DISSERTATIONES SCHOLAE DOCTORALIS SCIENTIAE CIRCUMIECTALIS,
ALIMENTARIAE, BIOLOGICAE. UNIVERSITATIS HELSINKIENSIS

ALPER TUNA KAVLAK

THE GENETIC PARAMETERS OF FEEDING BEHAVIOUR AND PRODUCTION TRAITS, SOCIAL INTERACTION AND DISEASE CONTROL IN FINNISH PIGS

ANIMAL SCIENCE
DEPARTMENT OF AGRICULTURAL SCIENCES
FACULTY OF AGRICULTURE AND FORESTRY
DOCTORAL PROGRAMME IN SUSTAINABLE USE OF RENEWABLE
NATURAL RESOURCES
UNIVERSITY OF HELSINKI

YEB
ALPER TUNA KAVLAK
THE GENETIC PARAMETERS OF FEEDING BEHAVIOUR AND PRODUCTION TRAITS, SOCIAL
INTERACTION AND DISEASE CONTROL IN FINNISH PIGS
8/2022

UNIVERSITY OF HELSINKI
DOCTORAL SCHOOL IN ENVIRONMENTAL, FOOD AND BIOLOGICAL SCIENCES
Doctoral programme in Sustainable Use of Renewable Natiral Resources
(Animal Science)

PUBLICATION No. 8/2022

THE GENETIC PARAMETERS OF FEEDING BEHAVIOUR AND PRODUCTION TRAITS, SOCIAL INTERACTION AND DISEASE CONTROL IN FINNISH PIGS

DOCTORAL THESIS

ALPER TUNA KAVLAK

ACADEMIC DISSERTATION

To be presented, with the permission of the Faculty of Agriculture and Forestry of the University of Helsinki, for public examination in Laboratory Building, Room 131 (K05145-131), Koetilantie 5, Helsinki, on August 19th, 2022, at 1 p.m.

HELSINKI, FINLAND
2022

Custos: **Professor Pekka Uimari**
Department of Agricultural Sciences
University of Helsinki, Finland

Supervisors: **Professor Pekka Uimari**
Department of Agricultural Sciences
University of Helsinki, Finland

Associate Professor Matti Pastell
Natural Resources Institute Finland (Luke)
Helsinki, Finland

Reviewers: **Dr. Eli Grindflek**
Norsvin R&D
Hamar, Norway

Professor Dr. Ir. Peter Groot Koerkamp
Department of Plant Sciences
Wageningen University & Research (WUR)
Wageningen, Netherland

Opponent: **Dr. Wendy Mercedes Rauw**
Department of Animal Breeding
The National Institute for Agricultural and Food Research and Technology (INIA), Spain

Cover photo: Figen Oy (Längelmäki, Finland)
Dissertationes Schola Doctoralis Scientiae Circimiectalis, Alimentariae, Biologicae
ISBN 978-951-51-8220-3 (PRINT)
ISBN 978-951-51-8221-0 (ONLINE)
ISSN 2342-5423 (PRINT)
ISSN 2342-5431 (ONLINE)
Electronic publication at http://ethesis.helsinki.fi

Unigrafia
Helsinki 2022

ABSTRACT

A major proportion of the costs of pork production is related to feed; thus, improving feed efficiency is one of the most important breeding goals in pig breeding programs. Feed efficiency measurements are based on the performance of pigs in a controlled test station environment with an automatic feeding system. During the test period, a large amount of feeding-related information, including time entering the feeder, time leaving the feeder and amount of feed consumed, are gathered. These data can be used to form different feeding behaviour traits (**FBT**) such as daily feed intake (**DFI**) or feed intake per visit (**FPV**). During the test period, pigs are group housed. Therefore, social behaviour between pen mates may affect feeding behaviour, production traits (**PT**) and welfare of other pen mates. Furthermore, unusual or radical changes in pigs' behaviour can be a sign of diseases or discomfort and can be used for the detection of diseases for timely intervention.

The thesis has three objectives. The first objective was to estimate the heritability of FBT and their correlation with PT in the Finnish Yorkshire pig population (I). The second objective was to estimate genetic parameters of feeding behaviour and PT in Finnish pig breeds using a model with a social genetic effect (**SGE**; study II). The third objective was to develop machine learning methods for detection of pigs that may need medical treatment or extra management based on changes in their feeding behaviour patterns (III).

The data consisted of purebred Finnish Yorkshire and Finnish Landrace pigs and their F1-crosses. Pigs were raised in a controlled test station environment located in Längelmäki, Finland (Figen Oy). Pigs arrived at the test station at an average age of 89 ± 10 days (mean ± standard deviation) and an average weight of 34.7 ± 6.4 kg. The slaughter age was 186 ± 10 days, and the slaughter weight was 121.2 ± 12.9 kg. The feeding, production and sickness data were collected from 10,275 pigs that had entered the test station between 2011 and 2016 (October). From obtained feeding observation, daily values were calculated as the number of visits per day (**NVD**), time spent in feeding per day (**TPD**), DFI, time spent feeding per visit (**TPV**), FPV and feed intake rate (**FR**). The final records of FBT were calculated as averages of the daily records separately for the five testing time periods: 0 to 20 days (period 1), 21 to 40 days

(period 2), 41 to 60 days (period 3), 61 to 80 days (period 4) and 81 to 93 days (period 5). The PT were average daily gain (**ADG**), feed conversion rate (**FCR**) and backfat thickness (**BF**) as well as residual feed intake (**RFI**), which was calculated as the difference between the observed and predicted DFIs. For the last objective, daily observations of the health status of pigs (794,509 daily observations) were used. The data included 13,018 observations of symptoms (cough, limp, loss of appetite, skin damage, bitten tail). Pigs with symptoms were classified as "sick" while the pigs with no symptoms were classified as "healthy".

For the first objective (I), a univariate animal model was applied for the estimation of heritability of FBT and PT and a bi-variate model for the estimation of genetic correlations between the traits. For the second objective, single-trait social genetic models were applied with two different group size parametrisations: a fixed group size with a random sampling of eight pigs for each pen and a variable group size model where the number of pen mates varied. The restricted maximum likelihood (**REML**) method and the DMU software were used for the variance component estimations (I and II). For the last objective, the machine learning algorithm (Xgboost) was designed with different window lengths (one-day, three-day and seven-day window lengths) using features calculated from the feeding data to predict the individual health status (sick or healthy).

Heritability estimates for FBT and PT were moderate (around 0.3). Only the DFI had strong genetic correlations with PT (e.g., over 0.8 with RFI). Interestingly, the FR and FPV were associated with BF in early testing time (I). The SGE was significant for FBT and the FCR but not for the ADG and BF (II). For those traits, the total heritable variation was considerably higher than the pure classical heritability. The fixed and variable group size models produced very similar estimates (II). Using the Xgboost machine learning method, the best performance for predicting pigs' daily health status was obtained with a seven-day window length (80% area under the Receiver Operator Characteristics (**ROC**) curve) [**AUC**], 7% F1-score, 67% sensitivity and 73% specificity). However, the precision was very low (0.04), possibly due to an imbalanced dataset (III).

In conclusion, results indicate that the FBT are moderately inherited but their correlation with PT are weak; thus, FBT are not important in breeding programs. The social genetic component was important in FBT but also in the FCR; thus, applying SGEs in breeding value estimation would increase the accuracy of selection for the FCR. Predictions based on changes in feeding behaviour can help in the detection of sick pigs, leading to better pig welfare and increased sustainability of pork production.

ACKNOWLEDGEMENTS

This thesis work was carried out at the Department of Agricultural Sciences, University of Helsinki and has been quite challenging but provided me with the opportunity to work with exceptionally talented, intelligent and dedicated professionals. I am grateful to Raisio Oyj Tutkimussäätiö (Finland) and Suomen Sianjalostuksen Säätiö (Finland) to fund my study.

First, I am very grateful to my supervisor Professor Pekka Uimari who has a vision, knowledge, sincerity and motivation that have deeply inspired me throughout this research. I would also like to express my deepest and sincere gratitude to my co-supervisor Associate Professor Matti Pastell who supported me in this process. It was a great privilege and honour to work and study under his invaluable guidance. I would like to acknowledge and express my appreciation to my co-authors Professor Ismo Strandén and Martin H. Lidauer from the Natural Resources Institute Finland (Luke) for their collaboration and for sharing their software expertise in my preparation of my second manuscript. I am also grateful to my thesis committee advisors Dr Timo Serenius (Figen Oy), Marja-Liisa Sevón-Aimonen (Luke) and Professor Jarmo Valaja (University of Helsinki) for their encouragement, insightful comments and difficult questions. I am grateful to reviewers Dr Eli Grindflek and Professor Dr Ir Peter Groot Koerkamp for their constructive comments and suggestions to improve my thesis. I also wish to express my sincere gratitude to Dr. Wendy Mercedes Rauw for honouring me by being my opponent and to Professor Pekka Uimari for being my custos.

I thank all my past and present colleagues, especially Professor Asko Mäki-Tanila and Professor Dr Zeynel Cebeci, who has always supported, redirected and encouraged me during my academic life. I owe my warmest gratitude to family and friends who have been supportive and encouraged me. In addition, I am grateful and feel lucky to have my cat Ozzy, who always motivated me with his cuteness. Finally, huge thanks to my dearest wife Esra, who has been my motivation and support and has always been with me in times of celebration and frustration.

CONTENTS

ABSTRACT 3

ACKNOWLEDGEMENTS 6

LIST OF ORIGINAL PUBLICATIONS 9

ABBREVIATIONS 10

1 INTRODUCTION 11

- 1.1 Feeding behaviour traits 12
- 1.2 Importance of feeding behaviour traits in pig breeding programs 13
- 1.3 Social interaction models 13
- 1.4 Disease detection by machine learning methods 15

2 OBJECTIVES 17

3 MATERIALS AND METHODS 18

- 3.1 Materials 18
  - 3.1.1 Figen test station (Längelmäki, Finland) 18
  - 3.1.2 Feeding beheaviour data 19
  - 3.1.3 Production data 20
  - 3.1.4 Sickness data 21
  - 3.1.5 Pedigree data 21
- 3.2 Methods 22
  - 3.2.1 Estimation of heritability and total heritable variation 22
  - 3.2.2 Genetic correlation 23

3.2.3 Xgboost algorithm..23

3.2.4 Features and window lengths selection for Xgboost...............................24

3.2.5 Cross validation, performance metrics and feature importance 25

4 RESULTS AND DISCUSSION 27

4.1 Estimated heritabilities of feeding behaviour and production traits 27

4.2 Genetic correlations between feeding behaviour and prodcution traits……………...……………………………………………………………… 30

4.3 Estimated social genetic parameters of feeding behaviour and prodcution traits……………………………………………………………………………36

4.4 Predicting pigs' health status using the feeding behaviour data.......................38

4.5 Implications and future researchs 41

5 CONCLUSION...43

REFERENCES 44

LIST OF ORIGINAL PUBLICATIONS

This thesis is based on the following publications, which are referred to in the text by their Roman numerals:

I	Kavlak, A.T. and Uimari, P. 2019. Estimation of heritability of feeding behaviour traits and their correlation with production traits in Finnish Yorkshire pigs. Journal of Animal Breeding and Genetics, 136: 484-494.
II	Kavlak, A.T., Strandén, I., Lidauer, M.H. and Uimari, P. 2021. Estimation of social genetic effects on feeding behaviour and production traits in pigs. Animal, 15(3): 100168.
III	Kavlak, A.T., Pastell, M. and Uimari, P. 2022. Disease detection in pigs based on feeding behaviour traits using machine learning. Biosystems Engineering. (Submitted on 25.12.2021)

Contribution of the author to studies I to III:

The author participated in study planning, data editing, statistical analyses, interpretation of results and dissemination of research outcomes to the journals as the main author of studies I, II and III.

ABBREVIATIONS

ADG	Average daily gain
AUC	Area under the ROC curve
BF	Backfat thickness
CV	Cross-validation
DBV	Direct breeding value
DFI	Daily feed intake
DGE	Direct genetic effect
DRDFI	Daily rank of daily feed intake within the group of pigs
DRTPD	Daily rank of time spent in feeding per day within the group of pigs
FBT	Feeding behaviour traits
FCR	Feed conversion rate
FE	Feed efficiency
FPV	Feed intake per visit
FR	Feed intake rate
ML	Machine learning
NH_DE	Non-heritable direct effects
NH_SE	Non-heritable social effects
NVD	Number of visits per day
PT	Production traits
ResTPD	Residuals of time spent in feeding per day
ResDFI	Residuals of daily feed intake
RFI	Residual feed intake
ROC	Receiver Operator Characteristics
SBV	Social breeding value
SE	Standard error
SGE	Social genetic effect
REML	Restricted maximum likelihood
TBV	Total breeding value
TPD	Time spent in feeding per day
TPV	Time spent in feeding per visit

1 INTRODUCTION

In a test station, pigs are routinely evaluated for growth, feed conversion rate (**FCR**) and carcass traits. In addition, detailed feeding records are available for these pigs and can be used for feeding behaviour studies. Feeding behaviour traits (**FBT**) do not have direct economic value. However, if the FBT are correlated with production traits (**PT**), then FBT can be used as auxiliary traits for PT. For example, FBT may provide early information about PT that are difficult to measure or can be measured only late in life (Schulze et al., 2003). Therefore, in study I, in addition to heritability estimates for FBT, correlations between FBT and average daily gain (**ADG**), FCR, backfat thickness (**BF**) and residual feed intake (**RFI**) were estimated.

Pigs are housed in groups in commercial farms and in test stations. When the animals are group housed, their performance depends on direct genetic effects (**DGEs**), environmental effects and their social interaction with pen mates (Griffing, 1967; Muir, 2003). Therefore, the social behaviour between pen mates may have positive or negative effects on the feeding behaviour, production and wellbeing of the pigs sharing the same pen (Wolf et al., 1998). Some social behaviours such as competition and aggression generally have negative effects on pen mates (Ellen et al., 2014), while cooperation and mothering ability have generally positive effects on pen mates (Canario et al., 2017). These social behaviour patterns are partly inherited. This part is called the social genetic effect (**SGE**; Nielsen et al., 2018). Therefore, if the animals' performance also depends on social interactions between groupmates, the SGE should be considered in both breeding value estimation and selection to achieve better genetic response than with programs that are based purely on additive genetic effects of the breeding candidates (Bijma et al., 2007b). Therefore, in study II, the effect of pen mates on FBT and PT was explored.

In recent years, there has been increased concern about pig welfare under farming systems with the scientific concurrence that animals' welfare should be improved (Mellor, 2016). Animals express their internal condition through their behaviours (feeding/drinking behaviour, social behaviour, etc.), and unexpected changes in pigs' behaviour may indicate possible health problems (Matthews et al., 2017). Changing patterns in feeding behaviour can be used for the early detection of possible health and welfare problems in pigs. Therefore, in study III, a prediction method was developed to detect possible health problems based on changes in FBT.

1.1 Feeding behaviour traits

FBT are based on feeding data collected from automatic feeding systems with identification of animals, time entering and leaving the feeder and amount of feed consumed during the visit. Feeding behaviour can be described as the number of visits per day (**NVD**), time spent in feeding per day (**TPD**), daily feed intake (**DFI**), time spent in feeding per visit (**TPV**), feed intake per visit (**FPV**) and feed intake rate (**FR**; von Felde et al., 1996; Hall et al., 1999). There are also some other FBT that can be measured that are not used in this study, such as time intervals between meals (Remus et al., 2020). FBT can be used to estimate the genetic and phenotypic relationship between the FBT and other traits (Do et al., 2013), estimate the effects of social interaction (Bijma 2007a, b; Bergsma et al., 2008) and predict sickness (Matthews et al., 2017; Gertz et al., 2020). In general, genetic variance exists in FBT, and heritabilities of FBT vary from moderate to high estimates of heritability (de Haer and de Vries, 1993; Do et al., 2013; von Felde et al., 1996; Labroue et al., 1997; Schulze et al., 2003; Table 1). For example, de Haer and de Vries (1993) reported low to moderate heritabilities for the NVD, TPD, DFI, TPV, FPV and FR of 0.38, 0.24, 0.16, 0.27, 0.35 and 0.28, respectively, in group-housed pigs. On the other hand, von Felde et al. (1996) reported moderate to high heritability estimates for the NVD, TPD, DFI, TPV, FPV and FR of 0.43, 0.43, 0.22, 0.42, 0.51 and 0.44, respectively, in group-housed Landrace and Large White boars. Moreover, there is a genetic correlation between FBT and PT (von Felde et al., 1996; Labroue et al., 1997; Hall et al., 1999; Schulze et al., 2003; McSweeny et al., 2003; Young et al., 2012). For example, von Felde et al. (1996) reported both low and high estimations of genetic correlation between the DFI and the FCR, RFI, ADG and BF (0.68, 0.45, 0.13 and 0.97, respectively). Labroue et al. (1997) also reported high genetic correlations between the DFI and ADG in the Large White (0.87) and French Landrace (0.81).

Apart from the genetic factors, FBT is also affected by environmental factors, and housing should provide the most optimum conditions possible to ensure good performance as well as good welfare of pigs. For example, Gonyou et al. (1998)

reported a reduction in the DFI from 2.49 to 2.21 kg when the group size increased from 3 to 15 pigs. However, Nielson et al. (1995) and Wolter et al. (2001) reported that FBT were not affected by group size. In addition, heat stress influences the DFI and weight gain of growing pigs (McGlone et al., 1985; Feddes et al., 1989; McGlone et al., 1994; Hyun et al., 1998). Feddes et al. (1989) reported that if the temperature varied from 26 to 40°C, the greatest feed intake was observed either during early morning or late evening compared to pigs growing in an environment with a constant temperature (33°C).

1.2 Importance of feeding behaviour traits in pig breeding programs

Efficient pig production relies on large and viable litters and pigs' ability to efficiently produce high-quality meat (Knap et al., 2001). Since most of the costs of pork production are related to feed, feed efficiency (**FE**) is a significant interest in pig breeding programs, and traits like the FCR and RFI are commonly included in breeding programmes. FBT are not usually of direct interest in breeding programmes, but if a positive genetic correlation exists between FBT and PT, using FBT in selection programmes may increase the accuracy of the selection of PT depending on the correlation between FBT and PT (Do et al., 2013). Previous studies have shown the phenotypic and genetic correlation between feeding behaviour and PT (Labroue et al., 1997; Schulze et al., 2003; Rauw et al., 2006; Do et al., 2013); for example, animals that eat more per visit tend to grow faster (Labroue et al., 1997). In addition, in Dekkers et al.'s (2014) study, pigs with reduced RFI ate less feed, spent less time at the feeder, made fewer daily visits to the feeder, and ate at a faster rate than the control group pigs. Thus, FBT could be utilised at least as an auxiliary trait in genetic evaluation of pigs for PT.

1.3 Social interaction models

Social interaction has an impact on group-housed animals sharing the same pen. In classical quantitative genetics, observed trait values (P) can be expressed as the sum of a heritable component (A, breeding value) and a non-heritable or environmental component (E) where P = A + E (Falconer and Mackay, 1996; Lynch and Walsh, 1998). When the trait is affected by interactions with the group or pen mates, then the classical model needs to be expanded with social effects of pen mates (Dickerson, 1947; Wilham, 1963; Griffing, 1967; Cheverud, 1984; Wolf et al., 1998). In a social genetic model, an animal's phenotype depends on the genotype of the animal itself (referred to as DGE), the genotypes of the other animals in the same pen (referred to as either the social or indirect genetic effects, SGEs) and the corresponding non-heritable direct and social effects (Griffing, 1967; Moore et al., 2007; Bijma et al., 2007a,b). The total breeding value (**TBV**) of an animal is the sum of its direct breeding value (**DBV**) and its own social breeding value (**SBV**) towards its group mates. DBV is a part of the DGE that is transmitted to progeny. Similarly, SBV is a part of an animal's SGE that is transmitted to progeny. Because TBV is the part of an animal's genotype that can be utilised in selection programmes to improve the direct performance of the progeny and improve the social effects of the progeny towards pen mates, it is preferred over the traditional DBV. Therefore, the proportion of variance of TBV over the total phenotypic variance is more important than the classical heritability in traits that are affected by SGEs.

To illustrate P, DBV and TBV, an example of three pigs is used below:

$P_1 = DGE_1 + NH\_DE_1 + SGE_2 + NH\_SE_2 + SGE_3 + NH\_SE_3$

$P_2 = DGE_2 + NH\_DE_2 + SGE_1 + NH\_SE_1 + SGE_3 + NH\_SE_3$

$P_3 = DGE_3 + NH\_DE_3 + SGE_1 + NH\_SE_1 + SGE_2 + NH\_SE_2$

$TBV_1 = DBV_1 + SBV_1$

$TBV_2 = DBV_2 + SBV_2$

$TBV_3 = DBV_3 + SBV_3$

where Non-heritable direct effects (**NH_DE)** and Non-heritable social effects (**NH_SE)** relate to non-heritable direct and social effects, respectively.

The variance of TBV depends on the (co)variance components of DBV and SBV but also on group size (Bijma et al., 2007a, b). The number of pen mates may vary for several reasons (e.g., removal due to sickness), causing variation in group size. Varying group size complicates the estimation of variance components using existing variance component estimation programs. When group sizes vary in the population, simply fitting a fixed effect for group size may not be enough because variance components may also vary between group sizes. For instance, when the group size is large, social interaction between pigs may be less intense, which means that variance of SGEs decreases with group size. Several approaches have been proposed as a solution for coping with variable group sizes (e.g., Arango et al., 2005; Hadfield and Wilson, 2007; Bijma et al., 2007a; Nielsen et al., 2018a). One approach to account for the variable group size in a statistical model is to use a so-called dilution factor (*d*) where the effect of the SGE of the pen mates in large groups on an individual's phenotype is relatively smaller than the effect in small groups (Bijma, 2010; Hadfield and Wilson, 2007). In addition, the group size and the relatedness of the animals within a group affect the accuracy of SGE variance estimates. The accuracy can be improved with large groups with highly related pen mates (Chu et al., 2021). Finally, the social interaction may depend on environmental conditions such as a restricted feeding system that creates a competition between the pen mates if the total amount of feed within a pen is fixed. Competition due to restricted feeding may cause negative correlations between direct and social effects as shown by Bergsma et al. (2008) for ADG.

1.4 Disease detection by machine learning methods

Early disease control is important in increasing welfare and treatment success among pigs. Animals express their internal situation through behaviour; thus, observing changes in behaviour (e.g., feeding) could pinpoint animals that are potentially sick. These changes can be observed either by examining the animals or using statistical modelling based on the data received from the automated sensors and feeding systems. These behavioural changes can be correlated to any sickness (e.g., limp, loss of appetite, bitten tail). For example, Wallenbeck and Keeling (2013) identified a low frequency of feeding at a group level up to nine weeks prior to the first injury. Additionally, Munsterhjelm et al. (2015) observed the reduction of feed intake 20 days prior to the tail being bitten. In practise, animals are not examined in detail for possible sickness on a daily basis. Instead, farm workers eyeball the animals during their routine management work. This permits detection of the most obvious and severe health issues, at which point it may be too late for effective and treatable intervention. Furthermore, microphones have been used to monitor the sounds of coughing of pigs to build an intelligent alarm system to detect the disease in its early stage (Guarino et al., 2008), 3D-cameras to predict tail biting outbreaks by identifying lowered tail postures (D'Eath et al., 2018), and deviations in typical feeding patterns to monitor overall welfare of pigs (e.g., Brown-Brandl et al., 2013; Munsterhjelm et al., 2015; Bus et al., 2021).

Machine learning (**ML**) is a type of artificial intelligence, and ML algorithms such as extreme gradient boosting, random forests and support vector machines are used to predict different outcomes from the sample data (or "training data"). The algorithms have been applied in a wide variety of fields. Recently, ML has become a popular and powerful method that helps to detect behavioural changes related to health problems (Liaskos et al., 2018). Although the ML algorithms are quite a powerful method for classification as well as clustering, the quality of the data has a crucial impact on the performance of the algorithm; thus, before applying the ML algorithm to data, the data should be filtered to avoid unwanted outcomes and increase the robustness of the performance (Alsaaod et al., 2012).

2 OBJECTIVES

This thesis had separate but linked goals, namely to increase knowledge of the genetics of feeding behaviour, estimate the genetic part of social interaction between the pen mates and find possible patterns in feeding behaviour that are correlated with animal health. Therefore, the main objective was divided into three specific goals as outlined below (with the study number provided in parentheses):

1) To estimate the heritability of FBT and their correlation with PT in Finnish Yorkshire pigs (I)
2) To estimate the SGEs on feeding behaviour and PT in pigs (II)
3) To detect disease in pigs based on FBT using ML (III)

The proposed project was based on the hypothesis that pigs display genetic differences in FE and FBT that can be used in the selection of pigs with improved FE and health. The second hypothesis is that revealing the genetic basis of social interaction between pigs allows the selection of pigs that have a positive effect on pen mates' health, welfare and productivity. The third hypothesis is that changes in behaviour patterns can be used to detect pigs that may need medical treatment or extra management to avoid possible diseases such as diarrhoea or lameness.

3 MATERIALS AND METHODS

3.1 MATERIALS

For all objectives, the pigs were either purebred Finnish Yorkshire (I–III), Finnish Landrace (II, III) or F1-crosses between the two breeds raised in a controlled test station environment located in Längelmäki, Finland (Figen Oy). Pigs were related to each other through common ancestors. The production, feeding and sickness data were collected from 10,275 pigs that entered the test station between 2011 and 2016 (October). The pigs arrived at the test station at an average age of 89 ± 10 days (mean ± standard deviation) and an average weight of 34.7 ± 6.4 kg in total testing time (on average, 95 ± 3 days). Moreover, the slaughter age was 186 ± 10 days while the slaughter weight was 121.2 ± 12.9 kg.

In study I, only Yorkshire pigs (3,235 pigs) were used for estimation of the genetic parameters of FBT and their correlation with PT. In study II, both feeding behaviour and production data from all animals were used in the data. However, the groups which had less than 8 and more than 12 pigs were removed. The final data in study II included 3,075 Finnish Yorkshire, 3,351 Finnish Landrace and 968 F1-crosses. In study III, feeding behaviour and sickness data from both breeds and F1-crosses were used. The data included 10,261 animals with 794,509 daily feeding behaviour and sickness observations.

No animal experiment was needed in this project; thus, no ethical approval was required. The study followed the University of Helsinki's ethical research guidelines, good scientific practice and applicable laws and regulations.

3.1.1 Figen test station (Längelmäki, Finland)

The data was received from the pig test station located in Längelmäki in central Finland. The tested breeds are Finnish Landrace and Finnish Yorkshire, which belong to the breeding program run by Figen. The station conducts individual tests on approximately 2,500 purebred boars each year. Boar candidates come from the nucleus farms at 30 kg of weight for the 14-week test period. During the test period, feeding information, weight, BF and conformation are recorded. After slaughtering, carcass and meat quality information are also collected. The piggery has automated ventilation based on the age of the pigs and the outdoor temperature. Artificial light is on from 7 am to 3 pm, and the dimension of the pen is 16.8 m^2 with one third of concrete slats. In this study, the data included the feeding and production records of pigs from the central test station from 2011 to 2016.

Pigs arrive at the test station on Tuesday or Wednesday, and the test starts on Saturday. The grouping of pigs into different pens is done according to the arrival weight (same size) and sex (only boars or a combination of gilts and castrates) of the pigs. Feeding is recorded automatically using the Schauer Spotmix with Schauer MLP electronic feeders and MLP manager data management software (Schauer Agrotronic GmbH). Feeding type (dry feeding) is *ad libitum* consisting of two commercial feedstuffs. Water was available *ad libitum*. The proportion of the two feedstuffs is based on the growth rate curve of an average pig from the previous test periods.

3.1.2 Feeding behaviour data

The raw feeding behaviour data consisted of 28,964,641 individual feeding visits observations with ear tag transponder-id, date, time entering the feeder, time leaving the feeder and FPV. The feed intake was measured as a weight of the feed before and after the pig has been in feeder. From the raw records, time spent per visit and feeding rate per visit (g/min) were calculated. Some visits may have been missed due to ear tag issues (e.g., incorrect records, ear tags falling out) or sickness, and some individual records could be erroneous due to malfunction of the feeder or extraordinary feeding behaviour of some pigs on some days. Therefore, prior to statistical analyses, the thresholds presented in Casey et al. (2005) were applied to remove potentially erroneous data. The proportion of the outliers was less than 1%, and the remaining data included 28,826,029 daily values. From the daily values, the NVD (counts), TPD (min), DFI (g), TPV (min), FPV (g) and FR (g/min) were calculated over periods 1 to 5 (0–20, 21–40, 41–60, 61–80 and 81–93 days in the test). The descriptive statistics of FBT (I, II) are shown in Figure 1.

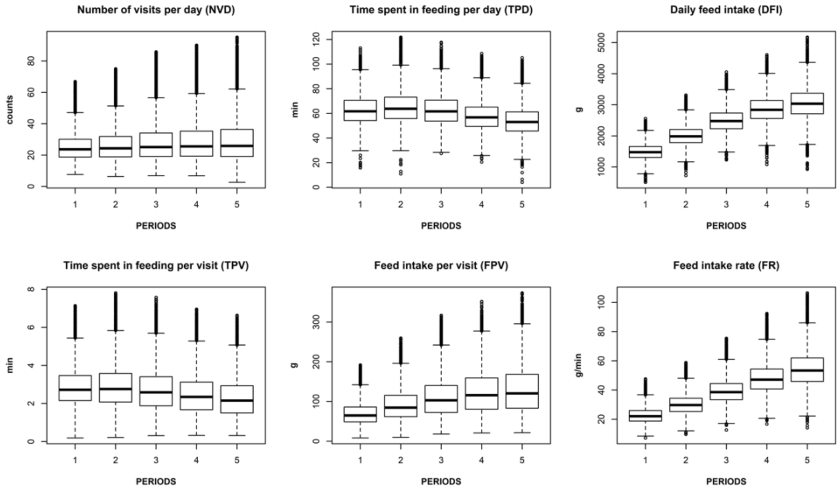


Figure 1. Box plots of feeding behaviour traits in pigs over the five periods.

3.1.3 Production data

The PT analysed in studies I and II were the ADG, FCR and BF. The ADG was calculated as the total growth during the test period (finishing body weight – initial body weight) divided by the number of test days. The FCR was defined as the feed consumption during the test period measured in feed units (1 feed unit is 9.3 MJ net energy) divided by the total growth during the test period. The BF was an average of two Hennessy Grading Systems (type GP4, Auckland, New Zealand) measurements: one at 8 cm off the midline of the carcass behind the last rib and one at 6 cm off the midline between the third and fourth ribs. In addition, the RFI was calculated using the difference between the observed and predicted DFI based on the initial body weight, ADG, BF, sex and year-season effect (I). The descriptive statistics of PT except RFI are shown in Figure 2 (I and II).

