Livestock Science 261 (2022) 104961

Contents lists available at ScienceDirect

Livestock Science

journal homepage: www.elsevier.com/locate/livsci

Data-driven decision making in pig farming: A review of the literature

Thomas van Klompenburg, Ayalew Kassahun *

Information Technology Group, Wageningen University & Research, Wageningen, the Netherlands

HIGHLIGHTS

- Importance in data analysis in the pork sector increased steadily since 2006.
- Use of machine learning in the pork sector started mainly in 2018 and increased rapidly.
- Main applications of machine learning are related to diseases, DNA analysis and feeding strategies.
- Most data analysis studies are done on limited number of animals and on experimental basis.
- Future research should focus on the use of real-time data in a realistic business context.

ARTICLE INFO

Keywords:
Decision support
Data analysis
Data processing
Machine learning
Pig farming

ABSTRACT

Applications of data analytics, and recently machine learning, in pig farming have been investigated in literature and the results indicate great potential for data-driven decision support at various scales of the sector—from farm to the management of entire supply chains. However, there is insufficient overview of the studies conducted so far. Particularly, there is little insight into the extent of studies conducted in the context of actual business cases. In this study we conducted a systematic literature review to shed light on the state-of-the-art knowledge about data-driven decision making in the pig sector. In order to cover both classical data analysis techniques and machine learning, we used two separate search strings to search the literature. The results show that the various attributes of live pigs and slaughter data are used in analytics. Most studies focus on the occurrence and prevention of diseases, followed by DNA-related analysis and the effect of feeding strategies on growth. Among the studies we analysed, there was a large variation in herd size under study. Most studies used a selected group of pigs in an experimental environment; fewer studies used a larger number of pigs. Notably, all studies except two focussed on real-life business contexts where real-time data is used. The application of machine learning, mainly the use of *random forest* and *neural network* algorithms, took off since 2018. Current studies focus on isolated and one-off problems, and we suggest future research to consider the complexity encountered in real-life business circumstances and routine decision making through the integration of data analytics within farm information management systems.

1. Introduction

Companies in the livestock sector are increasingly investigating the potential of data-driven decision making for improving production (Koketsu and Iida, 2020). In the pork sector, the growing demand for pork meat and the outbreak of diseases that threatens pig farming have forced pig farmers to implement precision farming practices. The rapid implementation of precision farming following the outbreak of African swine fever and the subsequent shortage of pork presents an illustrative example of the value of data-driven decision making (Wang, 2020).

Data-driven decision making at pig farms means that decisions that are made will depend on predictions made using the information gathered at the farm and across the supply chain. To perform successful predictions and help in decision making, data analytics and machine learning (ML) techniques can be used. Recently, ML models are being used to predict various variables of interest to decision making, such as sales and feed performance (Neethirajan, 2020). Data gathering and analytics are thus becoming crucial for gaining a competitive advantage in the agricultural sector in general, and in the pork sector in particular. The availability and accessibility of structured data are crucial for the

* Corresponding author.
E-mail address: ayalew.kassahun@wur.nl (A. Kassahun).

https://doi.org/10.1016/j.livsci.2022.104961
Received 28 April 2021; Received in revised form 18 March 2022; Accepted 9 May 2022
Available online 10 May 2022

Table 1
Algorithms used for disease detection at a livestock farm (Neethirajan, 2020).

Disease	Algorithm	Parameter detected
Mastitis	Bag of Words (BoW), Gradient Boosted Trees (GBT)	Somatic cell Count (SSC), Electrical Conductivity (EC)
Lameness	Fog computing, Classification and regressive tree (CART) XGBoost algorithm	Leg movement, Neck movement and Image/Video data
Postpartum disease	Random Forest Algorithm (RFA)	Lactose yield, Protein production, Milk yield
Coccidiosis	Principal Component Analysis (PCA)	Volatile Organic Compounds (VOC) in air
African Swine Flu	Optical flow algorithm	Mobility, speed, direction

Table 2
Algorithms used to help farmers monitor feed intake vs feed efficiency (Neethirajan, 2020).

Algorithms	Data collection method
Top-down induction of decision trees (TDIDT) algorithm	Weight and concentration of food, drones and manual entries
Random Forest Algorithm elastic net (ENET) and nearest shrunken centroid algorithm	Metabolic rate, Gene expression, Average daily weight gain, and Average Back fat gain
Sing Shot multi-box Detector (SSD) algorithm	Body Condition Score (BCS) through the multicamera system
Auto-Regressive Moving Average model (ARIMA)	Feeding weight of dry and concentrate and milk production by each cow
Convolutional Neural Networks (CNNs)	RGB camera, RFID for cow feed intake and milk production measurement and frequency

Table 3
ML use in livestock management, derived from (Sharma et al., 2020).

Algorithm	Objective
ANN	Supply chain management (Vlontzos and Pardalos, 2017)
Bayesian network	Animal welfare tracking and production management (Liakos et al., 2018)
Clustering	Supply chain management (Vlontzos and Pardalos, 2017)
Deep learning	Feeding and marketing decisions of growing pigs (Pourmoayed et al., 2016)
Decision tree	Supply chain management (Vlontzos and Pardalos, 2017)
Instance-based learning	Supply chain management (Vlontzos and Pardalos, 2017)
Neural networks	Pig cough recognition (Mucherino et al., 2009)
Regression	Automatically and continuously monitoring animal health (Yazdanbakhsh et al., 2017)

performance of ML models (Lee and Shin, 2020). It is therefore important to investigate the entire pipeline of data analytics in order to apply improved data analytics in practice (Pääkkönen and Pakkala, 2015).

There have been only a few systematic reviews of the literature with respect to data analytics in the pork sector. Koketsu and Iida (2020) reviewed the literature on the data analysis used for the performance of sows and predictors in breeding herds. The review defined four components of sow lifetime performance, namely lifetime efficiency, sow longevity, fertility, and prolificacy. Further, they proposed two different lifetime performance trees, one for piglets weaned and one for pigs born alive, which have relationships with the performance components. They also described predictors for high lifetime performance. Nguyen Thi Thuy et al. (2020) did a review on the pork value chain in Vietnam. They concluded that the presence of middlemen and traders is important for facilitating the nationwide distribution of pigs and pork, but they also observed that, in general, no official contracts are used in transactions in the current supply chain, and as a result, it is hard to trace the origin of pork in Vietnam. Newer pork value chains are developing in Vietnam but currently represent only a fraction of the whole pork supply chain.

The purpose of this study is to perform a Systematic Literature Review (SLR) and find out what research has been done on the application of data analytics, data-driven management, and data-driven decision support in the pork sector and identify the associated data used. To achieve these goals we defined five research questions:

RQ1: What data are used for data analytics in the pork sector?

RQ1A: To what extent is the available data adequate and of good quality?

RQ1B: To what extent is the available data accessible?

RQ2: From which information systems are the data over pigs and pork retrieved?

RQ3: What data can be retrieved from these information systems?

RQ3A: What are the key requirements for data gathering?

RQ4: What are the steps of data analytics pipeline and data management in relation to data analytics for animal nutrition, which is the biggest cost factor in pig production?

RQ5: What kind of data analytics can be done to support data-driven management to support operational, tactical, and strategic decisions?

By performing this SLR, we aim to present the state-of-the-art with regard to data analytics in the pork sector.

2. Background

Animal-based products have a large and increasing contribution to the human food supply (Bradford, 1999). The increasing demand is driven by the increase of both the world population and the wealth of households. Optimizing animal-based products requires optimizing the feed intake of animals, reducing the risk of diseases, and improving the animals' living environment and their general welfare state. Optimizing animal-based production systems requires, therefore, coordination among farmers, feed companies, and other actors. The ultimate aim of optimized animal-based production is to attend to the precise needs of the individual animal, which is called precision livestock farming (PLF). The transition to PLF is achieved when extensive data on animal growth, production outputs, disease development, animal behaviour, and living environment are gathered and analysed for use in data-driven decision making (Bos et al., 2018). PLF is part of the general trend of precision farming which is practiced in diverse disciplines of agriculture, such as arable farming (van Klompenburg et al., 2020).

2.1. Machine learning

Machine learning (ML), which includes the application of deep learning and neural networks, is a branch of artificial intelligence (AI). AI is a key element of the fourth industrial revolution, and in agriculture this revolution is referred to as Agriculture 4.0. Different kinds of agricultural technology are being used for many years, but the introduction of IoT, cloud computing, robotics, and AI could change the way of farming significantly (Rose and Chilvers, 2018).

A variety of research has been done on the use of ML in the agricultural sector, including in the livestock sector. For example, cows' body weight was predicted using a support vector machine classification model, and several ML algorithms have been used in disease detection (Neethirajan, 2020). Table 1, derived from Neethirajan (2020), shows which diseases have been detected using ML.

Machine learning is also being used in optimizing feed efficiency and energy intake. Optimizing feed efficiency is crucial because feed is the biggest cost factor in animal production (Van der Meulen, 2020). Table 2 shows what algorithms have been used and how data was collected for optimizing feed consumption.

Besides feed efficiency and disease detection, ML has been used in the livestock sector for behaviour tracking, health monitoring, and the management of the supply chain (Sharma et al., 2020) as shown in Table 3.

When looking specifically at the research on pigs, most attention has been given to the growth of pigs. Much research is done to find patterns in (feeding) behaviour, the environment, and weight gains. For example, patterns have been found that the more pigs contract Pneumonic Pasteurellosis, a disease that occurs throughout the world in pig farming, the less will be their daily weight gain (Hill et al., 1992).

