Farm Animals’ Behaviors and Welfare Analysis with IA Algorithms: A Review

Olivier Debauche1,2,3*, Meryem Elmoul2, Said Mahmoudi1, Jérôme Bindelle2,4, Frédéric Lebeau2,3

1 University of Mons, Faculty of Engineering - ILIA / Infortech, Place du parc 20, Mons 7000, Belgium
2 University of Liège - GxABT, Terra, Passage des dépôts 2, Gembloux 5030, Belgium
3 University of Liège - GxABT, BioDynE - DEAL, Passage des dépôts 2, Gembloux 5030, Belgium
4 University of Liège - GxABT, Precision Livestock and Nutrition, Passage des dépôts 2, Gembloux 5030, Belgium

Corresponding Author Email: olivier.debauche@umons.ac.be

https://doi.org/10.18280/ria.350308

Received: 20 April 2021
Accepted: 5 June 2021

Keywords:
animal behavior, machine learning, artificial intelligence, livestock, cow, sheep, pig, chicken

ABSTRACT

Numerous bibliographic reviews related to the use of AI for the behavioral detection of farm animals exist, but they only focus on a particular type of animal. We believe that some techniques were used for some animals that could also be used for other types of animals. The application and comparison of these techniques between animal species are rarely done. In this paper, we propose a review of machine learning approaches used for the detection of farm animals’ behaviors such as lameness, grazing, rumination, and so on. The originality of this paper is matched classification in the midst of sensors and algorithms used for each animal category. First, we highlight the most implemented approaches for different categories of animals (cows, sheep, goats, pigs, horses, and chickens) to inspire researchers interested to conduct investigation and employ the methods we have evaluated and the results we have obtained in this study. Second, we describe the current trends in terms of technological development and new paradigms that will impact the AI research. Finally, we critically analyze what is done and we draw new pathways of research to advance our understanding of animal’s behaviors.

1. INTRODUCTION

With the increase of the world population, the global demand for various meat and animal products will increase by over 70% in the next 30 years [1]. Improving our production systems has become crucial to produce more animal products with limited natural resources, particularly in terms of soil and water. In addition to that, the increase in herd size is hindering the detection of sick animals. Thanks to sensors and massive data collection, it is now possible to detect individual changes in animal behavior in terms of feeding, fluid intake, usual body movements in pigs and sheep [1], or lameness in cows [2]. Environmental parameters can also be responsible for diseases like air quality, which predicts the onset of Coccidiosis in chickens [3]. Advanced technologies like Machine Learning (ML), Deep Learning (DL), and Artificial Intelligence (AI) have emerged with Big Data technologies and High-Performance Computing, which has opened new ways of research in data-intensive science [4, 5].

It is important to distinguish concepts of ML, DL, and AI. AI is a science that builds intelligent programs and machines to solve problems usually processed by humans. ML is a part of AI that provides systems that enable automatic learning and improve itself from experience. While DL is part of ML-based on neural networks to analyze various factors with a structure that mimics the human neural system.

These techniques allow us to extract meaningful information from dataset and improve our ability to understand complex animals’ systems that integrate genetics, environmental factors, and management priorities [1]. The coupling of sensors, Big Data, and ML help farmers to detect early signs of seeing diseases such as a lethargic body, slower movements, and decrease of activity [6]. IA applications in precision livestock farming mainly target the animals’ welfare and livestock production [4]. The ML allows for example to determine the number of animals grazing sustainably on a given pasture during a specific time [1]. ML can also on basis of Inertial Measurement Unit (IMU) [7], optical sensors [8], deep video cameras [9]. It also allows to classify animals’ macro behaviors like grazing, rumination, walking, stopping, resting, and breeding events such as estrus and health events like lameness [10].

The correct selection of optimal algorithm, sampling rate, window size and sensor position are crucial to optimize the energy consumption and the autonomy of the device [11].

Generally speaking, animal’s behaviors reviews focus only on one kind of animals but developed approaches could be transferred or adapted to other farms’ animals. Our motivation is to archive a transversal review to inspire researchers in terms of methodologies used in different categories of farms’ animals. This literature review describes machine learning and artificial intelligence algorithms used to identify behaviors of farm’ animals (cows, sheep, goats, pigs, and chickens).

Next sections of this paper are structured as follow: In section 2, we summarize recent papers about animals’ behaviors. Then, we develop challenges and opportunities in section 3. Afterwards, we present advanced technologies and argue the challenges and opportunities about them. In section 4, we discuss issues about used models. Finally in section 5, we conclude this review and draw perspectives and possible future applications using AI.
2. ANIMALS’ BEHAVIORS

In this section, we summarize our works describing the applications using distributed ML or IA according to different categories of farm animals (cows, sheep, goats, pigs, horses, and chickens). This review has been archived based on publications from 2016 to 2021 published on Google Scholar, Scopus, Science Direct, IEEE Xplore. Queries achieved are (cows, sheep, goats, pigs, horses, or chicken), (welfare or behavior), and (deep learning or machine learning).

2.1 Cows

Macro behaviors involve a significant movement of several parts of the body and are subdivided in individual behaviors (grazing and/or rumination), and social behaviors visible in cattle such as bulling during estrus, aggression, or domination. The understanding of behaviors helps farmers to verify the welfare state and the early detection of pathology symptoms or injury. Table 1 summarizes Macro behaviors of previous studies found in the literature.

