Letters in Animal Biology

Journal homepage: www.liabjournal.com

Animal affective-state analysis using Convolutional Neural Networks: Methods, datasets, applications, challenges, and future directions

Biswajit Mandal *, Dilip Kumar

Department of Computer Science and Engineering, National Institute of Technology Jamshedpur, Jamshedpur, India

Article info

Received: 25 September 2026
Received in revised form: 04 November 2026
Accepted: 06 November 2026
Published online: 10 November 2026

Keywords

Animal emotion states
Animal emotion recognition
Deep learning
Animal welfare
Convolutional Neural Networks
Facial expression analysis

* *Corresponding author:*
Biswajit Mandal
Email: emailtobiswajit@gmail.com

Reviewed by:

Rotimi-Williams Bello
Faculty of Information and Communication Technology, Tshwane University of Technology, Pretoria, South Africa

Consent to publish the name of second reviewer could not be obtained

Abstract

Animal emotions are inner affective and bodily states that make certain physiological, behavioral, and survival responses to environmental changes. Comprehension of the animal behavior and emotional states is pivotal a step to improve human-animal relationship and enhance animal welfare. Convolutional Neural Networks (CNNs) are among the popular methodologies to analyze the animal-emotion related data. However, CNNs face challenges such as dataset scarcity, inter-breed morphological variability, and challenges in associating subtle behavioral micro expressions with emotion labels. CNNs help predict animal emotions via successive convolutional layers, images pass through these layers and automatically extract features like eyebrow, mouth curvature, and classify them into different emotional states. Various methods, including AlexNet, ResNet, EfficientNet, MobileNet, have been commonly used for feature extraction and classification. Leveraging the potential of CNN in image analysis and data generation, it can be applied to animal emotion-state analysis. This review outlines the fundamental issues of emotion classification and information retrieval by CNN techniques; and critically evaluates the available literature on the CNN applications in animal affective-state analysis, advocates research gaps, and proposes future research direction. Combining CNNs with data augmentation technique can greatly enhanced model accuracy and prediction, and addresses the issues with small training datasets. Various methods for emotion detection and very few available datasets used by researchers or to used by other researchers for future research are there. The use of CNN-based approaches will greatly benefit the researchers to enhance their studies and help them develop and implement real-world applications.

This is an open access article under the CC Attribution license (http://creativecommons.org/licenses/by/4.0/)

1. Introduction

Animal emotions are short-term affective states, comprising behavioral, physiological, and presumably subjective components, that mediate responses to environmental events (Paul et al. 2005; Mendl et al. 2010). In non-verbal animals, subjective experience cannot be observed directly and must be inferred from indicators. In this review, animal emotional state therefore means an affective state inferred from observable behavioral, facial, vocal, or physiological indicators. These indicators are treated as proxies, not as direct measurements of subjective experience. The existing literature occasionally uses the terms emotion, affect, behaviour, stress, pain, welfare, and sentiment interchangeably. However, this review distinguishes them in two ways. First, affect is an umbrella term, described in terms of valence (positive to negative) and arousal (low to high); discrete labels such as fear or frustration are categories placed under this term. Second, this review does not use the term sentiment for animals, because it denotes opinion polarity in human language processing. Table 1 distinguishes five levels of what a model can actually recognize.

Classification accuracy measured against human-assigned labels shows that a model agrees with its annotators. It does not by itself establish that the label reflects the animal's emotional state. Human-centred labels such as "happy", "sad" or "angry" may reflect annotator interpretation rather than animal biology (Descovich et al. 2017). Each included study is therefore assigned a level in Table 2, and conclusions in this review are limited to that level.

This is a focused review of deep-learning approaches to animal affective-state recognition in which convolutional neural networks (CNNs) are the principal modelling backbone. Vision Transformers (ViTs), hybrid CNN–transformer models, temporal models and multi-modal models are included for comparison. Because related work differs in how directly it informs affective-state recognition, the literature is separated into three tiers and used only the first as evidence for the review questions. Table 2 shows the classification of studies according to the scope of this review.

2. Review methodology

This review followed the PRISMA 2020 statement (Page et al. 2021). The review addressed the following research questions: What CNN designs have been utilized for recognizing animal emotional states and which ones exhibited the best performance? What datasets, data acquisition methodologies, and emotion-tagging approaches have been employed in current research? To what extent can CNN-based models differentiate among various animal emotional states, including fear, aggression, pain, tension, and positive affect? What are the primary

Table 1. Different levels of information a model can recognize

Level	What the model recognizes	Typical ground truth	What can be claimed	What cannot be claimed
1. Behaviour recognition	Postures and activities (for example ear position, barking, lying)	Observer-coded behaviour	The behaviour occurred	Any internal state
2. Facial-expression or action-unit recognition	Facial configurations, coded action units, grimace-scale items	Trained coders using a coding system	A specific facial movement was present	The emotion behind it
3. Physiological-state recognition	Heart rate, heart-rate variability or stress hormone level predicted from images or video	Sensors or assays	A correlate of arousal	Valence or a specific emotion
4. Affective-state inference	Categories such as stress, pain, fear or relaxed	Context-defined or expert-judged labels	The model reproduces the reference labelling process	That the label equals the animal's subjective state
5. Validated emotion recognition	A category supported by independent, converging evidence, tested on unseen animals	Behaviour, physiology and context agreeing, with reliable raters	A reliable proxy under the conditions tested	Direct measurement of subjective experience

constraints of current CNN-based methodologies concerning dataset magnitude, class disparity, picture fidelity, and environmental fluctuations? The literature search was restricted to English-language publications published during 2021-2026 (Until 23-09-2026). Reference lists of the included studies and related reviews were also screened. The core search string was adapted to the syntax of each database and applied on 23-09-2026. The search string was: "CNN" OR "convolutional neural network" AND "animal emotion" OR "animal affect" OR "animal emotional state" OR "animal sentiment".