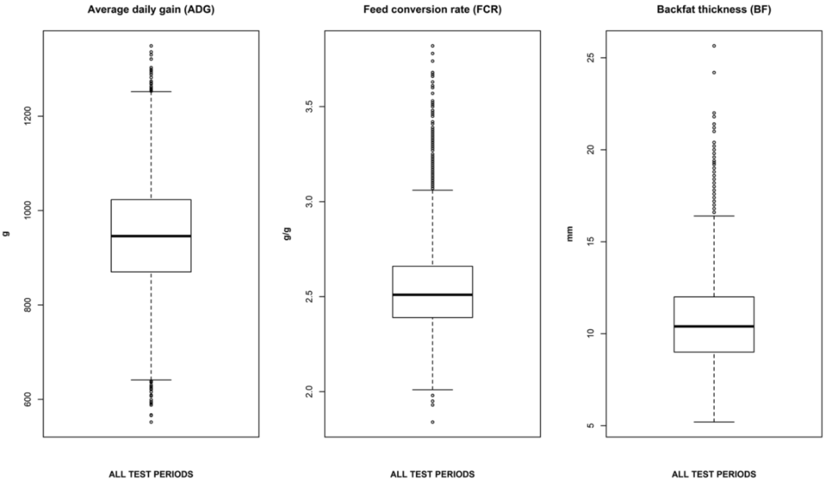


Figure 2. Box plots of the production traits in pigs over the entire test period.

3.1.4 Sickness data

The sickness data (III) were recorded daily by the test station staff members (Figen Oy) during the routine check (twice a day) and included the ID of the pig, the symptom(s) and the date. Out of 794,509 daily health observations, 13,018 were related to recorded symptoms, which were classified as cough (988), limp (6,603), loss of appetite (1,015), skin damage (1,471) and bitten tail (2,941). Pigs with symptoms were classified as "sick", and pigs that had no recorded symptoms were classified as "healthy" (781,491 observations). The prevalence of the sicknesses were 0.008, 0.001, 0.004, 0,001 and 0.002 for limp, cough, bitten tail, loss of appetite and skin damage, respectively. The total prevalence of sick pigs was 0.016. Antibiotics and other drugs were given only for the sick animals based on veterinary prescriptions.

3.1.5 Pedigree data

In study I, the pedigree data consisted of 5,396 Yorkshire pigs with observations and their ancestors down to four generations. In studies II and III, the pedigree data included 11,301 animals and their ancestors down to four generations as well. The average number of offspring with observations per sire was 16 in study I and 16.6 in studies II and III. The average additive genetic relationship between the pigs within a group was 0.08 (II). The pedigree data was checked and pruned using the Relax2 program (Strandén, 2014).

3.2 METHODS

3.2.1 Estimation of heritability and total heritable variation

The univariate mixed linear model (I) included sex and herd*year*season (*hys*) as fixed effects (except for RFI, which was corrected for these effects beforehand) and animal, litter and batch*pen (*bp*) as random effects for all FBT and PT. The heritability estimates (I) were calculated using the following formula:

1) $$h^2 = \frac{\sigma_a^2}{\sigma_a^2+\sigma_{bp}^2+\sigma_l^2+\sigma_e^2}$$

where σ_a^2 is the variance of the animal, σ_{bp}^2 is the variance of batch*pen (*bp*), σ_l^2 is the variance of litter and σ_e^2 is the residual variance of the trait.

For the social interaction study (II), single-trait SGE models were applied. Two SGE models were tested: a fixed group size (*n*) model with eight randomly selected pigs from each pen and a variable group size model (group size varied from 8 to 12, and the average group size was 9.8 ± 1.19). The linear mixed model included sex, breed and herd*year*season as fixed effects and group (batch*pen), litter, the genetic effects of the animal itself (DGEs) and those of its pen mates (SGEs) as random effects. The total heritable variation and classical heritability and the corresponding variance components were calculated using the following formulas:

The variance of the TBV (σ_{TBV}^2; Bijma et al., 2007a, b):

2) $$\sigma_{TBV}^2 = \sigma_{AD}^2 + 2(n-1)\sigma_{A_{DS}} + (n-1)^2\,\sigma_{AS}^2$$

where σ_{AD}^2 is the variance of DGEs, σ_{AS}^2 is the variance of SGEs and $\sigma_{A_{DS}}$ is the covariance between DGEs and SGEs.

The total phenotypic variance (σ_P^2):

3) $$\sigma_P^2 = \sigma_{AD}^2 + \sigma_g^2 + \sigma_l^2 + \sigma_e^2 + (n-1)\sigma_{AS}^2 + (n-1)r[2\sigma_{A_{DS}} + (n-2)\sigma_{AS}^2]$$

where σ_g^2 is the variance of group effect, σ_l^2 is the variance of litter effect and σ_e^2 is the residual variance. The average number of pigs in each group is *n*, and the average relatedness within groups is *r* (*r* = 0.08).

The total heritable variation (T^2):

4) $T^2 = \frac{\sigma_{TBV}^2}{\sigma_P^2}$

The classical heritability (h^2):

5) $h^2 = \frac{\sigma_{AD}^2}{\sigma_P^2}$

3.2.2 Genetic correlation

A bivariate model was used to estimate the genetic correlations between feeding behaviour and PT (I). For the bivariate model, the (co)variance matrix of the normally distributed additive genetic effects was $\mathbf{A}\otimes\mathbf{G}$, where **A** is the numerator relationship matrix, **G** is the genetic (co)variance matrices of the traits and $\otimes$ denotes the Kronecker product. The (co)variance matrix of the normally distributed litter, batch*pen and residual effects were $\mathbf{I}\otimes\mathbf{B}$, $\mathbf{I}\otimes\mathbf{C}$ and $\mathbf{I}\otimes\mathbf{R}$, respectively, where **I** is an identity matrix and **B**, **C** and **R** are the (co)variance matrices for litter, batch*pen and residual effects, respectively.

The variance and covariance components were estimated with **REML** method (Patterson and Thompson, 1971) using the DMU software (Madsen and Jensen, 2013).

3.2.3 Xgboost algorithm

Xgboost is a gradient boosted tree that attempts to predict a target variable by combining an ensemble of estimates from a set of simpler and weaker models which works powerfully in ML competitions by virtue of its robust handling of different data types as well as hyperparameters that can be tuned accordingly. Xgboost can be used in both regression and classification problems. Xgboost requires some implementation such as tuning hyper-parameters, which are optimisation parameters that tune the performance of ML algorithms (Bergstra and Bengio, 2012), and feature selection as predictors for the models to achieve robust and better performance. In this study, the hyper-parameters were selected as the number of boosting iterations (*nrounds*), maximum depth of a tree (*max-depth*) and *eta* that controls the learning rate. Xgboost combines weak base classifiers into a strong classifier. At each iteration of the training process, the residual of a base classifier is used in the next classifier to optimise the objective function. The objective of the classification model was binary *(binary:logistic)* and the model was fitted by minimizing the binary classification error rate. In addition, we used an additional parameter (scale_pos_weight) in the models to control the balance of classes weights due to the imbalanced data set. The parameter was calculated as the proportion of the number of sick observations to number of healthy observations. Therefore, in this study, the Xgboost algorithm was applied using the R package Xgboost (Chen et al., 2018) in R 3.6.1 software (R Core Team, 2019).

3.2.4 Features and window length selection for Xgboost

The features used in Xgboost (III) included the original daily observations (NVD, TPD and DFI), daily rank of TPD within the group of pigs (DRTPD), daily rank of DFI within the group of pigs (DRDFI) and residuals of TPD (ResTPD) and DFI (ResDFI). The daily ranks relate the rank of the animal's observation (NVD, TPD, or DFI) among the observations of the other pigs in the same pen, while the residuals of TPD and DFI were calculated by fitting a polynomial (quadratic) regression model to the whole dataset (III):

6) $$y_i = \mu_i + b_0 + b_1 * age_i + b_2 * age_i^2 + e_i,$$

where y_i is either the TPD or DFI of a pig i, b_0 is overall mean, $age_i$is the age of the pig i related to observation y_i, b_1 and b_2 are the linear and quadratic regression coefficients and e_i is the residual used in ML.

In addition, parameters form two normal distributions (one representing possible normal visits and the other representing possible visits without eating) were fitted to daily visit interval data using a maximum likelihood method and considered as features including p the proportion of intervals belonging to the first distribution, σ_1 and σ_2 the standard deviations and μ_1 and μ_2 the means of the distributions (e.g., Tolkamp et al., 1998). Logarithm to base e of the time between the feedings was used instead of the raw intervals.

Features used in ML were either from one, three or seven days:

a) one-day window length: The only features of the day of the health status classification were used in prediction.

b) three-day window length: The final features were the mean of the features of the day of the health status and the previous two days.

c) seven-day window length: The final features were the mean of the features of the day of the health status and the previous six days.

Based on the three- and seven-day window features, new features “delta” and “SD” were calculated for the seven-day window, with delta as a difference between the means of the features of the seven-day window and the three-day window and SD as a standard deviation of the features of the seven-day window. Windows were overlapping.

An alternative labelling of sick animals was tested with the seven-day window model. In the Alt-1 model, only the symptoms “limp” and “loss of appetite” were labelled as “sick”, and in the Alt-2 model, only the symptoms “bitten tail” and “skin damage” were labelled as “sick”.

3.2.5 Cross validation, performance metrics and feature importance

Prior to the performance measurement of the Xgboost algorithm, the data including sickness and feature information were split into the training (70%) and testing (30%) datasets (III). The observations were stratified by the symptoms and pig ID; thus, the proportion of sick and healthy observations was the same in both data sets.

As the first step in predictions, we applied 10-fold cross-validation (CV) to optimise the features and hyperparameters (the number of boosting iterations [nrounds], maximum depth of a tree [max-depth] and eta that controls the learning rate as well as gamma, lambda, subsample). CV was also applied to avoid under/overfitting of the models. For the 10-fold CV, the training data set were divided into 10 sets of equal size. In each validation step, nine of the sub-sets were used to train the model, and one sub-set was used to validate the model. The set of hyperparameters that gave the best performance metric (based on AUC) from the model was selected for the final performance testing.

The performance of the models was assessed based on sensitivity (TP / [TP + FN]), specificity (TN / [TN + FP]), precision (TP / [TP + FP]), F1-score ([2 x Precision x Sensitivity] / [Precision + Sensitivity]) and the AUC. The model was evaluated as non-informative with an AUC ≤ 0.50, weak with an AUC of 0.50 to 0.70, accurate with an AUC of 0.70 to 0.90 and highly accurate with an AUC ≥ 0.90 (Swets, 1988; Greiner et al., 2000).

In study III, we also obtained the feature importance using the ability of Xgboost to remove the non-informative or redundant predictors from the model (Chen et al., 2018). The importance matrix was produced from each model while fitting the Xgboost models. In this matrix, the "gain" metric indicates the relative contribution of the corresponding feature to the model calculated by taking each feature's contribution for each tree in the model. The "cover" metric indicates the relative number of observations related to this feature and the frequency, which is the percentage of the relative number of times a particular feature occurs in the trees of the model. An obtained score of each feature is based on how much more information about the class is gained when using that feature. We quantified the importance of features by "feature gain". R software (R Core Team, 2019) with the R packages xgboost and caret (Kuhn et al., 2018) were used in study III.

4 RESULTS AND DISCUSSION

This thesis consists of three studies that focus on the effects of FBT and PT in Finnish pig breeds as well as their welfare using feeding behaviour and health information. The studies include an estimation of genetic parameters using the classical animal model (I) and the social genetic model (II). Finally, since timely disease detection is important for animal welfare, the prediction models using the Xgboost-method were built based on feeding behaviour and health information data (III).

4.1 Estimated heritabilities of feeding behaviour and production traits

The heritabilities of FBT were estimated for five periods; however, only the smallest and largest estimates are presented in Table 1. Standard error (**SE**) of the estimates were between 0.05 and 0.07 for all traits. In general, the estimated heritabilities of FBT in average were moderate (0.30; Table 1). The highest heritability estimate was obtained for TPV (0.47) and the lowest for DFI (0.17). The range in estimates over the periods was largest for DFI and lowest for TPD. Even though slightly different estimates were obtained over the periods, the correlation between the estimates of the same trait over the periods was high (see study I and Section 4.2.); thus, the genetic basis of the trait is the same over periods. The lowest heritability estimates were observed during the first period of testing time where the young pigs may have been more vulnerable to environmental factors than older pigs. On the other hand, the statistical model with sex, herd*year*season, litter and batch*pen effects may not have been sensitive enough to explain all the environmental factors influencing the feeding behaviour of young pigs. Furthermore, some social factors such as adaptation to new pen mates and feeding systems may have influenced the estimates. The heritability estimates for the ADG, FCR and RFI were moderate, while for BF, the estimate was high (0.57).

Table 1. Estimated heritabilities (h^2) of feeding behaviour and production traits based on study I (first column) and the literature (second column).

Trait	h^2	h^2
FBT		
NVD	0.31-0.41[a]	0.31[1], 0.34[2,7], 0.38[6], 0.43[3,5], 0.47[4], 0.49[8]
TPD	0.28-0.37[a]	0.08[2], 0.17[8] , 0.24[6], 0.36[5], 0.37[1], 0.43[3], 0.45[4], 0.46[7]
DFI	0.17-0.32[a]	0.16[6], 0.20[4], 0.21[2,9], 0.22[3,8], 0.25[10], 0.39[7], 0.42[5]
TPV	0.34-0.47[a]	0.11[2], 0.27[6], 0.39[4], 0.42[3], 0.44[7], 0.45[5], 0.60[1]
FPV	0.32-0.44[a]	0.27[2], 0.28[8], 0.35[6], 0.44[7], 0.50[4], 0.51[3], 0.53[5]
FR	0.19-0.29[a]	0.04[2], 0.27[8], 0.29[6], 0.41[7], 0.44[3], 0.46[4], 0.49[5]
PT		
ADG	0.25[a]	0.05[9], 0.36[7], 0.37[10], 0.43[3]
BF	0.57[a]	0.41[10], 0.50[7], 0.53[9], 0.54[3], 0.56[1]
FCR	0.28[a]	0.19[3,5], 0.21[8], 0.28[7], 0.30[9], 0.45[10]
RFI	0.32[a]	0.19[12], 0.21[8], 0.26[9], 0.24[10], 0.35[11]

[a] Kavlak and Uimari (2019), the smallest and the largest values.

[1] Rohrer et al. (2013) – Landrace, Duroc and Yorkshire; [2] Hall et al. (1999) – Large White; [3] Von Felde et al. (1996) – Landrace and Large White; [4] Kalm et al. (1996) – Landrace and Large White; [5] Labroue et al. (1997) – Large White; [6] de Haer and de Vries (1993) – Landrace and Yorkshire; [7] Schulze et al. (2003) – Landrace and Large White; [8] Sermyagin et al. (2020) – Duroc; [9] Saintilan et al. (2013) – Large White; [10] Gilbert et al. (2007) – Large White; Foster et al. (1983) – Landrace[11] and Large White.[12]

The estimated heritabilities of FBT (I) are in line with estimates presented in the literature (Table 1). The highest heritability estimate for FBT presented in the literature is 0.6 for TPV (Rohrer et al., 2013), and the lowest estimate is 0.04 for the FR (Hall et al., 1999). In the present study, the heritability estimates for the NVD ranged from 0.31 to 0.41 within five test periods. These estimates are similar to estimates by Rohrer et al. (2013), Schulze et al. (2003) and de Haer and de Vries (1993; 0.31, 0.34 and 0.38, respectively). The heritability estimates for the TPD (0.28–0.37) and FPV (0.32–0.44) were also in line with other studies: the lowest estimates were reported by Hall et al. (1999; 0.34, 0.08, 0.21, 0.11,0.27 and 0.04 for the NVD, TPD, DFI, TPV, FPV and FR, respectively), and the highest were reported by Labroue et al. (1997; 0.43, 0.36, 0.42, 0.45, 0.53 and 0.49 for the NVD, TPD, DFI, TPV, FPV and FR, respectively). The heritability estimates for DFI (range 0.17–0.32) were similar to those reported by others (Table 1) except that Labroue et al. (1997) and Schulze et al. (2003) reported higher heritabilities (0.42 and 0.36, respectively). In addition, Von Felde et al. (1996) also estimated the heritability of DFI for different periods of the test (period 1: 0.16, period 2: 0.24, period 3: 0.30, period 4: 0.27, period 5: 0.26). They also obtained some variation in the estimated heritabilities over testing time, and the maximum value ($h^2 = 0.30$) was obtained for the middle of the test period (period 3). Finally, heritability estimates for the FR (0.19–0.29) were close to those of de Haer and de Vries (1993; 0.29).

The estimates obtained for PT (I) are in line with estimates presented in the literature (Table 1). In this study (I), the heritability estimate for the ADG (0.25) is a bit lower than reported in other studies, and heritability estimates for BF and RFI are generally higher than those reported in other studies (Table 1). However, the estimates obtained from other Large White and Yorkshire populations (Schulze et al., 2003; Saintilan et al., 2013) are quite similar to our estimates. Schulze et al. (2003) reported the heritabilities of ADG, BF and FCR at 0.36, 0.50 and 0.28 while Saintilan et al. (2013) obtained the heritability of BF, FCR and RFI at 0.53, 0.30 and 0.26, respectively. In contrast, our heritability estimate for RFI (0.32) is far greater than the estimate obtained for the Northern Irish Large White population (0.19; Foster et al., 1983) but similar to the estimate obtained for the Northern Irish Landrace (0.35) population (Foster et al., 1983). It has also been found that the heritability of RFI is lower for boars (0.14) compared to females and castrated males (0.24; Gilbert et al., 1997). In study I, sex was considered as a systematic fixed effect, and sex-specific heritabilities were not estimated. In general, PT that contained feed intake as a component (e.g., the FCR and RFI) showed similar estimates of heritability to the DFI over the periods.

Variance components related to FBT were estimated from the mean values of each of the five test periods (approximately 20 days long; studies I and II). An alternative way to analyse the data would have been to use daily records and a random regression model (Schaeffer, 2004). For the PT, the number of weight measurements was small (less than six during the test period, personal communication with Timo Serenius 01.09.2021); thus, the benefit of using all available weight measurements instead of a single ADG value is small (Mrode and Kennedy, 1993). For the FBT, daily values were available, but instead of estimating daily heritabilities, periodical ones (five periods) were used. The gain in the accuracy of heritability estimates from daily records or usability of the estimates compared to periodical values is expected to be small (Wetten et al., 2012).

4.2 Genetic correlations between feeding behaviour and production traits

FBT has no direct economic value; however, if FBT are correlated with PT, then FBT can be used as auxiliary traits in pig breeding. Genetic correlations were estimated both between FBT and between FBT and PT (I). In addition, the genetic correlations of the same trait between different test periods were estimated (I).

Genetic correlations between the same FBT over the testing periods were generally high (Table 2 and study I). The strongest genetic correlations were obtained between contiguous periods varying from 0.96 ± 0.01 to 0.99 ± 0.01 for all FBT and the weakest between periods 1 and 5 varying from 0.68 ± 0.09 (TPD1 vs. TPD5) to 0.83 ± 0.06 (NVD1 vs. NVD5). Based on the results, the genetic basis of the trait over time is the same (between adjacent periods) or similar (between the first and last periods).

Strong genetic correlations between all FBT were obtained (Table 2). For example, genetic correlations between the NVD and TPV ranged from −0.79 ± 0.05 to −0.88 ± 0.03, and between the NVD and FPV, genetic correlations were over −0.90 ± 0.02 in all testing periods. In addition, a strong positive correlation between FPV and TPV (0.83 ± 0.04 to 0.90 ± 0.02) was discovered. These strong correlations indicate that the NVD, FPV and TPV are similar traits; pigs that have genetic background tend to visit the feeder (NVD) less often than other pigs tend to eat more during the visit (TPV) and spend more time feeding per visit (TPV). This is natural feeding behaviour; pigs either consume feed often but in small portions or rarely in large portions (Rauw et al., 2006). In addition, the FR had a moderate genetic correlation with the DFI (0.22 ± 0.14 to 0.40 ± 0.12) over the periods, which indicates that faster consumers (g/min) aimed to have higher genetic potential for DFI. The SE of the estimates varied from 0.01 to 0.17 (I).

In previous studies (Table 3), Do et al. (2013) also reported highly negative genetic correlations between the NVD and FPV (−0.95 ± 0.02) as well as between the NVD and TPV (−0.91 ± 0.02) in Danish Yorkshire populations while the DFI and FR had a significant positive correlation (0.36 ± 0.06). Schulze et al. (2003) also reported highly negative genetic correlations between the NVD and FPV (−0.92 ± 0.01) and between the NVD and TPV (−0.81 ± 0.02), a highly positive correlation between the FPV and TPV (0.86 ± 0.01) and a moderately positive correlation between the DFI and FR (0.20 ± 0.05) in the population of Yorkshire and Landrace pigs based on dam lines.

Table 2. Range of the genetic correlation within the same trait between different periods (diagonal) and between the feeding behaviour traits (off-diagonal) and between feeding behaviour traits and production traits over the testing time.

Traits	**Range (min–max) of genetic correlation**					
FBT	NVD	TPD	DFI	TPV	FPV	FR
NVD	**0.83 — 1**	-0.03 — 0.50	0.01 — 0.18	-0.88 — -0.79	-0.97 — -0.90	-0.06 — 0.35
TPD		**0.68 — 1**	0.17 — 0.31	0.09 — 0.57	-0.36 — 0.26	-0.80 — -0.66
DFI			**0.81 — 1**	0.04 — 0.17	0.25 — 0.36	0.22 — 0.40
TPV				**0.80 — 1**	0.83 — 0.90	-0.63 — -0.31
FPV					**0.80 — 1**	-0.24 — 0.24
FR						**0.76 — 1**
PT						
ADG	-0.10 — -0.01	-0.04 — 0.31	0.67 — 0.91	0.00 — 0.33	0.17 — 0.46	0.29 — 0.38
BF	-0.22 — -0.05	-0.05 — 0.22	0.60 — 0.72	0.14 — 0.31	0.32 — 0.47	0.11 — 0.51
FCR	0.03 — 0.19	0.18 — 0.36	0.73 — 0.88	-0.01 — 0.10	0.08 — 0.21	0.04 — 0.31
RFI	0.08 — 0.17	0.34 — 0.45	0.83 — 0.89	-0.00 — 0.15	0.04 — 0.18	0.05 — 0.19

FBT: feeding behaviour traits, PT: production traits, NVD: number of visits per day, TPD: time spent in feeding per day, DFI: daily feed intake, TPV: time spent feeding per visit, FPV: feed intake per visit, FR: feed intake rate, ADG: average daily gain, BF: backfat thickness, FCR: feed conversion rate, RFI: residual feed intake.

Given the generally high SE (from 0.04 to 0.16), genetic correlations between FBT and PT traits did not differ from 0. Only DFI had strong positive genetic correlations with all PT (Table 2). In addition, the FPV had a moderately positive (unfavourable) correlation with BF, which shows that Finnish Yorkshire pigs that consume large quantities of feed per visit tend to gain more BF. The genetic correlation between FPV and BF increased from periods 1 to 5 from 0.14 ± 0.13 to 0.31 ± 0.12 (I). BF was also correlated with the FR; thus, animals with a high FR tend to gain more BF. Furthermore, the range of the genetic correlation between the TPD and RFI varied from 0.34 ± 0.13 to 0.45 ± 0.13 over the testing time; thus, animals with more TPD also show an increase in RFI.

Strong correlations between different FBT have also been reported in the literature (Table 3). The highest genetic correlations were reported by Do et al. (2013) and Ding et al. (2018) between DFI and RFI in the Duroc population at 0.95 ± 0.13 and 0.98 ± 0.13, respectively. Generally, the lowest genetic correlations were reported between the FR and PT (Do et al., 2013; Schulze et al., 2003; Young et al., 2012). The genetic correlations between the NVD and FCR and RFI estimated by Do et al. (2013; 0.50 ± 0.09 and 0.40 ± 0.07) were higher than our estimates, while Do et al. (2013) obtained similar estimates for the genetic correlations between the NVD and ADG (-0.26 ± 0.09) and BF (-0.05 ± 0.08). We obtained a lower genetic correlation between the TPD and PT compared to other studies (Table 3). Unlike our estimate (ranging from 0.73 ± 0.18 to 0.88 ± 0.08), McSweeny et al. (2003) reported a weak genetic correlation between the DFI and FCR (0.01 ± 0.18) while Do et al. (2013) and Ding et al. (2018) reported estimates similar to ours (0.67 ± 0.05 and 0.75 ± 0.06, respectively). For the genetic correlation between the FCR and FPV, similar estimates to ours were reported by Schulze et al. (2003; 0.05 ± 0.08) and McSweeny et al. (2003; 0.04 ± 0.16). Finally, most literature estimates of genetic correlation between the TPV and PT were similar to our estimates with the exception of those by Do et al. (2013).

Table 3. Genetic correlation between feeding behaviour traits and production traits based on the literature.

Traits	**Genetic correlation**					
FBT	NVD	TPD	DFI	TPV	FPV	FR
NVD		0.19[1], -0.21[4], 0.38[5]	0.27[1], 0.02[2], -0.14[4]	-0.91[1], -0.81[2]	-0.95[1], -0.92[2], -0.86[4]	-0.04[1]
TPD			0.29[1], 0.41[2], 0.97[4]	0.24[1], 0.62[4]	-0.16[1], 0.61[4]	-0.87[1], -0.89[3]
DFI				-0.16[1], 0.50[4]	0.19[1], 0.35[3], 0.48[4]	0.36[1], 0.20[2]
TPV					0.88[1], 0.86[2]	-0.30[1]
FPV						0.10[1]
FR						
PT						
ADG	-0.26[1], 0.01[2]	0.23[1], 0.31[2], 0.32[3]	0.45[1], 0.77[3], 0.78[4]	0.35[1], 0.51[4]	0.40[1], 0.39[3], 0.52[4]	0.02[1], 0.17[3]
BF	-0.05[1], 0.05[2]	0.10[1], 0.30[3], 0.39[4]	0.29[1], 0.52[3], 0.45[4]	0.26[1], 0.002[4]	0.29[1], 0.26[3], 0.01[4]	0.13[1] ,0.16[3]
FCR	0.50[1] ,0.04[2]	0.11[1], 0.12[4], 0.22[5]	0.67[1], 0.01[4], 0.75[5]	-0.43[1], 0.05[4]	-0.36[1], 0.05[2], 0.04[4]	0.25[1], 0.10[2]
RFI	0.40[1], 0.14[5]	0.25[1] ,0.39[3], 0.63[5]	0.95[1],0.65[3], 0.98[5]	-0.28[1], 0.07[3]	0.16[1], -0.07[3], 0.13[6]	0.28[1], -0.04[3]

[1]Do et al. (2013) – Duroc, [2]Schulze et al. (2003) – Landrace and Large White, [3]Young et al. (2012) – Yorkshire, [4]McSweeny et al. (2003), [5]Ding et al. (2018) – Duroc, [6]Von Felde et al. (1996) – Landrace and Large White.

All genetic correlations between the PT were statistically significant (Table 4). The FCR had a high genetic correlation with BF (0.79 ± 0.07) but a moderate one with the ADG (0.37 ± 0.16). Unlike the FCR, the genetic correlation between RFI and BF was moderate (0.38 ± 0.12) while RFI had considerable genetic correlation with the ADG (0.63 ± 0.13) and with the FCR (0.79 ± 0.06). Additionally, Do et al. (2013) reported a high genetic correlation between the ADG and BF in the Duroc (0.75 ± 0.05) population but quite a low one between the ADG and RFI (0.16 ± 0.09). Furthermore, similar to our estimate, Do et al. (2013) reported a high genetic correlation between RFI and the FCR (0.63 ± 0.13). Therefore, selection for RFI has a stronger effect on ADG than on BF, but on the other hand, selection for FCR has a stronger effect on BF than on the ADG. Many other studies have reported significant genetic correlations between PT including a very high correlation between RFI and the FCR (0.97 ± 0.01; Ding et al., 2018), a moderate correlation between RFI and the ADG (0.41 ± 0.10; Von Felde et al., 1996), a moderate correlation between the ADG and BF (0.55 ± -0.24; McPhee et al., 1979) and a relatively low correlation between the ADG and BF (0.23 ± 0.04; Labroue et al., 1997).

Table 4. Genetic correlation within the production traits. Standard errors are given in parentheses.

PT	ADG	BF	FCR	RFI
ADG	**1**	0.37 (0.12)	0.37 (0.16)	0.63 (0.13)
BF		**1**	0.79 (0.07)	0.38 (0.12)
FCR			**1**	0.79 (0.06)
RFI				**1**

PT: production traits, ADG: average daily gain, BF: backfat thickness, FCR: feed conversion rate, RFI: residual feed intake.

Based on the results (I), FBT are moderately inherited, but since only DFI had a high genetic correlation with PT in Finnish Yorkshire pigs, there may be no reason to include FBT in the breeding programme. However, as Herrera-Cáceres et al. (2019) present in their study, when the social interaction is included, the genetic correlation between FBT and PT becomes stronger. Therefore, FBT could still be considered in breeding programmes.

4.3 Estimated social genetic parameters of feeding behaviour and production traits

In study II, we expanded the animal model (I) by including the SGEs. Since the heritability estimates for FBT and PT were similar in study I based on Finnish Yorkshire and in a master's thesis by Riikimaki (2019) based on Finnish Landrace, we combined the data in study II to include both Finnish Yorkshire and Finnish Landrace animals and their F1-crosses. However, Bijma (2010) and Ødegård and Olesen (2011) reported that the estimated variance of SGEs would be more accurate if the data consisted only of one population and the groups were formed to include only a few families per pen. However, analysing breeds separately would not have been possible in our data because pigs with different breed origins shared the same pen.

The SGE effect was estimated using either a fixed ($n = 8$) or variable group size ($n = 9.8$) model (II). The group sizes can vary due to different reasons in testing. In our dataset, the animals were mostly removed from the test due to health issues (e.g., limp, loss of appetite, bitten tail, etc.). Unlike Ask et al. (2020), we did not implement a fine-tuned approach in our analysis. Instead, we first applied a simple model to randomly sample an equal number of pigs ($n = 8$) from each pen for a fixed group size model and compared these results to those from a variable group size model ($n = 9.8$). Because the models produced similar classical heritability and total heritability variation estimates, only those related to variable group size are presented here.

Based on the results (II), estimations of SGEs for FBT and PT were not statistically significant except for the TPV, FPV and FCR. In contrast with our findings, based on the data of 1,144 Duroc pigs, Herrera-Cáceres et al. (2019) reported a considerable SGE in the NVD and FR during the fattening period. Within socially affected PT, Bouwman et al. (2010) did not find a significant SGE in the ADG, and neither did we, but Chen et al. (2008) and Nielsen et al. (2018) found a significant SGE for the ADG. Herrera-Cáceres et al. (2019) reported a significant SGE in the FCR as we did, while Bergsma et al. (2008) reported a significant SGE in BF and DFI. In conclusion, based on study II and the literature, the SGE is either significant or insignificant for both FBT and PT, and even if the SGE is statistically significant, the SGE is usually small compared to the DGE.