Table 1. Macro behaviors of previous studies

<table>
<thead>
<tr>
<th>Behaviors</th>
<th>Methods</th>
<th>Accuracy</th>
<th>Sensitivity</th>
<th>Specificity</th>
<th>Precision</th>
<th>F-Score</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>Grazing</td>
<td>DT</td>
<td>91.0%</td>
<td>91.1%</td>
<td>90.9%</td>
<td>93.5%</td>
<td></td>
<td>[12]</td>
</tr>
<tr>
<td>Ruminating</td>
<td></td>
<td>96.5%</td>
<td>53.1%</td>
<td>99.4%</td>
<td>84.5%</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Other behaviors</td>
<td></td>
<td>87.6%</td>
<td>87.6%</td>
<td>87.5%</td>
<td>79.1%</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Grazing</td>
<td>RFA, LOOA, SCV</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>91.4%</td>
<td>[13]</td>
</tr>
<tr>
<td>Standing</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>89.0%</td>
<td></td>
</tr>
<tr>
<td>Ruminating</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>93.2%</td>
<td></td>
</tr>
<tr>
<td>Out of pen milking</td>
<td>DT</td>
<td>94.2%</td>
<td>95.6%</td>
<td>94.0%</td>
<td>59.9%</td>
<td></td>
<td>[14]</td>
</tr>
<tr>
<td>Non-feeding</td>
<td></td>
<td>80.8%</td>
<td>74.9%</td>
<td>91.3%</td>
<td>93.9%</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Feeding</td>
<td></td>
<td>83.2%</td>
<td>65.3%</td>
<td>93.0%</td>
<td>83.5%</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Grazing Nesting Grazing</td>
<td>RF</td>
<td>83.0%</td>
<td></td>
<td></td>
<td></td>
<td>77.0%</td>
<td>[15]</td>
</tr>
<tr>
<td>Jrip</td>
<td></td>
<td>85.0%</td>
<td></td>
<td></td>
<td></td>
<td>76.0%</td>
<td></td>
</tr>
<tr>
<td>3 Behaviors</td>
<td></td>
<td>100%</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>[16]</td>
</tr>
<tr>
<td>Walking</td>
<td>FMM DT</td>
<td>98%</td>
<td>96%</td>
<td>99%</td>
<td>91%</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Stationary</td>
<td></td>
<td>99%</td>
<td>99%</td>
<td>96%</td>
<td>99%</td>
<td></td>
<td></td>
</tr>
<tr>
<td>3 behaviors</td>
<td></td>
<td>99.2%</td>
<td>100.0%</td>
<td>100.0%</td>
<td></td>
<td></td>
<td>[17]</td>
</tr>
<tr>
<td>6 behaviors</td>
<td>XGB</td>
<td>97.0% / 98.0%</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>[18]</td>
</tr>
<tr>
<td></td>
<td>RF</td>
<td>97.0% / 97.0%</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>SVM</td>
<td>96.0% / 97.0%</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>ADA</td>
<td>95.0% / 95.0%</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>7 behaviors</td>
<td>GHD</td>
<td>86.3%</td>
<td></td>
<td>80.6%</td>
<td>88.8%</td>
<td>86.1%</td>
<td>[19]</td>
</tr>
<tr>
<td>Ruminating</td>
<td>SVM</td>
<td>83.2%</td>
<td>89.2%</td>
<td>88.8%</td>
<td>86.1%</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Eating</td>
<td>LADA</td>
<td>72.4%</td>
<td></td>
<td>75.0%</td>
<td></td>
<td></td>
<td>[20]</td>
</tr>
<tr>
<td>Drinking</td>
<td></td>
<td>76.6%</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Chewing</td>
<td></td>
<td>71.6%</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Walking</td>
<td></td>
<td>76.3%</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Social Behavior</td>
<td></td>
<td>76.7%</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Self-grooming</td>
<td></td>
<td>75.0%</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Other</td>
<td></td>
<td>77.0%</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Note: 1. Accuracy before smooth / accuracy after smooth. 2. Recall.

