Multi-Sensor IoT Platform for Farm Animal Welfare Monitoring: Rule-Based Edge Classification, Person Detection, Live Dashboard Visualisation and Exploratory Vision Model Evaluation

Chidike Vincent
Sheffield Hallam University
Sheffield, United Kingdom
v.vincent@shu.ac.uk

Aliyuda Ali
Sheffield Hallam University
Sheffield, United Kingdom
Aliyuda.Ali@shu.ac.uk

Pratikshya Sharma
Sheffield Hallam University
Sheffield, United Kingdom
Pratikshya.Sharma@shu.ac.uk

Sarah Wild
Heeley City Farm
Sheffield, United Kingdom
s.wild@heeleyfarm.org.uk

Abdel-Karim Al Tamimi
Sheffield Hallam University
Sheffield, United Kingdom
A.Al-Tamimi@shu.ac.uk

Abstract

Automated and continuous animal welfare monitoring remains challenging in community farming environments, where limited resources, manual observation constraints, and the need for public engagement restrict the adoption of advanced digital technologies. This paper presents a low-cost edge IoT platform deployed at Heeley City Farm, Sheffield, UK, for real-time welfare monitoring across five animal enclosures housing goats, sheep, cattle, ponies, and chickens. The system integrates multi-modal sensing, edge computing, privacy-preserving person detection, and a public-facing web dashboard to provide transparent welfare insights aligned with the Five Freedoms framework. Each enclosure is equipped with a Raspberry Pi-based sensing unit collecting environmental and behavioural indicators, including temperature, humidity, light, sound, and motion data, which are classified locally on-device and visualised through a live dashboard accessible to farm staff and visitors.

The deployed platform demonstrates practical value by identifying a sustained thermal welfare concern in the poultry enclosure, where the computed Temperature-Humidity Index consistently exceeds the mild heat stress onset threshold for domestic poultry. As a secondary exploratory contribution, farm fine-tuning on 13,692 labelled overhead images reveals that Freedom 4 classification achieves Matthews Correlation Coefficient scores of up to +0.732 for the pony specialist using EfficientNet-B0, with ResNet-50 leading the generalist at +0.540. The transformer structural advantage hypothesis is not supported for this task: Swin Transformer Tiny underperforms ResNet-50 on three of five specialist enclosures. An embedding distance analysis confirms that Freedom 4 welfare labels, derived from rolling sensor window readings, are not a single-frame visual property, motivating future work on temporal sequence models.

The findings demonstrate the feasibility of combining participatory design, edge IoT, rule-based analytics, and accessible data visualisation to support transparent animal welfare monitoring in community farm settings. The work provides a foundation for future integration of augmented reality-based educational experiences and wider evaluation of AI-driven welfare assessment.

CCS Concepts

• **Computer systems organisation** → **Embedded systems**; • **Computing methodologies** → *Object detection*; • **Information systems** → *Sensor networks*.

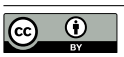
Keywords

animal welfare, IoT, edge computing, Five Freedoms, web dashboard, precision livestock farming, computer vision, transfer learning, convolutional neural network, vision transformer, community farm, multi-sensor platform

ACM Reference Format:
Chidike Vincent, Aliyuda Ali, Pratikshya Sharma, Sarah Wild, and Abdel-Karim Al Tamimi. 2026. Multi-Sensor IoT Platform for Farm Animal Welfare Monitoring: Rule-Based Edge Classification, Person Detection, Live Dashboard Visualisation and Exploratory Vision Model Evaluation. In *The 16th International Conference on the Internet of Things (IOT 2026), 2026, Newcastle, United Kingdom.* ACM, New York, NY, USA, 9 pages.

1 Introduction

Farm animal welfare is both a statutory obligation and a growing public concern in the United Kingdom. The Animal Welfare Act 2006 places a duty of care on those responsible for animals, and the Five Freedoms, codified by the Farm Animal Welfare Council from the foundational Brambell Report [4, 9], provide the internationally accepted framework for welfare assessment. The Five Freedoms specify that animals should be free from hunger and thirst, from discomfort, from pain and disease, from fear and distress, and should be able to express normal behaviour [4, 9]. In practice, compliance is assessed through periodic manual inspections, which are labour-intensive, intermittent, and inherently subjective. Staff fatigue and



This work is licensed under a Creative Commons Attribution 4.0 International License.
IOT 2026, Newcastle, United Kingdom
© 2026 Copyright held by the owner/author(s).

the impossibility of round-the-clock observation mean that welfare problems can go undetected for extended periods.

IoT technology offers a path towards continuous, objective, and scalable monitoring. Sensors can capture environmental and behavioural indicators associated with all five freedoms without requiring direct observer presence [8, 21]. Yet most deployed systems target large commercial operations or individual-animal metrics such as GPS and wearable accelerometers [1], leaving community and urban farms underserved. Heeley City Farm, a registered community charity in Sheffield housing goats, sheep, cattle, ponies, and chickens across five enclosures, exemplifies this gap: as an educational facility open to the public, it has both a statutory welfare obligation and an engagement opportunity that requires live, accessible welfare data at the point of visit.

Vision-based classification complements sensor data by capturing spatial and behavioural signals that point-in-time readings cannot. Overhead cameras have been used to detect pig postures, identify individual cattle, and assess body condition [3, 22]. A question that receives less attention is whether convolutional neural network (CNN) architectures, which extract local spatial features, are structurally better suited to multi-animal scene-level welfare classification than transformer architectures, which operate through global self-attention. When the welfare signal lies in the spatial distribution of animals across an enclosure, the relational reasoning capacity of attention mechanisms may offer a structural advantage over local convolution [14].

This paper makes the following contributions:

(1) **Rule-based edge welfare classification with on-device person detection**: a multi-sensor system that classifies enclosure welfare in real time against the Five Freedoms framework using rule-based thresholds, running entirely on a Raspberry Pi 4B. Human-containing frames are detected and permanently deleted on-device before any path is recorded.
(2) **Live web dashboard**: a publicly accessible dashboard that displays real-time welfare status and sensor readings for all five enclosures, updating every 30 seconds from live edge data.
(3) **Exploratory vision model evaluation**: a preliminary evaluation of lightweight CNN and transformer architectures: EfficientNet-B0, Swin Transformer Tiny, and ResNet-50 for overhead multi-animal welfare classification, conducted offline on the development machine. To the best of the authors' knowledge, simultaneous image-based welfare classification across five separate, species-specific, indoor farm enclosures each monitored by a fixed overhead camera has not previously been demonstrated in the literature.

