Proceedings of the 19th Conference on Computer Science and Intelligence Systems (FedCSIS) pp. 643–648 DOI: 10.15439/2024F2064 ISSN 2300-5963 ACSIS, Vol. 39

# Towards understanding animal welfare by observing collective flock behaviors via AI-powered Analytics

Savvas Karatsiolis[1,*], Pieris Panagi[2],
Vassilis Vassiliades[3], Andreas Kamilaris[4,5]
[1]0000-0002-4034-7709
[2]0009-0007-4027-1488
[3]0000-0002-1336-5629
[4]0000-0002-8484-4256
[1,2,3,4]Cyens Centre of Excellence
1 Plateia Dimarchou Lellou, Nicosia 1016, Cyprus
[5]Department of Computer Science, University of Twente,
7522 NB Enschede, The Netherlands
Email: {s.karatsiolis, p.panagi, v.vassiliades, a.kamilaris}@cyens.org.cy
* Corresponding author

Nicolas Nicolaou[1], Efstathios Stavrakis[2]
[1]0000-0001-7540-784X
[2]0000-0002-9213-7690
Algolysis Ltd,
Archiepiskopou Makariou III 200, Lakatamia 2311,
Nicosia, Cyprus
Email: {nicolas, stathis}@algolysis.com

***Abstract*—Animal farming has undergone significant transformation and evolved from small-scale businesses to large-scale commercial ventures. While maximizing productivity and profitability has always been a major concern in animal farming, during recent years there has been an increasing rise of concern regarding the welfare of the animals. In this context, the integration of artificial intelligence (AI) technologies offers immense potential for monitoring the well-being of chickens on farms and optimizing revenue streams simultaneously. Several works have integrated AI methodologies into everyday animal farming activities. Still, very few (if any) have proposed efficient and practical solutions that may facilitate farm owners in making impactful decisions regarding their business profitability and the welfare of the animals. In this direction, we propose a non-invasive chicken farm monitoring system that relies on onfield sound and video recordings integrated with sensory data acquired from the farm. The system consists of hardware that handles data acquisition and storage, a sensory data collection system and audio/video processing AI models. The last component of the system will be an inference engine that analyzes the collected data and infers useful facts about the flock's welfare and even psychological state.**

## I. Introduction

Ensuring the well-being of chickens on a farm is of paramount importance for ethical, environmental, and economic reasons. The welfare of chickens directly impacts the quality of the final product, e.g., the number and the quality of the daily produced eggs or the meat produced. Chicken flocks living in a clean environment, being fed adequately, having space to roam and being stress-free are less susceptible to diseases and premature death [1]. In this work, we explore monitoring the well-being of the chickens on the farm with technology to understand whether the flock is under stress and identify the source(s) of stress. This enables the farm management to deal with potential stress-inducing factors, keep the flock healthy and prevent catastrophic consequences. AI-driven monitoring systems can play a crucial role in this regard by continuously assessing implicit health or behavior cues like the flock's clucking or its daily motion index [12] along with environmental conditions. By leveraging machine learning algorithms, such systems can identify potential health issues or disease outbreaks, enabling proactive measures to be taken. Good, non-invasive information sources for inferring the well-being of the animals on a chicken farm are audio and video recordings of the chicken combined with the data acquired by sensors measuring ambient pollutants and the weather conditions on the farm. Chickens are especially vocal birds and tend to express their psychological and physical state through clucking to each other which enables the monitoring of their welfare and the detection of stressful conditions. Finally, the health and welfare of the birds on a farm rely on the conditions they live in. For example, high concentrations of ammonia on a farm [11], extreme (low or high) temperatures and humidity affect the health of the chickens and should be monitored carefully.

## II. Related Work

Many research works on AI-empowered chicken farm monitoring use computer vision to detect individual animals and track their motion on the farm. For example, [5] proposes applying optical flow in video recordings of chickens for identifying early signs of infection by the pathogen Campylobacter. Along similar lines, the authors of [9] identify early warnings of footpad dermatitis and hockburn in broiler chicken flocks.