Andriamandroso et al. evaluated the performance of IMU of iPhone 5s placed on the cow neck and proposed a Decision Tree (DT), which detect grazing behavior with an accuracy of 91%, sensitivity of 91.1%, specificity of 90.9%, and precision of 93.5%. Ruminination is predicted with an accuracy of 96.5%, a sensitivity of 53.1%, a specificity of 99.4%, and a prediction of 84.5%; while other behaviors are identified with an accuracy of 87.6%, a sensitivity of 87.6%, a specificity of 87.5%, and a prediction of 79.1% [12]. Rashman et al. have studied the impact of the position of accelerometer/ magnetometer (ear tag, collar (under neck), and halter) on the classification accuracy of grazing, standing, and ruminating. The accelerometer/magnetometer sample at 30 Hz for ear tag and halter, and the 3D accelerometer on collar at 12 Hz. The Random Forest Algorithm (RFA) was used and tested with Leave-Out-One-Animal (LOOA) and Stratified Cross Validation (SCV) approaches. Indeed, results show that halter with Stratified Cross Validation (SCV) F-Score are better with values of 91.4%, 89%, and 93.2% for grazing, standing and ruminating behaviors respectively [13]. Barker et al. evaluate a decision tree based on an accelerometer sampled at 12.5Hz and a position to classify on one hand behaviors (out of the pen for milking, non-feeding, and feeding); and on the other hand, (lame and non-lame). The window size used for the analysis was 2s. They obtained for the behavior classification performance for out of the pen for milking (accuracy: 94.2%, sensitivity: 95.6%, specificity: 94.0%, and precision: 59.9%). While parameters of non-feeding behavior performances are accuracy of 80.8%, precision: 93.9%, specificity: 91.3% and a sensitivity of 74.9%. Feeding behavior classification performances are accuracy of 83.2%, precision of 83.5%, specificity of 93%, and sensitivity of 65.3%. Moreover, they show that lame cows feed for less time in the afternoon and in total over a full day [14]. Williams et al. have combined data mining to extract features and 4 ML algorithms (Naï ve Bayes, JRip, J48, and Random Forest) to classify GPS data, sampled at 0.2 Hz, in grazing, resting, and walking behaviors. The evaluation was achieved with 10-fold cross-validation. The best classifiers were JRip and Random Forest with respectively an average accuracy of 85% and 83%, and F-
measure of 76% and 77% respectively [15]. Achour et al. classified 7 behaviors from IMU data sampled at 80 Hz and placed on the back of the cow. The classification model is based on univariate and multivariate Finite Mixture Models (FMM) and DT. First, the proposed algorithm identified lying on the left and on the right side, standing behavior, and changing between these behaviors with an accuracy of 100%. Second, walking behavior is classified with an accuracy of 98%, a sensitivity of 96%, a specificity of 99%, and a precision of 91%. Third, stationary behavior is classified with an accuracy of 99%, a sensitivity of 99%, a specificity of 96%, and a precision of 99% [16]. Brennam et al. have developed a collar coupling at low-cost GPS with a fix recorded each 1 minute and a 3D-accelerometer sampled at 12Hz. Moreover, accelerometric data are aggregated, and statistic parameters are calculated for each 1s interval. The previous authors also compared performances of 4 classification algorithms (RF, LDA, QDA, and SVM) to identify grazing or non-grazing behaviors. Based on their study, the best classifiers are RF and SVM. RF outperforms slightly SVM when it is trained on many data [22]. Tamura et al. used 12bit and 3D-accelerometer sampled at 20Hz with Classification and Regression Tree (CART) algorithm to classify eating, rumination, and lying behaviors. They obtained an accuracy of 99.2%, and 100% of sensibility and specificity [17]. Riaboff et al. compared the performance of behaviors’ classification (rumination-lying, resting-lying, resting-standing, rumination-standing, walking, and grazing) with xExtreme Gradient Boosting (XGB), RF, SVM, and Adaboost (ADA) with a window size of 10s. Hence, the results obtained from their analysis were reassessed on the temporal structure within the sequence of behaviors after smoothing with a Hidden Markov Model (HMM)-based Viterbi algorithm. Accuracies obtained before and after smoothing are XGB (97% and 98%), RF (97% and 97%), SVM (96% and 97%), and ADA (95% and 95%) respectively. XGB offers the best performances on all behavior’s classification except resting/standing where SVM is better [18]. Khanh et al. have evaluated 4 ML algorithms: Gradient Boosted Decision Tree (GBDT), SVM, RF, and KNN to classify 7 cow behaviors (feeding, lying, standing, lying down, standing up, normal walking, and active walking). Data were acquired with 3DOF accelerometer placed on cow leg and configured at a rate of 1Hz. GBDT provides better performance in terms of overall accuracy with 86.3% and sensibility with 80.6% for a window size of 16s [19]. Vanrell et al. have experimented with many variants of regularity-based acoustic foraging activity recognition (RAFAR) to segment foraging activities. Best average F1 scores are obtained with the gap merging before classification and partition of long block variant (RAFAR-MBBP) are respectively for activity segmentation (frame-based: 96.2%, block-based: 71.5%), rumination classification (frame-based: 89.1%, block-based: 87.3%), grazing classification (frame-based: 93.5%, block-based: 85.2%) [23]. Ayadi et al. compared performance of VGG16, VGG19, and ResNet152V2 for rumination detection. The best performance was obtained with VGG16 with an accuracy of 98.12% and a mean recall and precision of 98% [24]. Hamilton et al. used a bolus equipped of a real-time and 3D accelerometer/gyroscope are configured at 12.5Hz. A linear Support Vector Machine (SVM) model was implemented to detect rumination behavior. Consequently, the performance obtained from the study of Hamilton et al. are an accuracy of 83.2%, a recall of 89.2%, a specificity of 88.8%, and a F1 score of 86.1% [20]. Shen et al. have studied rumination characterization from the change of noseband pressure. Accuracies obtained respectively for the number of ruminations, the duration of rumination, and the number of cuds are 100%, 94.2%, and 94.45% [25]. Rodriguez-Baena et al. have suggested LADA an algorithm that determines the windows timeframe of behaviors in two steps: activity’s classification and detection. The Gap Threshold is a sensibility parameter that determine the number of false positives tolerated during a window frame. The best accuracies are obtained with a GapThreshold of 3, which are successive of eating: 72.4%, drinking: 76.6%, chewing: 71.6%, walking: 76.3%, social interaction: 76.7%, self-grooming: 75%, and other: 77% [21].

Micro behaviors are discreet movements of the body such as tail, eyes, ears, or jaws.

Chelotti et al. have admitted that only acoustic monitoring can distinguish jaws movements: chews, bites, and chew-bites. They present Chew-Bite Intelligent Algorithm (CBIA) based on patterns of recognition and ML. This algorithm achieves recognition of the 3 previously mentioned behaviors with an accuracy of 90.74%, a recall of 92.57%, and a precision of 92.21% in combining Empirical Mode Decomposition (EMD) with Support Vector Machine (SVM). On the contrary, the Least Mean Square filter (LMS) associated with Multiayer Perceptron (MLP) that offers a better compromise between recognition rate and computational cost [26]. Shen et al. have used a 3DOF-accelerometer sampled at 5Hz to monitor, and other behaviors. Thus, 3 algorithms were evaluated KNN, SVM, and Probabilistic Neural Network (PNN). Their study shows the best performances are obtained with KNN and segment length of 256. Results for feeding and rumination classification are (accuracy: 92.8%, recall: 95.6%, specificity: 96.1%) and (accuracy: 93.7%, recall: 94.3%, and specificity: 97.5%) [27].

Production parameters such as estrus, calving are special moments in the life of animals that require more attention from the breeder.

