This finding suggests that contrary to the conventional assumption, a decrease in stable temperature was associated with an increase in water consumption ($p = 0.02$). This unexpected outcome raises questions about the relationship between stable temperature and water intake in pigs. It is essential to acknowledge that this study's observation of an unexpected negative interaction between temperature and water consumption points to potential limitations in using stable temperature as the sole determinant for predicting pig water consumption. Other factors not considered in the study may be at play, influencing how pigs respond to temperature changes and their resulting water intake.

Therefore, while the temperature adjustment in the stable is intended to optimise the pigs' comfort, the study's results suggest that this may not be a straightforward predictor of water consumption. Further research and investigation are warranted to understand better the complex factors that govern water consumption in pigs and to avoid oversimplification in management practices.

3.6.3 Temperature and Fouling

From logistic regression, see Table 3.1, it can be observed that an increase in the temperature in the rest area is associated ($p=0.001$) with an increased risk of fouling, while variation in temperature in the excreting region has no significance on fouling ($p=0.06$).

The full linear regression model used to predict the temperature in the slatted and rest areas included all the dependent variables, including the DAYS and FOUL, as the factors. No significant difference in temperature in the rest and excreting areas was observed between fouling and non-fouling groups. However, it was interesting to observe that five,four and three days before fouling was observed, the temperature in the excreting area was significantly lower by 0.61 ^{o}C ($p=0.04$), 0.66 ^{o}C ($p=0.05$), 0.60 ^{o}C ($p=0.04$) compared to the day that fouling was observed. This increase in temperature in the excreting area indicates pigs starting to move away from their resting area to thermoregulate, as previously discussed and can be related to increases in activity in the rest area.

The significant terms were selected and summarised in the reduced model, see Table 3.2. As expected, the temperature in the rest and excreting areas are significantly influenced by changes in the overall section temperature. However, the temperature in the rest and excreting area is significantly different from the measured section temperature (temperature sensed by the climate controller) by 5.6 ^{o}C ($p<0.001$) and -0.64 ^{o}C ($p=0.02$), respectively. This temperature difference is normal (Børge and Kristensen, 2016) and results from the body heat dissipated by pigs in the resting area.

For every 10% more pig in the activity area, there is a -0.2 ^{o}C ($p=0.01$) and 0.62 ^{o}C ($p<0.001$) change in temperature in the rest and slatted area, respectively. This shows a strong relationship between the percentage of pigs lying in the different zones and the temperature in the slatted and rest areas.

Temperature difference (Temperature (Rest) minus Temperature (Excreting)) was analysed using a linear regression model. Pens, where fouling was observed, had a smaller difference in temperature between the two areas by 0.52 ^{o}C ($p<0.007$). This can be explained by the increase in some pigs in the Activity area before fouling, thus dissipating in areas other than the rest area. A 10% increase in pigs in the activity and excreting area resulted in a decrease in a temperature difference of 0.5 ^{o}C ($p<0.001$) and 0.6 ^{o}C ($p<0.001$), respectively. The average temperature difference also depends on the total number of pigs in the pen; for every additional pig, the temperature difference drops by 0.1 ^{o}C

($p<0.001$). Hence it can be concluded that Temperature difference is a good indication of the spatial distribution of pigs within their pens.

3.7 Conclusion

In conclusion, this chapter has identified early behavioural changes, particularly spatial positioning, as potential fouling indicators in conventionally housed slaughter pigs. Significant correlations were found between the increased presence of pigs in the activity and excreting areas and the rising risk of fouling. Moreover, the temperature difference between the Rest and Excreting areas emerged as a reliable predictor of fouling. The study also found that water consumption decreases with the increased risk of fouling, contradicting the common assumption of higher water intake under heat stress. This suggests the complex interplay of various factors influencing pigs' water consumption.

The results underscore the importance of monitoring these indicators for effectively managing pig fouling events, leading to better animal welfare and enhanced productivity. The findings also highlight the limitations of using stable temperature as the sole determinant for predicting pig water consumption and suggest further research to understand the complex factors governing water consumption in pigs.

The insights gained from this study can provide valuable input for developing machine learning models to accurately predict fouling events, marking a significant advancement in precision animal farming. Further research could explore temperature variation as a proxy for assessing pig activity, streamlining the monitoring process for farmers and researchers alike.

Chapter 4: Prediction of tail-biting, fouling and diarrhoea in pigs.

4.1 Abstract

Tail-biting, fouling, and diarrhoea affect the welfare condition of commercially farmed pigs and reduce the profitability and sustainability of commercial farms. Early detection of these unwanted events can be achieved by detecting behavioural changes. Pigs will likely change their diurnal behaviour when their health and welfare conditions are compromised. However, early detection of diurnal behavioural change is challenging due to the erratic nature of the pig's behaviour.

This study proposed using a stacked bidirectional long short-term memory and feed-forward neural network architecture to address the complexity of identifying behavioural changes in commercially farmed pigs. This study aims to train the proposed neural network to automatically learn and classify patterns from time series data to predict tail-biting, fouling and diarrhoea on commercial pig farms. Using the common neural network architecture, three separate models were trained to predict the occurrence of each event within specific pens.

To infer the behavioural state of each pen, pig water consumption, pen level temperatures, indoor climatic data (ventilation, cooling, heating, and relative humidity), and the pen characteristics were used. In this study, data were collected from 112 pens. A total of 7632 samples were used to train and test the proposed neural network. The network was trained using a stratified 10-fold cross-validation approach to minimise training bias.

The performance on the test set was measured using the area Under the curve of the Receiver Operating Characteristic (AUROC). Using a 7-day window, AUROC of 0.782, 0.775 and 0.820 was obtained to predict tail-biting, diarrhoea, and fouling. This study demonstrated that the proposed neural network architecture could successfully learn the behavioural changes that cause specific welfare and health problems in pigs.

To conclude, besides the ability to effectively learn to predict a range of health and welfare problems robustly, the approach taken in this study did not involve the laborious task of manual feature engineering that traditional machine learning often requires. The neural network proposed in this study can automatically learn complex non-linear relationships from temporal data, speeding the model development process for health and welfare problem detection. However, the features that neural network models learn are often abstract and difficult to

interpret. For more interpretability, future studies will investigate how machine learning techniques perform using a range of manually extracted features from temporal data.

Keywords: Neural Network, Behavioural Detection, Automatic Feature Learning.

4.2 Introduction

In recent years, there has been growing concern over pig welfare in intensive production systems (Mellor, 2016). Keeping high health and welfare of pigs reared for commercial consumption can have many implications, including production profitability and sustainability. Therefore, minimising health and welfare-related issues on a commercial pig farm has become increasingly important. Changes in behaviour may occur prior to clinical observation of a disease or an unwanted event and may thus indicate a compromised health and welfare state. Such changes often occur prior to signs of injury or disease outbreak (González et al., 2008; Kyriazakis et al., 1998; Tolkamp et al., 2011) or even outbreaks of tail biting (Larsen et al., 2016). However, manual quantification of behavioural changes by farm staff is impractical on commercial farms as they are subjective and time-consuming, especially for intensive production systems (Hemsworth et al., 2000). In addition, behavioural changes may be subtle and shown as alterations in the diurnal pattern (Andersen et al., 2016).

The development of automatic early detection systems relies on monitoring quantifiable variables associated with health and welfare problems. For example, changes in feeding behaviour have been associated with dehydration from diarrhoea (Madsen & Kristensen, 2005), stress (Averos et al., 2008) and high ambient temperature (Rushen et al., 2012). Changes in excreting behaviour can indicate fouling, while changes in diurnal activity patterns have been linked with outbreaks of infection (Escobar et al., 2007; Reiner et al., 2009), tail biting (Statham et al., 2009) and stress (Salak-Johnson et al., 2004).

In practice, behavioural data have a temporal structure, i.e., they are time-dependent. Hence, the sensor-based systems developed in pig research for event detection have been feature-based (Exadaktylos et al., 2008; Jensen et al., 2017; Madsen & Kristensen, 2005), i.e., models that parameterise a mathematical function to model the temporal structure of the data. Numerous methods have been used to model pigs' behaviour, such as linear regression, analysis of variance (Brown-Brandl, 2013; Quiniou et al., 2001), Gaussian models (Morgan et al., 2000), logistic models (Kyriazakis et al., 1998), feed-forward neural network (Cross et al., 2018), and dynamic linear models (Jensen et al., 2017). The parameters of these models (features) could then be passed through a classifier that will classify a set of features as malign or benign. Two major drawbacks of these types of approaches are their complexity, as they rely heavily on carefully engineered features and their dependency on expert knowledge of the relationships between the events to be predicted and the possible predictors of such events.

In many cases, the occurrence of a specific event is often a non-linear combination of several features (Larsen et al., 2016). Hence, the feature-based approaches currently available from the literature would require complex feature engineering to identify the features in the data that can be used to predict specific health and welfare problems. Feature-based models often fail to generalise on unseen data, as the features causing health and welfare problems often vary between herds and are hard to identify (Larsen et al., 2018b). Because of the complexity of event prediction, models available in the literature have generalised the events as one class, i.e., predicting that something unusual will happen without distinguishing what that specific event might be (Jensen et al., 2017).

Recently, recurrent neural networks (R.N.N.s), a type of artificial neural network (A.N.N), have excelled in the task of automatic feature learning and classification of temporal data. Compared to traditional classification algorithms, R.N.N.s can better generalise (Liu et al., 2017) as they can learn and classify highly non-linear features from the data. R.N.N.s are more accurate and robust in comparison with the more traditional Feed Forward Neural Networks (F.N.N.) in areas where the temporal structure of the data is important as they focus on learning directly from a sequence of data rather than individual time samples (Bengio et al., 1994). For this reason, R.N.N.s have been applied to solve problems where the temporal structure of the data is important, such as handwriting recognition (Graves & Schmidhuber, 2009), behaviour recognition (Williams & Zipser, 1989), speech recognition (Graves et al., 2013), image classification (Ciresan et al., 2012), natural language processing (Mikolov et al., 2011), and time series classification (Karim et al., 2018).

The use of the neural network in pig research for event detection has been very limited and restricted to feed-forward neural networks (Cross et al., 2018; Oczak et al., 2014) and machine vision (Hansen et al., 2018; Smulders et al., 2006; Wang et al., 2008). To combine the power of both R.N.N.s and FFNNs, stacked or a combination of both neural networks has often resulted in state-of-the-art performance (Schmidhuber, 2015). While there exist several variants of R.N.N.s, in this study, the more widely used Long-Short-Term-Memory (LSTM) recurrent neural network (Hochreiter & Schmidhuber, 1997) was used in combination with more advanced optimisation methods that have proven to increase the performance and robustness of the LSTM (Salehinejad et al., 2018).

This study aimed to provide an alternative solution to feature-based models that is robust and capable of generalising well on a range of event prediction tasks. As a solution, this study has used a stacked LSTM-FNN to extract non-linear temporal features to be used for predicting a range of distinct events in pig production (tail-biting, fouling, and diarrhoea), hence bypassing the laborious and expert-dependent feature engineering step that the current literature has adopted. The use of recurrent neural networks for event monitoring on commercial pig farms is considered novel.

The objectives of this study were:

- To develop an approach that can be used to automatically train a neural network to predict specific events on commercial pig farms,
- To demonstrate the ability of a neural network to perform well on a range of prediction tasks: tail-biting, diarrhoea, and fouling,
- To show the robustness of the neural network by applying the model to unseen data.

4.3. Materials and Methods

4.3.1 Animals, housing, and management

The data used in this study was recorded at the finisher unit at Aarhus University (Denmark) from 2015 to 2016 by a protocol approved by the Danish Animal Experiments Inspectorate (Journal no. 2015-15-0201-00593). The finisher unit housed three sections, each of 16 Finisher pens. A full description of the experimental setup can be found in the literature (Larsen et al., 2017, 2016).

The dataset comprised 112 pens (1624 slaughter pigs). The 112 pens were divided into four batches (batch 1, 3 and 4: 32 pens each; batch 2: 16 pens) from June 2015 to November 2016. The pigs were randomly inserted into their pens at an average weight of 31.6 ± 6.6 kg. The pens were 5.45 m by 2.48 m, with the floor equally divided into solid concrete (rest area), drained (activity Area) and slatted floor (excreting area). The gap between the slats in the slatted floor and the drained floor was both 2 cm, while the respective slats were 180 mm and 80m cm wide. An overview of a pen with sensor locations is given in Figure 4.1.

Artificial light was on from 0530 to 1830 h (182 lux). The finisher units' climate (temperature (°C) and humidity (%)) were controlled by activating the heating, cooling or ventilation system via the climate control systems (SKOV A/S, Roslev, DK). Each pen included an automatically controlled shower system (SKOV A/S, Roslev, Denmark) above the slatted floor. The pigs were fed ad libitum with a commercial dry feed, and the feeders were filled three times a day at 0300, 1000 and 1830 h. The educated farm staff performed the general farm management. Between 1000 and 1200 h, the general routine in the stables was performed, including cleaning, straw provision, and a general health check of each pen.

During the daily check-up performed by the farm staff, the occurrence of diarrhoea, fouling, and tail-biting events was recorded according to the set protocol (Larsen et al., 2016, 2018a). The events were given a binary score of 0, indicating an event not occurring. A fouling event was defined as when more than half of the solid floor was wet with excreta and/or urine. The occurrence of diarrhoea was recorded daily by visual inspection of the pen from outside. Diarrhoea was recorded when at least one dropping of faeces with a liquid or runny consistency was observed. Tail biting was recorded to occur when at least one pig was observed in the pen with visible fresh blood on the tail (Larsen et al., 2017).

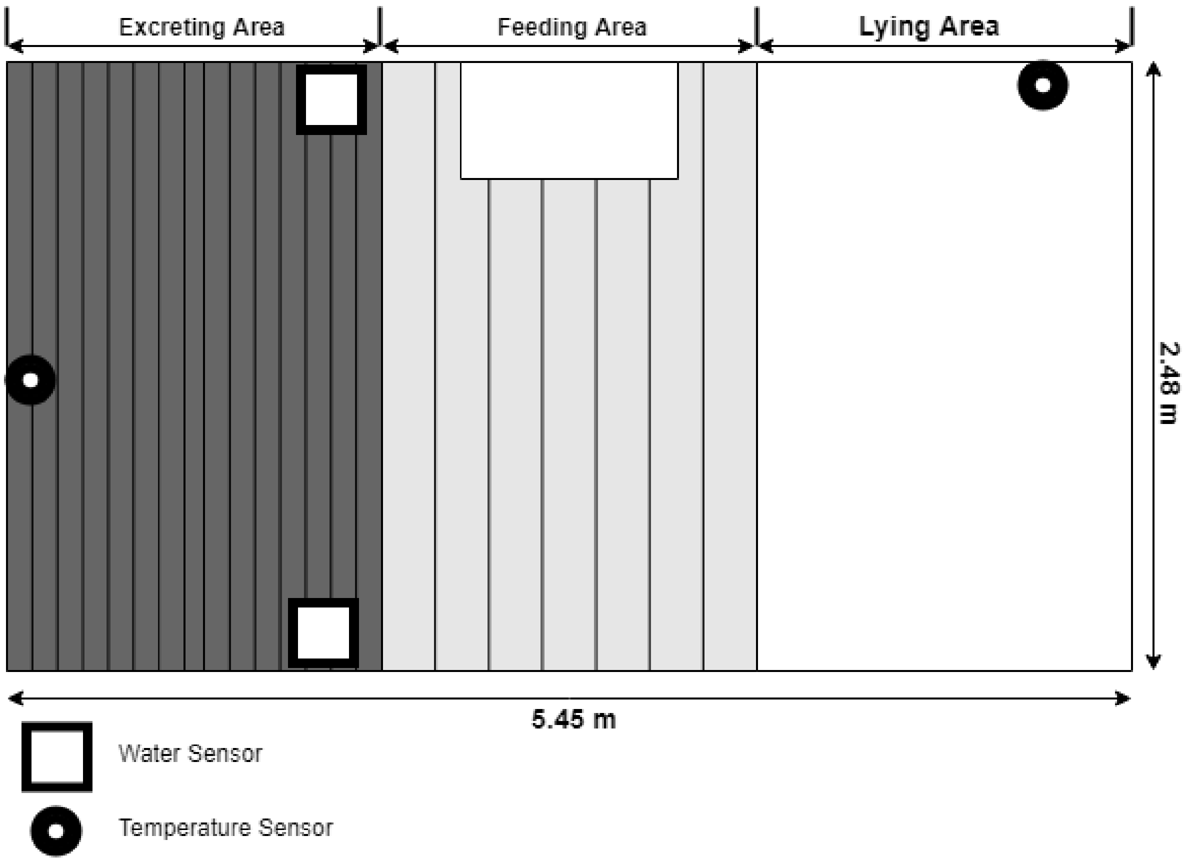

Figure 4.1: A top view of a pen with the location of different sensors and the

designated rest, feeding, and lying area.

4.3.2 Raw Data Description

Behavioural changes were monitored using a range of variables indicating drinking behaviour, the climate in individual finisher pens and the overall unit, and the characteristics of each pen. For clarification, a unit in this study refers to a stable section housing sixteen pens. The drinking behaviour was monitored using photo-electric flow sensors (RS V8189 15 mm Dia. Pipe) on the two drinkers in each pen (Figure 4.1). The temperature of each pen was monitored using two temperature probes (VE10 by Veng System) located 50 cm above floor level in the slatted and solid floor area (See Figure 4.1).

Table 4.1: Summary of Variables recorded with the raw sampling frequency and associated unit.

	Variable Name	Sampling Frequency	Unit
Input Variables	Temperature (Solid Floor)	1 Minute	Degrees
	Temperature (Slatted Floor)	1 Minute	Degrees
	Water (Drinker 1)	10 Seconds	Litres
	Water (Drinker 2)	10 Seconds	Litres
	Relative Humidity (Finisher Unit)	1 Minute	%
	Ventilation Output (Finisher Unit)	1 Minute	%
	Heating Output (Finisher Unit)	1 Minute	%
	Cooling Output (Finisher Unit)	1 Minute	%
	Temperature (Finisher Unit)	1 Minute	Sensor
	Age of Pigs	Daily	Days after Insertion
	Number of Pigs	Daily	N/A
	Straw	N/A (Constant)	Binary
	Tail Type	N/A (Constant)	Binary
Output Variables	Fouling	Daily	Binary
	Diarrhoea	Daily	Binary
	Tail Biting	Daily	Binary

Measurements from the Finisher Unit climate control system (SKOV A/S, Roslev, Denmark) were extracted: these measurements included the temperature and relative humidity of the Finisher unit. Each unit's output from the cooling, heating, and ventilation systems (SKOV A/S, Roslev, Denmark) was recorded as a percentage of the system's maximum output capacity. In addition, pen-level information on the age of the pig and the number of pigs were recorded daily. Categorical variables describing the individual pen characteristics (Tail type (Docked or undocked) and Pen type (Straw or No Straw)) were included in the dataset. The categorical variables were encoded as binary constants. A total of thirteen variables were recorded for each pen, as shown in Table 4.1. The raw sampling frequency of all the variables used in this study is shown in Table 4.1. The raw continuous sensor data were resampled to a 10-minute sampling interval.

The objective of the neural networks was to learn potential precursors of tail-biting, fouling or diarrhoea. Three different models were trained for the detection of each task. To train the neural network on detecting precursors for each task. The network was fed with labelled data samples. A data sample, X_m, corresponds to a day of continuous recording of the thirteen variables sampled at 10-minute intervals (144 time steps). Hence X_m is a matrix of dimension $13\ variables \times 144\ time\ steps$. Each vector $x_m^{<t>}$ in X_m represents the vector of variables measured at time $< t >$. The labelled data for X_m (i.e., fouling, tail-biting, and diarrhoea occurrence) were encoded as binary targets for each sample. 0 and 1 indicating negative and positive samples, respectively. A mean normalisation function was learned for each sensor using the training data. The learned functions were used when testing and validating the model. Hence keeping the normalisation procedure constant for all datasets.

4.3.3 Training, Test and Validation Dataset

Figure 4.2 shows a standard neural network-building workflow (Chollet, 2017; Gron, 2017) that was used in this study. The dataset was divided into a training, validation, and test data set. The training set was used to parameterise the weights (parameters) of the neural network; the validation set was used to find the optimum neural network architecture (hyperparameters). The test set was used to evaluate the performance of the trained neural network with the optimum architecture. A summary of the data sets used in this study is given in Table 4.2.

Table 4.2: Overview of data showing each group's start and end date, with the total number of pens, the number of samples and the number of fouling, tail-biting and diarrhoea occurrences recorded.

Start	End	No. of Pens	Purpose	Samples ($\mathbf{X_m}$)	Fouling	Tail Biting	Diarrhoea
16/06/2015	03/09/2015	16	Train	944	288	6	52
16/06/2015	03/09/2015	16	Train	944	182	16	80
14/09/2015	03/12/2015	16	Train	1072	158	7	35
12/01/2016	31/03/2016	16	Train	1040	76	18	59
12/01/2016	31/03/2016	16	validation	1040	84	14	72
07/09/2016	26/11/2016	16	Test	1296	175	9	8
07/09/2016	26/11/2016	16	Test	1296	163	8	10

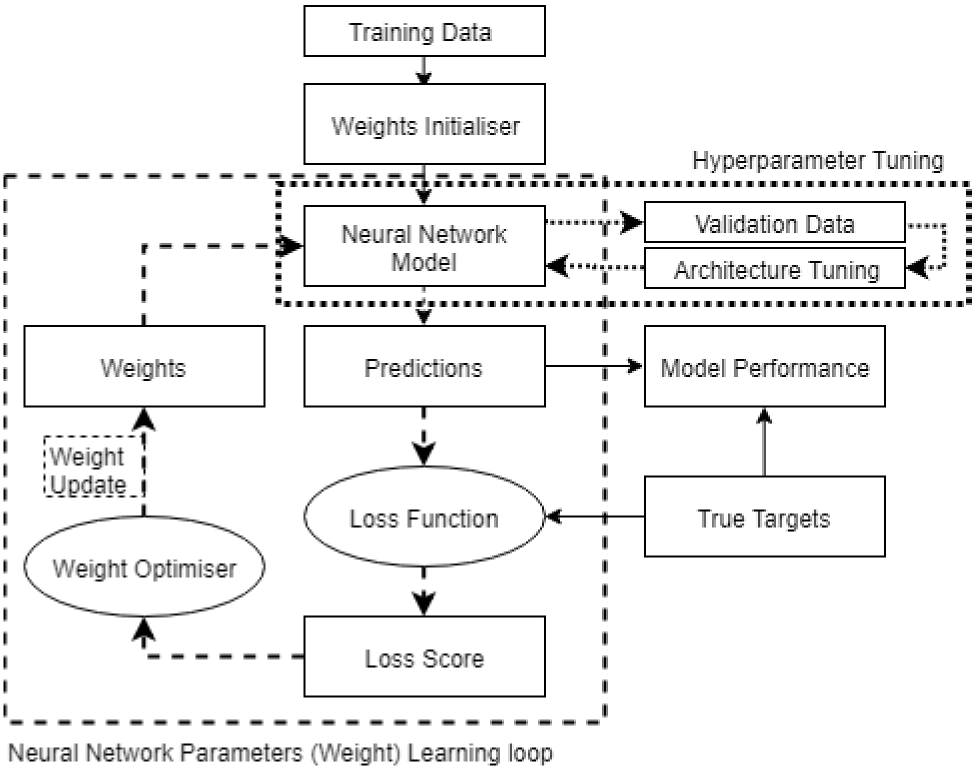

Figure 4.2: Neural network building workflow. The neural network parameter learning procedure and neural network Hyperparameter tuning procedure are also shown.

4.3.4 The Neural Network Architecture

The neural network models in this study were implemented using the Keras Application Programming Interface (Chollet, 2015) with a TensorFlow Backend (Abadi et al., 2016). The customised neural model architecture is programmatically defined in Python using Keras's Functional Application Programming Interface.

4.3.4.1 Temporal Feature Learning

A recurrent neural network (R.N.N.) learns to detect patterns in time-series data that indicate whether a signal belongs to the malign or benign class. These patterns are learned by looking at previous time series from both classes and memorising the useful patterns within the training data. R.N.N. is a class of neural networks and there exist several R.N.N. variants (Salehinejad et al., 2018). This study used the Long-Short-Term-Memory R.N.N. (LSTM), proposed by Hochreiter & Schmidhuber (1997). The LSTM is widely used because of its ability to learn non-linear, longer-term patterns from time-series data (Greff et al., 2015).

The LSTM transforms time-series data into a useful set of features using a series of gated functions controlling information flow in memory cells. An LSTM cell has two forms of memory: the short-memory cell $h^{<t>}$ and long-memory cell $C^{<t>}$. The short-term memory memorises temporal dependencies from the previous time step, $t-1$, while the long-term memory memorises meaningful temporal dependencies from all previous time steps. The mapping of the time series to a non-linear stationary feature vector is encoded in the weights (W) and biases (b) of the LSTM. Figure 4.3 shows how an LSTM processes information at time $< t >$.

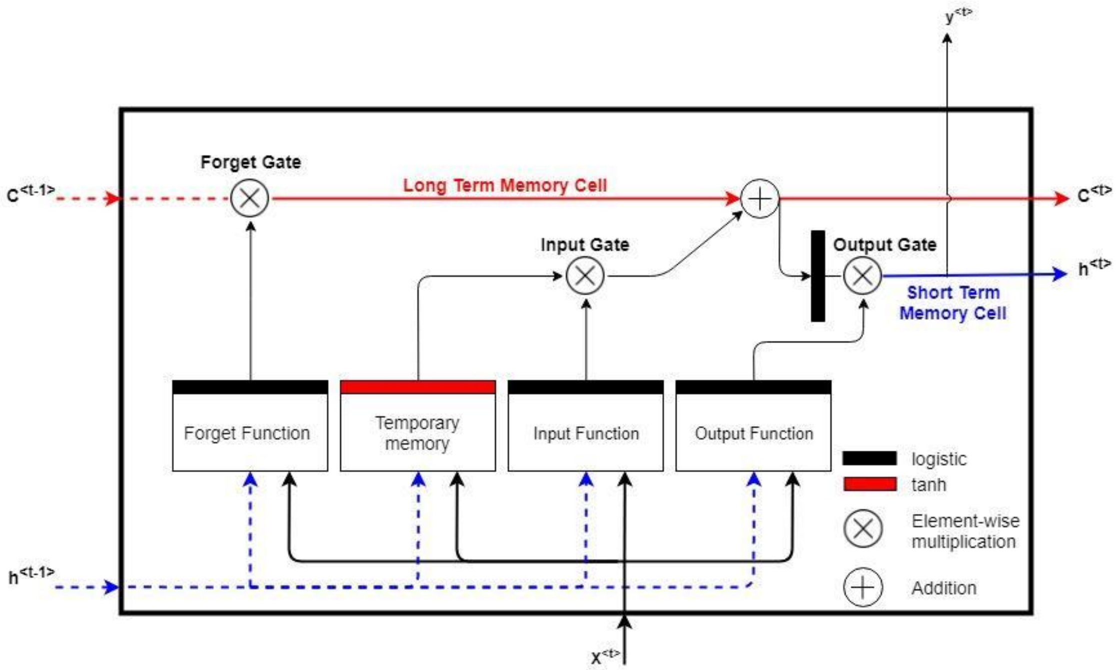

Figure 4.3: Illustration of a Long Short-Term Memory Cell. Short-term (h) and long-term (C) information from the previous time step (t-1) enters the current LSTM cell at the current time step (t). The information from the previous time step, $\boldsymbol{h^{<t-1>}}$ and $\boldsymbol{C^{<t-1>}}$, are filtered using a series of gated functions. The output from the LSTM cell is filtered short and long-term information for the current time step, denoted by $\boldsymbol{h^{<t>} \; and \; C^{<t>}}$ Respectively.

