LANDTECHNIK 76(1), 2021, 24–35
DOI:10.15150/lt.2021.3261

LANDTECHNIK
AGRICULTURAL ENGINEERING

Machine learning for predicting animal welfare risks in pig farming

Tobias Zimpel, Martin Riekert, Achim Klein, Christa Hoffmann

Animal welfare is a quality indicator of modern pig farming and increasingly important to society. Animal welfare risks have multiple factors and should be recognized and mitigated early on to prevent economic risks. In this work, we use machine learning models to predict animal welfare risks. Our dataset comprises data for over 57,000 pigs with indications of 10 animal welfare risks and 14 suckling phase features. We contribute a prediction model for suckling phase deaths with an accuracy of 80.4% – providing a sizeable improvement over a majority vote‘s accuracy of only 53.1%. The proposed model may help pig farmers to prevent deaths in the suckling phase of pigs at an early stage by taking countermeasures.

Keywords
Animal Welfare, birth-related animal welfare risks, pig farming, early influencing factors, machine learning

Animal welfare can be defined using the five freedoms according to the Farm Animal Welfare Council (2009): freedom from hunger, thirst, discomfort, pain, fears, suffering, injuries, illnesses, and the freedom to develop normal behavior. In practice, animal welfare risks are recognized, for example, by observing animal welfare indicators (Zapf et al. 2017). The procedure and information on the collection of animal welfare indicators in practice are included in the guidelines of the Kuratorium für Technik und Bauwesen in der Landwirtschaft e.V.

Also, sensors are used in the literature to monitor animal welfare indicators. For example, cameras and microphones are applied (Riekert et al. 2020, Matthews et al. 2016). Microphones can be used to determine a pig’s stress level. Thereby, the classification error is less than 5% (Manteuffel and Schön 2002). Chung et al. (2013) use microphones for diseases detection and report a classification error between 6 and 9%. However, sensor-based approaches for monitoring animal welfare risks require the installation and operation of sensors.

Another approach consists of using data as indicators that are already recorded in the piggery, e.g., the weight. Preliminary work on the prediction of animal welfare risks indicates the influence of early biological indicators (Díaz et al. 2017). Díaz et al. (2017) studied the relationship between biological indicators recorded at an early stage and the mortality and the health of pigs. According to Díaz et al. (2017), the probability of death is 28.4 % for pigs with a birth weight of less than 0.95 kg.

On the other hand, pigs with a birth weight greater than 0.95 kg survived with a probability of 87.1% (Díaz et al. 2017). However, due to the low lower limit, the procedure is very restrictive. Also, Díaz et al. determined relationships between early biological indicators and animal welfare risks that are related to diseases and inflammation of organs.

In particular, pigs with lower weaning weights have a higher risk of heart disease, pneumonia, and pleurisy than pigs with higher weaning weights. The knowledge of the higher risk of heart disease with low weaning weight can, at least for high blood pressure as heart disease, be transferred to low birth

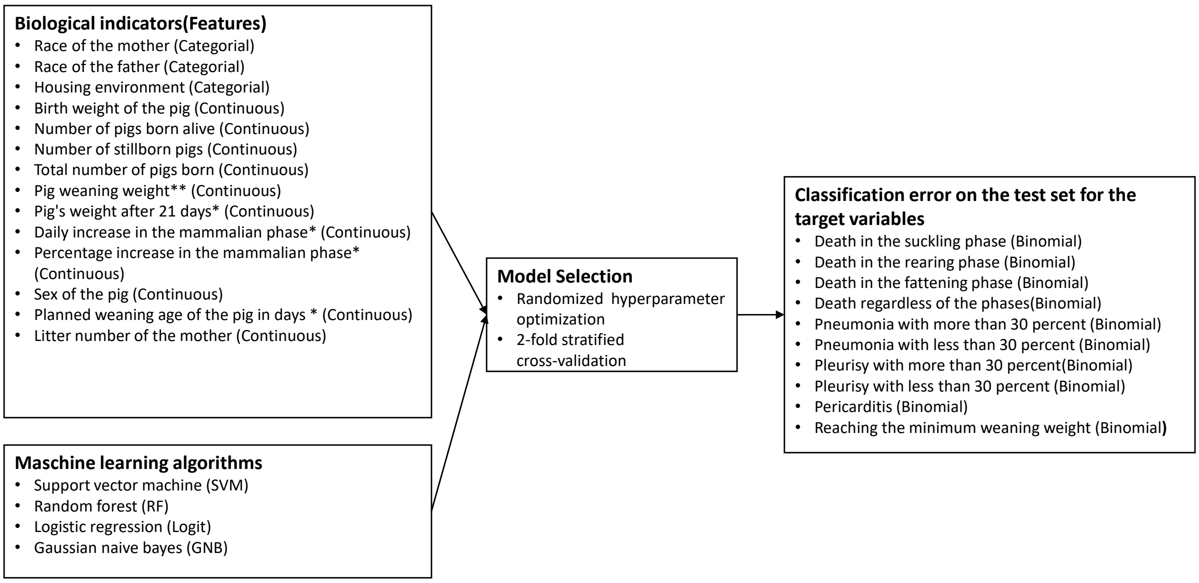
eingereicht 17 June 2019 | akzeptiert 6 August 2020 | veröffentlicht 11 March 2021