In the era of Deep Learning, several research works suggest using popular Neural Network architectures to conduct individual chicken detection from video footage collected in the field. For example, in [3] the authors use Faster R-CNN [15] architectures to develop object detection and instance segmentation models that operate on edge devices installed on the farm. They propose combining these models with the monitoring of environmental parameters for early disease detection which is the subject of future work in those papers. A similar work [19] proposes the use of the YOLOv5 architecture

[17] to detect cage-free chickens on the litter floor. Besides the provision of the technology for non-invasive inference based on video recordings, several researchers propose methodologies in the context of Precision Livestock Farming (PLF) which involves wireless and Radio Frequency Identification (RFID) sensors on the chickens. For example, the authors of [4] detect the movements of each chicken with an RFID system and classify them into active, normal, or sick, claiming they can detect sick chickens at early sickness stages before the whole flock is affected. Another interesting approach by [7] involves automated monitoring and quantification of feeding and nesting behaviors of individual hens.

Finally, animals' vocalizations can be exploited as they contain a wealth of biological information (e.g., reflecting their social interactions, communicating alarm signals and containing cues about their psychological state). In this context, animals' vocalization may be used as a welfare indicator [10]. The clucking of chickens is a form of communication within the flock; it conveys messages between the birds and can be used as a warning signal or a means of conveying messages of discomfort, stress, satisfaction and expressing social interactions among the chicken.

Our paper differs from related work in that we employ advanced AI techniques to combine multiple modalities of sensory observations (i.e., audio and video recordings) from the farm environment to enable the production of analytics.

## III. Tools and Methodology

### A. Hardware

We employ audio and video recordings of the flock to conduct inference on the condition of the animals on the farm. For data acquisition, we set up a network of sensory devices each equipped with a microphone, a camera, a speaker[1] and a local storage device (Solid State Drive-SSD) for temporarily storing collected data. The heart of the system is a centralized device that synchronizes the acquisition conducted by the sensory devices over the network via Application Programming Interface (API) calls. The centralized device also acts as a Network Access Server (NAS) and Processing Engine (PE) that runs the audio Neural Network (NN) and the motion detection algorithm on the data (audio and video) captured by the sensory devices and sent to the NAS. We designed the hardware so that important configuration parameters (e.g., acquisition interval and duration, external stimuli sound, network configuration like sensors' IP addresses, etc.) are configurable on the centralized device. The centralized device will be referred to as the Synchronization and Processing Engine (SPE). The SPE performs the following:

- Implements API calls for conducting synchronization.
- Records status/error logging from sensors' communication.
- Hosts the NAS service.
- Runs inference on the audio NN.

[1]The speakers will be used in the future for producing short sounds to assess the flock's response to external stimuli

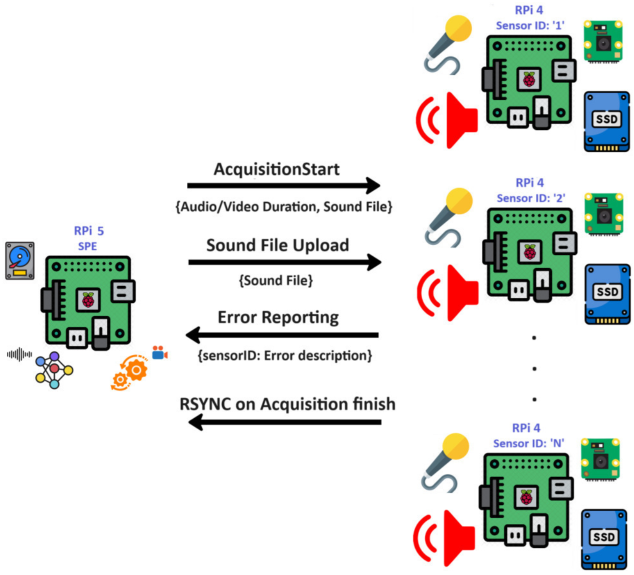


Fig. 1. The communication flow between the SPE and the sensors in the monitoring network. The SPE handles the administrative tasks and overall coordination. Every message exchange is carried out via API calls.

- Runs the motion detection algorithm.

The sensor devices capture audio/video streams in response to the synchronization messages sent by the SPE (the synchronization messages also instruct the sensors of the duration of the imminent video/audio acquisitions). For all devices in the monitoring network, we use Raspberry Pi (RPi) [13] modules because they are flexible, support audio/video acquisition and provide a fair amount of processing power. Due to the more demanding operations required by the SPE, it is built around RPi model 5 while the sensor modules are built around RPi model 4. The operational flow of the data acquisition hardware is shown in Figure 1. To efficiently handle the role of the NAS, the SPE is equipped with a 16TB Hard Disk Drive (HDD) to store several weeks' worth of data. It also supports a hot swap operation i.e., another HDD can replace the working HDD while the system is operational. This enables the operators of the monitoring system to transport the data to other premises and further process it. We will soon publish the audio/video datasets for general public use.