2 Related Work

2.1 IoT-Based Livestock Monitoring

Sensor-based IoT systems for livestock health have developed substantially over the past decade. Modak et al. [21] survey temperature, humidity, and motion sensor systems paired with Raspberry Pi, ESP32, and Arduino microcontrollers. Wathes [31] identified precision livestock farming as transformative for welfare monitoring. Matthews et al. [19] showed Raspberry Pi systems can infer cattle activity patterns with practical accuracy. Colombari et al. [6] deployed a six-month IoT framework tracking temperature, light, and humidity in commercial dairy housing. These approaches focus on intensive commercial operations; enclosure-level community farm monitoring with public engagement requirements has received little attention.

2.2 Computer Vision for Animal Welfare

Camera-based monitoring of livestock has accelerated with advances in deep learning. Bhujel et al. [3] used overhead cameras and learning-based detection to monitor pig physico-temporal activities under varying environmental conditions, producing a publicly available labelled overhead dataset. Oliveira et al. [22] reviewed deep learning applications for livestock, covering detection, classification, and body condition scoring across multiple species. Matthews et al. [18] showed that automated camera-based tracking of pig behaviour could detect welfare-relevant changes earlier than human observation alone. A systematic review of vision-based approaches to outdoor livestock monitoring found a clear shift toward deep learning methods and identified multi-species generalisation as an open challenge [32]. Cross-species welfare classification from overhead fixed cameras across a multi-enclosure community farm remains unexplored.

2.3 CNN and Transformer Architectures for Livestock

EfficientNet-B0 [26] achieves competitive accuracy through compound scaling of depth, width, and resolution at 5.3 M parameters. ResNet-50 [11] remains the most widely cited CNN baseline in livestock vision literature. The Swin Transformer [14] computes hierarchical shifted-window self-attention, enabling global context modelling that scales linearly with image resolution and can reason about spatial relationships across the full image rather than local patches [28]. For multi-animal enclosure scenes, where welfare signals may lie in the spatial distribution of animals, this structural difference presents a testable hypothesis. To the best of the authors' knowledge, no published work has comparatively evaluated CNN and transformer architectures for image-based welfare classification across simultaneously monitored, species-specific, indoor community farm enclosures.

2.4 IoT Dashboards and Real-Time Welfare Visualisation

Web-based dashboards that aggregate sensor data and present welfare indicators to farm operators have been explored in precision livestock farming [6]. Phupattanasilp and Tong [24] demonstrated that live sensor data could be surfaced through accessible interfaces for non-technical users. Translating Five Freedoms classifications into plain-language indicators accessible through a standard web browser without login represents a practical step toward transparency in community animal care.

3 System Design and Implementation

3.1 Overview

The system comprises five identical sensor units, one per enclosure, each running on a Raspberry Pi 4B with a 30-second polling loop. A participatory co-design workshop with farm staff established the welfare monitoring requirements, directly informing sensor selection, enclosure deployment, and dashboard design. Sensor readings are classified locally on the device against the Five Freedoms framework and written to a cloud database along with raw sensor values. A serverless API function serves the most recent classification and sensor data to any requesting client. A static web dashboard hosted on GitHub Pages displays live welfare data refreshed at 30-second intervals. The vision model pipeline runs offline on the development machine, pulling stored images and sensor logs from each Pi via SSH, and producing per-enclosure welfare predictions that appear on the dashboard alongside the rule-based output. Figure 1 shows the overall architecture.

3.2 Hardware

Each sensor unit is housed in a QWORK IP65-rated junction box measuring 150 mm × 100 mm × 90 mm, providing weather protection for outdoor enclosure mounting. A 170-point mini modular breadboard inside the box mounts the MCP3008 analogue-to-digital converter chip and distributes shared power and ground rails, avoiding multiple wires on a single GPIO pin. The unit is powered by an official Raspberry Pi USB-C supply with the cable entering through a sealed gland. The camera module faces through the transparent polycarbonate lid. A silica gel sachet inside the sealed box manages residual moisture. Table 1 summarises the sensor stack.

Table 1: Sensor unit hardware per enclosure

Component	Sensor / Purpose
Raspberry Pi 4B 4 GB	Edge compute and polling controller
AZDelivery DHT22	Temperature and humidity (single-wire)
GY-302 BH1750	Light intensity in lux (I^2C)
PIR sensor (7 m)	Passive Infrared (PIR) enclosure-level motion
Adafruit MAX9814	Microphone amplifier (analogue output)
MCP3008 ADC	10-bit SPI analogue-to-digital conversion
Pi Camera Module 3	Overhead image capture, 75° FOV

The DHT22 and BH1750 support Freedom 2 (Discomfort) and Freedom 4 (Normal Behaviour) classification respectively. The Passive Infrared (PIR) sensor detects enclosure-level motion, the primary proxy for Freedom 4 and Freedom 3 (Pain and Disease). The MAX9814 analogue microphone output is digitised at ten-bit resolution via hardware SPI and normalised to a 0–100 scale for Freedom 5 (Fear and Distress) and Freedom 3 assessment. The camera captures overhead images on a species-specific schedule, forming the vision model training dataset.

3.3 Sensing and Classification Approach

A Python service runs continuously on each Pi, executing a 30-second polling loop managed by a `systemd` unit that restarts automatically on failure or device reboot. At each cycle, all five sensors are sampled, the rule-based welfare classifier is applied, and both the raw sensor data and the resulting classification are written to Supabase PostgreSQL via an HTTPS REST call. In addition to Supabase writes, every sensor reading is appended to a local JSON log file (`readings.jsonl`) on the Pi SD card, providing a complete local record that persists through any connectivity interruption.

3.3.1 Rule-Based Welfare Classification. The classifier evaluates each of the Five Freedoms independently, producing a binary good or attention status for each and a single overall enclosure status. For any scalar sensor reading x_f associated with freedom f, the decision rule is:

$$s_f = \begin{cases} \texttt{attention} & \text{if } x_f \notin [t_f^{\text{low}},\, t_f^{\text{high}}] \\ \texttt{good} & \text{otherwise,} \end{cases} \tag{1}$$

where $[t_f^{\text{low}},\, t_f^{\text{high}}]$ is the species-specific acceptable range for freedom f. Welfare thresholds are applied at a fixed 0.5 decision boundary with no tuning performed; future work should evaluate threshold optimisation for imbalanced deployment data.

Freedom 2 (Discomfort) departs from this scalar formulation. Rather than applying independent temperature and humidity thresholds, the classifier computes the Temperature-Humidity Index (THI), which integrates both measurements into a single thermal stress score widely used in precision livestock farming literature [15, 27]:

$$\text{THI} = (1.8T + 32) - \left[(0.55 - 0.0055 \cdot \text{RH}) \times (1.8T - 26)\right], \tag{2}$$

where T is the dry-bulb temperature in degrees Celsius and RH is the relative humidity as a percentage. Freedom 2 flags attention when THI reaches or exceeds the species-specific mild stress onset threshold. THI thresholds are drawn from recent literature and applied per enclosure as shown in Table 2.