Estimates of classical heritability varied from 0.14 (DFI1) to 0.39 (TPD2 and TPD3) for FBT and from 0.29 (ADG and FCR) to 0.38 (BF) for PT (Table 5). The estimates are similar to those presented in study I. The SE of the estimates were around 0.04 for all traits. The estimates of the total heritable variation were generally higher than the estimates of the classical heritability for FBT, where the estimations of DFI range from 0.17 ± 0.10 (FR3) to 0.85 ± 0.16 (TPV3; Table 5). However, the estimates of total heritable variation were similar to the estimates of classical heritability for PT except for BF ($T^2 = 0.48 \pm 0.12$ vs. $h^2 = 0.38 \pm 0.07$) and the FCR ($T^2 = 0.37 \pm 0.12$ vs. $h^2 = 0.29 \pm 0.04$). In contrast with our findings, Herrera-Cáceres et al. (2019) reported high total heritabilities (compared to classical heritabilities) for the NVD ($T^2 = 0.93 \pm 0.49$ vs. $h^2 = 0.46 \pm 0.09$), DFI ($T^2 = 0.29 \pm 0.29$ vs. $h^2 = 0.25 \pm 0.08$), TPV ($T^2 = 0.67 \pm 0.\ 30$ vs. $h^2 = 0.47 \pm 0.09$) and FR ($T^2 = 0.39 \pm 0.29$ vs. $h^2 = 0.32 \pm 0.08$). However, for many of these estimates, SEs were high. Herrera-Cáceres et al. (2019) reported that the SGE was not important for the ADG ($T^2 = 0.22 \pm 0.18$ vs. $h^2 = 0.22 \pm 0.09$) and BF ($T^2 = 0.51 \pm 0.28$ vs. $h^2 = 0.35 \pm 0.11$) while they found a larger effect of the SGE for the FCR ($T^2 = 0.55 \pm 0.43$ vs. $h^2 = 0.24 \pm 0.09$) as we did (Table 5). Again, the SE were high for all estimates. Some other studies also investigated the impact of the SGE but mostly on the ADG and other PT. Bergsma et al. (2008) found a significant contribution of the SGE for the ADG ($T^2 = 0.71 \pm 0.08$ vs. $h^2 = 0.21 \pm 0.02$), DFI ($T^2 = 0.70 \pm 0.17$ vs. $h^2 = 0.17 \pm 0.03$) and BF ($T^2 = 0.41 \pm 0.04$ vs. $h^2 = 0.35 \pm 0.02$) from the data of 14,032 crossbred pigs while Nielsen et al. (2018) estimated the SGE for the ADG using separated bivariate models for Danish Landrace gilts and boars and found that the T^2 was higher for the boars (0.32 ± 0.02) than for the gilts (0.27 ± 0.01). We also tested the effect of dilution in the SGE model with two dilution parameters $d = 0.5$ and $d = 1$, but the effect of including the dilution parameter in the SGE model on estimated variance components was small. To conclude, our findings are similar to other published results indicating that the SGE has a considerable impact on FBT (at least in the TPV and FPV) and in the FCR, but it is less important in the BF and ADG.

Table 5. The range of classical heritability (h^2) and total heritable variation (T^2) of feeding behaviour traits and production traits (II) from periods 1 to 5 using the variable group size model. Standard errors are given in parentheses.

Trait	h^2	T^2
FBT		
NVD	0.26 (0.04) – 0.33 (0.04)	0.30 (0.11) – 0.50 (0.14)
TPD	0.28 (0.04) – 0.39 (0.04)	0.27 (0.10) – 0.49 (0.12)
DFI	0.14 (0.03) – 0.30 (0.04)	0.33 (0.09) – 0.73 (0.15)
TPV	0.28 (0.04) – 0.34 (0.04)	0.53 (0.15) – 0.85 (0.16)
FPV	0.28 (0.04) – 0.33 (0.04)	0.54 (0.13) – 0.77 (0.16)
FR	0.22(0.03) – 0.28 (0.04)	0.17 (0.10) – 0.44 (0.15)
PT		
ADG	0.29 (0.04)	0.29 (0.11)
BF	0.38 (0.07)	0.48 (0.12)
FCR	0.29 (0.04)	0.37 (0.12)

FBT: feeding behaviour traits, PT: production traits, NVD: number of visits per day, TPD: time spent in feeding per day, DFI: daily feed intake, TPV: time spent feeding per visit, FPV: feed intake per visit, FR: feed intake rate, ADG: average daily gain, BF: backfat thickness, FCR: feed conversion rate.

4.4 Predicting pigs' health status using the feeding behaviour data

Xgboost is as an ML method used for the designed models that combines weak base classifiers into a strong classifier. At each iteration of the training process, the residual of a base classifier is used in the next classifier to optimise the objective function. Therefore, we chose the Xgboost algorithm for our model predictions. The designed models were evaluated based on classification performance metrics including sensitivity, specificity, precision, F1-score and AUC. In general, the accuracy of the prediction based on the AUC ranged from 0.70 to 0.90; the best accuracy was achieved with the seven-day window model (Table 6). The accuracies with both the training and testing dataset were similar, which indicates that the over/underfitting was avoided by optimising the hyper-parameters in the models.

Table 6. Results from the models based on testing data and training (average from 10-fold CV given in the parentheses)

	Window length (days)				
Metrics	1	3	7	Alt-1	Alt-2
AUC	0.70 (0.71)	0.73 (0.75)	0.80 (0.81)	0.83 (0.85)	0.77 (0.80)
Precision	0.03 (0.03)	0.03 (0.03)	0.04 (0.04)	0.03 (0.03)	0.01 (0.01)
Sensitivity	0.60 (0.61)	0.63 (0.65)	0.67 (0.72)	0.67 (0.71)	0.67 (0.74)
Specificity	0.67 (0.67)	0.69 (0.69)	0.73 (0.73)	0.78 (0.81)	0.70 (0.70)
F1-score	0.06 (0.06)	0.06 (0.06)	0.07 (0.08)	0.05 (0.06)	0.02 (0.02)

The sensitivity and specificity of the models were reasonable in all window lengths; however, the models had a very low precision (the models predicted more sick animals than reported in the data) as well as low F1 scores. The seven-day window had the best performance with 67% sensitivity, 73% specificity, 4% precision and a 7% F1 score. In addition to our findings from basic models, other designed alternative models with the seven-day window model gave either better (Alt-1 model) or similar (Alt-2 model) performances based on the AUC than the performance of the actual seven-day window model (Table 6). Similar to our study, Thomas et al. (2021) analysed individual water and feed intake related to the weight of 102 weaned piglets as well as their link to diarrhoea using an ML approach with seven different methods; however, the methods that they used in their study failed to detect individually diarrheic pigs using water and feed intake related to weight due to substantial individual instability except for the Gaussian naïve Bayes classification, which performed the best with 73.7% sensitivity even though the authors were considered unacceptable since more than 25% of the sick piglets were not detected by the model. As an alternative approach, Maselyne et al. (2018) also investigated if unusual behavioural changes in the feeding pattern of 152 pigs can be detected automatically and utilised as an indicator for health, welfare as well as productivity problems. Therefore, they developed warning systems using variables of the feeding system based on historical data and then validated and compared these systems online by comparing the alerts with comprehensive observations. Although they had considerably high specificity (98.7%) and accuracy (96.7%), they had the performance of the warning system as promising, but still requires improvements for the sensitivity (58.0%) and the precision (71.1%) since false alerts could compose lack of confidence on farmers for the system. In contrast with our findings, Gertz et al. (2020) achieved better classification performance than our models using the Xgboost algorithm where locomotor-related diseases were predicted using locomotion data collected from leg and neck sensors in a commercial farm of 397 dairy cows with 86% AUC, 81% F-score and precision as well as 78% specificity and 81% sensitivity. Furthermore, Alsaaod et al. (2012) predicted lameness in dairy cows using features created from the pedometric activity and behaviour data on lying down where they had a high precision indicated by achieving an average of 80% with models with different window segments. Using a similar window length approach as in our study, Piette et al. (2020) developed systems using different sensor technologies to detect the lameness in 2,000 Holstein Friesian cows from Swedish dairy farms, and they achieved very good classification performance with a high AUC performance (approximately 85%). Their finding was also slightly better in a longer window segment, which indicates that predicting the models with longer window segments may be time consuming but may produce better model performance. In conclusion, based on study III and the literature, the Xgboost algorithm is relatively efficient according to the performance metrics for predicting diseases.

Setting features for predictions is crucial for improving the performance of the classification as well as its robustness. In this study, the features were calculated from the feeding behaviour data within either short or long window segmentations. After predicting the models , ResTPD and ResDFI in the one-day and three-day window models as well as SD_ResTPD in the seven-day window model were determined the most informative features (III). Overall, the importance of the features ranged from 20 to 35% by the models within these features. Besides these most important features, other calculated features were included in the models as predictors, and the importance of other features was approximately 10%. However, this does not mean that these other features are not important as predictors. In general, the residuals of FBT were more beneficial in predicting pig sickness than absolute values were. These derived new features from DFI and the TPD can be a better indication of a possible health problem than absolute values are. Therefore, we recommend using features that indicate a deviation in an animal’s feeding behaviour from that its pen mates (rank) or from that of pigs of the same age (residual) rather than raw observations (NVD, TPD, DFI). Additionally, Alsaaod et al. (2012) found it very effective to derive new features (by bout analyses) using absolute traits where the new features were more informative than absolute values in the classification of sickness problems in dairy cows.

4.5 Implications and future research

The main objective of this thesis was to estimate the heritability of FBT and their correlation with PT in Finnish Yorkshire pigs, to estimate the SGEs on feeding behaviour and PT in pigs and to detect disease in pigs based on FBT using ML.

Based on the results, we did not find additional benefit from including feeding behaviour-related traits in the breeding programmes. However, when the genetic model was expanded to account for the SGE, we found that the total heritable variation (T^2) was notably higher than the classical heritability (h^2). Thus, an important application would be to implement SGEs in the estimation of breeding values of the FCR to simultaneously selecting for animals that have a high direct genetic merit for the FCR and also a positive effect on the FCR of the other pigs sharing the same pen.

Sustainable pig production with high animal welfare, low labour input and low medical usage has become very important. Sustainable production also allows high economic output. Breeding organisations such as Norsvin have indicated that animal welfare has become an increasing concern for breeders and customers (Norsvin, 2018). Additionally, Danbred aims to improve animal welfare by improving sow longevity and controlling inbreeding as well as attending to the conditions where pigs are moved from one place to another to avoid unnecessary agitation. Evaluation of animal welfare involves a complete evaluation process of the animal's physiological, behavioural, physical and emotional condition (Brito et al., 2020). This complete assessment needs to be made based on a combination of multiple traits including feeding behaviour and on some key factors such as feeding, housing and health conditions. Management and care of animals with improved welfare are safer and easier and require less time and medication than that of animals that suffer from poor welfare (Sinclair et al., 2019).

Regarding future research, applying feeding data to improve the welfare of pigs through selection for resilient pigs would be an interesting expansion of the research presented in this thesis. Resilience is the animals' ability to recover from the effects of physical and environmental stressors and disease challenges (Albers et al., 1987). It is expected that developing resilience enhances production efficiency sustainably, where the animals with a high resilience can maintain their performance under infection and any other stressors (Albers et al., 1987). Specifically, disease resilience aids in improving the pig industry with healthier and more profitable pigs for producers.

Feed intake and feeding behaviour are sensitive to diseases such as lameness and diarrhoea. Recently, Putz et al. (2019) used routinely collected feed intake data to quantify the resilience of growing pigs under a multifactorial natural disease challenge that was created to copy a commercial environment with high disease pressure to maximise the expression of resilience. In this study, radical reductions and day-to-day variation in feed intake among pigs were used to quantify resilience. However, these reductions and variations in feed intake cannot quantify resilience specifically since it was not possible to confirm that all changes in feed intake can relate to infectious diseases. Knap (2009) suggests that day-to-day variation in feed intake for each animal could be used to quantify environmental sensitivity such as resilience to heat stress. Since improvement in resilience has a direct effect on the sustainability of pork production and animal welfare and maintains the public acceptance of pork production, it also has a positive effect on economics by reducing the days of impaired growth and health due to sudden environmental changes such as changes in feedstuff or management or due to possible disease outbreaks. Therefore, selection of pigs that have high resilience will improve the wellbeing of pigs in commercial farms as well as their performance under any conditions.

The data used is this thesis can be extended with more recent data and extract new phenotypes related to resilience based on fluctuations in DFI and TPD to estimate the heritability of these resilience-related traits and to estimate their genetic correlations with PT. If these resiliency traits are heritable and do not have strong negative genetic correlations with PT, they can be used in selection to improve animal resiliency and welfare, leading to better pig welfare and increased sustainability of Finnish pork production.

5 CONCLUSION

This study investigated the genetic parameters of FBT and PT with and without SGEs and disease detection based on FBT in Finnish pig breeds. Heritabilities of FBT were moderate (around 0.3) and had no strong correlation with PT with the exception of DFI having strong genetic correlations with RFI (over 0.8; study I). The SGE was significant for the FBT and FCR but not for the ADG and BF (II). For those traits, the total heritable variation was considerably higher than the pure classical heritability. The fixed and variable group size models produced very similar estimates (II). Using the Xgboost ML method, we obtained the best performance for predicting pigs' daily health status with a seven-day window length (80% AUC, 7% F1-score, 67% sensitivity and 73% specificity). However, the precision was very low (0.04), possibly due to an imbalanced dataset (III).

Based on the results, 1) since FBT have no direct economic value and are not strongly collated with PT, FBT do not provide additional value for the current breeding programmes of the Finnish pig breeds; 2) it would be beneficial to include SGEs in the routine breeding value estimation of the FCR and use a TBV instead of a DBV as a selection criterion; 3) and the application of an ML approach to detect sick pigs based on FBT was promising.

For the future line of research, variation in daily activities (day-to-day variation, off-feed visits) in feeding behaviour could be used to detect resilient animals, those that show small variation in daily activities even under disease break-out or under sub-optimal conditions and stress factors that are common in pig production. Therefore, we assume that phenotypes derived from the FBT can be used as indicator traits to select for better resilience. This approach may enhance the welfare of pigs on commercial farms under different environmental conditions.

REFERENCES

Aberle, E.D., Merkel, R.A., Forrest, J.C. and Alliston, C.W. 1974. Physiological responses of stress susceptible and stress resistant pigs to heat stress. Journal of Animal Science. 38(5): 954-959.

Alsaaod, M., Römer, C., Kleinmanns, J., Hendriksen, K., Rose-Meierhöfer, S., Plümer, L., Büscher, W. 2012. Electronic detection of lameness in dairy cows through measuring pedometric activity and lying behavior. Applied Animal Behaviour Science. 142: 134–141.

Arango, J., Misztal, I., Tsuruta, S., Cubertson, M. and Herring, W. 2005. Estimation of variance components including competitive effects of Large White growing gilts. Journal of Animal Science. 83: 1241–1246.

Ask, B., Christensen, O.F., Heidaritabar, M., Madsen, P. and Nielsen, H.M. 2020. The predictive ability of indirect genetic models is reduced when culled animals are omitted from the data. Genetics Selection Evolution. 52: 8.

Bates, R.O., Morrow-Tesch, J.L. and Neal, S.M. 1995. The Effect of Feeder Space Allocation on Pig Performance. The Professional Animal Scientist. 11 (3): 135-138.

Bijma, P., Muir, W.M. and Van Arendonk, J.A.M. 2007a. Multilevel selection 1: quantitative genetics in inheritance and response to selection. Genetics. 175: 277–288.

Bijma, P, Muir, W.M., Ellen, E.D., Wolfand, B.J. and Van Arendonk, J.A.M. 2007b. Multilevel selection 2: estimating the genetic parameters determining inheritance and response to selection. Genetics. 175: 289–299.

Bijma, P. 2010. Estimating indirect genetic effects: precision of estimates and optimum designs. Genetics. 186: 1013–1028.

Bergsma, R., Kanis, E., Knol, E.F. and Bijma, P. 2008. The contribution of social effects to heritable variation in finishing traits of domestic pigs (Sus scrofa). Genetics. 178: 1559-1570.

Bergstra, J. and Bengio, Y. 2012. Random search for hyper-parameter optimization. Journal of Machine Learning Research. 13: 281–305.

Bouwman, A.C., Bergsma, R., Duijvesteijn, N. and Bijma, P. 2010. Maternal and social genetic effects on average daily gain of piglets from birth until weaning. Journal of Animal Science. 88: 2883–2892.

Brito, L.F., Oliveira, H.R., McConn, B.R., Schinckel, A.P., Arrazola, A., Marchant-Forde, J.N. and Johnson, J.S. 2020. Large-Scale Phenotyping of Livestock Welfare in Commercial Production Systems: A New Frontier in Animal Breeding. Frontiers in genetics. 11: 793.

Brown-Brandl, T., Rohrer, G.A. and Eigenberg, R.A. 2013. Analysis of feeding behavior of group housed growing–finishing pigs. Computers and Electronics in Agriculture. 96: 246–252.

Bus, J.D., Boumans, I.J.M.M., Webb, L.E. and Bokkers, E.A.M. 2021. The potential of feeding patterns to assess generic welfare in growing-finishing pigs. Applied Animal Behaviour Science. 241: 105383.

Cammack, K.M., Leymaster, K.A., Jenkins, T.G. and Nielsen, M.K. 2005. Estimates of genetic parameters for feed intake, feeding behavior, and daily gain in composite ram lambs. Journal of Animal Science. 83(4): 777-785.

Canario, L., Lundeheim, N. and Bijma P. 2017. The early-life environment of a pig shapes the phenotypes of its social partners in adulthood. Heredity. 118: 534-541.

Casey, D.S., Stern, H.S. and Dekkers, J.C.M. 2005. Identification of errors and factors associated with errors in data from electronic swine feeders. Journal of Animal Science. 83: 969–982.

Chen, C.Y., Kachman, S. D., Johnson, R.K., Newman, S. and Van Vleck, L.D. 2008. Estimation of genetic parameters for average daily gain using models with competition effects. Journal of Animal Science. 86 (10): 2525–2530.

Chen, T., He, T., Benesty, M., Khotilovich, V., Tang, Y., Cho, H., Chen, K., Mitchell, R., Cano, I., Zhou, T., Li, M., Xie, J., Lin, M., Geng, Y. and Li, Y., 2018. xgboost: Extreme Gradient Boosting.

Cheverud, J.M. 1984. Evolution by kin selection: a quantitative genetic model illustrated by maternal performance in mice. Evolution. 38: 766–777.

Chu, T.T., Henryon, M., Jensen, J., Ask, B. and Christensen, O.F. 2021. Statistical model and testing designs to increase response to selection with constrained inbreeding in genomic breeding programs for pigs affected by social genetic effects. Genetics Selection Evolution. 53:1.

D'Eath, R.B., Jack, M., Futro, A., Talbot, D., Zhu, Q., Barclay, D. and Baxter, E.M. 2018. Automatic early warning of tail biting in pigs: 3D cameras can detect lowered tail posture before an outbreak. PLoS One. 13: 18.

Dickerson, G.E. 1947. Composition of hog carcasses as influenced by heritable differences in rate and economy of gain. Iowa Agric. Experiment Station Research Bulletin. 354: 492–524.

Dieters, M.J., White, T.L. and Littell, R.C. 1995. Application of approximate variances of variance components and their ratios in genetic tests. Theoretical and Applied Genetics. 91: 15-24.

Ding Rongrong, D., Ming, Y., Xingwang, W., Jianping, Q., Zhanwei, Z., Shenping, Z., Shaoyun, L., Zheng, X., Enqin, Z., Gengyuan, C., Dewu, L., Wen, H., Jie, Y. and Zhenfang, W. 2018. Genetic Architecture of Feeding Behavior and Feed Efficiency in a Duroc Pig Population. Frontiers in Genetics. 9: 220.

Do, D.N., Strathe, A.B., Jensen, J., Mark, T. and Kadarmideen, H.N. 2013. Genetic parameters for different measures of feed efficiency and related traits in boars of three pig breeds. Journal of Animal Science. 91: 4069–4079.

Ellen, E.D., Rodenburg, T.B., Albers, G.A.A., Bolhuis, J.E., Camerlink, I., Duijveste ijn, N. and Bijma, P. 2014. The prospects of selection for social genetic effects to improve welfare and productivity in livestock. Frontiers in Genetics. 5: 377–391.

Falconer, D.S. and Mackay, T.F.C. 1996 Introduction to Quantitative Genetics. Pearson Education. Harlow, UK.

Feddes, J.J. R., Young, B.A. and DeShazer, J.A. 1989. Influence of temperature and light on feeding behaviour of pigs. Applied Animal Behaviour Science. 23(3): 215-222.

Foster, W., Kilpatrick, D. and Heaney, I. 1983. Genetic variation in the efficiency of energy utilization by the fattening pig. Animal Science. 37(3): 387-393.

Garg, V. and Garg, K. 2016. Early stage disease detection platform in cattles. Journal of Emerging Technologies and Innovative Research. 3: 11.

Gertz, M., Große-Butenuth, K., Junge, W., Maassen-Francke, B., Renner C., Sparenberg, H., Krieter, J. 2020. Using the XGBoost algorithm to classify neck and leg activity sensor data using on-farm health recordings for locomotor-associated diseases. Computers and Electronics in Agriculture. 173.

Gilbert, H., Bidanel, J.P., Gruand, J., Cartitez, J.C., Billon, Y., Guillouet, P., Lagant, H., Noblet, J. and Sellier, P. 2007. Genetic parameters for residual feed intake in growing pigs, with emphasis on genetic relationships with carcass and meat quality traits. Journal of Animal Science. 85: 3182-3188.

Greiner, M., Pfeiffer, D. and Smith, R.D. 2000. Principles and practical application of the receiver-operating characteristic analysis for diagnostic tests. Preventive Veterinary Medicine. 45: 23-41.

Griffing, B. 1967. Selection in reference to biological groups. I. Individual and group selection applied to populations of unordered groups. Australian Journal of Biological Sciences. 20: 127–142.

Gonyou, H.W. and Stricklin, W.R. 1998. Effects of floor area allowance and group size on the productivity of growing/finishing pigs. Journal of Animal Science. 76(5): 1326- 1330.

Guarino, M., Jans, P., Costa, A., Aerts, J.M. and Berckmans, D. 2008. Field test of algorithm for automatic cough detection in pig houses. Computers and Electronics in Agriculture. 62: 22–28.

Hadfield, J.D. and Wilson, A.J. 2007. Multilevel selection 3: modeling the effects of interacting individuals as a function of group size. Genetics. 177: 667–8.

Haer, L.C.M. de and Vries, A.G. de. 1993. Effects of genotype and sex on the feed Intake pattern of group housed growing pigs. Livestock Production Science. 36: 223-232.

Hall, A., Hill, W., Bampton, P. and Webb, A. 1999. Genetic and phenotypic parameter estimates for feeding pattern and performance test traits in pigs. Animal Science. 68: 43–48.

Herrera-Cáceres, W., Ragab, M. and Sánchez, J. 2019. Indirect genetic effects on the relationships between production and feeding behaviour traits in growing Duroc pigs. Animal. 14: 233-242.

Hyun, Y., Ellis, M., McKeith, F.K. and Wilson E.R. 1997. Feed intake pattern of group-housed growing-finishing pigs monitored using a computerized feed intake recording system. Journal of Animal Science. 75:1443–51.

Hyun, Y., Ellis, M., Riskowski, G. and Johnson, R.W. 1998. Growth performance of pigs subjected to multiple concurrent environmental stressors. Journal of Animal Science. 76(3): 721-727.

Hyun, Y. and Ellis, M. 2001. Effect of group size and feeder type on growth performance and feeding patterns in growing pigs. Journal of Animal Science. 79: 803-810.

Jungst, S. B., Christian, L.L. and Kuhlers, D.L. 1981. Response to selection for feed efficiency in individually fed Yorkshire boars. Journal of Animal Science. 53: 323−331.

Kalm, E., Felde, A. and von and Röhe, R. 1996. [New traits in station testing of pigs.]. Züchtungskunde. 68: 474-482.

Kavlak, A.T. and Uimari, P. 2019. Estimation of heritability of feeding behaviour traits and their correlation with production traits in Finnish Yorkshire pigs. Journal of Animal Breeding and Genetics. 136: 484-494.

Kavlak, A.T., Strandén, I., Lidauer, M.H. and Uimari, P. 2021. Estimation of social genetic effects on feeding behaviour and production traits in pigs, Animal.15(3): 100168.

Kennedy, B.W., Van der Werf, J.H. and Meuwissen, T.H. 1993. Genetic and statistical properties of residual feed intake. Journal of Animal Science. 71: 3239–3250.

Knap, P.W., van der Steen, H. A.M. and Plastow, G.S. 2001. Developments in pig breeding and the role of research. Livestock Production Science. 72: 43–48.

Labroue, F., Gueblez, R., and Sellier, P. 1997. Genetic parameters of feeding behaviour and performance traits in group-housed Large White and French Landrace growing pigs. Genetic Selection Evolution. 29: 451–468.

Liakos, K.G., Busato, P., Moshou, D., Pearson, S and Bochtis, D. 2018. Machine learning in agriculture: A review. Sensors. 18: 2674.

Lynch, M. and Walsh, B. 1998. Genetics and Analysis of Quantitative Traits. Sinauer Associates, Sunderland, MA.

Madsen, P. and Jensen, J. 2013. A user’s guide to DMU. A package for analyzing multivariate mixed models. Version 6, release 5.2. University of Aarhus, Center for Quantitative Genetics and Genomics, Department of Molecular Biology and Genetics, Research Centre Foulum, Tjele, Denmark.

Maselyne, J., Van Nuffel, A., Briene, P., Vangeyte, J., De Ketelaere, B., Millet, S., Van den Hof, J., Maes, D. and Saeys, W. 2018. Online warning systems for individual fattening pigs based on their feeding pattern. Biosystems Engineering. 173: 143–156.

Matthews, S.G., Miller, A.M., Plötz, T. and Kyriazakis, I. 2017. Automated tracking to measure behavioural changes in pigs for health and welfare monitoring. Scientific Reports. 7: 1-12.

McGlone, J.J. 1985. A quantitative ethogram of aggressive and submissive behaviors in recently regrouped pigs. Journal of Animal Science. 61(3): 559-565.

McGlone, J.J. and Newby, B.E. 1994. Space requirements for finishing pigs in confinement: behavior and performance while group size and space vary. Applied Animal Behaviour Science. 39(3): 331-338.

McPhee, C.P., Brennan, P.J. and Duncalfe, F. 1979. Genetic and phenotypic parameters of Australian Large White and Landrace boars performance-tested when offered food ad libitum. Animal Production. 28(01): 79-85.

McSweeny, J.M., Hermesch, S. and Crump, R. 2003. Genetic analysis of feeding pattern traits in pigs. Animal Genetics and Breeding Unit, a Joint Venture of the University of New England and NSW Agriculture, University of New England, Armidale, NSW, 2351.

Mellor, D.J. 2016. Updating Animal Welfare Thinking: Moving beyond the "Five Freedoms" towards "A Life Worth Living". *Animals*. *6*: 21.

Moore, A.J., Brodie, E.D. and Wolf, J.B. 1997. Interacting phenotypes and the evolutionary process: I. Direct and indirect genetic effects of social interactions. Evolution. 51: 1352–1362.

Mrode, R.A. and B.W. Kennedy. 1993. Genetic-variation in measures of food efficiency in pigs and their genetic-relationships with growth-rate and backfat. Animal Science. 56(2): 225-232.

Muir, W.M. 2003. Indirect selection for improvement of animal well-being, pp. 247–256 in Poultry Genetics, Breeding and Biotechnology, edited by W. M. Muir and S. Aggrey. CABI Press, Cambridge, MA.

Munsterhjelm, C., Heinonen, M. and Valros, A. 2015. Effects of clinical lameness and tail biting lesions on voluntary feed intake in growing pigs. Livestock Science. 181: 210–219.

Nielsen, B.L., Lawrence, A.B. and Whittemore, C.T. 1996. Effect of individual housing on the feeding behaviour of previously group housed growing pigs. Applied Animal Behaviour Science. 47: 149-161,

Nielsen, H.M., Ask, B. and Madsen, P. 2018. Social genetic effects for growth in pigs differ between boars and gilts. Genetics Selection Evolution. 50: 4.

Patterson, H.D. and Thompson, R. 1971. Recovery of inter-block information when block sizes are unequal. Biometrika. 58: 545-554.

Piette, D., Norton, T., Exadaktylos, V., Berckmans, D. 2020. Individualised automated lameness detection in dairy cows and the impact of historical window length on algorithm performance. Animal. 14: 409-417.

Rauw, W.M., Soler, J., Tibau, J., Reixach, J., and Gomez Raya, L. 2006. Feeding time and feeding rate and its relationship with feed intake, feed efficiency, growth rate, and rate of fat deposition in growing Duroc barrows. Journal of Animal Science. 84: 3404–3409.

Remus, A., Hauschild, L., Létourneau-Montminy, M.P., Andretta, I. and Pomar, C. 2020. Feeding behavior of growing and finishing pigs fed different dietary threonine levels in a group-phase feeding and individual precision feeding system. Translational Animal Science. 4(4): 177

Riikimaki, A. 2019. MSc thesis, University of Helsinki, Helsinki, Finland. Syöntikäyttäytymisominaisuuksien periytyminen ja yhteydet tuotanto-ominaisuuksiin maatiaisrotuisella sialla. http://urn.fi/URN:NBN:fi:hulib-201906122715.

Rohrer, G.A., Brown-Brandl, T., Rempel, L.A., Schneider, J.F. and Holl, J. 2013. Genetic analysis of behavior traits in swine production. Livestock Science. 157: 28-37.

Ødegard, J. and Olesen, I. 2011. Comparison of testing designs for genetic evaluation of social effects in aquaculture species. Aquaculture. 317: 74–8.

Schaeffer, L.R. 2004. Application of random regression models in animal breeding. Livestock Production Science. 86: 35-45.

Schulze, V., Roehe, R., Bermejo, J.L., Looft, H. and Kalm, E. 2003. The influence of feeding behaviour on feed intake curve parameters and performance traits of station-tested boars. Livestock Production Science. 82: 105–116.

Sermyagin, A.A., Belous, A.A., Trebunskih, E.A. and Zinovieva, N.A. 2020. Feeding behavior as the new breeding traits in pigs. Sel'skokhozyaistvennaya Biologiya [Agricultural Biology]. 55: 1126-1138.

Saintilan, R., Mérour, I., Brossard, L., Tribout, T., Dourmad, J.Y., Sellier, P., Bidanel, J., van Milgen, J. and Gilbert, H. 2013. Genetics of residual feed intake in growing pigs: Relationships with production traits, and nitrogen and phosphorus excretion traits. Journal of Animal Science. 91(6): 2542-54.

Strandén I. RelaX2 program for pedigree analysis, version 1.65. 2014.

Swets, J.A., 1988. Measuring the accuracy of diagnostic systems. Science. 240:1285–1293.

Thomas, J. Rousselière, Y., Marcon, M. and Hémonic, A. 2021. Early Detection of Diarrhea in Weaned Piglets From Individual Feed, Water and Weighing Data. Frontiers in Animal Science. 2: 2673-6225.

Webb, A. and King, J. 1983. Selection for improved food conversion ratio on ad libitum group feeding in pigs. Animal Science. 37: 375-385.

Wetten, M., Odegard, J., Vangen, O. and Meuwissen, T. 2012. Simultaneous estimation of daily weight and feed intake curves for growing pigs by random regression. Animal. 6, 433–439.

Wilham, R.L. 1963. The covariance between relatives for characteristics composed of components contributed by related individuals. Biometrics. 19: 18–27.

Wolf, J.B., Brodie, E.D., Cheverud, J.M., Moore, A.J. and Wade, M.J. 1998. Evolutionary consequences of indirect genetic effects. Trends in Ecology & Evolution. 13: 64–69.

Wolter, B.F., Ellis, M., Curtis, S.E., Augspurger, N.R., Hamilton, D.N., Parr, E. N. and Webel, D. M. 2001. Effect of group size on pig performance in a wean-to-finish production system. Journal of Animal Science. 79(5): 1067-1073.

Vargas, J.V., Craig, J.V. and Hines, R.H. 1987. Effects of feeding systems on social and feeding behavior and performance of finishing pigs. Journal of Animal Science. 1987. 65(2):463-74.