### B. Deployment

We built a small-scale monitoring system comprising one SPE and two sensor modules and installed it on a chicken farm located in a rural area in the district of Nicosia, Cyprus. All monitoring system devices are placed in protective enclosures (fabricated with 3D printers) to deal with the harsh environmental conditions on the farm, i.e., large concentrations of sand and particles in the air, high humidity and extreme temperatures. Figure 2 shows images of the installation on site.

We configured the system to acquire video and audio recordings of 58 seconds per minute and allow 2 seconds per minute

Fig. 2. Photos of the monitoring system.

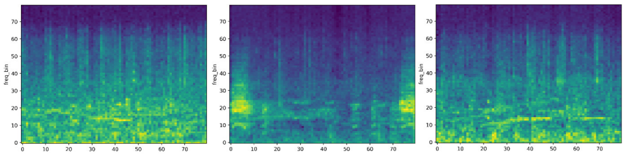

Fig. 3. The Mel spectrograms of 2-second chunks of audio streams collected on the farm. Each spectrogram shows a different frequency content in the audio files. We use these frequency representations to learn the features of the audio and infer the psychological condition of the chicken.

for system operations. We record audio data throughout the day (24 hours) and video for 12 hours per day only (from 06:00 in the morning to 6:00 p.m.) because our cameras operate in the visible light spectrum. Although chickens are mostly quiet at night, we keep recording audio during night hours to detect potential nocturnal predators' attacks. To train/develop our audio and video models, we gathered data for $50$ days which translates to a dataset comprising $75K$ audio and $37.5K$ video datapoints.

### C. Audio Anomaly Detection Model

We acknowledge the difficulty of creating an accurate dataset containing chicken vocalizations annotated with their psychological state, e.g., satisfied, calm, stressed, or panicked. Developing such a dataset requires the contribution of several experts and great effort. Alternatively, we use unsupervised learning for training the model which is much cheaper and faster and utilizes the acquired data more efficiently. Thus, the data is used to create a benchmark for the vocalization of chickens on a certain farm. Then, we exploit this benchmarked vocalization database to infer anomalies in future chicken vocalizations.

*1) Data pre-processing:* The collected audio streams are transformed into Mel spectrograms [14] and then processed by a Convolutional Denoising Autoencoder (Conv-DAE) [18]. Mel spectrograms apply a frequency-domain filter bank to audio signals that are windowed in time. We use Mel spectrograms because they offer a more perceptually relevant representation of audio signals being aligned with human sound perception. Essentially, the audio streams are transformed into frequency domain filter banks that describe the sound signal in terms of its frequency content. Some examples of the Mel spectrograms created from audio files acquired on the farm are shown in Figure 3. The Mel spectrograms provide a signature of the processed audio that reflects the psychological state of the chicken and thus enable the efficient learning of audio features in chicken clucking or any other sound made by the birds. To learn these audio features, we use unsupervised learning and thus no annotations are required. Specifically, we apply a Conv-DAE that learns to reconstruct the Mel spectrograms from their noisy versions.

Given a Mel spectrogram $x$ sampled from a dataset $X$, the model accepts an input $x' = x + \mathcal{N}(\mu, \sigma^2)$, with $\mathcal{N}$ being a Gaussian noise process, and outputs $\hat{x}$. The model learns to minimize the Mean Squared Loss (MSE) between the reconstructed spectrogram and the original spectrogram. Concretely, the optimization objective function is $min\, L = min \sum_i \frac{1}{2}(\hat{x}_i - x_i)^2, i \in [0, |X|]$. In our experiments, we use $\mu = 0, \sigma^2 = 0.25$ and compute the Mel Spectrograms on 2-second chunks of audio data.

Since each sensor module records an audio sequence, we concatenate all individual Mel spectrograms obtained from the sensor modules into one data structure that has multiple channels. The number of channels equals the number of obtained spectrograms (the number of the system's sensor modules). This technique allows using a single NN that accounts for all sound data recorded by the sensors. The main advantage of processing all sound streams concurrently instead of processing them separately is the inherent ability to combine sounds from different locations on the farm that represent the same event, i.e., the model gets recordings of the same observation from different standpoints which, in certain cases, may provide richer information. For example, a loud sound made by a chicken may overtake the signal recorded by a certain nearby microphone preventing the device from representing any other equally important sounds sourced from different locations inside the farm.

