Feature selection and intelligent livestock management

Alsahaf, Ahmad

DOI:
10.33612/diss.145238079

Chapter 5

Outlook and conclusions

In this thesis, we presented two applications of machine learning and computer vision in the livestock industry. The applications served as examples of two trends in the use of data in that field: first, using non-linear supervised learning for phenotype prediction or estimation; an approach that could address some of the shortcomings of the genetic-statistical animal models that are conventionally used in breeding programs; and second, using computer vision for improving farm logistics and practices, which fits within the broader trend of precision farming, and the application of IoT technology in agriculture. These ongoing developments could strongly impact the science and industry of animal breeding in the coming years.

For long, livestock breeding programs relied on population genetics, and on tried-and-tested mixed linear models, to reach their desired objectives. Those models were gradually augmented by emerging technologies in molecular genetics, like the lowering of genome sequencing costs in recent years. Albeit, the modelling approaches, and their underlying assumptions, remained largely unchanged. A useful alternative for the use of data in that sector could come through the application of algorithmic prediction models, or machine learning.

The differences between the statistical and machine learning approaches to predictive modelling were expounded by Leo Breiman [Breiman, 2001b], author of the bagging and random forest algorithms. The majority of statisticians at that period - by Leo Breiman's estimate - approached applied statistics problems by assuming that the data was generated by parametric models whose parameters were to be estimated. The models were then validated by goodness-of-fit tests and residual analysis, which often led to misleading conclusions [Breiman, 2001b].

Another approach, less popular at the time, was machine learning. In contrast to the data modelling approach, machine learning was algorithmic, validated by prediction accuracy on unseen examples, and made less a priori assumptions on the structures of input data. The dominance of the statistical data modelling approach at the time, and the aversion to algorithmic models, had led to an excess of theory, and a hindrance to progress on real-life prediction prediction challenges [Breiman, 2001b].

A lot has changed since Breiman's promotion of machine learning. Empirical successes and developments in theory have made the field more popular and well trusted in both science and industry. Nonetheless, the field still attracts critics and skeptics. And some of the critiques are perhaps justified.

Most notably, the "black box" characterization of machine learning, and more generally of artificial intelligence, often comes under scrutiny. Despite reaching impressive milestones - for example, with deep neural networks outperforming humans on several tasks - the opacity of most machine learning models has led many to doubt their utility as analytical tools, or as tools for scientific enquiry.

By not giving any insight into the inner-workings of a model, and how those relate to the phenomenon that the model emulates, an algorithm's performance on the prediction task could be considered an achievement of purely engineering nature, with no discernible scientific benefit.

This has created a demand for models that are interpretable or explainable by humans. The demand is also a consequence of machine learning being deployed in areas that affect people's personal affairs, like healthcare and credit scoring. Therefore, criteria that were once auxiliary to task performance, such as safety, privacy, and non-discrimination, have become highly desirable, if not necessary [Doshi-Velez and Kim, 2017, and references therein].

In chapter 4, we proposed a new feature selection method based on boosting, or sample re-weighting. Feature selection increases the parsimony of a model regardless of the used algorithm. This could be seen as a first step towards more interpretable models generally, since models with less variables are easier to explain. The boosting framework we proposed explicitly addresses the issue of feature redundancy: When multiple features have the same predictive power, the boosting mechanism ensures that only one of them is retained.

In our view, a pressing challenge in feature selection research is the lack of a standardized framework for performance evaluation. Feature selection problems are mostly unsupervised. The truly relevant features are often not known, so the performance on a proxy task, like the predictive ability of the selected feature subsets, is used to evaluate the performance of the feature selection algorithm. This can lead to inconsistencies and biases. For instance, if a wrapper feature selection method is validated using the same learner that is used for selection, the performance is likely to be exaggerated. Future research should focus on a framework that unifies evaluations based on proxy tasks, like prediction performance, with objective evaluations when possible, i.e., when the relevant features are known.

While the focus on interpretability could increase the trustworthiness of machine learning in general, some apprehensions towards it are domain-specific; having to do with a domain's history and its ingrained practices. The science of livestock breeding is a particularly interesting case thereof. Not only has the field had large successes with traditional statistics, but it has also been a driving force behind major developments in the statistics of the 20th century. The pioneers of population genetics, R. A. Fisher (1890-1620), Sewall Wright (1889-1988), and J. S. Haldane (1892-1865); in addition to being geneticists, were also prominent statisticians, with lasting contributions to the field [Thompson, 1990]. And given that the most practical application of population genetics was livestock breeding, the latter became strongly interlinked with traditional statistical analysis.

Thus, in order to clearly demonstrate the usefulness of machine learning algorithms to the practitioners, researchers, and shareholders of livestock breeding, more examples like the ones given in this thesis are needed. A survey of the literature shows that such studies are already taking place (Fig 5.1). In the future, still more ambitious efforts could be undertaken. This may take the form of breeding programs that are built from the ground up with the premise of using machine learning and big data. For instance, by measuring and storing phenotypic and environmental information with an even higher resolution than the current standards.