weight (Poore et al. 2002). Poore et al. (2002) stated that pigs with a birth weight of less than 1.4 kg had a higher risk of developing high blood pressure than pigs with a birth weight of more than 1.7 kg.

Against this background, we use machine learning (ML) to predict animal welfare risks with biological indicators as early as possible. ML algorithms are defined by their ability to improve performance with increasing experience concerning a task (Mitchell 1997). In this work, we use the machine learning algorithms support-vector machine (SVM) (Cortes and Vapnik 1995), random forest (RF) (Breiman 2001), gaussian naive bayes (GNB) (John and Langley 1995) and logistic regression (Logit) (Lin et al. 2008). Hyperparameters have an important influence on these algorithms. A suitable choice of the hyperparameters increases the performance of the ML algorithm. Therefore, testing different assignments of the hyperparameters is required (Bergstra, Yamins and Cox 2013).

Material and method

Our dataset was recorded at the Boxberg Teaching and Research Centre – Centre for pig rearing and pig breeding (LSZ Boxberg) for the analysis, consisting of 57,104 examples including different biological data of pigs (like weights, environmental forms, diseases). Based on the pigs' birthdate, the biological data were recorded in the period between October 2011 and May 2018. Our approach for generating models for predicting animal welfare risks is divided into the model selection, parameterization and training of the ML-methods on the training set, and determining the performance on an independent test set. Figure 1 shows our research model for ten different animal welfare risks as target variables.



** Not for death in the suckling phase and for reaching the minimum weaning weight
* Not for death in the suckling phase

Figure 1: Research model

These ten animal welfare risks consist of four deaths in different phases, five diseases, and reaching the minimum weaning weight. In this work, the minimum weaning weight is reached if the pig has a weight of at least 5 kg on the weaning date. The four forms of death consist of death in the suckling phase, death in the rearing phase, death in the fattening phase, and death regardless of the phase. The five diseases are: pneumonia with a disease degree of less than 30% of the lungs, pneumonia with a disease degree of more than 30% of the lungs, pericarditis, pleurisy affecting than 30% of the tissue layer, and pleurisy affecting more than 30% of the tissue layer. These ten animal welfare risks were chosen due to their influence on the development of pigs (Botreau et al. 2007). Each animal welfare risk is binary coded with the value 1 for "occurs" and 0 for "does not occur". Diseases are stored in the data when they occur. Hence, we assume that no record of the disease implies that the disease has not occurred. Deaths are categorized according to whether there is a death or a regular slaughter in the records.

We selected 14 biological indicators that are available during the suckling phase and are directly related to pigs (Figure 1):

1. Pigs' housing environment during the suckling, rearing, and fattening phase with the forms "alternative" and "conventional" housing
2. Eight races of the mother
3. Seven races of the father
4. The birth weight
5. The control weight on the 21st day after birth
6. The weaning weight
7. Gender with the form's male, female, and hermaphrodite
8. The percentage increase between weaning weight and birth weight
9. The planned weaning age in days.
10. The daily increase during the suckling phase
11. The mother's litter number (number of litters)
12. The number of pigs born alive in the same litter
13. The number of stillborn pigs in the same litter
14. The total number of pigs born in the same litter

The pig's housing environment describes either conventional housing on a slatted floor or alternative housing, which differs in terms of floor design and stable area for each animal (Federal Ministry of Food and Agriculture Germany 2018, Mauer 2014). The eight races of the mother (mainly a cross between Large White and the German Landrace as well as a cross between Large White, Leicoma, and Deutsche Landrace) and the seven races of the father (mainly German Piétrain pig) in the dataset correspond to the breed coding of the Bundesverband Rind und Schwein e.V.