*2) Modeling:* By learning to restore the original spectrograms, the model learns the features of the problem domain and thus becomes capable of identifying the peculiarities of the data. In other words, the model learns the manifold of the data and distills the low-level audio characteristics that comprise the data [2]. To infer the flock's psychological state we observe the relative location of audio samples mapped to the benchmarked feature space on the data manifold. Since different positions on the manifold reflect different acoustic characteristics, the psychological state of the chicken(s) expressed by certain vocal characteristics can be inferred based on the mapping of the data. To create the Mel spectrograms, we first split each 58-second audio recording into 2-second chunks. We chose a 2-second audio interval because it is a good fit for providing an audio signature: it is adequate for providing sufficient vocal information while not being big enough to cause severe audio frequency shifts that may harm the processing. We process each chunk with 80 spectral banks and fast Fourier transforms of size 2220 to produce spectrograms of size 80×80.

The Conv-DAE is comprised of an encoder-decoder architecture. The encoder maps the input (noisy Mel spectrograms) to a latent space and the decoder reconstructs the input. For

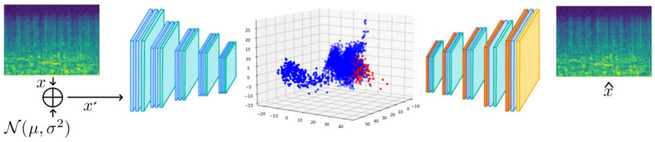


Fig. 4. The architecture of the Convolutional Autoencoder used to learn the features of the audio acquired on the farm. By learning to reconstruct noisy Mel spectrograms of recorded audio, the model builds a knowledge of the underlying elements of chickens' vocal cues.

the encoder, we use a Residual Network [6] and specifically the ResNet18 variant and for the decoder a sequence of Convolutional and Up-sampling layers. The model contains 26M trainable parameters and is shown in Figure 4.

*3) Training:* The encoder compresses the 80X80 spectrograms into a 512-D representation and the decoder reconstructs the original spectrogram by processing the 512-D vector. To train the model, we split the acquired audio dataset into a training and a testing set (80%-20% respectively) and the latter is used for assessing the model. The model training achieves a mean absolute error (MSE) of 0.126 while the test set error is 0.141. To evaluate the quality of the embeddings produced by the encoder we reduce the dimensionality of the test set embeddings from 512-D to 3-D with Principal Components Analysis (PCA) by projecting the data onto a much smaller subspace. This allows the visualization of the embedding space for evaluation purposes. We further identify the top 100 points with the highest reconstruction error and perceive them as anomalous cases: these are the data points that the model cannot adequately reconstruct which means that they are either outside the high probability regions of the distribution or they are under-represented in the dataset.

*4) Visualizations:* We exploit the low-dimensional (3-D) representations of the test set calculated by PCA to link the audio semantics of the data points with their distribution in space (left image in Figure 5). Interestingly, the data points in the test set are distributed in the 3-D latent space in such a way that data points with similar semantics are feature-mapped close to each other. For example, most of the audio data points that contain very soft clucking are mapped close to each other at a certain region of the latent space. Likewise, most of the audio data points that contain the clucking of rather stressed chicken are mapped close to each other in a certain region of the feature space. Most importantly, most of the anomalous points (data points with the highest reconstruction error) contain sounds of panicked birds that make distinct sounds of despair. Still, this 3-D feature mapping is imperfect because it does not distinguish between the data points in a definite and clear manner mostly because of the information loss during the dimensionality reduction and the confusing mix-up of chicken sounds and environmental noises (especially noises from machinery and feeding/watering equipment).

*5) Clustering:* We further cluster the 3-D embeddings with the $k$-means algorithm [8] into 5 regions. The choice of using 5 regions lies with the way the embeddings are spread onto the feature space. Figure 5 (right image) shows the 3-D embeddings of the test set and the clustering obtained with the $k$-means algorithm.