The long-term memory cell at the current time step $C^{<t>}$ drops' some' memory from the previous time step $C^{<t-1>}$ and then adds 'some' relatively new memories. All the information from the current time step is computed in the temporary cell. $\tilde{C}^{<t>}$. The filtration of useful information in $C^{<t>}$ is controlled by the forget Γ_f and input gates Γ_u. Equations 4.1 to 4.4 show the computations to update $C^{<t>}$.

$$C^{<t>} = \Gamma_u \otimes \tilde{C}^{<t>} + \Gamma_f \otimes C^{<t-1>} \quad \text{Equation 4.1}$$

$$\tilde{C}^{<t>} = tanh(W_C^T[h^{<t-1>}, x^t] + b_c) \quad \text{Equation 4.2}$$

$$\Gamma_u = \sigma(W_u^T[h^{<t-1>}, x^t] + b_u) \quad \text{Equation 4.3}$$

$$\Gamma_f = \sigma(W_f^T[h^{<t-1>}, x^t] + b_f) \quad \text{Equation 4.4}$$

The b and $\otimes$ operations represent bias and element-wise multiplication, while the term $[h^{<t-1>}, x^t]$ represents the horizontal concatenation of the matrices $h^{<t-1>}$ and x^t respectively. The gates in an LSTM cell are all bounded by the logistic sigmoid function $\sigma(\,.\,)$. The hyperbolic tangent function ($tanh$) is used for the temporary cell, $\tilde{C}^{<t>}$.$C^{<t>}$ is passed to the next time step without any further transformation. The short-term state $h^{<t>}$, also the current cell output $y^{<t>}$ are described by equations 4.5 and 4.6:

$$\Gamma_o = \sigma(W_o^T[h^{<t-1>}, x^t] + b_o) \quad \text{Equation 4.5}$$

$$y^{<t>} = h^{<t>}\Gamma_o \otimes tanh\,(C^{<t>}) \quad \text{Equation 4.6}$$

The subscripts u, f, o and c represent the input gate, forget gate, output gate and temporary cell, respectively.

An LSTM network consists of interconnected LSTM cells unrolled for each time step in a sequence. The original LSTM only processes information in the forward directions. In this study, the Bi-directional LSTM (BLSTM), proposed by Schuster & Paliwal (1997), was also tested to learn more complex features in both backward and forward directions. An unfolded BLSTM consisting of a forward and a backward LSTM layer is illustrated in Figure 4.4

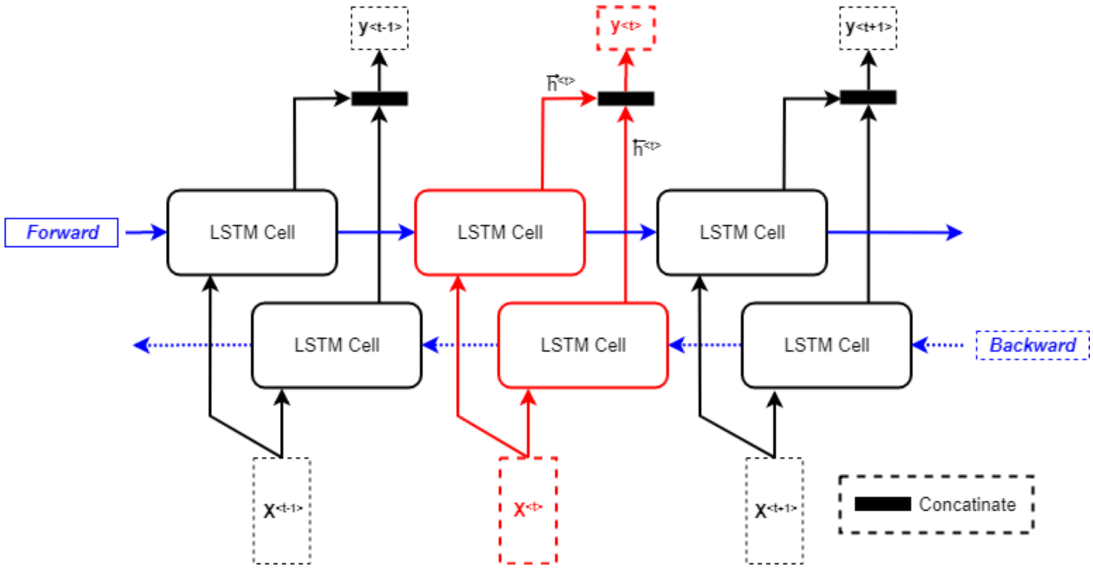

Figure 4.4: A Bidirectional Network with Long-Short-Term-Memory (LSTM) cells.

The BLSTM calculates the forward $\vec{h}$ and backward layer $h^{\leftarrow}$ outputs by going iteratively through the sequence of inputs in both directions. Both the forward and backward layer output is calculated using the standard LSTM operations, Equations 4.1 to 4.6. The output vector $Y_{BD}^{<t>}$ of the BLSTM concatenates the outputs from $\vec{h}$ and $h^{\leftarrow}$, see equation 4.7.

$$Y_{BD}^{<t>} = concat(\vec{h}, h^{\leftarrow})$$ Equation 4.7

4.4.4.2 Feature Classification

For feature classification, using the output from the last LSTM/BLSTM cell as the output of the LSTM/BLSTM network is recommended (Graves, 2012; Schuster & Paliwal, 1997). The number of hidden units controls the number of features learned,H_r , in the LSTM/BLSTM cell. Hence the vector of features from the LSTM/BLSTM network is $y^{<T>}$.

A feed-forward neural network (F.N.N.) was stacked on top of the LSTM/BLSTM network (see Figure 4.5). Stacked neural network layers can learn more complex features and improve overall model performance (Cui et al., 2018; Schmidhuber, 2015). The output vector from the LSTM/BLSTM network, $y^{<T>}$, with length H_r is fed into an F.N.N. layer with H_f hidden units (neurons in hidden layer). A more detailed description of the F.N.N. can be found in the literature (Chollet, 2017; Goodfellow et al., 2016).

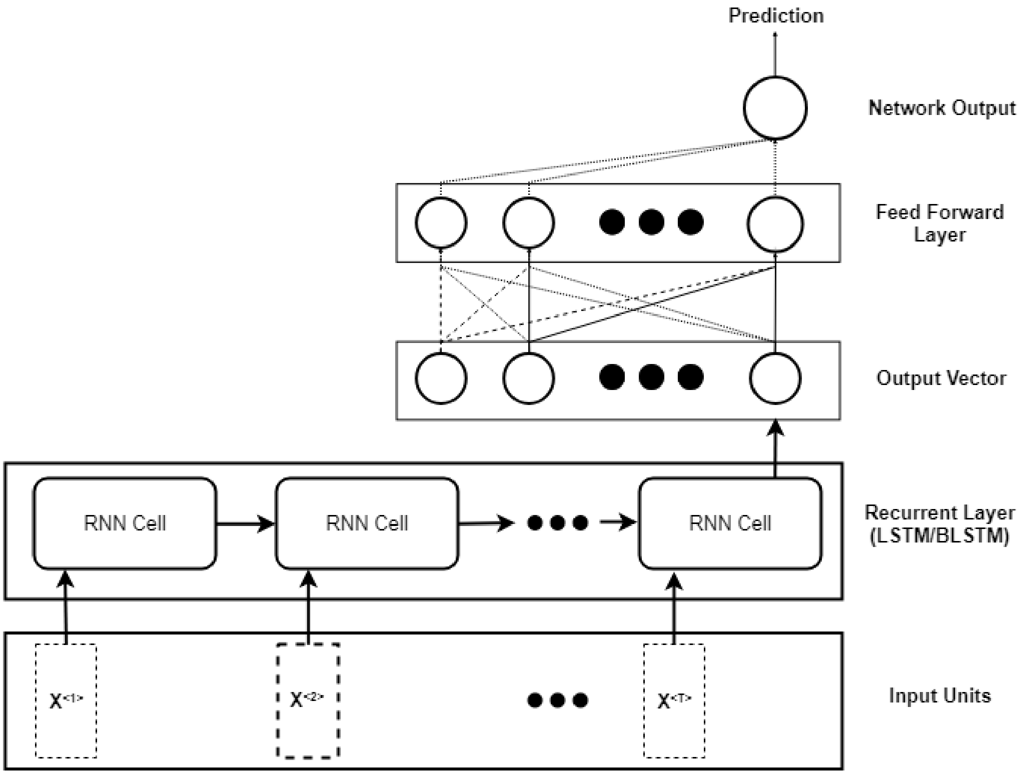

Figure 4.5: Neural Network Architecture was used for this study. The neural network classifies a matrix of input n by T, where n is the number of sensors and T is the fixed length of the time series of data for each sensor. The neural network automatically learns and classifies the n by T matrix as either malign or benign. T has a length of 144, sampled every 10 minutes daily.

4.4.5 Initialisation, Loss function, Optimisation, and Dropout

In the forward propagation step, a neural network maps an input x to a prediction output $\hat{y}$. During the propagation step, information at the output $\hat{y}$ is compared with the ground truth y through the scalar cost, $J(\theta)$. For classification problems, the cross-entropy loss function (equation 4.8) is best suited (Goodfellow et al., 2016) and was used in this study.

$$J(\theta) = -\frac{1}{n}\sum_{i=1}^{n}[y_i \, log\,(\hat{y}_i) + (1 - y_i)\, log\,\,(1 - \hat{y}_i)] \qquad \text{Equation 4.8}$$

where n is the total number of samples indexed i.

The proposed neural network architecture parameters (weights and biases) were randomly initialised using the Glorot initialisation (Glorot & Bengio, 2010). The cost function of the neural network was minimised using stochastic gradient descent (S.G.D.) with mini batches. In this study, Adam (Kingma & Ba, 2014), a more efficient gradient-based optimisation algorithm was used. The maximum number of iterations for the optimisation algorithms was fixed at two hundred for this study.

The neural network gradient was calculated through backpropagation (Rumelhart et al., 1986). Backpropagation allows the information from the loss function to flow backwards through the neural network to compute the gradient. The backpropagation through time-variant (Werbos, 1990) was used for recurrent layers.

The neural network can 'cheat' by learning complex co-adaptation that only solves the time-series classification problem for the training data while lacking the ability to generalise to unseen data. Dropout (Srivastava et al., 2014) was used to regularise the neural network model. Dropout is a regularisation technique that makes any hidden unit within the neural network unreliable. This technique has been used to improve the neural network's performance for many sequence classification problems (Cui et al., 2018; Srivastava et al., 2014; Zhang et al., 2018).

4.4.6 Hyperparameter Tuning

The neural network does all the learning through its weights and biases. One crucial step in building a neural network is the fine-tuning of the network architecture: the structure of the neural network is defined by its hyperparameters. There is no straightforward approach for hyperparameter tuning. In this study, the optimal architecture for each sequence classification task was found using the tree of Parzen Estimators optimisation (TPE) algorithm (Bergstra et al., 2011) from the Hyperopts python library (Bergstra et al., 2013). The TPE algorithm has proven to improve model performance when compared to the more traditionally used random hyperparameter search (Bergstra & Bengio, 2012; Pinto et al., 2009).

Numerous parameters can be tuned; however, in this study, only parameters that have a high importance on recurrent neural network performance (Reimers & Gurevych, 2017) were varied. The other 'insignificant' hyper-parameters were kept at their default settings. The hyperparameter space used to find the optimal model for each classification task is detailed in Table 4.3. The hyperparameter space was optimised over a range of dropout probabilities (D_P), Hidden Units in the recurrent and feed-forward layer. The optimisation algorithm was also given a choice of either the LSTM or BLSTM. The range of dropout probabilities varied around 0.5: the recommended setting (Srivastava et al., 2014).

4.4.7 Performance Evaluation

The trained neural models were tested an unseen dataset to access the model's performance. To be consistent with the existing literature (Jensen et al., 2017; Jensen and Kristensen, 2016), a time window of detection (Hogeveen et al., 2010) was used to test the performance of the model in predicting the occurrence of tail-biting, fouling, and diarrhoea. An alarm was raised when the probability of an event happening exceeded a set control limit C_L. In this study, since the output of the neural model was a logistic sigmoid function, C_L is the probability of an event occurring.

Table 4.3: The hyper-parameter space is searched during the optimisation process of the neural network.

Hyper-Parameters	Values
Dropout Probability D_P	0.3, 0.4, 0.5, 0.6, 0.7
Hidden Units (Recurrent Layer) H_r	16, 32, 64, 128, 256,512
Hidden Units (Feed Forward Layer) H_f	16, 32, 64, 128, 256,512
Recurrent Layer	LSTM, BLSTM
Epoch	200
Batch Size	32
Optimisation Algorithm	Adam
Output Layer activation	Sigmoid Function
Loss Function	Cross Entropy
Weight Initialisation	Glorot Initialisation

The true negative (T.N.), true positive (T.P.), false negative (F.N.) and false positive (F.P.) were defined as follows using the Time window of detection:

- T.P.: An event occurring W_{alarm} days after an alarm was raised.
- F.P.: No event occurring W_{alarm} days after an alarm was raised.
- F.N.: An alarm was raised, and no event occurred in W_{alarm}.
- TN: No alarm was raised, and no event occurred in.

Simply using accuracy as a measurement for binary classification algorithms has been argued to be misleading, and a Receiver Operator Characteristic (R.O.C.) curve was recommended (Provost et al., 1998). The R.O.C. curve shows how the number of correctly predicted positive samples varies with the number of incorrectly predicted samples. Each point of the R.O.C. space represents a confusion matrix corresponding to a control limit. C_L. The R.O.C. space is constructed by plotting the True Positive Rate (T.P.R.)/sensitivity on the y-axis and the False Positive Rate (1-Specificity) on the x-axis. The specificity and sensitivity are described by equations 4.9 and 4.10, respectively:

$$\text{Specificity} = \frac{TN}{TN+FP} \qquad \text{Equation 4.9}$$

$$\text{Sensitivity} = \frac{TP}{TP+FN} \qquad \text{Equation 4.10}$$

The area under the R.O.C. curve (AUROC) is often preferred for comparing classification algorithms as it is a single metric that defines how an algorithm is performing over the entire R.O.C. space (Bradley, 1997; Singla and Domingos, 2005). An AUROC of 0.5 is equivalent to a model that makes random guesses, while an AUROC of 1.0 is a perfect model. While the R.O.C. curves and AUROC give a global overview of the performance of an algorithm, it is common in the literature to define the sensitivity and specificity at the optimum control limit. C_L: the optimum decision threshold. In this study, the optimum decision threshold was the threshold with the highest Youden Index (Fluss et al., 2005). The Youden index is equivalent to $Sensitivity + Specificity$ (Krzanowski and Hand, 2009).

A 10-fold stratified cross-validation scheme was used to evaluate the AUROC using the training and test data to minimise any estimation bias. Stratification of the cross-validation splits ensures that each fold is representative of the dataset. This scheme has been shown to provide an adequate and accurate estimate of the prediction error and expected error on future, unseen samples (Breiman et al., 1984; McLachlan et al., 2005).

4.4 Results and Discussion

4.4.1 Neural Network Architecture

The final neural architectures were obtained using the TPE optimisation algorithm for a range of hyper-parameters (see Table 4.3). The algorithms iteratively trained different neural network architectures using the training data, and the performance of each trained architecture was tested on the validation data set to obtain the validation loss. The best model for each classification task corresponds to the model with the lowest validation loss score. A summary of the losses and a description of the best neural network architecture for each detection task is shown in Table 4.4.

Using an independent validation set to tune the hyperparameters of the neural network is common practice (Chollet, 2015; Larochelle et al., 2009) as it helps the model to learn a more general solution of the classification task, hence reducing any bias the model might have learned from the training dataset (Goodfellow et al., 2016).

Table 4.4: Summary of best model architecture for each detection task with the mean losses and standard deviation for all the model architecture tested and loss of the best model architecture.

		Fouling	**Tail - Biting**	**Diarrhoea**
Best Architecture	Dropout Probability D_P	0.4	0.6	0.6
	Hidden Units (H_r)	128	128	128
	Hidden Units (H_f)	64	64	256
	Recurrent Layer	BLSTM	BLSTM	BLSTM
	Mean Loss (Std)	$0.431\ (0.055)$	$0.860\ (0.124)$	$0.549\ (0.064)$
	Best Loss	0.375	0.684	0.411

The optimal solution for each neural network architecture, with a maximum number of twenty iterations, was found after 18 hours (approximately) search using an Nvidia Quadro 4000m graphics card with an Intel Xeon E3 1535M clocked at 2.90 GHz. The best model architecture for each of the classification tasks was further trained for a maximum of two hundred epochs. It is to be noted that the BLSTM was chosen as the best recurrent cell for all the classification tasks; this was an expected outcome as BLTSM can learn to extract more complex information.

4.4.2 Performance Evaluation

The model was evaluated using a 10-fold cross-validation scheme to assess the neural network's performance developed in this study. The sensitivity, specificity at the optimum control limit and the mean AUROC curve for the prediction of tail-biting, fouling and diarrhoea are shown in Table 4.5. The 95th Confidence interval for the AUROC is also displayed in Table 4.5. The minimum observed AUROC (1-day window) were 0.542, 0.619 and 0.719, while the maximum A.U.C. (7-day windows) were 0.782, 0.775 and 0.820, respectively, for tail-biting, diarrhoea and fouling. Since all the minimum AUROC are above the 0.500 ($P < 0.001$) threshold mark, indicative of a model making random guessing, it can be

concluded that the neural networks have successfully learned features to predict tail-biting, diarrhoea and fouling from temporal farm signals.
The novelty of this research lies in the specific discrimination abilities of the neural network model proposed in this study. While previous research by Jensen et al. (2017) yielded comparable performance in predicting either diarrhoea or fouling within 1-day and 7-day prediction windows, their approach was limited to indiscriminate event detection. In contrast, the model developed in the present study matches the performance of the Jensen et al. model and advances beyond its capabilities. Specifically, the neural network model presented can learn features specific to discriminate events. This ability to distinguish between different events, rather than detecting them indiscriminately, represents a significant advance in the application of machine learning to farming practices. This fine-grained analysis ability enhances the model's potential usability and adaptability in real-world farm settings.

Predicting what type of event might occur is important as it allows farm staff to apply the right treatment to prevent specific health and welfare problems. Using a single model architecture for the detection of multiple events is novel. The ability of the recurrent neural network to automatically learn complex features from temporal data simplifies the task of learning compared to more traditional feature-based learning approaches. Hence, the approach taken in this study removed the need for feature engineering and expert knowledge about specific events. By automating the pattern learning task using recurrent neural networks, this study has demonstrated that R.N.N. is an efficient solution when the objective is to train a single model architecture for the distinct detection of a range of events.

A prediction window (described in 2.7) was used for event detection. Figures 4.6, 4.7, and 4.8 show a relationship between the prediction windows and the model performance (AUROC, Sensitivity, and Specificity): longer prediction windows lead to higher model performance. Such a relationship is consistent with what has been previously reported in the literature for detection models used for cows (Hogeveen et al., 2010) and pigs (Jensen et al., 2017; Oczak et al., 2014). The use of detection windows is motivated by the difficulty associated with predicting unwanted events in animals by observing behavioural changes: visible signs of unwanted behaviours often occur a few days after a specific behavioural change has occurred.

Table 4.5: The Sensitivity (S.E.), Specificity (S.P.) and Area under the Receiver Operator Characteristic Curve (AUROC) of the optimal control limit for tail-biting, diarrhoea, and fouling are shown. The confidence interval for the AUROC was calculated using the results from the 10-fold cross-validation.

	Tail Biting			**Diarrhoea**			**Fouling**		
Window	**SE**	**SP**	**AUROC (5^{th}, 95^{th}) CI**	**SE**	**SP**	**AUROC (5^{th}, 95^{th}) CI**	**SE**	**SP**	**AUROC (5^{th}, 95^{th}) CI**
1	59.1%	65.8%	0.542 (0.540,0.544)	85.2%	57.2%	0.619 (0.615,0.623)	86.2%	65.2%	0.719 (0.718,0.720)
2	72.0%	64.0%	0.640 (0.545,0.644)	89.2%	57.2%	0.653 (0.650,0.657)	90.0%	65.2%	0.774 (.0.771,0.777)
3	80.6%	62.7%	0.710 (0.707,0.712)	82.0%	68.0%	0.678 (0.776,0.680)	90.0%	67.6%	0.812 (0.810,0.814)
4	86.1%	62.7%	0.742 (0.739,0.746)	84.5%	68.0%	0.705 (0.700,0.710)	91.8%	67.6%	0.822 (0.824,0.825)
5	89.7%	62.7%	0.787 (0.784,0.791)	86.8%	68.0%	0.738 (0.736,0.742)	91.2%	69.8%	0.830 (0.829,0.831)
6	91.7%	67.1%	0.791 (0.792,0.793)	84.0%	72.7%	0.755 (0.750,0.760)	92.3%	69.8%	0.832 (0.822,0.843)
7	92.5%	67.1%	0.782 (0.780,0.785)	85.8%	72.7%	0.775 (0.770,0.780)	93.0%	70.7%	0.820 (0.812,0.826)

The performance metric used to evaluate the model's performance with prediction is consistent with the literature. From Table 4.6, increasing prediction windows increases the chance of the model being 'correct', hence better model accuracy. It can be inferred that increasing the Time window of detection either leads to an overestimation of the T.N. and T.P., an underestimation of F.P. or F.N., or a combination of both. The effect of the time window on model performance will vary between different detection tasks. To compare model performance and mitigate the effect of the time window on model performance, models should be compared using a time window of 1 day. For practical implementations of such a model, the window of detection needs to be carefully identified for individual tasks to optimise model performance in line with the monitoring strategy of the farmers. Further investigation of the effect of the time window of detection on the model is beyond the scope of this study and should be investigated in future work.

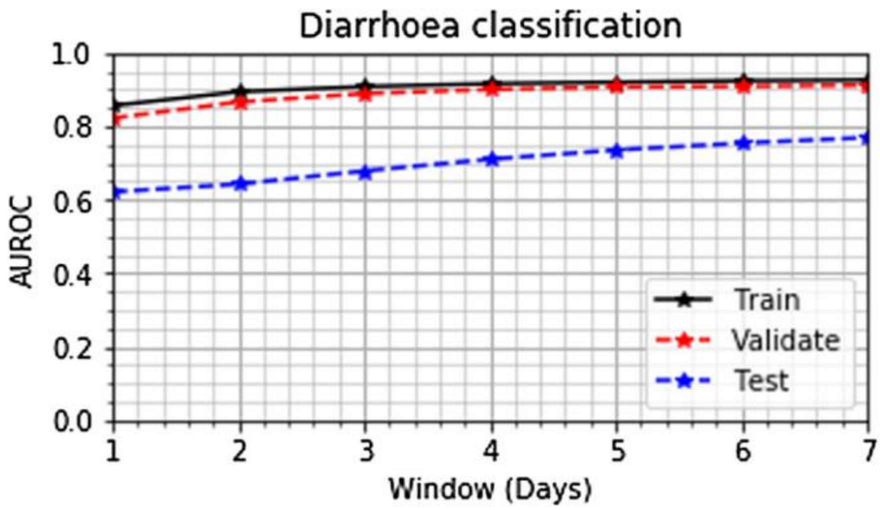

Figure 4.6: Area under the rate of recall curve (AUROC) plotted against different windows for the train, validation, and test set for Diarrhoea prediction.

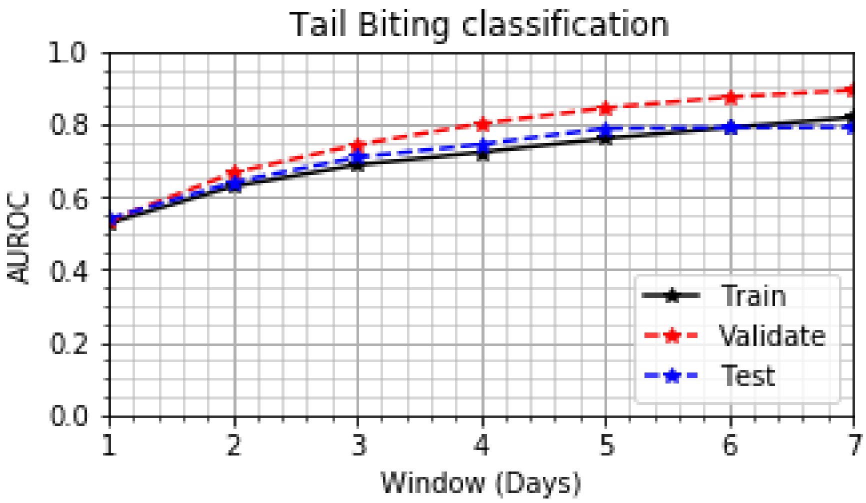

Figure 4.7: Area under the rate of recall curve (AUROC) plotted against different windows for the train, validation, and test set for tail-biting classification.

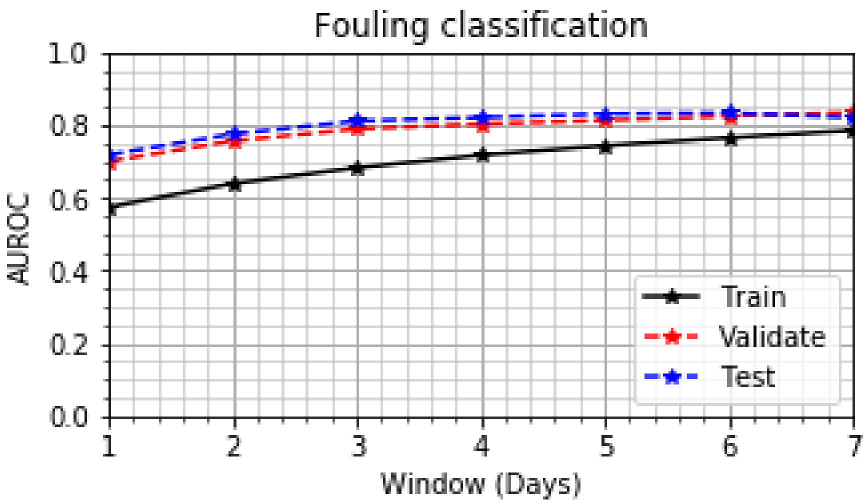

Figure 4.8: Area under the rate of recall curve (AUROC) plotted against different windows for the train, validation, and test set for fouling classification.

It has been demonstrated that the model's performance on the test set is comparable to the literature. The model's performance to generalise on unseen data can be evaluated by comparing the performance of the training and test data set. The difference between the AUROC scores of the training and test set varied from 0.010 to 0.200. Pigs' behaviour heavily depends on the management, outside climate and their genetic origin (Bolhuis et al., 2004, 2006). These factors vary between batches, hence affecting the way the animal reacts. The model's ability to generalise well will also depend on the predictability of specific events. While the models successfully learned features to predict specific events, performance varied for each event.

It is crucial to understand that while developing a universally applicable model is ideal, farming practices imply considerable diversity in management techniques, environmental conditions, and animal genetics. While the goal is a universally applicable model, each farm's unique environment and management may require some model re-training or fine-tuning. This would ensure that the model can adapt to the distinct data patterns present on each farm. Transfer learning techniques may be of value in such contexts.

Overfitting, where a model overly specialises in the training data at the expense of generalizability, is a crucial concern. This study used a 10-fold cross-validation scheme to help mitigate this risk, alongside considering other regularisation techniques that penalise overly complex model behaviours. Future strategies could include data augmentation, ensemble methods, or early stopping to prevent overfitting further.

In this study, it is anticipated that the approach outlined will have broad applicability to other farms. However, the necessity of making adjustments to accommodate specific farm conditions and the significance of preventing overfitting is acknowledged. The objective is to fine-tune the trade-off between customization and versatility through multiple methods to ensure that the model is resilient in various situations. Further research will help establish the effectiveness of the model in diverse farming environments.

4.4.3 Detecting tail-biting, fouling, and diarrhoea.

The challenge in detecting and predicting unwanted behavioural events is that these events are not clinical conditions: no clinical tests can be used to detect such events. For an expert human observer, it is impossible to predict the occurrence of such events in the pens without the help of specialised testing systems. For this reason, the literature has heavily relied on statistical inference to bring rationales into the occurrence of tail-biting (Hunter et al., 2001; Moinard et al., 2003; Ursinus et al., 2014), fouling (Aarnink et al., 2006, 2001; Bertelsen et al., 2017) and diarrhoea (Pluske et al., 1997; Rhouma et al., 2017).

Tail-biting and fouling can be viewed as behavioural manifestations because of environmental changes that stress the pigs. While diarrhoea is a recognised clinical manifestation due to the proliferation of bacteria in the intestine, the feeding behaviour and environment are also known contributing factors in diarrhoea (Rhouma et al., 2017).

The scope of this study was to automatically use observable changes within a pen ecosystem to predict tail-biting, fouling and diarrhoea. The drinking behaviour and micro-environment of pigs were used as predictors of unwanted events in this study as these factors are easily observable via sensor measurements and have shown a strong correlation in the literature with behavioural-related events (Aarnink et al., 2006, 2001; Larsen et al., 2017; Sällvik & Walberg, 1984; Ursinus et al., 2014). Feeding behaviour was not included in this study as previous studies have shown that drinking and feeding are highly correlated (Andersen et al., 2014; Bigelow & Houpt, 1988), and thus, drinking patterns may also reflect feeding patterns. The micro-environment was monitored by looking at the temperatures, humidity, ventilation output and heating output. Thermal stress, humidity, and draft within pens have often been associated with poor welfare conditions (Aarnink et al., 2001; Larsen et al., 2017) and hence were considered important for this study.