For data cleansing, the steps outlier detection, typo correction for text fields, and data type inspection were carried out. However, outlier detection was only performed for continuous features, i.e., biological indicators with a value from the real numbers (e.g., birth weight). Features are the input data into the ML algorithms. A feature value was marked as an outlier if the value is greater than the upper limit of 1.5 times the interquartile range plus 5% of the interquartile range to the third quartile (Q3) or smaller than the lower limit of 1.5 times the interquartile range minus 5% of the interquartile range to the first quartile (Q1). Then, we removed these from the dataset. Furthermore, we removed examples from the dataset if the actual data type of the feature did not match the target data type. Also, we deleted examples that did not have a valid filled death or slaughter feature. The minimum weaning weight was calculated and achieved if the weaning weight was greater than or equal to 5 kg.

We created a separate dataset for each animal welfare risk. Also, e.g., if pigs died in the suckling phase, these cannot appear in the dataset for death in the fattening phase. In sum, ten datasets with a size of up to 11,000 examples were created. Thereby, we chose the most recent examples according to the birthdate. Each of the datasets was divided into a training and test set before one-hot encoding. Here, we used the most recent 10,000 examples for the training set and the remaining examples as a test set. One-hot encoding describes the process of replacing a categorical feature by inserting a new column for each possible value of this categorical feature. Here, categorical features describe biological indicators for which the values are nominally scaled (e.g., gender). After the split into training and test set, examples were removed from the test set in which expressions of categorical features (e.g., housing "alternative") appear in the test set, but not in the training set. If a value of a categorical feature is not included in the training set is used in the test set, the meaning of the columns and the number of columns may change after one-hot encoding. Table 1 shows the number of examples in the training and testing set.

Table 1: Number of examples in the training and test set

Animal welfare risk	**D. SP**	**D. RP**	**D. FP**	**D.**	**P**	**PL1**	**PL2**	**PN1**	**PN2**	**MWW**
Examples in the training set	10,000	10,000	10,000	10,000	10,000	10,000	10,000	10,000	10,000	10,000
Examples in the test set	652	264	999	428	1,000	1,000	1,000	1,000	652	71

D: Death regardless of the phase, D. RP: Death in the rearing phase, D. FP: Death in the fattening phase, D. SP: Death in the suckling phase, MWW: Reaching the minimum weaning weight, P: Pericarditis, PL1: Pleurisy < 30 %, PL2: Pleurisy > 30 %, PN1: Pneumonia < 30 %, PN2: Pneumonia > 30 %

In this work, we used the algorithms (Scikit-learn Version 0.19.2) SVM (support vector machine), RF (random forest), GNB (gaussian naive bayes), and Logit (logistic regression). Because no method has a higher performance a priori, a comparison of the procedures is necessary (Wolpert 1996). For replication, we published our code at https://wi2.uni-hohenheim.de/analytics (Pineau et al. 2020).

Before the randomized hyperparameter optimization, the training data were normalized to the interval [0.1] and transformed into vectors for categorical features using the one-hot encoding method for each animal welfare risk (Bergstra and Bengio 2012). This normalization allows for a faster computation time by holding null values in comparison to non-null values for categorical features. The corresponding test set was also normalized using the same normalization setting from the underlying training set.

Subsequently, the model selection is carried out using randomized hyperparameter optimization. Different configurations of hyperparameters (Table 2) are tested on the training set to determine the parameters achieving the highest performance of the method (Bergstra and Bengio 2012).

Table 2: Values for randomized hyperparameter optimization

ML	Hyperparameter	Range
RF	Number of features for the split	Random integer between one and the number of features
	Lower limit based on examples of a splitt	Random integer between two and eleven
	Bootstrap	True, False
	Split criterion	Gini-Index, Entropy
	Number of estimators	10,100, 200
SVM	Kernel	linear, poly, rbf, sigmoid
	C	0.001, 0.01, 0.1, 1, 10, 50, 75, 100, 125, 150, 1,000, 1,500, 2,000, 3,000, 5,.000
Logit	C	0.001, 0.01, 0.1, 1, 10, 50, 75, 100, 125, 150, 1,000, 1,500, 2,000, 3,000, 5,000

ML: Machine learning method, Logit: Logistic regression, RF: Random forest, SVM: Support vector machine

For ML algorithm RF, we used a random natural number between one and the number of biological indicators for the number of features and a random natural number between one and eleven for the minimum number of samples. Also, we used the entire dataset for tree generation, Gini index and entropy as the split criterion, as well as the numbers 10, 100, and 200 for the number of trees. The Gini index uses the relative frequencies of the feature values as the basis for making decisions about the split. Entropy describes a measure for the level of information as the basis for decision-making during the split.