Studies were included if they (i) were peer-reviewed journal articles or full conference papers; (ii) used non-human animals; (iii) applied CNN-based model to classify or estimate an emotional, affective, pain or stress state, or a behavioral or facial indicator explicitly linked to one; and (iv) reported an empirical evaluation. Studies were excluded if they involved only humans, contained no animal data, performed detection or tracking without any state classification, were reviews or editorials or gave too little methodological detail to extract the data items. Following the inclusion of relevant studies, titles and abstracts were screened, full texts of eligible studies were assessed, and the reasons for exclusion at the full-text assessment stage were recorded. Data were extracted with a standard form covering: species, number of animals, number of images or clips, setting, modality, affective categories, label source and inter-rater agreement, model and architecture family, split type (random frame, clip, animal-level, cross-dataset), metrics, confidence intervals, explainability method, data availability, and stated limitations. The PRISMA 2020 flow chart used for the current systematic review is displayed in Fig. 1. Also, Fig. 2 shows the year-wise distribution of publications in this area. The search identified 141 records in Google Scholar and 22 additional records through citation searching, and after removal of duplicates the total number of records was reduced to 102. The screening of title and abstracts of available records further reduced the number to 84 and the full texts of only 66 records were assessed. Among the 66 full-text records 22 records were excluded and the remaining 44 full-text records were included in this review. However, given the rapid developments in this research area, the search strategy adopted in this review may have missed more recently published studies.

3. Evidence synthesis framework

The studies included in this review were combined thematically rather than presented as a study-by-study catalogue. This review followed PRISMA 2020 reporting principles, which emphasize transparent study selection, reproducible data extraction, and clear presentation of findings and limitations (Page et al. 2021). The evidence was arranged according to six themes: (1) study characteristics; (2) species and data modalities; (3) emotional-state labels and reference standards; (4) CNN architectures and learning strategies; (5) validation procedures and performance assessment; and (6) laboratory versus real-world performance. This structure is appropriate because animal emotional-

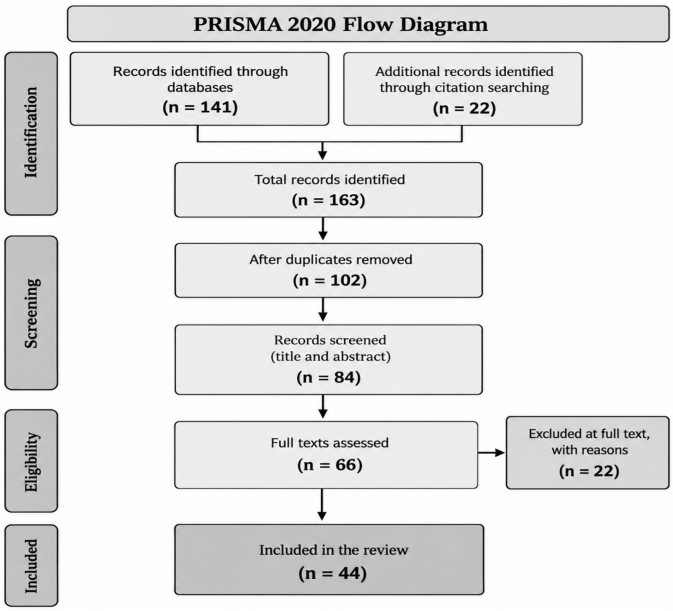


Fig. 1. PRISMA 2020 flow diagram of the selection process of the studies

Table 2. Scope and classification of the reviewed literature

Level	Content
1. Core evidence	Studies that classify emotional, affective, pain or stress states in animals
2. Adjacent animal-vision work	Animal detection, individual identification, livestock posture and activity recognition, general species classification
3. General background	Methods from human emotion recognition or other image domains

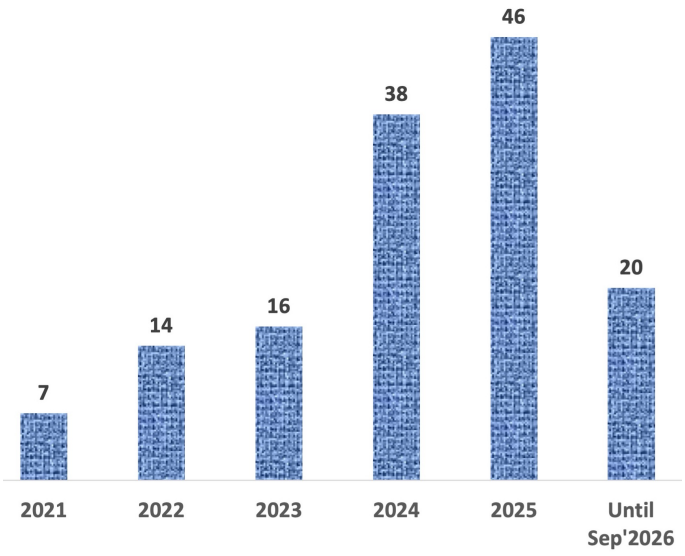


Fig. 2. Annual distribution of animal emotion analysis (AEA) related publications (Google Scholar)

state research includes a range of species, experimental circumstances, data formats, and biological outcomes. For example, Coutant et al. (2024) reviewed bio-acoustic methods for assessing farm-animal well-being and recognized applications related to health, behavior, physiology, and affect, highlighting the need for further research before wider practical application. The synthesis distinguishes between direct and indirect indicators of emotional states, such as facial expressions associated with pain, vocalizations associated with tension, and behavioral changes. These indicators should not be interpreted as definitive evidence of a clinical condition or as indicative of the underlying mental state of an animal. The strength of any result will therefore depend on the quality of the reference labels, the study design, and the level of independent validation.

3.1 Study characteristics and species coverage

For each study included in this review, the publication year, species and breed, sample size, experimental or field setting, study aim and outcome, data source, and study strategy were extracted. The species represented in the included studies included dogs, cats, horses, cattle, pigs, sheep, goats, laboratory rats, and other animals. Each species included in a study is recorded separately, and each study was given a unique identifier to prevent duplicate counting. The results are summarized using frequency tables and descriptive statistics. The number of animals, recordings, images, or video clips, as well as the number of studies for each species, are provided wherever available. These measures remain distinct since a large number of recordings does not always imply a large or representative animal sample. The synthesis also distinguishes controlled laboratory research from studies carried out in homes, farms, veterinary clinics, and other real-world contexts.