Building automated systems for behavioural event detection is very challenging as pigs' behaviour is chaotic in nature: there is no straightforward answer about which predictors or what pattern to use for the detection of specific events. The behavioural pattern leading to specific events will defer from different batches. The difficulty associated with the prediction of behavioural events is reflected in the specificities (57.2 % to 72.2 %) and sensitivities (57.2 % to 93.0 %) reported in this study (see Table 4.5).

No well-defined baseline is available in the literature to compare the specificities and sensitivities of behavioural prediction models. One key observation is that the sensitivities were consistently lower than the specificities for this study's behavioural prediction tasks (see Table 4.5). This is consistent with the literature (Jensen et al., 2017; Jensen & Kristensen, 2016): for a sensitivity of 80%, specificities of approximately 60% were reported for fouling and diarrhoea predictions. Low specificities indicate alarms raised when no event occurred (False positive); this is not necessarily indicative of poor model performance. The model learned features that can cause a specific event to happen. However, a low specificity could indicate all the risk factors being present, but the animals did not manifest any signs of the event occurring. Intervention at alarms may improve animal welfare in these cases since the underlying stressors are removed.

4.4.4 Feature learning with neural network

The underlying principle of building any detection system is to classify whether specific patterns extracted from time-series signals are malign or benign. Time series classification challenges are feature engineering static features to represent the time series signal. Such features include the time-series signal's mean, variance, and wavelet transforms. These features characterise the time-series data into a stationary space. An alternative and more common approach in the literature is using parametric functions to describe the time series signal (Jensen et al., 2017; Madsen & Kristensen, 2005). The parameters are used as a stationary representation of the time series signals. The stationary features can further train classification algorithms such as support vector machines (Hearst et al., 1998), logistic regression, and decision trees (Breiman, 2001) for prediction.

The neural network architecture used in this study automatically learns features by finding ways to transform the time-series data into a stationary space. The mapping from a non-stationary to a stationary space is encoded in the weights and biases of the neural network. The features learned by the neural network are abstract and very complex. This study has demonstrated the ability of neural networks to automate the task of learning; however, it is very difficult to interpret the meaning of the features learned by the neural network model.

4.5 Conclusion

This research introduced a framework for event detection using a stacked neural network that consists of a bi-directional long short-term memory recurrent and feed-forward network. The design of this neural network was aimed at automating the learning of pertinent temporal features indicative of tail-biting, fouling, and diarrhoea on commercial pig farms.

The model developed within this chapter has demonstrated commendable performance in identifying distinct events, delivering an area under the receiver operating characteristic curve of 0.782, 0.775, and 0.820 for predicting tail-biting, fouling, and diarrhoea, respectively. Despite these results suggesting a degree of accuracy that aligns with existing literature and the model's capacity to generalise across multiple events, it is acknowledged that the full extent of the model's generality requires additional exploration. The model's robustness was evaluated on unseen data (test set), and its performance indicates a potential for broader generalisation on fresh data. However, a more rigorous and comprehensive investigation is imperative to ascertain its broader applicability.

The methodology deployed in this study presents a substantial benefit to farm personnel by accurately anticipating specific events, thereby facilitating precise interventions. However, the research concurrently recognises the potential fluctuations in real-world agricultural environments, including the availability of the identical sensor array employed in this study. As a result, forthcoming research should contemplate the investigation of model adaptability or alternative machine learning approaches, such as Bayesian modelling, which may exhibit higher flexibility or ease of implementation across diverse farming contexts.

Furthermore, the neural network's output is a continuous signal that denotes the event's probability. This feature can be used as an indicator of welfare levels, integrated with climate control systems on commercial pig farms in the future to minimise the occurrence of specific events. While the neural network weights can be made available, it should be noted that the specific features learned by the model are abstract and complex, thus challenging to interpret.

To conclude, while the model presented in this study can learn and predict specific unwanted behavioural events, its limitations are also acknowledged, particularly regarding its generality and interpretability. Future research must focus on these areas, potentially investigating alternative machine learning models and exploring behavioural-related factors using more conventional and possibly more interpretable methodologies.

Chapter 5: A scalable, adaptive, and interpretable machine learning approach for the early detection of tail-biting, pen-fouling, and diarrhoea on pig farms

5.1 Abstract

This chapter of the thesis presents a novel approach to developing an Automatic Early Warning System (AEWS) for pig farming. The proposed system uses machine learning algorithms and multi-sensor data to predict and detect undesirable welfare issues such as tail-biting, pen fouling, and diarrhoea. The study includes the development of an efficient clustering strategy to manage vast sensor data, applying pattern extraction techniques to reveal trends and patterns, and using decision-tree-based classifiers for categorising extracted features and flagged-up fouling, diarrhoea, and tail-biting. A thorough analysis of F1 scores and the Area Under the Receiver Operating Characteristic Curve (AUROC) was conducted and benchmarked against existing literature. It was found that the methods used in this chapter were effective in predicting fouling but were unable to accurately predict diarrhoea and tail-biting. Although the modelling approach used in this chapter had its limitations, a foundation for future research was provided, which could include improving the model's precision through advanced sensors and exploring more advanced machine learning techniques.

5.2 Introduction

In sustainable pig farming, striking a balance between promoting pig well-being and ensuring economic viability is paramount. It's essential to intervene early when harmful behaviours arise, such as tail biting, fouling, or when there are signs of diseases like diarrhoea. While previous chapters underscored the importance of traditional human monitoring for early detection, these methods can be time consuming, susceptible to human biases, and financially challenging. This underscores the pressing need for innovative, cost-effective technological solutions.

Sensor-based Automatic Early Warning Systems (AEWS) are emerging as game-changers, marking a pivotal transition from traditional to technology-driven pig farming. These systems seek to augment health and welfare management practices with the aid of modern technology. Yet, AEWS also face hurdles. Managing extensive data from individual sensors and converting this information into accurate behavioural or disease forecasts remain challenging.

In the previous chapter, the use neural networks for predicting pig behaviours using data from farm sensors was explored. While potential was shown by these methods, the complexity of neural networks makes results difficult to interpret. These challenges are aimed to be addressed by this chapter through the development of transparent machine learning models that integrate multi-sensor data. The goal is to create a system that accurately predicts tail biting, fouling, and diarrhoea, and integrates these predictions into an effective early warning mechanism. Therefore, two main objectives are pursued by this chapter:

1. An enhanced machine-learning model is to be developed that combines diverse environmental sensor data sources, allowing a comprehensive understanding of pig behaviour.

2. The proposed model is to be ensured to be accurate and transparent so that stakeholders - from farmers to researchers - can easily understand and trust the system's predictions and actions.

The approach combines insights from traditional farming practices and emerging technological advancements. By integrating these perspectives, the model aims to bridge the gap between rich multi-sensor data and its real-world interpretation in the context of pig behaviour and health.

An important aspect of the proposed model is its adaptability. Given the dynamic nature of farming environments, any predictive system must be flexible enough to adjust to varying conditions and provide consistent results. Thus, the model is designed with adaptive algorithms that continually learn and refine predictions based on real-time feedback and environmental changes.

5.3 Materials and Methods

5.3.1 Experimental Setup and data collection

This chapter, like the other chapters presented, hinged on collecting and analysing a variety of data. A more detailed description of the experimental setup is provided in Chapter 2 of this thesis. Table 5.1 summarises the sensor information used in this chapter. A brief description of the sensor data used in this chapter is presented in the subsequent sections.

5.3.1.1 Temperature Sensors

Both solid and slatted floor temperature measurements in this chapter were obtained using precision thermocouples. These sensors, calibrated to an accuracy of ±0.5°C, were strategically placed within the pig housing units. The thermocouples were interfaced with data loggers that sampled temperature readings every minute.

5.3.1.2 Water Consumption Sensors

Flow meters were employed to measure water consumption in the two Drinkers (Drinker 1 and Drinker 2). These flow meters captured the water intake rates of the pigs and were calibrated to ensure an accuracy of ±1%. Data from these flow meters was logged every 10 seconds.

5.3.1.3 Relative Humidity, Ventilation, Heating, and Cooling Output Sensors

Relative Humidity, Ventilation Output, and the unit's average temperature were extracted from the climatic control systems.

Table 5.1: Summary of Data used in this chapter.

Variable	Sampling Frequency	Unit
Temperature (Solid Floor)	1 Minute	Degrees
Temperature (Slatted Floor)	1 Minute	Degrees
Water (Drinker 1)	10 seconds	Liters
Water (Drinker 2)	10 seconds	Liters
Relative Humidity (Finisher Unit)	1 Minute	%
Ventilation Output (Finisher Unit)	1 Minute	%
Temperature (Finisher Unit)	1 Minute	Degrees

5.3.2 Development of automatic early warning systems.

The AEWS development process involved data pre-processing, extracting patterns, and classifying key behavioural and health-related occurrences.

5.3.2.1 Data Clustering

Data clustering is an integral pre-processing step of the machine-learning application designed to deal with redundancy in sensor data. Due to the setup of the experimental environment, some sensors, based on their location or type, may capture similar information, leading to redundant data. This study used an automated approach to data clustering, ensuring accuracy, time efficiency, and reduced manual input.

The automatic clustering employed in this study leaned on the principles of hierarchical clustering, as Ward (1963) stipulated. This iterative clustering technique allowed the creation of several nested partitions by continuously merging similar data sets and groups. The process began by considering each data point as a separate group and gradually combining them based on their similarities, eventually forming fewer, larger clusters. The issue of ascertaining similarity between time series data sets was solved using a correlation coefficient (r).

For a pair of sensors to be considered a part of the same cluster, the correlation coefficient value (r) must surpass a set threshold of 0.95, indicating a very high similarity. These similar data sets, after identification, were integrated through a process known as Principal Component Analysis (PCA). The PCA is a statistical procedure that orthogonally transforms the original n coordinates of a data set into a new set of n coordinates known as the principal components. As a result, it condensed the high-dimensional data space into a low-dimensional space, reducing the overall complexity of the data while retaining most of the variability in the data.

Thus, the clustering and consolidation of similar sensor data dramatically streamlined the data size and reduced the complexity of the input data used for subsequent stages of pattern extraction and event classification. It also improved the interpretability and transparency of the data, making the system more efficient, responsive, and understandable.

5.3.2.2 Pattern Extraction

A suite of sophisticated algorithms was used to uncover the latent patterns in the time series signals. This exploration was segmented into four key techniques for pattern extraction:

5.3.2.2.1 Statistic-Based Pattern Extraction (Stats)

The statistic-based approach aimed at extracting statistical attributes from the time-series data. This technique incorporated entropy-related features—which convey the degree of randomness and unpredictability in the data—and computed statistical metrics such as order statistics. The latter assists in organising data points and analysing their relationships.

5.3.2.2.2 Spectral-Based Pattern Extraction (Spec)

In the spectral-based analysis, the time-series signal was transformed into its frequency domain, offering a distinct view of the data in terms of its periodic patterns. This technique hinged on two widely recognised spectral decomposition methods: the Discrete Wavelet Transform (DWT) and the Discrete Fourier Transform (DFT). For this research, the implementations of DFT and DWT as presented by Liu et al. (2004) and Frigo & Johnson (1998), were adopted.

5.3.2.2.3 Similarity-Based Pattern Extraction (Sim)

The similarity-based technique is particularly beneficial when the data exhibits repetitive patterns or baselines. It aims to discern the commonalities and deviations between standard behaviour and emerging patterns. For this purpose, the Dynamic Time Warping method, as mentioned by Balasubramanian et al. (2016), was chosen. This robust approach facilitated the assessment of discrepancies in form between the standard diurnal patterns and the newly observed ones.

5.3.2.2.4 Model-Based Pattern Extraction (MB)

The model-based technique operates on the premise that the patterns observed in a time series can be represented using specific models. By aligning these models with the data, features about the model's structure and parameters can be observed, granting a richer insight into the nature of the data.

All extraction techniques, except the similarity-based method, were executed using the "TsFresh" package (Christ et al., 2018). This involved a detailed analysis of a time series represented as T. The length of this series is designated as N. Mathematically, this can be expressed as T = {x1, x2, ..., xi}. Here, xi stands for the i^{th} sample in the ordered series.

By applying these diverse pattern extraction methods, a thorough sensor data analysis was accomplished, revealing crucial trends and patterns pivotal for predicting pig behavioural and health-related events.

5.3.2.3 Behaviour Classification

The final stage in developing the AEWS was categorising extracted features and designating behavioural events. To facilitate this process, tree-based models were applied due to their interpretability and effectiveness in multidimensional classification problems. These models made decisions by implementing a series of 'if-then' logical conditions, allowing us to categorise different behaviours accurately.

The models were trained using the gradient boosting method incorporated within the XGBoost algorithm, as Chen & Guestrin (2016) recommended. Gradient Boosting, sequentially adding predictors and correcting previous models enhanced the initial model's predictive performance. The relative contribution of each feature was determined through an index known as the Gini Importance or

Mean Decrease Impurity (MDI). The Gini Importance or MDI of a feature measures the total reduction of the criterion (impurity) brought by that feature. This provides a quantitative measure of feature significance for classification, helping to understand the underlying mechanisms for predicting fouling, diarrhoea, and tail-biting.

5.3.2.4 Training Methodology and Performance Evaluation

To avoid potential estimation biases in applying the models, a cross-validation technique was used, precisely a tenfold cross-validation method. This technique divides the dataset into ten equal parts, with a single part reserved for model validation and the rest for training. This process is iterated ten times, with each part used for validation once.

Model performance was quantified using the F1 score and Area Under the Receiver Operating Characteristic Curve (AUROC) score. The F1 score combines precision and recall of the model by calculating their harmonic mean. As such, the F1 score provided a balanced measure of the model's precision and robustness, crucial for verifying its predictive efficiency.

In Chapter 4, observations regarding tail-biting, diarrhoea, and fouling were analysed across different time windows. The most suitable and balanced choice for these behavioural indicators was found to be a 3-day window. Although a 4-day window showed slightly better AUROC and sensitivity in the context of tail-biting, the 3-day window was deemed as effective and had the added benefit of a shorter window. For diarrhoea, the 3-day window not only achieved maximum AUROC but also maintained a harmonious balance between sensitivity and specificity. Furthermore, the 3-day window yielded the highest AUROC values for fouling, indicating unparalleled model performance for this behaviour. The 3-day window offers a robust balance of sensitivity, specificity, and AUROC across all three behavioural events. Given that the metrics do not significantly improve beyond the 3-day window and considering the practicality of timely interventions, a 3-day window seems optimum.

5.4 Results and Discussion

5.4.1 Understanding Diurnal Patterns in Pig Sensors.

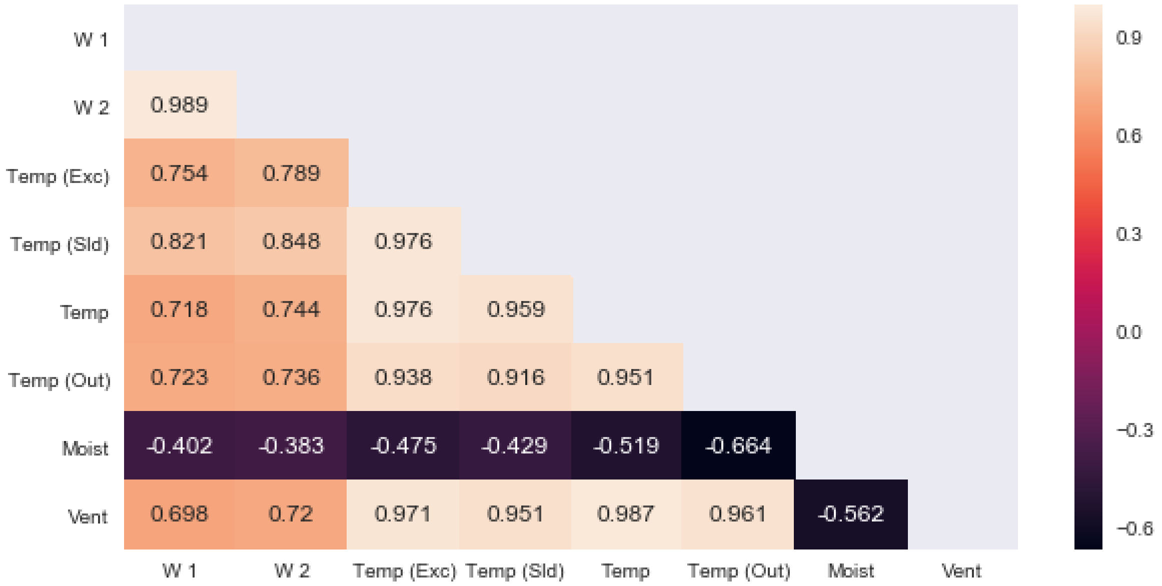

Figure 5.1: The correlation matrix for all sensors. W1 stands for water sensor at Drinker 1, while W2 stands for water sensor at Drinker 2. Temp stands for temperature, moist stands for moisture, and vent stands for ventilation. Exc represents the temperature of the excrement area, sld represents the temperature of the solid area and out represents the outside temperature.

In the pre-processing phase, a step was introduced to cluster sensors with similar information. Figure **5.**2 shows the outcomes of this hierarchical clustering. The similarity coefficient, r, had to exceed 0.95 to be classified as a cluster. The correlation matrix for the sensors is presented in Figure **5.**1.

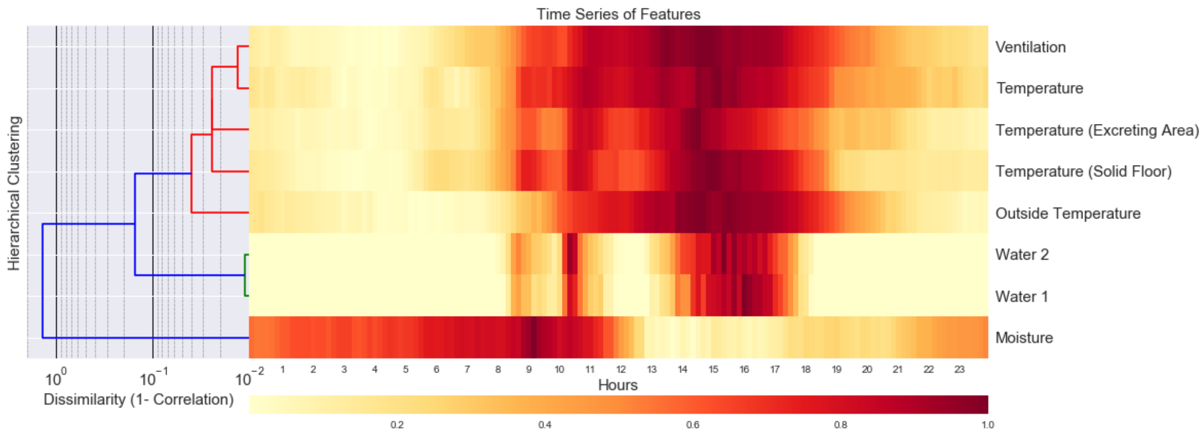

Figure 5.2: Hierarchical clustering outcomes for the sensors. Sensors with a dissimilarity greater than 0.05 are highlighted with blue links. The heatmap indicates daily average time series patterns.

Figure 5.2 presents sensor data categorized into three distinct groups: the Water Intake Group (WI), Moisture Group (M), and Temperature Group (TG). Each category captures distinct sets of behaviours or environmental conditions identified by the sensors, with their own range of values and units of measurement. Achieving valuable insights from this diverse data and ensuring consistency and comparability across these groups necessitated normalization of the data. Uniform conversion of time-series samples within the range of 0 to 1 resulted in several significant benefits:

- Simplified Computation: Algorithms, especially those used in machine learning, often converge faster and perform more efficiently when data is within a smaller and consistent scale. This normalisation, therefore, facilitated quicker and more robust processing by the AWES algorithm.
- Enhanced Comparability: The comparability of sensor outputs can be improved by standardizing the data through normalization. This practice allows for easier identification of patterns and relationships between different sensor readings, which would otherwise be difficult due to the varying nature of the raw data.
- Reduced Bias: Data based on magnitude from individual sensors was excluded. This was done because a high magnitude in one category could hide the subtler nuances in others. By doing this, it was ensured that the AWES algorithm remained sensitive to all categories and their subtle patterns and changes.
- Robustness Against External Variabilities: Pigs, as living beings, possess distinct physiological traits that can affect sensor readings even when placed under identical conditions. In addition, inconsistencies may arise in sensor readings due to modifications in experimental setups. To mitigate these innate variabilities, the algorithm normalizes the data, ensuring consistent performance

Valuable insights are provided by Figure 5.3, particularly in the data from the temperature sensor. Specific patterns can be anticipated due to the mechanistically controlled climate in the pig housing units. As previous studies have emphasized, potential stressors within the pig population may be suggested by elevated temperature levels (Barnett et al., 1984; Zone, 2003; Larsen et al., 2018). This information can be of significance to researchers and practitioners in the field of animal welfare and agriculture.

A high correlation (R-value of 0.99) was observed between the water sensors.

This suggests a consistent water consumption behaviour among the pigs. In examining the moisture readings, they appear to exhibit a less pronounced correlation with other sensor groups. However, the inverse relationships observed—expressly, r = 0.46 with the Temperature Group (TG) and r = -0.42 with Water Intake (WI)—warrant further investigation into the potential interactions or implications. Intriguingly, Figure **5.**3 shows distinct hydration peaks for the pigs at 08:30, 10:30, and 15:30. This is corroborated by analogous findings from Fernandez et al. (2011) and Madsen et al. (2005), even if the exact temporal coordinates demonstrate minor variances. Compared with existing literature, the consistency of these trends reinforces the validity of these observations.

Furthermore, the relationship between Temperature (TG) and Water Intake (WI) provides a deeper understanding of the pig's physiological responses. Heightened activity levels during feeding and hydration sessions may increase heat production. This notion is substantiated by Nahsiramadi et al. (2017), who identified a direct correlation between in-pen temperatures and pig activity levels. Additionally, Fernandez et al. (2011) posited a discernible increment in heat dissipation during diurnal periods compared to nocturnal spans.

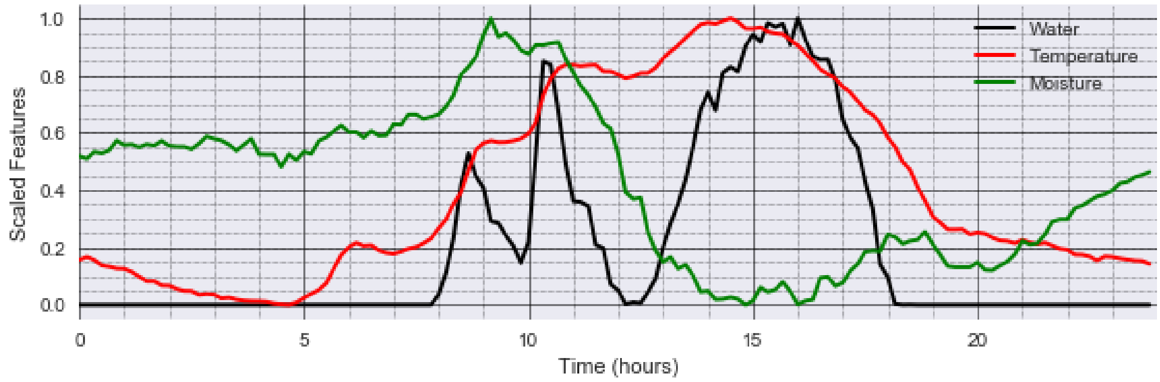

Figure 5.3: The median scaled time series for the three clustered sensors discovered in the pre-processing stage.

Based on the findings of this study, there is significant potential in categorizing sensors based on their daily patterns and recorded data. This method is crucial in gaining a deeper understanding of pig behaviour and their surrounding environment.

One major advantage of this approach is the ability to predict certain outcomes. When sensors are strategically grouped together, patterns like hydration spikes or temperature fluctuations become more apparent. This results in the extraction of predictive insights, which can assist in anticipating specific behaviours and conditions. This is valuable as it enables farmers to take proactive measures in preventing potential issues before they materialize.

Furthermore, grouping sensors leads to greater resource efficiency. Instead of processing each sensor separately, clustering them based on similarities allows for batch data processing. This not only conserves computational resources but also simplifies the analytics process.

Finally, when a clear baseline is established for each sensor group, it becomes simpler to identify deviations or anomalies. This can trigger alerts in an early warning system, enabling timely intervention and highlighting the significance of a systematic sensor grouping approach.

However, while the advantages of sensor grouping are manifold, there are potential limitations to consider:

1. Over-reliance on Correlation: It is important to keep in mind that while the R-value is a useful tool for identifying correlations, it is not necessarily indicative of causation. There may be other factors at play that are not immediately apparent when looking at the sensor data as a whole. False Positives/Negatives: An early warning system built on this approach could be susceptible to false alarms, especially if the system is not fine-tuned to account for natural variations in pig behaviours or minor sensor discrepancies.

2. Adaptive Changes Over Time: Pigs' behaviours and environmental conditions might evolve. Relying solely on historical data without periodic recalibration could render the early warning system less effective.

3. Technical Failures: Malfunctioning sensors or those affected by external interferences might provide skewed data. While the system might be built to detect significant anomalies, subtle biases introduced by such malfunctions might go unnoticed.

4. Complexity in Integration: As sensors capture varied data types, integrating them into a single cohesive early warning system might pose technical challenges. Differences in data granularity, update frequencies, or even data transmission methods could introduce complexities in building and maintaining the system.

In conclusion, while using clustered sensor data provides a promising foundation for building robust early warning systems, it is imperative to refine, test, and calibrate the system continually. Leveraging machine learning techniques that evolve with incoming data, combined with rigorous field testing, will be crucial in realising the full potential of such a system in real-world applications.

5.4.2 Comparison of Pattern Extraction Techniques

Understanding the underlying patterns is critical in predicting pig behaviours using time-series sensor data. This section delves deep into the comparative analysis of four prominent pattern extraction techniques employed in this study: Statistic-Based Pattern Extraction (Stats), Spectral-Based Pattern Extraction (Spec), Similarity-Based Pattern Extraction (Sim), and Model-Based Pattern Extraction (MB). By comparing their attributes and assessing their performance metrics, this analysis aims to comprehensively understand their capabilities and implications for automated early warning systems (AEWS) in pig farming.

Tables 5.2 and 5.3 show how different methods and sensors performed, measured by their f1 and AUROC scores. The p-values presented in Table 5.2 were obtained using a 10-fold cross-validation comparison. Cross-validation is a robust method for evaluating and validating the performance of models, especially in scenarios where overfitting is a concern. In 10-fold cross-validation:

1. The dataset is partitioned into ten equal-sized subsets.
2. Out of the ten subsets, a single subset is retained as the validation data, and the remaining nine subsets serve as training data.
3. The model is trained on the training set and validated on the validation set.
4. This process is repeated ten times, with each subset serving as the validation data exactly once.
5. Across these ten iterations, performance metrics (like the f1 score) are computed and averaged, resulting in a more robust and generalised assessment of the model's performance.

The p-values are derived by comparing these averaged metrics and the benchmark values.

When it came to fouling, the model-based method performed better with an f1 score difference of 0.039 compared to the benchmark of 0.674. While the statistical method wasn't as strong, it still had value with an increase of 0.019 and a p-value of 0.026. However, the similarity method decreased by 0.046 when compared to the benchmark. The complexity of capturing patterns for diarrhoea is evident in the substantial f1 score decrease of 0.329 when compared with the similarity method. In contrast, the statistical and model-based methods, with score differences of -0.012 and 0.008, respectively, presented relatively minor deviations from the benchmark of 0.357.

Tail-biting, inherently challenging, manifested divergent outcomes. Most notably, the similarity method, decreasing by 0.181 from the benchmark of 0.180, stands out as considerably less effective. The AUROC scores in Table 5.3 echo these patterns. The statistical method, for instance, had an AUROC score of 0.55 for diarrhoea, exhibiting an enhanced capability over the benchmark of 0.45.

The moisture sensor's deviation of -0.345 for fouling, compared to the benchmark of 0.674, suggests potential pitfalls in singularly depending on it for prediction. Meanwhile, temperature and water sensors, with score decreases of 0.041 and 0.061 for fouling, revealed deficits, though they are more moderate than the moisture sensor. Mirroring these sentiments, the AUROC values in Table 5.3 further spotlight the moisture sensor's limitations, as evidenced by its 0.20 score for diarrhoea against the benchmark of 0.45. Though not as drastically impacted, the temperature and water sensors also trailed the benchmarks with scores of 0.35 and 0.31, respectively, for the same behaviour.

The results offer nuanced perspectives on the efficacy of the varied extraction methods and sensors. Methods such as model-based and statistical have showcased potential, whereas the similarity method's performance invites scrutiny. From a sensor standpoint, the consistent underperformance of the moisture sensor necessitates caution and possibly combined usage with other sensors. Undoubtedly, these findings reinforce the intricate challenge of predicting pig behaviours. It underscores the imperative nature of meticulously selecting feature extraction techniques and sensor data combinations tailored for each behaviour.