Table 2: THI mild stress onset thresholds applied per enclosure for Freedom 2 classification, based on 2022–2025 literature.

Enclosure	Species	THI onset	Source
Chickens	Domestic poultry	74	[16]
Cattle	Beef cattle	72	[15]
Goats	Small ruminant	72	[23]
Sheep	Small ruminant	72	[23]
Ponies	Equine	72	[2]

If either temperature or humidity is unavailable due to a sensor fault, the THI cannot be computed and Freedom 2 defaults to good, consistent with the principle that sensor faults should not generate welfare alerts. Active hours are parameterised per enclosure based on species routines confirmed through the co-design workshop: chickens 03:00–08:00 and 16:00–21:00 UTC; goats, sheep, ponies, and cattle 03:00–09:00 and 14:00–21:00 UTC. A false motion reading

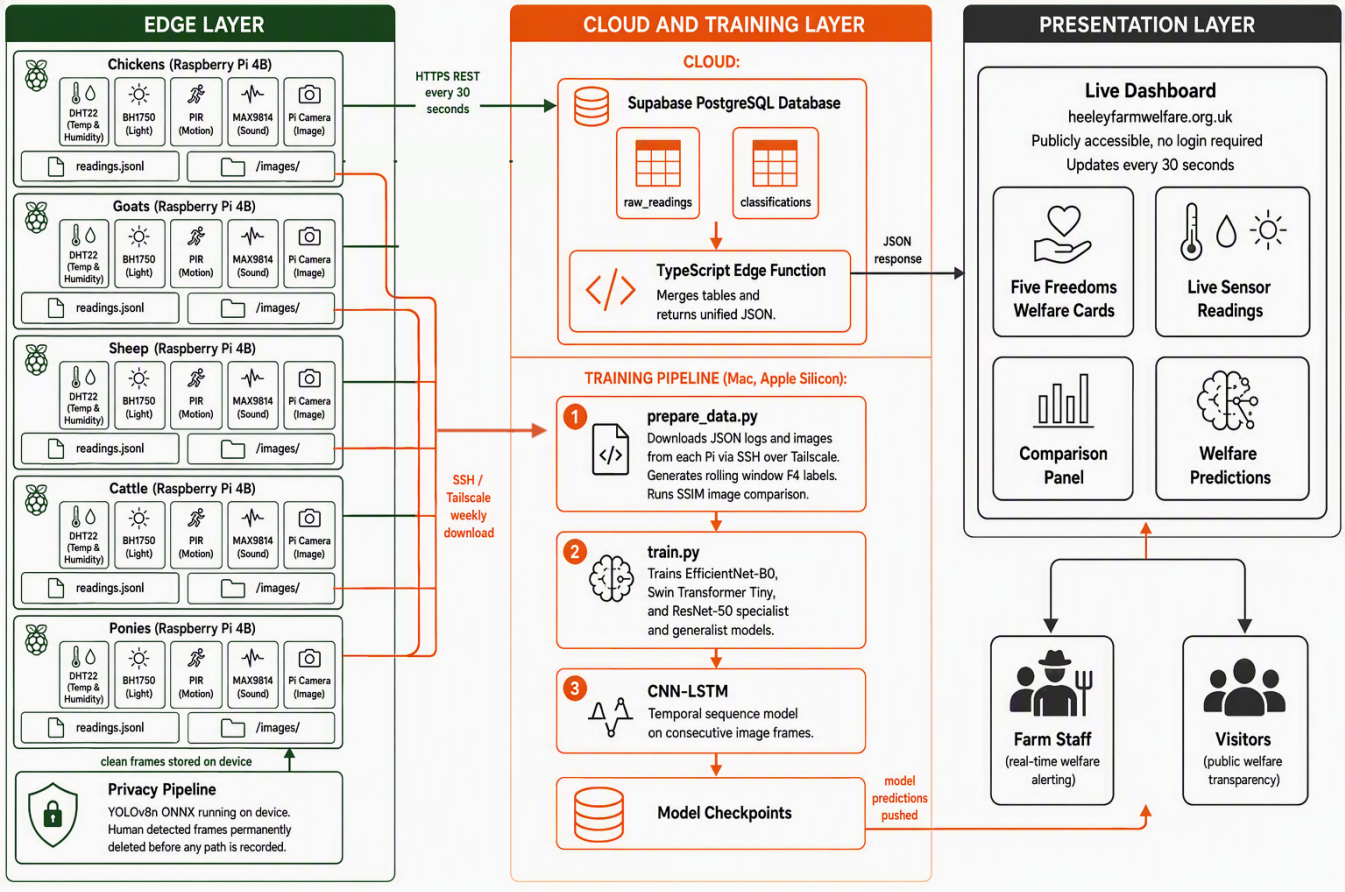


Figure 1: System architecture. The polling service runs on each Pi, classifying sensor readings locally and writing raw readings and classifications to Supabase PostgreSQL. A Supabase Edge Function returns merged JSON to the dashboard. The vision model pipeline runs on the development machine, pulling stored sensor readings and images from each Pi via SSH over Tailscale, classifying readings fresh at training time, and pushing updated predictions to the public dashboard.

during active hours triggers a Freedom 4 attention status; outside active hours the same reading produces a good status, as resting is expected.

The microphone monitors ambient sound within each enclosure. Sustained readings above 45 during active hours (observed in 0.6% of readings) contribute to Freedom 5 (Fear and Distress) and Freedom 3 assessments. Pre-written explanation templates accompany each classification, factual and deliberately conservative: the Freedom 4 attention template reads *Lower movement than expected. Animals may be resting, outside, or require attention*, acknowledging that a resting animal produces the same PIR reading as genuinely restricted movement [25].

3.3.2 Privacy-Preserving Image Approach. Capture windows are aligned to reliable daylight (lux ≥ 3.0) and confirmed animal presence per enclosure: chickens 06:00–08:00 and 16:00–19:00 UTC at one-minute intervals; goats, sheep, ponies, and cattle 06:00–09:00 and 14:00–19:00 UTC at two-minute intervals, yielding approximately 200 capture attempts per enclosure per day.

Immediately after capture, a human presence check runs using YOLOv8n [13] via ONNX Runtime [20] on-device, achieving 156 ms inference on the Pi 4B. This adapts the YOLO-based edge detection approach from prior work [29]: if a person is detected at confidence above 0.5 (COCO class 0), the frame is permanently deleted via `os.remove()` before any path is recorded. This human detection task is distinct from the animal presence filter discussed in Section 4: YOLOv8n performs well on persons because COCO training data includes abundant upright bipedal examples, whereas overhead animal detection represents a different visual domain that the same model does not generalise to. The ONNX model format is required because the ultralytics PyTorch runtime produces an illegal instruction fault on the Pi 4B ARM Cortex-A72.