Researchers at Wageningen University are currently working on the 'Feed-a-Gene' study, focusing on the improvement of feed conversion. Their latest findings showed that when pigs get adjusted nutrients per pig, the impact on the environment will be lower. An alternative feed system was developed and pigs were fed based on their individual characteristics (Wageningen University, 2020a).

Besides the livestock sector, ML has multiple other applications in the agricultural sector. ML has been used in arable farming for a variety of prediction problems. For example, crop yield prediction is done using ML and data gathered over the soil conditions, temperature, and rainfall (van Klompenburg et al., 2020). Precision farming is applied successfully in the arable farming sector (Wageningen University, 2020b).

2.2. Information systems used in the meat supply chain

Information systems are used in the meat supply chain sector for several reasons. Farmers use farm management information systems to manage their herds. Bigger companies, such as feed producers, slaughterhouses, and meat processors use enterprise systems. Enterprise systems are often complex systems and are thus modular and configurable.

The aspects of information systems relevant for this research are the use of data for decision support and data integration across the supply chain. A supply chain of a product is a set of coordinated entities that are involved in the realization of the product (Tekinerdogan, 2019). When looking at the supply chain from the perspective of one of the actors (for instance, a farmer), upstream and downstream flows can be defined. The upstream supply chain is focused on supplying the needs for production, and the downstream chain is focused on the flow of products to its end consumers.

2.3. Integration of ML and data analysis in meat supply chain information systems

With the implementation of more sensors and data sources, more possibilities arise for using ML in information systems. Manufacturing companies are often focused on static explanatory models, but there are many opportunities in predictive modelling where ML can be used and improve the decision making processes (Flath and Stein, 2018). For example, ML is used to predict limb conditions of pigs, using data collected at farm (Bakoev et al., 2020). Liang et al. (2020) predicted African swine fever (ASF) outbreaks, using ASF outbreak data and meteorological data. Xu et al. (2013) proposed a pork traceability framework, based on IoT monitoring and data mining techniques.

2.4. Data need for improved data analysis and machine learning

A prerequisite for applying improved data analytics, such as ML, is the availability of large and high quality dataset (Sessions and Valtorta, 2006). Businesses gather data for various purposes but not all of it is good enough for use in ML. The data often has inaccuracies and inconsistencies, also known as noise (Bose and Mahapatra, 2001). When data is collected from various databases in the supply chain, the data from different sources have to conform to a common standard. Otherwise, combining the data from the different sources will be difficult or even impossible. When certain data are gathered at one place (such as at a slaughterhouse) but the related data are not gathered at the other places (such as at farms), the desired objective of data analytics may not be achieved.

In the last decades, data were mostly used for the management of farm tasks, and to a limited extent for further analytics (Piñeiro et al., 2019). The variables that are routinely measured were batch data on growth (body weight), feed intake, and mortality. Feed efficiency and performance were determined using these data. Environmental data and data from slaughterhouses were not widely available or accessible and thus were not used widely.

Feeding costs have a significant share in the total production costs for animal production (Van der Meulen, 2020). When the feeding process could be optimized, the financial impact could be significant. When the collection and analysis of useful information is made available to farmers and integrated within their farm information management systems, productivity could become higher (Banhazi et al., 2012).

Technologies that identify the individual animal and collected data accordingly can help farmers to monitor individual animals and improve the welfare of the animals. When data is visualized in meaningful and easy-to-understand ways, social acceptance of the technologies could become higher (Van Hertem et al., 2017). Since feed has the largest share in the variable production costs in livestock farming, optimization of the feed use could result in more sustainable and controlled production.

2.5. Data preparation challenges

The available raw data can't be used directly within ML models most of the time. Data needs preparation to create a usable dataset for ML. Data preparation is time consuming; it is the most time-intensive task in data analytics with ML. When data are gathered from a variety of sources, they are often incompatible and there is often no alignment in the features within the datasets, which makes data preparation a necessity (Garner et al., 1999).

2.6. Use of machine learning in businesses

Machine learning is used in diverse business processes in a variety of companies. For example, data mining with ML algorithms is done in finance, telecom, marketing, and web analytics firms. The application of ML has produced useful results when used for the prediction of various variables but there are also many aspects that can be improved. There is a trade-off between the reward of using an ML application and the investment that comes with it, while successful prediction cannot always be guaranteed. In the finance sector, for instance, ML is more widely used to reduce uncertainty when making decisions but the accuracy of the predictions depends on a number of factors and the results should be used cautiously (Bose and Mahapatra, 2001).

3. Methodology

An SLR is an important research activity, and its results should be objective, replicable, and transparent. The SLR of this research is conducted following a well-known guideline of Kitchenham et al. (2007). The steps taken in this research are clearly described, and all search results along with the review data have been included.

3.1. Review protocol

Kitchenham et al. (2007) define the review process in three phases: planning the review, conducting the review, and reporting the interview.

In the first step of the planning phase, we identified the research questions which are presented in the introduction section. Then, a search protocol was defined, stating exclusion criteria, and selecting databases that will be searched. Subsequently, the initial search strings were identified. The search strings were iteratively improved in such a way that they captured all the relevant publications.

Table 4
Studies found with the broad search query, before and after exclusion criteria.

Database	Found	After EC1	After EC 2–7
Science Direct	118	50	12
Scopus	587	50	17
Web of Science	29	29	13
Springer Link	177,351	50	6
Wiley	908	50	2
Total	178,993	229	50

Table 5
Studies found with the narrow search query, before and after exclusion criteria.

Database	Found	After EC1	After EC 2–7
Science Direct	25	25	9
Scopus	90	50	10
Web of Science	69	50	4
Springer Link	858	50	3
Wiley	40	40	0
Total	1082	215	26

After finishing the planning phase, the review was conducted by searching the databases using search strings that are adjusted according to the options the databases offer. When the search in a particular database returned more than 50 publications, only the first 50 most relevant publications were selected. The publication details were saved, including title, author names, type of publication, and publication year, merged, duplicates removed, and filtered using exclusion criteria. After the selection of relevant publications, the selected publications have been analysed, also known as primary studies.

After conducting the review, the last step in the process was executed. The results of the review have extensively been reported using tables and used to answer the research questions.

3.2. Search strategy

The search is performed in five well-known scientific databases: Science Direct, Scopus, Web of Science, Springer Link, and Wiley. The search was done by choosing a broad search string and incrementally narrowing down the search. We started by searching for "Data use in agriculture" in Science Direct, which returned 23.217 results which shows this search query is too broad. Therefore, we limited the search string to make it focus on the livestock sector and the specific issue that the research questions addressed. However, when a search query was narrowed further no results were returned. Finally, we selected a broad enough search string and selected the 50 most relevant publications using the relevance order offered by the database. This was done to exclude less relevant publications; also, publications that focus on the end-of-chain consumer side were excluded since most of them focus on the storage of meat and sales related publications. When using a feasible search query, the abstract of the study was read to see if the source was of relevance. If this was the case, the complete study was read and the study was analysed based on the exclusion criteria. Finally, the data that

Table 6
Topics of research in primary studies, broad search query.

Topic of research	# of studies
Disease	9
DNA analysis	9
Feeding strategy	6
Fat composition	5
Pig performance	4
Meat quality	4
Boar taint	3
Meat origin	2
Muscle	2
Other	6

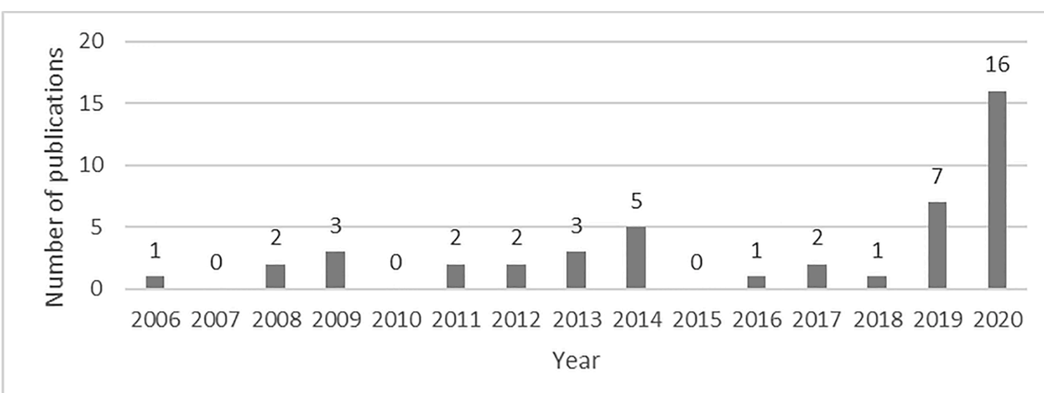

Fig. 1. Distribution of primary studies found through the broad search query per publication year.

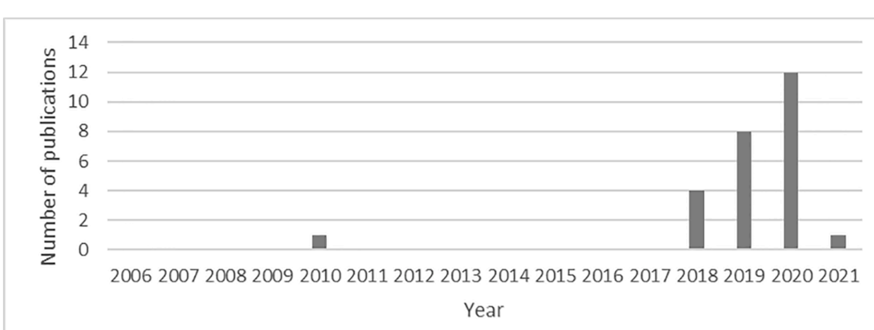

Fig. 2. Distribution of primary studies fund through the narrow search query per publication year.