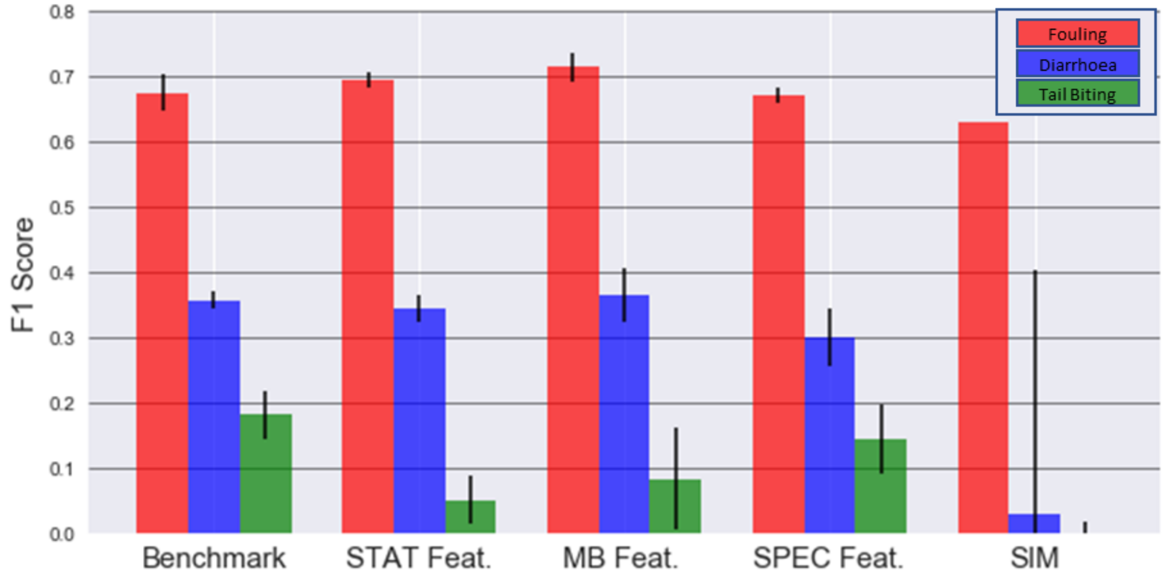

Figure 5.4: Bar chart showing a comparison of F1 scores for the various approaches to pattern extraction (which are the spectral features (SPEC), model-based features (BM), statistical-based features (STAT), and similarity-based features (SIM)) in relation to the benchmark performance,

In the literature, Jensen et al. (2017) used the AUROC (Area under the Receiver Operation Curve) to measure the performance of their predictive models, which focused on the undiscriminating prediction of diarrhoea and fouling in pigs. Their model reported an impressive AUROC of 0.76 over a three-day window. In contrast, as depicted in Figure 5.4, this study generated F1 scores of 0.67 for fouling, 0.18 for tail-biting, and 0.36 for diarrhoea. Furthermore, as evidenced by Table 5.3, the AUROC scores in this study for the same behaviours were 0.75, 0.45, and 0.30, respectively. The divergence in these metrics may hinge on the inherent challenges of forecasting these specific disease events. Notably, models relying on undiscriminating predictions tend to boast greater accuracy, primarily because of the increased probability of such events occurring after a warning signal—one critical distinction between the methodologies used. Many past research endeavours have opted for indiscriminate predictions. This means they provide generalised forecasts not specific to any events, such as diarrhoea, fouling, or tail-biting. In contrast, this research deploys a more discriminate prediction approach. This aims to specifically identify and predict individual events, making the prediction more targeted.

Indiscriminate predictions often yield higher accuracy rates. This is because by predicting events in a broad sense, without differentiating between them, the model has a higher likelihood of a correct prediction following an alert. However, the drawback is that while it might predict an event, it will not specify which one, thus potentially compromising on timely and specific interventions. On the other hand, discriminate predictions, like the one used in this study, provide more granular insights, targeting specific events. The advantage here is that it can lead to more targeted interventions, potentially saving resources and improving animal welfare. Given the complexities of differentiating between closely related behaviours or symptoms, the challenge is that it is harder to achieve high accuracy with such a targeted approach. Furthermore, this approach benefits from being adaptable to different monitoring systems in the literature. This adaptability is crucial given that most modern models that monitor pig behaviour are explicitly designed around data from specialised sensors. Kashiha et al. (2013) and Jensen et al. (2017) emphasise this point. The implication is that while the method used in this study offers a fresh perspective, it is essential to consider the kind of sensor data being used, as the model's efficacy could hinge on this factor.

Upon analysing the data presented in Figure 5.5, it becomes evident that certain features play a crucial role in predicting pig behaviour. Specifically, the features of water signal spectra and temperature prove noteworthy in forecasting fouling, tail-biting, and diarrhea, contributing to 78%, 84%, and 65% of the predictions, respectively. GINI importance was utilized to determine the significance of these features, which measures how often a feature is selected to split the data and how impactful those splits are. A high GINI importance indicates that the feature was pivotal in segregating the data and integral to the model's predictions.

In order to provide a clearer picture of which factors most influence the predictions, the values of 78%, 84%, and 65% were presented as a proportion of the most important feature, which was benchmarked at 100%. Upon further analysis of the sensor data, it was discovered that shifts in sinusoidal diurnal behaviour patterns often herald the emergence or intensification of certain adverse events in pigs, such as fouling or tail-biting.

This discovery is consistent with previous academic research. Larsen et al. (2016, 2018) and Dominiak & Kristensen (2017) have highlighted that short-term changes in animal behaviour are crucial for anticipating unfavourable events. They suggest that these momentary shifts can reveal more than long-standing behaviour patterns, which challenges traditional statistical models that rely heavily on order and moment. This perspective is further supported by the work of Moinard et al. (2003) and Aarnink et al. (2001, 2006), all indicating the importance of capturing and analysing these fleeting changes in behaviour for accurate predictions.

Table 5.2: Compares f1 scores associated with three pig behaviours: fouling, diarrhoea, and tail-biting. Scores are derived based on different feature extraction methods and specific sensor readings. The 'Benchmark' row establishes a reference f1 score for each behaviour. Subsequent rows under 'Methods' and 'Sensors' display the mean differences in f1 scores compared to this benchmark. The corresponding p-value, which results from a 10-fold cross-validation comparison alongside each mean difference, indicates the statistical significance of the observed mean differences.

		Fouling		**Diarrhoea**		**Tail Biting**	
		Mean	P Value	Mean	P Value	Mean	P Value
	Benchmark	0.674	-	0.357	-	0.180	-
Methods	Statistical	0.019	0.026	- 0.012	0.015	-0.131	<0.001
	Model-Based	0.039	<0.001	0.008	0.045	-0.098	0.009
	Spectral	-0.003	0.915	-0.056	<0.001	-0.037	<0.001
	Similarity	-0.046	<0.001	-0.329	<0.001	-0.181	<0.001
Sensors	Moisture	-0.345	<0.001	-0.344	<0.001	-0.181	<0.001
	Temperature	-0.041	<0.001	-0.146	<0.001	-0.088	<0.001
	Water	-0.061	<0.001	-0.164	<0.001	-0.118	<0.001

Table 5.3: Comparative Analysis of Area Under the Receiver Operating Characteristic Curve (AUROC) Score for Different Pig Behaviours Using Various Feature Extraction Methods and Sensor Data

		Fouling	**Diarrhoea**	**Tail Biting**
	Benchmark	**0.75**	**0.45**	**0.30**
Methods	Statistical	0.77	0.55	0.20
	Model-Based	0.78	0.46	0.25
	Spectral	0.75	0.41	0.28
	Moisture	0.50	0.20	0.15
	Temperature	0.73	0.35	0.23
	Water	0.71	0.31	0.20

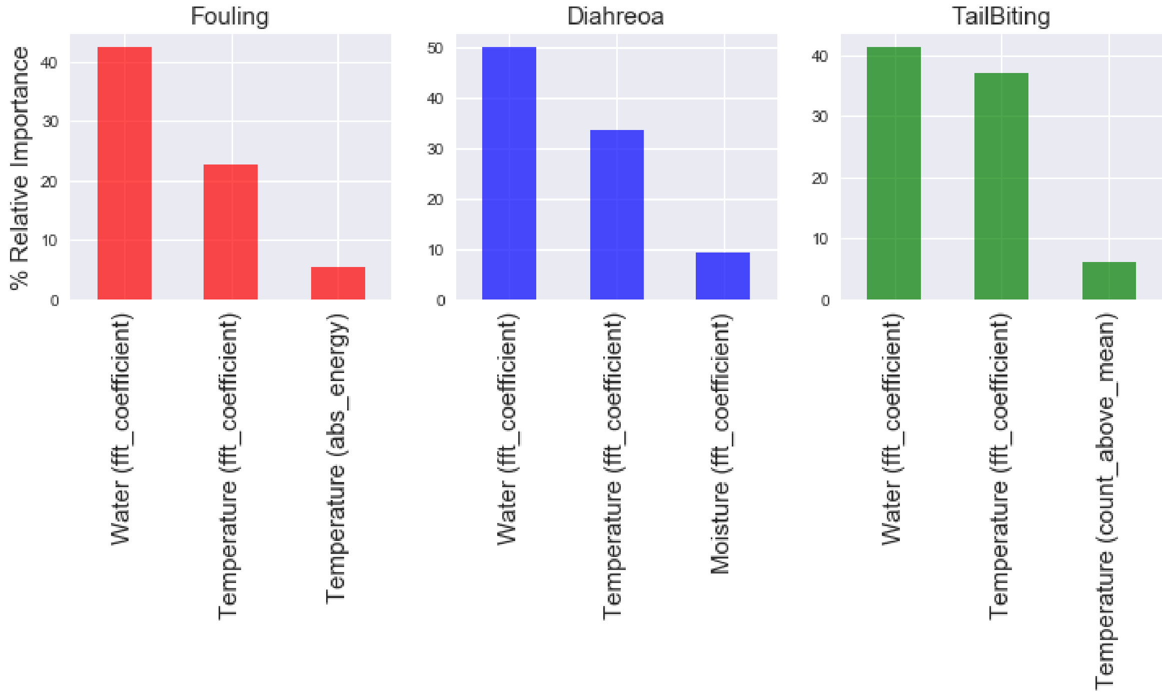

Figure 5.5: Bar charts showing three most significant features for predicting tail-biting, diarrhoea, and fouling. The convention of naming applied is in line with the one on the TsFresh package (Christ et al., 2018). All the features can be accessed at https://tsfresh.readthedocs.io/en/latest/text/list_of_features.html.

5.5 Conclusion

This work aimed to develop a multi-sensor-based Automatic Early Warning System (AEWS) that employs machine learning algorithms to detect major undesired behaviours–diarrhoea, tail-biting, and pen fouling–in pig farms.

The study achieved several significant milestones. First, it proposed an efficient clustering strategy to manage the vast data from disparate sensors. This innovative approach led to the formation of three distinct sensor groups, namely, Water Intake (WI), Moisture (M), and Temperature Group (TG), thereby significantly reducing data redundancy and enhancing the system's interpretability and efficiency.

Secondly, several pattern extraction techniques were effectively employed, revealing crucial trends and patterns for predicting pig behavioural and health-related events. The model-based method, coupled with the statistical method, delivered promising results in predicting specific pig behaviours. This comparison offers a novel understanding of the efficacy of varied extraction methods in behavioural prediction.

Lastly, the study demonstrated the utility of decision-tree-based classifiers in categorising extracted features and flagged-up behavioural events. The employed tree-based models leveraged the power of gradient boosting, enhancing the model's predictive performance.

Despite the accomplishments, the study faced limitations in predicting certain behaviours, especially diarrhoea and tail-biting. The deviation in these metrics echoed the inherent complexity of forecasting these specific behavioural events. However, despite these limitations, the model accurately predicted these events.

Future work will focus on refining the current machine-learning model. The precision can be enhanced by incorporating more advanced sensors that capture intricate details about the pigs' behaviour, movement, and physiology. The use of microbiome data to improve diarrhoea prediction also offers promising potential.

Moreover, future research must continually fine-tune the model to adapt to new data and trends. Exploring advanced machine learning techniques, including deep learning, could unfold unprecedented avenues in predicting pig behaviours.

Chapter 6: Using Bayesian Network to simulate intervention effects in pigs.

6.1 Abstract

This study presents a Bayesian Network model for predicting pen fouling in pigs and compares its performance with other models in the literature. The Bayesian Network model is designed to study the effect of different intervention scenarios but can also be used as a predictive classifier. Feature selection was performed using the LightGBM algorithm, and the top ten features were selected. The Bayesian Network model was trained and tested on discrete data and reported an Area Under the Receiver Operating Characteristic Curve (AUROC) of 0.778. The Light Gradient Boosting Machine (LightGBM) model trained and tested on raw data using 10-fold cross-validation achieved an AUROC of 0.940 and was the best performer amongst all available models. The study also highlights the relationship between temperature and fouling, showing a strong relationship between zonal temperatures and the number of pigs in the solid area. The result of this study shows the trade-off in performance between the Bayesian Network approach and traditional machine learning methods and suggests areas for future work, such as incorporating additional features or improving the model's performance.

Keywords: Pen fouling, welfare, hygiene, Bayesian Network, machine learning, pigs, randomised trials, temperature, ventilation, floor enrichment

6.2 Introduction

Pigs are naturally clean animals, keeping their rest and excretion areas separate. This behaviour has been observed in multiple studies, such as those conducted by Aarnink et al. (2006) and Stolba and Wood-Gush (1989), where pigs were seen to defecate as far away from their designated resting areas as possible. However, in conventional commercial pig farms, pigs are housed in indoor pens with either wholly or partly slatted floors to maximise space and minimise costs. Slatted floors are implemented to prevent the build-up of excrement within the pens.

Typically rectangular, commercial pens can be divided into three separate regions: rest, activity, and excretion areas. These regions are designed to be equidistant from each other, with the rest and excretion areas being the farthest apart and the activity area being located between them. This layout is chosen because pigs usually eat, drink, and defecate in that order (as documented by Randall, Armsby, and Sharp, 1983).

However, when pigs start to defecate in the designated rest area and vice versa, this behaviour is referred to as fouling and can lead to several negative consequences, such as poor air quality, poor hygiene, increased farmer workload, increased aggression among pigs, and poor resting behaviour. This has been documented by numerous studies, including those conducted by Aarnink et al. (1996), Hillmann et al. (2004), and Smulders et al. (2006). Thus, it is crucial to detect and prevent fouling to maintain the welfare and health of pigs.

The review by Larsen et al. (2017) provides valuable insights into the factors influencing fouling and suggests practical interventions for farmers to prevent it. The review identifies four primary factors that directly affect fouling in pig pens: insufficient space allowance, the flooring design of the pen, the thermal climate, and the pigs' earlier experience.

Insufficient space allowance can lead pigs not to make their way to a designated dunging area or differentiate between spaces (Hillmann et al., 2005). Pigs may be compelled to defecate in their resting area when housed in overcrowded conditions. Research has shown that when pigs are provided with adequate space, they are more likely to excrete in the defined excretion area, leading to a cleaner lying area (Zeng et al., (2023). However, excessive space allowance may result in pigs needing to be motivated to move away from the other pigs to perform their excretory behaviour (Aarnink et al., 1996). The flooring design of the pen can also affect fouling, with slatted floors being more effective at reducing fouling than solid floors (Huynh et al. (2004). A well-designed floor encourages pigs to utilise the designated dunging area and avoid fouling in their resting area (Rantzer et al., 1999). The thermal climate is another important factor, with high temperatures and humidity leading to increased fouling (Aarnink et al., 2001; Huynh et al., 2005; Spoolder et al., 2012). Finally, pigs' earlier experiences can also affect fouling, with pigs raised in a clean environment less likely to foul.

To prevent fouling, the review suggests several interventions for farmers. Optimising the pen climate by controlling temperature, humidity, and airflow is one of the most important interventions. Providing a separate dunging area and using slatted floors can also help reduce fouling. Additionally, providing pigs with a comfortable and stress-free environment can help reduce the likelihood of fouling. This can be achieved by providing adequate space allowance, using appropriate flooring design, and managing pigs' earlier experiences (Hacker et al., 1994).

Traditionally, research in understanding behaviour and prediction of fouling in pig pens has heavily relied on statistical methods to analyse data and identify correlations between various factors (Aarnink et al., 2001; Domun et al., 2019; Jensen et al., 2016; Larsen et al., 2016, 2017). While traditional statistical analyses have helped identify associations and successfully predict the event of fouling, they have limitations. They can only provide insights based on observed data and cannot effectively simulate scenarios or predict outcomes for unseen data. Causal inference is required to answer questions such as “What fraction of fouling events could have been prevented by maintaining a low temperature?” which cannot be answered using standard statistical techniques.

Bayesian Networks offer a powerful tool for causal inference. Bayesian Network is a probabilistic graphical model, first introduced by Friedman, Geiger, and Goldszmidt (1997), that aims to infer probabilities under static conditions and the dynamics of probabilities under changing conditions, such as those induced by treatments or external interventions. This method has been successfully applied in human medicine for health outcome research and medical decision analysis (Ońiśko & Druzdzel, 2013) and in agriculture for cause identification, decision support, and prediction (Drury et al., 2017).

Bayesian Networks have been applied in various aspects of animal farming, including disease vulnerability assessment, decision-making for sustainability, and aquaculture system analysis. (Manyweathers et al., 2021) used a Bayesian Network model to analyse the vulnerability of Australian sheep producers to a Foot and Mouth Disease (FMD) outbreak. The model was built using data from sheep farmers and categorised them into six risk-based typologies. The study found that vulnerability increases as property size and ewe numbers decrease, with exposure variables such as restricting visitor access and enforcing visitor biosecurity practices having the most influence. (Ferro et al., 2023) integrated the pillars of sustainability into a Bayesian Belief Network model to select the best techniques for reducing NH3 emissions from the agricultural sector. The Bayesian Network model was used to evaluate different scenarios’ potential effects, providing policymakers with recommendations on the most promising emission reduction techniques. Soriano et al., 2022) used the Bayesian Network to investigate fish pan-microbiomes and all variables in each aquaculture system. They introduced SAMBA, a software implementation of a Bayesian model, which integrates quantitative experimental data and qualitative stakeholder assessment to provide recommendations for policymakers.

Bayesian Networks are a highly effective tool for calculating probabilities in a variety of scenarios, allowing for the evaluation of different interventions and their impact on these probabilities. This process, referred to as 'scenario analysis,' employs the do-calculus (Pearl, 2009), which can modify a specific variable while keeping the rest of the network constant. This approach effectively identifies the impact of a particular manipulation on the probabilities under investigation. By leveraging Bayesian Networks, optimal strategies for reducing pen fouling in commercial pig farming can be determined with high accuracy.

The study aims to build upon these insights and utilise a Bayesian Network model to analyse the relationship between pen fouling, environmental conditions, and pig activity. The research intends to expand on existing studies by investigating the impact of potential interventions on fouling probability and unveiling any hidden relationships. By the conclusion of this study, the researchers aim to answer the following research questions:

1. What are the underlying relationships between the factors contributing to pen fouling?
2. How can the impact of each potential intervention on the fouling probability be measured?
3. Is it possible for the outcomes of unseen data or simulated scenarios to be predicted with accuracy using the Bayesian Network model?

The answers to these questions will provide valuable insights into the most effective strategies for reducing pen fouling in commercial pig farms, ultimately improving pig welfare and health.

6.3 Materials and Methods

6.3.1 Animals, housing, and management

The research study was carried out over two years, from 2015 to 2016, strictly adhering to a protocol approved by the Danish Animal Experiments Inspectorate (Journal no. 2015-15-0201-00593). During this period, four cohorts of pigs were sequentially introduced into each pen, marking distinct rounds of the experiment. The respective timelines for each round were as follows:

- Round 1: June 16th, 2015, to September 3rd, 2015.
- Round 2: September 14th, 2015, to December 3rd, 2015.
- Round 3: January 12th, 2016, to March 31st, 2016.
- Round 4: September 7th, 2016, to November 26th, 2016.

The experiment involved 112 pens (1624 slaughter pigs), which were divided into four batches (batch 1, 3 and 4: 32 pens each; batch 2: 16 pens). The pigs were randomly placed in their pens at an average weight of 31.6 ± 6.6 kg. Each pen had a size of 5.45 m by 2.48 m, and its floor was divided into three distinct areas - a solid concrete rest area, a drained activity area, and a slatted floor excreting area. The gap between the slats in the slatted floor and the drained floor was 2 cm, while the respective slats were 18 cm and 8 cm wide. A diagram of a pen with sensor locations can be found in Figure 6.1.

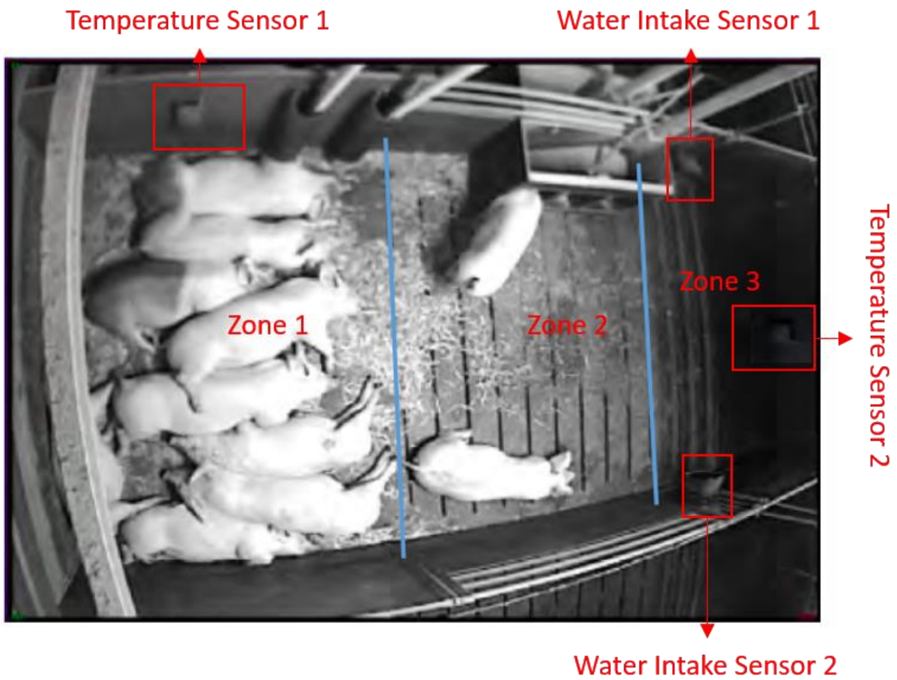

Figure 6.1: A top view of a pen with the location of different sensors. Zones 1, 2 and 3 are the designated Lying, activity and fouling areas.

The finisher pens were randomly assigned various treatments concerning tail docking, provision of straw, and stocking density. The tail docking procedure, performed within the first four days post-birth using a hot-iron cutter, conformed to Danish legislation. During the research, both docked and undocked pigs were placed in pens with or without straw. The pigs were allocated different space allowances, precisely 1.21 m²/pig and 0.73 m²/pig.

The artificial light was on from 0530 h to 1830 h, with a light intensity of 182 lux. The climate control system controlled the finisher unit's temperature and humidity (SKOV A/S, Roslev, DK). Each pen had an automatically controlled shower system (SKOV A/S, Roslev, Denmark) above the slatted floor. The pigs were fed ad libitum with a commercial dry feed, and the feeders were filled three times a day at 0300, 1000 and 1830 h. The general farm management was carried out by educated farm staff. Between 1000 and 1200 h, the farm staff performed daily routines in the stables, including cleaning, straw provision, and a general health check.

A protocol for scoring fouling was followed during the daily check-up between 1000 and 1200 h. The events were given a binary score, with a score of 0 indicating that the event did not occur. A fouling event was defined as a situation where more than half of the solid floor (Lying Area and Activity Area) was wet with excreta and/or urine.

6.3.2 Behavioural Video Observations

For each recorded fouling event, video data was extracted for five days prior to the event (including the fouling day). The average percentage of pigs lying in each of the three distinct areas (% Lying (Rest), % Lying (Activity), and % Lying (Excreting)) was calculated for all the frames extracted. Figure 6.2 provides an overview of the different areas.

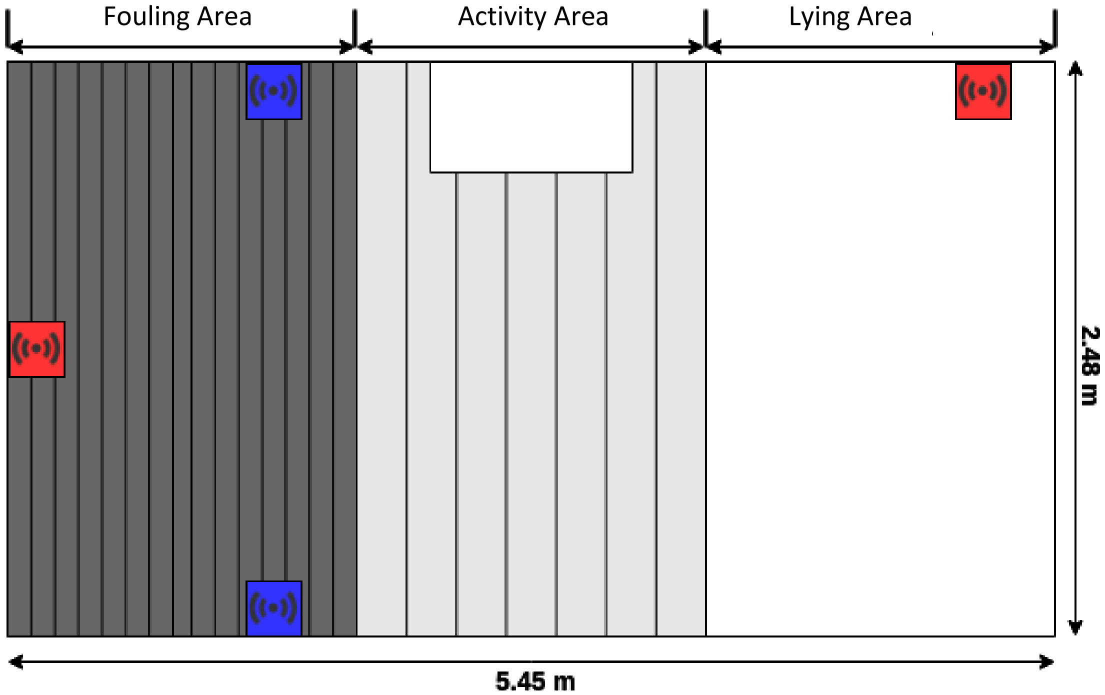

Figure 6.2: Schematic drawing of the pen design used. The location of the drinkers is indicated in blue, and the temperature sensors in red.

The video observation data were augmented with sensor information from the climate control system, temperature, and water intake sensors. When variables are continuous, Bayesian Networks can be challenging to manage. The continuous variables were discretised into groups to overcome this challenge by dividing the quantiles uniformly into four levels (Cobb, Rumí, and Salmerón, 2007; Chen and Pollino, 2012; Suzuki, 2014). Table 6.1 describes each variable and how it was discretised. For instance, Level 0 indicates that the value is in the 0 to 25th percentile range.

6.3.3 Data Summary and Discretisation.

Table 6.1 presents an overview of the variables utilised in constructing the Bayesian Network. Two types of constraints were imposed on specific variables:

Root Nodes: A variable designated as a root node is one that any other variables cannot influence in the network. This means that it cannot have any child nodes. This study identified the time and tail type of system input as root nodes, as they were deemed independent and not influenced by any other factors.

Leaf Nodes: In a network, a leaf node variable does not exert any influence on other variables due to its lack of children. In the current study, the event Variable was designated as a leaf node to enable an accurate analysis of fouling impact. By making the event Variable a leaf node, the effects of fouling were able to be studied with greater precision, without any interference from other variables.

It is important to note that no constraints were placed on the remaining variables in the network. This allowed for a more comprehensive and flexible examination of the relationships among the variables in the network.

6.3.4 Bayesian Network

The Bayesian Network is a powerful tool for modelling causal relationships between variables. It is a probabilistic graphical model that represents a set of variables and their probabilistic inter-dependencies. This model can be graphically represented by a Directed Acyclic Graph (DAG), a concept introduced by Bang-Jensen and Gutin (2009). The DAG comprises two fundamental elements: Nodes and Edges. A node symbolises a specific variable, while an edge, depicted as an arrow, signifies a directional relationship between two nodes. As Cummiskey & Lübke (2022) highlight, DAG serves as a key instrument for illustrating the causal structure of variables, often called a causal diagram (Dawid, 2002). In a Bayesian Network, each node in the graph represents a random variable, and the links between the nodes represent the conditional dependencies between the variables. The main idea behind the Bayesian Network is to express the joint distribution of a set of variables as a product of conditional distributions. The final DAG is shown in Figure 6.3.