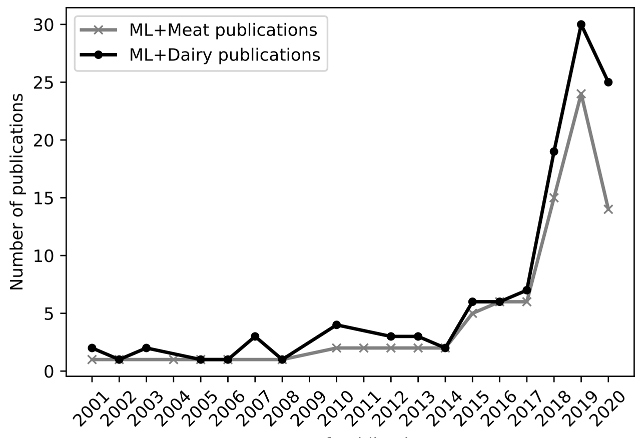

Figure 5.1: The number of publications containing the terms "machine learning" and "dairy" or "meat" in either their titles, abstracts, or keywords. Source: web of science database (July, 2020).

In chapter 2, we showed that random forest regression outperformed linear regression, a statistical linear model, in predicting a future phenotype in pigs, based on diverse types of input data (genetic and phenotypic). This framework could be adapted for predicting similar quantitative traits, in pigs and other livestock species.

For size related phenotypes in particular, like the one studied in chapter 2, the inclusion of more potentially relevant variables as predictors will likely improve prediction performance, and allow for more precise management of the animal's growth cycle. Examples of such variables are daily weight measurements and feed intake.

In chapter 3, we gave an example of estimating a different size related trait in pigs, namely, their live muscularity. The application used a combination of RGB-D computer vision and ensemble learning to provide an alternative to human evaluation of the trait. The flexibility of computer vision is likely to improve many practices in the livestock industry, related to quality assessment, farm logistics, and animal welfare.

Epilogue

In conclusion, while this thesis could be seen as an espousal of using machine learning technology in the livestock industry, this must be grounded in the reminder that the industry itself needs more than just technological tools, if it wished to be part of a prosperous future for humans.

Science and technology, along with the widely held ideology of humanity's dominion over nature, have led the livestock industry to its current mammoth proportions. Animals raised for food are massive in quantity; so much so, in fact, that they constitute a significant portion of all biological life on Earth.

A study of the biomass composition of the planet by Bar-On et al. [2018] showed that among mammals, livestock animals represent roughly 60% of the total biomass, while wild mammals represent a mere 4%. The remaining 36% are humans. Similarly, the biomass of domesticated poultry is threefold that of all wild birds.

The production of this large mass of sentient creatures ranks high among the list of anthropogenic activities that warm the climate, raise sea levels, pollute the soil, reduce biodiversity, and irreversibly deplete the planet of its energy sources [Steinfeld et al., 2006; O'Mara, 2011]. The human impact on the environment is no longer a peripheral issue, nor one that can be relegated to the fringes of ethical debate. Instead, it must be treated for what it truly is, "a threat to the perpetuation of organized human life" [Chomsky, 2019].

With that in mind, it could well be argued that the best thing the livestock industry could do going forward is to significantly curtail its own growth. And in many cases, contrary to the spirit of data-driven efficiencies, a return to traditional forms of livestock rearing may better serve people, animals, and the planet.

Parts of this thesis were written during the COVID-19 pandemic; the latest in a series of diseases caused by pathogens of animal origin. Another grim reminder of how fragile our relationship to nature is. And a reminder that instead of unbridled growth, in livestock or elsewhere, we should strive for rational and responsible custodianship of our planet and its resources.

Bibliography

Alsahaf, A., Azzopardi, G., Ducro, B., Veerkamp, R. F. and Petkov, N.: 2018a, Assigning pigs to uniform target weight groups using machine learning, *Proceedings of the World Congress on Genetics Applied to Livestock Production*, World Congress on Genetics Applied to Livestock Production, p. 112.

Alsahaf, A., Azzopardi, G., Ducro, B., Veerkamp, R. F. and Petkov, N.: 2018b, Predicting slaughter weight in pigs with regression tree ensembles., *APPIS*, pp. 1–9.

Amraei, S., Mehdizadeh, S. A. and Sallary, S.: 2017, Application of computer vision and support vector regression for weight prediction of live broiler chicken, *Engineering in agriculture, environment and food* **10**(4), 266–271.

Anaraki, F. P. and Hughes, S.: 2014, Memory and computation efficient pca via very sparse random projections, *International Conference on Machine Learning*, pp. 1341–1349.