3.3.3 Cloud and Serving Infrastructure. Sensor readings and classifications are written to two Supabase PostgreSQL tables: `raw_readings` and `classifications`. A TypeScript Edge Function merges both into a single JSON response served via a JWT-disabled public endpoint, exposing only welfare indicators and environmental sensor values.

3.4 Live Dashboard and Data Serving

A static dashboard hosted on GitHub Pages fetches the Edge Function every 30 seconds. It displays a Five Freedoms welfare card per enclosure with colour-coded status indicators, live sensor readings, and a side-by-side comparison panel. The dashboard is publicly accessible at `heeleyfarmwelfare.org.uk` without login. Figure 2 shows the interface.

3.5 Vision Model Approach

The vision model pipeline trains on overhead images paired with welfare labels generated fresh at training time. The `readings.jsonl` log on each Pi is downloaded via SSH over Tailscale; the current classifier is applied to each raw reading to produce labels, ensuring consistency regardless of when readings were captured or whether connectivity gaps interrupted Supabase writes. The cloud database serves real-time display only.

3.5.1 Architecture Selection and Research Hypothesis. Three architectures are evaluated in a preliminary exploratory comparison: EfficientNet-B0 [26], Swin Transformer Tiny [14], and ResNet-50 [11]. ResNet-50 is the most widely cited CNN baseline in livestock vision literature [22]. EfficientNet-B0 achieves strong accuracy at 5.3 million parameters through compound network scaling. Swin Transformer Tiny, with 28 million parameters, represents the transformer approach through hierarchical shifted window self-attention [14].

The central research hypothesis is that multi-animal enclosure welfare classification is a scene-level task. Welfare signals such as the spatial distribution of animals across a pen, clustering behaviour associated with fear [10], or complete enclosure vacancy cannot be reliably extracted from local receptive fields alone. A CNN identifies features within small image patches; a transformer's self-attention can reason about relationships between patches across the full image simultaneously [28]. The comparative evaluation tests whether this structural advantage translates to higher classification accuracy in practice.

3.5.2 Pre-Training on Pig Overhead Data. In the absence of labelled overhead imagery for the five target species at the time of writing, a two-phase transfer learning strategy is employed. Each model is first pre-trained on the publicly available pig overhead activity monitoring dataset by Bhujel et al. [3], comprising 667 labelled overhead frames released under a CC BY 4.0 licence. This establishes the domain of overhead livestock imagery before farm-specific fine-tuning begins. Per-class weighted binary cross-entropy is applied independently for each of the five freedom outputs to address label imbalance:

$$\mathcal{L} = -\frac{1}{N}\sum_{i=1}^{N}\left[w_+\, y_i \log \hat{y}_i + (1-y_i)\log(1-\hat{y}_i)\right], \tag{3}$$

where N is the number of training samples, $y_i \in \{0,1\}$ is the ground truth label, $\hat{y}_i$ is the predicted probability of the positive (attention) class, and w_+ is the positive class weight set to the ratio of negative to positive samples to compensate for the expected predominance of good readings during normal farm operation.

Training runs on Apple Silicon via the Metal Performance Shaders backend in PyTorch. Phase 1 freezes all backbone layers and trains only the classification head for ten epochs at a learning rate of 10^{-3}. Phase 2 unfreezes the full network and fine-tunes for a further ten epochs at 10^{-4}.

Pre-training results are shown in Table 3. EfficientNet-B0 converges stably at 91.8%. Swin Transformer Tiny reaches 74.6% with accuracy instability at epochs 8–9, consistent with transformer sensitivity to limited data [7]. ResNet-50 reaches 64.9% with a gradient spike at epoch 6. All models use the best frozen-phase checkpoint as the fine-tuning starting point.

Table 3: Pre-training results on the Bhujel et al. pig overhead dataset [3]. Phase 1 is frozen backbone; Phase 2 is full network fine-tune. Best validation accuracy is reported at epoch 10.

Model	Params	Val Acc (%)	Behaviour
EfficientNet-B0	5.3 M	91.8	Stable
Swin-Tiny	28 M	74.6	Collapse ep. 8–9
ResNet-50	25 M	64.9	Spike ep. 6

3.5.3 Farm Fine-Tuning and Evaluation. Farm fine-tuning uses 13,692 labelled image rows collected between 1 July and 21 July 2026 across all five enclosures. Five per-enclosure specialist models are trained separately to avoid mixing species-specific behavioural signatures. A sixth generalist model is trained on the combined five-enclosure dataset with training samples capped at the size of the smallest enclosure to prevent any single enclosure from dominating the learned representation, addressing the class imbalance effect on minority-class recall reported in prior work on edge classification [30]. Two training phases are reported: Phase 1 uses data collected between 1 July and approximately 15 July (pre-smudge for chickens), and Phase 2 uses the full combined dataset to 21 July. The chicken camera lens became partially obscured from 13 July 2026; a `camera_clean` boolean column is added to the training CSV to record image quality, enabling the smudge impact to be audited separately.

Evaluation reports the Matthews Correlation Coefficient (MCC) [5], which accounts for all four elements of the confusion matrix (true positives TP, true negatives TN, false positives FP, false negatives FN) and is particularly suited to imbalanced binary classification:

$$\mathrm{MCC} = \frac{TP \cdot TN - FP \cdot FN}{\sqrt{(TP+FP)(TP+FN)(TN+FP)(TN+FN)}}, \tag{4}$$

where a value of +1 indicates perfect prediction, 0 indicates performance no better than random guessing, and −1 indicates systematic inversion. Interpretive bands are adopted from the literature [5]: 0.0 to 0.3 indicates weak learning; 0.3 to 0.5 moderate; 0.5 to 0.7 good; 0.7 to 0.9 strong. A classifier that predicts good for every input would achieve high accuracy during a well-managed deployment period but score MCC 0.0, correctly exposing the absence of genuine discrimination. All models are evaluated on single runs with shared hyperparameters; results are indicative rather than confirmatory, and no confidence intervals are reported.

Figure 3 shows the validation accuracy curves from pre-training across ten epochs for all three architectures.