3.2 Data modalities and annotation methods

Studies were categorized using five modalities: still images, video, audio, physiological or sensor data, and multi-modal data. Physiological or activity measurements may be used in sensor-based research; movement and behavioral sequences may be examined in video studies; vocalizations and other acoustic signals may be investigated in audio studies; posture, facial expressions, or pain-related grimace patterns may be examined in still-image studies. Multi-modal studies are those that incorporate two or more input modalities. The labelling procedures used to identify emotional states or health-related outcomes in each study were documented in the review. Potential reference standards include expert behavioral assessments, validated pain scales, veterinary diagnoses, conditions identified through experimentation, and physiological data. Labels based on direct observation and those derived from proxy measures are distinguished in the review. This distinction is very important in bio-acoustic research, as vocal features may be linked to affect or welfare without causing a unique emotional state (Coutant et al. 2024).

3.3 CNN architectures and learning strategies

The architectural synthesis was used to classify the models used in the included research, including conventional CNNs, CNN–recurrent neural network combinations, pre-trained CNNs using transfer learning, and hybrid CNN-based classifiers. The review ascertains the role of each hybrid approach component, such as whether the CNN extracted features from spectrograms or images and whether the final prediction was made by a different classifier. The extracted data contained input representation, preprocessing, feature extraction, model initialization, augmentation, class-balancing strategies, and training methodology. The review describes how audio-based research used raw waveforms, spectrograms, mel-spectrograms, or mel-frequency cepstral coefficients. When reported for image and video investigations, it records temporal aggregation methods, frame-selection procedures, and image resolution. The synthesis differentiates CNN-based research from approaches that used only non-CNN architectures or traditional machine learning. These studies may provide useful contextual information, but they were not considered direct evidence of CNN performance.

3.4 Validation practices and performance evaluation

The methodological synthesis examined how each study divided its data into training, validation, and test sets; whether hyperparameter selection was done independently of the final test set; and whether validation was carried out across animals, recording sessions, facilities, or datasets. Particular attention was paid to data leakage when the same animal appeared in many recordings or frames across training and test sets. Accuracy, precision, recall, F1-score, sensitivity, specificity, and area under the receiver operating characteristic curve were among the performance metrics reported in the included studies. The results were compared only when the evaluation processes, class definitions, and prediction tasks were sufficiently similar. Furthermore, accuracy and F1-score may be supplemented when necessary by the Matthews correlation coefficient, particularly in imbalanced binary-classification settings (Chicco and Jurman 2020). When available, model interpretability was noted. Saliency-map sanity checks and gradient-based visual explanations, for instance, can be used to evaluate what features a CNN appears to employ for its predictions; nevertheless, interpretability visualizations alone do not establish biological validity or causal relevance (Adebayo et al. 2018; Chattopadhay et al. 2018).

3.5 Laboratory versus field performance

The model performance under controlled conditions and operational field setting was compared in the final theme. Consistent illumination, standardized recording equipment, precisely defined labels, and controlled stimuli can all be found in laboratory datasets. Background noise, inconsistent lighting, occlusion, breed variations, environmental variance, and ambiguous behavioral labels can all be present in field

Table 3. Different methods for label creation

Label source	Examples	Strength	Weakness
Validated coding systems	DogFACS, CatFACS, EquiFACS; grimace scales for horse, cat, sheep and rodents	Standardized, reproducible, training-based	Codes movements, not emotions; labour-intensive
Expert judgement	Ethologists or veterinarians rating behaviour	Domain knowledge	Subjective unless agreement is measured
Physiological measures	Cortisol, heart-rate variability, temperature	Independent of human interpretation	Reflects arousal more than valence; invasive or noisy
Context or stimulus-defined	Labels set by the situation (for example vet visit equals stress)	Objective assignment	Assumes the context produced the state
Owner or crowd labels	Owner surveys, online annotators	Large scale, cheap	Highest risk of anthropomorphic bias

datasets. The synthesis included documentation of external datasets, internal validation, animal-independent testing, and validation across farms, clinics, or other places. Performance evaluated using randomly selected samples from the same datasets was separated from evidence of external validity. This approach assisted in determining whether reported CNN results are likely to generalize outside of the context in which the models were created.

3.6 Species-wise evidence strength

The operational definitions used in this review were as follows. *Strong*: Several independent datasets, reliable reference standards, and proof of external or animal-independent validation. *Limited*: Although there is some pertinent data, there is not enough independent validation, sample variety, or methodological consistency. *Minimal*: The evidence is scant, tentative, or derived from a small number of research studies with significant methodological flaws. It is not appropriate to present these categories as a validated universal grading system; rather, they are suggested for this evaluation. Instead of focusing only on the number of publications, the evaluation took study independence and quality into account. When there was insufficient evidence to make a confident decision, the review clearly reflects ambiguity.

4. Label validity and ground truth

Object-class labels and emotional-state labels are fundamentally different. An image of an animal may reliably be identified as a dog but identifying an animal in the image as "happy" is an interpretation of its observed features. The methods used to create labels in the included research are summarized in Table 3. For categorical labels, inter-rater agreement should be recorded using Cohen's or Fleiss' kappa, and for scores, intraclass correlation (Cohen 1960; Landis and Koch 1977; Koo and Li 2016). Labels assigned by a single annotator without an agreement statistic or independent validation were treated as having very low evidential reliability in this review. Where possible, labels were triangulated using behavioural, physiological, and contextual evidence to described animal emotions.

4.1 Data leakage and validation hierarchy

Data leakage occurs when information from the test set influences training, so that performance overstates generalization (Kaufman et al. 2012; Kapoor and Narayanan 2023). In animal emotion recognition, a common source of data leakage is the splitting of image or video frames at random, which places near-identical frames of the same animal, session, and background in both training and test sets. The model then recognizes the individual or the setting, not the affective state. Other sources of data leakage are augmenting images before splitting, fitting normalization on the full datasets, repeated sessions in one environment, and tuning hyperparameters on the test set. Different sources of data leakage and their potential effects are shown in Table 4. Results from levels 0 and 1 should be interpreted as upper bounds. Grouped or leave-one-animal-out cross-validation, data augmentation applied after the split and within the training folds only, normalization fitted on training data only, and reporting the number of animals in each fold are generally recommended.