Von Felde, A., Roehe, R., Looft, H. and Kalm, E. 1996. Genetic association between feed intake and feed intake behaviour at different stages of growth of group-housed boars. Livestock Production Science. 47: 11–12.

Young, J.M., Dekkers, J., Gabler, N., Johnson, A., Nettleton, D., Rothschild, M. and Tuggle, C. 2012. The effect of selection for residual feed intake during the row/finish phase of production on feeding behavior traits and sow reproduction and lactation efficiency in Yorkshire pigs. PhD Dissertation. Iowa State University, Ames.

Received: 19 February 2019 | Revised: 6 May 2019 | Accepted: 7 May 2019
DOI: 10.1111/jbg.12408

ORIGINAL ARTICLE

Journal of Animal Breeding and Genetics WILEY

Estimation of heritability of feeding behaviour traits and their correlation with production traits in Finnish Yorkshire pigs

Alper T. Kavlak | **Pekka Uimari**

Department of Agricultural Sciences, University of Helsinki, Helsinki, Finland

Correspondence
Alper T. Kavlak, Department of Agricultural Sciences, University of Helsinki, Helsinki, Finland.
Email: alper.kavlak@helsinki.fi

Funding information
This study was funded by Raisio Oyj Tutkimussäätiö and Suomen Sianjalostuksen Säätiö.

Abstract
A major proportion of the costs of pork production is related to feed. The feed conversion rate (FCR) or residual feed intake (RFI) is thus commonly included in breeding programmes. Feeding behaviour traits do not directly have economic value but, if correlated with production traits, can be used as auxiliary traits. The aim of this study was to estimate the heritability of feeding behaviour traits and their genetic correlations with production traits in the Finnish Yorkshire pig population. The data were available from 3,235 pigs. Feeding behaviour was measured as the number of visits per day (NVD), time spent in feeding per day (TPD), daily feed intake (DFI), time spent feeding per visit (TPV), feed intake per visit (FPV) and feed intake rate (FR). The test station phase was divided into five periods. Estimates of heritabilities of feeding behaviour traits varied from 0.17 to 0.47. Strong genetic correlations were obtained between behaviour traits in all periods. However, only DFI was strongly correlated with the production traits. Interestingly, a moderate positive genetic correlation was obtained between FR and backfat thickness (0.1–0.5) and between FR and average daily gain (0.3–0.4), depending on the period. Based on the results, there is no additional benefit from including feeding-related traits other than those commonly used (FCR and RFI) in the breeding programme. However, if correlated with animal welfare, the feeding behaviour traits could be valuable in the breeding programme.

KEYWORDS
eating rate, feeding behaviour, feeding rate, genetic correlation, heritability, pigs

1 | INTRODUCTION

The profitability of pork production is dependent mainly on the production of lean meat by the efficient use of feed for growth. Thus, the most important production traits considered in pig breeding programmes are growth (average daily gain, ADG), feed conversion rate (FCR), and residual feed intake (RFI), the difference obtained between feed intake and predicted feed intake based on growth and maintenance (Kennedy, Werf, & Meuwissen, 1993). The genetic improvement of production traits in pigs is commonly based on the performance of group-housed pigs in a controlled test station environment. During the test period, feed intake can be measured automatically. Feeding behaviour can be measured early in life, and if it is correlated with production traits, it can be used as an early selection criterion, while pigs with unfavourable feeding behaviour can be removed from the test and replaced by other test pigs.

Estimates of the heritabilities of feeding behaviour traits range from low to high (e.g., Do, Strathe, Jensen, Mark, & Kadarmideen, 2013; Gilbert et al., 2007; Lu et al., 2017; Von Felde, Roehe, Looft, & Kalm, 1996). A positive genetic correlation was found between feeding behaviour and feed efficiency traits, in which animals that consume more feed per visit tend to grow faster (Labroue, Gueblez, & Sellier, 1997). Thus, genetic improvement in feed efficiency may also be dependent on the genetics of feeding behaviour traits, and therefore, including these traits in breeding programmes has been suggested (Hall, Hill, Bampton, & Webb, 1999; Labroue et al., 1997).

The objective of this study was to estimate the heritability of feeding behaviour traits and their genetic correlations with production traits in the Finnish Yorkshire pig population. In this study, we show that feeding behaviour traits are moderately heritable and highly correlated. We also show that feeding behaviour traits, except daily feed intake, do not have a strong genetic correlation with production traits.

2 | MATERIALS AND METHODS

2.1 | Data

The data were recorded at the central test station of Figen Oy (Pietarsaari, Finland) and were available from the beginning of 2011 until 2016 (October). The pigs arrived at the test station at an average age of 89 ± 10 days (mean ± standard deviation) and an average weight of 32.7 ± 5.4 kg. During the test period (on average 95 ± 3 days), pigs were fed ad libitum, and the feedings were recorded automatically. The feeding system was Schauer Spotmix with Schauer MLP electronic feeders and MLP-manager data management software (Schauer Agrotronic GmbH). The pigs were slaughtered (except those boars that we selected for artificial insemination) at an average age of 186 ± 10 days and an average weight of 117.7 ± 12.1 kg. The pigs were either purebred Finnish Landraces or Yorkshires or their F1-crosses. In this study, only purebred Yorkshire animals were used.

The raw data (28,964,641 observations) included transponder id, date, time of entering the feeder, time leaving the feeder and feed intake per visit. From the raw records, time spent per visit and feeding rate per visit (g/min) were calculated. All observations that did not fulfil the following thresholds were removed from the data (see Casey, Stern, & Dekkers, 2005): the feed intake per visit should be over −20 g and below 2 kg or below 20 g if the time spent per visit was 0; the time spent per visit should be more than 0 s and less than 1 hr if the feed intake per visit was <50 g; the feeding rate per visit should be <500 g/min if the feed intake per visit was more than 50 g; the feeding rate per visit should be <170 g/min if the feed intake was 0 g; the time spent per visit should be <500 s; the feeding rate per visit should be more than 2 g/min; and the time of entering and the time of the leaving the same feeder should not be overlapping. With these criteria, 0.5% of the raw observations were discarded, mainly because the feeding rate per visit was more than 170 g/min (given that the feed intake was over 50 g) or because the feeding rate per visit was <2 g/min.

From the remaining 28,826,029 observations, daily values were calculated as the number of visits per day (NVD, counts), time spent in feeding per day (TPD, min), daily feed intake (DFI, g), time spent feeding per visit (TPV, min), feed intake per visit (FPV, g) and feed intake rate (FR, g/min). The TPV, FPV and FR were average values of the daily records. The final records were calculated as averages of the daily records separately for the five test periods: 0–20 days in the test, 21–40, 41–60, 61–80 and 81–93 days.

The production traits analysed were ADG, FCR, RFI and backfat thickness (BF). FCR was defined as the feed consumption during the test period measured in feed units (one feed unit is 9.3 MJ net energy) divided by the total growth during the test period (finishing body weight − initial body weight). The BF was an average of two Hennessy Grading Systems (type GP4) measurements, one at 8 cm off the midline of the carcass behind the last rib and one at 6 cm off the midline between the third and fourth ribs. The RFI was computed as the difference between the observed and predicted DFIs; that is, the RFI was a residual term from the linear model:

$$\mathrm{ADFI}_{ijk} = \mathrm{sex}_i + \mathrm{hys}_j + b_1\,(\mathrm{IBW})_{ijk} + b_2\,(\mathrm{ADG})_{ijk} + b_3\,(\mathrm{BF})_{ijk} + e_{ijk}$$

where ADFI_{ijk} is the average daily feed intake over five periods, sex_i is the sex effect (boar, gilt, castrate), hys_j is the herd*year*season interaction (four seasons were defined as follows: January–March, April–June, July–September, October–December), b_1, b_2 and b_3 are partial regression coefficients of the initial weight at the beginning of the test period (IBW), ADG and BF, respectively.

The final data included records of 3,235 Yorkshire pigs (2,335 boars, 484 gilts and 416 castrates). All animals had observations for all the studied traits.

2.2 | Statistical analysis

The traits were analysed, using the following animal model:

$$\mathbf{y} = \mathbf{Xb} + \mathbf{Z_a a} + \mathbf{Z_l l} + \mathbf{Z_{bp} bp} + \mathbf{e},$$

where **y** is a vector of observation (feeding behaviour traits, ADG, BF, FCR and RFI), **b** a vector of fixed effects (sex_i and hys_j), **X** a incidence matrix relating records to fixed effects, **a** a vector of random additive genetic effects, **l** a vector of random litter effects, **bp** a vector of random batch*pen effects and **e** a vector of random residuals; the corresponding incidence matrices are $\mathbf{Z_a}$, $\mathbf{Z_l}$ and $\mathbf{Z_{bp}}$, respectively. Pigs were from 684 different batch*pen and 174 herd*year*season combinations. The number of observations in these batch*pens varied from 1 to 12, and from 5 to 62 in herd*year*seasons. Since RFI was already corrected for sex_i and hys_j effects, these were not included in the linear model of RFI.

A univariate model was used for the estimation of heritability and a bi-variate model for genetic correlations. For the bi-variate model, the (co)variance matrix of the normally distributed additive genetic effects was $\mathbf{A}\otimes\mathbf{G}$, where **A** is the numerator relationship matrix, **G** is the genetic (co)variance matrices of the traits, and ⊗ denotes the Kronecker product. The (co)variance matrix of the normally distributed litter, batch*pen and residual effects was $\mathbf{I}\otimes\mathbf{B}$, $\mathbf{I}\otimes\mathbf{C}$ and $\mathbf{I}\otimes\mathbf{R}$, respectively, where **I** is an identity matrix and **B**, **C** and **R** are the (co)variance matrices for litter, batch*pen and residual effects, respectively. The variance and covariance components were estimated by the restricted maximum-likelihood (REML) method (Patterson & Thompson, 1971) using the DMU software (Madsen & Jensen, 2013). The pedigree data (5,396 animals) included all animals with observations and their ancestors down to four generations. The average number of offspring with observations per each sire was 16.

3 | RESULTS

3.1 | Phenotypic description of the traits

The distributions of most of the traits were right-skewed (Figure 1). Prior to statistical analysis, extreme outliers (4 *SD* of the mean) were removed from the data. However, the data also included short visits at the feeder; thus, the minimum TPV was only 32 s and the minimum FPV 13 g. Generally, older pigs visited more often at the feed station than younger animals. In contrast, the TPV decreased radically from 2.5 min (TPV1) to 1.6 min (TPV5). The total TPD increased up to 66 min/day (TPD2) and then decreased to 53 min/day (TPD5). The FPV increased from 57 g per visit (FPV1) to 89 g per visit (FPV5) and DFI from 1,443 g (DFI1) to 2,990 g (DFI5) from period 1 to period 5 (Figure 1). The ADG varied in this data from 588 to 1,268 g/day, with mean ADG of 925 ± 108 g/day. The average BF was 9.9 ± 2.1 mm, average FCR 2.5 ± 0.2 and RFI 0 ± 135 g.

3.2 | Estimated heritabilities

The heritability and corresponding variance components for feeding behaviour traits are given in Table 1. The highest heritability estimate was obtained for TPV2 (0.47 ± 0.07) and the lowest for DFI1 (0.17 ± 0.05). In general, there was more variation in heritability between traits in period 1 than in later periods. In period 5, all the estimates were near 0.3; thus, the heritability of the feeding behaviour

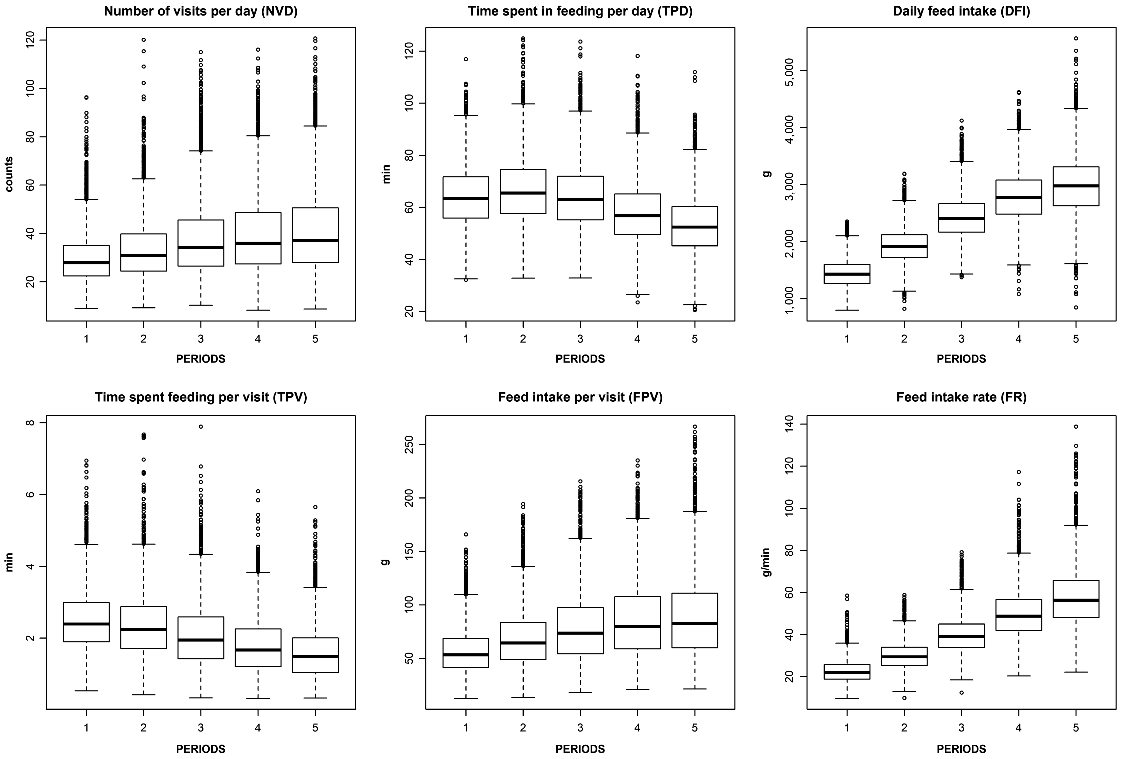


FIGURE 1 Box-plots of the feeding behaviour traits over the five periods

TABLE 1 Heritability (h^2) and variances of additive genetic (σ_a^2), litter (σ_1^2), batch*pen (σ_{bp}^2) and residual (σ_e^2) effects of feeding behaviour and production traits

Trait	h^2	*SE* (h^2)	σ_a^2	σ_1^2	σ_{bp}^2	σ_e^2
Period 1						
NVD1	0.36	0.06	44.4	12.3	20.0	41.8
TPD1	0.37	0.06	51.1	17.1	2.9	62.7
DFI1	0.17	0.05	9,572	9,814	10,691	22,881
TPV1	0.38	0.07	0.28	0.07	0.11	0.26
FPV1	0.44	0.07	198.3	44.7	56.6	142.7
FR1	0.19	0.06	5.6	2.8	9.9	11.0
Period 2						
NVD2	0.41	0.06	72.1	14.9	23.4	62.0
TPD2	0.36	0.06	58.3	17.1	9.6	76.4
DFI2	0.26	0.06	22,405	11,264	14,411	36,879
TPV2	0.47	0.07	0.41	0.04	0.12	0.28
FPV2	0.43	0.07	324.3	56.9	114.7	244.0
FR2	0.21	0.06	9.4	3.0	9.5	21.6
Period 3						
NVD3	0.37	0.06	95.6	17.6	41.7	99.6
TPD3	0.34	0.06	51.5	13.9	10.3	74.4
DFI3	0.30	0.06	36,238	10,854	17,839	53,845
TPV3	0.36	0.06	0.28	0.04	0.13	0.32
FPV3	0.34	0.06	368.3	95.3	186.0	423.1
FR3	0.26	0.06	22.9	5.4	12.6	39.9
Period 4						
NVD4	0.37	0.06	105.0	13.1	44.6	116.7
TPD4	0.36	0.06	47.9	8.4	7.7	64.9
DFI4	0.32	0.06	54,448	14,064	22,324	74,246
TPV4	0.37	0.06	0.23	0.03	0.09	0.26
FPV4	0.36	0.06	482.9	97.9	201.1	528.9
FR4	0.28	0.06	39.3	8.5	20.8	68.9
Period 5						
NVD5	0.31	0.06	99.4	17.7	47.9	147.0
TPD5	0.28	0.06	36.8	7.9	11.3	68.9
DFI5	0.29	0.06	71,959	16,457	25,757	125,164
TPV5	0.34	0.06	0.19	0.03	0.08	0.24
FPV5	0.32	0.06	493.7	116.8	218.6	694.9
FR5	0.29	0.06	61.1	11.1	27.5	110.6
Production traits						
ADG	0.25	0.06	2,661	1,576	697	5,439
BF	0.57	0.07	1.93	0.22	0.12	1.06
FCR	0.28	0.06	0.010	0.002	0.007	0.017
RFI	0.32	0.06	6,749	1,359	3,320	9,104

Abbreviations: ADG, average daily gain; BF, backfat thickness; DFI, daily feed intake; FCR, feed conversion rate; FPV, feed intake per visit; FR, feed intake rate; NVD, number of visits per day; RFI, residual feed intake;TPD, time spent in feeding per day; TPV, time spent feeding per visit.

traits was moderate. Overall, the lowest heritabilities were obtained for FR (from 0.19 to 0.29) than for other feeding behaviour traits. The heritability estimates of ADG, FCR and RFI were moderate (0.25 ± 0.06, 0.28 ± 0.06 and 0.32 ± 0.06, respectively) and high for BF (0.57 ± 0.07; Table 1).

3.3 | Correlations between the traits

As expected, the genetic correlations between the same behaviour trait over the various periods (1–5) were generally high (Figure 2). The strongest genetic correlations were obtained between adjacent periods varying from 0.96 ± 0.01 to 0.99 ± 0.01 for all behaviour traits and the weakest between periods 1 and 5, varying from 0.71 ± 0.09 (FR1 vs. FR5) to 0.83 ± 0.06 (NVD1 vs. NVD5). The phenotypic correlations within the traits were also very high throughout the periods. The strongest phenotypic correlation was obtained for FR (0.88) between periods 4 and 5, while the weakest (0.25) was obtained for DFI between periods 1 and 5.

The phenotypic and genetic correlations between the feeding behaviour traits within the same period are given in Table 2. The genetic correlations between NVD and both TPV and FPV were very high, varying from −0.79 ± 0.05 to −0.97 ± 0.01. The phenotypic correlations between these traits were also strong, varying from −0.76 to −0.83. In contrast, NVD did not correlate with the other traits. Thus, the frequency of feeding did not affect the DFI or FR. FR and TPD had a strong negative correlation; animals that show high FR have shorter TPD than animals with slow FR. The FR also had a moderately positive genetic correlation between the DFI varying from 0.22 ± 0.14 (period 4) to 0.4 ± 0.12 (period 5). Thus, animals with genetic background of high FR also tend to have a genetic background of higher DFI.

Most of the correlations between feeding behaviour traits and production (ADG, BF, FCR and RFI) traits did not differ from zero (Table 3). Only DFI had strong positive genetic correlations with all the production traits (Table 3 and Figure 3), a favourable correlation with ADG, but unfavourable correlations with BF, FCR and RFI. In addition, the FPV had a moderately positive (unfavourable) correlation with BF; animals that consume large quantities of feed per visit tend to gain more BF. The genetic correlation between FPV and BF increased from periods 1 to 5. BF was also correlated with FR; animals with high FR tend to gain more BF as well. This correlation became smaller from periods 1 to 5. In addition, the genetic correlation between TPD and RFI varied from 0.34 ± 0.13 (period 2) to 0.45 ± 0.13 (period 5); thus, animals with more TPD also show increased RFI.

All correlations between the production traits were significant. The genetic correlation between FCR and BF was 0.79 ± 0.07, but only 0.37 ± 0.16 between FCR and ADG (Table 3). In contrast, the genetic correlations between RFI and BF were only 0.38 ± 0.12 while between RFI and ADG 0.63 ± 0.13. Thus, selection for RFI more strongly affects ADG than BF, while selection for FCR more strongly affects BF than ADG.

4 | DISCUSSION

4.1 | Feeding behaviour and production traits

In this study, the performance of pigs at the test station was divided into five 20-day periods, starting 3 days after arrival at the test station. In the previous study of Von Felde et al. (1996), the feed intake was measured during five

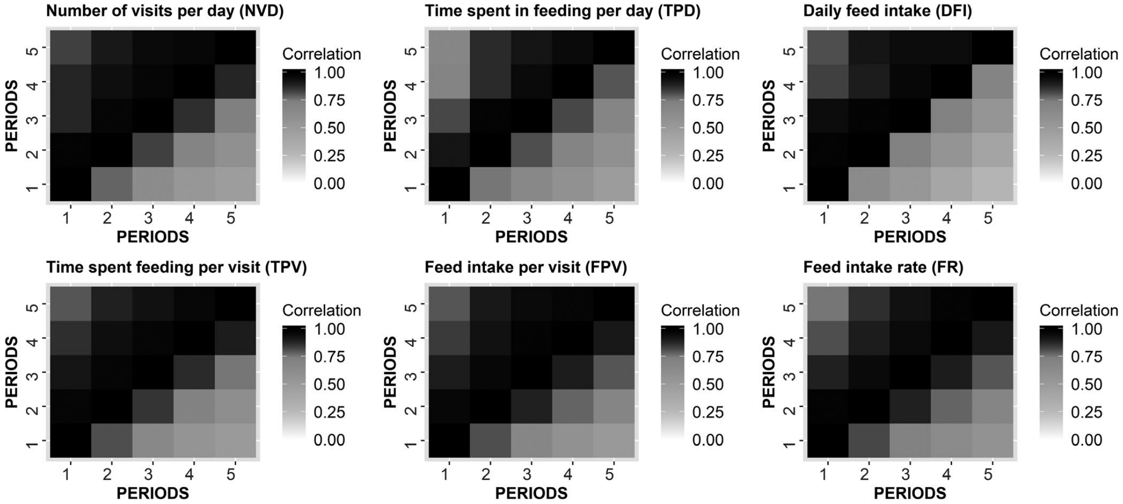


FIGURE 2 Phenotypic (below diagonal) and genetic correlations (above diagonal) between the same behaviour traits over the five periods

TABLE 2 Phenotypic (below diagonal) and genetic correlation (above diagonal) between feeding behaviour traits within the same period

Trait	NVD1	TPD1	DFI1	TPV1	FPV1	FR1
Period 1						
NVD1		0.50 (0.11)	0.18 (0.17)	−0.79 (0.05)	−0.91 (0.03)	−0.06 (0.15)
TPD1	0.20		0.17 (0.17)	0.09 (0.13)	−0.36 (0.12)	−0.74 (0.08)
DFI1	0.10	0.22		0.04 (0.17)	0.25 (0.16)	0.40 (0.16)
TPV1	−0.76	0.28	0.03		0.83 (0.04)	−0.31 (0.14)
FPV1	−0.77	−0.11	0.35	0.80		0.24 (0.14)
FR1	0.10	−0.52	0.51	−0.31	0.17	
Period 2	NVD2	TPD2	DFI2	TPV2	FPV2	FR2
NVD2		0.10 (0.13)	0.01 (0.14)	−0.83 (0.04)	−0.93 (0.02)	0.17 (0.14)
TPD2	0.05		0.27 (0.14)	0.45 (0.10)	0.08 (0.13)	−0.75 (0.07)
DFI2	0.05	0.26		0.17 (0.13)	0.36 (0.12)	0.29 (0.15)
TPV2	−0.77	0.39	0.07		0.89 (0.02)	−0.56 (0.10)
FPV2	−0.80	0.01	0.33	0.84		−0.09 (0.15)
FR2	0.20	−0.59	0.42	−0.42	0.01	
Period 3	NVD3	TPD3	DFI3	TPV3	FPV3	FR3
NVD3		−0.01 (0.13)	0.03 (0.14)	−0.86 (0.03)	−0.97 (0.01)	0.28 (0.13)
TPD3	0.01		0.21 (0.14)	0.50 (0.10)	0.11 (0.14)	−0.77 (0.06)
DFI3	0.05	0.24		0.11 (0.14)	0.32 (0.13)	0.29 (0.14)
TPV3	−0.78	0.40	0.05		0.87 (0.03)	−0.60 (0.10)
FPV3	−0.82	0.04	0.27	0.86		−0.16 (0.15)
FR3	0.27	−0.64	0.37	−0.49	−0.11	
Period 4	NVD4	TPD4	DFI4	TPV4	FPV4	FR4
NVD4		−0.07 (0.13)	0.01 (0.13)	−0.88 (0.03)	−0.94 (0.02)	0.32 (0.12)
TPD4	−0.01		0.28 (0.13)	0.51 (0.10)	0.17 (0.13)	−0.80 (0.06)
DFI4	0.11	0.29		0.15 (0.13)	0.34 (0.12)	0.22 (0.14)
TPV4	−0.80	0.40	0.03		0.89 (0.02)	−0.60 (0.09)
FPV4	−0.83	0.08	0.22	0.88		−0.20 (0.14)
FR4	0.34	−0.63	0.37	−0.53	−0.18	
Period 5	NVD5	TPD5	DFI5	TPV5	FPV5	FR5
NVD5		−0.03 (0.14)	0.06 (0.14)	−0.83 (0.04)	−0.90 (0.03)	0.35 (0.12)
TPD5	0.00		0.31 (0.13)	0.57 (0.10)	0.26 (0.13)	−0.66 (0.08)
DFI5	0.18	0.41		0.17 (0.14)	0.34 (0.13)	0.40 (0.12)
TPV5	−0.78	0.39	0.02		0.90 (0.02)	−0.63 (0.09)
FPV5	−0.81	0.12	0.18	0.90		−0.24 (0.14)
FR5	0.36	−0.57	0.35	−0.52	−0.21	

Note: Standard errors are given in brackets.

Abbreviations: DFI, daily feed intake; FPV, feed intake per visit; FR, feed intake rate; NVD, number of visits per day; TPD, time spent in feeding per day; TPV, time spent feeding per visit.

periods (every second week) over 10 weeks. Rauw, Soler, Tibau, Reixach, and Gomez Raya (2006) and Schulze, Roehe, Bermejo, Looft, and Kalm (2003) also used five periods, while Young, Cai, Nettleton, and Dekkers (2009) measured feeding behaviour traits over three periods: the entire test (from approximately 3–8 months of age), the first half of the test and the second half of the test. Chen, Misztal, Tsuruta, Zumbach, et al. (2010) also used three periods (85–106, 107–128 and 129–150 days of age). The data can also be analysed, using a random regression model as applied to body weight and feed intake by Wetten, Ødegård, Vangen, and Meuwissen (2012) and Coyne et al. (2017). However, based on our results and those of other authors, division of the test period into five parts enabled

TABLE 3 Phenotypic (r_p) and genetic correlation (r_g) between production and feeding behaviour traits

	ADG		BF		FCR		RFI	
Traits	r_p	r_g (*SE*)	r_p	r_g (*SE*)	r_p	r_g (*SE*)	r_p	r_g (*SE*)
Period 1								
NVD1	−0.02	−0.01 (0.15)	−0.08	−0.05 (0.13)	−0.02	0.18 (0.14)	0.13	0.13 (0.15)
TPD1	0.09	−0.04 (0.15)	0.01	−0.05 (0.13)	0.03	0.34 (0.13)	0.15	0.42 (0.13)
DFI1	0.45	0.67 (0.12)	0.33	0.65 (0.11)	0.34	0.88 (0.08)	0.30	0.89 (0.10)
TPV1	0.07	0.00 (0.15)	0.11	0.14 (0.13)	−0.02	0.10 (0.14)	−0.04	0.15 (0.14)
FPV1	0.21	0.17 (0.15)	0.23	0.32 (0.11)	0.11	0.21 (0.13)	0.01	0.18 (0.14)
FR1	0.18	0.38 (0.15)	0.15	0.51 (0.12)	0.25	0.31 (0.16)	0.14	0.12 (0.16)
Period 2								
NVD2	−0.02	−0.03 (0.14)	−0.08	−0.07 (0.12)	0.08	0.19 (0.13)	0.14	0.17 (0.13)
TPD2	0.12	0.23 (0.15)	0.10	0.17 (0.12)	0.04	0.18 (0.14)	0.14	0.34 (0.13)
DFI2	0.60	0.88 (0.05)	0.46	0.72 (0.07)	0.35	0.76 (0.09)	0.36	0.85 (0.08)
TPV2	0.08	0.10 (0.14)	0.13	0.22 (0.12)	−0.03	−0.01 (0.14)	−0.04	−0.00 (0.14)
FPV2	0.25	0.29 (0.13)	0.27	0.43 (0.10)	0.07	0.08 (0.14)	0.01	0.04 (0.14)
FR2	0.24	0.30 (0.16)	0.15	0.25 (0.14)	0.20	0.29 (0.16)	0.16	0.19 (0.16)
Period 3								
NVD3	−0.05	−0.04 (0.15)	−0.11	−0.18 (0.12)	0.09	0.07 (0.14)	0.17	0.11 (0.14)
TPD3	0.14	0.19 (0.15)	0.12	0.09 (0.13)	0.02	0.22 (0.14)	0.13	0.40 (0.13)
DFI3	0.68	0.91 (0.04)	0.51	0.64 (0.08)	0.34	0.74 (0.09)	0.44	0.88 (0.07)
TPV3	0.10	0.12 (0.15)	0.15	0.25 (0.12)	−0.05	0.03 (0.14)	−0.07	0.04 (0.14)
FPV3	0.27	0.32 (0.14)	0.28	0.47 (0.11)	0.03	0.17 (0.14)	−0.01	0.10 (0.14)
FR3	0.24	0.29 (0.15)	0.13	0.24 (0.13)	0.19	0.12 (0.16)	0.19	0.05 (0.16)
Period 4								
NVD4	−0.04	−0.10 (0.15)	−0.10	−0.22 (0.12)	0.09	0.03 (0.14)	0.17	0.09 (0.14)
TPD4	0.14	0.23 (0.15)	0.15	0.17 (0.12)	0.01	0.30 (0.13)	0.10	0.36 (0.13)
DFI4	0.71	0.88 (0.04)	0.52	0.60 (0.08)	0.30	0.75 (0.09)	0.46	0.85 (0.07)
TPV4	0.11	0.26 (0.14)	0.14	0.29 (0.12)	−0.06	0.07 (0.14)	−0.08	0.04 (0.14)
FPV4	0.28	0.22 (0.15)	0.27	0.44 (0.11)	0.02	0.17 (0.14)	−0.01	0.11 (0.14)
FR4	0.25	0.31 (0.15)	0.12	0.11 (0.13)	0.19	0.04 (0.15)	0.24	0.05 (0.15)
Period 5								
NVD5	−0.04	−0.08 (0.15)	−0.07	−0.20 (0.13)	0.08	0.06 (0.15)	0.17	0.08 (0.14)
TPD5	0.14	0.31 (0.15)	0.16	0.22 (0.13)	0.03	0.36 (0.14)	0.17	0.45 (0.13)
DFI5	0.61	0.86 (0.05)	0.50	0.60 (0.08)	0.26	0.73 (0.10)	0.49	0.83 (0.07)
TPV5	0.11	0.33 (0.14)	0.15	0.31 (0.12)	−0.03	0.10 (0.14)	−0.05	0.11 (0.14)
FPV5	0.27	0.46 (0.13)	0.27	0.46 (0.11)	0.02	0.20 (0.14)	0.02	0.11 (0.15)
FR5	0.25	0.30 (0.14)	0.13	0.15 (0.13)	0.17	0.05 (0.15)	0.24	0.11 (0.15)
ADG			0.49	0.37 (0.12)	−0.21	0.37 (0.16)	0.02	0.63 (0.13)
BF					0.24	0.79 (0.07)	0.02	0.38 (0.12)
FCR							0.79	0.79 (0.06)

Note: Standard errors (*SE*) are given in brackets.