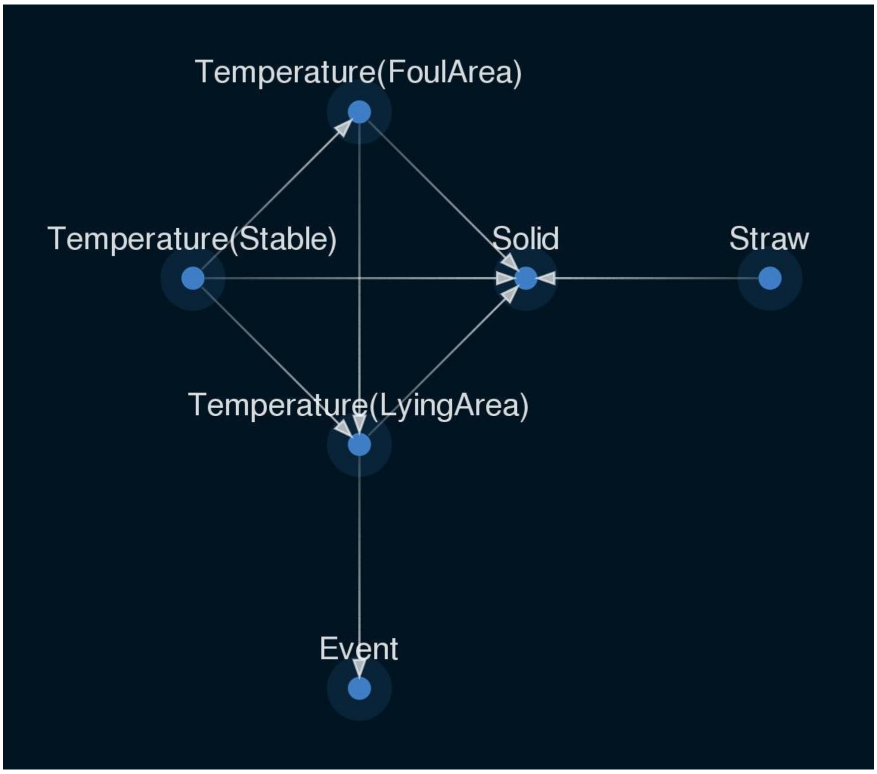

Figure 6.3: A subgraph containing the Temperature (Foul, Stable, Lying Areas), Straw, Solid and Event nodes.

Consider a joint distribution over three variables: a, b, and c. By applying the product rule of probability, the joint distribution can be expressed as:

$$p(a,b,c) = p(c|\, a,b)\, p(\, b \mid a\,)\, p\,(\,a\,) \qquad \text{Equation 6.1}$$

Equation 6.1 is valid for any choice of joint distribution and can be extended for K variables by the repeated application of the product rule of probability. The joint distribution $p(x_1,...,x_k)$ can be written as a product of conditional distributions for each variable:

$$p\left(x_1, \ldots, x_K\right)=p\left(x_K \mid x_1, \ldots, x_{K-1}\right) \ldots p\left(x_2 \mid x_1\right) p\left(x_1\right) \qquad \text{Equation 6.2}$$

For a given choice of K, Equation 6.2 represents a directed graph having Knodes, each corresponding to a conditional distribution on the right-hand side of Equation 6.2. This type of graph is called a fully connected graph, where every pair of nodes has a link between them.

This study uses DAG to study the causal interactions between variables. A DAG is a graph useful for studying causal relationships between variables, as it does not allow for any cycles.

Figure 6.5 presents an example of a DAG; the DAG describes the joint distribution over a set of variables. $x_1, \dots\ x_7$. The joint distribution of the DAG in Figure 6.5 is given by:

$$p(x_1)\, p(x_2)\, p(x_3)\, p(x_4 \mid x_1, x_2, x_3)\, p(x_5 \mid x_1, x_3)\, p(x_6 \mid x_4)\, p(x_7 \mid x_4, x_5) \quad \text{Equation 6.3}$$

The joint distribution of a DAG can be computed by taking the product, over all nodes in the graph, of a conditional distribution for each node conditioned on its parent variables in the graph. The joint distribution of a graph with K nodes can be represented as:

$$p(\mathbf{x}) = \prod_{k=1}^{K} p\left(x_k \mid \mathrm{pa}_k\right) \quad \text{Equation 6.4}$$

Where pa_k denotes the set of parents of variable x_k, and $x = \{x_1, \dots x_k\}$. A comprehensive treatment of graphical models can be found in the book by Whittaker (2009).

6.3.5 Structure Learning with the NOTEAR Algorithm

In the context of Bayesian Networks, structure learning refers to the process of determining the underlying structure or relationships between the variables in a network which a DAG represents. It involves discovering the connections between variables, such as which variables are dependent on which other variables and how these dependencies are related. This structure is crucial for accurately modelling the relationships between variables and making predictions about the system's behaviour described by the Bayesian Network. The previous section discussed a DAG's joint distribution and structure learning's crucial role in building a Bayesian Network. However, learning the structure of DAGs from data is an NP-hard problem (Chickering, 1996; Chickering, Heckerman and Meek, 2004) due to the exponential growth of the space of possible graphs with the increase in the number of nodes. In computational science, "NP-hard" refers to problems as hard as the most difficult problems in the class NP (nondeterministic polynomial time), for which a solution can be verified in polynomial time. If a problem is NP-hard, no known algorithm can solve all instances of the problem quickly.

To overcome this, the literature contains several exact algorithms for structure optimisation (Cooper and Herskovits, 1992; Jaakkola et al., 2010; Cussens, 2012; Malone et al., 2018). However, the traditional approaches, described by equation 5, for solving the combinatorial optimisation problem are inefficient due to the acyclicity of DAGs.

$$min\ F(W), subject\ to\ G(W) \in DAGs \qquad \text{Equation 6.5}$$

where $G(W)$is the d-node graph induced by the adjacency matrix $W, F: R^{d \times d}$, where R is a scoring function.

In this paper, the NOTEAR algorithm (Zheng et al., 2018) that formulates the structure learning problem as a purely continuous optimisation problem over real was used. This bypasses the traditional combinatorial constraint imposed by traditional Structure learning approaches and leads to improved performance. The optimisation problem using NOTEARS is characterised by Equation 6.6.

$$min\ F(W), subject\ to\ G(W) \in DAGs \qquad \text{Equation 6.6}$$

Equation 6.5 minimises a scoring function F(W) while satisfying the constraint G(w) = 0, where G is a smooth function over real matrices that exactly characterises acyclic graphs.

6.3.6 Learning Conditional Probability Distribution (CPD)

The Conditional Probability Distribution (CPD) is a crucial aspect to consider when constructing a Bayesian Network. The CPD of the links in the DAG, which represents the relationships between variables in the Bayesian Network, is learned through the use of a Bayesian Estimator with K2 Priors

In the case of pen fouling, understanding the CPD of the different variables' relationships can help identify the factors most likely to influence pen fouling. This information is useful for developing more precise predictive models of pen fouling and for evaluating the impact of various management practices and environmental factors.

To efficiently learn the CPD of the links in the DAG based on available data, the Bayesian Estimator employs a commonly used prior for Bayesian Network structure learning called the K2 Prior. The implementation of the Bayesian Estimator is available through the pgmpy python package, an open-source toolkit for working with Bayesian Networks (Ankan & Panda, 2015).

By combining structure learning with the NOTEAR algorithm and learning the

CPD with the Bayesian Estimator, a Bayesian Network can be accurately constructed. This network represents the relationships between variables and predicts the likelihood of pen fouling in pigs. This information can aid in informed decision-making in pig farm management and reduce the occurrence of pen fouling.

6.3.7 Scenario Simulation with Do-Calculus

Conducting experimental or randomised trials may not always be feasible or ethical, especially in animal welfare cases, such as the study of pen fouling in pigs. While traditional statistics provide insights into observational relationships, i.e., $p(y|x)$, they are not equipped to handle interventions and estimate the effect of intervening on a distribution, i.e., $p(y|do(x))$. For instance, if y represents the outcome of fouling, $do(x)\ denotes$ an action taken in the model.

The do-calculus theory was introduced by Pearl (1995) to address the issue of limited observation when studying causal effects of interventions. Predictions can be made using this concept regarding how a system will respond to hypothetical interventions, and the causal impact of such interventions can be estimated. In this research, the principles of do-calculus are applied to simulate different intervention scenarios and assess their effect on fouling.

The theorem of do-calculus states that if a set of variables Z satisfies the front-door criterion relative to the ordered pair of variables $(x, y)$and if $P(x, z)\ > 0$, then the causal effect of Xon Y can be determined using the following formula:

$$P(y \mid do(x)) = \sum_{z} P(z \mid x) \sum_{x'} P\left(y \mid x', z\right) P\left(x'\right) \qquad \text{Equation 6.7}$$

The Front-Door criterion is defined as follows: A set of Variables Z satisfies the front-door criterion relative to (X, Y) if:

- Z intercepts all directed paths from X to Y.
- There is no unblocked path from X to Z.
- All backdoor paths from Z to Y are blocked by X.

In simpler terms, Z serves as a mediator in the causal relationship between X and Y. If one can control for Z (the mediator), the causal effect of X on Y can be estimated, even in the presence of unobserved confounding variables that affect both X and Y. The Front Door Criterion proves useful in situations where it is not feasible to control for all confounding variables, a common scenario in real-world observational studies. The causal effect of interest can still be estimated by identifying a suitable mediator variable that satisfies the Front Door Criterion.

The do-calculus method was used for the study to create hypothetical intervention scenarios and analyse their effect on pig fouling. The discretised data was subjected to these interventions, as outlined in subsequent sections. The simulation was conducted using Python and the CausalNex library.

6.3.8 Feature Selection and Node Regularisation.

To improve the efficiency and accuracy of the Bayesian Network, two simplification steps were taken in this study. The first step was the use of LightGBM (Ke et al., 2017), a gradient-boosting framework that uses tree-based learning algorithms for feature selection. This reduces the dimension of the dataset and makes the task of fitting the Bayesian Network more efficient. This reduced the number of features in the dataset and made the task of fitting the Bayesian Network more manageable. The second step involved incorporating node regularisation in the NOTEARS algorithm using L2-Regularization (also known as Lasso). The steps taken in this study were summarised in Figure 6.4.

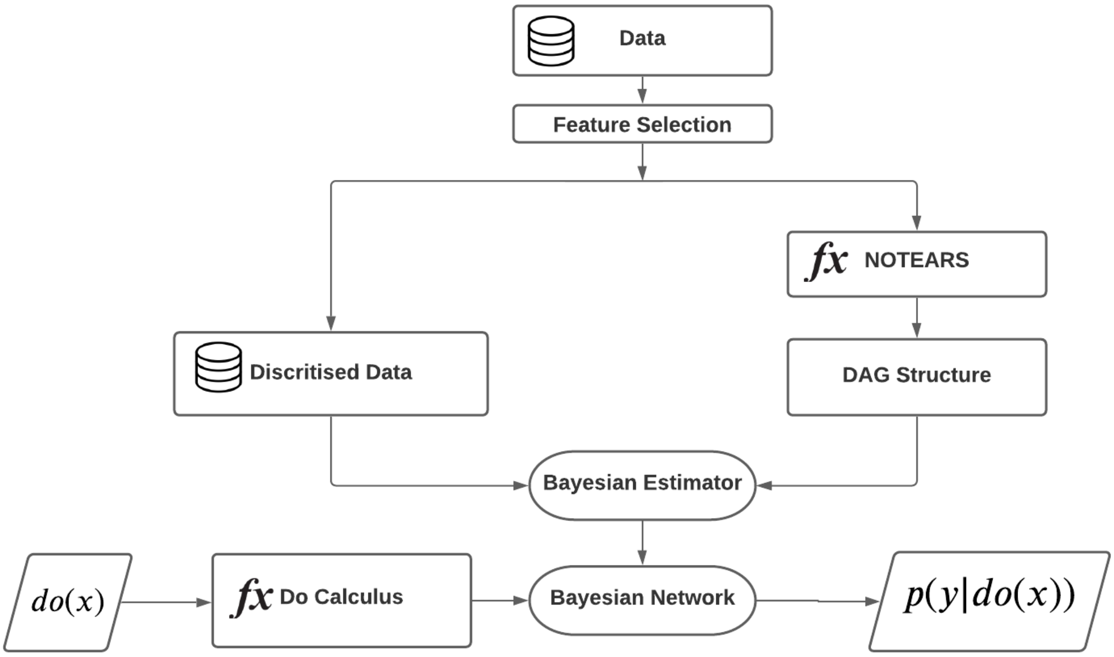

Figure 6.4: Summary of Data Processing and Simulation Pipeline.

6.3.9 Model Performance

To evaluate the performance of the Bayesian Network, the data was divided into a training and test set. The test set consisted of pigs in pens and time points that were not included in the training set. The pens were randomly selected, with 80% being used to fit the CPD of the Bayesian Network and 20% being used for testing the accuracy in predicting fouling. The pens were classified as either fouled or not fouled.

The performance of the Bayesian Network was measured using the area under the receiver operating characteristic curve (AUROC). The AUROC is a single metric that provides a comprehensive evaluation of the performance of an algorithm over the entire ROC space (Bradley, 1997). An AUROC of 0.5 indicates a model that makes random guesses, while an AUROC of 1.0 represents a perfect model.

6.3.10 Feature Selection

In this study, feature selection was performed using the LightGBM algorithm. The algorithm uses gradient-boosting tree-based learning algorithms. The feature selection process involved ranking the features based on their importance and selecting the top ten features that significantly impacted the prediction of pen fouling. In addition, the movements of pigs in the solid area were also selected to include a complete representation of the data. The selected features are shown in Figure 6.5, describing why they were deemed significant.

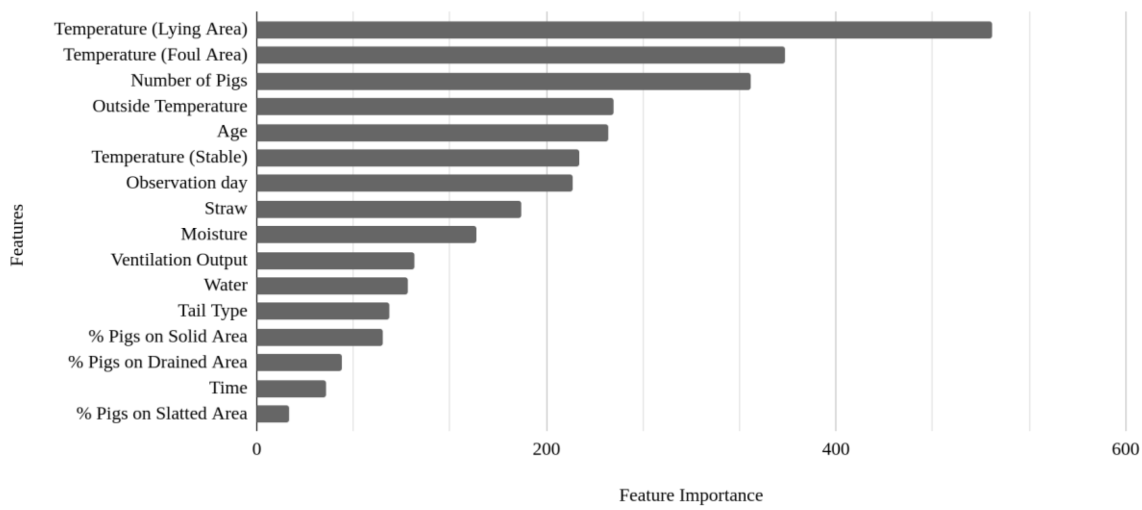

Figure 6.5: Summary of Feature Importance

6.3.11 Bayesian Structure

The Bayesian Network's Directed Acyclic Graph (DAG) was constructed using the NOTEARS algorithm and adhering to the constraints outlined in Table 6.1. Figure 6.6 displays the resulting DAG, including key variables identified through the feature selection process. The accuracy of the Bayesian Network was assessed by testing it against a pen data set, resulting in an AUROC value of 0.778. This indicates the network's efficacy in predicting whether pens belonged to the fouling or control group.

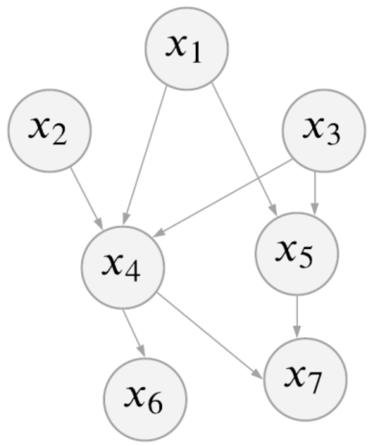

Figure 6.6: Example of a directed acyclic graph describing the joint distribution over variables $x_1, \ldots\ x_7$.

6.3.12 Scenarios Probability

The objective of this study was to evaluate the influence of various interventions on pen fouling. These interventions were based on the preventative measures recommended to mitigate fouling, as outlined in the work of Larsen et al. (2018). A total of eleven scenarios were tested, which are delineated in Table 6.2. The results of these scenarios, in comparison to the likelihood of fouling absent any intervention, are provided in Table 6.3a & 6,3b, and these outcomes are summarised in Table 6.4. To enhance the comprehensibility of the results, bar graphs and heat maps were employed to depict the impact of each intervention scenario on the probability of pen fouling. Furthermore, limitations or constraints in the interpretation of the findings were discussed.

6.4 Results and Discussion

The results of this study are presented in five tables, each providing a different perspective on the data and the outcomes of the scenarios tested.

Table 6.1: The table below provides a comprehensive overview of all variables used in the study. It includes information on the units of measurement, categorization, and hierarchical structure of each variable. For example, the 'Tail Type' variable has two options, 'Docked' represented by '0' and 'Undocked' represented by '1'. Additionally, the 'DAG Constraints' column indicates the relationship between variables, with 'Root' variables being independent and 'Leaf' variables having no dependencies. The 'Stable Temperature' sensor measures the ambient temperature controlled by the climate system, while 'Uniform Quantile (4 Levels)' categorizes continuous data into four equal-sized bins, ranging from 0 to 3.

DAG Constraints	Variable	Unite	Definition
Root	Tail Type	Binary	0 = Docked, 1= Undocked
Root	Time	Hour	Morning (0600 to 1200), Afternoon (1200 to 2200), night (2200 to 0600)
Root	Temperature (Stable)	Degrees	Uniform Quantile (4 Levels)
	Temperature (Fouling Area)	Degrees	Uniform Quantile (4 Levels)
	Temperature (Lying Area)	Degrees	Uniform Quantile (4 Levels)
Root	Moisture	% Relative Humidity	Uniform Quantile (4 Levels)
	Water	Litres	Uniform Quantile (4 Levels)
Root	Ventilation Need		Uniform Quantile (4 Levels)
Root	Age	Weeks after Insertion	Weaner (Age < 21 weeks), Grower (21 weeks <= Age, 41 weeks), Finisher (Age > 41 Weeks)
Root	Straw	Binary	0 =Yes, 1 = No
	% in Slated Area	%	Uniform Quantile (4 Levels)
	% in Lying Area	%	Uniform Quantile (4 Levels)
	% in Solid Area	%	Uniform Quantile (4 Levels)
Leaf	Event	Binary	0 = No Fouling, 1 = Fouling
	Group size (No Pigs)	Binary	Large (more than eleven pigs), Small (less or equal to 11)
	Obsday	Days Prior to Fouling	1 = 3-5 days before event, 2 = 1-2 days before event, 0 = day event occurred.

Table 6.1 defines all variables used in the study. For instance, the variable “Tail Type” is a binary variable, meaning it can take one of two values: zero for docked tails and 1 for undocked tails. Similarly, “Time” is measured in hours and is categorised into three periods: morning, afternoon, and night. This table is key to understanding the rest of the data in the subsequent tables. It is important to note that the variables are discretised into binary or quantile levels, simplifying the analysis and allowing for a clear interpretation of the results.

Table 6.2 outlines the various scenarios that were tested in the study. The scenarios are designed to test the effects of different environmental and management conditions on the probability of fouling. Each scenario represents a unique combination of conditions for the root variables. The ‘do(X)’ column indicates the specific intervention applied in each scenario. For example, in Scenario 1, the stable temperature was always kept below the 25th percentile. This is represented as $P(T_{Stable} = 0) = 1$,, meaning the probability of the stable temperature being in the lowest quantile (0) is one, or 100%.

Table 6.3a & 6.3b presents the conditional probabilities for all scenarios, including the baseline case. These probabilities represent the likelihood of a given variable assuming a particular value under each scenario. For example, row $P\left(T_{Lying} = 0\right)$ represents the conditional probability that the temperature in the lying area (T_{Lying}) is at level 0, given the conditions specified in each scenario. In the baseline scenario, P ($T_{Lying} = 0$) is 0.246, which means there is a 24.6% chance that the temperature in the lying area is at level 0 under normal conditions. In Scenario 1, P ($T_{Lying} = 0$) is 0.439, which means there is a 43.9% chance that the temperature in the lying area is at level 0 when the stable temperature is always kept below the 25th percentile. This suggests that lowering the stable temperature increases the likelihood of the temperature in the lying area being at level 0.

In Scenario 2, $P\left(T_{Lying} = 0\right)$ is 0.208, which means there is a 20.8% chance that the temperature in the lying area is at level 0 when the ventilation output is always kept below the 25th percentile. This suggests that lowering the ventilation output decreases the likelihood of the temperature in the lying area being at level 0. In Scenarios 3, 4, and 9, $P\left(T_{Lying} = 0\right)$ is one, which means there is a 100% chance that the temperature in the lying area is at level 0 under these scenarios. This indicates that these scenarios involve significant cooling of the lying area. Valuable insights can be gained from the probabilities listed here regarding the influence of environmental and management factors on the temperature of the lying area. By comparing these probabilities across various scenarios, the impact of different interventions on the temperature in that area can be assessed.

Table 6.2: Scenario Descriptions and Interventions: This table explains each test scenario and corresponding interventions, represented by do(X). For example, in Scenario 1, the stable temperature was consistently adjusted to remain below the 25th percentile.

Scenarios	**Description**	**do(X)**
Scenario 1	The stable temperature was lowered to always be below the 25th percentile.	$P(T_{Stable} = 0) = 1$
Scenario 2	The Ventilation Output was lowered to always be below the 25th percentile.	$P(V = 0) = 1$
Scenario 3	A temperature gradient was created between the Lying and Fouling Area. The temperature in the Lying Area lower to be below the 25th percentile, and the Temperature in Fouling was set to be above the 75th percentile)	$P(T_{Lying} = 0) = 1 \; AND \; P(T_{Foul} = 3) = 1$
Scenario 4	The temperature in both the Lying and Fouling areas was reduced to be below the 25th percentile.	$P(T_{Lying} = 0) = 1 \; AND \; P(T_{Foul} = 0) = 1$
Scenario 5	The number of pigs was lowered to always be equal to 11 pigs.	$P(NoPigs = Small) = 1$
Scenario 6	Extreme ventilation output was avoided by setting the ventilation output to always be between the 25th and 75th percentile values.	$P(V = 1) = 0.5 \; AND \; P(V = 2) = 0.5$
Scenario 7	A temperature gradient was created between the Lying and Fouling Area. The temperature in the Lying Area was lower to be below the 50th percentile, and the Temperature in Fouling was set to be above the 50th percentile).	$P(T_{Lying} = 0 \; or \; 1) = 1 \; AND \; P(T_{Foul} = 2 \; or \; 3) = 1$
Scenario 8	Limit temperature Lying and Fouling (between 25th and 75th percentile)	$P(T_{Lying} = 1 \; or \; 2) = 1 \; AND \; P(T_{Foul} = 1 \; or \; 2) = 1$
Scenario 9	Apply scenario (6 and 3)	
Scenario 10	All pens were set to contain straw.	$P(Straw = 1) = 1$
Scenario 11	Extreme Stable Temperature was avoided by setting the temperature between the 25th and 75th percentile.	$P(T_{Stable} = 1) = 0.5 \; AND \; P(T_{Stable} = 2) = 0.5$

The first four rows of the table represent the probabilities of the temperature in the lying area (T_{Lying}) being at different levels (0, 1, 2, 3). The baseline scenario shows an even distribution across all four levels, indicating that the temperature in the lying area can vary widely under normal conditions. However, under Scenarios 3, 4, and 9, the temperature in the lying area is always at level 0, suggesting that these scenarios involve significant cooling of the lying area. This could potentially be a strategy to reduce fouling, as lower temperatures might discourage pigs from fouling in the lying area. The next four rows represent the probabilities of the solid area being at different levels (0, 1, 2, 3). Again, the baseline scenario shows an even distribution across all four levels. None of the scenarios seems to significantly alter these probabilities, suggesting that the conditions specified in the scenarios do not strongly impact the solid area.

The rows for P (Event = Control) and P (Event = Foul) represent the probabilities of a control event (no fouling) and a fouling event, respectively. The baseline scenario shows a higher probability for a control event (0.557 or 55.7%) than for a fouling event (0.443 or 44.3%). However, under Scenario 2, the probability of a fouling event increases to 0.45 or 45%, indicating that reducing the ventilation output to always be below the 25th percentile (as specified in Scenario 2) may increase the risk of fouling. Conversely, under Scenario 6, the probability of a fouling event decreases to 0.381 or 38.1%, suggesting that avoiding extreme ventilation output (as specified in Scenario 6) may reduce the risk of fouling. The rows for P (NoPigs = Large) and P (NoPigs = Small) represent the probabilities of the number of pigs being large (more than 11) and small (less or equal to 11), respectively. The baseline scenario shows a higher probability for many pigs (0.615 or 61.5%) than for a small number of pigs (0.385 or 38.5%). However, under Scenario 5, the number of pigs is always small, suggesting that this scenario significantly reduces the number of pigs. This could be another strategy to reduce fouling, as fewer pigs might result in less fouling. Overall, the results from Tables 6.3a and 6.3b provide valuable insights into how different environmental and management conditions can influence the likelihood of various outcomes in pig pens, including the temperature in different areas, fouling events, and the number of pigs. These insights could be used to develop strategies to manage and reduce fouling in pig pens.

Table 6.3a: Conditional Probabilities for Scenarios and Baseline: The table presents the conditional probabilities for each variable across all scenarios, including the baseline. The notation p (K = x) is used to denote the probability of variable K taking on value x.

	Baseline	Scenario one	Scenario 2	Scenario 3	Scenario 4	Scenario 5	Scenario 6	Scenario 7	Scenario 8	Scenario 9	Scenario 10	Scenario 11
$P(T_{Lying} = 0)$	0.246	0.439	0.208	1	1	0.246	0.274	0.5	0.5	1	0.246	0.183
$P(T_{Lying} = 1)$	0.263	0.234	0.225	0	0	0.255	0.302	0.5	0.5	0	0.263	0.318
$P(T_{Lying} = 2)$	0.253	0.171	0.255	0	0	0.251	0.247	0	0	0	0.253	0.314
$P(T_{Lying} = 3)$	0.238	0.156	0.311	0	0	0.247	0.176	0	0	0	0.238	0.185
$P(Solid = 0)$	0.249	0.24	0.226	0.253	0.246	0.239	0.233	0.259	0.25	0.239	0.247	0.248
$P(Solid = 1)$	0.25	0.262	0.24	0.25	0.251	0.239	0.251	0.251	0.25	0.254	0.255	0.243
$P(Solid = 2)$	0.257	0.266	0.267	0.248	0.257	0.25	0.262	0.245	0.25	0.26	0.253	0.255
$P(Solid = 3)$	0.244	0.232	0.267	0.248	0.246	0.272	0.254	0.244	0.25	0.247	0.245	0.254
$P(Event = Control)$	0.557	0.563	0.55	0.606	0.606	0.566	0.619	0.572	0.586	0.644	0.557	0.556
$P(Event = Foul)$	0.443	0.437	0.45	0.394	0.394	0.434	0.381	0.428	0.414	0.356	0.443	0.444
$P(T_{Foul} = 0)$	0.262	0.539	0.214	0	1	0.275	0.362	0	0	1	0.262	0.188
$P(T_{Foul} = 1)$	0.293	0.186	0.265	0	0	0.279	0.344	0	0	0	0.293	0.424
$P(T_{Foul} = 2)$	0.233	0.169	0.251	0	0	0.266	0.209	0.5	0.5	0	0.233	0.279
$P(T_{Foul} = 3)$	0.211	0.106	0.27	1	0	0.181	0.085	0.5	0.5	0	0.211	0.109
$P(NoPigs = Large)$	0.615	0.615	0.615	0.615	0.615	0	0.615	0.615	0	0.615	0.615	0.615
$P(NoPigs = Small)$	0.385	0.385	0.385	0.385	0.385	1	0.385	0.385	1	0.385	0.385	0.385

Table 6.3b: Conditional Probabilities for Scenarios and Baseline: The table presents the conditional probabilities for each variable across all scenarios, including the baseline. The notation p (K = x) is used to denote the probability of variable K taking on value x.