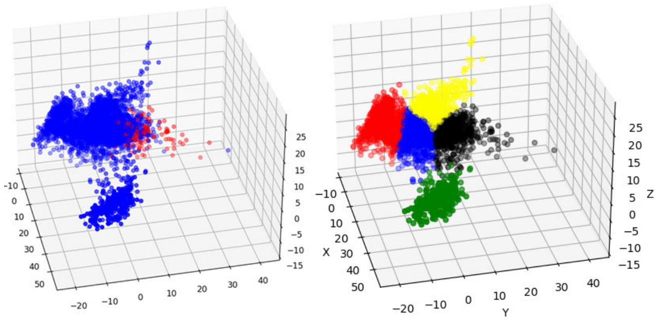


Fig. 5. Left: The 3-D embeddings of the test dataset after applying dimensionality reduction for visualization purposes. The points shown in red are perceived as anomalous points because they have high reconstruction errors. Right: The 3-D embeddings are clustered with the $k$-means algorithm. The resulting clusters shown with various colors represent different sound semantics.

The clusters calculated by the $k$-means contain semantically different sounds:

- Red cluster: Low-intensity sounds (flock resting and being very calm).
- Blue cluster: Normal soft clucking.
- Yellow cluster: Calm clucking and ambient noises (like food-delivery-machinery).
- Black cluster: Flock noises ranging from clucking of medium intensity to extremely loud flock sounds (panic sounds).
- Green cluster: Very soft clucking and ambient noise (mainly fans blowing air in the farm to cool down the flock)

Most importantly, we observe that the anomalous points (the ones with the highest reconstruction loss, shown in red color in the left image of Figure 5) are located at the extremities of the black cluster (medium to extreme noises). We provide a video that demonstrates this analysis with sound to show the different sound semantics of the various clusters at *https://sworld.cyens.org.cy*.

### D. Motion Detection

Similar to the case of audio recordings, video can also be used to detect whether the chickens in a farm are calm or under stress. The primary indicator of the stress level in video recordings is the chickens' motion: stressed chickens make rapid movements, wander around loudly and become very jerky. We detect chickens' motion with an algorithm based on background removal. Background removal or subtraction is commonly used to segment moving parts from static scenes (background and foreground). The motion is detected by subtracting the current frame of the video from a reference static background calculated by a background modeling technique which is continuously updated. One of the most popular background subtraction algorithms is the Mixture of Gaussians (MOG). According to MOG [16], for each background pixel,

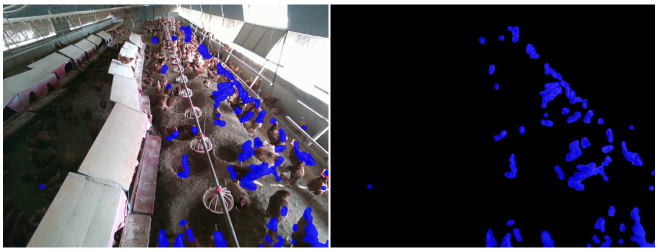

Fig. 6. Left: A single frame from a video recording from the farm. Right: The motion mask produced by MOG2.

a mixture of Gaussian distributions and a weighting parameter are utilized to "save the lifetime of pixels in the scene". Pixels with a long lifetime are interpreted as background pixels while the rest are characterized as foreground or pixels belonging to objects that move around. An improved version of MOG is called MOG2 [20] and determines the appropriate number of distributions for the modeling.

We apply MOG2 on the video streams we acquired from the sensor modules, to obtain the motion masks (images showing where motion is detected) of the frames comprising the videos. We use a relatively high video frame rate (30 frames per second) to facilitate MOG2 in detecting subtle chicken movements. Figure 6 shows the output of the motion detection process applied on a single random frame.

The motion masks produced by the MOG2 algorithm are binary images indicating whether there is motion at a certain pixel of the input frame. To generate a meaningful metric that reflects chickens' motion, we aggregate the masks by adding them together and thus compute a single number representing the magnitude of the flock's motion during the interval of the processed recording: a higher value means a more active flock. A video demonstration of the motion detection of the chicken can be found at *https://sworld.cyens.org.cy*.

Quantifying the flock's motion makes it possible to set an activity threshold which, when exceeded, could indicate that the flock is reacting to a threat or stressful condition. Symmetrically, another activity threshold could be set which, when it is not reached, could indicate that the flock is indolent and may suffer from a disease or being stressed by abnormal environmental conditions or underfeeding. Furthermore, the flock motion could be used to calculate motion statistics at different intervals (hourly, daily, weekly, etc.) and thus infer the activity level and the stress of the chicken on the farm from different perspectives.