Abbreviations: ADG, average daily gain; BF, backfat thickness; DFI, daily feed intake; FCR, feed conversion rate; FPV, feed intake per visit; FR, feed intake rate; NVD, number of visits per day; RFI, residual feed intake; TPD, time spent in feeding per day; TPV, time spent feeding per visit.

us to demonstrate possible differences in heritabilities and genetic covariances between various growth phases of fattening pigs.

The mean TPD (61 min), DFI (2.2 kg) and FR (40 g/min) over five periods were similar to those reported by Do et al. (2013) (62 min, 2.1 kg, and 37 g/min) in Yorkshire

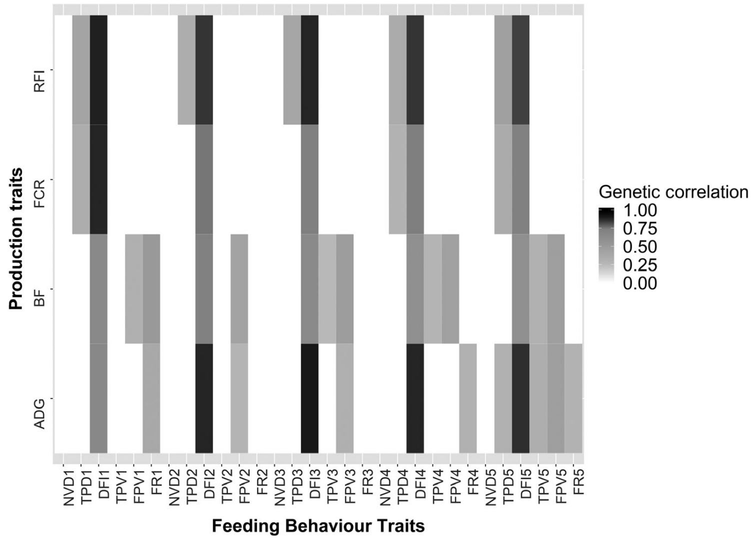


FIGURE 3 Genetic correlations between the feeding behaviour traits and the production traits. Only statistically significant (*p*-value <0.05) correlations are shown

and slightly larger than those reported by Labroue, Gueblez, Meunier-Salaun, and Sellier (1999) also in Yorkshire (50 min, 1.7 kg and 35 g/min). In addition, the tendency for feeding behaviour traits to vary over time in our study was similar to that reported by other authors. Reyer et al. (2017) reported that Maxgro line (Hermitage Genetics, Ireland) pigs tended to visit the feeder more often, but spent less time there as they aged, while older pigs showed higher FRs than younger pigs.

The average production performances (ADG, BF and FCR) of the test pigs in this study from 2011 to 2016 were 925 g, 9.9 mm and 2.5, respectively. Similar test station performances have also been reported in other recent studies in different breeds, for example, Jiao, Maltecca, Gray, and Cassady (2014), Do et al. (2013), Bahelka, Tomka, Bucko, and Hanusova (2015) and Godinho et al. (2018). Thus, Finnish Yorkshires have similar growth performance than other recently studied pig breeds. The RFI was calculated as the difference between the observed and predicted DFI based on IBW, ADG and BF (also corrected for sex and herd*year*season effects). The RFI varied from −505 to 878 g, with *SD* of 135 g. Similar RFI SDs were also obtained by Cai, Casey, and Dekkers (2008) and Dekkers and Gilbert (2010) in Yorkshire.

4.2 | Estimates of heritability

The heritability of the feeding behaviour traits varied over time (from period 1 to 5), but the genetic correlations between the same trait over time were high for all traits (Figure 2). Thus, the genetic basis for feeding behaviour at the beginning of the test was similar to that at the end. The estimated heritabilities were slightly lower in the first period than in later periods, especially for DFI. The young pigs at the beginning of the test may have been more vulnerable to environmental factors than older pigs. Our statistical model with sex, herd*year*season, litter and batch*pen effects may not have been sensitive enough to record all the environmental factors influencing the feeding behaviour of young pigs. Factors such as adaptation to new pen mates and feeding systems may have influenced the estimates.

In addition, the estimated heritabilities between traits varied more at the beginning of the test (period 1) than at the end (period 5). The heritabilities of all the feeding behaviour traits in period 5 converged to an approximate value of 0.3. Similar heritabilities have been obtained in other studies. For example, Hall et al. (1999) obtained heritabilities varying from 0.27 (FPV) to 0.34 (NVD) in Yorkshire, Schulze et al. (2003) from 0.34 (NVD) to 0.46 (TPD) in a combined data of Yorkshire- and Landrace-based dam lines, and Chen, Misztal, Tsuruta, Herring, et al. (2010) from 0.18 (TPD) to 0.42 (FR) in Duroc. Our estimates for DFI were higher than those obtained by Coyne et al. (2017) (0.07 to 0.25), in which a random regression model and combined Finnish Landrace and Yorkshire data were used. This could have been due to the difference in statistical approaches (logistic regression vs. periodic approach) and data (all breed data vs. single breed data). However, our estimates for DFI were similar to those presented by Hall et al. (1999), in which the heritabilities ranged from 0.18 to 0.26 over the four test periods. The estimates of the heritability of production traits (ADG, BF and FCR) were moderate and similar to those obtained in other studies (see Clutter, 2011).

4.3 | Genetic correlations

Strong negative genetic correlations were obtained between NVD and FPV (over −0.9 in all periods) and between NVD and TPV (−0.79 to −0.88). Also, a strong positive correlation was obtained between FPV and TPV (0.83–0.90). As discussed above, these strong correlations were reasonable. To obtain sufficient energy and nutrients for maintenance and growth, the pigs either consume feed often but in small portions or more seldom and in large portions. In addition, FR had a moderately genetic correlation between DFI (0.22 to 0.40) in all periods; thus, faster eaters (g/min) tended to have higher genetic potential for DFI. A strong genetic correlation was also obtained with FR and TPD especially early in the test period. Do et al. (2013) also reported highly negative genetic correlations between NVD and FPV (−0.95) and between NVD and TPV (−0.91) in Danish Yorkshire populations, as well as a significant positive correlation between DFI and FR (0.36). Schulze et al. (2003) also reported highly negative genetic correlations between NVD and FPV (−0.92) and between NVD and TPV (−0.81), a highly positive correlation between FPV and TPV (0.86) and a moderately positive correlation between DFI and FR (0.20) in a combined data of Yorkshire- and Landrace-based dam lines.

Among the feeding traits examined, the DFI showed the highest genetic correlation between all production traits. These correlations were exceptionally high in the first period (0.65–0.89) and slightly lower in the last period (0.60–0.86). The genetic correlation between DFI and ADG increased from period 1 (0.67) to period 5 (0.86), while the remaining genetic correlations (between DFI and BF, FCR and RFI) decreased over time. A positive genetic correlation between DFI and ADG was also reported by Jiao et al. (2014) (0.32), Chen, Misztal, Tsuruta, Herring, et al. (2010) (0.46), Do et al. (2013) (0.84), Cai et al. (2008) (0.88) and between DFI and BF by Jiao et al. (2014) (0.36), Do et al. (2013) (0.68) and Cai et al. (2008) (0.57). Given that a high genetic correlation was obtained only between DFI and production traits, there is no reason to include feeding behaviour traits in the breeding programme. However, if there exist correlation between feeding behaviour traits and animal welfare-related traits such as tail biting, as indicated by Wallenbeck and Keeling (2013), then using feeding behaviour traits as auxiliary traits to improve animal welfare should be considered in breeding programmes.

One of the most interesting genetic correlations was that between FR1 and BF (0.51). Thus, animals with the genetic potential for fast FR early in the test (period 1) also gain more fat. Later in the test, the genetic correlation between FR and BF decreased to 0.15 (period 5). In contrast to FR, the genetic correlation between FPV and BF strengthened from periods 1 (0.32) to 5 (0.46). A positive genetic correlation between FR and BF was found by Schulze et al. (2003) (0.16) and Do et al. (2013) (0.26), but none between FPV and BF (Schulze et al., 2003) (0.09), while a weaker correlation (0.25) was found by Do et al. (2013).

Interestingly in humans, a positive association was obtained between eating speed and overweight (Lee et al., 2016; Tanihara et al., 2011) and eating speed and metabolic syndrome among other health problems (Nohara et al., 2015; Tajima et al., 2014; Tao et al., 2018; Zhu, Haruyama, Muto, & Yamazaki, 2015). Fast eaters tend to be more obese, have higher blood pressure and be more susceptible to metabolic diseases than slow eaters. The estimated odds ratio between normal and fast-eating individuals for overweight in the Japanese adult population was 1.9 (Lee et al., 2016) and between slow and fast-eating individuals for metabolic syndrome in the Chinese adult population 2.3 (Tao et al., 2018). Given the similar metabolic system in pigs and humans, pigs can be used as an animal model for further investigation of the mechanism behind the unfavourable association between eating (or feeding) speed and accumulation of body fat.

5 | CONCLUSIONS

In this study, the heritability of feeding behaviour traits and their correlation with production traits were investigated in the Finnish Yorkshire pig population. The results indicated moderate heritability for all studied feeding behaviour traits. High genetic correlations were obtained only between DFI and production traits. The most interesting correlation was obtained between FR (and FPV) and BF. High breeding values for FR (g/min) early in the test or large portion sizes late in the test indicate a genetic potential to gain BF. However, the potential utilization of this correlation is limited in pig breeding programmes even though selection against FR does not seem to affect other production traits. BF is highly heritable; thus, direct selection of BF based on information from relatives or ultrasound measurement of the animal itself is more effective than indirect selection based on early FR. In the selection of sow replacement at the farm level, favouring piglets that consume feed rapidly may in turn favour sows that have greater fat reservoirs for farrowing and feeding the litter. Finally, since a positive correlation between eating speed and overweight has been demonstrated in human populations, studies in pigs may aid in revealing the genetic basis of this unfavourable association.

ACKNOWLEDGEMENTS

The authors thank Raisio Oyj Tutkimussäätiö and Suomen Sianjalostuksen Säätiö for funding this research. We are grateful to Figen Oy for providing the data and M.Sc. Anna Riihimäki for help in editing the data.

DATA ACCESSIBILITY

The data that support the findings of this study are available from Figen Oy, Finland. Restrictions may apply to the availability of these data, which were used under license for this study.

ORCID

Alper T. Kavlak https://orcid.org/0000-0001-6829-4242

REFERENCES

Bahelka, I., Tomka, J., Bucko, O., & Hanusova, E. (2015). Growth performance and carcass quality of entire males, surgical castrates and gilts. *Slovak Journal of Animal Science*, *48*, 116–121.

Cai, W., Casey, D. S., & Dekkers, J. C. M. (2008). Selection response and genetic parameters for residual feed intake in Yorkshire swine. *Journal of Animal Science*, *86*, 287–298.

Casey, D. S., Stern, H. S., & Dekkers, J. C. M. (2005). Identification of errors and factors associated with errors in data from electronic swine feeders. *Journal of Animal Science*, *83*, 969–982.

Chen, C. Y., Misztal, I., Tsuruta, S., Herring, W. O., Holl, J., & Culbertson, M. (2010). Influence of heritable social status on daily gain and feeding pattern in pigs. *Journal of Animal Breeding and Genetics*, *127*, 107–112. https://doi.org/10.1111/j.1439-0388.2009.00828.x

Chen, C. Y., Misztal, I., Tsuruta, S., Zumbach, B., Herring, W. O., Holl, J., & Culbertson, M. (2010). Estimation of genetic parameters of feed intake and daily gain in Durocs using data from electronic swine feeders. *Journal of Animal Breeding and Genetics*, *127*, 230–234. https://doi.org/10.1111/j.1439-0388.2009.00833.x

Clutter, A. C. (2011). Genetics of performance traits. In M. F. Rotschild, & A. Ruvinsky (Eds.), *The genetics of the pig* (2nd ed., pp. 463–510). New York, NY: CAB International.

Coyne, J. M., Donagh, P. B., Matilainen, K., Sevon-Aimonen, M.-L., Mäntysaari, E. A., Juga, J. K., … McHugh, N. (2017). Genetic covariance functions for live weight, feed intake, and efficiency measures in growing pigs. *Journal of Animal Science*, *95*, 0021–8812. https://doi.org/10.2527/jas2017.1408

Dekkers, J. C. M., & Gilbert, H. (2010). *Genetic and biological aspect of residual feed intake in pigs*. Proceedings of the 9th world congress on genetics applied to livestockproduction: 1–6 August 2010; Leipzig.

Do, D. N., Strathe, A. B., Jensen, J., Mark, T., & Kadarmideen, H. N. (2013). Genetic parameters for different measures of feed efficiency and related traits in boars of three pig breeds. *Journal of Animal Science*, *91*, 4069–4079.

Gilbert, H., Bidanel, J. P., Gruand, J., Cartitez, J. C., Billon, Y., Guillouet, P., … Sellier, P. (2007). Genetic parameters for residual feed intake in growing pigs, with emphasis on genetic relationships with carcass and meat quality traits. *Journal of Animal Science*, *85*, 3182–3188.

Godinho, R. M., Bastiaansen, J. W. M., Sevillano, C. A., Silva, F. F., Guimarães, S. E. F., & Bergsma, R. (2018). Genotype by feed interaction for feed efficiency and growth performance traits in pigs. *Journal of Animal Science*, *96*, 4125–4135.

Hall, A., Hill, W., Bampton, P., & Webb, A. (1999). Genetic and phenotypic parameter estimates for feeding pattern and performance test traits in pigs. *Animal Science*, *68*, 43–48. https://doi.org/10.1017/S1357729800050062

Jiao, S., Maltecca, C., Gray, K. A., & Cassady, J. P. (2014). Feed intake, average daily gain, feed efficiency, and real-time ultrasound traits in Duroc pigs: I. Genetic parameter estimation and accuracy of genomic prediction. *Journal of Animal Science*, *92*, 2377–2386.

Kennedy, B. W., Van der Werf, J. H., & Meuwissen, T. H. (1993). Genetic and statistical properties of residual feed intake. *Journal of Animal Science*, *71*, 3239–3250. https://doi.org/10.2527/1993.71123239x

Labroue, F., Gueblez, R., Meunier-Salaun, M. C., & Sellier, P. (1999). Feed intake behaviour of group-housed Pietrain and Large White growing pigs. *Annales de Zootechnie*, *48*, 247–261.

Labroue, F., Gueblez, R., & Sellier, P. (1997). Genetic parameters of feeding behaviour andperformance traits in group-housed Large White and French Landrace growing pigs. *Genetic Selection Evolution*, *29*, 451–468. https://doi.org/10.1186/1297-9686-29-4-451

Lee, J. S., Mishra, G., Hayashi, K., Watanabe, E., Mori, K., & Kawakubo, K. (2016). Combined eatingbehaviors and overweight: Eating quickly, late evening meals, and skippingbreakfast. *Eating Behaviors*, *21*, 84–88. https://doi.org/10.1016/j.eatbeh.2016.01.009

Lu, D., Jiao, S., Tiezzi, F., Knauer, M., Huang, Y., Gray, K. A., & Maltecca, C. (2017). Therelationship between different measures of feed efficiency and feeding behavior traits in Duroc pigs. *Journal of Animal Science*, *95*, 3370–3380.

Madsen, P., & Jensen, J. (2013). *A user's guide to DMU*. A package for analyzing multivariate mixed models. Version 6, release 5.2. University of Aarhus, Center for Quantitative Genetics and Genomics, Department of Molecular Biology and Genetics. Denmark: Research Centre Foulum, Tjele.

Nohara, A., Maejima, Y., Shimomura, K., Kumamoto, K., Takahashi, M., Akuzawa, M., … Shimomura, Y. (2015). Self-awareness of fast eating and its impact on diagnostic components of metabolic syndrome among middle-aged Japanese males and females. *Endocrine Regulations*, *49*, 91–96. https://doi.org/10.4149/endo_2015_02_91

Patterson, H. D., & Thompson, R. (1971). Recovery of inter-block information when blocksizes are unequal. *Biometrika*, *58*, 545–554. https://doi.org/10.1093/biomet/58.3.545

Rauw, W. M., Soler, J., Tibau, J., Reixach, J., & Gomez Raya, L. (2006). Feeding time andfeeding rate and its relationship with feed intake, feed efficiency, growth rate, and rate of fat deposition in growing Duroc barrows. *Journal of Animal Science*, *84*, 3404–3409.

Reyer, H., Shirali, M., Ponsuksili, S., Murani, E., Varley, P. F., Jensen, J., & Wimmers, K. (2017). Exploring the genetics of feed efficiency and feeding behaviour traits in a pig line highly selected for performance characteristics. *Molecular Genetics and Genomics*, *292*, 1001–1011. https://doi.org/10.1007/s00438-017-1325-1

Schulze, V., Roehe, R., Bermejo, J. L., Looft, H., & Kalm, E. (2003). The influence of feeding behaviour on feed intake curve parameters and performance traits of station-tested boars. *Livestock Production Science*, *82*, 105–116. https://doi.org/10.1016/S0301-6226(03)00034-4

Tajima, M., Lee, J. S., Watanabe, E., Park, J. S., Tsuchiya, R., Fukahori, A., … Kawakubo, K. (2014). Association between changes in 12 lifestyle behaviors and the development of metabolic syndrome during 1 year among workers in the Tokyo metropolitan area. *Circulation Journal*, *78*, 1152–1159. https://doi.org/10.1253/circj.CJ-13-1082

Tanihara, S., Imatoh, T., Miyazaki, M., Babazono, A., Momose, Y., Baba, M., … Une, H. (2011). Retrospective longitudinal study on the relationship between 8-year weight change andcurrent eating speed. *Appetite*, *57*, 179–183. https://doi.org/10.1016/j.appet.2011.04.017

Tao, L., Yang, K., Huang, F., Liu, X., Li, X., Luo, Y., … Guo, X. (2018). Associationbetween self-reported eating speed and metabolic syndrome in a Beijing adult population: A cross-sectional study. *BMC Public Health*, *18*, 855. https://doi.org/10.1186/s12889-018-5784-z

Von Felde, A., Roehe, R., Looft, H., & Kalm, E. (1996). Genetic association between feed intake and feed intake behaviour at different stages of growth of group-housed boars. *Livestock Production Science*, *47*, 11–12. https://doi.org/10.1016/S0301-6226(96)01006-8

Wallenbeck, A., & Keeling, L. J. (2013). Using data from electronic feeders on visit frequency and feed consumption to indicate tail biting outbreaks in commercial pig production. *Journal of Animal Science*, *91*, 2879–2884.

Wetten, M., Ødegård, J., Vangen, O., & Meuwissen, T. (2012). Simultaneous estimation of daily weight and feed intake curves for growing pigs by random regression. *Animal*, *6*, 433–439. https://doi.org/10.1017/S1751731111001832

Young, J., Cai, W., Nettleton, D. S., & Dekkers, J. C. M. (2009). Feeding behavior of Yorkshire pigs selected for residual feed intake. *Animal Industry Report*, *655*, R2454.

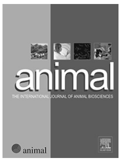
Zhu, B., Haruyama, Y., Muto, T., & Yamazaki, T. (2015). Association between eating speed and metabolic syndrome in a three-year population-based cohort study. *Journal of Epidemiology*, *25*, 332–336. https://doi.org/10.2188/jea.JE20140131

How to cite this article: Kavlak AT, Uimari P. Estimation of heritability of feeding behaviour traits and their correlation with production traits in Finnish Yorkshire pigs. *J Anim Breed Genet*. 2019;136:484–494. https://doi.org/10.1111/jbg.12408

Animal 15 (2021) 100168

Contents lists available at ScienceDirect

Animal
The international journal of animal biosciences



Estimation of social genetic effects on feeding behaviour and production traits in pigs

A.T. Kavlak [a,*], I. Strandén [b], M.H. Lidauer [b], P. Uimari [a]

[a] *Department of Agricultural Sciences, University of Helsinki, Helsinki 00014, Finland*
[b] *Natural Resources Institute Finland (Luke), Jokioinen, 31600, Finland*

ARTICLE INFO

Article history:
Received 8 May 2020
Received in revised form 16 December 2020
Accepted 18 December 2020
Available online 20 January 2021

Keywords:
Feed conversion rate
Heritability
Indirect genetic effect
Social interaction
Swine

ABSTRACT

Pigs are housed in groups during the test period. Social effects between pen mates may affect average daily gain (**ADG**), backfat thickness (**BF**), feed conversion rate (**FCR**), and the feeding behaviour traits of pigs sharing the same pen. The aim of our study was to estimate the genetic parameters of feeding behaviour and production traits with statistical models that include social genetic effects (**SGEs**). The data contained 3075 Finnish Yorkshire, 3351 Finnish Landrace, and 968 F1-crossbred pigs. Feeding behaviour traits were measured as the number of visits per day (NVD), time spent in feeding per day (TPD), daily feed intake (DFI), time spent in feeding per visit (TPV), feed intake per visit (FPV), and feed intake rate (FR). The test period was divided into five periods of 20 days. The number of pigs per pen varied from 8 to 12. Two model approaches were tested, i.e. a fixed group size model and a variable group size model. For the fixed group size model, eight random pigs per pen were included in the analysis, while all pigs in a pen were included for the variable group size model. The linear mixed-effects model included sex, breed, and herd*year*season as fixed effects and group (batch*pen), litter, the animal itself (direct genetic effect (**DGE**)), and pen mates (SGEs) as random effects. For feeding behaviour traits, estimates of the total heritable variation ($T^2 \pm SE$) and classical heritability ($h^2 \pm SE$, values given in brackets) from the variable group size model (e.g. period 1) were 0.34 ± 0.13 (0.30 ± 0.04) for NVD, 0.41 ± 0.10 (0.37 ± 0.04) for TPD, 0.40 ± 0.15 (0.14 ± 0.03) for DFI, 0.53 ± 0.15 (0.28 ± 0.04) for TPV, 0.66 ± 0.17 (0.28 ± 0.04) for FPV, and 0.29 ± 0.13 (0.22 ± 0.03) for FR. The effect of social interaction was minimal for ADG ($T^2 = 0.29 \pm 0.11$ and $h^2 = 0.29 \pm 0.04$), BF ($T^2 = 0.48 \pm 0.12$ and $h^2 = 0.38 \pm 0.07$), and FCR ($T^2 = 0.37 \pm 0.12$ and $h^2 = 0.29 \pm 0.04$) using the variable group size model. In conclusion, the results indicate that social interactions have a considerable indirect genetic effect on the feeding behaviour and FCR of pigs but not on ADG and BF.

Implications

Social interaction between group-housed pigs is a very important component for traits related to feeding behaviour, productivity, and well-being. Genetic variation in traits incorporating social interactions can be estimated using models that include both direct genetic and social genetic effects between pen mates. According to our results, the social genetic effect was important for certain feeding behaviour traits. In particular, the social genetic effect was significant for the feed conversion trait. Thus, accounting for social genetic effect in selection is beneficial for improving the feed conversion rate.

Introduction

Social interactions between animals may affect the health and productivity of livestock housed in groups. For example, cooperation has positive effects on the well-being and productivity of group members, while competition and aggression have adverse effects (Ellen et al., 2014). Aggressive behaviour is usually caused by either ranking disputes in group hierarchy or by competition for limited resources (e.g. feeding). Even though skin injuries and stress are the most notable outcomes of aggression in group-housed pigs, aggressive behaviour also decreases production (Marchant-Forde and Marchant-Forde, 2005; Rydhmer et al., 2013; Camerlink, 2014).

The effects of social interactions may be partly genetic and modelled by a social genetic effect (**SGE**) (also referred to as indirect genetic effects) (Moore et al., 1997) or associate effect (Griffing, 1967). Pigs are a typical example of livestock housed in groups both in commercial settings and at test stations. Several studies have shown that SGE is important in pigs. For example, according to Canario et al. (2010), SGE contributes 44% of the heritable variation in average daily gain (**ADG**) also Nielsen et al. (2018a) found a significant SGE for ADG. In contrast, Bouwman et al. (2010) did not find a find a significant SGE in ADG. Herrera-Cáceres et al. (2019) reported a significant SGE in feed conversion rate (**FCR**), and Bergsma et al. (2008) reported a significant SGE in

---

* Corresponding author.
E-mail address: alper.kavlak@helsinki.fi (A.T. Kavlak).

https://doi.org/10.1016/j.animal.2020.100168
1751-7311/

backfat thickness (**BF**) and daily feed intake (DFI). However, little is known of the importance of SGE on feeding behaviour traits. Feeding behaviour traits, such as the number of visits to a feeder per day and feeding speed, are expected to depend on the social behaviour of pen mates. In a recent study, Herrera-Cáceres et al. (2019) found a considerable SGE in the number of visits per day (NVD) and feed intake rate (FR) but not in the time spent in feeding per day (TPD).

Modelling SGE in the context of quantitative genetics has been presented by Griffing (1967), Muir and Schinkel (2002), and Bijma et al. (2007a). In an SGE model, the phenotypic value of an animal depends on its direct breeding value and the sum of the social breeding values of its group mates plus the corresponding non-heritable direct and social effects. Total breeding value of an animal is the sum of its direct breeding value and its own social breeding values towards group mates. The variance of total breeding value depends on the (co)variance components of direct breeding value and social breeding values but also on group size (Bijma et al., 2007a, 2007b). In socially affected traits, the proportion of total breeding value variance over the total phenotypic variance is the measure for inheritable variation preferred over the classical heritability.

Statistical modelling and estimation of variance components in an SGE model are relatively easy with existing variance/covariance estimation programmes when all groups have the same number of animals, i.e. the group size is constant. In practice, group size often varies in commercial settings within and between farms but also at test stations. Even when the aim is to include groups of equal size during the test period, group (batch*pen) size may vary due to several reasons. Variable group sizes pose challenges in estimating variance components for the direct and SGEs using existing variance component estimation programmes. Several solutions have been proposed for coping with variable group sizes (e.g. Arango et al., 2005; Hadfield and Wilson, 2007; Bijma et al., 2007a; Nielsen et al., 2018a).

Given the limited number of publications related to the importance of SGE in feeding behavioural traits in general and in Finnish commercial pig breeds, the objective of our study was to estimate genetic parameters of feeding behaviour and production traits in Finnish pig breeds using a model with an SGE. Estimation and comparison of the magnitude of T^2 and h^2 were our main interests. We also compared the fixed group size SGE model to the more complicated variable group size SGE model.

Material and methods

Data

The data were provided by Figen Oy (Pietarsaari, Finland) and included the feeding and production records of pigs from the central test station from 2011 to 2016. Table 1 shows the ages and weights of the animals during their test periods with the slaughter records.

Pigs arrived at the test station on Tuesday or on Wednesday, and the test started on Saturday. The grouping of pigs to different pens was done according to the arrival weight (same size) and sex (only boars or combination of gilts and castrates) of the pigs. Feeding was *ad libitum* consisting of two commercial feedstuffs. The proportion of the two feedstuffs was based on the growth rate curve of an average pig from the previous test periods. The piggery has automated air conditioning and ventilation based on the age of the pigs and outdoor temperature. Artificial light is on from 7 am to 3 pm. The dimension of the pen is 16.8 m^2 with one-third of concrete slats. The feedings were recorded automatically using the Schauer Spotmix with Schauer MLP electronic feeders and MLP manager data management software (Schauer Agrotronic GmbH).

Table 1
Descriptive statistics of the pigs.

	Mean ± SD
Arrival age at test station	89 ± 10 days
Weight at arrival age	34.4 ± 6.4 kg
Total test time	95 ± 3 days
Slaughter age	186 ± 10 days
Slaughter weight	121.2 ± 12.9 kg

The raw data consisted of 28 964 641 observations made from Finnish Yorkshire, Finnish Landrace, and F1-crossbred pigs and included transponder id, date, time of entering the feeder, time leaving the feeder, and feed intake per visit (FPV). Some visits may be missed due to ear tag-related problems, either the feeding system was not able to record the tag properly or some ear tags might have fallen out from pigs. However, these problems are rare (personal communication with the personnel of the test station). In addition, some pigs were removed from the tests due to sickness (e.g. lameness, loss of appetite, etc.). Otherwise, the thresholds presented in Casey et al. (2005) were applied to remove possible erroneous data. The proportion of outliers was less than 1%. The remaining data contained 28 826 029 observations.

These separate visit observations were used to calculate daily values for the NVD (counts), TPD (min), DFI (g), time spent in feeding per visit (TPV, min), FPV (g), and FR (FR = FPV/TPV, g/min). The final records used in variance component estimation were calculated as averages of the daily records for five test periods of 20 days each: 0–20, 21–40, 41–60, 61–80, and 81–93 days (for more information, see Kavlak and Uimari, 2019). In the following, we use abbreviation where the number indicates the corresponding test period, e.g. NVD1 is NVD from the test period 1, and NVD2 is NVD from the test period 2, etc.

Production traits were ADG (g), FCR (g/g), and BF (mm). For an individual pig, ADG was calculated as the ratio of the difference between last day weight and first day weight within a test period over the duration of the test period in days. Feed conversion rate was measured as feed consumption during the test period measured in feed units divided by the total growth during the test period. Lastly, BF was calculated using the average measurement result of a Hennessy grading probe (type GP4, Auckland, New Zealand), where one sample was taken 8 cm off the midline of the carcass behind the last rib and another 6 cm off the midline between the third and fourth ribs.

Before the statistical analyses, outliers (4 standard deviations of the mean) were removed from the data. The final data included records from 7 394 pigs. The number of animals in pens varied from 8 to 12. Average group size was 9.8. The average additive genetic relationship (r) between the pigs within a group was 0.08, which was estimated using the pedigree data of 11 301 animals with ancestors down to four generations. The average number of offspring with observations per each sire was 16.6.

Statistical analyses

Variance component estimation used a single trait SGEs model. For each trait, two data sets with different requirements from the statistical model were analysed. In the first data, the number of pen mates was restricted to the same fixed number of observations by randomly sampling 8 pen mates with records from each pen. We call this the fixed group size model. In the second data, the number of pen mates was allowed to vary according to the size of the original pen. Only pens where all pen mates had records were included in the analysis. Consequently, numbers of accepted records and pens varied by trait. We call this the variable group size model. Restricted maximum likelihood (**REML**) estimates of variance components were calculated using average information algorithm or AI-REML as implemented in *DMU* software (Madsen and Jensen, 2013). Standard errors to the estimates were taken from the DMU output except for estimates to h^2 and T^2 where Taylor series expansion were used (Dieters et al., 1995).