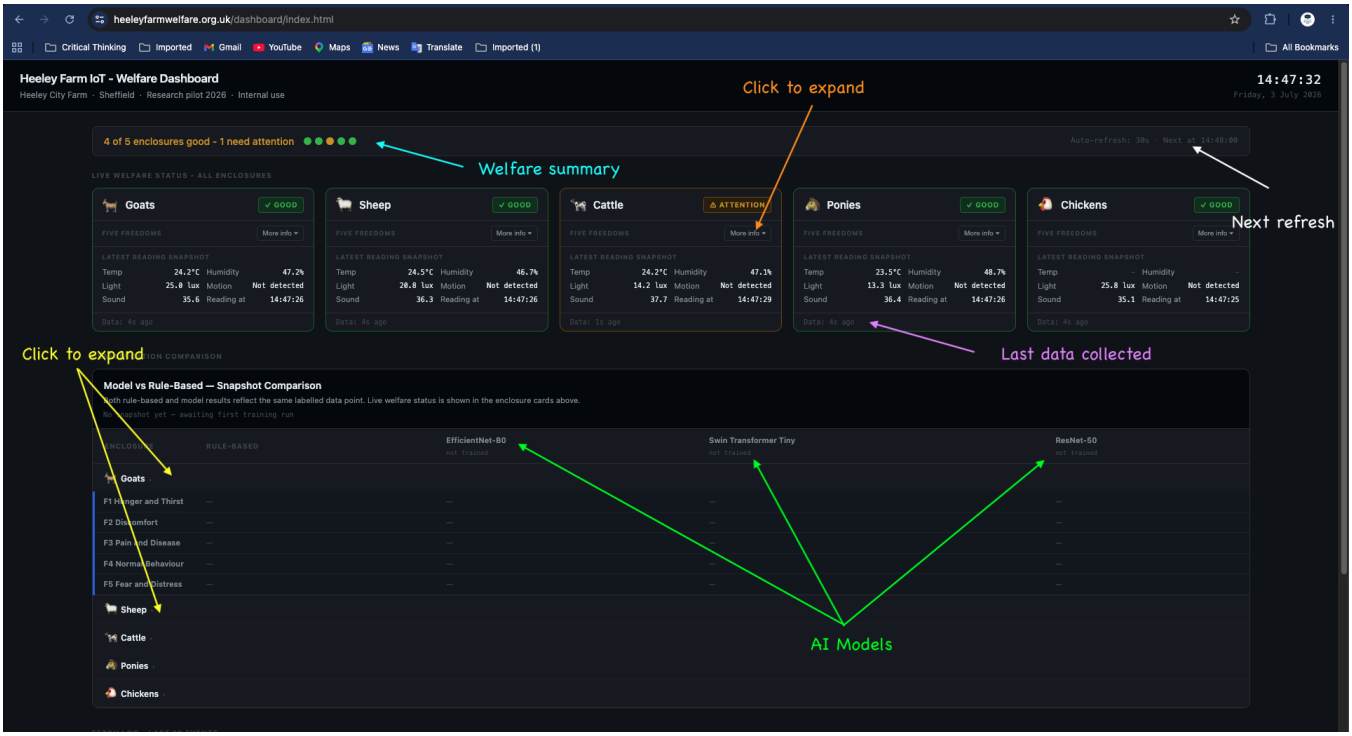


Figure 2: Live dashboard at `heeleyfarmwelfare.org.uk` showing welfare status cards for all five enclosures in desktop view. Each card displays the overall Freedom status, individual Freedom indicators with colour-coded dots (green for good, amber for attention), and a live sensor reading snapshot. A side-by-side comparison panel allows farm staff to assess all enclosures at a glance. The dashboard updates every 30 seconds and is publicly accessible without login.

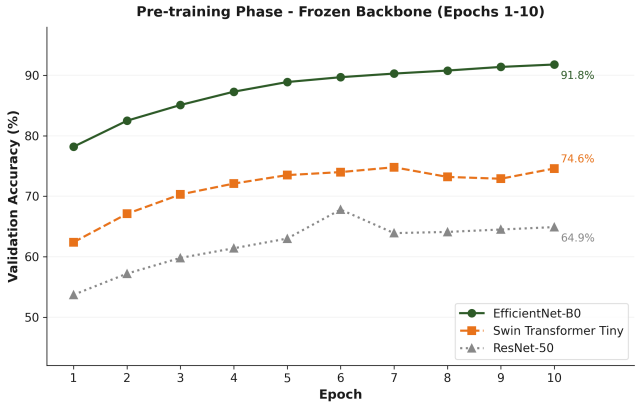


Figure 3: Validation accuracy over ten fine-tuning epochs on the pig overhead dataset [3] for EfficientNet-B0, Swin Transformer Tiny, and ResNet-50. EfficientNet-B0 converges stably at 91.8%. Swin Transformer Tiny collapses at epochs 8–9 before recovering to 74.6%. ResNet-50 exhibits a gradient spike at epoch 6 before settling at 64.9%.

4 Preliminary Deployment Findings

The system was deployed at Heeley City Farm in June 2026 with all five enclosures operational by 1 July 2026. Official data collection commenced at 14:14 UTC on 1 July 2026. Capture windows were subsequently refined from lux analysis; the minimum threshold was lowered from 10.0 to 3.0, confirmed usable by physical image inspection. The rule-based classifier has processed over 88,000 polling cycles across all five enclosures.

Thermal welfare concern. At a representative reading of 29 °C and 43% relative humidity, the computed THI is approximately 80, placing the poultry enclosure in the moderate heat stress zone (onset THI 74 [16]). Freedom 2 is flagged continuously; farm staff were made aware through the dashboard. This illustrates a welfare signal that periodic manual inspection would not reliably surface. The THI-based classification is more scientifically grounded than temperature alone, since humidity substantially affects avian thermoregulatory capacity [16].

Animal presence filter. A YOLOv8n-based animal presence filter was evaluated during training data preparation but produced rejection rates exceeding 89% across all enclosures due to a mismatch between COCO training perspective (side-on, ground-level) and the overhead enclosure view. This is a distinct failure mode from the on-device human detection described in Section 3.3.2, which performs reliably: person detection benefits from consistent bipedal COCO examples, while overhead animal detection is a different

visual domain the same model does not generalise to. The filter was disabled; a domain-specific overhead classifier is identified as future work.

Deployment limitations. A wild pigeon triggered PIR capture in the poultry enclosure on one occasion, illustrating the limitation of fixed PIR sensors with unrestricted openings. Animal enclosure assignments changed during the collection period, with a different species observed in the sheep pen from 8 July 2026, introducing label inconsistency. Capture windows concentrated on staff-provided schedules; animals do not follow fixed routines, and time-correlated image distributions may reduce model generalisation. Intermittent connectivity loss affected two Pi units; the local `readings.jsonl` log preserved all data, and JSON files contained approximately seven times more training-eligible rows than the Supabase record, confirming the value of the local-first data approach.