For ML algorithms SVM and Logit, we used the values 0.0001, 0.01, 0.1, 1, 10, 50, 75, 100, 125, 150, 1,000, 1,500, 2,000, 3,000 or 5,000 for hyperparameter C. These values are based on Hsu et al. (2003). For the hyperparameter “kernel”, the SVM we chose a linear kernel (linear), a polynomial kernel (poly), a kernel with a radial basis function (rbf), and a kernel with a sigmoid function (sigmoid). We conducted the randomized hyperparameter optimization with ten runs using random parameters of the predefined set of parameters and two-fold stratified cross-validation for each ML algorithm (Kohavi 1995). No hyperparameter optimization was conducted for GNB, because no hyperparameters of the procedure can be specified without limiting GNB (e. g., costs of the SVM).

For each animal welfare risk, together with the hyperparameter optimization, the performance of all possible combinations of biological indicators was compared with one another for each ML-method. (e. g., birth weight, weaning weight and the combination of birth weight and weaning weight). Thus, with 13 biological indicators there exist over 16,000 combinations for the animal welfare risks death in the resting phase, death in the fattening phase, and for the diseases. Also, nine biolocial indicators generating over 500 combination for the death in the suckling phase and death regardless of the phase. In contrast, over 4,000 combinations with 10 features were compared for the risk reaching the minimum weaning weight.

Accuracy describes the percentage of a classifier to correctly assign objects to a class (Witten et al. 2011, Nguyen and Armitage 2008). The classification of the ML is divided into the following categories:

- True positive (tp): The class was recognized, and the class is present.
- False positive (fp): The class was recognized, but this class is not present.
- False negative (fn): The class was not recognized, but the class is present.
- True negative (tn): The class was not recognized and does not present.

The following metrics were calculated using the confusion matrix terminology (Manning and Schütze 1999):

$$Recall = \frac{tp}{tp + fn}$$

$$Precision = \frac{tp}{tp + fp}$$

$$Accuracy = \frac{tp + fp}{tp + fp + tn + fn}$$

The majority vote describes a decision rule in which the most common class is selected as the true class. The majority vote is calculated as follows (Tian and Zhu 2015):

$$\hat{y}_i = \underset{d \in [D]}{\text{argmax}} \sum_{j=1}^{N} \mathbb{I}\left(x_{ij} = d\right), \forall i \ \in [M]$$

Thereby, $\hat{y}_i$ is the selected class for an example i of a dataset with M examples, [D] the set of occurring classes (here: $[D] = \{0, 1\}$), j a method of the N methods (here: $|N| = 1$) for proposals of the class, and x_{ij} represents the proposed class of j for $\mathbb{I}(\cdot)$ is an indicator function that takes the value 1 if the proposed class corresponds to the actual class and takes the value 0 otherwise (Tian and Zhu 2015).

Accuracy is used to compare the combinations of biological indicators during the parameterization and training of the methods for each animal welfare risks. Then, we calculated the metrics recall, precision, and accuracy on the independent test using the trained ML model of the combinations of biological indicators with the highest accuracy on the validation amount for each animal welfare risk.

Results

Table 3 shows the accuracy of the animal welfare risks after hyperparameter optimization in the model selection using two-fold stratified cross-validation. On the training set for the death regardless of the phases we achieve an accuracy of 0.736 and an accuracy of 0.837 for death in the suckling phase. However, no precision and no recall were calculated on the validation set. No improvement of the accuracy concerning the majority vote in the model selection that is greater than 0.02 is achieved for the animal welfare risks pneumonia, pericarditis, pleurisy, reaching the minimum weaning weight, death in the rearing phase, and death in the fattening phase. For the remaining configurations the model configurations are not reported, because no improvement in accuracy is achieved in the model selection or on the test set.