Table 7
Topics of research in primary studies, narrow search query.

Topic of research	# of studies
Pig performance	6
Disease	5
Feeding strategy	3
Pig detection	3
Behaviour analysis	3
Origin detection	3
Gene	2
Other	2

Table 8
data used in primary studies, broad search query.

Data used	# of studies
Data collected after slaughtering	16
Live pig data measurements	10
Data from the living environment	8
Blood sample data	8
Feed data	6
Reproduction performance measures	5
Animal weight data	2
Other	8

Table 9
features used in primary studies, narrow search query.

Features used	# of studies
Pork meat information	6
Images and recordings	5
Growth	5
Occupancy and reproduction	4
Indoor climatic conditions	3
Genetics	3
Outdoor climatic conditions	1
Other	6

Table 10
Data gathering location primary studies, broad search query.

Data gathering location	# of studies
Farm	21
Experimental farm	8
Slaughterhouse	5
External data	5
Market	3
Laboratory	2
Other	4

Table 11
Data analysis used in primary studies broad search query.

Data analysis method used	# of studies
Least squares	6
Logistic regression	5
General linear (mixed) model	4
Multivariate data analysis	4
Linear discriminant analysis	4
Linear regression	4
Other	13

Table 12
Machine Learning algorithms used in primary studies, narrow search query.

Algorithm	# of studies
Random forest	12
Neural networks	9
Convolutional neural networks	2
Support Vector Machine	8
Gradient Boosting tree	4
Decision tree	3
Bayesian	2
Linear discriminant analysis	2
Elasticnet	2
Logistic regression	2
K-means clustering	1
K-Nearest neighbours	1
Generalized linear modelling	1
Bagging	1

Table 13
Data used in primary studies, broad search query.

Research group	# of studies
# of pigs < 25	4
# of pigs from 25 to 50	2
# of pigs from 50 to 100	3
# of pigs from 100 to 500	7
# of pigs from 500 to 1000	6
# of pigs > 1000	4
Farms	3
Meat	6
Faecal samples	2
Other	4

is required for answering the research questions were extracted and analysed.

3.3. Search strings

The search strings were chosen to ensure that they are broad enough and will result in all relevant publications with respect to data-driven decision making in pig farming. Two search strings were formulated that will allow us to answer all the research questions. The first is a broad search string focusing on the use of data in pig farming, and the second is a narrow search string focusing specifically on the application of ML in pig farming. Different databases offer different search features, and thus we adapted the search strings according to the features offered by the databases.

3.3.1. Broad search string

The broad search string focussed on the use of data analytics in pork chains in general. The goal is to retrieve publications that cover the monitoring, prediction, and other types of data analytics in pig farming and the pork industry. The following list provides the broad search string per database:

Science Direct: (“Data driven” OR “data analysis”) AND (monitoring OR prediction OR analysis) AND (pig? OR pork), [title, abstract or author-specified keywords].

Scopus: (“Data driven” OR “data analysis”) AND (monitoring OR prediction OR analysis) AND (pig? OR pork), [article title, abstract, keywords].

Web of Science: (“Data driven” OR “data analysis”) AND (monitoring OR prediction OR analysis) AND (pig? OR pork), [title, abstract, author keywords, and keywords plus, subcategory “Agriculture animal science”].

Springer Link: (“Data driven” OR “data analysis”) AND (monitoring OR prediction OR analysis) AND (pig? OR pork), [anywhere].

Wiley: (’pig AND pork AND "data analysis" AND (monitoring OR

Table 14
selected publications.

Key	Title	Retrieved from	Authors	Year
P2	NIR hyperspectral imaging as non-destructive evaluation tool for the recognition of fresh and frozen-thawed porcine longissimus dorsi muscles	Science Direct	Barbin, Douglas F; Sun, Da Wen; Su, Chao	2013
P5	Immunocastration with Improvac reduces aggressive and sexual behaviours in male pigs	Science Direct	Brewster, Veronica; Nevel, Amanda	2013
P6	Classification of some heat-treated liver pastes according to container type, using heavy metals content and manufacturer's data, by principal components analysis and potential curves	Science Direct	Brito, G.; Andrade, J. M.; Havel, J.; Díaz, C.; García, F. J.; Peña-Méndez, E. M.	2006
P7	A sensory description of boar taint and the effects of crude and dried chicory roots (Cichorium intybus L.) and inulin feeding in male and female pork	Science Direct	Byrne, Derek V.; Thamsborg, Stig M.; Hansen, Laurits L.	2008
P11	Development of technology for the early post mortem prediction of water holding capacity and drip loss in fresh pork	Science Direct	Forrest, John C.; Morgan, Mark T.; Borggaard, Claus; Rasmussen, Allan J.; Jespersen, Bo L.; Andersen, Jan R.	2000
P12	Assessing the effect of season, montanera length and sampling location on Iberian pig fat by compositional data analysis and standard multivariate statistics	Science Direct	Garrido-Fernández, Antonio; León-Camacho, Manuel	2019
P16	Geographical origin authentication of pork using multi-element and multivariate data analyses	Science Direct	Kim, Jae Sung; Hwang, In Min; Lee, Ga Hyun; Park, Yu Min; Choi, Ji Yeon; Jamila, Nargis; Khan, Naeem; Kim, Kyong Su	2017
P21	Seroprevalence and genetic characteristics of five subtypes of influenza A viruses in the Chinese pig population: A pooled data analysis	Science Direct	Liu, Wei; Wei, Mao Ti; Tong, Yigang; Tang, Fang; Zhang, Lei; Fang, Liqun; Yang, Hong; Cao, Wu Chun	2011
P25	Predicting intramuscular fat content variations in boiled pork muscles by hyperspectral imaging using a novel spectral pre-processing technique	Science Direct	Ma, Ji; Pu, Hongbin; Sun, Da Wen	2018
P36	Development of a biosecurity assessment tool and the assessment of biosecurity levels by this tool on Japanese commercial swine farms	Science Direct	Sasaki, Yosuke; Furutani, Aina; Furuichi, Tomohiro; Hayakawa, Yuiko; Ishizeki, Sayoko; Kano, Rika; Koike, Fumiko; Miyashita, Mali; Mizukami, Yoshihiro; Watanabe, Yugo; Otake, Satoshi	2020
P38	A method for nondestructive prediction of pork meat quality and safety attributes by hyperspectral imaging technique	Science Direct	Tao, Feifei; Peng, Yankun	2014
P44	From the application of antibiotics to antibiotic residues in liquid manures and digestates: A screening study in one European center of conventional pig husbandry	Science Direct	Widyasari-Mehta, Arum; Hartung, Susen; Kreuzig, Robert	2016
P51	Fatty acid composition of the intramuscular fat in the longissimus thoracis muscle of Apulo-Calabrese and crossbreed pigs	Scopus	Aboagye, Gizella; Zappaterra, Martina; Pasini, Federica; Dall'Olio, Stefania; Davoli, Roberta; Nanni Costa, Leonardo	2020
P53	Effects of interactions between feeding practices, animal health and farm infrastructure on technical, economic and environmental performances of a pig-fattening unit	Scopus	Cadéro, A.; Aubry, A.; Dourmad, J. Y.; Salaün, Y.; Garcia-Launay, F.	2020
P55	A bivariate genomic model with additive, dominance and inbreeding depression effects for sire line and three-way crossbred pigs	Scopus	Christensen, Ole F.; Nielsen, Bjarne; Su, Guosheng; Xiang, Tao; Madsen, Per; Ostersen, Tage; Velander, Ingela; Strathe, Anders B.	2019
P58	Genetic characterization of porcine circovirus 3 strains circulating in sardinian pigs and wild boars	Scopus	Dei Giudici, Silvia; Franzoni, Giulia; Bonelli, Piero; Angioi, Pier Paolo; Zinellu, Susanna; Deriu, Viviana; Carta, Tania; Sechi, Anna Maria; Salis, Francesco; Balzano, Francesca; Oggiano, Annalisa	2020
P59	Genomic Characterization of mcr-1-carrying Salmonella enterica Serovar 4,[5],12:i:- ST 34 Clone Isolated From Pigs in China	Scopus	Elbediwi, Mohammed; Wu, Beibei; Pan, Hang; Jiang, Zenghai; Biswas, Silpak; Li, Yan; Yue, Min	2020
P61	Identification of metabonomics changes in longissimus dorsi muscle of finishing pigs following heat stress through LC-MS/MS-based metabonomics method	Scopus	Gao, Jie; Yang, Peige; Cui, Yanjun; Meng, Qingshi; Feng, Yuejin; Hao, Yue; Liu, Jiru; Piao, Xiangshu; Gu, Xianhong	2020
P66	Farm data analysis for lifetime performance components of sows and their predictors in breeding herds	Scopus	Koketsu, Yuzo; Iida, Ryosuke	2020
P68	Chilling control of beef and pork carcasses in a slaughterhouse based on causality analysis by graphical modelling	Scopus	Kuzuoka, Kumiko; Kawai, Kohji; Yamauchi, Syunpei; Okada, Ayaka; Inoshima, Yasuo	2020
P69	African swine fever in two large commercial	Scopus	Lamberga, Kristine; Olševskis, Edvins;	2020

(continued on next page)

Table 14 (*continued*)