Wang et al. have deployed accelerometric and location data to detect estrus (heat). The Back Propagation Neural Network (BPNN) with a window size of 30 minutes provides the best results with respectively an accuracy of 95.36%, a sensitivity of 99.36%, a specificity of 53.33%, a precision of 95.76%, and a F1 score of 97.51% [28]. Keceli et al. proposed an automated solution based on Bi-directional Long Short-Term Memory (Bi-LSTM) to predicting accurately calving days. While RusBoosted Tree classifier allows to predict the remaining 8h before calving. The results they have gotten are respectively for BiLSTM and RusBoosted Tree classifier, with an accuracy of 83.34% and 84.16%, a sensitivity of 81.9% and of 80.51%, and finally a specificity of 98.72% and 85.74% [29]. Other researchers like Shahriar et al. have used 3D-accelerometer sampled at 10Hz and attached to a collar to detect the heat from high activity index derived from time series by means of k-means algorithm. The sensitivity is 100%, overall accuracy lies from 82% to 100% while the specificity lies between 82% to 100% [30]. Higaki et al. evaluated performances of DT, SVM, and Artificial Neural Network (ANN) for estrous detection from measure of vaginal temperature and conductivity. ANN algorithm performs with a sensitivity, a precision and F1-score equal to 0.94 [31].

The Body Condition Score (BCS) evaluates the nutritional status of dairy cow and is closely associated with health and breeding management [32].

Rodriguez Alvarez et al. applied a CNN based on
SqueezeNet where the 3 channels are: (1) The depth rescaled on 8 bits (0 to 255 value); (2) The image processed with a discrete Fourier Transform (FT) and followed by a high pass filtering and an inverse FT; (3) The body contour obtained with the Canny algorithm. Their best results were obtained with an error range of 0.5; in addition to that, the model using depth and contour channels. The accuracy, the recall, and the F1-score were all at 97% [33]. Shigeta et al. called out a Kinect V2 (Microsoft) to acquire point cloud which was then converted into 2D grayscale image. CaffeNet, a network based on AlexNet predicts BCS from images. The average accuracy obtained was 89.1% (97.5% with an error range of 0.5), a precision of 79.2%, a recall of 76.8%, and a F-measure of 77.7% [32].

Welfare depends on environmental conditions as heat stress but also lameness.

Heat stress impacts the health and performance of grazing animals.

Davison et al. used a neck-mounted temperature and humidity sensor from which the Temperature Humidity Index was calculated in order to detect signs of heat stress [34].

The lameness is an abnormal gait due to painful foot or limb lesions [35], which impacts the milking production that leads to weight loss caused by a reduction in feed intake [2, 34], reduces fertility [35], and increases risk of injury [1]. It is costly for dairy farmers in terms of time, veterinary expenditures, medication and treatment, and loss of production.

Barwick et al. evaluate tri-axial accelerometer (sampled at 12 Hz) ability to classify with QDA sheep behaviors (Walking, Standing, Grazing, and Lying). The accelerometer was placed at the neck with a collar, front leg, and ear level. Ear position has given the better accuracy (walking: 99%, grazing: 97%, standing: 97%), sensitivity (walking: 96%, grazing: 92%, standing: 98%), specificity (walking:100%, grazing: 98%, standing: 95%), and precision (walking: 99%, grazing: 94%, standing: 96%). No lying was observed in ear position, the best lying accuracy was obtained in leg position with a value of 97% [36].

Taneja et al. evaluated the accuracy of several classification algorithms [Support Vector Machine (SVM), Random Forest (RF), K-Nearest Neighbors (KNN), and Decision Trees (DT)] to detect lameness. These authors found out that the K-NN provided great equilibrium between accuracy and early detection of 3 days notification with an accuracy of 87%, a sensitivity of 89.7%, and a specificity of 72.5% [2]. Alsaood et al. analyzed automatic lameness detection systems (ALDSs) relying on three types of methods or combinations of them: (1) kinematic methods based on image processing technique, pressure-sensitive walkway, accelerometer with low or high frequency data collection; (2) kinetic methods using ground reaction force systems, four-scale weighting platform, kinetic variables of accelerometers; (3) indirect methods such as thermography, feeding behavior, automatic milking system, and milk production [35].

### 2.2 Sheep

Monitoring sheep behaviors is important because it allows detection of sick animals through reduced locomotion, food intake, or social behaviors. The table below summarizes sheep behaviors of previous studies found in the literature. Table 2 summarizes sheep behaviors of previous studies found in the literature.