5 Results and Evaluation

5.1 Rule-Based Classifier Performance

Freedoms F1, F2 (excluding the poultry thermal event), F3, and F5 produced zero attention events across all 13,692 image rows during the 21-day pilot. This reflects consistently good welfare management and is a positive finding in its own right. Freedom 4 (F4) is the only freedom with variation, producing 1,200 attention rows across all five enclosures, and is the sole target for vision model training and evaluation.

5.2 Vision Model Evaluation

5.2.1 Phase 1: EfficientNet-B0 Specialists. Table 4 reports F4 classification results for EfficientNet-B0 specialist models trained on the first training batch (predominantly July 1–15 data; chicken images prior to the camera smudge event of 13 July). Each specialist is trained on its enclosure's data only. The generalist is trained on all five enclosures combined, capped at the size of the smallest enclosure to balance species representation.

Table 4: EfficientNet-B0 F4 classification results: Phase 1 training (clean chicken data). MCC: 0.0 to 0.3 weak; 0.3 to 0.5 moderate; 0.5 to 0.7 good; 0.7 to 0.9 strong.

Enclosure	MCC	TP	TN	FP	FN	Band
Chickens	+0.619	258	54	43	8	Good
Goats	0.000	350	0	1	0	Degenerate
Sheep	+0.451	332	7	0	25	Moderate
Cattle	+0.225	324	6	21	12	Weak
Ponies	+0.732	284	22	1	15	Strong
Generalist	+0.423	1364	110	30	102	Moderate

Ponies achieves the strongest result at MCC +0.732, indicating clear F4 signal in the pony enclosure visual patterns. Chickens scores MCC +0.619 on pre-smudge data. Goats is degenerate (MCC 0.0, TN = 0) due to only three attention rows in the training set; no class weighting can compensate for a near-absent minority class.

5.2.2 Phase 2: Three-Model Comparison. Table 5 reports F4 specialist results for all three architectures trained on the combined 13,692-image dataset (1 July to 21 July 2026). Figure 4 compares generalist model performance across all three architectures. Figure 5 shows the MCC comparison across all three models per enclosure.

Table 5: F4 specialist MCC, Phase 2 combined dataset. Bold = best per enclosure. Goats degenerate across all models (only 3 attention rows in training; not learnable).

Enclosure	EffNet-B0	Swin-Tiny	ResNet-50	Best
Chickens	+0.379	+0.547	**+0.656**	ResNet
Goats	0.000	0.000	0.000	Degenerate
Sheep	**+0.442**	+0.254	+0.401	EffNet
Cattle	+0.306	**+0.310**	+0.167	Swin
Ponies	**+0.629**	+0.535	+0.569	EffNet
Avg*	+0.439	+0.412	**+0.448**	ResNet

*Average excludes Goats.

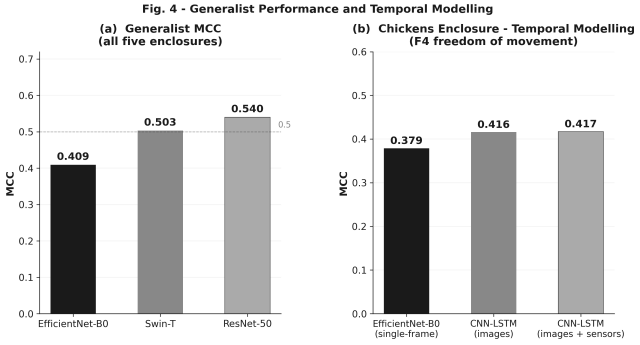


Figure 4: Left: generalist F4 MCC for all three architectures, Phase 2 combined dataset. ResNet-50 leads at MCC +0.540. Right: CNN-LSTM temporal sequence model compared with single-frame EfficientNet-B0 on chickens F4. CNN-LSTM (images only) achieves MCC +0.416; sensor fusion adds no further value at +0.417.

ResNet-50 achieves the highest average specialist MCC (+0.448) and highest generalist MCC (+0.540). The transformer structural advantage hypothesis is not supported: Swin underperforms ResNet-50 on three of five specialists. F4 is a temporal PIR construct; convolutional hierarchies capture this signal more effectively than patch-level self-attention. EfficientNet-B0 leads on ponies and sheep, suggesting strong domain transfer for those enclosures. ResNet-50 achieves MCC +0.656 on chickens, the highest single specialist score, though 63% of Phase 2 chicken images carry camera-lens obscuration from 13 July; lower-level convolutional filters are more robust to uniform image haze than attention mechanisms, which may partially explain the gap. These rankings are indicative only:

each model was run once with shared hyperparameters and no confidence intervals are available.

5.2.3 Temporal Sequence Modelling. To test whether temporal context across consecutive frames improves F4 classification, a CNN-LSTM model is evaluated on the chicken enclosure. The CNN-LSTM combines a Convolutional Neural Network (CNN) feature extractor with a Long Short-Term Memory (LSTM) [12] recurrent network. The frozen EfficientNet-B0 backbone (chickens specialist weights) extracts per-frame embeddings, which the LSTM then processes as a sequence, allowing the model to reason about changes in animal position across multiple consecutive captures rather than classifying each frame in isolation. Sequences of five consecutive chicken enclosure images are constructed from the combined dataset.

The CNN-LSTM achieves MCC +0.416 on the Phase 2 combined dataset, outperforming single-frame classification at +0.379, confirming that temporal context across consecutive frames adds genuine but modest value for F4 detection. The model stabilises at MCC +0.416 with 129 validation attention sequences in Phase 2, having been unstable in Phase 1 with only 40 validation attention sequences.

Appending the rolling PIR motion window readings to the image embeddings (sensor fusion) produces MCC +0.417, statistically identical to images only. The LSTM already captures temporal stillness patterns from the visual stream; the PIR readings are redundant. Sensor fusion is not the appropriate direction for F4 classification with the current sensor set.

An embedding distance analysis further establishes the fundamental nature of the challenge. Using the trained EfficientNet-B0 backbone to extract 1,280-dimensional image embeddings, the mean distance of F4 attention images from the centroid of F4 good images is 9.27, compared with 9.15 for good images themselves; an effect size of $d = 0.102$ (negligible) with $p = 0.020$ (statistically significant only due to sample size). The two distributions are essentially co-located in embedding space: F4 attention is not a single-frame visual property. The F4 label derives from sustained PIR inactivity across a rolling window of sensor readings; a still image cannot encode that temporal history regardless of the architecture used to process it.

5.3 Dashboard Evaluation

The dashboard has been operational and publicly accessible at `heeleyfarmwelfare.org.uk` since July 2026. The system updates every 30 seconds from live edge data. Quantitative latency and uptime measurements and user-centred evaluation via farm staff questionnaire form part of the extended deployment evaluation beyond the submission period.