	Baseline	Scenario one	Scenario 2	Scenario 3	Scenario 4	Scenario 5	Scenario 6	Scenario 7	Scenario 8	Scenario 9	Scenario 10	Scenario 11
$P(T_{Stable} = 0)$	0.248	1	0.248	0.248	0.248	0.248	0.248	0.248	0	0.248	0.248	0
$P(T_{Stable} = 1)$	0.248	0	0.248	0.248	0.248	0.248	0.248	0.248	0.5	0.248	0.248	0.5
$P(T_{Stable} = 2)$	0.252	0	0.252	0.252	0.252	0.252	0.252	0.252	0.5	0.252	0.252	0.5
$P(T_{Stable} = 3)$	0.252	0	0.252	0.252	0.252	0.252	0.252	0.252	0	0.252	0.252	0
$P(Obsday = 0)$	0.47	0.47	0.47	0.47	0.47	0.47	0.47	0.47	0.47	0.47	0.47	0.47
$P(Obsday = 1)$	0.326	0.326	0.326	0.326	0.326	0.326	0.326	0.326	0.326	0.326	0.326	0.326
$P(Obsday = 2)$	0.204	0.204	0.204	0.204	0.204	0.204	0.204	0.204	0.204	0.204	0.204	0.204
$P(Straw = 0)$	0.733	0.733	0.733	0.733	0.733	0.733	0.733	0.733	0	0.733	0	0.733
$P(Straw = 1)$	0.267	0.267	0.267	0.267	0.267	0.267	0.267	0.267	1	0.267	1	0.267
$P(Moisture = 0)$	0.251	0.251	0.251	0.251	0.251	0.251	0.251	0.251	0.251	0.251	0.251	0.251
$P(Moisture = 1)$	0.247	0.247	0.247	0.247	0.247	0.247	0.247	0.247	0.247	0.247	0.247	0.247
$P(Moisture = 2)$	0.251	0.251	0.251	0.251	0.251	0.251	0.251	0.251	0.251	0.251	0.251	0.251
$P(Moisture = 3)$	0.251	0.251	0.251	0.251	0.251	0.251	0.251	0.251	0.251	0.251	0.251	0.251
$P(V = 0)$	0.252	0.252	1	0.252	0.252	0.252	0	0.252	0	0	0.252	0.252
$P(V = 1)$	0.248	0.248	0	0.248	0.248	0.248	0.5	0.248	0.5	0.5	0.248	0.248
$P(V = 2)$	0.248	0.248	0	0.248	0.248	0.248	0.5	0.248	0.5	0.5	0.248	0.248
$P(V = 3)$	0.251	0.251	0	0.251	0.251	0.251	0	0.251	0	0	0.251	0.251

Table 6.4 summarises the effects of the different interventions on the probability of fouling. Compared to the baseline, the effects are measured as the increase in the probability of fouling when a certain intervention is applied. For example, in Scenario 6, where the ventilation need was modified, the probability of fouling decreased by 6.2% compared to the baseline scenario.

Table 6.5 compares the predictive performance of the Bayesian network model used in this study with other machine learning models reported in the literature. The performance is measured using the AUROC. For instance, the Bayesian network model used in this study achieved an AUROC of 0.778, indicating a 77.8% chance of correctly distinguishing between a fouling and a non-fouling event. However, the Gradient Boosted Decision Tree (LightGBM) model achieved a higher AUROC of 0.940, indicating a 94% chance of correctly distinguishing between a fouling and non-fouling event, thus outperforming the Bayesian network model.

Table 6.4: Effect Summary of Interventions on Probability of Fouling: The table summarises the impact of different interventions on the Probability of Fouling. The ‘Effect’ column provides the change in the Probability of Fouling when the respective intervention was applied, compared to the baseline. The table is arranged such that the top row indicates the intervention that caused the most significant decrease in the probability of fouling.

	Control Variable	**Effect**
Scenario 6	Ventilation Need	-6.2
Scenario 3	Temperature (Lying and Fouling Area)	-4.9
Scenario 4	Temperature (Lying and Fouling Area)	-4.9
Scenario 8	Temperature (Lying and Fouling Area)	-2.9
Scenario 7	Temperature (Lying and Fouling Area)	-1.5
Scenario 5	Pig Number	-0.9
Scenario 1	Temperature (Stable)	-0.6
Scenario 10	Straw	0
Scenario 11	Temperature (Stable)	0.1
Scenario 2	Ventilation Need	0.7
Scenario 9	Ventilation Need & Temperature (Lying and Fouling Area)	2.3

6.4.1 Comparing Model Performance with Literature

This study used two models to predict pen fouling in pigs: a Bayesian Network model and a LightGBM (Gradient Boosted Decision Tree) model. Both models have their strengths and weaknesses, reflected in their performance.

Table 6.5: Summary of predictive performance (AUROC) between several models. In this study, only show the best AUROC reported in each study are shown.

Models	Source	Climate Variables	Pen Level Information	Pen Characteristics	Activity	AUROC
Bayesian Network	This Study	Yes	Yes	Yes	Yes	0.778
Gradient Boosted Decision Tree (LightGBM)	This Study	Yes	Yes	Yes	Yes	0.940
Recurrent Neural Network	Literature (Domun et al. 2019) Chapter 4	Yes	Yes	Yes	No	0.820
Ensemble Learning	Literature (Jensen et al. 2020)	No	Yes	Yes	Yes	0.780

The graphical model known as Bayesian Networks is used to represent the dependencies among variables, providing a useful tool for modelling complex relationships and handling uncertainty. In the case of predicting pen fouling, the model was trained and tested on discrete data, resulting in a simplification of the probability space by reducing the number of possible outcomes or states that the variables in the model could take on. However, this simplification led to a trade-off in predictive performance, as evidenced by the AUROC of 0.778. While this score is respectable, it suggests that the model's ability to distinguish between fouling and non-fouling events could be further improved. It is possible that the use of discrete data may have limited the model's ability to capture the full complexity of the relationships among the variables, which could have contributed to the lower performance.

On the other hand, the LightGBM model is a gradient-boosting framework that uses tree-based learning algorithms. This model is known for its high performance and efficiency, and it can handle different types of data, including raw data. The LightGBM model was trained and tested on raw data using 10-fold cross-validation, a robust method for estimating the model's performance on unseen data. The LightGBM model achieved an AUROC of 0.940, indicating a high level of performance. Using raw data and a larger feature set likely contributed to this high performance, as these factors would allow the model to capture more complex relationships in the data. Using data up to 5 days before a fouling event would also increase the prediction accuracy, as it provides the model with more information about the conditions leading up to the event. While both models have their strengths, the LightGBM model outperformed the Bayesian Network model in this study.

Several studies have utilised different approaches and features to predict pen fouling in the literature. Jensen et al. (2020) used machine learning methods, specifically random forests, and artificial neural networks, to predict pen fouling days in advance based on the position of pigs within the pen at specific times of the day. This approach uses a similar feature set to the Bayesian Network model developed in this study but differs in the machine learning techniques. Their model achieved high predictive performance, although the exact AUROC was not reported.

Dominiak et al. (2019) presented a multivariate spatial dynamic linear model (DLM) predicting diarrhoea and pen-fouling outbreaks amongst growing pigs. This dual prediction approach differs significantly from this study, where the models were designed to predict only pen fouling. The DLM model's ability to predict two outcomes could impact its performance. On the one hand, the model might be more complex due to the need to distinguish between two different outcomes, which could decrease performance if the model becomes too complex to capture the underlying patterns in the data accurately. On the other hand, the ability to predict two outcomes might also provide the model with additional information that could improve its performance. For instance, if common factors influence diarrhoea and pen fouling, the model might leverage this information to make more accurate predictions. These factors could partly explain the difference in performance between the models from this study and the DLM model. The LightGBM model, designed to predict only pen fouling, achieved an AUROC of 0.940, slightly lower than the highest AUROC obtained by the DLM model (0.98 for weaners and 0.94 for finishers). This suggests that the ability to predict two outcomes might have provided the DLM model with additional information that improved its performance. However, it is also important to note that the DLM model uses sensor-based water data and applies a standardised two-sided Cusum on forecast errors generated by the model, which is a different approach from the one used in this study. These differences in data and methodology could also contribute to the differences in performance between the models.

In the study by Domun et al. (2019), a stacked bidirectional long short-term memory (LSTM) and feedforward neural network architecture were used to predict pen fouling, tail-biting, and diarrhoea in pigs. This approach fundamentally differs from this study, using a Bayesian Network model and a LightGBM (Gradient Boosted Decision Tree) model to predict only pen fouling. The LSTM model used by Domun et al. (2019) is a type of recurrent neural network capable of learning patterns in time-series data. This is particularly useful for predicting events like pen fouling, tail-biting, and diarrhoea, which are likely to be influenced by temporal patterns in the data. However, one potential drawback of this approach is that it requires a large amount of data to train effectively, and the features used by the LSTM models are abstract and can be challenging to interpret.

In contrast, the Bayesian Network model is a probabilistic graphical model representing conditional dependencies among variables. This approach allows us to model complex relationships and handle uncertainty, but the training process is computationally complex and is an NP-Hard problem. The LightGBM model, on the other hand, is a gradient-boosting framework that uses tree-based learning algorithms, which are more interpretable and can handle different types of data. In terms of performance, the LightGBM model achieved an AUROC of 0.940, which is higher than the AUROC obtained by the LSTM model used by Domun et al. (2019) to predict fouling (0.775). This suggests that the LightGBM model may be more effective at predicting pen fouling. However, it is worth noting that the LSTM model could also predict tail-biting and diarrhoea with high AUROC scores, demonstrating its versatility.

In conclusion, this study has demonstrated the potential of both the Bayesian Network and LightGBM models in predicting pen fouling in pigs, with the LightGBM model showing superior performance. The Bayesian Network model, while not as performant, offers valuable insights due to its ability to model complex relationships and handle uncertainty. However, the computational complexity of the Bayesian Network model, an NP-Hard problem, presents a significant challenge. Comparisons with other studies in the literature reveal a variety of approaches to predicting pen fouling, each with its strengths and weaknesses. Each approach has its strengths and weaknesses, and the choice between them may depend on the application's specific requirements, such as data availability, the need for interpretability, and the computational resources available.

6.4.2 The effect of temperature.

Our study highlights the critical role of temperature and ventilation in managing pen fouling, a significant factor impacting pig welfare and productivity. This research enhances our understanding of the environmental and management factors contributing to pen fouling and offers strategies to mitigate these factors effectively.

The relationship between temperature and fouling established in this study aligns with prior literature, thus further consolidating temperature as a pivotal determinant of fouling. Studies conducted by Huynh et al. (2005), Aarnink et al. (2006), Savary et al. (2009), and Spoolder et al. (2012) have all previously emphasised the role of temperature, particularly high ambient temperature, in contributing to pen fouling. Pigs' behavioural adaptation to elevated temperatures, precisely their propensity to lie on cooler surfaces such as slatted floors, has been recognised as a primary trigger for pen fouling. Our findings reaffirm this understanding, substantiating the temperature-fouling connection with empirical evidence.

In the first scenario when the temperature dropped below the 25th percentile, there was a decrease in the frequency of pen fouling incidents. This aligns with prior studies that highlight the role of temperature in mitigating pen fouling. Additionally, in Scenario 2, a slight uptick in fouling incidents was noticed when ventilation was reduced. This underscores the importance of ventilation in preventing fouling, an aspect that had been overlooked in earlier research. Our data strongly indicate that ventilation is a crucial factor in averting pen fouling.

Interestingly, Scenarios 3 and 7, which involved creating a temperature gradient between the Lying and Fouling Area, yielded a decrease in fouling events. These results align with prior studies, such as Hillmann et al. (2004), which highlighted the importance of effective temperature regulation across different areas within a pen. Furthermore, Scenario 4, which entailed a simultaneous reduction in temperature in both the Lying and Fouling areas, also resulted in a decrease in 'Fouling events. This further emphasised the role of temperature regulation in reducing pen fouling, a factor that is often underplayed but influential. Scenario 5, which involved always limiting the number of pigs to 11, did not yield a significant change in the probability of fouling events. This suggests that the occupancy rate of the pen might not be as impactful as the factors of temperature and ventilation in our experiment, based on our data set.

However, Scenarios 6 and 8, which involved controlling ventilation output and temperature in the Lying and Fouling area to always be between the 25th and 75th percentile, resulted in a substantial decrease in fouling events. This validates our previous observations about the significant role of temperature and ventilation in reducing fouling events. Interestingly, Scenario 9, a combination of Scenarios 6 and 3, yielded the most drastic reduction in fouling events. This illustrates the crucial need for a comprehensive approach to managing temperature and ventilation. In contrast, Scenarios 10 and 11, which involved always keeping straw in all pens and avoiding extremely stable temperatures, did not induce any significant changes in the probability of fouling events. This hints at the overriding influence of temperature and ventilation over other factors, such as straw availability and extreme stable temperatures.

In conclusion, the conditional probability table (Table 6.3a & 6.3b) results for all scenarios, including the Baseline case, reinforce the significant influence of temperature and ventilation in mitigating pen fouling. When combined with the outcomes from the various scenarios, the models' predictive capabilities provide practical and actionable strategies that farm managers can implement to boost pig welfare and productivity by effectively controlling these environmental factors.

While this study provides valuable insights into the relationships between temperature, ventilation, and pen fouling in pig farming, its limitations should be noted. These include a potential lack of generalizability due to the single-farm observational nature of this study and limitations in our modelling approach that preclude causal inferences or the consideration of interaction effects between variables. Measurement inaccuracies and unaccounted variables may have also influenced our findings. Lastly, this study did not entirely capture the complexity of pen fouling, influenced by multiple factors beyond temperature and ventilation, such as diet, pen design, and individual pig behaviours. These limitations underline the need for more comprehensive, experimental, and varied setting studies in the future.

6.4.3 Comfort Level of Lying Area and its Impact on Fouling

The impact of the comfort level of the designated lying area on pig behaviour and subsequent pen fouling is a critical consideration. Pigs will naturally gravitate towards more appealing environments; when the lying area is uncomfortable, pigs will seek out alternative spaces. Such behaviours often culminate in fouling events (Aarnink et al., 1996; Huynh and Aarnink, 2004). In certain situations, if the fouling area appears more inviting than the designated lying area, pigs may rest there, increasing pen fouling (Hillmann et al., 2004; Opderbeck et al., 2020).

Existing literature (Larsen, Bertelsen, and Pedersen, 2018) has pinpointed three primary strategies for enhancing the appeal of the lying area: (1) altering the flooring material (Minvielle and le Roux, 2009; Savary et al., 2009), (2) minimising draughts (Hacker et al., 1994; Huynh et al., 2005), and (3) rendering the dunging area less attractive (Aarnink et al., 1997). This study evaluated the effects of two techniques: supplying all pens with straw and reducing the ventilation needed to mitigate excessive drought conditions. The Bayesian Network analysis unveiled an interesting pattern: the introduction of straw influences the number of pigs in the lying area, but it did not directly affect the incidence of fouling.

Previous research has indicated that avoiding draughty areas is a common behaviour in pigs (Randall, Armsby, and Sharp, 1983). The analysis of the Bayesian Network supports these findings, as it demonstrates that proper ventilation directly impacts the temperature and fouling of pens. By minimising the need for ventilation to between the 25th and 75th percentiles, a reduction of 6.2% in the probability of fouling was observed. However, this also increased the probability of the temperature dropping below the 50th percentile in the lying and fouling areas by 6.8% and 15.1%, respectively. It is possible that the decrease in turbulence and increase in hot air and pen-level temperature, brought on by the limited need for ventilation, is responsible for this reduction in fouling.

Our study suggests that the combined strategies of reducing draught can effectively mitigate pig pen fouling. Nonetheless, future research is warranted to delve deeper into the most pragmatic and efficacious techniques to alleviate fouling and augment the comfort level of the lying area, aiming to improve pig welfare and productivity.

6.5 Conclusion and Future Works

This study presents a Bayesian Network model to predict pen fouling in pigs and compares its performance with other models in the literature. Results showed that the LightGBM model achieved the best performance, but the Bayesian Network approach has some strengths, such as the ability to model complex relationships and handle uncertainty. The Bayesian Network analysis revealed a strong relationship between the zonal temperatures and the number of pigs in the solid area. It showed that local temperature control is more effective in reducing the probability of fouling in pigs than overall stable temperature control.

The comfort level of the designated lying area also plays a crucial role in the behaviour of pigs, and literature has identified several techniques to make the lying area more appealing.

Future work can aim to improve the performance of the Bayesian Network model by incorporating additional features or improving the model's training process. Additionally, more research is needed to understand better the causal relationship between temperature and fouling in pigs and to explore other effective temperature control strategies, such as conducting experiments to compare the effects of different temperature control methods on pigs and to understand better the complex relationships between temperature and other variables such as ventilation and moisture. Further research could also aim to understand the impact of different flooring materials and pen design on the comfort level of the designated lying area and its effect on pen fouling.

Chapter 7 General Discussion and Conclusion

This section's examination of how specific behavioural events in commercially farmed pigs can be predicted by machine learning models and provide early warning signals is concluded. In the preceding chapters, various models that focused on important welfare issues, such as pen fouling, tail-biting, and health-related issues such as diarrhoea, were developed, validated, and compared. Informative and promising results were obtained, revealing previously unknown behavioural relationships and demonstrating the potential of advanced machine learning algorithms to predict specific events.

This final discussion summarises the insights gained from this research, and the strengths, limitations, and implications of the work done so far are considered. The reliability and accuracy of the models are evaluated, the methodologies used are critiqued, and the hypotheses are assessed based on the results. The challenges encountered, and the lessons learned throughout our journey are also explored based on these reflections.

Furthermore, the potential for future work in this area is discussed. What innovative techniques, novel sensors, or adaptive machine-learning models could improve the accuracy and robustness of event detection is considered. At the same time, the trajectory of these predictions towards an integrated early warning system with robust algorithms, state-of-the-art sensor arrays, and actionable forecasts, which will contribute significantly to farm management and animal welfare, is envisioned.

Evidence of the applicability of the research is provided, interpreting the implications of the findings of this thesis for commercial pig farming practices, and speculation on the future of precision animal farming enriched by machine learning and intelligent sensors is made.

7.1 Comparative Assessment of Different Approaches

This thesis uses distinctive data processing and analysis methodologies, contributing to the study’s forecast and interpretation of specific pig behaviours. These methodologies are compared and assessed herein, relating to their efficacy, robustness, and practical value in enhancing data interpretation and outcome prediction in precision pig farming.

The frequency analysis employed in Chapter 2 and system identification techniques revealed two distinct behavioural modes in pigs based on their water consumption and pen activity patterns. While these algorithms offer valuable insights into pig behaviour dynamics, they only serve as preliminary tools that skim the surface of the intricate connections between behavioural variables, lacking explicative power to capture deeper, more complex correlations.

Complementing fundamental behavioural analysis, advanced machine learning algorithms, introduced in Chapter 4, impart greater predictive accuracy. The stacked bidirectional long short-term memory and feedforward neural network architecture demonstrated high performance in automatically learning and classifying patterns in time series data related to pig behaviour. However, the architecture's high complexity and abstractness limit its accessibility and interpretability, which may pose challenges for its practical application in routine livestock management. This underlines the importance of balancing model complexity and interpretability in applied machine learning for precision livestock farming (Wang et al., 2018; Dominiak & Kristensen, 2017).

Chapter 5 introduces a hierarchical clustering approach, adopting a novel yet practical angle to handle multi-sensor data. The approach simplifies data analysis yet efficiently captures essential behavioural trends critical to pig welfare. However, while clustering offers an efficient way of dealing with extensive sensor data, it may still overlook subtle variations inherent to individual pigs or pens, potentially limiting the precision of predictions (Rashidi & Salah, 2021).

The Bayesian Network adopted in Chapter 6 is a significant leap towards integrating transparency into machine learning. Unlike the black-box nature that typifies many machine learning algorithms, Bayesian Networks provide an interpretable cause-and-effect relationship that is much easier to grasp. However, building a robust Bayesian Network requires expertise and an understanding of the variables under study, potentially limiting its broader application (Kalet, 2021).

This research's innovative approach lies in its ability to fuse different analytical methods - traditional statistical analyses with cutting-edge machine learning and Bayesian methods. Appropriate application of these methods can provide accurate behavioural predictions and decipher complex relationships between variables to provide usable and interpretable insights.

In conclusion, while each methodology has its merits and limitations, their combination in this thesis maximises complementary advantages and constitutes a significant step towards comprehensively predicting and understanding pig behaviours. These methodologies collectively set a cornerstone for an integrated early warning system for pig welfare issues, ushering in a transformative movement in smart pig farming. Future research should aim to refine these approaches further, developing more sophisticated, robust, and practicable predictive models that can be seamlessly integrated into regular farm management.

7.2 Advances in Machine Learning

The thesis explored using machine learning algorithms to predict and understand animal welfare issues in commercially farmed pigs. The models developed focused on three pivotal problems: pen fouling, tail-biting, and diarrhoea. Given behavioural data's intricate and non-stationary nature, these models were designed to extract features over time, revealing relationships between variables such as water consumption, temperature differences, and the spatial positioning of pigs with the welfare above issues.

The machine learning algorithms employed in this research were adept at processing extensive data from sensors monitoring water intake, temperature differences, and other farm-level metrics, generating valuable insights. These insights can potentially be instrumental for early predicting welfare problems, thereby bolstering animal health and welfare in pig farms. However, like many pioneering studies, this research also faced challenges, particularly concerning the complexity of data interpretation in machine learning models and the substantial data requirements for practical algorithm training.

Recent advancements in precision livestock farming (PLF) have been highlighted in a comprehensive review by Jiang et al., 2023, underscoring the evolutionary trends and development processes in PLF research, emphasising the growing significance of machine learning. The future of this domain holds immense promise. Enhanced feature engineering and extraction methodologies can refine the precision of machine learning algorithms. As demonstrated by Sozzi et al. (2022), the potential of deep learning in animal behaviour analysis suggests that more intricate and meaningful features can be extracted from collected data. Incorporating more granular data sources, such as individual pig-level data or specialised environmental conditions, could further augment the prediction accuracy of these models.

The rapid advancements in sensor technologies and the Internet of Things (IoT) are poised to bring transformative changes to data collection and precision farming practices. As highlighted by Ali et al. (2023), the increasing sophistication and affordability of on-farm sensors enable the integration of a more comprehensive range of sensor data in large-scale agricultural settings.

When combined with exploring alternative machine learning techniques such as reinforcement learning, this integration can pave the way for developing more nuanced and efficient models for smart farming.

Dynamic modelling approaches, which adapt and evolve, are gaining traction in precision farming. As Rokade et al. (2022) emphasised, integrating real-time data analytics, especially in environments like farming where conditions are in perpetual flux, is crucial. Such evolving systems, leveraging the power of regression-based supervised machine learning, can recalibrate their parameters based on the most recent data. This ensures robustness and enhances the accuracy of predictions, making them more attuned to the dynamic nature of agricultural settings.

The generalizability of these machine-learning models across diverse farms and environments is another frontier to be explored. Extensive research can be undertaken to test these models across varied farms, pig breeds, management practices, and environmental conditions. Such endeavours can amplify the models' real-world applicability and efficacy.

The interdisciplinary approach of this research, amalgamating insights from engineering, biology, and data science, sets a precedent for future studies. Such holistic approaches are indispensable for addressing the multifaceted animal welfare issue in controlled environments. In summation, while the models developed in this research exhibit promising predictive capabilities, the discussion underscores the vast potential for further exploration and creation of even more advanced models to enhance pig welfare in commercial farms.

The research's findings and prospects illuminate the path towards sustainable and ethical animal farming practices rooted in data insights and precision management. The developed models can be integrated into real-time on-farm early warning systems, facilitating timely interventions and enhancing overall animal welfare. The potential transformation of these algorithms into practical on-farm solutions can be immensely beneficial for both animals and farmers.

In conclusion, while this research has laid a solid foundation for using machine learning and neural networks for predicting specific pig behaviours, the complex nature of these methodologies can sometimes render outcomes challenging to interpret. This research paves the way for future studies to delve deeper into these methodologies, aiming to develop a reliable system capable of real-time predictions about animal welfare in commercial farming environments.

7.3 Sensor Analysis: Cost vs. Benefit

The increasing use of sensor technology in pig farming has opened a new era of possibilities for increasing productivity, improving animal welfare, and reducing environmental impact. However, the question of cost versus benefit remains a significant consideration for farmers in deciding whether to adopt these technologies.

Farmers have diverse needs and priorities, which can influence their perception of the value of different sensors. While weight and growth are primary concerns, other factors such as animal health and welfare, feed efficiency, and environmental sustainability are increasingly important. For instance, sensors that can monitor individual pig's feed intake and growth can provide valuable data for optimising feeding strategies, reducing feed costs, and improving growth rates. Similarly, sensors that detect early signs of disease or stress can help farmers take proactive measures to prevent disease spread, reduce mortality, and improve overall animal welfare.

The economic implications of addressing problems like pen fouling, tail biting, and diarrhoea are significant. These issues were highlighted because they are common problems in pig farming that can lead to significant economic losses. Pen fouling can lead to poor pen hygiene, increased risk of disease, and reduced growth rates. Tail biting can result in injury and infection, reducing growth and carcass quality. Diarrhoea can result in poor nutrient absorption, weight loss, and, in severe cases, death.

A cost/benefit analysis of using sensors to address these problems would need to consider the costs of the sensors, installation, and maintenance against the potential benefits of improved productivity and reduced losses due to disease. For example, if the use of sensors can reduce the incidence of diarrhoea by early detection and treatment, the savings in terms of reduced medication costs, improved growth rates, and reduced mortality could potentially outweigh the costs of the sensors.

In addition to these problems, farmers face other economically significant challenges, such as diseases that lead to poor growth, reproductive problems, and respiratory conditions. Sensors that can monitor individual pig's behaviour, physiological parameters such as heart rate and temperature, and environmental conditions such as temperature and humidity can provide valuable data for early detection of these problems.

In conclusion, while the cost of sensor technology can be high, the potential benefits of improved productivity, animal welfare, and environmental sustainability could make it a worthwhile investment for many farmers. However, more research is needed to fully understand the cost/benefit ratio of different sensors in different farming systems and to develop cost-effective sensor-based solutions that meet the specific needs of farmers.

7.4 Can Advance Algorithm make up for cheap sensors?

The interplay between advanced algorithms and inexpensive sensors is a thought-provoking subject in precision animal farming. Can a sophisticated algorithm compensatory bridge the gap created by the limitations of basic, economical sensors? Would farms be better served by investing in advanced analytics rather than splurging on high-end sensor technology?

From the research conducted in this thesis, a compelling case surfaces advocating the effectiveness of robust algorithms in synergising with simple, affordable sensors. Several attributes make advanced algorithms a worthy cornerstone of cost-effective precision farming strategies.

Advanced algorithms, such as machine learning models and predictive analytics, have demonstrated their potential to leverage the data provided by inexpensive sensors to extract meaningful information and make reliable predictions. These algorithms can adapt, learn, and improve their performance over time. Contrarily, high-end sensors, although providing higher precision measurements and more nuanced data, are a more static investment and could be susceptible to redundancy over time when technology advances rapidly.

Advanced algorithms can augment this noise limitation when inexpensive sensors generate high noise-level data. Notably, the applied algorithms in this research showed the impressive capability to filter noise and isolate patterns within the data, thereby drawing upon the existing sensor information more comprehensively. It suffices to say that robust algorithms can mitigate the lack of sensor precision to a substantial extent.

Expanded scalability is another advantage yielded by advanced algorithms. As algorithms can be manipulated, updated, and fine-tuned virtually, they offer ideal scalability. For instance, the machine learning model used in this thesis could learn as more data became available, facilitating continuous improvement and adaptation. Compared to the static nature of sensors, the scalability of algorithms skews the cost-benefit analysis in their favour.

From a financial perspective, investing in advanced algorithm development may be more economical than acquiring high-end sensors, particularly for small to medium-scale producers. Algorithm development incurs costs, but these models can be scaled once developed, with maintenance costs relatively minor. Conversely, procuring high-end sensors can represent a daunting upfront cost often not feasible for smaller-scale operations.

However, one should not rush to avoid the benefits offered by high-end sensors. They offer high-quality data, broader ranges of measurements, and greater precision — elements that can undeniably feed into more accurate predictions and insights. If available and economically viable, they could undeniably enhance the farm’s data ecosystem and, consequently, the accuracy of predictions.

In the real world, choices often boil down to resource availability. It is thus essential to evaluate where the investment would bring the most significant improvement. If resources permit, investing in high-end sensors and advanced algorithms would push the boundaries of precision farming and optimise results. Combining high-quality data from advanced sensors with predictive modelling introduces a powerful synergy offering extensive insights and predictions. However, under constrained resources, the course of investment becomes less straightforward.