Table 3: Accuracy in the model selection

	D	D. SP	MWW	D. RP	D. FP	P	PL1	PL2	PN1	PN2
Majority Vote	0.502	0.717	0.980	0.951	0.984	0.980	0.983	0.998	0.929	0.993
SVM	0.736	0.836	0.996	0.951	0.984	0.980	0.983	0.998	0.929	0.993
RF	0.696	0.830	0.993	0.951	0.984	0.981	0.983	0.999	0.929	0.993
Logit	0.729	0.837	0.994	0.951	0.984	0.980	0.983	0.998	0.929	0.993
GNB	0.733	0.663	0.966	0.934	0.984	0.980	0.983	0.998	0.925	0.993

D: Death regardless of the phases, D. RP: Death in the rearing phase, D. FP: Death in the fattening phase, D. SP: Death in the suckling phase, GNB: Gaussian naive bayes, Logit: Logistic regression, MWW: Reaching the minimum weaning weight, P: Pericarditis, PL1: Pleurisy < 30 %, PL2: Pleurisy > 30 %, PN1: Pneumonia < 30 %, PN2: Pneumonia > 30 %, RF: Random forest, SVM: Support vector machine

Subsequently, for each animal welfare risk the ML model that achieved the highest accuracy was selected. Then, we tested these models on the test set. For example, we selected the SVM as an ML model for death regardless of the phases and Logit as an ML model for death in the suckling phase. The corresponding accuracy, precision, and recall using these ML models on the test set are shown in Table 4.

Table 4: Metrics of the selected ML algorithms on the test set

	D	D. SP	MWW**	D. RP*	D. FP*	P*	PL1*	PL2*	PN1*	PN2*
Accuracy majority vote	0.661	0.531	0.972	0.909	0.990	0.981	0.992	0.995	0.956	0.988
Accuracy selected ML model	0.755	0.804	0.986	0.909	0.990	0.981	0.992	0.995	0.956	0.988
Precision selected ML model	0.964	0.807	0.986	0.807	0	0	0	0	0	0
Recall selected ML model	0.682	0.765	1	0.765	0	0	0	0	0	0
Number of examples of the test set	428	652	71	652	999	428	1,000	1,000	1,000	652

** Test set less than 1% of the validation set
* No improvement towards the majority vote
D: Death regardless of the phases, D. RP: Death in the rearing phase, D. FP: Death in the fattening phase, D. SP: Death in the suckling phase, MWW: Reaching the minimum weaning weight, P: Pericarditis, PL1: Pleurisy < 30 %, PL2: Pleurisy > 30 %, PN1: Pneumonia < 30 %, PN2: Pneumonia > 30 %

In Table 4, the precision and recall value of zero indicates that the ML model cannot distinguish the class and therefore classifies using the majority vote “does not occur”. Table 5 shows the hyperparameters of the ML models for death in the suckling phase, death regardless of the phases, and reaching the minimum weaning weight.

Table 5: Configuration of the ML model resulting from the model selection

	D	**D. SP**	**MWW**
Biological indicators	BW, LNM, sex, TNP	BW, H, LNM, NPA, RaM, RaF, sex, TNP	BW, NPA, PWA, H, DI
Selected ML method	SVM	Logit	SVM
Hyperparameters of the selected ML method	C: 0.1 kernel: rbf	C: 1	C: 5,000 kernel: rbf

BW: Birth weight of the pig, MWW: Reaching the minimum weaning weight, D: Death regardless of the phases, DI: Daily increase in the mammalian phase, DSP: Death in the suckling phase, GNB: Gaussian naive bayes, H: Housing environment, LNM: Litter number of the mother, Logit: Logistic regression, NPA: Number of pigs born alive, NPS: Number of stillborn pigs, PWA: Planned weaning age of the pig in days, RaM: Race of the Mother, RaF: Race of the Father, SVM: Support vector machine, TNP: Total number of pigs born

For death in the suckling phase, the accuracy can be increased from 0.531 to 0.804. This increase is accompanied by a precision of 0.807 and a recall of 0.765. The increase of 0.273 in accuracy corresponds to an increase in accuracy of 51.41% compared to using the majority vote. Using the ML model increases the accuracy of the death regardless of the phases from 0.661 while using the majority vote by 0.094 to 0.755. This difference of 0.094 corresponds to an increase in accuracy of 14.22%. Thereby, we achieve a precision of 0.964 and a recall of 0.682 for the risk of death regardless of the phases. However, we cannot achieve an improvement in accuracy concerning the majority vote on the test set for the diseases, the death in the rearing phase, and the death in the fattening phase. Additionally, we achieve an improvement in accuracy by using the ML model compared to the majority vote for the risk of reaching the minimum weaning weight. However, the test set is smaller than 1% of the training set for the risk of reaching the minimum weaning weight. We do not provide the configuration of ML models achieving the same accuracy as using the majority vote. Thereby, the use of these ML models is not necessary.