Andrew, W., Greatwood, C. and Burghardt, T.: 2017, Visual localisation and individual identification of holstein friesian cattle via deep learning, *Proceedings of the IEEE International Conference on Computer Vision Workshops*, pp. 2850–2859.

Apichottanakul, A., Pathumnakul, S. and Piewthongngam, K.: 2012, The role of pig size prediction in supply chain planning, *biosystems engineering* **113**(3), 298–307.

Avogadri, R. and Valentini, G.: 2009, Fuzzy ensemble clustering based on random projections for dna microarray data analysis, *Artificial Intelligence in Medicine* **45**(2-3), 173–183.

Baccianella, S., Esuli, A. and Sebastiani, F.: 2009, Evaluation measures for ordinal regression, *2009 Ninth international conference on intelligent systems design and applications*, IEEE, pp. 283–287.

Bar-On, Y. M., Phillips, R. and Milo, R.: 2018, The biomass distribution on earth, *Proceedings of the National Academy of Sciences* **115**(25), 6506–6511.

Barbon, S., Costa Barbon, A. P. A. d., Mantovani, R. G. and Barbin, D. F.: 2018, Machine learning applied to near-infrared spectra for chicken meat classification, *Journal of Spectroscopy* **2018**.

Barddal, J. P., Enembreck, F., Gomes, H. M., Bifet, A. and Pfahringer, B.: 2019, Boosting decision stumps for dynamic feature selection on data streams, *Information Systems* **83**, 13–29.

Bhole, A., Falzon, O., Biehl, M. and Azzopardi, G.: 2019, A computer vision pipeline that uses thermal and rgb images for the recognition of holstein cattle, *International Conference on Computer Analysis of Images and Patterns*, Springer, pp. 108–119.

Bingham, E. and Mannila, H.: 2001, Random projection in dimensionality reduction: applications to image and text data, *Proceedings of the seventh ACM SIGKDD international conference on Knowledge discovery and data mining*, ACM, pp. 245–250.

Bolón-Canedo, V. and Alonso-Betanzos, A.: 2019, Ensembles for feature selection: A review and future trends, *Information Fusion* **52**, 1–12.

Borboudakis, G. and Tsamardinos, I.: 2019, Forward-backward selection with early dropping, *The Journal of Machine Learning Research* **20**(1), 276–314.

Boutsidis, C., Zouzias, A., Mahoney, M. W. and Drineas, P.: 2014, Randomized dimensionality reduction for k-means clustering, *IEEE Transactions on Information Theory* **61**(2), 1045–1062.

Breiman, L.: 1996, Bagging predictors, *Machine learning* **24**(2), 123–140.

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

Breiman, L.: 2001b, Statistical modeling: The two cultures (with comments and a rejoinder by the author), *Statistical science* **16**(3), 199–231.

Breiman, L.: 2002, Manual on setting up, using, and understanding random forests v3. 1, *Statistics Department University of California Berkeley, CA, USA* **1**, 58.

Cangar, Ö., Leroy, T., Guarino, M., Vranken, E., Fallon, R., Lenehan, J., Mee, J. and Berckmans, D.: 2008, Automatic real-time monitoring of locomotion and posture behaviour of pregnant cows prior to calving using online image analysis, *Computers and electronics in agriculture* **64**(1), 53–60.

Carabús, A., Martinell, M. G. and Furnols, M. F.: 2016, Imaging technologies to study the composition of live pigs: A review, *Spanish journal of agricultural research* **14**(3), 11.

Carabús, A., Sainz, R., Oltjen, J., Gispert, M. and Font-i Furnols, M.: 2015, Predicting fat, lean and the weights of primal cuts for growing pigs of different genotypes and sexes using computed tomography, *Journal of animal science* **93**(3), 1388–1397.

Carlotto, M. J.: 2009, Effect of errors in ground truth on classification accuracy, *International Journal of Remote Sensing* **30**(18), 4831–4849.

Chen, T. and Guestrin, C.: 2016, Xgboost: A scalable tree boosting system, *Proceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining*, ACM, pp. 785–794.

Chomsky, N.: 2019, *Internationalism or Extinction*, Routledge.

Christin, C., Hoefsloot, H. C., Smilde, A. K., Hoekman, B., Suits, F., Bischoff, R. and Horvatovich, P.: 2013, A critical assessment of feature selection methods for biomarker discovery in clinical proteomics, *Molecular & Cellular Proteomics* **12**(1), 263–276.

Cobzas, D., Birkbeck, N., Schmidt, M., Jagersand, M. and Murtha, A.: 2007, 3d variational brain tumor segmentation using a high dimensional feature set, *2007 IEEE 11th International Conference on Computer Vision*, IEEE, pp. 1–8.

Coroiu, A. D. C. A. and Coroiu, A.: 2018, Interchangeability of kinect and orbbec sensors for gesture recognition, *2018 IEEE 14th International Conference on Intelligent Computer Communication and Processing (ICCP)*, IEEE, pp. 309–315.