The social model included sex (5 035 boars, 1 296 gilts, and 1 063 castrates), breed (3075 Yorkshire, 3351 Landrace, and 968 F1 crosses of the two breeds), and herd*year*season (348 classes) as fixed effects, and batch*pen (766 groups), litter (2 862 groups), animal (direct genetic effect (**DGE**) and SGE), and residual as random effects.

$$\mathbf{y} = \mathbf{Xb} + \mathbf{Z}_\mathrm{D}\mathbf{a}_\mathrm{D} + \mathbf{Z}_\mathrm{S}\mathbf{a}_\mathrm{S} + \mathbf{Z}_\mathrm{l}\mathbf{l} + \mathbf{Z}_\mathrm{g}\mathbf{g} + \mathbf{e}$$

where **y** is a vector of observations (feeding behaviour or production), **b** is a vector of fixed effects, **X** is an incidence matrix relating records to fixed effects, $\mathbf{a}_\mathrm{D}$ is a vector of random DGE, $\mathbf{a}_\mathrm{S}$ is a vector of random SGE, **l** is a vector of random litter effects, **g** is a vector of random group (batch*pen) effects, and **e** is a vector of random residuals; the corresponding incidence matrices are $\mathbf{Z}_\mathrm{D}$, $\mathbf{Z}_\mathrm{S}$, $\mathbf{Z}_\mathrm{l}$, and, $\mathbf{Z}_\mathrm{g}$, respectively. The random effects were considered independent except between DGE and SGE:

$$\begin{bmatrix} \mathbf{a}_D \\ \mathbf{a}_S \end{bmatrix} = \mathbf{G} \otimes \mathbf{A}, \quad \text{where } \mathbf{G} = \begin{bmatrix} \sigma^2_{AD} & \sigma_{A_{DS}} \\ \sigma_{A_{DS}} & \sigma^2_{AS} \end{bmatrix}$$

A is the relationship matrix, σ^2_{AD} is the variance of DGE, σ^2_{AS} is the variance of SGE, $\sigma_{A_{DS}}$ is the covariance between DGE and SGE, and ⊗ is the Kronecker product. Litter, batch*pen (group), and residual effects were independently and normally distributed with expected values of **0** and variances of σ^2_l, σ^2_g, and σ^2_e for litter, group, and residual effects, respectively.

The variance of total breeding value (σ^2_{TBV}) (Bijma et al., 2007a, 2007b) is

$$\sigma^2_{TBV} = \sigma^2_{AD} + 2(n-1)\sigma_{A_{DS}} + (n-1)^2\,\sigma^2_{AS}$$

and the total phenotypic variance (σ^2_P) is

$$\sigma^2_P = \sigma^2_{AD} + \sigma^2_g + \sigma^2_l + \sigma^2_e + (n-1)\sigma^2_{AS} + (n-1)r\left[2\sigma_{A_{DS}} + (n-2)\sigma^2_{AS}\right]$$

where n is the average number of pigs in each group and r is the average relatedness within groups (r = 0.08). Total heritable variation (T^2) is the ratio $T^2 = \sigma^2_{TBV}/\sigma^2_P$) and classical heritability (h^2) is the ratio $h^2 = \sigma^2_{AD}/\sigma^2_P$.

Results

Phenotypic description

Descriptive statistics of feeding behaviour and production traits are shown in Figs. 1 and 2. In general, the older pigs became, the more frequently they visited the feed station (from NVD1 to NVD5). In contrast, time spent in feeding decreased slightly from 2.9 min (TPV1) to 2.2 min (TPV5). Thus, TPD increased up to 65.2 min/day (TPD2) and then decreased to 54.1 min/day (TPD5). In contrast to NVD and TPD, the rest of the feeding behaviour traits (TPV, DFI, FPV, and FR) increased constantly from period 1 to period 5 (Fig. 1). Mean ADG for the pigs was 946 ± 113 g/d (mean and SD). Corresponding mean values for BF and FCR were 10.7 ± 2.3 mm and 2.5 ± 0.2 g/g, respectively.

Classical heritability

Both estimation models gave the same range of estimates of classical heritability for the feeding behaviour traits: 0.14–0.39 (Tables 2 and 3). For the production traits estimates of classical heritability varied from 0.25 to 0.41 when a fixed group size model was applied and from 0.29 to 0.38 when a variable group size model was applied (Tables 2 and 3). Standard errors of the estimates varied from 0.03 to 0.05. The highest estimates for feeding behaviour traits were obtained for periods 2 and 3.

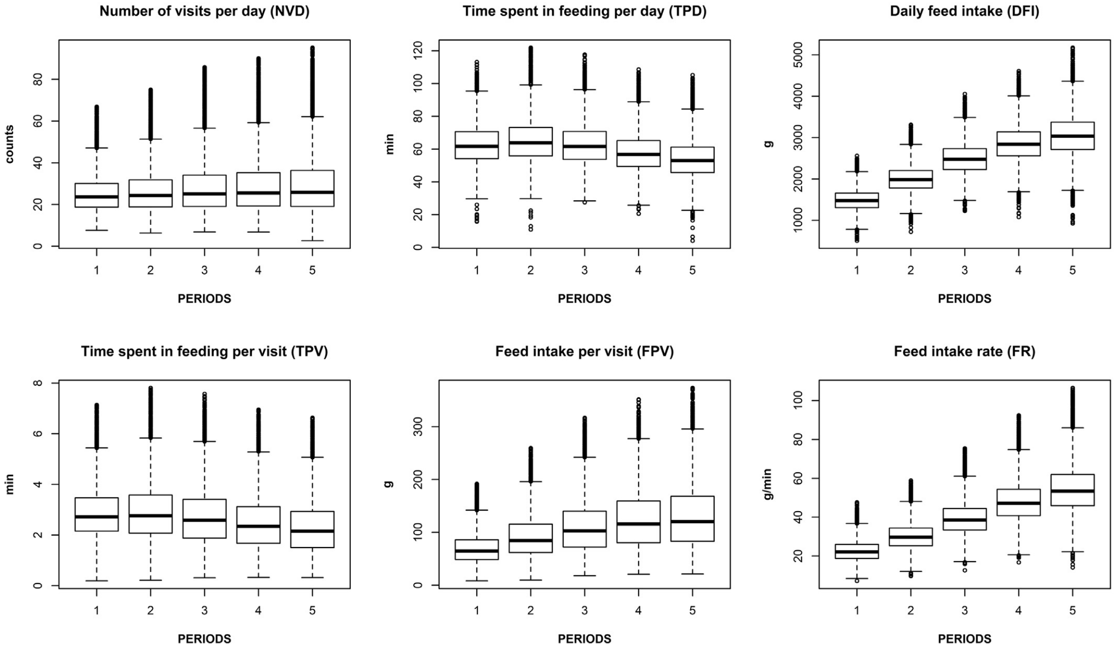


Fig. 1. Box plots of feeding behaviour traits over the five periods in pigs.

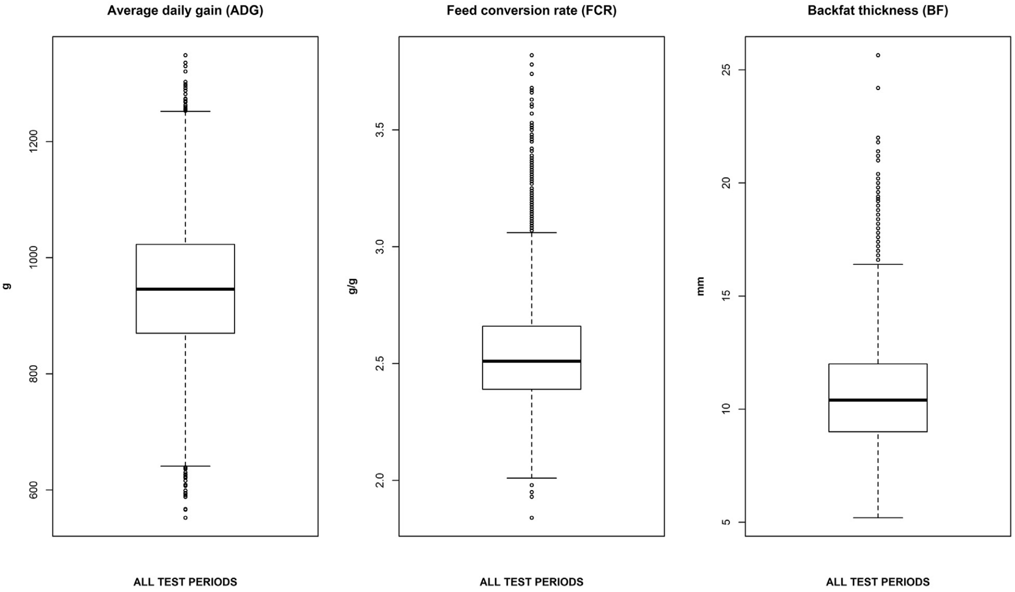


Fig. 2. Box plots of the production traits over the entire test period in pigs.

Variance of social genetic effect and correlation between direct genetic effect and social genetic effect

Variance of SGE (σ^2_{AS}) was generally smaller than the variance of DGE (σ^2_{AD}, Tables 2 and 3). Also, the SEs of the σ^2_{AS} estimates were large. Thus, most of the σ^2_{AS} estimates did not differ from 0 except for TPV and FPV for some traits during certain periods (the significant estimates are bolded in Tables 2 and 3). Most of the significant σ^2_{AS} estimates were obtained in test period 2 (days 21–40 of test). In addition, the SGE variance was small but significant for FCR but not for ADG or BF.

Estimated correlation between DGE and SGE was mostly positive in feeding behaviour traits but not statistically significant (Tables 2 and 3). When a fixed group size model was applied, only TPV1, TPV2, TPV3, TPV4, FPV3, and FPV4 indicated a statistically significant and positive correlation between DGE and SGE. When a variable group size model was applied, the correlation between DGE and SGE was significant only for TPV2 and TPV3. Based on our results, pigs that visit feeders for longer time periods (a genetic potential) also have a positive SGE on the TPV of the pen mates. This relationship was also true for FPV. No statistically significant correlation between DGE and SGE was obtained for the production traits.

Total heritable variation

For feeding behaviour traits, the estimates of T^2 were generally higher than estimates of h^2, ranging from 0.28 ± 0.08 ($h^2 = 0.28 \pm 0.04$) for DFI4 to 0.77 ± 0.13 ($h^2 = 0.36 \pm 0.04$) for TPV3 when a fixed group size model was applied and from 0.17 ± 0.10 ($h^2 = 0.27 \pm 0.04$) for FR3 to 0.85 ± 0.16 ($h^2 = 0.34 \pm 0.03$) for TPV3 when a variable group size model was applied (Tables 2 and 3). For production traits, the estimates of T^2 were similar to the estimates of classical heritability, except for FCR in both the fixed ($T^2 = 0.75 \pm 0.17$ vs. $h^2 = 0.25 \pm 0.04$) and variable ($T^2 = 0.37 \pm 0.12$ vs. $h^2 = 0.29 \pm 0.04$) group size models.

Differences between the models

Estimates and SEs of h^2 were almost identical by the fixed group size and variable group size models (Tables 2 & 3). In addition, estimates of T^2 were similar between the models (the differences between the estimates were within SEs). In production traits, some differences were observed in T^2 between the models, e.g. for FCR the T^2 estimate was 0.75 ± 0.17 using the fixed group size model and 0.37 ± 0.12 using the variable group size model. The difference in estimates of T^2 for FCR is most probably due to different data size; for the variable group size model only pens where all animals had records were included in the analysis (in the fixed group size model all pens with at least eight animals having a record were included in the analysis). In general, a slightly higher correlation between DGE and SGE was obtained from the fixed group size model than from the variable group size model.

Discussion

In this study, we investigated the importance of SGE on feeding behaviour and production traits in Finnish pig breeds. The data were obtained from the test station where animals of two Finnish pig breeds (Finnish Landrace, Finnish Yorkshire) and their F1-crosses were distributed randomly into each pen. Thus, the data did not allow separate analysis of the breeds. This may affect our results if the two breeds differ from each other drastically in terms of the studied traits. However, both are white breeds with similar breeding goals and programmes. Also, based on a previous study of Finnish Yorkshire by Kavlak and Uimari (2019) and Finnish Landraces by Riikimaki (2019), feeding behaviour and estimated heritabilities of feeding behaviour traits were similar between these two breeds. In addition, prior to the SGE model analysis, the same model used in Kavlak and Uimari (2019) was applied for the combined data. The estimated heritabilities were close to those presented in Kavlak and Uimari (2019). Thus, despite the data having two unrelated populations analysed together, we trust that the data provide reasonable and reliable estimates of SGE for feeding

Table 2
Estimates of the genetic parameters from the fixed group size model for feeding behaviour and production traits in pigs.

	N	σ^2_{AD}	σ^2_{AS}	$\sigma_{A_{DS}}$	σ^2_P	$r_{A_{DS}}$	σ^2_{TBV}	h^2	T^2
FBT									
Period 1									
NVD1	7349	22.2 ± 3.1	0.18 ± 0.10	0.32 ± 0.45	74.6	0.16 ± 0.23	35.5	0.30 ± 0.04	0.48 ± 0.12
TPD1	7370	48.7 ± 6.4	0.23 ± 0.14	−0.29 ± 0.72	136.0	−0.08 ± 0.21	55.9	0.36 ± 0.04	0.41 ± 0.09
DFI1	7388	7523 ± 1745	84 ± 83	330 ± 267	53279	0.41 ± 0.35	16259	0.14 ± 0.03	0.31 ± 0.12
TPV1	7372	0.26 ± 0.03	0.002 ± 0.001*	0.01 ± 0.005	0.89	0.57 ± 0.24*	0.49	0.29 ± 0.04	0.55 ± 0.13
FPV1	7390	170.7 ± 25.7	2.2 ± 1.1	4.3 ± 4.0	627.9	0.22 ± 0.20	338.7	0.27 ± 0.04	0.54 ± 0.14
FR1	7378	5.5 ± 0.9	0.08 ± 0.04	−0.12 ± 0.14	25.2	−0.18 ± 0.22	7.7	0.22 ± 0.03	0.31 ± 0.11
Period 2									
NVD2	7352	29.6 ± 3.8	0.15 ± 0.09	0.89 ± 0.55	94.0	0.42 ± 0.25	49.4	0.31 ± 0.04	0.53 ± 0.11
TPD2	7380	65.1 ± 8.2	0.31 ± 0.20	0.73 ± 0.91	166.8	0.16 ± 0.20	90.5	0.39 ± 0.04	0.54 ± 0.11
DFI2	7389	20204 ± 3401	688 ± 208*	781 ± 544	84182	0.21 ± 0.14	64850	0.24 ± 0.04	0.77 ± 0.16
TPV2	7364	0.34 ± 0.04	0.003 ± 0.001*	0.02 ± 0.006*	1.06	0.60 ± 0.20*	0.76	0.32 ± 0.04	0.70 ± 0.13
FPV2	7376	320.1 ± 45.8	4.2 ± 1.9*	10.0 ± 7.0	1121.4	0.27 ± 0.19	665.9	0.29 ± 0.04	0.59 ± 0.13
FR2	7382	10.5 ± 1.6	0.16 ± 0.07*	−0.01 ± 0.25	41.1	−0.01 ± 0.19	18.2	0.25 ± 0.04	0.43 ± 0.13
Period 3									
NVD3	7351	40.4 ± 5.2	0.22 ± 0.13	1.12 ± 0.71	123.4	0.37 ± 0.24	66.9	0.33 ± 0.04	0.54 ± 0.11
TPD3	7380	58.9 ± 7.1	0.26 ± 0.17	0.78 ± 0.81	150.1	0.19 ± 0.20	82.5	0.39 ± 0.04	0.55 ± 0.11
DFI3	7390	34086 ± 5068	375 ± 212	775 ± 698	114742	0.21 ± 0.19	63311	0.30 ± 0.04	0.55 ± 0.14
TPV3	7383	0.37 ± 0.04	0.003 ± 0.001*	0.02 ± 0.006*	1.02	0.56 ± 0.17*	0.79	0.36 ± 0.04	0.77 ± 0.13
FPV3	7386	576.4 ± 76.2	5.7 ± 2.5*	24.0 ± 10.7*	1702.3	0.42 ± 0.19*	1191.7	0.34 ± 0.04	0.70 ± 0.13
FR3	7372	19.1 ± 2.9	0.28 ± 0.15	−0.15 ± 0.43	70.1	−0.06 ± 0.19	30.7	0.27 ± 0.04	0.44 ± 0.14
Period 4									
NVD4	7354	39.3 ± 5.2	0.25 ± 0.16	0.36 ± 0.74	134.4	0.11 ± 0.23	56.6	0.29 ± 0.04	0.42 ± 0.10
TPD4	7377	46.1 ± 5.8	0.42 ± 0.21	0.23 ± 0.75	125.9	0.05 ± 0.17	69.9	0.37 ± 0.04	0.56 ± 0.12
DFI4	7392	42234 ± 6575	90 ± 146	−310 ± 769	152098	−0.16 ± 0.38	42304	0.28 ± 0.04	0.28 ± 0.08
TPV4	7380	0.28 ± 0.03	0.002 ± 0.001*	0.01 ± 0.005*	0.80	0.57 ± 0.18*	0.52	0.35 ± 0.04	0.65 ± 0.13
FPV4	7390	689.3 ± 91.0	6.2 ± 2.8*	27.0 ± 12.3*	2068.9	0.41 ± 0.19*	1371.1	0.33 ± 0.04	0.66 ± 0.12
FR4	7369	29.5 ± 4.4	0.40 ± 0.22	0.25 ± 0.66	106.0	0.07 ± 0.19	52.6	0.28 ± 0.04	0.50 ± 0.15
Period 5									
NVD5	7366	41.2 ± 5.8	0.36 ± 0.21	0.84 ± 0.86	158.6	0.22 ± 0.22	70.6	0.26 ± 0.03	0.44 ± 0.11
TPD5	7376	37.1 ± 5.4	0.17 ± 0.16	0.57 ± 0.70	130.9	0.22 ± 0.26	53.4	0.28 ± 0.04	0.41 ± 0.11
DFI5	7387	55050 ± 8688	1157 ± 515*	1635 ± 1347	226984	0.20 ± 0.17	134633	0.24 ± 0.04	0.59 ± 0.15
TPV5	7382	0.26 ± 0.03	0.001 ± 0.001	0.006 ± 0.004	0.75	0.30 ± 0.22	0.39	0.34 ± 0.04	0.52 ± 0.12
FPV5	7390	802.8 ± 108.6	7.1 ± 3.4*	17.1 ± 14.7	2499.5	0.22 ± 0.19	1390.1	0.32 ± 0.04	0.56 ± 0.12
FR5	7371	39.8 ± 5.9	0.07 ± 0.33	1.1 ± 0.91	143.7	0.22 ± 0.16	58.6	0.28 ± 0.04	0.41 ± 0.16
PT									
ADG	7394	3111.1 ± 492.2	13.5 ± 13.4	−53.0 ± 57.1	10841	−0.25 ± 0.29	3030.6	0.29 ± 0.04	0.28 ± 0.09
BF	6526	1.6 ± 0.22	0.005 ± 0.004	−0.02 ± 0.02	3.9	−0.28 ± 0.17	1.51	0.41 ± 0.05	0.38 ± 0.09
FCR	7340	0.01 ± 0.001	0.0002 ± 0.00009*	0.0004 ± 0.0002*	0.04	0.30 ± 0.17	0.03	0.25 ± 0.04	0.75 ± 0.17

FBT = feeding behaviour traits; PT = production traits; NVD = number of visits per day; TPD = time spent in feeding per day; DFI = daily feed intake; TPV = time spent in feeding per visit; FPV = feed intake per visit; FR = feed intake rate; ADG = average daily gain; BF = backfat thickness; FCR = feed conversion rate; σ^2_{AD} = variance of direct genetic effect; σ^2_{AS} = variance of social genetic effect; $\sigma_{A_{DS}}$ = covariance between direct genetic effect and social genetic effect; σ^2_P = total phenotypic variance; σ^2_{TBV} = variance of total breeding value; $r_{A_{DS}}$ = genetic correlation between direct genetic effect and social genetic effect; h^2 = classical heritability; T^2 = total heritable variation. N = number of observations for each trait. ± indicates the SEs of the estimates. *Statistically significant estimates ($P < 0.05$) of σ^2_{AS}, $\sigma_{A_{DS}}$, and $r_{A_{DS}}$.

behaviour and production traits in these two Finnish pig breeds. However, more precise estimates of SGE variance could have been obtained if the data consisted only one population and the groups were formed with few families per group (Bijma, 2010a; Ødegård and Olesen, 2011).

According to our results, SGE was not statistically significant for most of the feeding behaviour and production traits, except for TPV, FPV, and FCR. The estimates of the total heritable variation for TPV and FPV were generally over 0.5 (±0.17) (depending on the period and method) and 0.4–0.7 (±0.17) for FCR. Using data from a Duroc pig population, Herrera-Cáceres et al. (2019) obtained a significant contribution of SGE for NVD ($T^2 = 0.93 \pm 0.49$ vs. $h^2 = 0.46 \pm 0.09$), DFI ($T^2 = 0.29 \pm 0.29$ vs. $h^2 = 0.25 \pm 0.08$), TPV ($T^2 = 0.67 \pm 0.\ 30$ vs. $h^2 = 0.47 \pm 0.09$), and FR ($T^2 = 0.39 \pm 0.29$ vs. $h^2 = 0.32 \pm 0.08$), while a high negative correlation between DGE and SGE was obtained for TPD (−0.78 ± 0.09), leading to a smaller estimate of T^2 than h^2 ($T^2 = 0.20 \pm 0.19$ vs. $h^2 = 0.27 \pm 0.10$). For the production traits, SGE was less important (for ADG $T^2 = 0.22 \pm 0.18$ vs. $h^2 = 0.22 \pm 0.09$ and for BF $T^2 = 0.51 \pm 0.28$ vs. $h^2 = 0.35 \pm 0.11$) except for FCR ($T^2 = 0.55 \pm 0.43$ vs. $h^2 = 0.24 \pm 0.09$) (Herrera-Cáceres et al., 2019). Based on the data of 14 032 crossbred pigs, Bergsma et al. (2008) obtained a significant contribution of SGE for ADG ($T^2 = 0.71 \pm 0.08$ vs. $h^2 = 0.21 \pm 0.02$), DFI ($T^2 = 0.70 \pm 0.17$ vs. $h^2 = 0.17 \pm 0.03$), and BF ($T^2 = 0.41 \pm 0.04$ vs. $h^2 = 0.35 \pm 0.02$). Nielsen et al. (2018a) estimated SGE of ADG separately for Danish Landrace gilts and boars using a bivariate model and found that the T^2 is stronger between boars (0.32 ± 0.02) than between gilts (0.27 ± 0.01). Thus, our results are in line with other published results and confirm the pattern that SGE is important in feeding behaviour traits (at least in TPV and FPV) and in FCR but less important in BF and ADG.

Although the aim is to have an equal number of pigs in all pens during the test period, the group sizes vary, in our case, mainly because of animals being removed from the test due to health issues (various diseases and leg problems). Also culling of animals from the groups and changing the composition during the test period may cause the variability in SGEs of the animals unless it is taken into account in the model, e.g. by a regression of mean proportion of time spent in a pen or relative space allowance on the studied trait (Ask et al., 2020). However, the effect of including the regression or omitting it did not have a significant effect on estimates of the variance components or h^2 or T^2 (Ask et al., 2020). We did not apply this fine-tuned approach in our analysis. Instead, we first applied a simple model for randomly sampling an equal number of pigs ($n = 8$) from each pen for a fixed group size model and compared these results to those from a variable group size model ($n = 9.8$). The fixed group size model is simpler than the variable

Table 3
Estimates of the genetic parameters from the variable group size model for feeding behaviour and production traits in pigs.

	N	σ^2_{AD}	σ^2_{AS}	$\sigma_{A_{DS}}$	σ^2_P	$r_{A_{DS}}$	σ^2_{TBV}	h^2	T^2
FBT									
Period 1									
NVD1	7091	20.7 ± 2.9	0.15 ± 0.08	−0.50 ± 0.40	69.0	−0.28 ± 0.22	23.7	0.30 ± 0.04	0.34 ± 0.13
TPD1	7208	49.8 ± 6.5	0.18 ± 0.10	−0.46 ± 0.68	134.8	−0.15 ± 0.22	55.8	0.37 ± 0.04	0.41 ± 0.10
DFI1	7335	7660.6 ± 1767.5	93.3 ± 68.2	360.2 ± 261.0	53049	0.43 ± 0.30	21224	0.14 ± 0.03	0.40 ± 0.15
TPV1	7204	0.24 ± 0.03	0.001 ± 0.008	0.007 ± 0.005	0.87	0.43 ± 0.29	0.46	0.28 ± 0.04	0.53 ± 0.15
FPV1	7269	170.4 ± 25.8	2.00 ± 0.88*	4.6 ± 3.8	617.2	0.25 ± 0.20	407.2	0.28 ± 0.04	0.66 ± 0.17
FR1	7350	5.5 ± 0.9	0.06 ± 0.03*	−0.17 ± 0.13	25.0	−0.29 ± 0.24	7.4	0.22 ± 0.03	0.29 ± 0.13
Period 2									
NVD2	7052	28.3 ± 3.8	0.10 ± 0.06	−0.58 ± 0.50	87.5	−0.34 ± 0.30	25.9	0.32 ± 0.04	0.30 ± 0.11
TPD2	7266	63.8 ± 8.2	0.19 ± 0.14	0.11 ± 0.84	164.8	0.03 ± 0.24	80.9	0.39 ± 0.04	0.49 ± 0.12
DFI2	7344	19231 ± 3329	432.3 ± 151.4	392.2 ± 497	83637	0.14 ± 0.17	59616	0.23 ± 0.04	0.71 ± 0.18
TPV2	7136	0.32 ± 0.04	0.002 ± 0.001*	0.02 ± 0.006*	1.05	0.58 ± 0.22*	0.77	0.31 ± 0.04	0.73 ± 0.16
FPV2	7240	315.5 ± 45.7	3.4 ± 1.5*	5.4 ± 6.7	1100.2	0.16 ± 0.20	676.9	0.29 ± 0.04	0.62 ± 0.16
FR2	7334	10.4 ± 1.6	0.13 ± 0.05*	−0.12 ± 0.24	42.2	−0.10 ± 0.21	18.4	0.25 ± 0.04	0.44 ± 0.15
Period 3									
NVD3	7046	39.1 ± 5.2	0.18 ± 0.10	0.17 ± 0.69	118.5	0.06 ± 0.26	55.9	0.33 ± 0.04	0.47 ± 0.13
TPD3	7276	57.6 ± 7.1	0.17 ± 0.12	−0.10 ± 0.73	147.0	−0.03 ± 0.23	68.9	0.39 ± 0.04	0.47 ± 0.11
DFI3	7354	34236 ± 5082	301 ± 167	824 ± 667	115020	0.26 ± 0.21	72060	0.30 ± 0.04	0.63 ± 0.16
TPV3	7292	0.35 ± 0.04	0.003 ± 0.001*	0.02 ± 0.006*	1.02	0.49 ± 0.18*	0.87	0.34 ± 0.04	0.85 ± 0.16
FPV3	7330	558.6 ± 75.2	5.2 ± 2.1*	18.6 ± 10.4	1685.8	0.35 ± 0.19	1292.2	0.33 ± 0.04	0.77 ± 0.16
FR3	7219	18.1 ± 2.8	0.02 ± 0.04	−0.45 ± 0.32	67.9	−0.75 ± 0.93	11.6	0.27 ± 0.04	0.17 ± 0.10
Period 4									
NVD4	7039	34.9 ± 5.0	0.18 ± 0.12	−0.51 ± 0.68	127.9	−0.20± 0.27	40.1	0.27 ± 0.04	0.31 ± 0.11
TPD4	7252	45.1 ± 5.7	0.10 ± 0.09	−0.75 ± 0.59	121.7	−0.35 ± 0.29	39.6	0.37 ± 0.04	0.33 ± 0.10
DFI4	7373	40623 ± 6432	48 ± 104	342 ± 688	152074	0.24 ± 0.61	50399	0.27 ± 0.04	0.33 ± 0.09
TPV4	7275	0.26 ± 0.03	0.002 ± 0.001*	0.01 ± 0.005	0.81	0.51 ± 0.21*	0.59	0.32 ± 0.04	0.73 ± 0.15
FPV4	7353	674.1 ± 90.5	5.4 ± 2.3*	14.3 ± 11.9	2052.4	0.24 ± 0.20	1342.2	0.33 ± 0.04	0.65 ± 0.14
FR4	7204	28.7 ± 4.3	0.04 ± 0.06	−0.23 ± 0.52	102.7	−0.20 ± 0.46	28.2	0.28 ± 0.04	0.28 ± 0.11
Period 5									
NVD5	7146	40.4 ± 5.8	0.29 ± 0.16	0.73 ± 0.82	153.7	0.21 ± 0.24	76.3	0.26 ± 0.04	0.50 ± 0.14
TPD5	7265	35.5 ± 5.2	0.07 ± 0.09	−0.42 ± 0.59	125.6	−0.27 ± 0.40	33.4	0.28 ± 0.04	0.27 ± 0.10
DFI5	7321	59630 ± 9162	24 ± 156	1212 ± 944	222910	1.00 ± 2.95	82895	0.27 ± 0.04	0.37 ± 0.11
TPV5	7296	0.24 ± 0.03	0.001 ± 0.0008	0.004 ± 0.004	0.76	0.22 ± 0.24	0.42	0.32 ± 0.04	0.55 ± 0.13
FPV5	7352	788.1 ± 108.1	6.0 ± 2.6*	4.5 ± 14.3	2471.0	0.07 ± 0.21	1329.1	0.32 ± 0.04	0.54 ± 0.13
FR5	7193	38.1 ± 5.8	0.17 ± 0.12	0.31 ± 0.73	144.7	0.12 ± 0.29	57.1	0.26 ± 0.04	0.39 ± 0.12
PT									
ADG	7394	3118.2 ± 492.4	17.9 ± 11.1	−76.5 ± 57.3	10833.1	−0.32 ± 0.23	3162.1	0.29 ± 0.04	0.29 ± 0.11
BF	3542	1.6 ± 0.32	0.01 ± 0.008	−0.03 ± 0.04	4.2	−0.22 ± 0.26	2.03	0.38 ± 0.07	0.48 ± 0.12
FCR	6901	0.01 ± 0.001	$0.20\times10^{-5} \pm 3.5\times10^{-5}$*	0.0001 ± 0.0002	0.04	1.00 ± 8.95	0.01	0.29 ± 0.04	0.37 ± 0.12

FBT = feeding behaviour traits; PT = production traits; NVD = number of visits per day; TPD = time spent in feeding per day; DFI = daily feed intake; TPV = time spent in feeding per visit; FPV = feed intake per visit; FR = feed intake rate; ADG = average daily gain; BF = backfat thickness; FCR = feed conversion rate; σ^2_{AD} = variance of direct genetic effect; σ^2_{AS} = variance of social genetic effect; $\sigma_{A_{DS}}$ = covariance between direct genetic effect and social genetic effect; σ^2_P = total phenotypic variance; σ^2_{TBV} = variance of total breeding value; $r_{A_{DS}}$ = genetic correlation between direct genetic effect and social genetic effect; h^2 = classical heritability; T^2 = total heritable variation. N = number of observations for each trait. ± indicates the SEs of the estimates. *Statistically significant estimates ($P < 0.05$) of σ^2_{AS}, $\sigma_{A_{DS}}$, and $r_{A_{DS}}$.

group size model. In theory, random sampling of pen mates from the complete data for the fixed group size model can be expected to decrease the amount of information for the data analysis. However, simpler models tend to behave better numerically and allow the use of a wider range of software. When the variable group size model was used to analyse the full data, we observed that traits DFI5, TPD4, and FCR showed no convergence using AI-REML, nor using EM-REML, in sufficient time after several restarts using either already reached or different starting values. In these traits, the correlation between direct and SGEs was estimated to be one or almost one. No such convergence problems were observed using the fixed group size model. Overall, both models gave similar estimates and SEs of the variance components; the small differences between estimates were within the SEs. As a tendency, the fixed group size model gave higher estimates of the correlation between DGE and SGE than the variable group size model. Again, the differences in estimates between the two models were within the SE. Based on our results, a simple fixed group size model with a random selection of pen mates is useable for estimation of variance components even when the actual group size varies.