The SVM used the value 0.1 for the hyperparameter "C" and "rbf" as the kernel for the risk of death regardless of the phases. Additionally, the value 1.0 was used for the hyperparameter "C" of the method logit to predict the death in the suckling phase. The corresponding used biological indicators for death regardless of the phases consists of the total number of pigs born, sex, birth weight, and litter number. Furthermore, the biological indicators used to predict death in the suckling phase consist of the number of pigs born alive, the number of stillborn pigs, the total number of pigs born, birth weight, sex, housing environment, the race of the mother, the race of the father, and the litter number. Finally, sex, the litter number, the total number of pigs born, and the race of the mother are included in the two combinations of biological indicators, which produce the highest accuracy gain across the animal welfare risks.

Discussion

We contribute ML models for the prediction of the death in the suckling phase and the prediction of the death regardless of the phases in the pig breeding. The relative increase in accuracy compared to the majority vote is 51.41% for death in the suckling phase and 14.22% for death regardless of the phases. However, the used biological indicators in the ML models do not describe any explanatory connection between the indicators and the animal welfare risks. Díaz et al. (2007) describe the influence of sex and race of the mother on animal welfare risks. As in Diaz et al., our ML model for the death in the suckling phase uses the mother's sex and race as biological indicators. The used biological indicators litter number and the total number of pigs born may describe the influence of previous litters, especially concerning the age of the sow (litter number) or the size of the current litter. In fact, it is conceivable that the total number of pigs born influences the birth weight of just one pig. Thereby, the influence of birth weight on death in the suckling phase is already known in the literature (Díaz et al. 2017).

However, our ML models could not achieve an increase in accuracy compared to the majority vote for the death in the rearing phase, death in the fattening phase, the diseases, and reaching of the minimum weaning weight. We think it is conceivable that these risks are caused by more complex influences, which were not appropriately included in our dataset, or that indicators of later phases are missing. Also, this follows our observation regarding a lower accuracy for death in the suckling phase than for death regardless of the phases. In addition, the death corresponds to a generalization of the death in the various phases. Furthermore, the hyperparameters differ between the model for death in the suckling phase and the model for death regardless of the phases.

Deaths after the suckling phase and illnesses may be influenced by more complex relationships than those included in the data of the biological indicators. Therefore, this constitutes in starting points for further research. For example, the use of other biological indicators is conceivable, like including sensor data (e.g., water flow rates, feed quantities, or housing temperature). Changing the hyperparameter values of the ML models for the same dataset could be another approach. Other approaches for future research may be to transfer the ML models to risks of other animals or to use data from other sources.

Our research has following limitations. The test set of death in the suckling phase equals 4.28% of the training set. Also, the test amount of death regardless of the phases equals 6.52% of the corresponding training set. Thus, both test sets are below the targeted 10% or 1,000 examples. Thereby, a larger test set would increase the validity of the predictive power for new examples. The smaller test set than targeted were caused by dividing the entire dataset into a training set and a test set before one-hot encoding. On the one hand, this generated descriptive statistics for both the training and the test set. On the other hand, this leads to values in categorical features that are included in the test set and not included in the training set. Therefore, we removed examples with these values from the test set. For this reason, no examples would have been removed in the case of a division according to one-hot encoding.

Another restriction results from the imbalanced class frequency of the target variables. As a result, the methods are optimized for a prediction of the majority vote since the majority vote leads to an almost errorless classification. In this way, the procedures achieve the highest accuracy, however, in the operational context, a high recall may be desirable to prevent animal welfare risks at an early stage. For this purpose, there exist other methods that can be examined in future research on animal welfare risks (Haixiang et al. 2017).

Instead of one dataset for all animal welfare risks, we used separate datasets for each animal welfare risk. However, this leads to different examples of pigs in separate datasets. This prevents a pig that died in the suckling phase from inclusion in either of the datasets of death in the rearing or fattening phase. Likewise, up to 11,000 most recent examples were used in the training and test set and not selected randomly, based on the assumption that the circumstances during the recording (e.g., the pen equipment) are closer to the actual current circumstances and therefore generating a positive effect for the forecasting performance. However, other circumstances like epidemics or feed changes are not considered, which may result in lower prediction performance.