Cross, H., Gilliland, D., Durland, P. and Seideman, S.: 1983, Beef carcass evaluation by use of a video image analysis system, *Journal of Animal Science* **57**(4), 908–917.

Das, S.: 2001, Filters, wrappers and a boosting-based hybrid for feature selection, *Icml*, Vol. 1, pp. 74–81.

Dash, M. and Liu, H.: 1997, Feature selection for classification, *Intelligent data analysis* **1**(1-4), 131–156.

Deegalla, S. and Bostrom, H.: 2006, Reducing high-dimensional data by principal component analysis vs. random projection for nearest neighbor classification, *2006 5th International Conference on Machine Learning and Applications (ICMLA'06)*, IEEE, pp. 245–250.

Do, D. N., Ostersen, T., Strathe, A. B., Mark, T., Jensen, J. and Kadarmideen, H. N.: 2014, Genome-wide association and systems genetic analyses of residual feed intake, daily feed consumption, backfat and weight gain in pigs, *BMC genetics* **15**(1), 27.

Doeschl, A., Green, D., Whittemore, C., Schofield, C., Fisher, A. and Knap, P.: 2004, The relationship between the body shape of living pigs and their carcass morphology and composition, *Animal Science* **79**(1), 73–83.

Doshi-Velez, F. and Kim, B.: 2017, Towards a rigorous science of interpretable machine learning, *arXiv preprint arXiv:1702.08608* .

Dutta, R., Smith, D., Rawnsley, R., Bishop-Hurley, G., Hills, J., Timms, G. and Henry, D.: 2015, Dynamic cattle behavioural classification using supervised ensemble classifiers, *Computers and Electronics in Agriculture* **111**, 18–28.

Early, K., Fienberg, S. and Mankoff, J.: 2016, Cost-effective feature selection and ordering for personalized energy estimates, *Workshops at the Thirtieth AAAI Conference on Artificial Intelligence*.

Eklund, A.: 2012, Beeswarm: the bee swarm plot, an alternative to stripchart, *R package version 0.1* **5**.

Ferre, A. J. C., Raya, M. A., Gomez, V., Balasch, S., Chueca, J. R. M., Colomer, V. G. and Torres, A.: 2009, An automatic colour-based computer vision algorithm for tracking the position of piglets, *Spanish Journal of Agricultural Research* (3), 535–549.

Font-i Furnols, M., Carabús, A., Pomar, C. and Gispert, M.: 2015, Estimation of carcass composition and cut composition from computed tomography images of live growing pigs of different genotypes, *animal* **9**(1), 166–178.

Fonti, V. and Belitser, E.: 2017, Feature selection using lasso, *VU Amsterdam Research Paper in Business Analytics* .

Freund, Y. and Schapire, R. E.: 1997, A decision-theoretic generalization of on-line learning and an application to boosting, *Journal of computer and system sciences* **55**(1), 119–139.

Galelli, S. and Castelletti, A.: 2013, Tree-based iterative input variable selection for hydrological modeling, *Water Resources Research* **49**(7), 4295–4310.

Garrick, D.: 2010, An animal breeding approach to the estimation of genetic and environmental trends from field populations, *Journal of animal science* **88**(suppl_13), E3–E10.

Genuer, R., Poggi, J.-M. and Tuleau-Malot, C.: 2010, Variable selection using random forests, *Pattern Recognition Letters* **31**(14), 2225–2236.

Geurts, P., Ernst, D. and Wehenkel, L.: 2006, Extremely randomized trees, *Machine learning* **63**(1), 3–42.

Gil Gómez, G. L., Nybacka, M., Drugge, L. and Bakker, E.: 2018, Machine learning to classify and predict objective and subjective assessments of vehicle dynamics: the case of steering feel, *Vehicle system dynamics* **56**(1), 150–171.

Gonyou, H.: 1998, Sorting and mixing of grower/finisher pigs.

Gorczyca, M. T., Milan, H. F. M., Maia, A. S. C. and Gebremedhin, K. G.: 2018, Machine learning algorithms to predict core, skin, and hair-coat temperatures of piglets, *Computers and Electronics in Agriculture* **151**, 286–294.

Guyon, I. and Elisseeff, A.: 2003, An introduction to variable and feature selection, *Journal of machine learning research* **3**(Mar), 1157–1182.

Guyon, I., Li, J., Mader, T., Pletscher, P. A., Schneider, G. and Uhr, M.: 2006, Feature selection with the clop package, *Technical report*, Technical report.

Guyon, I., Weston, J., Barnhill, S. and Vapnik, V.: 2002, Gene selection for cancer classification using support vector machines, *Machine learning* **46**(1-3), 389–422.

Han, J., Shao, L., Xu, D. and Shotton, J.: 2013, Enhanced computer vision with microsoft kinect sensor: A review, *IEEE transactions on cybernetics* **43**(5), 1318–1334.