Social genetic effects among group mates may depend on group size: social interaction between animal pairs are weaker in a large group than in a small group (Arango et al., 2005). The dependency between group sizes and variances has been discussed in Bijma (2010b), where he also proposed a method to account for this dependence with a dilution parameter (d) that is trait dependent and has a value of 0 if SGE does not depend on group size and a value of 1 if SGEs are inversely proportional to group size, i.e. group member influences are diluted in large groups. Because the minimum group size in our data was 8 and the maximum was 12, our model may benefit from a dilution parameter. Therefore, we tested two dilution parameters: d = 0.5 and d = 1 (results not shown). In general, the estimated heritability stayed about the same. However, small differences (within SEs) appeared in the covariance between DGEs and SGEs leading to changes in T^2. Moreover, correlations between the diluted and nondiluted heritability and total heritable variation estimates were both over 90%. Certain field data analyses with the dilution parameter have been published (Canario et al., 2010; Nielsen et al., 2018b), but Duijvesteijn et al. (2012) did not find a dependency between the magnitude of SGE and the group size. Also, precision of the estimates of the dilution parameter may be low if the coefficient of variation in the group size is low (Heidaritabar et al., 2019).

When SGE has a significant contribution to the total heritable variance (T^2), it is important to consider social interactions between animals in a selection programme. Selection methods that ignore SGE may lead to unexpected response to selection, especially if DGE and SGE are negatively correlated (Bijma et al., 2007a). As an example of a benefit of including SGE in a breeding programme, Bijma et al. (2007b) estimated that when properly accounting for social interactions, a three-fold higher total heritable variation and potential response to selection may be obtained in a layer chicken population in terms of mortality caused by pecking behaviour than with a breeding programme that ignores SGE. Also, Ellen et al. (2014) demonstrated that selection for SGE is promising in many species, e.g. cattle, cod, deer, mice, mink, and pigs. To optimize SGE estimation, the number of groups should be reasonably large (Bijma, 2010a). Thus, SGE estimation is difficult in breeding programmes where data are collected from farms with a single group such as cattle (Ellen et al., 2014). Based on our results, including SGE in the estimation model for FCR, a higher response to selection may be obtained compared to the currently applied selection based purely on DGE.

Conclusion

In this study, we investigated the importance of accounting for SGE using two statistical models (fixed or variable group size) for feeding behaviour and production traits in Finnish pig breeds. The two models gave similar estimates of the variance components. Among the studied feeding behaviour traits, SGE was significant only for TPV and FPV. For these traits, the difference between T^2 (around 0.7) and h^2 (around 0.3) was large. However, these traits are generally not important in breeding programmes. The opposite is true for FCR, which is very important in most pig breeding programmes. Based on our and other published results, SGE should be accounted for in breeding value estimation of FCR and exploited in selection. Fortunately, there is no evidence of negative correlation between DGE and SGE for FCR. Thus, ignoring SGE in breeding programmes and selecting purely on DGE does not cause negative effects on social interactions that may affect FCR in the long run.

Ethics approval

The data were received from the test station. No animal experiment was needed in this project, so no ethical approval is required. The project follows the University of Helsinki research ethical guidelines, good scientific practice, and all applicable laws and regulations.

Data and model availability statement

The data set and software used in this study are not publicly available but may be available from the owner of the data (Figen Oy, Finland) and from the owner of the software (Natural Resources institute Finland, LUKE) on reasonable request.

Author ORCIDs

Alper Tuna Kavlak https://orcid.org/0000-0001-6829-4242.
I. Strandén https://orcid.org/0000-0003-0161-2618.
M. H. Lidauer https://orcid.org/0000-0003-0508-9991.
P. Uimari https://orcid.org/0000-0002-5349-3932.

Author contributions

Alper Tuna Kavlak: Conceptualization, Methodology, Formal analysis, Investigation, Data curation, Software, Writing - original draft. Ismo Strandén: Methodology, Formal analysis, Software, Writing - review & editing. Martin H. Lidauer: Methodology, Software, Writing - review & editing. Pekka Uimari: Supervision, Conceptualization, Methodology, Formal analysis, Data curation, Writing - review & editing.

Declaration of interest

The authors declare no conflict of interest.

Acknowledgements

The authors are grateful to Timo Serenius (Figen Oy), Marja-Liisa Sevón-Aimonen (Luke), and Jarmo Valaja (University of Helsinki) for their opinions as thesis advisory committee members.

Financial support statement

The study was funded by Raisio Oyj Tutkimussäätiö (Finland) and Suomen Sianjalostuksen Säätiö (Finland).

References

Arango, J., Misztal, I., Tsuruta, S., Cubertson, M., Herring, W., 2005. Estimation of variance components including competitive effects of large White growing gilts. Journal of Animal Science 83, 1241–1246.

Ask, B., OF, Christensen, Heidaritabar, M., Madsen, P., Nielsen, H.M., 2020. The predictive ability of indirect genetic models is reduced when culled animals are omitted from the data. Genetics, Selection, Evolution 52, 8.

Bergsma, R., Kanis, E., Knol, E.F., Bijma, P., 2008. The contribution of social effects to heritable variation in finishing traits of domestic pigs (Sus scrofa). Genetics 178, 1559–1570.

Bijma, P., 2010a. Estimating indirect genetic effects: precision of estimates and optimum designs. Genetics 186, 1013–1028.

Bijma, P., 2010b. Multilevel selection 4: modeling the relationship of indirect genetic effects and group size. Genetics 186, 1029–1031.

Bijma, P., Muir, W.M., Van Arendonk, J.A.M., 2007a. Multilevel selection 1: quantitative genetics in inheritance and response to selection. Genetics 175, 277–288.

Bijma, P., Muir, W.M., Ellen, E.D., Wolfand, B.J., Van Arendonk, J.A.M., 2007b. Multilevel selection 2: estimating the genetic parameters determining inheritance and response to selection. Genetics 175, 289–299.

Bouwman, A.C., Bergsma, R., Duijvesteijn, N., Bijma, P., 2010. Maternal and social genetic effects on average daily gain of piglets from birth until weaning. Journal of Animal Science 88, 2883–2892.

Camerlink, I., 2014. Sociable swine: Indirect genetic effects on growth rate and their effect on behaviour and production of pigs in different environments. PhD thesis. Wageningen University, Wageningen, the Netherlands.

Canario, L., Lundeheim, N., Bijma, P., 2010. Pig growth is affected by social genetic effects and social litter effects that depend on group size. Proceedings of the 9th world congress on genetics applied to livestock production, 10 January 2010. German Society of Animal Science, Leipzig, Germany, p. 87.

Casey, D.S., Stern, H.S., Dekkers, J.C.M., 2005. Identification of errors and factors associated with errors in data from electronic swine feeders. Journal of Animal Science 83, 969–982.

Dieters, M.J., White, T.L., Littell, R.C., 1995. Application of approximate variances of variance components and their ratios in genetic tests. Theoretical and Applied Genetics 91, 15–24.

Duijvesteijn, N., Knol, E.F., Bijma, P., 2012. Direct and associative effects for androstenone and genetic correlations with backfat and growth in entire male pigs. Journal of Animal Science 90, 2465–2475.

Ellen, E.D., Rodenburg, T.B., Albers, G.A.A., Bolhuis, J.E., Camerlink, I., Duijvesteijn, N., Bijma, P., 2014. The prospects of selection for social genetic effects to improve welfare and productivity in livestock. Frontiers in Genetics 5, 377–391.

Griffing, B., 1967. Selection in reference to biological groups I. Individual and group selection applied to populations of unordered groups. Australian Journal of Biological Sciences 20, 127–140.

Hadfield, J.D., Wilson, A.J., 2007. Multilevel selection 3: modeling the effects of interacting individuals as a function of group size. Genetics 177, 667–668.

Heidaritabar, M., Bijma, P., Janss, L., Bortoluzzi, C., Nielsen, H.M., Madsen, P., Ask, B., OF, Christensen, 2019. Models with indirect genetic effects depending on group sizes: a simulation study assessing the precision of the estimates of the dilution parameter. Genetics, Selection, Evolution 51, 24.

Herrera-Cáceres, W., Ragab, M., Sánchez, J., 2019. Indirect genetic effects on the relationships between production and feeding behaviour traits in growing Duroc pigs. Animal 14, 233–242.

Kavlak, A.T., Uimari, P., 2019. Estimation of heritability of feeding behaviour traits and their correlation with production traits in Finnish Yorkshire pigs. Journal of Animal Breeding and Genetics 136, 484–494.

Madsen, P., Jensen, J., 2013. A user's guide to DMU. A package for analyzing multivariate mixed models. Version 6, release 5.2. University of Aarhus, Center for Quantitative Genetics and Genomics, Department of Molecular Biology and Genetics, Research Centre Foulum, Tjele, Denmark.

Marchant-Forde, J.N., Marchant-Forde, R.M., 2005. Minimizing inter-pig aggression during mixing. Pig News and Information 26, 63–71.

Moore, A.J., Brodie, E.D., Wolf, J.B., 1997. Interacting phenotypes and the evolutionary process: I. Direct and indirect genetic effects of social interactions. Evolution 51, 1352–1362.

Muir, W.M., Schinkel, A., 2002. Incorporation of competitive effects inbreeding programs to improve productivity and animal wellbeing. 7th world congress on genetics applied to livestock production, 19–23 August 2002, Montpellier, France, Communication 14-07.

Nielsen, H.M., Ask, B., Madsen, P., 2018a. Social genetic effects for growth in pigs differ between boars and gilts. Genetics, Selection, Evolution 50, 4.

Nielsen, H.M., Ask, B., Christensen OF, Janss, L., Heidaritabar, M., Madsen, P., 2018b. Social genetic effects for growth in Landrace pigs with varying group sizes. Proceedings of the 11th world congress on genetics applied to livestock production, 11–16 February 2018, Auckland, New Zealand, Communication 11.212.

Ødegård, J., Olesen, I., 2011. Comparison of testing designs for genetic evaluation of social effects in aquaculture species. Aquaculture 317, 74–78.

Riikimaki, A., 2019. Syöntikäyttäytymisominaisuuksien periytyminen ja yhteydet tuotanto-ominaisuuksiin maatiaisrotuisella sialla, URN:NBN:fi:hulib-201906122715. MSc thesis. University of Helsinki, Helsinki, Finland.

Rydhmer, L., Hansson, M., Lundström, K., Brunius, C., Andersson, K., 2013. Welfare of entire male pigs is improved by socialising piglets and keeping intact groups until slaughter. Animal 7, 1532–1541.

Disease detection in pigs based on feeding behaviour traits using machine learning

A. T. Kavlak[a], M. Pastell[b], and P. Uimari[a]

[a]Department of Agricultural Sciences, University of Helsinki, 00014 Helsinki, Finland

[b]Production Systems, Natural Resources Institute Finland (Luke), 00790 Helsinki, Finland

Abstract

Disease detection is crucial for timely intervention to increase treatment success and reduce negative impacts on pig welfare. The objective of this study was to monitor changes in feeding behaviour patterns to detect pigs that may need medical treatment or extra management. The data included 794,509 observation days related to the feeding behaviour and health information of 10,261 pigs. Feeding behaviour traits were calculated including the number of visits per day (NVD), time spent feeding per day (TPD), and daily feed intake (DFI). The health status (sick or healthy) of pigs were predicted based on the features including the original feeding behaviour traits and features derived from those using a machine-learning algorithm (Xgboost). The predictions were based either on the features from the same day (one-day window), from the same day and two previous days (three-day window), or from the same day and six previous days (seven-day window). The model based on the seven-day window gave the most robust results and achieved an 80% AUC, 7% F1-score, 67% sensitivity, 73% specificity, and 4% precision. The analyses indicated that the features related to the deviation of a pig's observed TPD and DFI from the expected TPD and DFI were the most informative, as they gained the highest importance score. In conclusion, the feeding behaviour-based features gave good sensitivity and specificity in predicting sickness. However, the precision of the method was very low, possibly due to low prevalence of the monitored sickness symptoms, limiting the application of the approach in real life.

Keywords: welfare, disease detection, pigs, machine learning, feeding behaviour

1. Introduction

Pig welfare has gained more and more attention in recent years and should be improved, according to general consensus (Mellor, 2016). Animals express their wellbeing through feeding, drinking, social behaviour etc. Changes in behaviour can be used as early signs of discomfort and sickness (Matthews et al., 2017). In a commercial farm, only limited time is available to observe the individual behavioural changes in pigs, which only permits detecting considerable behavioural changes. This may lead to the too late treatment of the sick animal or intervention in conditions that create discomfort to animals.

Although small changes in daily behaviour are not easy to quantify, data collected automatically from sensors and feeders may include valuable information concerning signs of welfare problems. As an example, increased restlessness among pigs can signal an outbreak of tail biting up to six days prior, which on a commercial scale would be impossible to detect during daily checks (Matthews et al., 2017). In addition, microphones have been used to monitor the sounds of coughing of pigs to build an intelligent alarm system to detect the disease in its early stage (Guarino et al., 2008), 3D-cameras to predict tail biting outbreaks by identifying lowered tail postures (D'Eath et al., 2018), and deviations in typical feeding patterns to monitor overall welfare of pigs (e.g., Brown-Brandl et al., 2013; Munsterhjelm et al., 2015; Bus et al., 2021).

The data collected from the sensors and feeders create challenges to finding the true signals of behavioural changes out of the noise. The complexity of big data with non-linear dependencies and unknown interactions across multiple variables challenges the assumptions of many standard statistical methods (Valletta et al., 2017). Machine learning (ML) methods are highly efficient at determining non-linear relationships between variables in the data (Hastie et al., 2009). As an example, Pandey et al. (2021) collected data on movements, vocal sound, and temperature of pigs using ear sensors and applied ML models to predict the health and welfare status of pigs based on the collected data. Based on their results, the ML approach is a powerful tool for monitoring the health status of pigs leading to reduced medical treatments, cost savings and enhanced animal welfare. Thus, ML methods, such as eXtreme Gradient Boosting, Random Forest, and Support Vector Machine, provide a promising approach for detecting behavioural changes in farm animals that are associated with possible welfare problems (Liakos et al., 2018). Regardless of the method, data quality is important to avoid unwanted outcomes and to gain as robust results as possible. Setting criteria for outliers and applying data filtering prior to applying ML methods to data are therefore important (Alsaaod et al., 2012).

The objective of our study was to develop ML methods based on feeding data to observe changes in feeding behaviour patterns that can be used to detect pigs that may need medical treatment or extra management.

2. Material and Methods

2.1. Feeding behaviour data and pig housing

The feeding behaviour data were provided by Figen Oy (Pietarsaari, Finland) from their central test station, spanning from 2011 to 2016. Pigs arrived at the test station either on a Tuesday or a Wednesday, and the tests began on a Saturday. The pigs were grouped into different pens according to their arrival age (89 ± 10 days), weight (34.4 ± 6.4 kg), and sex (only boars or a combination of gilts and castrates). The average daily gain was 946 ± 113 g/day, and the average slaughter weight and age were 121.2 ± 12.9 kg and 186 ± 10 days, respectively. The average number of piglets in a pen was 9.8 (± 1.19). Water was available ad libitum. Also, feeding type (dry feeding) was ad libitum, consisting of two commercial feedstuffs, and the proportion of the two feedstuffs was based on the growth rate curve of an average pig from the previous test periods. Antibiotics and other drugs were given only for the sick animals based on veterinary prescriptions. The facility had automated ventilation based on pig age and outdoor temperature, and artificial lighting was on from 7 a.m. to 3 p.m. The size of pen was 16.8 m^2 with a concrete floor (2/3 solid, 1/3 slatted). Feedings were recorded automatically using the Schauer Spotmix with Schauer Multilayer Perceptron (MLP) electronic feeders and MLP manager data management software (Schauer Agrotronic GmbH). For further information see Kavlak et al. (2021).

The raw data consisted of 28,826,029 individual feeding visits from 10,261 pigs (Finnish Yorkshire, Finnish Landrace, and F1-crossbred), and included ear tag transponder id, date, time of entering the feeder, time leaving the feeder, and feed intake per visit. The feed intake was measured as a weight of the feed before and after the pig has been in feeder. The number of visits per day (NVD), time spent feeding per day (TPD), and daily feed intake (DFI) were calculated from the recorded observations. Observations from the first testing day were not included due to the DFI exhibiting some of the pigs as extreme outliers, may have been caused by the feeding recording system. Similar extreme DFIs were not observed on a large scale on any other testing days.

2.2. Sickness data

The sickness data were recorded daily by the test station staff members during routine checks (twice a day) and included the ID of the pig, the symptom(s), and the date. The symptoms were classified as a cough, a limp, loss of appetite, skin damage, and a bitten tail. Out of 794,509 daily health observations, 13,018 were related to the recorded symptoms. Within any given day a pig could suffer from several symptoms. In the ML models, pigs with any of the recorded symptoms were classified as “sick” for that given day and pigs with no recorded symptoms were classified as “healthy”.

2.3. Feature processing

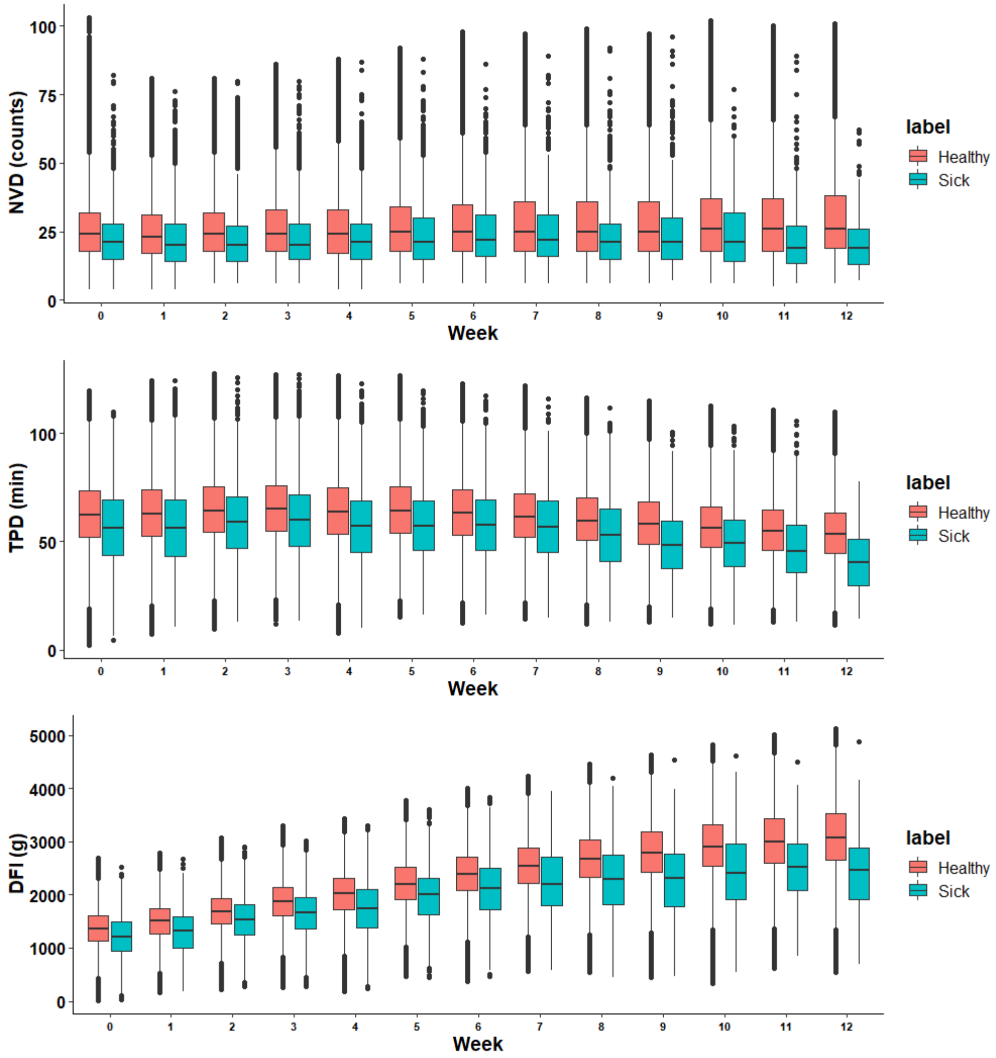
The absolute values of TPD and DFI may not be optimal features for predicting the sickness status of an animal, as they are strongly related to the animal's age. Therefore, we created new features, including daily ranks and residuals. Daily ranks relate the rank of an animal's observation (DRTPD and DRDFI) compared to other pigs within a pen in a given day, and residuals (ResTPD and ResDFI) describe that animal's difference from the expected value of TPD and DFI for a pig of same age. The residuals of TPD and DFI were calculated by fitting a polynomial (quadratic) regression model to the whole data set:

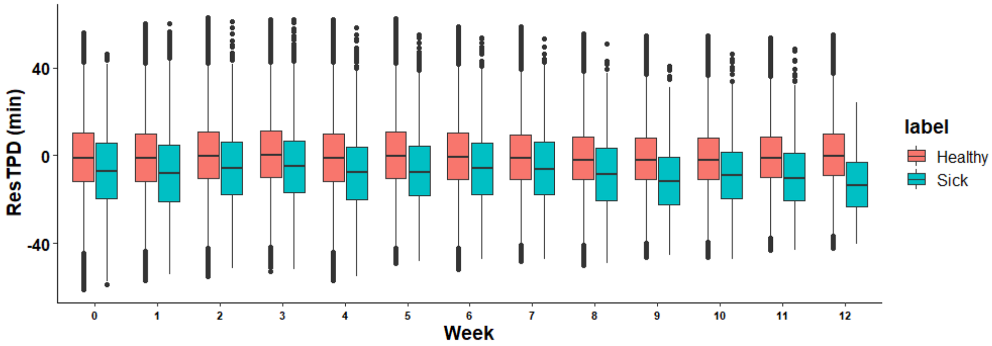
$$y_i = b_0 + b_1 * age_i + b_2 * age_i^2 + e_i \tag{1}$$

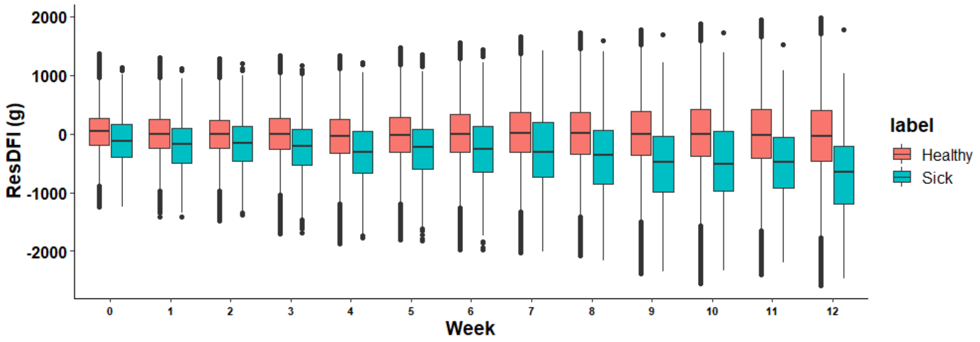
where y_i is either the TPD or DFI of pig i, b_0 is overall mean, $age_i$is the age of pig *i* related to observation y_i, b_1 and b_2 are linear and quadratic regression coefficients, and e_i is the residual used in ML.

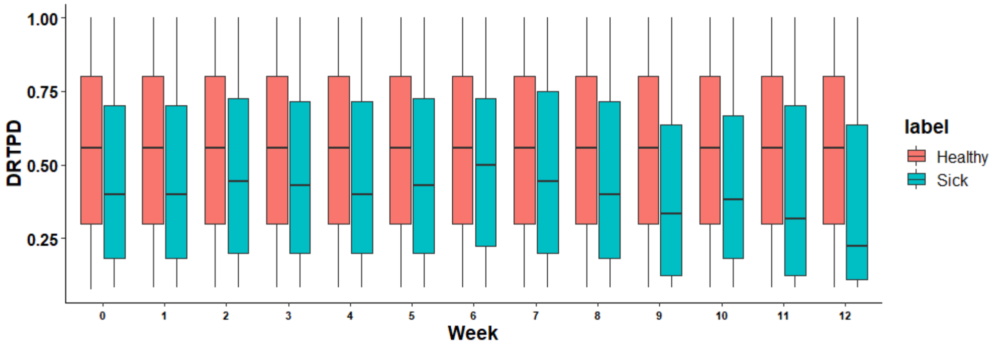
Regarding animal welfare based on feeding behaviour, short-term visits have been considerably challenging to interpret in animal behaviour analyses when conventional methods are used (Young and Lawrence, 1994). The frequency of visits without eating and intervals between visits can be informative feeding patterns that can contribute to predict the health status of animals (Garrido-Izard et al., 2020). Tolkamp et al. (1998) proposed log-normal distribution to model within and between feeding events. In this study, the intervals between feeding visits were calculated as the difference between the time of entering the feeder and the time leaving the feeder during the previous feeding. The intervals were first log transformed and then a mixture of two normal distributions was fitted for the log-transformed intervals of each pig for each day. The first distribution relates to short visits during one meal and second representing intervals between meals. The daily features for each pig from this mixture distributions were the proportion of intervals belonging to the first distribution (p), σ_1 (s1) and σ_2 (s2) the standard deviations, and μ_1 and μ_2 the means of the distributions.

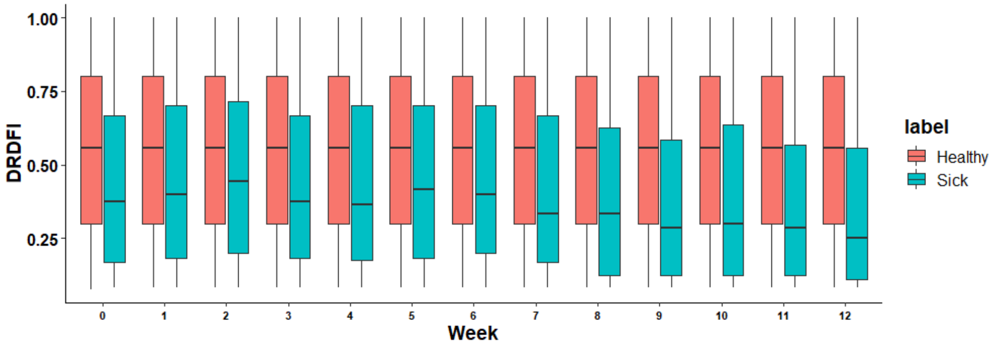
Finally, the features used in ML were NVD, TPD, DFI, DRTPD, DRDFI, ResTPD, ResDFI, $\mu_1, \mu_2, \sigma_1, \sigma_2$ and p. The mean and the distribution of the features in the healthy and sick groups over time (Week) are presented in Fig. 1. Prior to creating the features and ML models, extreme values of the features (outside quantiles 0.5% and 99.5% corresponding to likely registration errors from the feeders) were removed from the data.

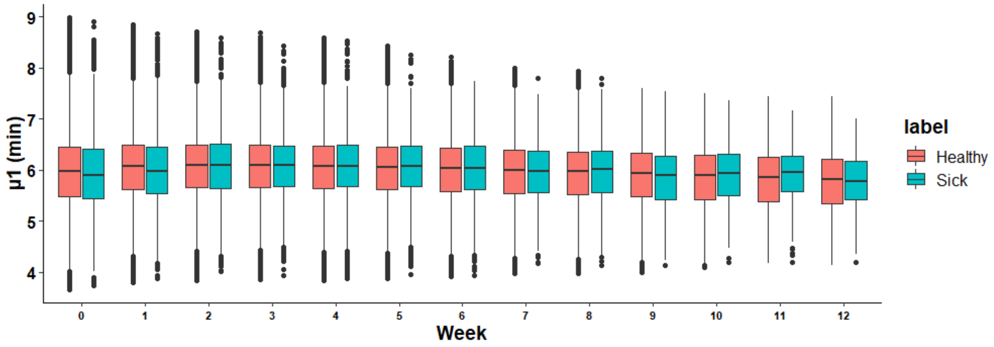


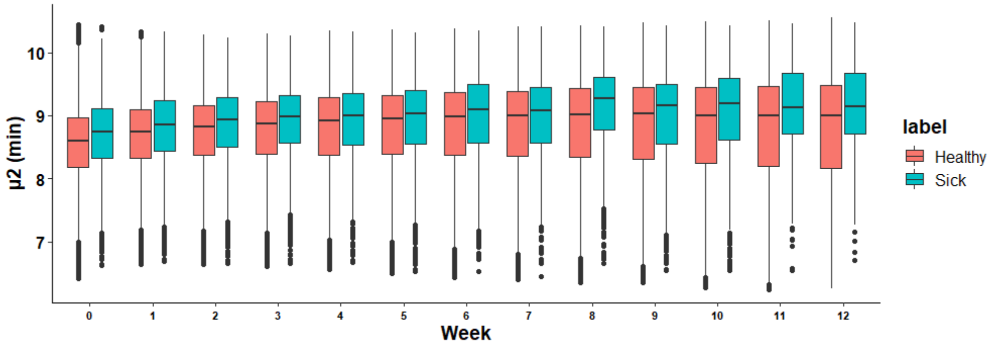

ResTPD (min)
40
0
-40
0
1
2
3
4
5
6
7
8
9
10
11
12
Week
label
Healthy
Sick

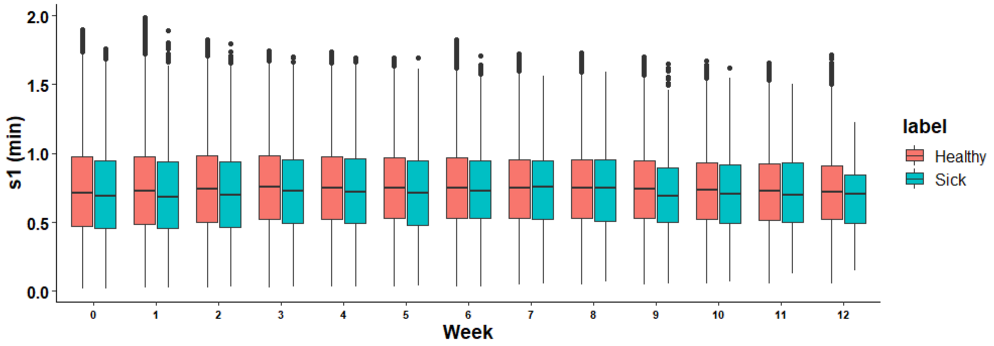

ResDFI (g)
2000
1000
0
-1000
-2000
0
1
2
3
4
5
6
7
8
9
10
11
12
Week
label
Healthy
Sick

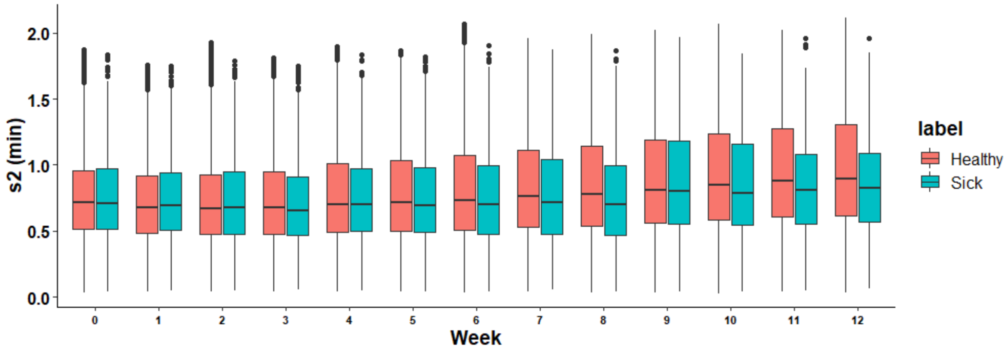

DRTPD
1.00
0.75
0.50
0.25
0
1
2
3
4
5
6
7
8
9
10
11
12
Week
label
Healthy
Sick


DRDFI
1.00
0.75
0.50
0.25
0
1
2
3
4
5
6
7
8
9
10
11
12
Week
label
Healthy
Sick


μ1 (min)
9
8
7
6
5
4
0
1
2
3
4
5
6
7
8
9
10
11
12
Week
label
Healthy
Sick


μ2 (min)
10
9
8
7
0
1
2
3
4
5
6
7
8
9
10
11
12
Week
label
Healthy
Sick


s1 (min)
2.0
1.5
1.0
0.5
0.0
0
1
2
3
4
5
6
7
8
9
10
11
12
Week
label
Healthy
Sick


s2 (min)
2.0
1.5
1.0
0.5
0.0
0
1
2
3
4
5
6
7
8
9
10
11
12
Week
label
Healthy
Sick

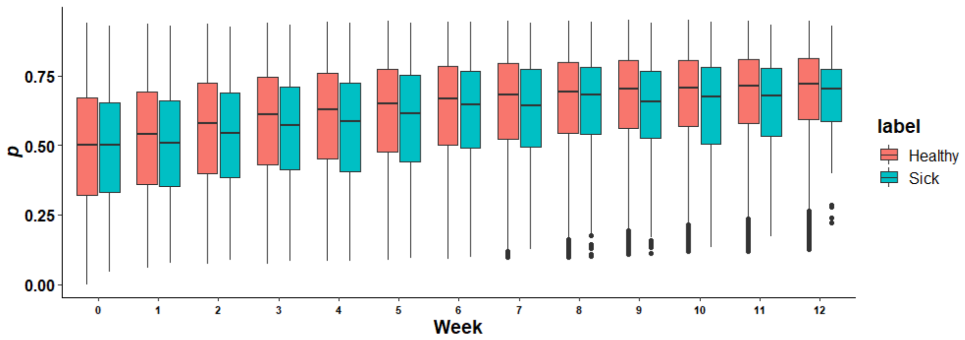


Fig. 1. Boxplots of the features for the testing period (in weeks) grouped by the disease status of the pigs (13,018 daily sick observations and 781,491 daily non-sick observations). The abbreviations of the features are explained in the text.