4.2 Dataset Taxonomy

Table 5 shows a taxonomy of the datasets used in the selected studies, including species, number of animals, numbers of images and recordings, categories, annotation method, data source and setting, modalities, and expert involvement.

4.3 Temporal and multi-modal approaches

Emotional expression unfolds over time and across channels, so a single still image may capture it poorly. This section compares image-based CNNs with temporal and multimodal approaches reported in the included studies. Table 6 presents a summary of different models and their advantages.

5. A methodological framework for animal emotion recognition

Table 4. Grouping of validation designs according to the level of testing

Label	Design	What it tests	How to read the results
0	Random frame-level split	Interpolation between near-duplicates	An optimistic upper-bound estimate only
1	Clip- or session-level split	Robustness to new moments	Still allows identity leakage
2	Animal-level (subject-independent) split	Generalization to unseen individuals	Minimum credible validation standard
3	Cross-breed or cross-group validation	Generalization across morphology	Stronger evidence
4	Cross-dataset or cross-site validation	Robustness to camera, environment and labelling	Strong evidence
5	Prospective field validation	Real-world performance	Required before welfare use

Table 5. Dataset Taxonomy

Dataset / source	Species	Animals	Items	Categories	Annotation method	Setting	Modality	Expert involvement
Bremhorst et al. (2019) dataset used by Boneh-Shitrit et al. (2022)	Dog	29	248 clips	Positive anticipation; frustration	Induced context; DogFACS	Laboratory	Video	Coders (DogFACS)
DogFLW (Martvel et al. 2025)	Dog (many breeds)	Not reported	3,732 images	46 landmarks	Manual landmark annotation	Web images	Images	Not reported
Owner-filmed fireworks videos (Gähwiler et al. 2020; Martvel and Riemer 2025)	Dog	42 (11 analyzed)	Not reported	Fireworks vs control	Context	Home	Video	Not reported
Kopczyński and Czubenko (2026)	Dog (120 breeds)	Not reported	About 10,000 (30,000 augmented)	Five classes	Human annotators; κ about 0.90	Web photographs	Images	5% expert review
Bhave et al. (2024)	Dog (10 breeds)	Not reported	2,184 images	Seven Panksepp categories	Researcher labels	Curated images	Images	Not reported
Feighelstein et al. (2022)	Cat	29	Not reported	Pain / no pain	Time point after surgery	Clinical	Images	Veterinary team
Feighelstein et al. (2024)	Cat	84	Not reported	Pain / no pain	Glasgow composite measure pain scale	Clinical	Images	Veterinary experts
Lencioni et al. (2021)	Horse	7	Not reported	Pain: none, moderate, obvious	Horse Grimace Scale	Feeder station, clinical	Images from video	Researcher with professionals
Briefer et al. (2022)	Pig	411	7,414 calls	Positive vs negative valence; context	Context of production	Commercial farms	Audio	Ethologists

Fig. 3 shows the proposed framework for animal emotion recognition, which potentially provides a more specific workflow for an effective animal affective-state recognition. Each stage explains the required steps, quality-control measures, and potential shortcomings observed

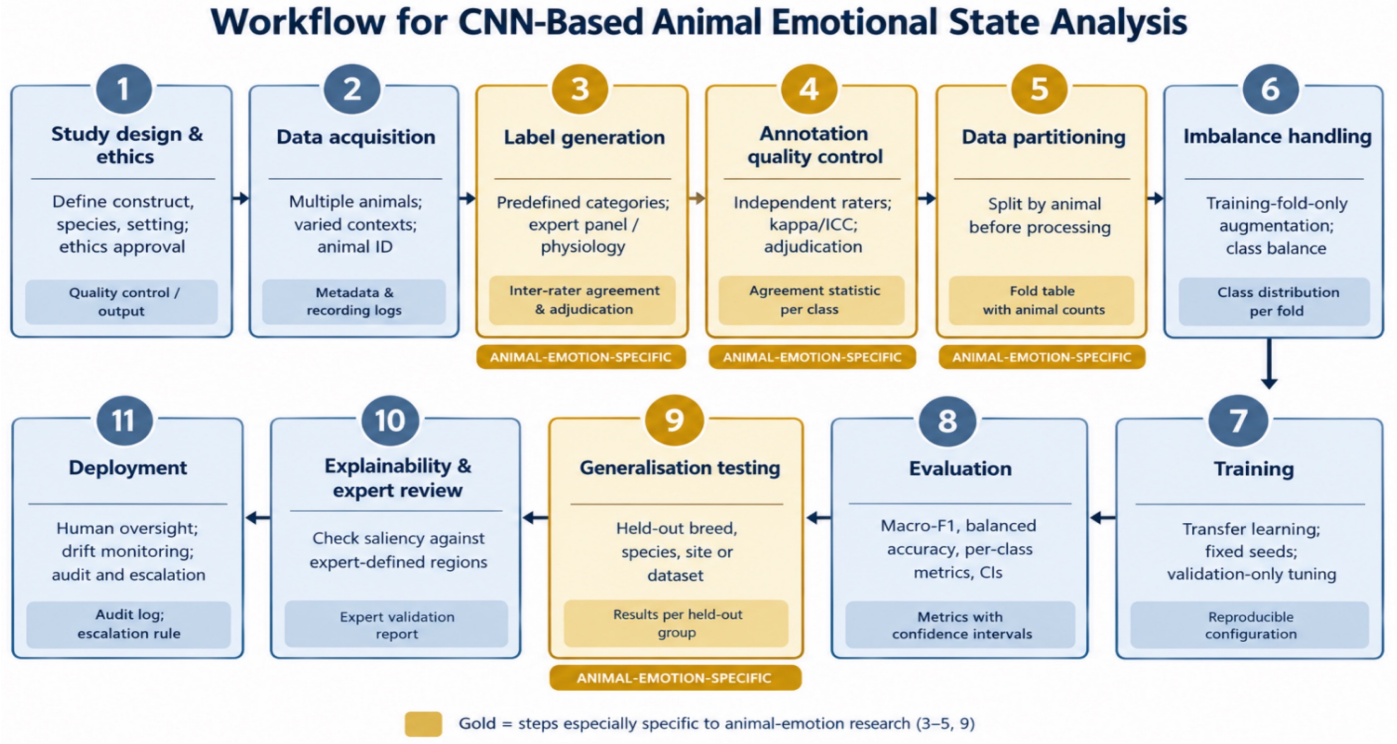


Fig. 3. Methodological framework for animals emotional-state analysis using CNN-based approaches