Hansen, M. F., Smith, M. L., Smith, L. N., Salter, M. G., Baxter, E. M., Farish, M. and Grieve, B.: 2018, Towards on-farm pig face recognition using convolutional neural networks, *Computers in Industry* **98**, 145–152.

Hara, S. and Hayashi, K.: 2016, Making tree ensembles interpretable, *arXiv preprint arXiv:1606.05390* .

Haralick, R. M., Sternberg, S. R. and Zhuang, X.: 1987, Image analysis using mathematical morphology, *IEEE transactions on pattern analysis and machine intelligence* (4), 532–550.

Hastie, T., Rosset, S., Zhu, J. and Zou, H.: 2009, Multi-class adaboost, *Statistics and its Interface* **2**(3), 349–360.

Hempstalk, K., McParland, S. and Berry, D. P.: 2015, Machine learning algorithms for the prediction of conception success to a given insemination in lactating dairy cows, *Journal of dairy science* **98**(8), 5262–5273.

Johnson, N.: 2009, A study of the nips feature selection challenge.

Johnson, W. B. and Lindenstrauss, J.: 1984, Extensions of lipschitz mappings into a hilbert space, *Contemporary mathematics* **26**(189-206), 1.

Kaiser, H. F.: 1960, The application of electronic computers to factor analysis, *Educational and psychological measurement* **20**(1), 141–151.

Kashiha, M., Pluk, A., Bahr, C., Vranken, E. and Berckmans, D.: 2013, Development of an early warning system for a broiler house using computer vision, *Biosystems Engineering* **116**(1), 36–45.

Kawasue, K., Win, K. D., Yoshida, K. and Tokunaga, T.: 2017, Black cattle body shape and temperature measurement using thermography and kinect sensor, *Artificial Life and Robotics* **22**(4), 464–470.

Kazemitabar, J., Amini, A., Bloniarz, A. and Talwalkar, A. S.: 2017, Variable importance using decision trees, *Advances in Neural Information Processing Systems*, pp. 426–435.

Kim, S., Park, S. and Choe, Y.: 2014, Effects using information system in pig production, *Adv. Sci. Technol. Lett* **49**, 206–212.

Kira, K. and Rendell, L. A.: 1992, A practical approach to feature selection, *Machine Learning Proceedings 1992*, Elsevier, pp. 249–256.

Kohavi, R. and John, G. H.: 1997, Wrappers for feature subset selection, *Artificial intelligence* **97**(1-2), 273–324.

Kongsro, J.: 2014, Estimation of pig weight using a microsoft kinect prototype imaging system, *Computers and Electronics in Agriculture* **109**, 32–35.

Kursa, M. B., Rudnicki, W. R. et al.: 2010, Feature selection with the boruta package, *J Stat Softw* **36**(11), 1–13.

Lakkaraju, H., Kamar, E., Caruana, R. and Leskovec, J.: 2017, Interpretable & explorable approximations of black box models, *arXiv preprint arXiv:1707.01154* .

Lee, J., Jin, L., Park, D. and Chung, Y.: 2016, Automatic recognition of aggressive behavior in pigs using a kinect depth sensor, *Sensors* **16**(5), 631.

Lee, W., Kim, S. H., Ryu, J. and Ban, T.-W.: 2017, Fast detection of disease in livestock based on deep learning, *Journal of the Korea Institute of Information and Communication Engineering* **21**(5), 1009–1015.

Legarra, A., Aguilar, I. and Misztal, I.: 2009, A relationship matrix including full pedigree and genomic information, *Journal of dairy science* **92**(9), 4656–4663.

Leroy, T., Vranken, E., Van Brecht, A., Struelens, E., Sonck, B. and Berckmans, D.: 2006, A computer vision method for on-line behavioral quantification of individually caged poultry, *Transactions of the ASABE* **49**(3), 795–802.

Li, H., Leung, K.-S., Wong, M.-H. and Ballester, P. J.: 2014, Substituting random forest for multiple linear regression improves binding affinity prediction of scoring functions: Cyscore as a case study, *BMC bioinformatics* **15**(1), 291.

Li, P., Hastie, T. J. and Church, K. W.: 2006, Very sparse random projections, *Proceedings of the 12th ACM SIGKDD international conference on Knowledge discovery and data mining*, ACM, pp. 287–296.

Li, R.-H. and Belford, G. G.: 2002, Instability of decision tree classification algorithms, *Proceedings of the eighth ACM SIGKDD international conference on Knowledge discovery and data mining*, ACM, pp. 570–575.

Liakos, K. G., Busato, P., Moshou, D., Pearson, S. and Bochtis, D.: 2018, Machine learning in agriculture: A review, *Sensors* **18**(8), 2674.

Liu, H., Liu, L. and Zhang, H.: 2009, Boosting feature selection using information metric for classification, *Neurocomputing* **73**(1-3), 295–303.