<table>
<thead>
<tr>
<th>Behaviors</th>
<th>Methods</th>
<th>Accuracy</th>
<th>Sensitivity</th>
<th>Specificity</th>
<th>Precision</th>
<th>F-Score</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>Walking</td>
<td>QDA</td>
<td>99%</td>
<td>96%</td>
<td>100%</td>
<td>99%</td>
<td></td>
<td>[36]</td>
</tr>
<tr>
<td>Grazing</td>
<td></td>
<td>97%</td>
<td>92%</td>
<td>98%</td>
<td>94%</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Standing</td>
<td></td>
<td>97%</td>
<td>98%</td>
<td>95%</td>
<td>96%</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Lying</td>
<td></td>
<td>100%</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Bite</td>
<td>DT</td>
<td>98.1%</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>[37]</td>
</tr>
<tr>
<td>Chewing</td>
<td></td>
<td>95.1%</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Other</td>
<td></td>
<td>95.8%</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Walking</td>
<td>KNN</td>
<td>92.93%</td>
<td></td>
<td>98.87%</td>
<td>54.56%</td>
<td>26.18%</td>
<td>[11]</td>
</tr>
<tr>
<td>Standing</td>
<td></td>
<td>78.35%</td>
<td></td>
<td>58.16%</td>
<td>79.11%</td>
<td>84.11%</td>
<td></td>
</tr>
<tr>
<td>Lying</td>
<td></td>
<td>84.25%</td>
<td></td>
<td>91.48%</td>
<td>76.03%</td>
<td>70.92%</td>
<td></td>
</tr>
<tr>
<td>Active State</td>
<td>CART</td>
<td>98.1%</td>
<td>97.4%</td>
<td>98.5%</td>
<td>96.9%</td>
<td></td>
<td>[38]</td>
</tr>
<tr>
<td>Inactive State</td>
<td>LDA</td>
<td>90.6%</td>
<td>80.7%</td>
<td>100.0%</td>
<td>100.0%</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Upright</td>
<td></td>
<td>90.6%</td>
<td>100.0%</td>
<td>80.8%</td>
<td>79.0%</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Prostate</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Grazing</td>
<td>RF</td>
<td>92%</td>
<td>93% &lt;sup&gt;1&lt;/sup&gt;</td>
<td>98%</td>
<td>96%</td>
<td>95%</td>
<td>[39]</td>
</tr>
<tr>
<td>Non-Eating</td>
<td></td>
<td>95% &lt;sup&gt;1&lt;/sup&gt;</td>
<td></td>
<td>91%</td>
<td>89%</td>
<td>92%</td>
<td></td>
</tr>
<tr>
<td>Ruminating</td>
<td></td>
<td>87% &lt;sup&gt;1&lt;/sup&gt;</td>
<td></td>
<td>97%</td>
<td>92%</td>
<td>89%</td>
<td></td>
</tr>
<tr>
<td>Foraging</td>
<td>RF</td>
<td>97.7%</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>[10]</td>
</tr>
<tr>
<td>Walking</td>
<td></td>
<td>91.3%</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Running</td>
<td></td>
<td>90.0%</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Standing</td>
<td></td>
<td>80.5%</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Lying</td>
<td></td>
<td>100.0%</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Urination</td>
<td></td>
<td>72.2%</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Grazing</td>
<td>MLP, RF,</td>
<td>97.66%</td>
<td></td>
<td>97.74%</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Lying</td>
<td>XGB, and</td>
<td>93.22%</td>
<td></td>
<td>99.76%</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Biting</td>
<td></td>
<td>95.70%</td>
<td></td>
<td>99.74%</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Standing</td>
<td>KNN</td>
<td>97.32%</td>
<td></td>
<td>98.50%</td>
<td></td>
<td></td>
<td>[40]</td>
</tr>
<tr>
<td>Walking</td>
<td></td>
<td>96.23%</td>
<td></td>
<td>99.53%</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Note: 1. Recall

Table 2. Sheep behaviors of previous studies
100% [36]. Alvarenga et al. discriminated biting and chewing behavior by means of 3D-accelerometer (25 Hz) attached on the underside of the halter positioned on the under-jaw of the sheep. The classification algorithm was a decision tree algorithm whose parameters have been calculated over time intervals of 5s. They respectively obtained an accuracy of 98.1%, 95.1%, and 95.8% for bite, chewing and other behaviors [37]. Vázquez-Diosdado et al. combined an offline KNN algorithm with an online k-means algorithm applied with a common time window and an online algorithm based on decision rules and two prior output to produce classification labels. Their approach aims to address the non-stationarity of the learning problem on long term. Performances obtained are for walking (accuracy: 92.93%; specificity: 98.87%; recall: 17.22%; precision: 54.56%; F-score: 26.18%), standing (accuracy: 78.35%; specificity: 58.16%; recall: 89.79%; precision: 79.11%; F-score: 84.11%), and lying (accuracy: 84.25%; specificity: 91.48%; recall: 66.45%; precision: 76.03%; F-score: 70.92%) [11]. Fogarty et al. benchmarked performances of Classification and Regression Trees (CART), SVM, Linear Discriminant Analysis (LDA), and Quadratic Discriminant Analysis (QDA) on sheep behavior classification with epochs of 5s, 10s, and 30s. Accelerometric data were collected at 12.5Hz and attached to the ear tag. The best result obtained with SVM with 10s epoch for the classification of grazing, lying, standing, and walking behavior, and an accuracy of 76.9%. While for classification between active/inactive states CART with 30s performs with 98.1% of overall accuracy and for active state (sensitivity: 97.4%, specificity: 98.5%, precision: 96.9%) and for inactive state (sensitivity: 98.5%, specificity: 97.4%, precision: 98.6%). The highest prediction rate of upright or prostrate posture was obtained with LDA with 30s epoch and overall accuracy of 90.6% and for upright (sensitivity: 80.7%, specificity: 100%, precision: 100%) and for posture prostate (sensitivity: 100%, specificity: 80.8%, precision: 79.0%) [38]. Mansbridge et al. compared the performances of RF, SVN, KNN, and adaptive boosting (Adaboost) for the classification of grazing and rumination behaviors with a window size of 7s, from data collected by means of an accelerometer/gyroscope sampled at 16Hz and placed in two locations such as the ear and the collar. The overall accuracy of 92% was obtained with the sensor placed at collar level with RF algorithm. Performances obtained for grazing are precision: 96%, recall: 93%, F-score: 95%, specificity: 98%; for non-eating behavior are precision: 89%, recall: 95%, F-score: 92%, specificity: 91%; for rumination are precision: 92%, recall: 87%, F-score: 89%, and specificity: 97% [39]. Kuznicka and Gburzyński used data from 3D-accelerometer sampled at 140Hz to predict lamb sucking (a series of rapid, sharp, and jerky movements). The developed method detects the sucking of ewes by lambs with an accuracy of 95% [41]. Lush et al. utilized a RF algorithm to identify (foraging, walking, running, standing, lying, and urination) behaviors from 3DOF accelerometer sampled at 40 Hz. They obtained with a window size of 5s except for urination with a window size of 10s. The inspection showed the best accuracies of 97.7%, 91.3%, 90.0%, 80.5%, 100%, and 72.2% for foraging, walking, running, standing, lying and urination respectively [10]. Klethous et al. evaluated performances of 4 classifiers MLP, RF, XGB, and KNN to classify grazing, lying, scatching/biting, standing and walking. The highest incomes were obtained with RF with an overall accuracy of 96.47%, a sensitivity of 97.66% and a specificity of 97.74% for grazing; a sensitivity of 93.22% and of 95.70%, and a specificity of 99.76% and of 99.74% for scratching or biting; a specificity of 97.32%, and a sensitivity of 98.50% respectively for standing; a sensitivity of 96.23%, and a specificity of 99.53% for walking [40].