## IV. Discussion-Conclusions

This paper introduced two methods for chicken flock monitoring, which could potentially lead to animal welfare indicators. Both the audio-based anomaly detection and the video-based motion detection methods directly provide the means to generate alerts to the farm personnel regarding the status of the flock. Besides the real-time detection of unpleasant situations, the combination of multi-modal data (audio streams, video streams and sensory data) into a unified system can be utilized to assess the welfare of the chicken. By combining the data originating from multiple modalities and collaborating with experts in the field, we expect to get reliable indicators that can then be exploited by farm owners to manage their farms better, improving the welfare of their flock, increasing revenue and preventing catastrophic events.

Analyzing the outputs of our proposed system would provide powerful analytics that humans cannot easily observe while managing the farm. To our knowledge, those analytics are not available by any commercial system to date. Some examples of the analytics we aspire to provide to the farm managers/owners in the future are the following:

- The average flock motion this week was 30% less than the average daily flock motion during last week. *This may be a sign of underfeeding, extreme temperature, or environmental pollution like high ammonia.*
- The average flock motion during the last three days was 70% less than last month's average. *This may indicate a serious condition like illness. Immediate action needs to be taken.*
- Average flock motion during feeding time is 140% higher than usual. *Maybe, this is a sign of underfeeding.*

These examples reflect our reasoning that the outputs of our system need to be translated into domain-specific analytics by experts in poultry farming.

**Funding:** The work was funded by the European Union Recovery and Resilience Facility of the NextGenerationEU instrument, through the Research and Innovation Foundation (CODEVELOP-ICT-HEALTH/0322/0061), as well as the European Union's Horizon 2020 Research and Innovation Programme under Grant Agreement No. 739578, and the Government of the Republic of Cyprus through the Deputy Ministry of Research, Innovation and Digital Policy.

## References

[1] Haitham G. Abo-Al-Ela, Seham El-Kassas, Karima El-Naggar, Safaa E. Abdo, Ali Raza Jahejo, and Rasha A. Al Wakeel. "Stress and immunity in poultry: light management and nanotechnology as effective immune enhancers to fight stress". In: *Cell Stress and Chaperones* 26.3 (2021), pp. 457–472. ISSN: 1355-8145. DOI: https://doi.org/10.1007/s12192-021-01204-6. URL: https://www.sciencedirect.com/science/article/pii/S1355814523005187.

[2] Guillaume Alain and Yoshua Bengio. "What regularized auto-encoders learn from the data-generating distribution". In: *Journal of Machine Learning Research* 15.1 (2014), pp. 3563–3593. DOI: 10.5555/2627435.2750359. URL: https://dl.acm.org/doi/10.5555/2627435.2750359.

[3] Stevan Cakic, Tomo Popovic, Srdjan Krco, Daliborka Nedic, Dejan Babic, and Ivan Jovovic. "Developing Edge AI Computer Vision for Smart Poultry Farms Using Deep Learning and HPC". In: *Sensors* 23.6 (2023). ISSN: 1424-8220. DOI: 10.3390/s23063002. URL: https://www.mdpi.com/1424-8220/23/6/3002.

[4] Zhang Feiyang, Hu Yueming, Chen Liancheng, Guo Lihong, Duan Wenjie, and Wang Lu. "Monitoring behavior of poultry based on RFID radio frequency network". In: *International Journal of Agricultural and Biological Engineering* 9.6 (2016), pp. 139–147.

[5] Colles Frances et al. "Monitoring chicken flock behaviour provides early warning of infection by human pathogen". In: *Royal Society B: Biological Sciences* 283 (1822 2016).

[6] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. "Deep Residual Learning for Image Recognition". In: *2016 IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2016, Las Vegas, NV, USA, June 27-30, 2016*. IEEE Computer Society, 2016, pp. 770–778. DOI: 10.1109/CVPR.2016.90. URL: https://doi.org/10.1109/CVPR.2016.90.

[7] Lihua Li, Yang Zhao, Jofran Oliveira, Wilco Verhoijsen, Kai Liu, and Hongwei Xin. "A UHF RFID system for studying individual feeding and nesting behaviors of group-housed laying hens". In: *Transactions of the ASABE* 60.4 (2017), pp. 1337–1347.

[8] Stuart P. Lloyd. "Least squares quantization in PCM". In: *IEEE Trans. Inf. Theory* 28.2 (1982), pp. 129–136. DOI: 10.1109/TIT.1982.1056489. URL: https://doi.org/10.1109/TIT.1982.1056489.