Key	Title	Retrieved from	Authors	Year
	pig farms in LATVIA-Estimation of the high risk period and virus spread within the farm		Seržants, Martinš; Berzinš, Aivars; Viltrop, Arvo; Depner, Klaus	
P71	Prevalence and phylogenetic analysis of hepatitis e virus in pigs in Vietnam	Scopus	Lee, Hu Suk; Dao, Duy Tung; Bui, Vuong Nghia; Bui, Ngoc Anh; Le, Thanh Duy; Nguyen-Viet, Hung; Grace, Delia; Thakur, Krishna K.; Hagiwara, Katsuro	2020
P74	Effect of Fermented Corn-Soybean Meal on Serum Immunity, the Expression of Genes Related to Gut Immunity, Gut Microbiota, and Bacterial Metabolites in Grower-Finisher Pigs	Scopus	Lu, Junfeng; Zhang, Xiaoyu; Liu, Yihao; Cao, Haigang; Han, Qichun; Xie, Baocai; Fan, Lujie; Li, Xiao; Hu, Jianhong; Yang, Gongshe; Shi, Xin'e	2019
P75	Effect of RNF4-SacII gene polymorphism on reproductive traits of Landrace — Large White crossbred sows	Scopus	Menčik, Sven; Vuković, Vlado; Jiang, Zhihua; Ostović, Mario; Sušić, Velimir; Žura Žaja, Ivona; Samardžija, Marko; Ekert Kabalin, Anamaria	2020
P78	Postnatal guinea pig brain development, as revealed by magnetic resonance and diffusion kurtosis imaging	Scopus	Mullins, Roger J.; Xu, Su; Zhuo, Jiachen; Roys, Steve; Pereira, Edna F. R.; Albuquerque, Edson X.; Gullapalli, Rao P.	2020
P79	Modelling the feed intake response of growing pigs to diets contaminated with mycotoxins	Scopus	Nguyen-Ba, H.; Taghipoor, M.; Van Milgen, J.	2020
P91	A between-herd data-driven stochastic model to explore the spatio-temporal spread of hepatitis e virus in the French pig production network	Scopus	Salines, Morgane; Andraud, Mathieu; Rose, Nicolas; Widgren, Stefan	2020
P96	Effect of supplementing hydrolysable tannins to a grower-finisher diet containing divergent PUFA levels on growth performance, boar taint levels in back fat and intestinal microbiota of entire males	Scopus	Tretola, Marco; Maghin, Federica; Silacci, Paolo; Ampuero, Silvia; Bee, Giuseppe	2019
P102	Effect of gonadotropin releasing factor suppression with an immunological on growth performance, Estrus activity, Carcass characteristics, and meat quality of market gilts	Web of Science	Bohrer, B. M.; Flowers, W. L.; Kyle, J. M.; Johnson, S. S.; King, V. L.; Spruill, J. L.; Thompson, D. P.; Schroeder, A. L.; Boler, D. D.	2014
P106	High prevalence of F4+ and F18+ *Escherichia* coli in Cuban piggeries as determined by serological survey	Web of Science	de la Fé Rodríguez, Pedro; Yoelvys Coddens, Annelies; Del Fava, Emanuele; Abrahantes, José Cortiñas; Shkedy, Ziv; Martin, Luis O.Maroto; Muñoz, Eduardo Cruz; Duchateau, Luc; Cox, Eric; Goddeeris, Bruno Maria	2011
P109	Prediction of genetic growth curves in pigs	Web of Science	Haraldsen, M.; Ødegård, J.; Olsen, D.; Vangen, O.; Ranberg, I. M.A.; Meuwissen, T. H. E.	2009
P111	The use of a covariate reduces experimental error in nutrient digestion studies in growing pigs	Web of Science	Jacobs, B. M.; Patience, J. F.; Lindemann, M. D.; Stalder, K. J.; Kerr, B. J.	2013
P112	Genetic relationships of body composition, serum leptin, and age at puberty in gilts	Web of Science	Kuehn, L. A.; Nonneman, D. J.; Klindt, J. M.; Wise, T. H.	2009
P113	A splicing mutation in PHKG1 decreased its expression in skeletal muscle and caused PSE meat in Duroc — Luchuan crossbred pigs	Web of Science	Liu, Y.; Liu, Y.; Ma, T.; Long, H.; Niu, L.; Zhang, X.; Lei, Y.; Wang, L.; Chen, Y.; Wang, Q.; Zheng, Z.; Xu, Xuewen	2019
P116	Estimating challenge load due to disease outbreaks and other challenges using reproduction records of sows	Web of Science	Mathur, P. K.; Herrero-Medrano, J. M.; Alexandri, P.; Knol, E. F.; Napel, J. Ten; Rashidi, H.; Mulder, H. A.	2014
P118	The dietary protein content slightly affects the body temperature of growing pigs exposed to heat stress	Web of Science	Morales, Adriana; Valle, J. Alan; Castillo, Gilberto; Antoine, Duckens; Avelar, Ernesto; Camacho, Reyna L.; Buenabad, Lorenzo; Cervantes, Miguel	2019
P119	Effect of dietary tryptophan to lysine ratio on growth of young pigs fed wheat-barley or corn based diets	Web of Science	Naatjes, M.; Htoo, J. K.; Walter, K.; Tölle, K. H.; Susenbeth, A.	2014
P123	Effects of oral administration of sodium citrate or acetate to pigs on blood parameters, postmortem glycolysis, muscle pH decline, and quality attributes of pork	Web of Science	Stephens, J. W.; Dikeman, M. E.; Unruh, J. A.; Haub, M. D.; Tokach, M. D.; Dritz, S. S.	2008
P124	A pilot study on transcriptome data analysis of folliculogenesis in pigs	Web of Science	Tosser-Klopp, G.; L Cao, K. A.; Bonnet, A.; Gobert, N.; Hatey, F.; Robert-Grani, C.; Djean, S.; Antic, J.; Baschet, L.; SanCristobal, M.	2009
P126	Seroprevalence of porcine reproductive and respiratory syndrome, Aujeszky's disease, and porcine parvovirus in replacement gilts in Thailand	Web of Science	Tummaruk, Padet; Tantilertcharoen, Rachod	2012
P129	The Asp298Asn polymorphism of melanocortin-4 receptor (MC4R) in pigs: evidence for its potential effects on MC4R constitutive activity and cell surface expression	Web of Science	Zhang, J.; Li, J.; Wu, C.; Hu, Z.; An, L.; Wan, Y.; Fang, C.; Zhang, X.; Li, J.; Wang, Y.	2020
P131	Vitamin D enhanced pork from pigs exposed	Springer Link	Barnkob, Line Lundbaek; Petersen, Paul Michael; Nielsen,	2019

(*continued on next page*)

Table 14 (*continued*)