Welfare / Health is often relied at early detection of lameness. Barwich et al. proposed to use 3D-accelerometer sampled at 12Hz placed on ear, collar, and leg to detect lame locomotion with Quadratic discriminant analysis (QDA) with respective accuracies of 82%, 35%, and 87%; sensitivity of 82%, 35%, and 87%; specificity of 99%, 90%, and 98%; precision of 82%, 35%, and 87% [42]. Noor et al. analyzed the performances of VGG16, ResNet50, DenseNet201, GoogleNet, DarkNet, InceptionV3, and AlexNet to identify sheep in pain from face images. The best accuracy is obtained with VGG16 with 100% of accuracy, precision, and F1 score [43]. Fuentes et al. coupled infrared thermal to measure skin temperature and RGB videos to assess Heat Rate (HR) and respiration rate (RR). A first model based on parameters extracted with RVAm from RGB Video to train a Bayesian Regularization algorithm to classify RR in low, medium, high frequencies. Furthermore, classified data were reanalyzed by three different functions and a second model using Bayesian Regularized algorithm employed these inputs to predict HR. These models performed an accuracy of 85% and 84% respectively [44].

2.3 Goats

Welfare / Health of goats is achieved through the analysis of nutrition behaviors.

Rao et al. presented a welfare monitoring for goats using IoT and ML to automatically classify and quantify behaviors. The faster R-CNN to localize goats (resting and walking), and recognition of eating and drinking behavior based on the part of the area beyond food and water lines [45].

Behavior of herds of goats is often carried out using drones. Sakai et al. exploited a 9-DOF IMU where the accelerometer and gyroscope sample at 100Hz while the magnetometer samples at 2Hz and placed behind withers. They studied the effect of imbalanced datasets on DT and KNN to classify lying, standing, and grazing behaviors. They showed that a magnetometer in addition to an accelerometer are useful and improve accuracy. Best global accuracy obtained respectively for KNN and DT are 81% and 87%. KNN performances for each behavior are (precision: 91%, sensitivity: 91%, F1-score: 91%) for lying, (precision: 47%, sensitivity: 61%, F1-score: 53%) for standing, (precision: 90%, sensitivity: 83%, F1-score: 86%) for grazing respective. While DT performances are (precision: 95%, sensitivity: 93%, F1-score: 94%) for lying, (precision: 69%, sensitivity: 49%, F1-score: 57%) for standing, (precision: 88%, sensitivity: 95%, F1-score: 91%) for grazing. After sampling of data, overall accuracy obtained are 79% and 84% [46]. Jiang et al. implemented YOLOv4 to detect behaviors of group houses. Accuracies obtained are 98.87%, 98.27%, 96.86%, and 96.92% for eating, drinking, active and inactive behaviors at 17fps respectively [47]. Bocaj et al. tested performances of 7 ConvNets to classify standing, walking, trotting, running, and eating behaviors from data collected with a 3D accelerometer and gyroscope placed on the neck of the animals and sampled at 100 Hz. The best ConvNet is composed of 4 layers. The three first layers are composed of 16 filters of 1x15, 25 filters of 1x11, 32 filters of 3x7 respectively, each one is followed by a ReLu activation, a striped 1D max pooling with a size of 1x4.
and a dropout of 0.5. The fourth layer contains a dense layer with the number of classes followed by a Softmax function [48]. Wang et al. proposed a performant goat detection based on an improved Faster R-CNN from surveillance video. The proposed method is twice quicker than Faster R-CNN with an accuracy of 92.49% [49].

2.4 Pigs

Behaviors classically studied for pigs are standing, lying, mounting, and aggressivity.

Zhang et al. proposed two-stream pigs behaviors recognition based on RestNet101, which classify feeding, lying, walking, scratching, and mounting behaviors with a global accuracy of 98.99% [50]. Li et al. used Mask R-CNN, an extension of Faster R-CNN to segment pigs followed by kernel-extreme learning machine to detect mounting behavior with an accuracy of 91.47%, a sensitivity of 95.2%, and a specificity of 88.34% [51]. Li et al. used SlowFast network architecture (PMB-SCN) to classify feeding, scratching, mounting, lying, motoring with an accuracy of 96.35% [52]. Chen et al. used a VGG16 to extract spatial features which are input of LSTM to extract temporal features to identify aggressive behaviors with an accuracy 97.2% [53]. Abozar et al. combined a region-based fully convolutional network (RCF) with ResNet101 to detect standing, lying on side, and lying on belly postures. The best accuracies were obtained with a learning rate of 0.003, respectively 93% for standing, 95% of lying side, and 92% for lying on belly [54]. Yang et al. proposed an algorithm based on ZF-net to extract features, a regional proposed network and a Faster R-CNN algorithm to recognize feeding behavior of pigs with a precision of 99.6% and a recall of 86.93% [55]. Chen et al. coupled a ResNet50 which extracts spatial features with a LSTM, which extracts temporal features and fully connected layer with Softmax function classify drinker and drinker players on video. Results obtained are accuracy: 87.2%, sensitivity: 84.9%, specificity: 89.5%, and precision: 89% for body and accuracy: 92.5%, sensitivity: 91.2%, specificity: 93.8%, precision: 93.6% for head [56]. Alameer et al. described a method to distinct between feeding and non-nutritive visit based on GoogLeNet architecture with gray images with an accuracy of 99.4% [57]. Rodriguez-Baena et al. described Livestock Activity Detection Algorithm (LADA) in two steps classification of data and detecting activity temporal windows. The first step consists in an identification of time periods where the subject is in activity. The second step extracts activity or inactivity windows. The model uses the GapThreshold, a parameter that determine the number of false positive event tolerated during the windows timeframe. Best results are obtained with a GapThreshold of 3 with respective accuracy: no social interaction (73.2%), social interaction (73.2%), exploring (74.4%), M. material (73.2%), eating (78%), drinking (73.3%), and others (73.3%) [21].