The min-max normalization used ensures efficient use of the ML in one-hot encoding but is also vulnerable to extreme values. Therefore, we controlled every continuous biological indicator for outliers. In the case of standardization using the z-transformation, the rarity of the non-zero features would not be retained, so we used the min-max normalization. The z-transformation describes the transformation of features in such a way that they have an expected value of zero and a variance of one. The one-hot encoding of the race of the mother and the race of the father combined with the other biological indicators and the number of different animal welfare risks made the efficient use of ML necessary. Efficient concerning the required time to train all these ML algorithms on the training set. We assumed the min-max normalization is faster than the z-transformation in our setting.

Conclusions

We propose a machine learning model that increases the accuracy for predicting deaths of pigs in the suckling phase from to 80.4% compared to a 53.1% majority vote. The proposed model achieves a precision of 80.7% and a recall of 76.5%. Additionally, we propose a machine learning model for predicting deaths regardless of the life phase, increasing the accuracy compared to the majority vote from 66.1 to 75.5%. This model achieves a precision of 96.4% and a recall of 68.2%. Moreover, for predicting death in the suckling phase we used biological indicators that are already available at birth time and can be determined without special sensors. Thus, in practice, potential deaths of pigs can be predicted with high accuracy and little effort close after birth. Thus, we contribute not only new models with high accuracy but also help pig farmers to retain a good reputation and to mitigate economic risks associated with animal welfare.

References

Bergstra, J.; Bengio Y. (2012): Random Search for Hyper-Parameter Optimization., Journal of Machine Learning Research 13, pp. 281-305

Bergstra, J.; Yamins, D.; & Cox, D.D. (2013): Hyperopt: A python library for optimizing the hyperparameters of machine learning algorithms. Proceedings of the 12th Python in science conference (13), pp. 13-19, https://doi.org/10.25080/Majora-8b375195-003

Botreau, R.; Veissier, I.; Butterworth, A.; Bracke M.B.M.; Keeling, L.J. (2007): Definition of criteria for overall assessment of animal welfare. Animal Welfare 16, pp. 225-228

Breiman, L. (2001): Random Forests. Machine Learning 45, pp. 5-32

Bundesministerium für Ernährung und Landwirtschaft (2018): Landwirtschaft verstehen - Fakten und Hintergründe. www.bmel.de/SharedDocs/Downloads/Broschueren/Landwirtschaft-verstehen.pdf;jsessionid=4672B911DF-00C56AF288442A5EA518E0.1_cid367?__blob=publicationFile, accessed on 23 Dec 2018

Bundesverband Rind und Schwein e. V. (2020): LPA-Rassencodes. https://www.rind-schwein.de/brs-schwein/lparassencodes.html, accessed on 17 Apr 2020

Chung, Y.; Oh, S.; Lee, J.;Park, D.; Chang, H-H; Kim, S. (2013): Automatic Detection and Recognition of Pig Wasting Diseases Using Sound Data in Audio Surveillance Systems. Sensors 13(10), pp. 12929-12942

Cortes, C.; Vapnik, V. (1995): Support-Vector Networks. Machine Learning 20, pp. 273-297

Díaz, J. A. C.; Boyle, L. A.; Diana, A.; Leonard, F. C.; Moriarty, J. P.; McElroy, M. C.; McGettrick, S.; Kelliher, D.; Manzanilla, E. G. (2017): Early life indicators predict mortality, illness, reduced welfare and carcass characteristics in finisher pigs. Preventive Veterinary Medicine 146, pp. 94-102, https://doi.org/10.1016/j.prevetmed.2017.07.018

Farm Animal Welfare Council (2009): Farm Animal Welfare in Great Britain: Past, Present and Future. https://www.gov.uk/government/publications/fawc-report-on-farm-animal-welfare-in-great-britain-past-present-and-future, accessed on 17 Apr 2020

Haixiang, G.; Yijing, L.; Shang, J.; Mingyun, G.; Yuanyue, H.; Bing, G. (2017): Learning from class-imbalanced data: Review of methods and applications, Expert Systems with Applications 73, pp. 220-239, https://doi.org/10.1016/j.eswa.2016.12.035

Hsu, C.-W.; C.-C. Chang; Lin, C.-J. (2003): A Practical Guide to Support Vector Classification, Technischer Bericht, National Taiwan University

John, G. H.; Langley, P. (1995): Estimating continuous distributions in Bayesian classifiers, in: Hg. Besnard, P.; Hanks, S., Proceedings of the Eleventh Conference on Uncertainty in Artificial Intelligence, Montréal Qué, Morgan Kaufmann Publishers, 1. Auflage., pp. 338-345

Kohavi, R. (1995): A Study of Cross-Validation and Bootstrap for Accuracy Estimation and Model Selection. IJCAI‘95 Proceedings of the 14th international joint conference on Artificial intelligence (2), pp. 1137-1143