Table 6. Different models used in animal affective-state analysis

CNN architecture	Animal/ datasets used	Accuracy	Key highlights	Reference
EfficientNetB5	dogs and cats	98.2% accuracy in training and 91.24% during validation	Suggested an altered EfficientNetB5 with a channel-recalibration/SE method for nuanced facial-emotion detection in household pets. Attained a precision of 90.98%, a recall of 92.14%, an F1-score of 91.20%, and a Cohen's kappa of 91.88%. The model was evaluated against MobileNet, VGG16, InceptionV3, AlexNet, DenseNet, Xception, and ResNet50.	Sumon et al. (2025)
YoloV3, a Faster YoloV4-based facial detection platform and an ensemble Convolutional Neural Networks (RCNN)	Cows and pigs; 7,800 images and >150 videos from 235 pigs and 210 dairy cows, collected from six farms	R-CNN's accuracy score (93.11% on training and 89.19% on validation set) outperforms both YOLOv4 (89.96% on training and 86.45% on validation set) and YOLOv3 (85.21% on training and 82.33% on validation set)	Created WUR Wolf, a real-time face coding system for recognizing emotions in farm animals. Identifies 13 facial movements and deduces nine emotional states through the analysis of the eyes, ears, snout, and overall face expression. YOLOv4 attained an average F1 score of 88.33% and offered a harmonious blend of precision and computational efficacy.	Neethirajan (2021)
YOLOv8	Dogs; Shetland Sheepdog Emotion Dataset and Stanford Dogs Dataset; five emotions: anger, joy, anxiety, neutral, trust	accuracy of 90% on the test set	Identifies eyes, ears, and muzzle utilizing YOLOv8 and amalgamates their emotional forecasts via confidence/count weighting. CSPDarkNet53 attained a mean Average Precision (mAP) of 96.4% for the detection of face features. The research explicitly examines generalization among various dog breeds and datasets.	Li et al. (2026)
EfficientNet-B0	Dogs; canine facial-emotion dataset	85% validation; 90% training	A lightweight approach for recognizing four canine emotions: anger, happiness, relaxation, and sadness. Employed data augmentation, AdamW optimization, and adaptive learning rate modification. The model comprises over 5.3 million parameters and was engineered for computational efficiency.	Li et al. (2025)
VGG 16 Model with transfer learning	dogs	accuracy of 96.33%	Proposed emotion recognition across species via a VGG16 framework with Ada-Sig-Max optimization. The described algorithm utilizes facial expression data and aims to broaden emotion identification from humans to dogs. Demonstrated superior classification efficacy across emotional categories.	Vikramathithan et al. (2025)
VGG16-EmotionNet	Human + canine/cross-species emotion images	96.33%	Utilizing transfer learning; seven categories of emotions; a cross-species framework for emotion recognition.	Vikramathithan et al. (2025)
faster R-CNN	440 images of horses were collected from private sources	accuracy of 80% on the validation set and 65% on the test set,	Created a horse emotion recognition proof-of-concept. A Fast R-CNN detector identifies horses, then a CNN emotion classifier using head, neck, ears, muzzle, and eye positions. Validated automatic equine emotion classification.	Corujo et al. (2021)
YOLOv3	Dogs; 3 video datasets for tracking and 2 video datasets for emotion recognition	81.73% and 76.02% for emotion recognitionv	Engineered an automated monitoring system that incorporates canine detection, tracking, and emotional assessment. The precision of tracking attained 93.02% for individual dog videos and 86.45% for footage featuring numerous dogs. The performance was influenced by blurriness, low resolution, angle of view, and swift or subtle motion.	Chen et al. (2023)
Custom CNN	Dogs; Kaggle dog-emotion image dataset	accuracy of 99.60%	Categorized four emotions: sorrow, joy, tranquility, and rage. CNN trained with SGD significantly surpassed the established baselines of SVM (89%) and KNN (81%). The research illustrates the capability of CNN-driven facial emotion categorization.	Tanwar (2024)
Multi-QuadEmoNet (LSTM–GRU); CNN10/CNN14 used for comparison	Dogs and cats; DMAD and CMAD vocalization datasets; 2,000 dog and 6,000 cat vocalizations	95% DMAD; 90% CMAD	Audio-driven sentiment identification utilizing MFCC attributes and a multi-tiered LSTM-GRU framework. The suggested model surpassed CNN10, CNN14, and YAMNet. CNN10 had a performance of 90% on DMAD and 72.5% on CMAD.	Pearline et al. (2026)
VGG16 with transfer learning	Dogs; facial-expression image dataset	88.05% validation; 96.72% training	Categorized four canine emotions: rage, joy, tranquility, and sorrow. Transfer learning utilizing VGG16 was employed for the extraction of facial features. The F1-score during testing was 84.30%, suggesting that validation accuracy should not be the exclusive metric for evaluating performance.	Tribethran et al. (2024)
DER-TLNN / transfer learning; 3D-CNN/VDERNet	Dogs; 400 static images plus small video-clip dataset	89% for static-image emotion classification; video model also evaluated	Examined wrath, anxiety, and happiness through the utilization of compact databases. Two veterinary specialists annotated the static photographs. The research notably tackles the issue of emotion identification with constrained training data and advances image-centric recognition to video-oriented 3D-CNN evaluation.	Hernández-Vanegas and Rodríguez-Romo (2026)