6 Discussion

The system demonstrates that the Five Freedoms framework can be operationalised on low-cost edge hardware without cloud compute at the classification step. Readings and classifications are logged locally through network outages; no data is transmitted before welfare assessment completes on the device. The local JSON log proved essential: connectivity gaps left Supabase records covering only a fraction of actual data, while the JSON files preserved everything. Welfare labels are therefore generated fresh at training time from the raw log, accommodating classifier refinements without retroactive database changes.

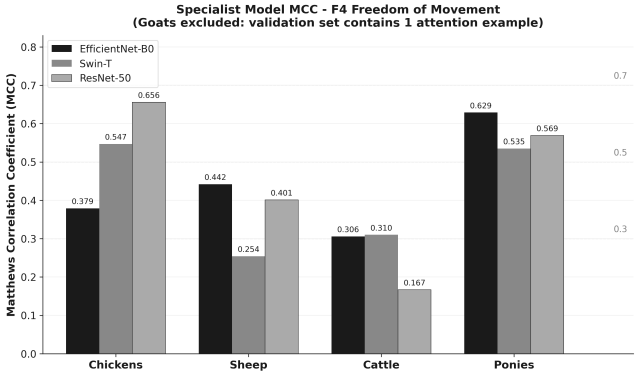


Figure 5: F4 MCC comparison across EfficientNet-B0, Swin Transformer Tiny, and ResNet-50 specialist models per enclosure, Phase 2 combined dataset. Goats excluded as all three models are degenerate. ResNet-50 achieves the highest specialist score on chickens (+0.656); EfficientNet-B0 leads on ponies and sheep.

Farm fine-tuning results confirm that the transformer structural advantage hypothesis is not supported for F4 classification: Swin Transformer Tiny underperforms ResNet-50 on three of five specialists. The F4 label is a temporal PIR motion construct; convolutional feature hierarchies capture this signal more effectively than patch-level self-attention. Because the spatial hypothesis requires F4 to be a visually separable property, and the embedding distance analysis (Section 5.2.3) shows the two classes are essentially co-located in embedding space ($d = 0.102$, negligible), the hypothesis is not testable from single-frame images alone; transformer advantage for appearance-based welfare signals remains an open question as deployment data grows.

The PIR sensor's single-sample limitation is the most significant structural constraint [25]. A rolling window classifier addresses this: F4 flags attention only after N consecutive false motion readings within the active window ($N = 5$ min for most species; 10 min for ponies [2, 17, 25]). As reported in Section 5.2.3, the CNN-LSTM edges out single-frame classification (MCC +0.416 versus +0.379), confirming that consecutive frames carry more F4 signal than any single frame, while sensor fusion adds no further value.

The microphone reports a normalised relative scale rather than absolute dB SPL, precluding direct comparison with published acoustic thresholds; it remains effective for detecting anomalous sound events within each enclosure. The live dashboard (Figure 2) provides continuous visibility across all five enclosures; each card displays Five Freedoms indicators with plain-language explanations and a live sensor snapshot accessible to farm staff and visitors without login.

7 Conclusion and Future Work

A multi-sensor IoT platform for real-time farm animal welfare monitoring has been deployed at Heeley City Farm, Sheffield, structured

against the Five Freedoms framework and developed through participatory co-design with farm staff. The edge classifier processes five sensor modalities every 30 seconds, with human-containing frames detected and permanently deleted on-device before any path is recorded. As an exploratory secondary contribution, farm fine-tuning on 13,692 labelled overhead images demonstrates genuine F4 discrimination across four of five enclosures, with ResNet-50 leading the generalist at MCC +0.540.

The poultry THI finding illustrates the system's practical value: a thermal welfare concern continuously flagged and communicated to farm staff that periodic manual inspection would not reliably surface.

Six directions for future work are identified: (1) extended deployment to generate F1–F3 and F5 attention events for full five-freedom evaluation; (2) longer collection to make the goat enclosure trainable; (3) a domain-specific overhead animal presence classifier to replace the unsuitable general-purpose YOLO filter; (4) CNN-LSTM training at larger scale to confirm temporal sequence advantages; (5) cross-species generalisation evaluation as enclosure datasets grow; and (6) an augmented reality overlay interface at each pen sign for public engagement.

Acknowledgments

The authors thank Heeley City Farm, Sheffield, for hosting the deployment and for providing access to the enclosures throughout the study period. This work is supported by Sheffield Hallam University.

References

[1] Rawan S. Abdulsadig et al. 2025. Pervasive Sensing for Livestock Health and Activity Monitoring: Current Methods and Techniques. *arXiv preprint arXiv:2503.15674* (2025).

[2] American Association of Equine Practitioners. 2024. Heat Stress Guidelines for Horses. US Equestrian Federation. https://www.usef.org

[3] Anil Bhujel, Elanchezhian Arulmozhi, Byeong-Eun Moon, and Hyeon-Tae Kim. 2021. Deep-Learning-Based Automatic Monitoring of Pigs' Physico-Temporal Activities at Different Greenhouse Gas Concentrations. *Animals* 11, 11 (2021), 3089. doi:10.3390/ani11113089

[4] Roger Brambell. 1965. *Report of the Technical Committee to Enquire into the Welfare of Animals Kept under Intensive Livestock Husbandry Systems.* Technical Report. Her Majesty's Stationery Office, London.

[5] Davide Chicco and Giuseppe Jurman. 2020. The Advantages of the Matthews Correlation Coefficient (MCC) over F1 Score and Accuracy in Binary Classification Evaluation. *BMC Genomics* 21, 1 (2020), 6. doi:10.1186/s12864-019-6413-7

[6] Andrea Colombari et al. 2025. An Internet of Things Framework for Monitoring Environmental Conditions in Livestock Housing to Improve Animal Welfare and Assess Environmental Impact. *Animals* 15, 5 (2025), 644. doi:10.3390/ani15050644

[7] Alexey Dosovitskiy, Lucas Beyer, Alexander Kolesnikov, Dirk Weissenborn, Xiaohua Zhai, Thomas Unterthiner, Mostafa Dehghani, Matthias Minderer, Georg Heigold, Sylvain Gelly, Jakob Uszkoreit, and Neil Houlsby. 2021. An Image is Worth 16x16 Words: Transformers for Image Recognition at Scale. In *International Conference on Learning Representations (ICLR).* https://openreview.net/forum?id=YicbFdNTTy

[8] Carly A. Evans, Mark G. Trotter, and Joanne K. Manning. 2022. Opportunities to Monitor Animal Welfare Using the Five Freedoms with Precision Livestock Management on Rangelands. *Frontiers in Animal Science* 3 (2022). doi:10.3389/fanim.2022.928514

[9] Farm Animal Welfare Council. 1979. Press Statement, 5 December 1979. Farm Animal Welfare Council.