Key	Title	Retrieved from	Authors	Year
	to artificial UVB light in indoor facilities		Jens Peter; Jakobsen, Jette	
P135	16S rRNA gene-based association study identified microbial taxa associated with pork intramuscular fat content in feces and cecum lumen	Springer Link	Fang, Shaoming; Xiong, Xingwei; Su, Ying; Huang, Lusheng; Chen, Congying	2017
P145	On the Compositional Analysis of Fatty Acids in Pork	Springer Link	Ros-Freixedes, Roger; Estany, Joan	2014
P146	Developing science-industry collaborations into a transdisciplinary process: a case study on improving sustainability of pork production	Springer Link	Schodl, Katharina; Leeb, Christine; Winckler, Christoph	2015
P150	Effects of lactobacillus plantarum ZJ316 on pig growth and pork quality	Springer Link	Suo, Cheng; Yin, Yeshi; Wang, Xiaona; Lou, Xiuyu; Song, Dafeng; Wang, Xin; Gu, Qing	2012
P155	A pork traceability framework based on Internet of Things	Springer Link	Xu, Baocai; Li, Jingjun; Wang, Yun	2013
P182	Investigation of Ebolavirus exposure in pigs presented for slaughter in Uganda	Wiley	Atherstone, Christine; Diederich, Sandra; Pickering, Bradley; Smith, Greg; Casey, Graham; Fischer, Kerstin; Ward, Michael P.; Ndoboli, Dickson; Weingartl, Hana; Alonso, Silvia; Dhand, Navneet; Roesel, Kristina; Grace, Delia; Mor, Siobhan M.	2020
P223	Prediction of nutrient digestibility in grower-finisher pigs based on faecal microbiota composition	Wiley	Verschuren, Lisanne M. G.; Schokker, Dirkjan; Bergsma, Rob; Jansman, Alfons J. M.; Molist, Francesc; Calus, Mario P. L.	2020
P236	Feed-forward and generalised regression neural networks in modelling feeding behaviour of pigs in the grow-finish phase	Science Direct	Cross, Amanda J.; Rohrer, Gary A.; Brown-Brandl, Tami M.; Cassady, Joseph P.; Keel, Brittney N.	2018
P237	Automatic prediction of stress in piglets (Sus Scrofa) using infrared skin temperature	Science Direct	da Fonseca, Felipe Napolitano; Abe, Jair; Minoro de Alencar Nääs, Irenilza; da Silva Cordeiro, Alexandra Ferreira; do Amaral, Fábio Vieira; Ungaro, Henry Costa	2020
P238	Learning patterns from time-series data to discriminate predictions of tail-biting, fouling and diarrhoea in pigs	Science Direct	Domun, Yuvraj; Pedersen, Lene Juul; White, David; Adeyemi, Olutobi; Norton, Tomas	2019
P243	Predicting pen fouling in fattening pigs from pig position	Science Direct	Jensen, Dan Børge; Larsen, Mona Lillian Vestbjerg; Pedersen, Lene Juul	2020
P247	Dual Stage Image Analysis for a complex pattern classification task: Ham veining defect detection	Science Direct	Lopes, Jessica F.; Barbon, Ana Paula A. C.; Orlandi, Giorgia; Calvini, Rosalba; Lo Fiego, Domenico P.; Ulrici, Alessandro; Barbon, Sylvio	2020
P248	Frequency analysis of the sneeze caused by swine influenza virus strains: Automatic sneeze around-the-clock detection using a support vector machine	Science Direct	Mito, Misaki; Aoki, Takuya; Mizutani, Koichi; Zempo, Keiichi; Wakatsuki, Naoto; Maeda, Yuka; Takemae, Nobuhiro; Tsunekuni, Ryota; Saito, Takehiko	2020
P250	Geographic origin discrimination of pork from different Chinese regions using mineral elements analysis assisted by machine learning techniques	Science Direct	Qi, Jing; Li, Yingying; Zhang, Chen; Wang, Cheng; Wang, Juanqiang; Guo, Wenping; Wang, Shouwei	2021
P252	A machine learning approach for the identification of population-informative markers from high-throughput genotyping data: Application to several pig breeds	Science Direct	Schiavo, G.; Bertolini, F.; Galimberti, G.; Bovo, S.; Dall'olio, S.; Nanni Costa, L.; Gallo, M.; Fontanesi, L.	2019
P253	Determining pig holding type from British movement data using analytical and machine learning approaches	Science Direct	Smith, R. P.; Gavin, C.; Gilson, D.; Simons, R. R. L.; Williamson, S.	2020
P258	PigLeg: Prediction of swine phenotype using machine learning	Scopus	Bakoev, Siroj; Getmantseva, Lyubov; Kolosova, Maria; Kostyunina, Olga; Chartier, Duane R.; Tatarinova, Tatiana V.	2020
P269	Comparison of data analytics strategies in computer vision systems to predict pig body composition traits from 3D images	Scopus	Fernandes, Arthur F.A.; Darea, Joao R.R.; Valente, Bruno Dourado; Fitzgerald, Robert; Herring, William; Rosa, Guilherme J.M.	2020
P275	Comparison of bacterial populations in the ceca of swine at two different stages and their functional annotations	Scopus	Kumar, Himansu; Park, Woncheol; Srikanth, Krishnamoorthy; Choi, Bong Hwan; Cho, Eun Seok; Lee, Kyung Tai; Kim, Jun Mo; Kim, Kwangmin; Park, Junhyung; Lim, Dajeong; Park, Jong Eun	2019
P279	Prediction for global African swine fever outbreaks based on a combination of random forest algorithms and meteorological data	Scopus	Liang, Ruirui; Lu, Yi; Qu, Xiaosheng; Su, Qiang; Li, Chunxia; Xia, Sijing; Liu, Yongxin; Zhang, Qiang; Cao, Xin; Chen, Qin; Niu, Bing	2020
P284	Predicting Growth and Carcass Traits in Swine Using Microbiome Data and Machine Learning Algorithms	Scopus	Maltecca, Christian; Lu, Duc; Schillebeeckx, Constantino; McNulty, Nathan P.; Schwab, Clint; Shull, Caleb; Tiezzi, Francesco	2019
P286	Investigation of muscle transcriptomes using gradient boosting machine learning identifies molecular predictors of feed efficiency in growing pigs	Scopus	Messad, Farouk; Louveau, Isabelle; Koffi, Basile; Gilbert, Hélène; Gondret, Florence	2019
P293	Machine learning applied to transcriptomic data to identify genes	Scopus	Piles, Miriam; Fernandez-Lozano, Carlos; Velasco-Galilea, María; González-	2019

(continued on next page)

Table 14 (*continued*)

Key	Title	Retrieved from	Authors	Year
	associated with feed efficiency in pigs		Rodríguez, Olga; Sánchez, Juan Pablo; Torrallardona, David; Ballester, Maria; Quintanilla, Raquel	
P299	Regression trees in genomic selection for carcass traits in pigs	Scopus	Silveira, L. S.; Lima, L. P.; Nascimento, M.; Nascimento, A. C. C.; Silva, F. F.	2020
P302	Machine Learning Prediction of Crossbred Pig Feed Efficiency and Growth Rate From Single Nucleotide Polymorphisms	Scopus	Tusell, Llibertat; Bergsma, Rob; Gilbert, Hélène; Gianola, Daniel; Piles, Miriam	2020
P305	Applications of Support Vector Machine in Genomic Prediction in Pig and Maize Populations	Scopus	Zhao, Wei; Lai, Xueshuang; Liu, Dengying; Zhang, Zhenyang; Ma, Peipei; Wang, Qishan; Zhang, Zhe; Pan, Yuchun	2020
P315	Using near-infrared spectroscopy in the classification of white and iberian pork with neural networks	Springer Link	Guillén, Alberto; del Moral, F. G.; Herrera, L. J.; Rubio, G.; Rojas, I.; Valenzuela, O.; Pomares, H.	2010
P339	CNN-Based Individual Ghungroo Breed Identification Using Face-Based Image	Springer Link	Mukherjee, Kaushik; Dan, Sanket; Roy, Kunal; Roy, Subhojit; Mustafi, Subhranil; Ghosh, Pritam; Mandal, Satyendra Nath; Hajra, Dilip Kumar; Banik, Santanu; Naskar, Syamal	2020
P353	Video-based pig recognition with feature-integrated transfer learning	Springer Link	Wang, Jianzong; Liu, Aozhi; Xiao, Jing	2018
P358	Prediction of slaughter age in pigs and assessment of the predictive value of phenotypic and genetic information using random forest	Web of Science	Alsahaf, Ahmad; Azzopardi, George; Ducro, Bart; Hanenberg, Egiel; Veerkamp, Roel F.; Petkov, Nicolai	2018
P367	Automated Individual Pig Localisation, Tracking and Behaviour Metric Extraction Using Deep Learning	Web of Science	Cowton, Jake; Kyriazakis, Ilias; Bacardit, Jaume	2019
P374	Machine learning algorithms to predict core, skin, and hair-coat temperatures of piglets	Web of Science	Gorczyca, Michael T.; Milan, Hugo Fernando Maia; Maia, Alex Sandro Campos; Gebremedhin, Kifle G.	2018
P382	Analysis of Growth Performance in Swine Based on Machine Learning	Web of Science	Lee, Woongsup; Ham, Younghwa; Ban, Tae Won; Jo, Ohyun	2019

prediction OR analysis OR data-driven)'), [anywhere, subcategory agriculture].

3.3.2. Narrow search string

To focus on the specific use of ML in the pork sector, a second and narrow search string was defined. The narrow search string is used for searching additional publications that specifically focus on the use of ML in the pig production sector. This search string is used to get a clear view of the state-of-the-art about application ML in the pork sector. The following list provides the narrow search string per database:

Science Direct: ("Machine learning") AND (monitoring OR prediction

Table 15
Broad and narrow search query, topic of research primary publications.

Key	Topic of research
P2	Meat quality
P5	Immunocastration against boar taint
P6	Meat origin detection
P7	Feeding strategy to prevent boar taint
P11	Meat quality
P12	Fat composition, seasonality influences
P16	Meat origin detection
P21	Disease in pigs
P25	Fat composition improvement
P36	Disease and biosecurity level
P38	Meat quality
P44	Antibiotic spread in manure
P51	Fat composition in muscle
P53	Farm performance
P55	Breeding analysis model
P58	DNA characterization
P59	Gene characterization
P61	Muscle profile under heat stress
P68	Meat quality
P69	Disease in pigs
P71	Disease in pigs
P74	Feeding strategy, internal climate influences
P75	Gene characterization
P78	Pig robustness characterization
P79	Pig robustness characterization, feed intake with perturbations monitoring
P91	Disease in pigs
P96	Fat composition, relation with boar taint
P102	Pig performance
P106	Disease in pigs
P109	Gene characterization
P111	Feeding strategy
P112	Gene characterization
P113	DNA characterization
P116	Disease in pigs
P118	Feeding strategy, heat stress and body temp
P119	Feeding strategy
P123	Feeding strategy
P124	Folliculogenesis research
P126	Disease in pigs
P129	Gene characterization
P131	Vitamin D increase in pigs
P135	DNA characterization
P145	Fat composition
P150	Pig performance
P182	Disease in pigs
P223	Gene characterization
P236	Feeding strategy/ behaviour analyses
P237	Stress detection
P238	Behaviour analyses
P243	Behaviour analyses
P247	Pork quality assessment
P248	Disease detection
P250	Origin detection
P252	Disease detection
P253	Pig origin detection
P258	Disease detection
P269	Performance estimation
P275	Disease detection
P279	Disease detection
P284	Performance estimation
P286	Feeding strategy
P293	Feeding strategy
P299	Genome selection
P302	Performance estimation
P305	Genome selection
P315	Origin detection
P339	Pig detection
P353	Pig detection
P358	Performance estimation
P367	Pig detection
P374	Performance estimation
P382	Performance estimation

Table 16
Broad search query, data gathering location.