Table 6 (Continued). Different models used in animal affective-state analysis

CNN architecture	Animal/ datasets used	Accuracy	Key highlights	Reference
CReToNeXt-YOLOv5	Pigs; self-labelled pig facial-expression image dataset	89.4% mAP	Identified four distinct facial expressions in pigs: joy, neutrality, fear, and anger. Implemented Coordinate Attention, CReToNeXt module, and EIOU loss within YOLOv5. Improvements to YOLOv5-based pig facial-expression recognition. Attention mechanisms and architectural changes improved joyful, neutral, frightened, and furious expression recognition Attained a 6.7% enhancement compared to the original YOLOv5, with an F1-score of 0.84.	Nie et al. (2024)
CNN-based models including ResNet, MobileNet, EfficientNet; also ConvNeXt and YOLOv11	Dogs; newly constructed canine-emotion image dataset	84% with YOLOv11; 75% for MobileNet + EfficientNet + ResNet50 ensemble	Integrated transfer learning with Explainable AI (Grad-CAM) to pinpoint image areas accountable for emotion forecasting. A total of ten categorization methodologies were assessed. YOLOv11 attained an accuracy of 84% and an F1-score, whereas DINOv2-SVM reached 75%.	Kopczyński and Czubenko (2026)
Deep-learning image classification models; CNN-based deep learning framework	Dogs and cats; 52 dog species and 23 cat species, >300 images/ species	-	Created the HappyPets mobile application for concurrent classification of species and emotions. The dataset encompassed a wide array of canine and feline breeds/species. The practical implementation produced over 113,000 predictions; emotion classification typically outperformed species classification, while the dataset was predominantly composed of happy/ neutral instances.	Sinnott et al. (2021)
ResNet50, GoogleNet/ Inception, EfficientNet; YOLOv5	Dogs and cats; 16,000-image test dataset	91% ResNet50; 89% GoogleNet; 89% EfficientNet; 87% YOLOv5	Evaluated CNN classification models against YOLOv5 object detection for recognizing emotions in dogs and cats. ResNet50 achieved the most elevated documented classification accuracy of 91%. The research examined feelings like anger, relaxation, sadness, and happiness.	Wu (2023)
CNN/deep-learning emotion and barking recognition framework	Dogs; image/audio-based pet social-network data	75%, improved to 95.70% using weighted combination of emotion and barking predictions	Created an emotional recommendation algorithm for a pet social platform. Integrating canine emotional recognition with bark identification significantly enhanced the overall predictive accuracy. The authors emphasized the necessity for multimodal datasets and validation across various dog breeds and settings.	Cheng et al. (2022)
CNN/deep-learning video model	Horse; controlled facial-expression videos	76%	Four emotional states; video-based deep learning surpassed EquiFACS-based machine learning; anticipation and frustration continued to pose challenges.	Feighelstein et al. (2024)
ResNet50 CNN	Domestic short-haired cats; 29 cats, facial images collected during ovariohysterectomy at different pain intensities	>72%	Created facial image-based cat pain recognition. ResNet50 vs. catFACS-inspired geometric-landmark method. Both methods were accurate above 72%. CNNs can objectively assess animal welfare and suffering, according to the study.	Feighelstein et al. (2022)
MobileNetV2	Kitten; posture dataset		Recognition of defensive and offensive aggression; a lightweight framework tailored for mobile implementation.	Sui et al. (2024)
Faster R-CNN	Cattle; Hartpury University Farm dataset	94/92/89%	Identification of stationary, recumbent, and ambulating states; instantaneous observation of livestock behavior.	Ahmad et al. (2023)
CNN	Dogs/cats/pets; facial-expression dataset		CNN-based facial emotion recognition for pets.	Jamalpur et al. (2024)
IFR-CNN + RNN	Human speech; EMODB & GEES	89.5/94.82%	Audio sentiment identification; MFCC combined with spectral and temporal characteristics; valuable methodological guide for multimodal investigation of animal emotions.	Suneetha and Anitha (2025)
ProtoNet + ResNet34	Dogs/cats; few-shot pet-emotion dataset	92.5%	Few-shot learning; classifications of anger, happiness, and sadness; tackles constrained labeled datasets of animal emotions.	Gilhotra et al. (2025)
EfficientNet + Transfer Learning	Dogs; dog facial-emotion image dataset	78%	Created an EfficientNet-based transfer-learning system for dog emotion recognition. Focuses on parameter optimization and training parameter effects on recognition.	Xie (2023)
Deep CNN / transfer-learning-based facial analysis	Elephants; elephant facial-expression images	——	Examines elephant emotional facial expressions. The study is useful for distinguishing emotions from very small facial changes in a species with extremely distinct face morphology from domestic animals.	Priya et al. (2026)
Custom CNN	Goats; publicly available goat-vocalization dataset collected at diverse farms	95.8% average classification rate	From goat vocalizations, classified emotions. Using pitch shifting and temporal stretching for augmentation. A key contribution is XAI analysis of CNN decision-making time-frequency areas.	Ntalampiras and Pesando Gamacchio (2025)