[10] Björn Forkman, Alain Boissy, Marie-Christine Meunier-Salaün, Elisabetta Canali, and Richard B. Jones. 2007. A Critical Review of Fear Tests Used on Cattle, Pigs, Sheep, Poultry and Horses. *Physiology & Behavior* 92, 3 (2007), 340–374. doi:10.1016/j.physbeh.2007.03.016

[11] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. 2016. Deep Residual Learning for Image Recognition. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR).* 770–778. doi:10.1109/CVPR.2016.90

[12] Sepp Hochreiter and Jürgen Schmidhuber. 1997. Long Short-Term Memory. *Neural Computation* 9, 8 (1997), 1735–1780. doi:10.1162/neco.1997.9.8.1735

[13] Glenn Jocher, Ayush Chaurasia, and Jing Qiu. 2023. Ultralytics YOLO (Version 8.0.0). GitHub. https://github.com/ultralytics/ultralytics DOI: 10.5281/zenodo.7347926.

[14] Ze Liu, Yutong Lin, Yue Cao, Han Hu, Yixuan Wei, Zheng Zhang, Stephen Lin, and Baining Guo. 2021. Swin Transformer: Hierarchical Vision Transformer Using Shifted Windows. In *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV).* 9992–10002. doi:10.1109/ICCV48922.2021.00986

[15] T.L. Mader, M.S. Davis, and T. Brown-Brandl. 2006. Environmental Factors Influencing Heat Stress in Feedlot Cattle. *Journal of Animal Science* 84, 3 (2006), 712–719. doi:10.2527/2006.843712x

[16] A.C. Mancinelli et al. 2024. Changes in Rectal Temperature as a Means of Assessing Heat Tolerance and Sensitivity in Chickens. *Animals* 14 (2024). doi:10.3390/ani14233389 PMC11584500.

[17] Simon G. Matthews, Andrea L. Miller, Jonathan Clapp, Thomas Plötz, and Ilias Kyriazakis. 2016. Early Detection of Health and Welfare Compromises through Automated Detection of Behavioural Changes in Pigs. *Veterinary Journal* 217 (2016), 43–51. doi:10.1016/j.tvjl.2016.09.005

[18] Simon G. Matthews, Andrea L. Miller, Thomas Plötz, and Ilias Kyriazakis. 2017. Automated Tracking to Measure Behavioural Changes in Pigs for Health and Welfare Monitoring. *Scientific Reports* 7, 1 (2017), 17582. doi:10.1038/s41598-017-17451-2

[19] Simon G. Matthews, Andrea L. Miller, Thomas Plotz, and Ilias Kyriazakis. 2017. IoT at the Grassroots: Exploring the Use of Sensors for Livestock Monitoring. In *Proceedings of the 2017 IEEE International Conference on Internet of Things (IThings).* doi:10.1109/iThings-GreenCom-CPSCom-SmartData.2017.44

[20] Microsoft. 2019. ONNX Runtime: Cross-Platform, High Performance Machine Learning Inferencing and Training Accelerator. GitHub. https://github.com/microsoft/onnxruntime

[21] Mrinmoy Modak, Muin Mustahasin Pritom, Sajal Chandra Banik, and Md Sanaul Rabbi. 2025. Internet of Things-Based Health Surveillance Systems for Livestock: A Review of Recent Advances and Challenges. *IET Wireless Sensor Systems* (2025), e70013. doi:10.1049/wss2.70013

[22] Dario A. B. Oliveira, Luiz G. R. Pereira, Tiago Bresolin, Rodrigo E. P. Ferreira, and Joao R. R. Dorea. 2021. A Review of Deep Learning Algorithms for Computer Vision Systems in Livestock. *Sensors* 21, 22 (2021), 7540. doi:10.3390/s21227540

[23] OSU Small Ruminant Team. 2022. Heat Stress in Small Ruminants. Ohio State University Extension. https://u.osu.edu/sheep/2022/05/24/heat-stress-in-small-ruminants-2/

[24] Patchayut Phupattanasilp and Sao-Ren Tong. 2019. Augmented Reality in the Integrative Internet of Things (AR-IoT): Application for Precision Farming. *Sustainability* 11, 9 (2019), 2658. doi:10.3390/su11092658

[25] C. J. Rutten, A. G. J. Velthuis, W. Steeneveld, and H. Hogeveen. 2013. Invited Review: Sensors to Support Health Management on Dairy Farms. *Journal of Dairy Science* 96, 4 (2013), 1928–1952. doi:10.3168/jds.2012-6107

[26] Mingxing Tan and Quoc V. Le. 2019. EfficientNet: Rethinking Model Scaling for Convolutional Neural Networks. In *Proceedings of the 36th International Conference on Machine Learning (ICML).* PMLR, 6105–6114.

[27] P. Thornton, G. Nelson, D. Mayberry, and M. Herrero. 2022. Impacts of Heat Stress on Global Cattle Production During the 21st Century: A Modelling Study. *The Lancet Planetary Health* 6, 3 (2022), e192–e201. doi:10.1016/S2542-5196(22)00002-X

[28] Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N. Gomez, Lukasz Kaiser, and Illia Polosukhin. 2017. Attention is All You Need. *Advances in Neural Information Processing Systems* 30 (2017).

[29] Chidike Vincent, Adriana Crainic, Jonathan Zasada-James, and James Baldwin. 2025. Real-Time Edge Intelligence in UAV for Search-and-Rescue: Onboard Energy-Efficient Video Summarisation with Reduced Data Transmission. In *Proceedings of the 15th International Conference on the Internet of Things (IOT 2025).* ACM, 52–58. doi:10.1145/3770501.3770508

[30] Chidike Vincent, Muhammad Aslam Jarwar, Najam Ul Hasan, Laurence Hirsch, Jonathan Zasada-James, and Hem Raj Dadral. 2026. Privacy-Preserving ECG-HRV Emotion Classification for Mood Swing Detection using Federated Learning on Constrained Edge Devices. In *2026 IEEE Conference on Computational Intelligence in Bioinformatics and Computational Biology (CIBCB).* https://shura.shu.ac.uk/id/eprint/37632

[31] Christopher M. Wathes. 2009. Precision Livestock Farming for Animal Health, Welfare and Production. In *Sustainable Animal Production: The Challenges and Potential Developments for Professional Farming*, A. Aland and F. Madec (Eds.). Wageningen Academic Publishers, Wageningen, 411–419.

[32] Junfeng Xu, Kristina Kacmorova, Dan Tulpan, and Minglun Gong. 2024. Systematic Literature Review of Vision-Based Approaches to Outdoor Livestock Monitoring with Lessons from Wildlife Studies. *arXiv preprint arXiv:2410.05041* (2024).