Lin, C.-J.; Wenig, R. C.; Keerthi, S. S. (2008): Trust Region Newton Method for Large-Scale Logistic Regression. Journal of Machine Learning Research 9, pp. 627-650

Manning, C. D.; Schütze, H.; (1999): Foundations of Statistical Natural Language Processing. Cambridge, The MIT Press

Manteuffel, G.; Schön, P.-Ch. (2002): Measuring welfare of pigs by automatic monitoring of stress sounds. Measurement Systems for Animal Data - Bornimer Agrartechnische Berichte 29, pp. 110-118

Matthews, S. G.; Miller ,A. L.; Clapp, J.; Plötz, T.; Kyriazakis, I. (2016): Early detection of health and welfare compromises through automated detection of behavioural changes in pigs. The Veterinary Journal 217, pp. 43-51, https://doi.org/10.1016/j.tvjl.2016.09.005

Mauer, J. (2014): Vergleich der Schweinemast in Stallungen der konventionellen und alternativen Bauweise. www.lszbw.de/pb/site/lel/get/documents/MLR.LEL/PB5Documents/lsz/pdf/Fachinformationen/Schweinemast%20-%20Haltung/LSZ_Vergleich%20Schweinemast%20in%20Stallungen.pdf?attachment=true, accessed on 23 Dec 2018

Mitchell, T. M. (1997): Machine Learning. New York, The McGraw-Hill Companies

Nguyen, T. T.; Armitage, G. (2008): A survey of techniques for internet traffic classification using machine learning. IEEE Communications Surveys & Tutorials 10 (4), pp. 56–76, https://doi.org/10.1109/SURV.2008.080406

Pineau, J.; Vincent-Lamarre, P.; Sinha, K.; Larivière, V.; Beygelzimer, A.; d‘Alché-Buc, F.; Fox, E.; Larochelle, H. (2020): Improving Reproducibility in Machine Learning Research (A Report from the NeurIPS 2019 Reproducibility Program). arXiv preprint, arXiv:2003.12206

Poore, K. R.; Forhead, A. J.; Gardner, D. S.; Giussani, D. A.; Fowden, A. L. (2002): The effects of birth weight on basal cardiovascular function in pigs at 3 months of age. Journal of Physiology 539(3), pp. 969-978, http://dx.doi.org/10.1013/jphysiol.2001.012926

Riekert, M.; Klein, A.; Adrion, F.; Hoffmann, C.; Gallmann, E. (2020): Automatically detecting pig position and posture by 2D camera imaging and deep learning. Computers and Electronics in Agriculture 174, p. 105391, https://doi.org/10.1016/j.compag.2020.105391

Tian, T. und Zhu, J. (2015): Max-Margin Majority Voting for Learning from Crowds. Advances in neural information processing systems, pp. 1621-1629

Witten, I. H.; Frank, E.; Hall, M. A. (2011): Data Mining: Practical Machine Learning Tools and Techniques. Amsterdam, Morgan Kaufmann Publishers, 3. Auflage

Wolpert, D. H. (1996): The Lack of A Priori Distinctions Between Learning Algorithms. Neural Computation 8(7), pp. 1341-1390

Zapf, R.; Schultheiß, U.; Knierim, U.; Brinkmann, J.; Schrader, L. (2017): Assessing farm animal welfare – guidelines for on-farm self-assessment, Landtechnik 72(4), pp. 214-220, https://doi.org/10.15150/lt.2017.3166

Authors

M. Sc. Tobias Zimpel, M. Sc. Martin Riekert and **Dr. oec. Achim Klein** are research associates at the chair for information systems at the University of Hohenheim, Schwerzstraße 35, 70599 Stuttgart.
E-Mail: tobias.zimpel@uni-hohenheim.de

Dr. Christa Hoffmann was Head of Unit at the State Institute for Pig Farming during the conduct of the study. Landesanstalt für Schweinezucht Boxberg, Seehöfer Str. 50, 97944 Boxberg-Windischbuch

Acknowledgements

The authors thank the staff at the Boxberg Teaching and Research Centre or their advice during the experiments and the provision of the dataset, especially Andrea Wild, Günter Lenkner, Lilljana Pyrkotsch, and Hansjörg Schrade.

This work was supported by the project “Landwirtschaft 4.0: Info-System”, funded by the Ministry of Rural Development and Consumer Protection of Baden-Württemberg, Germany.