Loh, W.-Y.: 2011, Classification and regression trees, *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery* **1**(1), 14–23.

Loughrey, J. and Cunningham, P.: 2005, Using early stopping to reduce overfitting in wrapper-based feature weighting, *Technical report*, Trinity College Dublin, Department of Computer Science.

Louppe, G., Wehenkel, L., Sutera, A. and Geurts, P.: 2013, Understanding variable importances in forests of randomized trees, *Advances in neural information processing systems*, pp. 431–439.

Lu, Y., He, X., Wen, Y. and Wang, P. S.: 2014, A new cow identification system based on iris analysis and recognition, *International journal of biometrics* **6**(1), 18–32.

Lu, Y., Mahmoud, M. and Robinson, P.: 2017, Estimating sheep pain level using facial action unit detection, *2017 12th IEEE International Conference on Automatic Face & Gesture Recognition (FG 2017)*, IEEE, pp. 394–399.

Luckner, M., Topolski, B. and Mazurek, M.: 2017, Application of xgboost algorithm in fingerprinting localisation task, *IFIP International Conference on Computer Information Systems and Industrial Management*, Springer, pp. 661–671.

Lundberg, S. M., Erion, G. G. and Lee, S.-I.: 2018, Consistent individualized feature attribution for tree ensembles, *arXiv preprint arXiv:1802.03888* .

Ma, C., Li, Y., Yin, G. and Ji, J.: 2012, The monitoring and information management system of pig breeding process based on internet of things, *2012 Fifth International Conference on Information and Computing Science*, IEEE, pp. 103–106.

Ma, M., Meyer, B. J., Lin, L., Proffitt, R. and Skubic, M.: 2018, Vicovr-based wireless daily activity recognition and assessment system for stroke rehabilitation, *2018 IEEE International Conference on Bioinformatics and Biomedicine (BIBM)*, IEEE, pp. 1117–1121.

Maddalena, L. and Petrosino, A.: 2018, Background subtraction for moving object detection in rgbd data: A survey, *Journal of Imaging* **4**(5), 71.

Marchant, J., Schofield, C. and White, R.: 1999, Pig growth and conformation monitoring using image analysis, *Animal Science* **68**(1), 141–150.

Matthews, S. G., Miller, A. L., Clapp, J., Plötz, T. and Kyriazakis, I.: 2016, Early detection of health and welfare compromises through automated detection of behavioural changes in pigs, *The Veterinary Journal* **217**, 43–51.

Mocanu, D. C., Pokhrel, J., Garella, J. P., Seppänen, J., Liotou, E. and Narwaria, M.: 2015, No-reference video quality measurement: added value of machine learning, *Journal of Electronic Imaging* **24**(6), 061208.

Molina, L. C., Belanche, L. and Nebot, À.: 2002, Feature selection algorithms: A survey and experimental evaluation, *2002 IEEE International Conference on Data Mining, 2002. Proceedings.*, IEEE, pp. 306–313.

Morales, I. R., Cebrián, D. R., Blanco, E. F. and Sierra, A. P.: 2016, Early warning in egg production curves from commercial hens: A svm approach, *Computers and Electronics in Agriculture* **121**, 169–179.

Murauer, B. and Specht, G.: 2018, Detecting music genre using extreme gradient boosting, *Companion Proceedings of the The Web Conference 2018*, pp. 1923–1927.

Mutanga, O., Adam, E. and Cho, M. A.: 2012, High density biomass estimation for wetland vegetation using worldview-2 imagery and random forest regression algorithm, *International Journal of Applied Earth Observation and Geoinformation* **18**, 399–406.

Nguyen, T.-T., Huang, J. Z. and Nguyen, T. T.: 2015, Unbiased feature selection in learning random forests for high-dimensional data, *The Scientific World Journal* **2015**.

Nir, O., Parmet, Y., Werner, D., Adin, G. and Halachmi, I.: 2018, 3d computer-vision system for automatically estimating heifer height and body mass, *Biosystems Engineering* **173**, 4–10.

Oldenbroek, K. and van der Waaij, L.: 2014, Textbook animal breeding: animal breeding andgenetics for bsc students.

Oliveira, S., Oehler, F., San-Miguel-Ayanz, J., Camia, A. and Pereira, J. M.: 2012, Modeling spatial patterns of fire occurrence in mediterranean europe using multiple regression and random forest, *Forest Ecology and Management* **275**, 117–129.

O'Mara, F. P.: 2011, The significance of livestock as a contributor to global greenhouse gas emissions today and in the near future, *Animal Feed Science and Technology* **166**, 7–15.

Patience, J., Engele, K., Beaulieu, A., Gonyou, H. and Zijlstra, R.: 2004, Variation: costs and consequences, *Advances in Pork Production* **15**, 257–266.

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V. et al.: 2011, Scikit-learn: Machine learning in python, *Journal of machine learning research* **12**(Oct), 2825–2830.