Key	Data gathering location
P2	Slaughterhouse
P5	Farm
P6	Market
P7	Experimental set-up
P11	Slaughterhouse
P12	Farm
P16	Slaughterhouse
P21	External data
P25	Market
P36	Farm
P38	Market
P44	Farm
P51	Farm
P53	Simulation
P55	Experimental farm
P58	Unknown
P59	Farm, slaughterhouse
P61	Farm
P68	Distribution centre
P69	Farm
P71	Farm, slaughterhouse
P74	Farm
P75	Farm
P78	Experimental farm
P79	External data
P91	Farm, external data
P96	Farm
P102	Farm
P106	Laboratory
P109	External data
P111	Experimental farm
P112	Farm
P113	Farm
P116	Farm
P118	Farm
P119	Experimental farm
P123	Experimental farm
P124	Laboratory
P126	Farm
P129	Experimental farm
P131	Experimental farm
P135	Farm
P145	External data
P150	Farm
P182	Abattoir
P223	Experimental farm

Table 17
Research group size, broad search query.

Key	Research group size
P2	48 meat samples
P5	76 pigs
P6	115 pasts
P7	288 pigs
P11	180 carcasses
P12	200 pigs
P16	3232 meat samples
P21	Externally gathered data
P25	104 muscle/meat samples
P36	141 farms
P38	31 meat samples
P44	21 farms
P51	103 pigs
P53	Virtual, 96 scenarios
P55	218 pigs
P58	263 pigs
P59	337 salmonella(faecal) samples
P61	8 litters
P68	Unknown
P69	204 meat samples
P71	475 sera and 250 faecal samples
P74	48 pigs
P75	461 records from 101 pigs
P78	5 pigs
P79	155 pigs
P91	8 farms
P96	44 pigs
P102	570 pigs
P106	1044 pigs
P109	2777 pigs
P111	72 pigs
P112	924 pigs
P113	614 pigs
P116	908,582 pigs
P118	9 pigs
P119	880 pigs
P123	54 pigs
P124	7 RNA pools
P126	7030 pigs
P129	4 pigs
P131	24 pigs
P135	500 pigs
P145	971 pigs
P150	150 pigs
P182	658 pigs
P223	160 pigs

OR analysis) AND (pig? OR pork), [title, abstract or author-specified keywords].

Scopus: (“Machine learning”) AND (monitoring OR prediction OR analysis) AND (pig? OR pork), [article title, abstract, keywords].

Web of Science: (“Machine learning”) AND (monitoring OR prediction OR analysis) AND (pig? OR pork), [title, abstract, author keywords, and keywords plus].

Springer Link: (“Machine learning”) AND (monitoring OR prediction OR analysis) AND (pig? OR pork), [anywhere, subcategory artificial intelligence].

Wiley: (“Machine learning”) AND (monitoring OR prediction OR analysis) AND (pig? OR pork), [anywhere, subcategory agriculture].

3.4. Selection criteria

To select the most appropriate publications for analysis and exclude less relevant publications, seven Exclusion Criteria (EC) were formulated. A publication was considered as a primary study when the publication passed all criteria.

EC1 – Publication is not in the 50 most relevant results,

EC2 - Publication is not related to the livestock sector and data analysis,

EC3 – Publication is not written in English,

EC4 – Publication that is a duplicate or already found at other databases,

EC5 – Publication does not have full text available,

EC6 – Publication is a review or survey paper, and

EC7 – Publication has been published before 2005.

3.5. Data extraction and synthesis

Data were extracted from the primary studies based on the requirements for answering the research questions. We check each primary study thoroughly first to ensure that the publication passes the exclusion criteria and should belong to the primary studies and then read the publication carefully to extract the required data. The data are analysed to answer the research questions.

4. Results

Table 4 shows the number of publications found after performing the broad search query. Table 5 shows the number of studies found by performing the search with the narrow search query. The large result of the search queries indicates that we chose our search string in such a way that we obtained a broad perspective on the research done in the pork chain. This large initial result allowed us to browse the results and

Table 18
Broad search query, type of data used.

Key	Type of data used
P2	Slaughter data
P5	Environment data, behaviour data
P6	Slaughter data
P7	Environment data, behaviour data
P11	Slaughter data
P12	Environment data, slaughter data
P16	Slaughter data
P21	Live pig data
P25	Slaughter data, images of muscle samples
P36	Reproduction performance measures
P38	Slaughter data, images
P44	Other – liquid manure
P51	Slaughter data
P53	Environment data, reproduction performance measures
P55	Live pig data
P58	Blood sample data
P59	Other - salmonella isolates
P61	Slaughter data, muscle
P66	Unknown
P67	Unknown
P68	Environment data, slaughter data
P69	Environment data, disease spread
P71	Blood sample data, other – faecal samples
P74	Blood sample data
P75	Reproduction performance measures
P78	Feed data
P79	Feed data
P91	Environment data, live pig data, movement, farm specifications
P96	Slaughter data, feed data
P102	Slaughter data, live pig data, estruss activity
P105	Unknown
P106	Blood sample data
P107	Unknown
P109	Live pig data, weight
P110	Unknown
P111	Blood sample data, other – faecal samples
P112	Live pig data, slaughter data, leptin, genotyping
P113	Live pig data, DNA, ear tissue
P116	Reproduction performance measures
P118	Live pig data, environment data, temperature, relative humidity, weight, body temperature
P119	Live pig data, feed data, diet information, body weight
P121	Unknown
P123	Blood sample data, slaughter data
P124	Other – nylon microarrays
P125	Unknown
P126	Blood sample data, reproduction performance measures, live pig data
P127	Unknown
P129	Other – hypothalamus tissue
P131	Live pig data, slaughter data, feed data
P135	Feed data, environment data, slaughter data
P144	Unknown
P145	Slaughter data
P146	Unknown
P147	Unknown
P150	Live pig data, slaughter data
P155	Unknown
P182	Blood sample data
P185	Unknown
P186	Unknown
P194	Unknown
P196	Unknown
P202	Unknown
P203	Unknown
P205	Unknown
P207	Unknown
P214	Unknown
P219	Unknown
P223	Other
P225	Unknown

Table 19
Broad search query, data analysis method used.

Key	Data analysis method used
P2	Partial least squares discriminant analyses
P5	Generalized linear model
P6	Principle component analysis, FA, linear discriminant analysis
P7	Partial least squares
P11	Partial least squares, MLR
P12	CoDa, hierarchical clustering, DA
P16	Principle component analysis, linear discriminant analysis
P21	Multi-level logistic regression
P25	Partial least squares
P36	Shapiro wilk test, SAS
P38	Gompert logistic model
P44	Unknown
P51	SAS
P53	Multivariate data analyses, variance analyses
P55	Bivariate model
P58	Chi-squared
P59	Unknown
P61	Multivariate data analyses
P68	Unknown
P69	Unknown
P71	Unknown
P74	Unknown
P75	Chi-squared
P78	nlsLM
P79	B-spline
P91	Logistic regression
P96	Least squares, multivariate data analyses
P102	SAS
P106	ELISA
P109	Random regression, Bayesian information criterion
P111	SAS
P112	Multitrait variance components
P113	Unknown
P116	Logistic regression
P118	Regression, least significant difference
P119	SAS, generalized linear model
P123	SAS, least squares
P124	Generalized linear model
P126	SAS, generalized linear model, logistic regression, chi-squared
P129	Unknown
P131	One-way ANOVA
P135	Linear Discriminant Analyses
P145	Linear regression
P150	One-way ANOVA, LSD
P182	Logistic regression
P223	Leave-one-out cross validation

get an impression of the research done. We assessed that the first 50 publications contain sufficient information for answering the research questions and thus were used as exclusion criteria. The primary studies with their authors and publication year are included in the appendix (Appendix A).

To get insight into the relevance of the topic for this SLR, the results of the broad and narrow search results are grouped by year of publication and presented in Fig. 1 and Fig. 2, respectively. When looking at the number of primary studies published for the broad search query, it seems that there are downward trends in 2010 and between 2014 and 2018. This is however due to the high number of publications that were found and the exclusion criteria that affected publication in those years disproportionately. Many publications were excluded because they were outside of the 50 most relevant publications. However, there is a consistent and sharp increase in the number of publications since 2018 both for the narrow and broad search queries.

To get insight into the focus of the primary studies, we have categorized the studies by topic (see Table 6). The categories clearly indicate that majority of the studies were on pig health. Most studies focussed on the presence, spread, or outbreak of diseases, and few studies focussed on the causes and prevention of boar taint.

In the selected primary studies, pig health and the performance of pig production were researched often. The production performance was

Table 20
Narrow search query, algorithms used.

Key	NN	RF	BY	DT	SVM	KMC	GBT	KNN	GLM	LDA	EN	LR	BG	CNN
P236	1													
P237														
P238	1													
P243	1	1	1											
P245				1										
P247		1			1									
P248					1									
P250	1													
P252		1												
P253						1								
P258	1	1	1		1		1	1	1	1				
P269	1										1			
P275		1								1				
P279		1										1		
P284		1					1							
P286							1							
P293		1			1						1			
P299		1		1			1						1	
P302					1									
P305					1									
P315	1				1									
P339														1
P353	1													
P358		1												
P367														1
P374	1	1												
P382		1		1	1							1		

Table 21
Algorithm legend.

Abbreviation	Algorithm
NN	Neural networks
RF	Random forest
BY	Bayesian
DT	Decision tree
SVM	Support vector machine
KMC	K-means clustering
GBT	Gradient boosting tree
KNN	K-nearest neighbour
GLM	Generalized linear model
LDA	Linear discriminant analysis
EN	Elasticnet
LR	Logistic regression
BG	Bagging
CNN	Convolutional neural networks

investigated by analysing various parameters, such as the DNA of pigs, fat composition of pig meat, meat quality, and muscle thickness. This shows that the production performance of pig farming was mostly measured using the attributes of the meat measured after the animals are slaughtered.