Table 6 (Continued). Different models used in animal affective-state analysis

CNN architecture	Animal/ datasets used	Accuracy	Key highlights	Reference
Deep-learning CNN approach	Dogs; 29 Labrador Retrievers, controlled experimental videos/ images	>89%	Automated dog facial expression recognition for eagerness and frustration. Compared deep learning to DogFACS (>71%). Heatmaps illustrated facial regions affecting predictions.	Boneh-Shitrit et al. (2022)
ResNet50; ViT; DINO-ResNet50; DINO-ViT	Dogs; 248 videos from 29 Labrador Retrievers, producing 19,392 labelled facial frames	85.3% validation with DINO-ViT	Comparison of self-supervised versus supervised dog facial emotion recognition. The most accurate validation was DINO-ViT. Task contrasted positive expectation with frustration. A subject-independent train/test split was utilized to clip dog faces with Mask R-CNN.	Boneh-Shitrit et al. (2022)
MoCo contrastive-learning CNN encoder + downstream classifier	Horses; 3,929 images from five wild horse breeds	42.1–65.56%, depending on emotion	Unsupervised/self-supervised learning to recognize horse emotions without emotion labels during representation learning. Assessed care, fear, sadness, lust, pain, play, fury, and seeking. Lust had the highest emotion accuracy (65.56%); wanting 42.1%.	Bhave et al. (2025)
MoCo with ResNet18/ResNet34 encoders; comparison with supervised ResNet50	Dogs; 2,184 images, 10 breeds, seven Panksepp-inspired emotion categories	43.2% (MoCo); 74.32% supervised ResNet50	Introduced facial expression and body posture-based Momentum Contrast (MoCo) for unsupervised canine emotion recognition. The unsupervised model outperformed the 14% random baseline but lagged ResNet50. A second public dataset was 48.46% accurate.	Bhave et al. (2024)
ResNet50-based DeepLabCut	healthy adult BALB/ c mice (both from the Beijing Vital River Laboratory Animal Technology Co., Ltd., 8-12 weeks old)	86.7%	Created multimodal fear-emotion recognition. Facial expression, posture, and trajectory obtained 73.3%, 78.3%, and 69.0%, respectively, whereas multimodal fusion achieved 86.7%, indicating complementing information.	Wang et al. (2025)
YOLOv5, YOLOv8, YOLOv9	Elephants; labelled elephant-emotion dataset	93%, 95%, 95% test accuracy	Anger, happiness, and sadness elephant emotions. YOLOv5, 8, and 9 obtained 93%, 95%, and 95% test accuracy. Eigen-CAM was used to describe and investigate image resolution.	Subair and Benedict (2024)
EquiFACS, Yolov5	Horses; controlled facial-expression video dataset	76%	Automatic baseline, positive expectation, disappointment, and frustration recognition. Video-based deep learning beat EquiFACS-based ML (76% vs. 69%). As a two-class task, anticipation and frustration were hard to separate with 61% accuracy.	Feighelstein et al. (2024)
Xception CNN	Pigs; PER dataset from Wageningen University and Research	92.9% IAF; 94.7% PAF after pig detection	Proposed Semi-Shuffling and Pig Detector–Pig Emotion Recognition (SSPD-PER). After deleting irrelevant background pixels with pig detector, Xception was utilized. Semi-shuffled testing significantly lowered apparent performance, revealing conventional evaluation data-leakage/bias issues.	Kim et al. (2023)
ResNet50 transfer learning; compared ResNet18, VGG16, AlexNet, MobileNet and GoogLeNet	Dogs; 19,921 Kaggle images, four classes: angry, happy, relaxed and sad	57.6% test accuracy after ResNet50 fine-tuning; 53% with UDA	Contrast pretrained CNNs versus transfer-learning. Although VGG16 has the highest training accuracy (58%), ResNet50 was chosen due to its lower computational cost. Fine-tuning improved test accuracy from 53.3% to 57.6%. Subjective labeling, breed/scene variance, and domain shift are highlighted in the study.	Du et al. (2023)
Improved YOLOv5 with Filter_Attention and Pooling_Module	Cattle; 1,723 images, with body, head, legs and tail manually annotated	90.74%	Recommended improving YOLOv5 for cattle key-part detection. Noise is reduced using Filter_Attention, and SoftPool-based Pooling_Module improves leg and tail detection. It enhanced AP/F1 for specific body sections and obtained 90.74% mAP.	Shao et al. (2023)
MobileNetV2	Animal image dataset; 26,179 10 classes of Kaggle images: dog, horse, elephant, butterfly, chicken, cat, cow, sheep, spider, squirrel.	95%	Tried lightweight MobileNetV2 for real-world animal picture classification. The focus is on computationally efficient categorization for mobile/embedded applications with good recognition performance.	Dhenia et al. (2025)

Table 6 (Continued). Different models used in animal affective-state analysis

CNN architecture	Animal/ datasets used	Accuracy	Key highlights	Reference
CNN–LSTM for video; GRU for audio; autoencoder for biometric data; hybrid fusion	Wildlife monitoring; multimodal video, acoustic and biometric data; edge/NB-IoT system	94.7% hybrid fusion; CNN–LSTM video 92.4%; GRU audio 88.1%; autoencoder 85.3%	Suggested an emotion-sensitive animal protection system that integrates multimodal sensing, edge computing, NB-IoT, adaptive geofencing, and interpretable AI. The device is designed to identify anxiety, anguish, and suffering while facilitating proactive wildlife conservation.	Mariadas et al. (2026)
ResNet-50 CNN, transfer learning using spectrograms	Pigs; 7,414 vocalizations from 411 pigs, covering 19 production contexts	91.5% valence; 81.5% context	Engineered an automated system for identifying the emotional valence of pigs from their vocalizations. The CNN examined spectrograms of both low-frequency and high-frequency vocalizations. It attained significantly superior classification efficacy compared to pDFA and exhibited promise for real-time welfare assessment.	Briefer et al. (2022)
Adjusted Siamese Network; CNN classification/regression baseline	Dogs; EmotionalCanines vocalization dataset; Husky and Shiba Inu	CNN classification: 60.45% / 43.35% arousal and 56.34% / 31.91% valence for Husky/Shiba; adjusted Siamese: 55.93% arousal, 48.07% valence	Reconceptualized canine bark identification as a continuous arousal-valence regression model instead of traditional categorical classification. The modified Siamese model diminished severe valence inaccuracies (TAP) in comparison to the regression baseline.	Song et al. (2025)
ConvNeXt, EfficientNet-V2, MobileNet-V3, RegNet; comparison with ViT	Pregnant sows; facial-image dataset from SRUC, categorized as low-stress/high-stress using behavioural assessment and cortisol	75% average accuracy for best overall model (ViT); F1 = 0.78 for high-stress detection	Assessed various deep-learning frameworks for the classification of stress based on swine facial photos. ViT attained an overall average accuracy of 75%, but CNN models exhibited significant fluctuations among batches, underscoring the difficulties of generalization.	Yunas et al. (2025)
YOLOv4 neural-network detector	Red foxes; video images of foxes in experimental housing	99.91% mAP	Engineered instantaneous computer vision for the observation of red-fox posture and behavior. The model categorized lying, sitting, and standing positions, attaining a precision of 98.61%, a recall of 95.12%, and a mean Average Precision (mAP) of 99.91% at roughly 73 ms per frame.	Schütz et al. (2022)

in the literature. Table 7 describes animal emotion recognition at different stages.