Pegorini, V., Zen Karam, L., Pitta, C. S. R., Cardoso, R., Da Silva, J. C. C., Kalinowski, H. J., Ribeiro, R., Bertotti, F. L. and Assmann, T. S.: 2015, In vivo pattern classification of ingestive behavior in ruminants using fbg sensors and machine learning, *Sensors* **15**(11), 28456–28471.

Peker, M., Arslan, A., Şen, B., Çelebi, F. V. and But, A.: 2015, A novel hybrid method for determining the depth of anesthesia level: Combining relieff feature selection and random forest algorithm (relieff+ rf), *2015 International Symposium on Innovations in Intelligent SysTems and Applications (INISTA)*, IEEE, pp. 1–8.

Pereira, D. F., Miyamoto, B. C., Maia, G. D., Sales, G. T., Magalhães, M. M. and Gates, R. S.: 2013, Machine vision to identify broiler breeder behavior, *Computers and electronics in agriculture* **99**, 194–199.

Pereira, S., Meier, R., McKinley, R., Wiest, R., Alves, V., Silva, C. A. and Reyes, M.: 2018, Enhancing interpretability of automatically extracted machine learning features: application to a rbm-random forest system on brain lesion segmentation, *Medical image analysis* **44**, 228–244.

Petkovic, D., Altman, R. B., Wong, M. and Vigil, A.: 2018, Improving the explainability of random forest classifier-user centered approach., *PSB*, World Scientific, pp. 204–215.

Pezzuolo, A., Guarino, M., Sartori, L., González, L. A. and Marinello, F.: 2018, Onbarn pig weight estimation based on body measurements by a kinect v1 depth camera, *Computers and electronics in agriculture* **148**, 29–36.

Piccardi, M.: 2004, Background subtraction techniques: a review, *2004 IEEE International Conference on Systems, Man and Cybernetics (IEEE Cat. No. 04CH37583)*, Vol. 4, IEEE, pp. 3099–3104.

Porto, S. M., Arcidiacono, C., Anguzza, U. and Cascone, G.: 2013, A computer vision-based system for the automatic detection of lying behaviour of dairy cows in free-stall barns, *Biosystems Engineering* **115**(2), 184–194.

Post, M. J., van der Putten, P. and van Rijn, J. N.: 2016, Does feature selection improve classification? a large scale experiment in openml, *International Symposium on Intelligent Data Analysis*, Springer, pp. 158–170.

Qi, H. and Hughes, S. M.: 2012, Invariance of principal components under low-dimensional random projection of the data, *2012 19th IEEE International Conference on Image Processing*, IEEE, pp. 937–940.

Rao, H., Shi, X., Rodrigue, A. K., Feng, J., Xia, Y., Elhoseny, M., Yuan, X. and Gu, L.: 2019, Feature selection based on artificial bee colony and gradient boosting decision tree, *Applied Soft Computing* **74**, 634–642.

Ribeiro, M. T., Singh, S. and Guestrin, C.: 2016, Why should i trust you?: Explaining the predictions of any classifier, *Proceedings of the 22nd ACM SIGKDD international conference on knowledge discovery and data mining*, ACM, pp. 1135–1144.

Robinson, G. K. et al.: 1991, That blup is a good thing: the estimation of random effects, *Statistical science* **6**(1), 15–32.

Roush, W., Dozier 3rd, W. and Branton, S.: 2006, Comparison of gompertz and neural network models of broiler growth, *Poultry Science* **85**(4), 794–797.

Schofield, C.: 2007, Portable image-based pig weight monitoring systems, *Precision Livestock Farming* **7**.

Scholz, A., Bünger, L., Kongsro, J., Baulain, U. and Mitchell, A.: 2015, Non-invasive methods for the determination of body and carcass composition in livestock: dual-energy x-ray absorptiometry, computed tomography, magnetic resonance imaging and ultrasound: invited review, *Animal* **9**(7), 1250–1264.

Shahinfar, S., Mehrabani-Yeganeh, H., Lucas, C., Kalhor, A., Kazemian, M. and Weigel, K. A.: 2012, Prediction of breeding values for dairy cattle using artificial neural networks and neuro-fuzzy systems, *Computational and mathematical methods in medicine* **2012**.

Shahinfar, S., Page, D., Guenther, J., Cabrera, V., Fricke, P. and Weigel, K.: 2014, Prediction of insemination outcomes in holstein dairy cattle using alternative machine learning algorithms, *Journal of dairy science* **97**(2), 731–742.

Shao, B. and Xin, H.: 2008, A real-time computer vision assessment and control of thermal comfort for group-housed pigs, *Computers and electronics in agriculture* **62**(1), 15–21.

Sharifi, S., Pakdel, A., Ebrahimi, M., Reecy, J. M., Farsani, S. F. and Ebrahimie, E.: 2018, Integration of machine learning and meta-analysis identifies the transcriptomic bio-signature of mastitis disease in cattle, *PloS one* **13**(2).