Welfare can be identified by means of pigs’ posture.

Nasirahmadi et al. used a linear SVM classifier to distinguish lateral and sternal lying postures of pigs, which are then scored. The performance of the classifier are an accuracy of 94.2%, a sensitivity of 94.4%, a specificity of 94%, and the performances for the scoring are an accuracy of 94.0%, a sensitivity of 94.5%, and a specificity of 93.4% [58]. Riekert et al. designed a deep learning system for position and posture detection based on 2D images. The pipeline is based on Faster R-CNN object detection and a Neural Architecture Search (NAS) for features extraction. The best accuracy obtained for position detection was 87.4% and for position detection with posture classification that was 80.2% [59]. Arulmozi et al. evaluated performance of multiple linear regression (MLR), multilayered perceptron (MLP), decision tree regression (DTR), and support vector regression (SVR) to predict indoor air temperature (IAT) and indoor relative humidity (IRH). RFR performs well with IAT and IRH prediction with $R^2 > 0.98$ for IAT and $R^2 > 0.93$ for IRH [60].

2.5 Horses

Behaviors of horses is particularly import of racehorses. Eerdekens et al. have trained CNN to classify 7 horse behaviors (Stand, Walk, Trot, Canter, Roll, Pow, and Flank watching). The model was trained on 400 epochs with Adam optimizer, early stopping with a patience of 60 on dataset divided 2/3 train, 1/3 test. They have shown that an accuracy of 99% can be reached with sampling at 25 Hz and an interval size of 2.1s [61]. Nunes et al. trained a RNN with a Bi-LSTM to classify chews and bites behaviors from a micro camera equipped of microphone (0-18kHz). Accuracies obtained for bite and chew are 83.93% and 88.91%, recall are 93.91% and 100% respectively, and F1 score are 88.64% and 94.13% respectively [62]. Bocaj et al. used 3D accelerometric and gyroscope data sampled at 100 Hz and magnetometer sampled at 12 Hz to classify eating, standing, and lying down behaviors. They tested 7 ConvNets composed of 4 layers and shown that ConvNet composed of 4 layers. The three first layers are composed of 16 filters of 1x15, 25 filters of 1x11, 32 filters of 3x7 respectively, each one is followed by a ReLU activation, a striped 1D max pooling with a size of 1x4 and a dropout of 0.5. The fourth layer contains a dense layer with the number of classes followed by a Softmax function [48]. Welfare of horse is impacted by the stress.

Norton et al. used a wearable sensor and an ARX model to evaluate the stress of police horses. The model performs with an accuracy of 78% and a sensitivity of 77% [63].

2.6 Chickens

Behaviors’ analysis ensures that there is no anomaly during growth time of chickens.

Li et al. developed algorithms to detect feeding objection behavior. Their proposition coupled a faster R-CNN to detect object, a tracker of bird, and an SVN-based algorithm to classify behaviors and tested them on 4 stocking density [27,29;33;39] kg/m². Object detection performances for eating bird are (precision: 97.9% to 98.9%; recall: 99.7% to 99.9%; F1 score: 98.9% to 99.4%), bird around feeder (precision: 92.5% to 94.5%; recall: 96.5% to 98.7%; F1 score: 95.1 to 95.7%). Performances obtained for behaviors classification are (precision: 92.7% to 94.6%; recall: 95.1% to 97.1%; F1 score: 94.3% to 95.3%) for walking and for other behaviors (precision: 92.6% to 97.5%; recall: 96.9% to 98.4%; F1 score: 95.4% to 97.5%) [64].

Welfare directly impacts the mortality in the chicken coop. The lameness is often linked to multifactorial causes and the consequence of reducing well-being, inducing poor growth and increased mortality.

de Alencar Nääs et al. created several decisions tree to detect lameness in broiler chickens. Results have shown that best was a binary decision tree (sound and lameness) based on velocity criterion. They obtained global accuracy of 91% (86%
for sound and 92% for lame) and recall of 84% and 94% respectively [65].

Health of chickens is directly impacted by environmental conditions. The control of these parameters allows to preventing of certain diseases. Xiao et al. used binocular vision to monitor the health of caged chickens. They implemented the Chan-Vese (CV) model improved with Region Scalable Fitting (RSF) to segment respectively the body and the head of the chickens. Average accuracies obtained for head and body detection are 91.3% and 94.6% respectively [66]. Debauche et al. have used a Gated recurrent unit (GRU) algorithm to predict the evolution of air quality in chickens coop; because it directly impacted the welfare and the emergence of disease in broiler chickens [67].

3. CHALLENGES AND OPPORTUNITIES

In this section, we present the development that pave the way to future research in particular the integration of AI algorithms in devices.