Besides pig health and pig production results, other topics were researched. Several feeding strategies have been developed and evaluated. Further, research was done to test if the origin of meat could be traced and on-farm environmental influencers, such as disease spread in manure and temperature effects on pigs.

The narrow search query focused more on the application of ML and the topics of research are summarised in Table 7. According to the search results, most of the research focussed on performance related issues, such as pig performance, feeding strategy, and genetics.

Pig health is the second most researched area, focusing mainly on diseases. The trend of using ML for disease detection through image and video analysis is reflected in the number of primary studies found on this topic. Pig detection and behaviour analysis were also performed using ML. Feeding strategy was analysed using basic statistical analysis and using ML. There were also primary studies that focus on detecting the origin of meat, detecting stress in pigs. A detailed description of the topics of research per primary study is shown in Appendix B.

Table 22
Features used during research, narrow search query.

Key	Features used
P236	Time of day, relative humidity
P237	Infrared skin temperature
P238	Water consumption, temperatures on pen level, indoor climate data
P243	Video recordings, positioning pigs
P245	Litter number, stillborn, born, alive, weaned, mortality, re-breeding
P247	Raw ham vein composition
P248	Number of sneezes per pig
P250	Pork samples, mineral elements
P252	SNP genotyping data
P253	Pig movement data, pig population data
P258	Muscle thickness, backfat amount, and average daily gain
P269	Image features such as volume, area, length, widths, heights, polar image descriptors, and polar fourier transforms
p275	Cecum of crossbred korean native pigs
P279	Outbreak data and meteorological features
P284	Microbiome
P286	Microarrays data from muscles
P293	Mean metabolic body weight, average daily gain, and average backfat gain
P299	Molecular markers
P302	Average daily gain and residual feed intake records and genotypes
P305	Genomic data
P315	Exterior and interior left cheek and interior and exterior right cheek spectra
P339	Pig images/records
P353	Pig images/records
P358	Phenotypic data, estimated breeding values (EBVs), along with pedigree and pedigree-genetic relationships
P367	Pig images/records
P374	Environmental data such as air temperature (Ta), black-globe temperature (Tg), and relative humidity (RH)
P382	Temperature, humidity, initial age (IA), initial body weight (IBW), number of pigs (NU) and stocking density (SD)

As can be seen in Table 8, a variety of data were used for research in the pork industry. Data were largely collected from body and tissue (16 cases of dataset after slaughter, and 8 cases of blood sample data from live animals, which add up to 24 primary studies). The rest of the data reported refers largely to measurements of the external characteristics of the animals (10 cases of various measurements about live pigs and 2 cases of animal weight data) and their living environment or feed (8 cases of growing environment and 6 cases of feed data). There were also primary studies that used data on manure and microarrays. The type of data used per primary study is shown in Appendix E.

The results of the analysis of data from the second search query focused on the application of ML to pig farming are shown in Table 9. The table shows the widely used features used in research. The features used can be grouped into features on pigs and features on climatic conditions. Features that focus on pigs vary from features of live pigs to features of meat. Pork meat information, images and recordings, growth, occupancy and reproduction, genetics, and pig movement have been used to build and validate models. The features on climatic data consist of data about indoor or outdoor climatic conditions, or both. Data used in those primary studies consist of outside temperature, indoor temperature, and relative humidity. The features used per primary study are shown in Appendix H.

The location of data gathering provides insight into the main sources of research data. We have categorized the locations where data are gathered into groups (see Table 10), which shows that most of the data is gathered at farms and slaughterhouses. The farms were sometimes conventional farms where a specific group of animals was monitored closely; other times, they were experimental farms used for precisely monitoring growing conditions and pig behaviour.

Slaughterhouses serve as one of the main data gathering locations. The quality and composition of meat are important indicators of production performance; therefore, data from slaughterhouses where these indicators are routinely measured, are important sources of data.

Other locations of data gathering for research include laboratories (where meat samples from markets are collected and analysed), externally collected data, previous research, or large database where data are shared. Further, data was collected at distribution centres, urban abattoirs, and there was a case of data from a virtual simulation. The data gathering location per primary study is shown in Appendix C.

To get an overview of data analysis methods used in the primary studies, the algorithms or principles used were calculated as shown in Table 11. Least squares regression was used most often, followed by logistic regression. The primary studies found by the broad search query did not often use complex algorithms; instead, they applied variations of the linear regression model. All data analysis methods used are shown in Appendix F.

All primary studies found by the second search query used ML algorithms. Random forest algorithm was used most, followed by neural networks and support vector machine, as is shown in Table 12. Compared to the primary studies found by the broad search query, those found by the narrow search query used more complex analysis models. The algorithms used per publication are shown in Appendix G. Eight studies using models based on random forest and/or neural networks recorded an R-squared accuracy scoring higher than 0.7. The challenges stated include difficulty in scaling into large-scale farms due to differences in the environment between the experimental setup and large scale farm, and other features that influence the performance.

Table 13 shows what has been studied, and if the research subjects were pigs, it shows the size of the group of pigs studied. A total of 26 primary studies considered live pigs, from which measurements were made. The number of pigs used for research varies significantly, from 5 to 908,582 pigs. Three primary studies considered farms as a whole, focusing on the growing environment of the pigs. Further, data on meat quality and faecal samples were used by some studies. There were also studies that used externally collected data, virtual experimental data, RNA pools, and liver pasts. Appendix D shows the research group types and sizes per primary study.

Much research has been using data on genetics, pig breeds, meat quality after processing, disease recognition, and factors influencing diseases. Research that focused on the prevention of negative aspects of the final product is, for example, research done on boar taint and the attempt to reduce it. Boar taint in meat results in a lower price for the carcass.

A variety of challenges have been stated in the studies regarding ML. The largest challenge is making the algorithms perform consistently high for all situations. Many studies reported that it was feasible to define a well performing model for the specific research setup and condition the model is trained and tested on but it was in general not possible to guarantee the same level of performance in practice. Mostly, more data and more features are required in order to improve the overall performance of the models developed.

5. Discussion and conclusion

The original aim of this study was to review the literature on the use of advanced data analytics techniques in real-life business scenarios. We tried diverse search strings in order to find studies that are not based on experimental set-up be based on the use of data sources from real-life production chains. But the number of results we obtained was low. We were unable to find any study that fused data from the whole production chain except one by Ma et al. (2019). Ma et al. (2019) developed an intelligent feeding equipment and network service platform, using sensors, software, and monitoring units. This system helped in optimizing feeding methods and improving pig management.

We considered, therefore, two search strings, a broad and a narrow search string: the former focussing on general data analytics and the second on machine learning. In both cases, we used more generic terms and relied on exclusion criteria to select the relevant studies. We consulted five scientific databases in this SLR, namely: Science Direct, Scopus, Web of Science, Springer Link, and Wiley. We did not consult Google Scholar initially which returned 16,300 publications, a vast majority of which were irrelevant, and thus filtering out the relevant publications was infeasible.

The exclusion criteria we selected ensured that only publications that are accessible and of high quality were considered. The exclusion criterion on year of publication (set at 2006) was based on our initial test search results. While research performed 20 years back is likely outdated, the set year allowed us to get a clear view of the trend in the studies published. This made it clear, for instance, that the number of primary studies increased substantially from 2018 onwards, particularly on the application of machine learning in the pork sector.

When analysing the results of the SLR, we grouped the attributes such as features, data types, or algorithms we found. Grouping the results allowed us to derive clear conclusions. The detailed results are however included as appendices to this study. As can be seen in the increasing number of primary studies since 2018, the application of ML will be important in the pork sector in the future. When looking at the algorithms used in this context, studies that used features and data on images and recordings often used neural networks, specifically convolutional neural networks algorithm. For example, Cowton et al. (2019), Mukherjee et al. (2020), and Fernandes et al. (2020) focused on pig recognition and deviations in behaviour. Studies that used pork meat information often had random forest as best performing algorithm. The focus of the research using pork meat information was to investigate growth performance (Piles et al., 2019) or the state of health (Bakoev et al., 2020).

What is striking in our review is that most of the research conducted use only groups of pigs selected for an experiment. Most of the studies are conducted in large pig farms; however, the numbers of animals studied within the farms were often small, indicating the studies were conducted on an experimental basis and not on the entire herd. While the focus on a selected group of pigs helps to accurately track and monitor pigs, it does not reflect the real-life situation. Therefore, it remains unclear whether the solutions found in research could be applied in real-life business cases. Out of the 41 primary studies we considered through the broad search query, only 10 involve more than 500 pigs, and of which only two, performed by Naatjes et al. (2014) and Haraldsen et al. (2009), used real-time data from farm management systems. Research based on real-life data (instead of data collected specifically for research purposes) is scarce and most studies focussed mainly on isolated and one-off problems. We suggest that future research should consider the complexity encountered in real-life circumstances and the integration of data analytics within the management systems so that the analytical results can be used within routine business processes.

Supplementary materials

Supplementary material associated with this article can be found, in the online version, at doi:10.1016/j.livsci.2022.104961.

Appendices

Appendix A: Selected publications

The selected publications after passing the exclusion criteria for both the broad and narrow search query are shown in Table 14. P2 until P223 are publications found from the broad search query, P236 until P283 are publications found with the narrow search query.

Appendix B: Topic of research

The topic of research per publication is shown in Table 15.

Appendix C: Data gathering location

The place of data collection per publication for the broad search query is shown in Table 16.

Appendix D: Research group size

Table 17 shows the research group size per publication, for the broad search query.