The decision insidiously becomes a trade-off between the quality of data and the quality of analysis. In the scenario that only inexpensive sensors can be afforded, investment in advanced algorithms would be a wise course of action. The machine learning models developed in this research offer a testament to the prowess of algorithmic analysis. They derived meaningful patterns and predictive insights from data generated by relatively basic sensors, showcasing the potential of sophisticated algorithms to compensate for sensor constraints.

Contrarily, the situation could be more evident when the choice is between advanced sensors and basic analytics. While high-end sensors guarantee superior data, without powerful algorithms, the understanding that can be collected from this data may remain superficial. While advanced sensors provide more detailed, nuanced information, transferring this additional information into actionable insights may be hindered if accompanied only by basic analytical methodologies.

In most cases, an optimised approach may lie in striking a balance between both aspects - investing in moderately advanced sensors and developing corresponding analytical capabilities. As precision farming evolves, exploiting the benefits of improved sensing technology and increasingly sophisticated analytical algorithms becomes crucial.

In this age of data-driven agriculture, the synthesis of advanced analytics with sensor technology has the potential to trigger notable advancements in animal welfare and overall operational efficiency in pig farming. This work has underscored the potential of combining powerful algorithms with even inexpensive sensors to weave a better understanding of pig behaviours and predict future outcomes.

However, it is essential to note that this is not a case of ‘either-or.’ Advanced sensors and complex algorithms have unique advantages, and their combined efforts offer the highest potential to enhance animal welfare and farm productivity. Each farming operation must examine its unique context, goals, and resource constraints and decide to invest in sensor hardware and/or analytical sophistication to achieve optimal benefits.

7.5 Conclusion

In conclusion, this thesis significantly contributes to the growing discourse on the application of machine learning in precision livestock farming. It introduced and validated various algorithms’ predictive proficiency for welfare and health issues such as pen fouling, tail-biting, and diarrhoea within commercial pig farms. Although each examined model and methodology demonstrated distinctive strengths and limitations, collectively, they fostered a comprehensive understanding of pig behaviours.

Furthermore, this research underscored the value of future exploration of machine learning models, enhanced sensor technologies and IoT devices to upgrade the scope and accuracy of intelligent pig farming. The endeavour to develop a reliable real-time early warning system for pig welfare issues may benefit from the building blocks laid by this study. Detailed and extensive testing across diverse farm environments remains a promising area for future research.

The study also raised pertinent questions regarding the economic feasibility of sensor technology in commercial pig farming, highlighting the need to analyse cost against potential benefits in yield productivity, improved animal welfare, and environmental sustainability.

This research, hence, serves as a reliable foundation and steppingstone for the future development of precision pig farming, propelled by machine learning and enhanced sensor technology.

References

Aarnink, A. J. A., Schrama, J. W., Verheijen, E. J. E., & Stefanowska, J. (2001). Pen fouling in pig houses affected by temperature. In 6th Int. Symp. on Livestock Environment, St. Joseph Mich. USA / ASAE (pp. 180-186). https://doi.org/10.2527/jas.2005-521

Aarnink, A. J. A., Schrama, J. W., Heetkamp, M. J. W., Stefanowska, J., & Huynh, T. T. T. (2006). Temperature and body weight affect fouling of pig pens. Journal of Animal Science. https://doi.org/10.2527/jas.2005-521

Aarnink, A. J. A., van den Berg, A. J., Keen, A., Hoeksma, P., & Verstegen, M. W. (1996). Effect of Slatted Floor Area on Ammonia Emission and on the Excretory and Lying Behaviour of Growing Pigs. Journal of Agricultural Engineering Research, 64, 299–310. https://doi.org/10.1006/JAER.1996.0071

Aarnink, A. J. A., et al. (1997). Effect of Type of Slatted Floor and Degree of Fouling of Solid Floor on Ammonia Emission Rates from Fattening Piggeries. Journal of Agricultural Engineering Research, 66(2), 93–102. https://doi.org/10.1006/jaer.1996.0121

Aerts, J.-M., Berckmans, D., Saevels, P., Decuypere, E., & Buyse, J. (2000). Modelling the static and dynamic responses of total heat production of broiler chickens to step changes in air temperature and light intensity. British Poultry Science, 41(5), 651–659. https://doi.org/10.1080/0007166002000920 7

Aerts, J.-M., Wathes, C. M. M., & Berckmans, D. (2003). Dynamic data-based modelling of heat production and growth of broiler chickens: Development of an integrated management system. Biosystems Engineering, 84(3), 257–266. https://doi.org/10.1016/S1537-5110(02)00285-4

Ali, M. Y., Alsaeedi, A., Shah, S. A. A., Yafooz, W. M. S., & Malik, A. W. (2023). Energy Efficient Data Dissemination for Large-Scale Smart Farming Using Reinforcement Learning. Electronics 2023, 12(5), 1248. https://doi.org/10.3390/ELECTRONICS12051248

Andersen, H. M.-L., Dybkjær, L., & Herskin, M. S. (2014). Growing pigs' drinking behaviour: number of visits, duration, water intake and diurnal variation. Animal, 8, 1881–1888. https://doi.org/10.1017/S175173111400192X

Ankan, A., & Panda, A. (2015). pgmpy: Probabilistic graphical models using python. In Proceedings of the 14th Python in Science Conference (SCIPY 2015). Citeseer. Available at: http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.986.409&rep=rep1&ty

pe=pdf.

ASABE. (2012). Design of Ventilation Systems for Poultry and Livestock Shelters (Vol. ASAE EP270.5 DEC1986 (R2012) pp. 33).

Averos, X., Herranz, A., Sanchez, R., Comella, J. X., & Gosalvez, L. F. (2008). Serum stress parameters in pigs transported to slaughter under commercial conditions in different seasons. Veterinary Medicine (Praha), 52, 333–342. https://doi.org/10.17221/1874-VETMED

Baker, J. E. E. (2004). Effective environmental temperature. Journal of Swine Health and Production, 12(3), 140–143.

Bang-Jensen, J., & Gutin, G. Z. (2009). Digraphs. https://doi.org/10.1007/978-1-84800-998-1

Banhazi, T. M., Dunn, M., Cook, P., Black, J., Durack, M., & Johnnson, I. (2007). Development of precision livestock farming (PLF) technologies for the Australian pig industry. Paper Presented at the Precision Livestock Farming ’07, Skiathos, Greece.

Banhazi, T. M., Seedorf, J., Rutley, D. L., & Pitchford, W. S. (2008). Identification of risk factors for sub-optimal housing conditions in Australian piggeries: Part 2. Airbone pollutants. Journal of Agricultural Safety and Health, 14(1), 21–39.

Bench, C.J., Rioja-Lang, F.C., Hayne, S.M., Gonyou, H.W., (2013). Group gestation sow housing with individual feeding-II: How space allowance, group size and composition, and flooring affect sow welfare. Livest. Sci. 152, 218–227. https://doi.org/10.1016/j.livsci.2012.12.020

Bengio, Y., Simard, P., Frasconi, P., (1994). Learning long-term dependencies with gradient descent is difficult. IEEE Trans. Neural Networks 5, 157–166. https://doi.org/10.1109/72.279181

Bergstra, J. & Bengio, Y. (2012). Random Search for Hyper-Parameter Optimization. Journal of Machine Learning Research, 13, 281-305. https://dl.acm.org/doi/10.5555/2188385.2188395

Bergstra, J., Bardenet, R., Bengio, Y., & Kégl, B. (2011). Algorithms for hyper-parameter optimisation. In Proceedings of the 24th International Conference on Neural Information Processing Systems (NIPS'11) (pp. 2546–2554). Red Hook, NY, USA: Curran Associates Inc. https://papers.nips.cc/paper_files/paper/2011/hash/86e8f7ab32cfd12577bc2619bc635690-Abstract.html

Bergstra, J., Yamins, D., Cox, D.D., (2013). Hyperopt: A Python Library for Optimising the Hyperparameters of Machine Learning Algorithms. PROC. 12th PYTHON Sci. CONF. https://github.com/hyperopt/hyperopt

Bertelsen, M., Pedersen, L.J., Larsen, M.L. V., (2017). Review. Factors affecting fouling in conventional pens for slaughter pigs. Submitt. Publ. 1–7. https://doi.org/10.1017/S1751731117001586

Bigelow, J.A., Houpt, T.R., (1988). Feeding and drinking patterns in young pigs. Physiol. Behav. 43, 99–109. https://doi.org/10.1016/0031-9384(88)90104-7

Black, J. L. L. (2014). Brief history and future of animal simulation models for science and application. Animal Production Science, 54(11–12), 1883–1895. https://doi.org/10.1071/AN14650

Black, J. L., Campbell, R. G., Williams, I. H., James, K. J., & Davies, G. T. (1986). Simulation of energy and amino acid utilization in the pig. Research and Development in Agriculture, 3(3), 121–145.

Bolhuis, J.E., Schouten, W.G.P., Leeuw, J.A. De, Schrama, J.W., Wiegant, V.M., (2004). Individual coping characteristics, rearing conditions, and behavioural flexibility in pigs. Behav. Brain Res. 152, 351–360. https://doi.org/10.1016/j.bbr.2003.10.024

Bolhuis, J.E., Schouten, W.G.P., Schrama, J.W., Wiegant, V.M., (2006). Effects of rearing and housing environment on behaviour and performance of pigs with different coping characteristics. Appl. Anim. Behav. Sci. 101, 68–85. https://doi.org/10.1016/j.applanim.2006.01.001

Bond, T. E., Heitman Jr., H., & Kelly, C. F. (1965). Effects of increased air velocities on heat and moisture loss and growth of swine. Transactions of the ASAE, 8(2), 167–174.

Bond, T. E., Kelly, C. F., & Heitman Jr., H. (1952). Heat and moisture loss from swine. Agricultural Engineering, 33, 148–152.

Bond, T. E., Kelly, C. F., & Heitman Jr., H. (1959). Hog house air conditioning and ventilation data. Transactions of the ASAE, 2(1), 1–4.

Børge, D., Kristensen, A.R., (2016). Temperature as a predictor of fouling and diarrhoea in slaughter pigs. Livest. Sci. 183, 1–3. https://doi.org/10.1016/j.livsci.2015.11.007

Bornett, H. L. I., Morgan, C. A., Lawrence, A. B., & Mann, J. (2000). The effect of group housing on feeding patterns and social behaviour of previously individually housed growing pigs. Applied Animal Behaviour Science, 70(2), 127–141. https://doi.org/10.1016/S0168-1591(00)00146-5

Bracke, M. B. M. (2011). Review of wallowing in pigs: Description of the behaviour and its motivational basis. Applied Animal Behaviour Science, 132(1–2), 1-13. ISSN 0168-1591. https://doi.org/10.1016/j.applanim.2011.01.002

Bradley, A. P. (1997) 'The use of the area under the ROC curve in the evaluation of machine learning algorithms', Pattern recognition, 30(7), pp. 1145–1159.

Brehme, U., Stollberg, E., Holz, R., Schleusener, T., (2004). Safer oestrus detection with sensor aided ALT-pedometer. Pap. Present. Proc. B. Abstr. Third Int. Work. Smart Sensors Livest. Monit. Leuven, Belgium.

Breiman, L. Random Forests. Machine Learning 45, 5–32 (2001). https://doi.org/10.1023/A:1010933404324

Breiman, L., Friedman, J.H., Olshen, R.A., Stone, C.J., (1984). Classification and Regression Trees, The Wadsworth statisticsprobability series. https://doi.org/10.1371/journal.pone.0015807

Bridges, T. C., & Gates, R. S. (2009). Modeling of animal bioenergetics for environmental management applications. Livestock Energetics and Thermal Environmental Management, 151–179.

Brody, S. (1945). Bioenergetics and Growth, with Special Reference to the Efficiency Complex in Domestic Animals. Bioenergetics and Growth, with Special Reference to the Efficiency Complex in Domestic Animals.

Brown-Brandl, T. M. M. (2013). Managing thermal stress in feedlot cattle: Environment, animal susceptibility and management options from a US perspective. In Livestock Housing: Modern Management to Ensure Optimal Health and Welfare of Farm Animals (pp. 189–208). https://doi.org/10.3920/978-90-8686-771-4_10

Brown-Brandl, T. M. M., Nienaber, J. A. A., Xin, H., & Gates, R. S. S. (2004). A literature review of swine heat production. Transactions of the American Society of Agricultural Engineers, 47(1), 259–270. https://doi.org/10.13031/2013.15867

Brown-Brandl, T.M.M., (2013). Managing thermal stress in feedlot cattle: Environment, animal susceptibility and management options from a U.S. perspective, in: Livestock Housing: Modern Management to Ensure Optimal Health and Welfare of Farm Animals. pp. 189–208. https://doi.org/10.3920/978-90-8686-771-4_10

Bruce, J. M. M., & Clark, J. J. J. (1979). Models of heat production and critical temperature for growing pigs. Animal Production, 28(3), 353–369. https://doi.org/10.1017/S0003356100023266

Buffington, D. E., Collazo-Arocho, A., Canton, G. H., Pitt, D., Thatcher, W. W., & Collier, R. J. (1981). Black globe-humidity index (BGHI) as comfort equation for dairy cows. Transaction of the ASAE, 24(3), 711–714. https://doi.org/10.13031/2013.34325

Chen, S. H. and Pollino, C. A. (2012) 'Good practice in Bayesian network modelling', Environmental Modelling & Software, pp. 134–145. doi: 10.1016/j.envsoft.2012.03.012.

Chepete, H. J., & Xin, H. (2001). Heat and moisture production of poultry and their housing systems – A literature review. Paper Presented at the 6th International Symposium on Livestock Environment, Louisville, KY, USA. https://dr.lib.iastate.edu/handle/20.500.12876/179

Chickering, D. M. (1996) Learning Bayesian Networks is NP-Complete, in Fisher, D. and Lenz, H.-J. (eds) Learning from Data: Artificial Intelligence and Statistics V. New York, NY: Springer New York, pp. 121–130.

Chickering, D. M., Heckerman, D. and Meek, C. (2004) 'Large-Sample Learning of Bayesian Networks is NP-Hard', Journal of machine learning research: JMLR, 5(Oct), pp. 1287–1330. https://arxiv.org/abs/1212.2468

Chollet, F., & others. (2015). Keras. GitHub. Retrieved from https://github.com/fchollet/keras.

Chollet, F., (2017). Deep Learning with Python, 1st ed. Manning Publications Co., Greenwich, CT, U.S.A.

Ciresan, D., Meier, U., Schmidhuber, J., (2012). Multi-column deep neural networks for image classification, in 2012 IEEE Conference on Computer Vision and Pattern Recognition. IEEE, pp. 3642–3649. https://doi.org/10.1109/CVPR.2012.6248110

Close, W. H. H., & Mount, L. E. E. (1978). The effects of plane of nutrition and environmental temperature on the energy metabolism of the growing pig. British Journal of Nutrition, 40(3), 413–421. https://doi.org/10.1079/BJN19780142

Cobb, B. R., Rumí, R. and Salmerón, A. (2007). Bayesian Network Models with Discrete and Continuous Variables, in Lucas, P., Gámez, J. A., and Salmerón, A. (eds) Advances in Probabilistic Graphical Models. Berlin, Heidelberg: Springer

Berlin Heidelberg, pp. 81–102. https://link.springer.com/chapter/10.1007/978-3-540-68996-6_4

Cooley, J. W., & Tukey, J. W. (1964). An Algorithm for the Machine Calculation of Complex Fourier Series. https://doi.org/10.2307/2003354

Cooper, G. F. and Herskovits, E. (1992). A Bayesian method for the induction of probabilistic networks from data, Machine Learning, pp. 309–347. doi: 10.1007/bf00994110.

Cross, A.J., Rohrer, G.A., Brown-Brandl, T.M., Cassady, J.P., Keel, B.N., (2018). Feed-forward and generalised regression neural networks in modelling feeding behaviour of pigs in the grow-finish phase. Biosyst. Eng. https://doi.org/10.1016/J.BIOSYSTEMSENG.2018.02.005

Cui, Z., Ke, R., Pu, Z., & Wang, Y. (2019). Deep Bidirectional and Unidirectional LSTM Recurrent Neural Network for Network-wide Traffic Speed Prediction. arXiv preprint arXiv:1801.02143.

Cussens, J. (2012) Bayesian network learning with cutting planes, arXiv [cs.AI]. Available at: http://arxiv.org/abs/1202.3713.

Da Silva, R. G. G., Morais, D. A. E. F., & Guilhermino, M. M. M. (2007). Evaluation of thermal stress indexes for dairy cows in tropical regions. Revista Brasileira de Zootecnia, 36(4 SUPPLEMENT), 1192–1198. https://doi.org/10.1590/S1516-35982007000500028

Daniel McFadden, (1975). Conditional logit analysis of qualitative choice behaviour., Frontiers in Economics. P. Zarembka, eds., New York.

Darr, M., & Epperson, W. (2009). Embedded sensor technology for real time determination of animal lying time. Computers and Electronics in Agriculture, 66(1), 106–111. https://doi.org/10.1016/j.compag.2009.01.004

Dawid, A. P. (2002). Influence Diagrams for Causal Modelling and Inference. International Statistical Review, 70(2), 161–189. https://doi.org/10.1111/J.1751-5823.2002.TB00354.X

Deaton, J. W., Reece, F. N., & Bouchillon, C. W. (1969). Heat and moisture production of broilers: Part 2. Winter conditions. Poultry Science, 48, 1579–1582. https://doi.org/10.3382/ps.0481579

Domun, Y. et al. (2019) Learning patterns from time-series data to discriminate predictions of tail-biting, fouling and diarrhoea in pigs, Computers and Electronics in Agriculture, 163, p. 104878.

https://doi.org/10.1016/j.compag.2019.104878

Drury, B. et al. (2017) A survey of the applications of Bayesian networks in agriculture', Engineering applications of artificial intelligence, 65, pp. 29–42. https://doi.org/10.1016/j.engappai.2017.07.003

Escobar, J., Van Alstine, W.G., Baker, D.H., Johnson, R.W., (2007). Behaviour of pigs with viral and bacterial pneumonia. Appl. Anim. Behav. Sci. 105, 42–50. https://doi.org/10.1016/j.applanim.2006.06.005

Exadaktylos, V., Silva, M., Aerts, J.-M., Taylor, C.J., Berckmans, D., (2008). Real-time recognition of sick pig cough sounds. Comput. Electron. Agric. 63, 207–214. https://doi.org/10.1016/j.compag.2008.02.010

Ferro, N. D., Mencaroni, M., Fabbri, G., Gottardo, F., Lazzaro, B., & Morari, F. (2023). How to select Best Available Techniques to reduce NH3 emissions from the agricultural sector? Results from integrating the pillars of sustainability into a Bayesian Belief Network model. EGU23. https://doi.org/10.5194/EGUSPHERE-EGU23-11378

Fluss, R., Faraggi, D., & Reiser, B. (2005). Estimation of the Youden Index and its Associated Cutoff Point. Biometrical Journal, 47(4), 458-472. https://doi.org/10.1002/bimj.200410135

Fournel, S., Rousseau, A. N., & Laberge, B. (2017). Rethinking environment control strategy of confined animal housing systems through precision livestock farming. Biosystems Engineering, 155, 96–123. https://doi.org/10.1016/j.biosystemseng.2016.12.005

Fraser, D. (2003). Assessing Animal Welfare at the Farm and Group Level: The Interplay of Science and Values. Animal Welfare, 12(4), 433–443. https://doi.org/10.1017/S0962728600026038

Fraser, D., (1985). Selection of bedded and unbedded areas by pigs in relation to environmental temperature and behaviour. Appl. Anim. Behav. Sci. 14, 117–126. https://doi.org/10.1016/0168-1591(85)90023-1

Friedman, N., Geiger, D. and Goldszmidt, M. (1997) 'Bayesian Network Classifiers', Machine learning, 29(2), pp. 131–163. https://doi.org/10.1023/a:1007465528199

Frost, A. R., Schofield, C. P., Beaulah, S. A., Mottram, T. T., Lines, J. A., & Wathes, C. M. (1997). A review of livestock monitoring and the need for integrated systems. Computers and electronics in agriculture, 17(2), 139-159.

https://doi.org/10.1016/S0168-1699(96)01301-4

Gallo, C., Véjar, L., Galindo, F., Huertas, S. M., & Tadich, T. (2022). Animal welfare in Latin America: Trends and characteristics of scientific publications. Frontiers in Veterinary Science, 9, 1030454. https://doi.org/10.3389/FVETS.2022.1030454/BIBTEX

Gaughan, J. B., Mader, T. L., & Gebremedhin, K. G. (2012). Rethinking Heat Index Tools for Livestock. In Environmental Physiology of Livestock (pp. 243–265). https://doi.org/10.1002/9781119949091.ch14

Gebremedhin, K. G., & Wu, B. (2005). Simulation of flow field of a ventilated and occupied animal space with different inlet and outlet conditions. Journal of Thermal Biology, 30(5), 343–353. https://doi.org/10.1016/j.jtherbio.2004.10.001

Geers, R., & Bosschaerts, P. W. L. (1989). Integrated control of air and floor temperature in piglet houses: animal and engineering aspects. https://doi.org/10.1051/animres:19900103

Geers, R., Goedseels, V., Parduyns, G., & Vercruysse, G. (1986). The group postural behaviour of growing pigs in relation to air velocity, air, and floor temperature. Applied Animal Behaviour Science, 16(4), 353–362. https://doi.org/10.1016/0168-1591(86)90007-9

Glorot, X. & Bengio, Y. (2010). Understanding the difficulty of training deep feed-forward neural networks. Proceedings of the Thirteenth International Conference on Artificial Intelligence and Statistics. Proceedings of Machine Learning Research. Available from https://proceedings.mlr.press/v9/glorot10a.html.

González, L.A., Tolkamp, B.J., Coffey, M.P., Ferret, A., Kyriazakis, I., (2008). Changes in Feeding Behavior as Possible Indicators for the Automatic Monitoring of Health Disorders in Dairy Cows. J. Dairy Sci. 91, 1017–1028. https://doi.org/10.3168/jds.2007-0530

Goodfellow, I., Bengio, Y., & Courville, A. (2016). Deep Learning. MIT Press. https://doi.org/10.1007/s10710-017-9314-z

Graves, A., & Schmidhuber, J. (2008). Offline Handwriting Recognition with Multidimensional Recurrent Neural Networks. In D. Koller, D. Schuurmans, Y. Bengio, & L. Bottou (Eds.), Advances in Neural Information Processing Systems (Vol. 21). Curran Associates, Inc. https://doi.org/10.1007/978-1-4471-4072-6_12

Graves, A., (2012). Supervised sequence labelling with recurrent neural networks. Springer. https://doi.org/10.1007/978-3-642-24797-2_2

Graves, A., Mohamed, A., Hinton, G., (2013). Speech recognition with deep recurrent neural networks, in 2013 IEEE International Conference on Acoustics, Speech and Signal Processing. IEEE, pp. 6645–6649. https://doi.org/10.1109/ICASSP.2013.6638947

Green, A. R., & Xin, H. (2009). Effects of stocking density and group size on thermoregulatory responses of laying hens under heat-challenging conditions. Transactions of the ASABE, 52(6), 2033–2038. https://www.scopus.com/inward/record.uri?eid=2-s2.0-76249093368&partnerID=40&md5=36b8ec03e8626a9229e060aa5faa58a4

Green, D. M., & Parsons, D. J. (2006). The place of models in the new technologies of production systems. In Mechanistic Modelling in Pig and Poultry Production (pp. 305–324). https://doi.org/10.13031/2013.29205

Greff, K., Srivastava, R.K., Koutník, J., Steunebrink, B.R., Schmidhuber, J., (2015). LSTM: A Search Space Odyssey. https://doi.org/10.1109/TNNLS.2016.2582924

Gron, A., (2017). Hands-On Machine Learning with Scikit-Learn and TensorFlow: Concepts, Tools, and Techniques to Build Intelligent Systems, 1st ed. O'Reilly Media, Inc.

Hacker, R. R. et al. (1994) 'Factors affecting excretory behavior of pigs', Journal of animal science, 72(6), pp. 1455–1460. https://doi.org/10.2527/1994.7261455x

Hahn, G. L., Gaughan, J. B., Mader, T. L., & Eigenberg, R. A. (2009). Thermal indices and their applications for livestock environments. Livestock Energetics and Thermal Environmental Management, 113–130.