3.1 Edge AI

The increase of capabilities of microcontrollers and the development of Soc dedicated to AI trace the convergence between Edge Computing and Artificial Intelligence in Edge AI. Debauche et al., have proposed an Edge AI-IoT architecture to deploy and train AI algorithms of connected sensors [68]. The most important challenge is to ensure a limited loss in model accuracy after model optimization. The methods that can be used to optimize models are parameters like pruning and sharing, quantization, knowledge distillation, low-rank factorization, and transferred/compact convolution filters [69]. The opportunity of Edge AI is the possibility to early alerted of a potential problem in avoiding false warning.

3.2 5G-Mobile Edge Computing

The future deployment of 5G network coupled with Mobile Edge Computing (MEC) opens the fields to new applications such as the monitoring of the real-time behavior of animals. Indeed, 5G will allow collecting massive data from monitoring devices at low cost, and processing at ultra-low latency and high throughput thanks to MEC. The combination of these two technologies offers the possibilities to monitor massively animals in field and process quickly collected data to propose new services. Nevertheless, the availability of 5G in rural zone remains linked to the will of suppliers on the one hand and to the adoption of sensors using 5G by farmers on the other hand.

3.3 Federated learning

The Federated Learning (FL) coupled with edge computing allows distributing learning of artificial algorithms without the transfer of data in the cloud. It is also possible to implement a continuous learning strategy to improve the global model over time with a limited transfer of pertinent data. Finally, AI algorithms can be distributed between edge where features are extracted, and the rest of the algorithm is trained in the cloud [69]. The opportunity of the FL helps us to maintain the confidentiality of farmers’ production data while providing farmers with better models.

3.4 UAV monitoring

UAVs (drones) are means of automated and programmed monitoring of herds.

Barbedo et al. showed that NasNet Large has the best accuracy to identify cattle on UAVs images with an image size of 56x56 pixels. Performances obtained are an accuracy: 96.4%, a precision: 96.5%, a recall: 96.5% and a F1 Score: 96.5%. Nevertheless, the authors argued that Xception offers an alternative with a better training time and an accuracy slightly inferior of 95.5%, precision: 95.3%, recall: 95.3%, and F1 score: 95.5% with images of 112x112 pixels. While MobileNet is adapted for embedded devices with an accuracy of 93.7%, precision: 94.3%, recall: 93.8%, F1 score: 93.8% for an image size of 112x112 pixels [70]. Xu et al used an UAV Mavic PRO (DJI, China) and Mask R-CNN algorithm to classify and count cattle and sheep. Classification accuracies obtained are 90.4% and 93.5% for cattle and sheep respectively while counting accuracies are respectively of 94.7% and 97.3% [71]. UAVs are widely used in Smart Farming but need the transfer of data to the cloud where they are processed. The challenge is the transmission of data in rural areas where high throughput networks can be unavailable [6].

3.5 Virtual fence

Animal behaviors can also be controlled to better manage the link between animals and their environment. Marini et al. studied the impact of virtual fences on grazing behavior of sheep. Garmin TT15 and Garmin Alpha 100 were installed on each sheep, and a patented CSIRO algorithm was implemented coupling 2s audio cue and electrical stimulus and showed that sheep were able to associate the audio with the virtual fence [72]. While Lomax et al. demonstrated the feasibility of virtual fence using dairy cows [73]. The coupling of grazing and walking behavior analysis, growing plants models, animal positioning and virtual fence is an opportunity to automatically manage herds. The challenge is to determine at which moment displace the herd to avoid overgrazing and/or conserve biodiversity of pastures.

4. DISCUSSION

The major issue to address in long term monitoring is the non-stationarity of the problem [30]. Indeed, in classical supervised classification, models are trained on dataset and the assumption that data are randomly selected with the same distribution that the future data. Biologic systems are by essence dynamic, a high performance on a validation dataset do not guarantee that a model will perform on future data.

Research conducted is limited to few majors’ behaviors leaving a large part of them in favor of displacement and eating habits. For example, up to 40 different behaviors can be observed in dairy cows [74] and only 5 or 6 of them are really studied.

Moreover, most models are established on a limited number of individuals, which impacts the variability within the datasets which are used to establish these models that impacts their robustness.

Developed models often require important mathematical
resources that makes them difficult to implement on constrained devices [16].

5. CONCLUSIONS

A better understanding of the interactions between animals and their environment and the influence on their behavior improves their well-being and their state of health. Behavior changes are also early indicators of illness, the presence of injuries, or a problem in their environment. The precise identification of these behaviors is therefore crucial to ensure high monitoring quality that affects farm decision-making.

Many researchers have used machine learning algorithms to classify behaviors, but they first require the extraction of features. Convolutional Neural Network (CNN) offers the advantage to automatically extract features. Moreover, Deep Learning (DL)-based classifier algorithms provides often better accuracy than ML.

Despite recent advance, there are still practical and technical challenges in terms of computational power, energy consumption, and data transmission. These ones should be addressed to obtain a complete real-time and long-term system to monitor farm animals. Afterward, the next step should be the return of farm animals to high value pastures in terms of biodiversity. The use of a virtual fence would allow better management of the link between animals and their natural environment.

Finally, we argue that is important to collect massively farms animals’ behavior data of various species evolving in different environment to establish more robust models usable at a large scale [75]. Nowadays, experimentations are generally achieved on reduced datasets of animals which allows obtaining high accuracy because of the homogeneity of the training data. These models are very specific and for the most part cannot be used in a context other than the one in which they were established and trained. Moreover, animals evolving in research center are in optimal and controlled conditions that differ from those of large herds grazing extensively where competition between animals is more important.

ACKNOWLEDGMENT

This work is supported and partially funded by Infortech / Numediart Institutes.

REFERENCES


252