Appendix E: Type of data used in research

Table 18 shows the type of data that was used during research for the selected publications.

Appendix F: Data analysis methods used

Table 19 shows the data analysis methods used for the selected publications.

Appendix G: Algorithms used

Table 20 shows the algorithms used in the publications that resulted from the narrow search query, Table 21 shows the legend for all the algorithm abbreviations.

Appendix H: Features used

Table 22 shows the type features and data that have been used in the selected publications.

References

Bakoev, S., Getmantseva, L., Kolosova, M., Kostyunina, O., Chartier, D.R., Tatarinova, T. V., 2020. PigLeg: prediction of swine phenotype using machine learning. PeerJ. https://doi.org/10.7717/peerj.8764, 2020.

Banhazi, T.M., Babinszky, L., Halas, V., Tscharke, M., 2012. Precision livestock farming: precision feeding technologies and sustainable animal production. Int. J. Agric. Biol. Eng. 5, 54–61. https://doi.org/10.3965/j.ijabe.20120504.00.

Bos, J.M., Bovenkerk, B., Feindt, P.H., Van Dam, Y.K., 2018. The quantified animal: precision livestock farming and the ethical implications of objectification. Food ethics 2, 77–92. https://doi.org/10.1007/s41055-018-00029-x.

Bose, I., Mahapatra, R.K., 2001. Business data mining - a machine learning perspective. Inf. Manag. 39, 211–225. https://doi.org/10.1016/S0378-7206(01)00091-X.

Bradford, G.E., 1999. Contributions of animal agriculture to meeting global human food demand. Livestock Production Science. Elsevier, pp. 95–112. https://doi.org/10.1016/S0301-6226(99)00019-6.

Cowton, J., Kyriazakis, I., Bacardit, J., 2019. Automated individual pig localisation, tracking and behaviour metric extraction using deep learning. IEEE Access 7, 108049–108060. https://doi.org/10.1109/ACCESS.2019.2933060.

Fernandes, A.F.A., Dórea, J.R.R., Valente, B.D., Fitzgerald, R., Herring, W., Rosa, G.J.M., 2020. Comparison of data analytics strategies in computer vision systems to predict pig body composition traits from 3D images. J. Anim. Sci. 98 https://doi.org/10.1093/jas/skaa250.

Flath, C.M., Stein, N., 2018. Towards a data science toolbox for industrial analytics applications. Comput. Ind. 94, 16–25. https://doi.org/10.1016/j.compind.2017.09.003.

Garner, S.R., Holmes, G., Mcqueen, R.J., Witten, I.H., 1999. Machine learning from agricultural databases: practice and experience.

Haraldsen, M., Ødegård, J., Olsen, D., Vangen, O., Ranberg, I.M.A., Meuwissen, T.H.E., 2009. Prediction of genetic growth curves in pigs. Animal 3, 475–481. https://doi.org/10.1017/S1751731108003807.

Hill, M.A., Scheidt, A.B., Teclaw, R.F., Clark, L.K., Knox, K.E., Jordan, M., 1992. Association between growth indicators and volume of lesions in lungs from pigs at slaughter. Am. J. Vet. Res. 53, 2221–2223.

Kitchenham, B., Charters, S., Budgen, D., Brereton, P., Turner, M., Linkman, S., Jorgensen, M., Mendes, E., Visaggio, G., 2007. Guidelines for Performing Systematic Literature Reviews in Software Engineering. Keele & Durham, UK.

Koketsu, Y., Iida, R., 2020. Farm data analysis for lifetime performance components of sows and their predictors in breeding herds. Porc. Health. Manag. https://doi.org/10.1186/s40813-020-00163-1.

Lee, I., Shin, Y.J., 2020. Machine learning for enterprises: applications, algorithm selection, and challenges. Bus. Horiz. 63, 157–170. https://doi.org/10.1016/j.bushor.2019.10.005.

Liakos, K., Busato, P., Moshou, D., Pearson, S., Bochtis, D., 2018. Machine learning in agriculture: a review. Sensors 18, 2674. https://doi.org/10.3390/s18082674.

Liang, R., Lu, Y., Qu, X., Su, Q., Li, C., Xia, S., Liu, Y., Zhang, Q., Cao, X., Chen, Q., Niu, B., 2020. Prediction for global African swine fever outbreaks based on a combination of random forest algorithms and meteorological data. Transbound. Emerg. Dis. 67, 935–946. https://doi.org/10.1111/tbed.13424.

Ma, W., Fan, J., Zhao, C., Wu, H., 2019. The realization of pig intelligent feeding equipment and network service platform. IFIP Advances in Information and Communication Technology. Springer, New York LLC, pp. 473–484. https://doi.org/10.1007/978-3-030-06179-1_47.

Mucherino, A., Papajorgji, P., Pardalos, P.M., 2009. A survey of data mining techniques applied to agriculture. Oper. Res. https://doi.org/10.1007/s12351-009-0054-6.

Mukherjee, K., Dan, S., Roy, K., Roy, S., Mustafi, S., Ghosh, P., Mandal, S.N., Hajra, D.K., Banik, S., Naskar, S., 2020. CNN-based individual ghungroo breed identification using face-based image. Lecture Notes in Electrical Engineering. Springer Science and Business Media Deutschland GmbH, pp. 241–252. https://doi.org/10.1007/978-981-15-7031-5_23.

Naatjes, M., Htoo, J.K., Walter, K., Tölle, K.H., Susenbeth, A., 2014. Effect of dietary tryptophan to lysine ratio on growth of young pigs fed wheat-barley or corn based diets. Livest. Sci. 163, 102–109. https://doi.org/10.1016/j.livsci.2014.02.004.

Neethirajan, S., 2020. The role of sensors, big data and machine learning in modern animal farming. Sens. Bio Sens. Res. https://doi.org/10.1016/j.sbsr.2020.100367.

Nguyen Thi Thuy, M., Dorny, P., Lebailly, P., Le Thi Minh, C., Nguyen Thi Thu, H., Dermauw, V., 2020. Mapping the pork value chain in Vietnam: a systematic review. Trop. Anim. Health Prod. https://doi.org/10.1007/s11250-020-02338-y.

Pääkkönen, P., Pakkala, D., 2015. Reference architecture and classification of technologies, products and services for big data systems. Big Data Res. https://doi.org/10.1016/j.bdr.2015.01.001.

Piles, M., Fernandez-Lozano, C., Velasco-Galilea, M., González-Rodríguez, O., Sánchez, J. P., Torrallardona, D., Ballester, M., Quintanilla, R., 2019. Machine learning applied to transcriptomic data to identify genes associated with feed efficiency in pigs. Genet. Sel. Evol. 51 https://doi.org/10.1186/s12711-019-0453-y.

Piñeiro, C., Morales, J., Rodríguez, M., Aparicio, M., Manzanilla, E.G., Koketsu, Y., 2019. Big (pig) data and the internet of the swine things: a new paradigm in the industry. Anim. Front. 9, 6–15. https://doi.org/10.1093/af/vfz002.

Pourmoayed, R., Nielsen, L.R., Kristensen, A.R., 2016. A hierarchical Markov decision process modeling feeding and marketing decisions of growing pigs. Eur. J. Oper. Res. 250, 925–938. https://doi.org/10.1016/j.ejor.2015.09.038.

Rose, D.C., Chilvers, J., 2018. Agriculture 4.0: broadening responsible innovation in an era of smart farming. Front. Sustain. Food Syst. 2, 87. https://doi.org/10.3389/fsufs.2018.00087.

Sessions, V., Valtorta, M., 2006. He effects of data quality on machine learning algorithms. [WWW Document].

Sharma, R., Kamble, S.S., Gunasekaran, A., Kumar, V., Kumar, A., 2020. A systematic literature review on machine learning applications for sustainable agriculture supply chain performance. Comput. Oper. Res. 119 https://doi.org/10.1016/j.cor.2020.104926.

Tekinerdogan, B., 2019. Organizational IS and their impact - engineering and management of information systems [WWW Document].

Van der Meulen, H., 2020. Agrimatie - economisch resultaat varkenshouderij [WWW Document].

Van Hertem, T., Rooijakkers, L., Berckmans, D., Peña Fernández, A., Norton, T., Berckmans, D., Vranken, E., 2017. Appropriate data visualisation is key to precision livestock farming acceptance. Comput. Electron. Agric. 138, 1–10. https://doi.org/10.1016/j.compag.2017.04.003.

van Klompenburg, T., Kassahun, A., Catal, C., 2020. Crop yield prediction using machine learning: a systematic literature review. Comput. Electron. Agric. https://doi.org/10.1016/j.compag.2020.105709.

Vlontzos, G., Pardalos, P.M., 2017. Data Mining and Optimisation Issues in the Food Industry. Inderscience Publishers, pp. 44–64. https://doi.org/10.1504/IJSAMI.2017.082921.

Wageningen University, 2020. Precision feed for pigs reduces environmental impact [WWW Document].

Wageningen University, 2020. Precision agriculture - smart farming [WWW Document].

Wang, X., 2020. Behind China's 'pork miracle': how technology is transforming rural hog farming environment the guardian [WWW Document].

Xu, B., Li, J., Wang, Y., 2013. A pork traceability framework based on Internet of Things. Communications in Computer and Information Science. Springer Verlag, pp. 159–166. https://doi.org/10.1007/978-3-642-41629-3_15.

Yazdanbakhsh, O., Zhou, Y., Dick, S., 2017. An intelligent system for livestock disease surveillance. Inf. Sci. 378, 26–47. https://doi.org/10.1016/j.ins.2016.10.026 (Ny).