Sloan, C., Harte, N., Kelly, D., Kokaram, A. C. and Hines, A.: 2017, Objective assessment of perceptual audio quality using visqolaudio, *IEEE Transactions on Broadcasting* **63**(4), 693–705.

Sokol, K. and Flach, P.: 2020, Explainability fact sheets: a framework for systematic assessment of explainable approaches, *Proceedings of the 2020 Conference on Fairness, Accountability, and Transparency*, pp. 56–67.

Song, Q., Ni, J. and Wang, G.: 2013, A fast clustering-based feature subset selection algorithm for high-dimensional data, *IEEE transactions on knowledge and data engineering* **25**(1), 1–14.

Stavrakakis, S., Li, W., Guy, J. H., Morgan, G., Ushaw, G., Johnson, G. R. and Edwards, S. A.: 2015, Validity of the microsoft kinect sensor for assessment of normal walking patterns in pigs, *Computers and Electronics in Agriculture* **117**, 1–7.

Steinfeld, H., Gerber, P., Wassenaar, T., Castel, V., Rosales, M., Rosales, M. and de Haan, C.: 2006, *Livestock's long shadow: environmental issues and options*, Food & Agriculture Org.

Strobl, C., Boulesteix, A.-L., Kneib, T., Augustin, T. and Zeileis, A.: 2008, Conditional variable importance for random forests, *BMC bioinformatics* **9**(1), 307.

Strobl, C., Boulesteix, A.-L., Zeileis, A. and Hothorn, T.: 2007, Bias in random forest variable importance measures: Illustrations, sources and a solution, *BMC bioinformatics* **8**(1), 25.

Tabus, I. and Astola, J.: 2005, Gene feature selection, *Genomic Signal Processing and Statistics* pp. 67–92.

Taheri-Garavand, A., Fatahi, S., Omid, M. and Makino, Y.: 2019, Meat quality evaluation based on computer vision technique: A review, *Meat science* .

Tang, J., Alelyani, S. and Liu, H.: 2014, Feature selection for classification: A review, *Data classification: algorithms and applications* p. 37.

Thanapongtharm, W., Linard, C., Chinson, P., Kasemsuwan, S., Visser, M., Gaughan, A. E., Epprech, M., Robinson, T. P. and Gilbert, M.: 2016, Spatial analysis and characteristics of pig farming in thailand, *BMC veterinary research* **12**(1), 218.

Thompson, E.: 1990, Ra fisher's contributions to genetical statistics, *Biometrics* pp. 905–914.

Tieu, K. and Viola, P.: 2004, Boosting image retrieval, *International Journal of Computer Vision* **56**(1-2), 17–36.

Trokielewicz, M. and Szadkowski, M.: 2017, Iris and periocular recognition in arabian race horses using deep convolutional neural networks, *2017 IEEE International Joint Conference on Biometrics (IJCB)*, IEEE, pp. 510–516.

Tuv, E., Borisov, A., Runger, G. and Torkkola, K.: 2009, Feature selection with ensembles, artificial variables, and redundancy elimination, *Journal of Machine Learning Research* **10**(Jul), 1341–1366.

Urbanowicz, R. J., Meeker, M., La Cava, W., Olson, R. S. and Moore, J. H.: 2018, Relief-based feature selection: introduction and review, *Journal of biomedical informatics* .

Wang, S. and Summers, R. M.: 2012, Machine learning and radiology, *Medical image analysis* **16**(5), 933–951.

Yakubu, A., Oluremi, O. and Ibrahim, Z.: 2018, Modelling egg production in sasso dual-purpose birds using linear, quadratic, artificial neural network and classification regression tree methods in the tropics, *Livest. Res. Rural Devel* **30**(10).

Yang, Y.-l., Rong, Z. and Kui, L.: 2017, Future livestock breeding: Precision breeding based on multi-omics information and population personalization, *Journal of integrative agriculture* **16**(12), 2784–2791.

Yao, C., Spurlock, D., Armentano, L., Page Jr, C., VandeHaar, M., Bickhart, D. and Weigel, K.: 2013, Random forests approach for identifying additive and epistatic single nucleotide polymorphisms associated with residual feed intake in dairy cattle, *Journal of dairy science* **96**(10), 6716–6729.

Yu, R., Leung, P. and Bienfang, P.: 2006, Predicting shrimp growth: artificial neural network versus nonlinear regression models, *Aquacultural Engineering* **34**(1), 26–32.

Zhang, Z.: 2012, Microsoft kinect sensor and its effect, *IEEE multimedia* **19**(2), 4–10.

Zhuang, X., Bi, M., Guo, J., Wu, S. and Zhang, T.: 2018, Development of an early warning algorithm to detect sick broilers, *Computers and Electronics in Agriculture* **144**, 102–113.