The health status of a pig was predicted using three different window lengths for determining features: a one-day window, a three-day window, and a seven-day window (number of observations are given in Table 1). In the one-day window approach, the health status of a pig was predicted based on the features from the same day. In the three-day window approach, the health status of a pig was predicted based on the mean of the features from the same day and the previous two days. Similarly, for the seven-day window approach, the health status of a pig was predicted based on the mean of the features from the same day and the previous six days. Based on the three- and seven-day window features, a new features “delta” and “SD” were calculated; delta as a difference between the means of the same feature from the seven- and three-day windows and SD as a standard deviation of the features within the seven-day window. Windows were overlapping.

The number of daily sick and healthy observations are given in Table 1. The number of observations varies between the models because in three- and seven-day window models if any of the daily features within a tree or seven days, respectively, were missing for a given pig, the pig was not included into analysis. In addition, various combinations of symptoms were used;

in Alt-1 -model, “a limp” and “loss of appetite” were treated as “sick”, while the other symptoms (cough, bitten tail, skin damage) were omitted and in Alt-2 -model “bitten tail” and “skin damages” were treated as “sick” (Table 1). For any given pig, on average there were 7.1 consecutive sick days (an average length of the sickness period).

Table 1 Number of symptoms and “sick” and “healthy” observations (pigs x days) for each model.

Model	Limp	Cough	Bitten tail	Loss of appetite	Skin damage	Sick observations	Healthy observations	Total
1-day	6,603	988	2,941	1,015	1,471	13,018	781,491	794,509
3-day	6,377	956	2,888	968	1,357	12,546	761,477	774,023
7-day	5,747	846	2,787	824	1,060	11,264	722,070	733,334
Alt-1[a]	5,747	–	–	824	–	6,571	722,070	728,641
Alt-2[b]	–	–	2,787	–	1,060	3,847	722,070	725,917
Prevalence	0.008	0.001	0.004	0.001	0.002	0.016		

N = Number of observations; Prevalence = Proportion of total symptoms labelled as “sick” out of the total observations in the data based on the 1-day model; [a] In the Alt-1 model, only “a limp” and “loss of appetite” were labelled as “sick” with the seven-day window model and all other symptoms were omitted; [b] In the Alt-2 model, only “a bitten tail” and “skin damage” were labelled as “sick” with the seven-day window model, and all other symptoms were omitted.

2.4. Xgboost algorithm

eXtreme Gradient Boosting (Xgboost) is an ML method similar to Random Forest, decision tree, boosting, gradient boosting, etc. It is an ensemble classifier derived from the gradient boosting decision tree. Xgboost combines weak base classifiers into a strong classifier. At each iteration of the training process, the residual of a base classifier is used in the next classifier to optimize the objective function. In this study, the Xgboost algorithm was applied using the R package Xgboost (Chen et al., 2018) in R 3.6.1 software (R Core Team, 2019).

Hyperparameters are optimization parameters that tune the performance of ML algorithms (Bergstra & Bengio, 2012). In this study, the hyperparameters were chosen using a grid search of the number of boosting iterations (*nrounds*), maximum depth of a tree (*max-depth*), *eta* that controls the learning rate as well as gamma, lambda, subsample. The value grid used for the hyperparameters is given in Table 2 and the final (best) hyperparameters are given in Table 3 based on training data. The objective of the classification model was binary *(binary:logistic)* and the model was fitted by minimizing the binary classification error rate.

Table 2 Range of the values of the hyperparameters.

Hyperparameter	Description	Range of values
nrounds	number of boosting iterations	10–20
max_depth	maximum depth of a tree	3–6
eta	controls the learning rate	0.05–0.5
gamma	controls the minimum reduction in the loss function	0–5
lambda	ridge regularization to prevent overfitting	1.0–2.0
subsample	subsample ratio of the training observations	0.5–1.0

2.5. Performance testing and cross-validation

For the performance of the Xgboost algorithm, the data were split into training and testing data sets. In this study, 70% of the observations were used in training the model and 30% in testing it (Fig. 2). A random sampling of observations was stratified according to the symptoms and pig ID to ensure that the proportion of sick and healthy observations was the same in both data sets and that data from different pigs were used for training and testing the model.

To optimize the hyperparameters and the features and to avoid overfitting the models, we applied 10-fold cross-validation (CV) during model training. The training data set was divided into 10 sets (folds) of equal size. In each validation step, nine of the sub-sets were used for training the model and one sub-set was used for testing the model (Fig. 2). In addition, we used an additional parameter (scale_pos_weight) in the models to control the balance of classes weights due to the imbalanced data set. The parameter was calculated as the proportion of the number of sick observations to number of healthy observations. From each validation step, the area under the ROC curve (AUC) was calculated from the holdout cross-fold (Validation-fold) (Hastie et al., 2009). The set of hyperparameters that gave the best performance metric (AUC) of the model were selected for the final performance testing (Testing data set in Fig. 2).

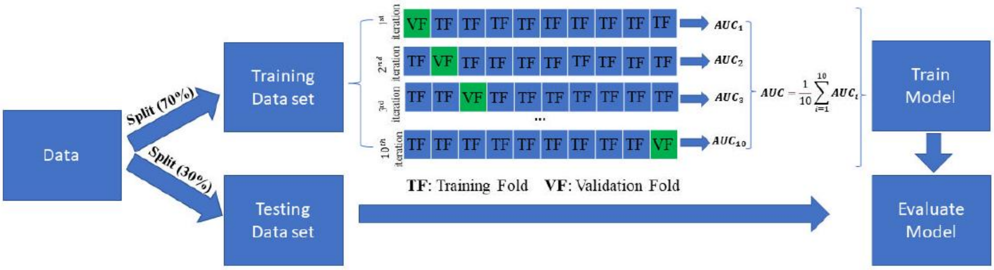


Fig. 2. Overview of 10-fold cross-validation and model testing.

Using the test data set, the models were evaluated based on precision (proportion of predicted true positives (an animal predicted as sick) out of all positive predictions; TP/(TP + FP)), sensitivity (proportion of positives (sick) that were identified correctly; TP/(TP + FN)), and specificity (proportion of negatives (healthy) that were identified correctly; TN/(TN + FP)). In addition, the harmonic means of the precision and sensitivity (F1-score = 2 x precision x sensitivity/(precision + sensitivity)), and AUC (receiver operating characteristics) curve were calculated. The model was considered non-informative with an AUC ≤ 0.50, weak with an AUC of 0.50–0.70, accurate with an AUC of 0.70–0.90, and highly accurate with an AUC ≥ 0.90 (Swets, 1988; Greiner et al., 2000).

We also calculated feature importance using the ability of Xgboost to remove the non-informative or redundant predictors from the model (Chen et al., 2018). While fitting the Xgboost models, an importance matrix was produced from each model. The "gain" metric indicates the relative contribution of the corresponding feature to the model calculated by taking each feature's contribution for each tree in the model. The "cover" metric indicates the relative number of observations related to this feature and the frequency, which is the percentage of the relative number of times a particular feature occurs in the trees of the model. An obtained score of each feature is based on how much more information about the class is gained when using that feature. We quantified the importance of features by "feature gain" (Fig. 3). The steps given above were carried out with the R package caret (Kuhn et al., 2018) in R software (R Core Team, 2019).

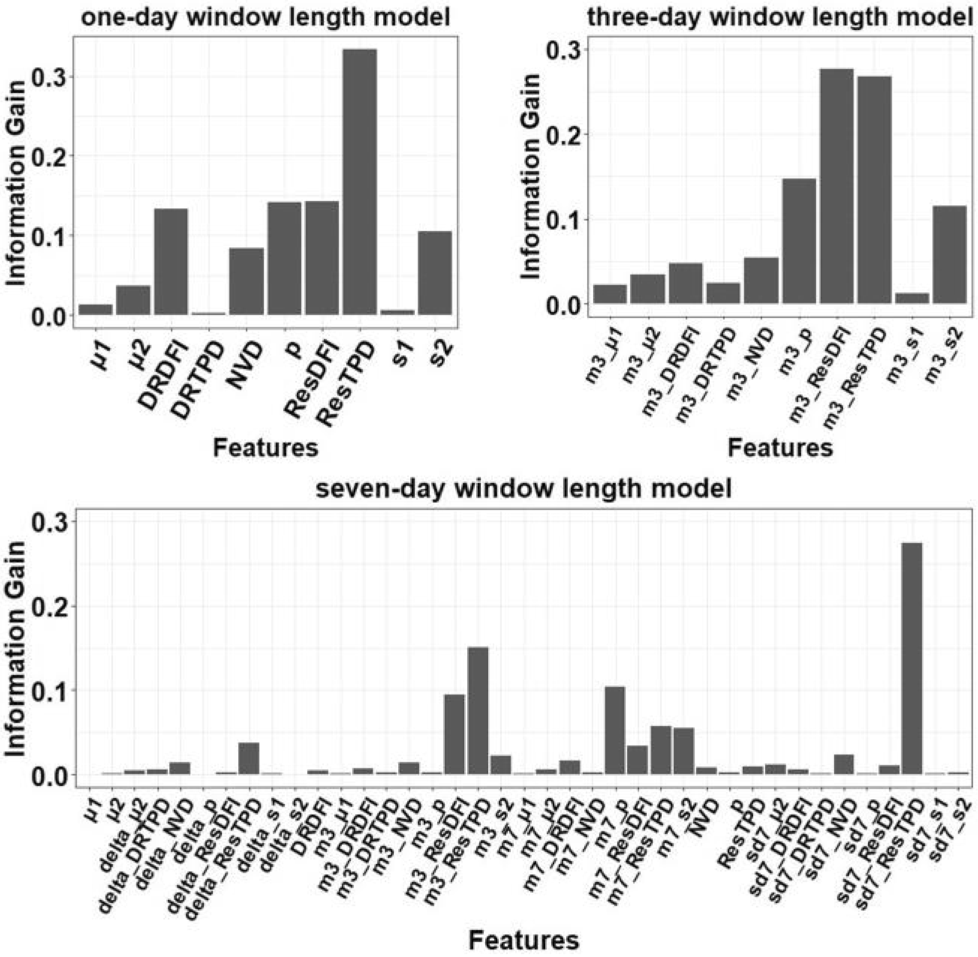


Fig. 3. Importance of the Xgboost features for the different window length. The "Information Gain" implies the relative contribution of the corresponding feature to the model. The abbreviations of the features are explained in the text.

3. Results

3.1. Classification performance of the models

The models were evaluated based on classification performance metrics, including AUC. The best performance according to AUC was obtained with the model applied in the seven-day window (Table 4). In addition, the accuracy differences with the training and testing data sets were small, indicating that over- or under-parametrization of the models was avoided. The best hyperparameters were chosen based on data that provided during training and used in prediction of the models (Table 3).

Table 3 The final values (best) of hyperparameters based on training data

	Window length (day)				
Final Hyperparameters	1	3	7	Alt-1[a]	Alt-2[b]
max_depth	4	4	4	3	4
Eta	0.45	0.45	0.45	0.45	0.45
Gamma	4	5	4	2	3
Lambda	1	2	1.4	2	2
Subsample	1	0.9	0.8	0.9	1

The sensitivity and specificity of the models were acceptable with all window lengths. However, precision and F1-score were quite low. Again, the best performance (67% sensitivity and 73% specificity) was obtained with the model applied in the seven-day window. Unlike the seven-day window model, other models performed at slightly lower efficiency according to the performance metrics. Overall, the results show that by increasing the window length, the performance of the classification models increases.

Alternative labelling of sick animals was tested with two alternative models. For the first alternative model (Alt-1 model), we only labelled “a limp” and “loss of appetite” as “sick” and omitted all other symptoms (cough, bitten tail and skin damage). This model gave 3–4% better performance based on AUC than the performance of the actual seven-day window model (Table 4). On the other hand, the second alternative model (Alt-2 model), where “a bitten tail” and “skin damage” were categorized as “sick” and omitted all other symptoms (cough, limp and loss of appetite), gave a similar performance as labelling all symptoms as “sick” (Table 4). A detailed distribution of assessments for Alt-1 is presented in Table 5, showing that the proportion of animals predicted as being sick was approximately 22%, despite the prevalence based on observed data being around 1%, resulting in low precision.

Table 4 Results from the models based on testing data and training (average from the 10-fold CV) (given in parentheses).

	Window length (day)				
Metrics	1	3	7	Alt-1[a]	Alt-2[b]
AUC	0.70 (0.71)	0.73 (0.75)	0.80 (0.81)	0.83 (0.85)	0.77 (0.80)
Precision	0.03 (0.03)	0.03 (0.03)	0.04 (0.04)	0.03 (0.03)	0.01 (0.01)
Sensitivity	0.60 (0.61)	0.63 (0.65)	0.67 (0.72)	0.67 (0.71)	0.67 (0.74)
Specificity	0.67 (0.67)	0.69 (0.69)	0.73 (0.73)	0.78 (0.81)	0.70 (0.70)
F1-score	0.06 (0.06)	0.06 (0.06)	0.07 (0.08)	0.05 (0.06)	0.02 (0.02)

[a] In the Alt-1 model, only “a limp” and “loss of appetite” were labelled as “sick” with the seven-day window model and all other symptoms were omitted.

[b] In the Alt-2 model, only “a bitten tail” and “skin damage” were categorized as “sick” with the seven-day window model, and all other symptoms were omitted.

Table 5 Tabular visualization of observed versus predicted values from the ALT-1 model based on testing data.

		Observed values		
		Sick	**Healthy**	**Total**
Predicted Values	**Sick**	TP (1,321)	FP (46,682)	48,003
	Healthy	FN (650)	TN (169,939)	170,589

Total	1,971	216,621	218,592

*The observation number of the observed and predicted values are given in parentheses. TP = true positives, TN = true negatives, FP = false positives, FN = false negatives

3.2. The most important features

The most informative features were those related to daily feeding time and daily feed intake: ResTPD and ResDFI in the one-day and three-day window models and SD_ResTPD in the seven-day window model. They alone explained between 20% to 35% of the information gain (Fig. 3). The importance of the other features was less than 10% (Fig. 3). In general, the new features calculated from the NVD, TPD, and DFI were more important in predicting the health status than the absolute values of NVD, TPD, and DFI. As we expected, using the seven-day window model with only the best 10 important features instead of all 40 features increased model performance (based on AUC) slightly (by 1–2%) and reduced the model's run time (results are not shown).

4. Discussion

In this study, the Xgboost algorithm, with features based on feeding station records, was applied to predict the possible sickness of pigs in a test station environment. The considered symptoms were limping, coughing, a bitten tail, loss of appetite, and skin damage, or any combination of these symptoms. In general, the models reached relatively high AUC (0.7 – 0.83). However, model precision was very low (the models predict more sick animals than are reported in the data). Similar to our study, Thomas et al. (2021) predicted diarrhea based on weight dependent water and feed intake using a machine learning approach with seven different methods. Most of the tested methods failed to detect diarrheic pigs due to substantial individual instability on feeding or water related to weight. Even with the best model, 25% of the sick piglets were not detected. Similar to our study, Maselyne et al. (2018) investigated if unusual behavioural changes in the feeding pattern in pigs can be utilised as an indicator of health, welfare, and productivity problems. Although they had considerably high specificity (98.7%) and accuracy (96.7%), sensitivity (58.0%) and precision (71.1%) were lower causing false alerts of health problems and lack of confidence of the farmers for the system. A higher precision (an average of 80%) has been achieved also in some other studies, e.g., in Alsaaod et al. (2012) and Gertz et al. (2020). Gertz et al. (2020) reported very good classification performance (86% AUC, 81% F1-score, 78% specificity, and 81% sensitivity) using the Xgboost algorithm, where locomotion-related diseases were predicted using locomotion data collected from leg and neck sensors in a commercial farm of 397 dairy cows. The health status of cows was monitored by on-farm staff and veterinarians during their daily routine. Based on their findings, using various models with different features and window segments increased model performance and sickness-related behaviours were accurately identified. Moreover, Alsaaod et al. (2012) reported better classification accuracy (76%) for predicting lameness in dairy cows using features created from the pedometric activity and behaviour data on lying down compared to classification accuracy (65%) achieved with the raw data by using the Support vector machine classification model. Thus, in line with our findings, creative new features calculated from the raw data are more informative than the actual sensor data in predicting the sickness of animals.

In our study, pig health was monitored by station staff during the daily routine check. It is possible that only the most severe cases were detected by the staff and some milder ones were missed, and thus the true prevalence of symptoms may be higher than the observed 2% (depending on what symptoms were classified as “sick”) in the data (Table 1). Thus, some of the true negatives (indicated as healthy in the data) could have been sick instead. Higher actual prevalence is supported by Munsterhjelm et al. (2015), where 2,672 pigs in the same test station (Längelmäki, Finland) were monitored in detail for symptoms three to four times daily by farm staff, who were supervised by a herd veterinarian, between November 2007 and December 2008. During that period, the prevalence of tail biting was 13%, 11% for limping, 2% for skin damage, and 6.1% for other symptoms (including diarrhoea, weight loss, vomiting etc.). Another possible explanation for the low precision in our study could be that the classifier did not learn the optimal decision boundary with our highly imbalanced data set despite the weighing we used for the samples from the minority class. Any real dataset may have several imbalanced classes causing biased classification in machine learning. Various techniques have been developed to deal with this problem such as undersampling methods, oversampling methods, ensemble methods etc. that improve the performance of classifiers (Provost, 2000; Japkowicz and Stephen, 2002; He and Garcia, 2009). Although we scaled the class weights according to the prevalence of observations in each class to solve the imbalanced classification problem, we should try out other suggested methods and find the best one in future for our dataset. However, the most effective technique still may vary depending on the dataset.

Unusual behavioural changes in pigs may indicate sickness. These behavioural changes may be rapid and indicate sickness immediately after the behavioural changes have occurred or the changes may begin several days prior to sickness. Therefore, we applied models with different window lengths. We found a clear tendency that considering records from several previous days instead of a single day was beneficial (AUC increased from 0.70 to 80). Gertz et al. (2020) also reported that using various window lengths allows the classifier to select the amount of data leading to the best prediction performance. However, the Xgboost preferably selected shorter window lengths in their study compared to our study. Thus, it is always a good practise to test several window lengths because method performance depends on the features and nature of the data, and the long window may not always be optimal. Also, in other behavioural studies (e.g., Smith et al., 2016; Piette et al., 2020; Riaboff et al., 2020), the sliding window length approach has had a positive impact on algorithm performances.

Selecting the optimal hyperparameters is important for successful model performance, as the ML methods have a high risk of under/overfitting the training data. However, there is no optimal way to tune the hyperparameters. In our study, the hyperparameters were tuned using the grid search method (Bergstra & Bengio, 2012) with 10-fold CV, and the best hyperparameters were selected for further analyses. Thus, even though tuning the hyperparameters requires extra computing time, obtaining good prediction performance is recommended.

Finally, the set of features available for prediction is crucial for improving the performance of the classification. In our study, the features were calculated from the feeding behaviour data with short and long window segmentations. The most important features were ResTPD and ResDFI along with SD_ResTPD in the seven-day window length model (Fig. 3), which indicate that using the residuals of feeding behaviour traits is more beneficial in predicting pig sickness than absolute values. Thus, a deviation from a typical daily feeding time or daily feed intake compared to the feeding time and daily feed intake of an average pig at the same age is a good indication of a possible health problem. Similarly, daily ranks of TPD and DFI were informative and are easier to calculate than the residuals of TPD and DFI. Hoy et al. (2012) also suggested that daily ranks based on feeding must be classified because many pigs have access to one feeding place in a pen. Therefore, we propose using features that indicate a deviation of an animal's feeding behaviour from its pen mates (rank) or from pigs of the same age (residual) rather than raw observations (NVD, TPD, DFI). Furthermore, solely using the most important features in the model instead of all available features improved algorithm performance slightly (1–2%).

5. Conclusions

Based on the performance metrics (AUC, sensitivity, and specificity), pig sickness is detectable by applying the Xgboost algorithm to the feeding behaviour data. However, very low precisions were obtained, possibly due to imbalanced data. Using the observations from several days (seven days) gave more accurate predictions than predictions based on a single day, even though the results did not differ considerably. When the prediction was based on one- or three-day observations (one- and three-day windows), the most important features were ResTPD and ResDFI. We examined a vast, but limited set of features, and our results can be improved by calculating new features, considering interactions between features, using different window length(s), different methods etc. From a practical standpoint, high sensitivity is more important than high precision because the final assessment of an animal's sickness would be based on a re-check by the management staff if the applied algorithm suggests that the animal may be sick. The cost of re-checking additional animals should be smaller than treating a sick animal that was not detected early enough. Despite this, the precision should be far higher than what was achieved here to gain trust in users of the algorithm on a routine basis. Features derived from other automatic data recording systems, such as locomotion sensors, could improve the predictive performance of the method. This would require more research.

CRediT authorship contribution statement

Alper Tuna Kavlak: Conceptualization, Methodology, Formal analysis, Investigation, Data curation, Software, Writing- Original draft. **Matti Pastell**: Methodology, Formal analysis, Software, Data curation, Writing- Reviewing and Editing. **Pekka Uimari**: Supervision, Conceptualization, Methodology, Formal analysis, Data curation, Writing- Reviewing and Editing.

Declaration of competing interest

The authors declare no conflict of interest.

Acknowledgement

The study was funded by Raisio Oyj Tutkimussäätiö (Finland) and Suomen Sianjalostuksen Säätiö (Finland). The authors are grateful to Natural Resources Institute Finland (Luke) for collaborating in this project and to Timo Serenius (Figen Oy), Marja-Liisa Sevón-Aimonen (Luke), and Jarmo Valaja (University of Helsinki) for their opinions as thesis advisory committee members.

References

Alsaaod, M., Römer, C., Kleinmanns, J., Hendriksen, K., Rose-Meierhöfer, S., Plümer, L., & Büscher, W. (2012). Electronic detection of lameness in dairy cows through measuring pedometric activity and lying behavior. *Applied Animal Behaviour Science*, 142, 134–141.

Bergstra, J., & Bengio, Y. (2012). Random search for hyper-parameter optimization. *Journal of Machine Learning Research*, 13, 281–305.

Brown-Brandl, T., Rohrer, G.A., & Eigenberg, R.A. (2013). Analysis of feeding behavior of group housed growing–finishing pigs. *Computers and Electronics in Agriculture,* 96, 246–252.

Bus, J.D., Boumans, I.J.M.M., Webb, L.E., & Bokkers, E.A.M. (2021). The potential of feeding patterns to assess generic welfare in growing-finishing pigs, *Applied Animal Behaviour Science,* 241, 105383.

Chen, T., He, T., Benesty, M., Khotilovich, V., Tang, Y., Cho, H., Chen, K., Mitchell, R., Cano, I., Zhou, T., Li, M., Xie, J., Lin, M., Geng, Y., & Li, Y. (2018). xgboost: Extreme Gradient Boosting.

D'Eath, R.B., Jack, M., Futro, A., Talbot, D., Zhu, Q., Barclay, D., & Baxter, E.M. (2018). Automatic early warning of tail biting in pigs: 3D cameras can detect lowered tail posture before an outbreak. *PLoS One*, 13, 18.

Garrido-Izard, M., Correa, E.C., Requejo, J.M., & Diezma, B. (2020). Continuous monitoring of pigs in fattening using a multi-sensor system: behavior patterns. *Animals,* 10, 17.

Gertz, M., Große-Butenuth, K., Junge, W., Maassen-Francke, B., Renner C., Sparenberg, H., & Krieter, J. (2020). Using the XGBoost algorithm to classify neck and leg activity sensor data using on-farm health recordings for locomotor-associated diseases. *Computers and Electronics in Agriculture,* 173.

Greiner, M., Pfeiffer, D., & Smith, R. D. (2000). Principles and practical application of the receiver-operating characteristic analysis for diagnostic tests. *Preventive Veterinary Medicine*, 45, 23–41.

Guarino, M., Jans, P., Costa, A., Aerts, J.M., & Berckmans, D. (2008). Field test of algorithm for automatic cough detection in pig houses. *Computers and Electronics in Agriculture*, 62, 22–28.

Hastie, T., Tibshirani, R., & Friedman, J. H. (2009). The Elements of Statistical Learning: Data Mining, Inference, and Prediction. Springer, New York, NY.

He, H., & Garcia, E. A. (2009). Learning from Imbalanced Data. *Knowledge and Data Engineering, IEEE,* 21(9), 1263–1284.

Hoy, S., Schamun, S., & Weirich, C. (2012). Investigations on feed intake and social behaviour of fattening pigs fed at an electronic feeding station, *Applied Animal Behaviour Science*, 139, 58-64.

Japkowicz, N., & Stephen, S. (2002). The class imbalance problem: A systematic study. *Intelligent Data Analysis*, 6(5), 429–449.

Kavlak, A. T., Strandén, I., Lidauer, M. H., & Uimari, P. (2021). Estimation of social genetic effects on feeding behaviour and production traits in pigs. *Animal*, 15(3), 100168.

Kuhn, M., Wing, J., Weston, S., Williams, A., Keefer, C., Engelhardt, A., Cooper, T., Mayer, Z., Kenkel, B., Benesty, M., Lescarbeau, R., Ziem, A., Scrucca, L., Tang, Y., Candan, C., & Hunt, T. (2018). caret: Classification and Regression Training.

Liakos, K. G., Busato, P., Moshou, D., Pearson, S., & Bochtis, D. (2018). Machine learning in agriculture: A review. *Sensors*, 18, 2674.

Maselyne, J., Van Nuffel, A., Briene, P., Vangeyte, J., De Ketelaere, B., Millet, S., Van den Hof, J., Maes, D., & Saeys, W. (2018). Online warning systems for individual fattening pigs based on their feeding pattern. *Biosystems Engineering,* 173, 143–156.

Matthews, S. G., Miller, A. L., Plötz, T., & Kyriazakis, I. (2017). Automated tracking to measure behavioural changes in pigs for health and welfare monitoring. *Scientific Report*. 7(1), 17582.

Mellor, D. J. (2016). Updating Animal Welfare Thinking: Moving beyond the "Five Freedoms" towards "A Life Worth Living". *Animals, 6*, 21.

Munsterhjelm, C., Heinonen, M., & Valros, A. (2015). Effects of clinical lameness and tail biting lesions on voluntary feed intake in growing pigs. *Livestock Science*, 181, 210–219.

Pandey, S., Kalwa, U., Kong, T., Guo, B., Gauger, P. C., Peters, D. J., & Yoon, K. J. (2021). Behavioral Monitoring Tool for Pig Farmers: Ear Tag Sensors, Machine Intelligence, and Technology Adoption Roadmap. *Animals*, 11(9), 2665.

Piette, D., Norton, T., Exadaktylos, V., & Berckmans, D. (2020). Individualised automated lameness detection in dairy cows and the impact of historical window length on algorithm performance. *Animal*, 14, 409–417.

Provost, F. (2000). Machine Learning from Imbalanced Data Sets 101. Workshop on Learning from Imbalanced Data Sets, Texas, US: AAAI.

R Core Team (2019). R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria.

Riaboff, L., Poggi, S., Madouasse, A., Couvreur, S., Aubin, S., Bedere, N., Goumand, E., Chauvin, A., & Plantier, G. (2020). Development of a methodological framework for a robust prediction of the main behaviours of dairy cows using a combination of machine learning algorithms on accelerometer data. *Computers and Electronics in Agriculture*, 169, 105179.

Smith, D., Rahman, A., Bishop-Hurley, G. J., Hills, J., Shahriar, S., Henry, D., & Rawnsley, R. (2016). Behavior classification of cows fitted with motion collars: Decomposing multi-class classification into a set of binary problems. *Computers and Electronics in Agriculture*, 131, 40–50.

Swets, J. A. (1988). Measuring the accuracy of diagnostic systems. *Science,* 240, 1285–1293.

Thomas, J. Rousselière, Y., Marcon, M., & Hémonic, A. (2021). Early Detection of Diarrhea in Weaned Piglets From Individual Feed, Water and Weighing Data. *Frontiers in Animal Science,* 2, 2673-6225.

Tolkamp, B. J., Allcroft, D. J., Austin, E. J., Nielsen, B. L., & Kyriazakis, I. (2016). Satiety Splits Feeding Behaviour into Bouts. *Journal of Theoretical Biology*, 194, 235–250.

Valletta, J. J., Torney, C., Kings, M., Thornton, A., & Madden, J. (2017). Applications of machine learning in animal behaviour studies. *Animal Behaviour,* 124, 203–220.

Young, R.J., & Lawrence, A.B. (1994). Feeding behaviour of pigs in groups monitored by a computerized feeding system. *Animal Science*, 58, 145–152.

ISBN 978-951-51-8220-3 (PRINT)
ISBN 978-951-51-8221-0 (ONLINE)
ISSN 2342-5423 (PRINT)
ISSN 2342-5431 (ONLINE)
http://ethesis.helsinki.fi

HELSINKI 2022