Hansen, M.F., Smith, M.L., Smith, L.N., Salter, M.G., Baxter, E.M., Farish, M., Grieve, B., (2018). Towards on-farm pig face recognition using convolutional neural networks. Comput. Ind. 98, 145–152. https://doi.org/10.1016/j.compind.2018.02.016

Hearst, M.A., Dumais, S.T., Osuna, E., Platt, J., Scholkopf, B., (1998). Support vector machines. IEEE Intell. Syst. their Appl. 13, 18–28. https://doi.org/10.1109/5254.708428

Heitman, H., Kelly, C. F., & Bond, T. E. (1958). Ambient air temperature and weight gain in swine. J. Anim. Sci., 17(1), 62–67. https://doi.org/10.2527/jas1958.17162x

Hemsworth, P.H., Coleman, G.J., Barnett, J.L., Borg, S., (2000). Relationships between human-animal interactions and productivity of commercial dairy cows. J. Anim. Sci. 78, 2821. https://doi.org/10.2527/2000.78112821x

Hillman, P. E., Gebremedhin, K. G., Aneshansley, D., & Landers, A. (2000). Design of a new cooling system for dairy cows in freestall facilities. 2000 ASAE Annual International Meeting, Technical Papers: Engineering Solutions for a New Century, 2,4705–4715.https://www.scopus.com/inward/record.uri?eid=2-s2.0-2342459115&partnerID=40&md5=d9488eaa4ef6b0e97b1da074dd3a0a74

Hillmann, E., Mayer, C., Gygax, L., & Schrader, L. (2005). Effects of space allowance on behavioural and adrenocortical reactions to elevated temperatures in fattening pigs**.** https://api.semanticscholar.org/CorpusID:86848124

Hillmann, E., Mayer, C., Schön, P.C., Puppe, B., Schrader, L., (2004). Vocalisation of domestic pigs (Sus scrofa domestica) as an indicator for their adaptation towards ambient temperatures. Appl. Anim. Behav. Sci. 89, 195–206. https://doi.org/10.1016/j.applanim.2004.06.008

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

Hogeveen, H., Kamphuis, C., Steeneveld, W., Mollenhorst, H., (2010). Sensors and Clinical Mastitis—The Quest for the Perfect Alert. Sensors 10, 7991–8009. https://doi.org/10.3390/s100907991

Holmes, C. W., & Close, W. H. (1977). The influence of climatic variables on energy metabolism and associated aspects of productivity in the pig. Nutrition and the Climatic Environment, 51–74. https://doi.org/10.1016/B978-0-407-00339-2.50005-4

Hunter, E.J.J., JONES, T.A.A., Guise, H.J.J., PENNY, R.H.C.H.C., Hoste, S., (2001). The Relationship between Tail Biting in Pigs, Docking Procedure and Other Management Practices. Vet. J. 161, 72–79. https://doi.org/10.1053/tvjl.2000.0520,

Huynh, T. T. T. and Aarnink, A. J. A (2004) Effects of floor cooling during high ambient temperatures on the lying behavior and productivity of growing finishing pigs, Transactions of the ASAE. American Society of Agricultural Engineers, 47(5), pp. 1773–1782. https://doi.org/10.13031/2013.17620

Huynh, T. T. T., Aarnink, A. J. A., Gerrits, W. J. J., Heetkamp, M. J. H., Canh, T. T., Spoolder, H. A. M., Kemp, B., & Verstegen, M. W. A. (2005). Thermal behaviour of growing pigs in response to high temperature and humidity. Applied

Animal Behaviour Science, 91(1), 1–16. https://doi.org/10.1016/j.applanim.2004.10.020

Huynh, T. T. T. Aarnink, A. J. A., Truong, C. T., Kemp, B., & Verstegen, M. W. A. (2006). Effects of tropical climate and water-cooling methods on growing pigs' responses. Livestock Science, 104(3), 278–291. https://doi.org/10.1016/j.livsci.2006.04.029

Huynh, T.T.T., Aarnink, a J. a, Verstegen, M.W. a, Gerrits, W.J.J., Heetkamp, M.J.W., Kemp, B., Canh, T.T., (2005). Effects of increasing temperatures on physiological changes in pigs at different relative humidities the online version of this article, along with updated information and services, is located on the World Wide Web at: Effects of increasing temperature 1385–1396. https://doi.org/10.2527/2005.8361385x

Jaakkola, T. et al. (2010) 'Learning Bayesian network structure using LP relaxations', in Proceedings of the Thirteenth International Conference on Artificial Intelligence and Statistics, pp. 358–365. http://hdl.handle.net/1721.1/66317

Jacobson, L. D. (1983). Validation of a swine growth model for early weaned pigs. https://www.elibrary.ru/item.asp?id=7389738

Jarissa Maselyne, Annelies Van Nuffel, Petra Briene, Jürgen Vangeyte, Bart De ketelaere, Sam Millet, Janne, & Van den HofdDominiekMaesdWouterSaeysb. (2017). Online warning systems for individual fattening pigs based on their feeding pattern. Biosystems Engineering. https://doi.org/10.1016/J.BIOSYSTEMSENG.2017.08.006

Jensen, D. B. (2016). Automatic learning and pattern recognition using sensor data in livestock farming. https://doi.org/10.1016/j.livsci.2014.06.028

Jensen, D. B., Kristensen, A. R., Børge, D., & Kristensen, A. R. (2016). Temperature as a predictor of fouling and diarrhoea in slaughter pigs. Livestock Science, 183, 1–3. https://doi.org/10.1016/j.livsci.2015.11.007

Jensen, D. B., Larsen, M. L. V. and Pedersen, L. J. (2020) Predicting pen fouling in fattening pigs from pig position', Livestock science, 231, p. 103852. https://doi.org/10.1016/j.livsci.2019.103852

Jensen, D.B., Toft, N., Kristensen, A.R., (2017). A multivariate dynamic linear model for early warnings of diarrhoea and pen fouling in slaughter pigs. Comput. Electron. Agric. 135, 51–62. https://doi.org/10.1016/J.COMPAG.2016.12.018

Jiang, B., Tang, W., Cui, L., & Deng, X. (2023). Precision Livestock Farming Research: A Global Scientometric Review. Animals: An Open Access Journal from MDPI, 13(13), 2096. https://doi.org/10.3390/ANI13132096

Jones, R., & Nicol, C. J. (1998). A note on the effect of control of the thermal environment on the well-being of growing pigs. Applied Animal Behaviour Science, 60, 1–9. https://doi.org/10.1016/S0168-1591(98)00148-8

Karim, F., Majumdar, S., Darabi, H., Chen, S., (2018). LSTM Fully Convolutional Networks for Time Series Classification. IEEE Access 6, 1662–1669. https://doi.org/10.1109/ACCESS.2017.2779939

Kashiha, M., Bahr, C., Haredasht, S. A., Ott, S., Moons, C. P. H. H., Niewold, T. A., Ödberg, F. O., & Berckmans, D. (2013). The automatic monitoring of pig's water uses by cameras. Computers and Electronics in Agriculture, 90, 164–169. https://doi.org/10.1016/j.compag.2012.09.015

Ke, G. et al. (2017) 'LightGBM: A Highly Efficient Gradient Boosting Decision Tree', in Guyon, I. et al. (eds) Advances in Neural Information Processing Systems. Curran Associates, Inc., pp. 3146–3154. https://dl.acm.org/doi/10.5555/3294996.3295074

Kingma, D. P., & Ba, J. (2014). Adam: A method for stochastic optimisation. arXiv preprint arXiv:1412.6980. Krzanowski, W.J., & Hand, D.J. (2009). ROC Curves for Continuous Data (1st ed.). Chapman and Hall/CRC. https://doi.org/10.1201/9781439800225

Kyriazakis, I., TOLKAMP, B.J.J., HUTCHINGS, M.R.R., (1998). Towards a functional explanation for the occurrence of anorexia during parasitic infections, Animal Behaviour. https://doi.org/10.1006/anbe.1998.0761

Larochelle, H., Bengio, Y., Louradour, J., Lamblin, P., (2009). Exploring Strategies for Training Deep Neural Networks. J. Mach. Learn. Res. 1, 1–40 ttps://api.semanticscholar.org/CorpusID:996073

Larsen, M. L. V., Andersen, H. M. L. M.-L., & Pedersen, L. J. (2018). Which is the most preventive measure against tail damage in finisher pigs: Tail docking, straw provision, or lowered stocking density? Animal, 12(6), 1260–1267. https://doi.org/10.1017/S175173111700249X

Larsen, M. L. V., Andersen, H. M.-L., & Pedersen, L. J. (2016). Can tail damage outbreaks in the pig be predicted by behavioural change? The Veterinary Journal, 209, 50–56. https://doi.org/10.1016/J.TVJL.2015.12.001

Larsen, M. L. V., Bertelsen, M. and Pedersen, L. J. (2018) Review: Factors affecting fouling in conventional pens for slaughter pigs', Animal: an international journal of animal bioscience, 12(2), pp. 322–328. https://doi.org/10.1017/S1751731117001586

Larsen, M. L. V., Bertelsen, M. and Pedersen, L. J. (2019) Pen Fouling in Finisher Pigs: Changes in the Lying Pattern and Pen Temperature Prior to Fouling', Frontiers in veterinary science, 6, p. 118. https://doi.org/10.3389/fvets.2019.00118

Larsen, M. L. V., Bertelsen, M., & Juul Pedersen, L. (2017). How do stocking density and straw provision affect fouling in conventionally housed slaughter pigs? Livestock Science, 205, 1–4. https://doi.org/10.1016/j.livsci.2017.09.005

Lemay, S. P., Guo, H., Barber, E. M., & Zyla, L. (2001). A procedure to evaluate humidity sensor performance under livestock housing conditions. CANADIAN BIOSYSTEMS ENGINEERING, 1979, 13–21.

Liu, W., Wang, Z., Liu, X., Zeng, N., Liu, Y., Alsaadi, F.E., (2017). A survey of deep neural network architectures and their applications. Neurocomputing 234, 11–26. https://doi.org/10.1016/j.neucom.2016.12.038

Longhouse, A. D. (1967). Design of poultry laying-house ventilation and insulation requirements based on calorimetric data and psychrometric relationships. Transactions of the ASAE, 10(4), 512–514.

Lyderik, K. K., Andersen, M.-L., Kristensen, H. H., & Pedersen, L. J. (2016). Protocol for the daily observations in the herds of the PigIT project. PigIT Report, 8. http://pigit.ku.dk/publications/

Madsen, T. N., & Kristensen, A. R. (2005). A model for monitoring the condition of young pigs by their drinking behaviour. Computers and Electronics in Agriculture, 48(2), 138–154. https://doi.org/10.1016/j.compag.2005.02.014

Madsen, T. N., Andersen, S., & Kristensen, A. R. (2005). Modelling the drinking patterns of young pigs using a state space model. Computers and Electronics in Agriculture, 48(1), 39–61. https://doi.org/10.1016/j.compag.2005.01.001

Malone, B. et al. (2018) 'Empirical hardness of finding optimal Bayesian network structures: algorithm selection and runtime prediction', Machine learning, 107(1), pp. 247–283. https://link.springer.com/article/10.1007/s10994-017-5680-2

McLachlan, G.J., Do, K.-A., Ambroise, C., (2005). Analysing microarray gene expression data. Wiley-Interscience.

https://onlinelibrary.wiley.com/doi/book/10.1002/047172842X

Meiszberg, A.M., Johnson, A.K., Sadler, L.J., Carroll, J.A., Dailey, J.W., Krebs, N., (2009). Drinking behavior in nursery pigs: Determining the accuracy between an automatic water meter versus human observers. J. Anim. Sci. 87, 4173–4180. https://doi.org/10.2527/jas.2008-1737

Mellor, D.J., (2016). Updating Animal Welfare Thinking: Moving beyond the " 'Five Freedoms'" towards 'A Life Worth Living'. Anim. an open access J. from MDPI 6. https://doi.org/10.3390/ani6030021

Mendl, M., Burman, O. H. P., & Paul, E. S. (2010). An integrative and functional framework for the study of animal emotion and mood. Proceedings of the Royal Society B: Biological Sciences, 277(1696), 2895–2904. https://doi.org/10.1098/RSPB.2010.0303

Mikolov, T., Kombrink, S., Burget, L., Cernocky, J., Khudanpur, S., (2011). Extensions of recurrent neural network language model, in 2011 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). IEEE, pp. 5528–5531. https://doi.org/10.1109/ICASSP.2011.5947611

Minvielle, B. and le Roux, A. (2009) ‘Influence of floor type during fattening on pig cleanliness and microbiological contamination of pigs and carcasses’, International Conference on the Epidemiology and Control of Biological, Chemical and Physical Hazards in Pigs and Pork. doi: 10.31274/safepork-180809-798.

Moinard, C., Mendl, M., Nicol, C.J., Green, L.E., (2003). A case control study of on-farm risk factors for tail biting in pigs. Appl. Anim. Behav. Sci. 81, 333–355. https://doi.org/10.1016/S0168-1591(02)00276-9

Morgan, C.A., Emmans, G.C., Tolkamp, B.J., Kyriazakis, I., (2000). Analysis of the feeding behavior of pigs using different models. Physiol. Behav. 68, 395–403. https://doi.org/10.1016/S0031-9384(99)00195-X

Morrison, S. R., Bond, T. E., & Heitman, H. (1967). Skin and lung moisture loss from swine. Transactions of the ASAE, 10(5), 691–696.

Mount, L.E., 1975. The assessment of thermal environment in relation to pig production. Livest. Prod. Sci. 2, 381–392. https://doi.org/10.1016/0301-6226(75)90121-9

Moura, D. J., Naas, I. A., Sevegnani, K. B., & Corria, M. E. (1997). The use of enthalpy as a thermal comfort index. Paper Presented at the 5th International

Livestock Environment Symposium.

Müller, R., & Schrader, L. (2003). A new method to measure behavioural activity levels in dairy cows. Applied Animal Behaviour Science, 83(4), 247–258. https://doi.org/10.1016/S0168-1591(03)00141-2

Nääs, I. A., Carvalho, V. C., Moura, D. J., & Mollo, M. N. (2006). Precision livestock production. Handbook of Agricultural Engineering, VI, 313–325.

Nienaber, J. A., & Hahn, G. L. (2000). Feeding behavior and energetics of growing-finishing swine as influenced by environmental temperature. Paper Presented at the International Winter Meeting of American Society of Agricultural Engineers (ASAE), Chicago, IL, USA.

Norton, T., Grant, J., Fallon, R., & Sun, D. W. (2009). Assessing the ventilation effectiveness of naturally ventilated livestock buildings under wind dominated conditions using computational fluid dynamics. Biosystems Engineering, 103(1), 78–99. https://doi.org/10.1016/j.biosystemseng.2009.02.007

Norton, T., Grant, J., Fallon, R., & Sun, D. W. (2010). Optimising the ventilation configuration of naturally ventilated livestock buildings for improved indoor environmental homogeneity. Building and Environment, 45(4), 983–995. https://doi.org/10.1016/J.BUILDENV.2009.10.005

Norton, T., Sun, D.-W., Grant, J., Fallon, R., & Dodd, V. (2007). Applications of computational fluid dynamics (CFD) in the modelling and design of ventilation systems in the agricultural industry: a review. Bioresource Technology, 98(12), 2386–2414. https://doi.org/10.1016/j.biortech.2006.11.025

Oczak, M., Viazzi, S., Ismayilova, G., Sonoda, L.T., Roulston, N., Fels, M., Bahr, C., Hartung, J., Guarino, M., Berckmans, D., Vranken, E., (2014. Classification of aggressive behaviour in pigs by activity index and multilayer feed forward neural network. Biosyst. Eng. 119, 89–97. https://doi.org/10.1016/j.biosystemseng.2014.01.005

Oniśko, A. and Druzdzel, M. J. (2013) ‘Impact of precision of Bayesian network parameters on accuracy of medical diagnostic systems’, Artificial intelligence in medicine, 57(3), pp. 197–206. https://doi.org/10.1016/j.artmed.2013.01.004

Opderbeck, S. et al. (2020) Influence of A Cooled, Solid Lying Area on the Pen Fouling and Lying Behavior of Fattening Pigs, Collection FAO: Agriculture, 10(7), p. 307. https://doi.org/10.3390/agriculture10070307

Oppenheim, A. V., & Schafer, R. W. (2014). Discrete-time signal processing.

1056. https://dl.acm.org/doi/book/10.5555/1795494

Ouellet, V., Vasseur, E., Heuwieser, W., Burfeind, O., Maldague, X., & Charbonneau, É. (2016). 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. https://doi.org/10.3168/jds.2015-10057

Patience, J.F., Umboh, J.F., Chaplin, R.K., Nyachoti, C.M., (2005). Nutritional and physiological responses of growing pigs exposed to a diurnal pattern of heat stress. Livest. Prod. Sci. 96, 205–214. https://doi.org/10.1016/j.livprodsci.2005.01.012

Pearl, J. (1995) Causal diagrams for empirical research, Biometrika, 82(4), pp. 669–688. https://doi.org/10.1093/biomet/82.4.669

Pearl, J. (2009) Causal inference in statistics: An overvie, Statistics surveys, 3, pp. 96–146. http://ftp.cs.ucla.edu/pub/stat_ser/r350.pdf

Philippe, F. X., & Nicks, B. (2015). Review on greenhouse gas emissions from pig houses: Production of carbon dioxide, methane and nitrous oxide by animals and manure. Agriculture, Ecosystems and Environment, 199, 10–25. https://doi.org/10.1016/j.agee.2014.08.015

Philippe, F. X., Cabaraux, J. F., & Nicks, B. (2011). Ammonia emissions from pig houses: Influencing factors and mitigation techniques. Agriculture, Ecosystems and Environment, 141(3–4), 245–260. https://doi.org/10.1016/j.agee.2011.03.012

Phillips, V. R., Holden, M. R., Sneath, R. W., Short, J. L., White, R. P., Hartung, J., ... & Wathes, C. M. (1998). The development of robust methods for measuring concentrations and emission rates of gaseous and particulate air pollutants in livestock buildings. Journal of Agricultural Engineering Research, 70(1), 11-24 https://doi.org/10.1006/jaer.1997.0283

Pinto, N., Doukhan, D., DiCarlo, J.J., Cox, D.D., (2009). A High-Throughput Screening Approach to Discovering Good Forms of Biologically Inspired Visual Representation. PLoS Comput. Biol. 5, e1000579. https://doi.org/10.1371/journal.pcbi.1000579

Pluske, J.R., Hampson, D.J., Williams, I.H., (1997). Factors influencing the structure and function of the small intestine in the weaned pig: a review. Livest. Prod. Sci. 51, 215–236. https://doi.org/10.1016/S0301-6226(97)00057-2

Provost, F., Provost, F., Fawcett, T., Kohavi, R., (1998). The Case Against Accuracy Estimation for Comparing Induction Algorithms. Proc. FIFTEENTH Int. Conf. Mach. Learn. 445--453. https://api.semanticscholar.org/CorpusID:13463620

Quiniou, N., Noblet, J., van Milgen, J., Dubois, S., (2001). Modelling heat production and energy balance in group-housed growing pigs exposed to low or high ambient temperatures. Br. J. Nutr. 85, 97. https://doi.org/10.1079/BJN2000217

Randall, J.M., Armsby, A.W., Sharp, J.R., (1983). Cooling gradients across pens in a finishing piggery: II. Effects on excretory behaviour. J. Agric. Eng. Res. 28, 247–259. https://doi.org/10.1016/0021-8634(83)90073-2

Rantzer, D., Westroem, B., & Svendsen, J. (1999). Effects of pen hygiene and moving the pigs at weaning on pig production and health during the first 8 weeks. Rapport-Sveriges Lantbruksuniversitet, Institutionen foer Jordbrukets Biosystem och Teknologi (JBT)(Sweden).

Reece, F. N., & Lott, B. D. (1982a). The effect of environmental temperature on sensible and latent heat production of broiler chickens. Poultry Science, 61(8), 1590–1593. https://doi.org/10.3382/ps.0611590

Reece, F. N., & Lott, B. D. (1982b). Typical broiler chicken growth rates. Poultry Science, 61, 1013–1014. https://doi.org/10.3382/ps.0611013

Reimers, N., & Gurevych, I. (2017). Optimal hyperparameters for deep lstm-networks for sequence labelling tasks. arXiv preprint arXiv:1707.06799.

Reiner, G., Hübner, K., Hepp, S., (2009). Suffering in diseased pigs as expressed by behavioural, clinical, and clinical–chemical traits, in a well-defined parasite model. Appl. Anim. Behav. Sci. 118, 222–231. https://doi.org/10.1016/j.applanim.2009.02.010

Rhouma, M., Fairbrother, J.M., Beaudry, F., Letellier, A., (2017). Post weaning diarrhoea in pigs: risk factors and non-colistin-based control strategies. Acta Vet. Scand. 59, 31. https://doi.org/10.1186/s13028-017-0299-7

Robert, B., White, B. J., Renter, D. G., & Larson, R. L. (2009). Evaluation of three-dimensional accelerometers to monitor and classify behavior patterns in cattle. Computers and Electronics in Agriculture, 67(1–2), 80–84. https://doi.org/10.1016/j.compag.2009.03.002

Rodrigues, V. C., da Silva, I. J. O., Vieira, F. M. C., & Nascimento, S. T. (2011).

A correct enthalpy relationship as thermal comfort index for livestock. International Journal of Biometeorology, 55(3), 455–459. https://doi.org/10.1007/s00484-010-0344-y

Rokade, A., Singh, M., Malik, P. K., Singh, R., & Alsuwian, T. (2022). Intelligent Data Analytics Framework for Precision Farming Using IoT and Regressor Machine Learning Algorithms. Applied Sciences 2022, Vol. 12, Page 9992, 12(19), 9992. https://doi.org/10.3390/APP12199992

Roque, K., Lim, G. D., Jo, J. H., Shin, K. M., Song, E. S., Gautam, R., Kim, C. Y., Lee, K., Shin, S., Yoo, H. S., Heo, Y., & Kim, H. A. (2016). Epizootiological characteristics of viable bacteria and fungi in indoor air from porcine, chicken, or bovine husbandry confinement buildings. Journal of Veterinary Science, 17(4), 531–538. https://doi.org/10.4142/JVS.2016.17.4.531

Rumelhart, D.E., Hinton, G.E., Williams, R.J., (1986). Learning representations by back-propagating errors. Nature 323, 533–536. https://doi.org/10.1038/323533a0

Rushen, J., Chapinal, N., de Passillé, A., (2012). Automated monitoring of behavioural-based animal welfare indicators. Anim. Welf. 21, 339–350. https://doi.org/10.7120/09627286.21.3.339

Salak-Johnson, J.L., Anderson, D.L., McGlone, J.J., (2004). Differential dose effects of central C.R.F. and effects of C.R.F. astressin on pig behavior. Physiol. Behav. 83, 143–150. https://doi.org/10.1016/j.physbeh.2004.08.005

Salehinejad, H., Sankar, S., Barfett, J., Colak, E., & Valaee, S. (2017). Recent advances in recurrent neural networks. arXiv preprint arXiv:1801.01078.

Sällvik, K., Walberg, K., (1984). The effects of air velocity and temperature on the behaviour and growth of pigs. J. Agric. Eng. Res. 30, 305–312. https://doi.org/10.1016/S0021-8634(84)80031-1

Savary, P., Gygax, L., Wechsler, B., & Hauser, R. (2009). Effect of a synthetic plate in the lying area on lying behaviour, degree of fouling and skin lesions at the leg joints of finishing pigs. Applied Animal Behaviour Science, 118(1-2), 20-27. https://doi.org/10.1016/j.applanim.2009.02.006

Schmidhuber, J., (2015). Deep learning in neural networks: An overview. Neural Networks 61, 85–117. https://doi.org/10.1016/J.NEUNET.2014.09.003

Schuster, M., Paliwal, K.K., (1997). Bidirectional recurrent neural networks. IEEE Trans. Signal Process. 45, 2673–2681. https://doi.org/10.1109/78.650093

Shao, Z., Yin, X., Bi, J., Ma, Z., & Wang, J. (2019). Spatiotemporal Variations of Indoor PM2.5 Concentrations in Nanjing, China. International Journal of Environmental Research and Public Health, 16(1). https://doi.org/10.3390/IJERPH16010144

Singla, P., Domingos, P., (2005). Discriminative training of Markov logic networks. Proc. 20th Natl. Conf. Artif. Intell. - Vol. 2 868–873. https://homes.cs.washington.edu/~pedrod/papers/aaai05.pdf

Smulders, D. et al. (2006). Validation of a behavioral observation tool to assess pig welfare', Physiology & behavior, 89(3), pp. 438–447. https://doi.org/10.1016/j.physbeh.2006.07.002

Soldatos, A. G., Arvanitis, K. G., Daskalov, P. I., Pasgianos, G. D., & Sigrimis, N. A. (2005). Nonlinear robust temperature-humidity control in livestock buildings. Computers and Electronics in Agriculture, 49(3), 357–376. https://doi.org/10.1016/j.compag.2005.08.008

Soriano, B., Hafez, A., Naya-Català, F., Moldovan, R. A., Toxqui-Rodríguez, S., Piazzon, M. C., Arnau, V., Llorens, C., & Perez-Sánchez, J. (2022). SAMBA: Structure-Learning of Aquaculture Microbiomes Using a Bayesian-Network Approach. BioRxiv, 2022.12.30.522281. https://doi.org/10.1101/2022.12.30.522281

Sozzi, M., Pillan, G., Ciarelli, C., Marinello, F., Pirrone, F., Bordignon, F., Bordignon, A., Xiccato, G., & Trocino, A. (2022). Measuring Comfort Behaviours in Laying Hens Using Deep-Learning Tools. Animals 2023, Vol. 13, Page 33, 13(1), 33. https://doi.org/10.3390/ANI13010033

Spoolder, H.A.M., Aarnink, A.A.J., Vermeer, H.M., van Riel, J., Edwards, S.A., (2012). Effect of increasing temperature on space requirements of group housed finishing pigs. Appl. Anim. Behav. Sci. 138, 229–239. https://doi.org/10.1016/j.applanim.2012.02.010

Srivastava, N., Hinton, G., Krizhevsky, A., Sutskever, I., & Salakhutdinov, R. (2014). Dropout: a simple way to prevent neural networks from overfitting. The journal of machine learning research, 15(1), 1929-1958. Available at: https://jmlr.org/papers/v15/srivastava14a.html

Statham, P., Green, L., Bichard, M., & Mendl, M. (2009). Predicting tail-biting from behaviour of pigs prior to outbreaks. Applied Animal Behaviour Science, 121(3–4), 157–164. https://doi.org/10.1016/j.applanim.2009.09.011

Stolba, A. and Wood-Gush, D. G. M. (1989) ‘The behaviour of pigs in a semi - natural environment, Animal production, (September 2010), pp. 419–425. https://doi.org/10.1017/S0003356100040411

Sundararajan, D. (2001). The Discrete Fourier Transform. https://doi.org/10.1142/4610

Suzuki, J. (2014). Learning Bayesian Network Structures When Discrete and Continuous Variables Are Present’, Probabilistic Graphical Models, pp. 471–486. https://10.1007/978-3-319-11433-0_31.

Taylor, C. J., Leigh, P., Price, L., Young, P. C., Vranken, E., & Berckmans, D. (2004). Proportional-integral-plus (PIP) control of ventilation rate in agricultural buildings. Control Engineering Practice, 12(2), 225–233. https://doi.org/10.1016/S0967-0661(03)00060-1

Teter, N. C., DeShazer, J. A., & Thompson, T. L. (1973a). Operational characteristics of meat animals: Part I. Swine. Transactions of the ASAE, 16(1), 157–159. https://doi.org/10.13031/2013.37473

Teter, N. C., DeShazer, J. A., & Thompson, T. L. (1973b). Operational characteristics of meat animals: Part III. Broilers. Transactions of the ASAE, 16(6), 1165–1167. https://doi.org/10.13031/2013.37723

Timmons, M. B., Gates, R. S., Bottcher, R. W., Carter, T. A., Brake, J., & Wineland, M. J. (1995). Simulation Analysis of a New Temperature Control Method for Poultry Housing. Journal of Agricultural Engineering Research, 62(4), 237–245. https://doi.org/10.1006/jaer.1995.1082

Tolkamp, B.J., Allcroft, D.J., Barrio, J.P., Bley, T.A.G., Howie, J.A., Jacobsen, T.B., Morgan, C.A., Schweitzer, D.P.N., Wilkinson, S., Yeates, M.P., Kyriazakis, I., (2011). The temporal structure of feeding behavior. Am. J. Physiol. Integr. Comp. Physiol. 301, R378–R393. https://doi.org/10.1152/ajpregu.00661.2010

Transtrum, M. K., & Qiu, P. (2016). Bridging Mechanistic and Phenomenological Models of Complex Biological Systems. PLOS Computational Biology, 12(5), e1004915. https://doi.org/10.1371/JOURNAL.PCBI.1004915

Tsamandouras, N., Rostami-Hodjegan, A., & Aarons, L. (2015). Combining the ‘bottom up’ and ‘top down’ approaches in pharmacokinetic modelling: fitting PBPK models to observed clinical data. British Journal of Clinical Pharmacology, 79(1), 48–55. https://doi.org/10.1111/BCP.12234

Tucci, R. R. (2013) ‘Introduction to Judea Pearl’s Do-Calculus’, arXiv [cs.AI].

Available at: http://arxiv.org/abs/1305.5506.

Ursinus, W.W., Van Reenen, C.G., Kemp, B., Bolhuis, J.E., (2014). Tail biting behaviour and tail damage in pigs and the relationship with general behaviour: Predicting the inevitable? Appl. Anim. Behav. Sci. 156, 22–36. https://doi.org/10.1016/j.applanim.2014.04.001

Usry, J. L., Turner, L. W., Bridges, T. C., & Nienaber, J. A. (1992). Modeling the physiological growth of swine part III: heat production and interaction with environment. Transactions of the American Society of Agricultural Engineers, 35(3), 1035–1042. https://doi.org/10.13031/2013.28698

Venables, W.N., Ripley, B.D., (2002. Modern Applied Statistics with S, Statistics and Computing. Springer New York, New York, NY. https://doi.org/10.1007/978-0-387-21706-2

Walker, J. W., Heitschmidt, R. K., & Dowhower, S. L. (1985). Evaluation of pedometers for measuring distance traveled by cattle on two grazing systems. Journal of Range Management, 38(1), 90–93.

Wang, Y., Yang, W., Winter, P., Walker, L., (2008. Walk-through weighing of pigs using machine vision and an artificial neural network. Biosyst. Eng. 100, 117–125. https://doi.org/10.1016/j.biosystemseng.2007.08.008

Wathes, C. M., Kristensen, H. H., Aerts, J. M., & Berckmans, D. (2008). Is precision livestock farming an engineer's daydream or nightmare, an animal's friend or foe, and a farmer's panacea or pitfall? Computers and Electronics in Agriculture, 64(1), 2–10. https://doi.org/10.1016/J.COMPAG.2008.05.005

Werbos, P.J., (1990). Backpropagation through time: what it does and how to do it. Proc. IEEE 78, 1550–1560. https://doi.org/10.1109/5.58337

Whittaker, J. (2009) Graphical Models in Applied Multivariate Statistics. Wiley Publishing. https://doi.org/10.2307/2532177

Williams, R.J., Zipser, D., (1989). A Learning Algorithm for Continually Running Fully Recurrent Neural Networks. Neural Comput. 1, 270–280. https://doi.org/10.1162/neco.1989.1.2.270

Yang, T. S., Howard, B., & Macfarlane, W. V. (1981). Effects of food on drinking behaviour of growing pigs. Applied Animal Ethology, 7(3), 259–270. https://doi.org/10.1016/0304-3762(81)90082-1

Yang, T.S., Howard, B., Macfarlane, W.V., 1981. Effects of food on drinking behaviour of growing pigs. Appl. Anim. Ethol. 7, 259–270.

https://doi.org/10.1016/0304-3762(81)90082-1

Young, P. C., & Jakeman, A. J. (1979). Refined instrumental variable methods of time-series analysis: Part I, SISO systems. International Journal of Control, 29, 1–30. https://doi.org/10.1080/00207178008961080

Zervanos, S.M., Hadley, N.F., (1973). Adaptational Biology and Energy Relationships of the Collared Peccary (Tayassu Tajacu). Ecology 54, 759–774. https://doi.org/10.2307/1935671

Zhang, J., Zhu, Y., Zhang, X., Ye, M., Yang, J., (2018). Developing a Long Short-Term Memory (LSTM) based model for predicting water table depth in agricultural areas. J. Hydrol. 561, 918–929. https://doi.org/10.1016/J.JHYDROL.2018.04.065

Zheng, X. et al. (2018) .DAGs with NO TEARS: Continuous Optimisation for Structure Learning, in Bengio, S. et al. (eds) Advances in Neural Information Processing Systems. Curran Associates, Inc., pp. 9472–9483. https://papers.nips.cc/paper/2018/hash/e347c51419ffb23ca3fd5050202f9c3d-Abstract.html