[9] Dawkins M.S., Roberts S.J., Cain R.J., Nickson T., and Donnelly C. A. "Early warning of footpad dermatitis and hockburn in broiler chicken flocks using optical flow, bodyweight and water consumption". In: *The Veterinary record of the British Veterinary Association* 180.20 (2017), p. 499.

[10] Gerhard Manteuffel, Birger Puppe, and Peter C Schön. "Vocalization of farm animals as a measure of welfare". In: *Applied Animal Behaviour Science* 88.1 (2004), pp. 163–182. ISSN: 0168-1591. DOI: https://doi.org/10.1016/j.applanim.2004.02.012. URL: https://www.sciencedirect.com/science/article/pii/S0168159104000565.

[11] Sadia Naseem and Annie J King. "Ammonia production in poultry houses can affect health of humans, birds, and the environment-techniques for its reduction during poultry production". In: *Environmental science and pollution research international* 25.16 (June 2018), pp. 15269–15293. ISSN: 0944-1344. DOI: 10.1007/s11356-018-2018-y. URL: https://doi.org/10.1007/s11356-018-2018-y.

[12] Suresh Neethirajan. "ChickTrack – A quantitative tracking tool for measuring chicken activity". In: *Measurement* 191 (2022), p. 110819. ISSN: 0263-2241. DOI: https://doi.org/10.1016/j.measurement.2022.110819. URL: https://www.sciencedirect.com/science/article/pii/S0263224122001154.

[13] Raspberry Pi. *Raspberry Pi hardware*. https://www.raspberrypi.com/. Accessed: 7th May, 2024. 2024.

[14] Lawrence Rabiner and Ronald Schafer. *Theory and Applications of Digital Speech Processing*. 1st. USA: Prentice Hall Press, 2010. ISBN: 0136034284.

[15] Shaoqing Ren, Kaiming He, Ross B. Girshick, and Jian Sun. "Faster R-CNN: Towards Real-Time Object Detection with Region ProposalNetworks". In: *Advances in Neural Information Processing Systems 28: Annual Conference on Neural Information Processing Systems 2015, December 7-12, 2015, Montreal, Quebec, Canada*. Ed. by Corinna Cortes, Neil D. Lawrence, Daniel D. Lee, Masashi Sugiyama, and Roman Garnett. 2015, pp. 91–99. URL: https://proceedings.neurips.cc/paper/2015/hash/14bfa6bb14875e45bba028a21ed38046-Abstract.html.

[16] Chris Stauffer and W. Eric L. Grimson. "Adaptive Background Mixture Models for Real-Time Tracking". In: *1999 Conference on Computer Vision and Pattern Recognition (CVPR '99), 23-25 June 1999, Ft. Collins, CO, USA*. IEEE Computer Society, 1999, pp. 2246–2252. DOI: 10.1109/CVPR.1999.784637. URL: https://doi.org/10.1109/CVPR.1999.784637.

[17] Ultralytics. *ultralytics/YOLOv5*. https://github.com/ultralytics/yolov5. Accessed: 7th May, 2024. 2024.

[18] Pascal Vincent, Hugo Larochelle, Isabelle Lajoie, Yoshua Bengio, and Pierre-Antoine Manzagol. "Stacked Denoising Autoencoders: Learning Useful Representations in a Deep Network with a Local Denoising Criterion". In: *J. Mach. Learn. Res.* 11 (2010), pp. 3371–3408. DOI: 10.5555/1756006.1953039. URL: https://dl.acm.org/doi/10.5555/1756006.1953039.

[19] Xiao Yang, Lilong Chai, Ramesh Bahadur Bist, Sachin Subedi, and Zihao Wu. "A Deep Learning Model for Detecting Cage-Free Hens on the Litter Floor". In: *Animals* 12.15 (2022). ISSN: 2076-2615. DOI: 10.3390/ani12151983. URL: https://www.mdpi.com/2076-2615/12/15/1983.

[20] Zoran Zivkovic. "Improved Adaptive Gaussian Mixture Model for Background Subtraction". In: *17th International Conference on Pattern Recognition, ICPR 2004, Cambridge, UK, August 23-26, 2004*. IEEE Computer Society, 2004, pp. 28–31. DOI: 10.1109/ICPR.2004.1333992. URL: https://doi.org/10.1109/ICPR.2004.1333992.