6. Model families used in animal affect studies

The models used in the included studies are not all CNN classifiers. Table 8 distinguishes them based on differences in task and architecture family. The historical development of CNNs and the basic definitions of convolution and pooling were not discussed because they are well-established textbook concepts and do not bear on the evidence reviewed. Only model families used in the included studies or needed

Table 7. Stages of animal emotion recognition

Stage	Required step	Quality control or output	Common shortcomings
1. Study design and ethics	Define the construct, species, setting; obtain ethical approval	Written protocol; welfare impact assessment	Collecting data that stresses animals
2. Data acquisition	Include several animals per class; varied backgrounds, lighting and camera angles; record animal identity	Metadata: animal ID, session, context	One animal or one room per class
3. Label generation	Use a coding system, expert panel or physiology; define each category beforehand	Uses at least two or more independent raters; kappa or ICC; adjudication	Single-rater, owner or crowd labels
4. Annotation quality control	Reassess a subset of samples; discard ambiguous items; document disagreement	Agreement statistic reported per class	Hiding ambiguous cases
5. Data partitioning	Split by animal (and by breed or site when possible) before any processing	Fold table with animal counts	Random frame split
6. Imbalance handling	Class weights, balanced sampling or focal loss; augmentation inside training folds only	Report class distribution for each fold	Augmenting before the split
7. Training	Transfer learning, fixed seeds, tuning on a validation set only	Reproducible configuration	Tuning on the test set
8. Evaluation	Macro-F1, balanced accuracy, per-class results, confidence intervals over animals	As per defined metrics	Reporting accuracy only
9. Generalization testing	Cross-breed, cross-species and cross-dataset validation	Results per held-out group	Claiming generalization untested
10. Explainability and expert review	Saliency maps checked against expert-defined regions	Expert sign-off	Treating heat maps as proof
11. Deployment	Human-in-the-loop use, drift monitoring, error-cost review	Maintain audit log; establish escalation procedures	Automated welfare decisions

Table 8. Comparison of different architecture models, applications, strengths, and limitations in animal emotional-state analysis

Family	Examples	Role in the pipeline	Strengths	Limitations
CNN classification backbones	VGG, ResNet, DenseNet, Inception	Classifies a facial, body, or frame image crops	Strong pretrained weights, easy to fine-tune	Heavy; ImageNet features are not animal-specific
Efficient CNNs	MobileNet, EfficientNet	Classification under compute limits	Small and fast; suited to edge devices	Lower capacity on subtle expressions
Attention-augmented CNNs	Squeeze-and-Excitation (SE) networks, CBAM	Re-weight channels or regions inside a CNN	Cheap gain on fine facial cues	Gains rarely tested across animals
Detection frameworks	YOLO (You Only Look Once), Faster R-CNN	Localise the animal, face or body part	Rapid detection; provide image crops for a classifier	Detect objects; they do not classify emotion unless a classifier follows
CNN with temporal modelling	CNN-LSTM, 3D-CNN, ConvLSTM, two-stream	Model movement over frames	Capture dynamics and onset	LSTM and GRU are recurrent, not convolutional; more data needed
Transformer and self-supervised models	ViT, DINO-ViT	Classify or extract features from whole images	Global context; strong self-supervised features	Large training datasets required; limited animal-affect evidence
Hybrid CNN–transformer	CNN stem with ViT blocks	Combine local and global features	Balances locality and context	Fewer supporting studies; more difficult to interpret

for comparison are retained. Transfer learning using ImageNet models was used in different studies. Whether it improves generalization across animals, breeds or species was tested; this remains unclear and should not be assumed, because anatomy and expression differ between species. When the same architecture is evaluated under multiple validation designs, the validation design explains more variation in reported performance than the choice of architecture, indicating that data quality is more important than architecture choices.

7. Explainability and its validation

Grad-CAM, Grad-CAM++, and SHAP are the techniques helping recognize the input features or regions influencing a model's predictions (Selvaraju et al. 2017; Chattopadhay et al. 2018; Lundberg and Lee 2017). They do not show that those regions carry emotional information. An attribution map highlighting the face shows where the model looked; it does not demonstrate that the region reflects a real emotional mechanism. Three separate questions must be kept apart and Table 9 provides the approaches to assess model attention and validity of visual explanations.

8. Ethical considerations

The ethical issues specific to CNN-based animal emotion recognition are as follows.

Data collection: Animals should not be subjected to stress unnecessarily during recording, nor should they be restrained, isolated, or subjected to intrusive measurements. A welfare justification and ethical approval are required for studies that experimentally induce emotional states. Ownership, consent, and privacy concerns are raised by farm, shelter, and owner datasets.

Anthropomorphism and labelling: When human emotion terms are used to describe animal imagery, human interpretations may influence the assigned labels. This risk may be decreased by supporting physiological evidence and validated coding schemes.

Errors and their repercussions: If a false positive occurs frequently, it can lead to alert fatigue as well as an unwarranted alarm or intervention. An animal in pain or anguish may go undetected in case of a false negative result. The decision threshold needs to be carefully selected and reported because the effects of each type of error varies, depending on the species and environment.

Automated oversight and decision-making: Model outputs are not measurements of subjective experience, but rather indicators for a trained human. Human oversight should be involved in animal welfare choices, and there should be an audit record.

Validity and transparency: Systems should not be used outside of the species, breeds, circumstances, and validation level for which they were evaluated. Table 10 describes the associated risks and the mitigation strategies in CNN-based animal emotion recognition analysis.

9. Performance metrics

Because emotion datasets are typically imbalanced, overall accuracy of classification by the model can be readily misleading. For uncommon, welfare-critical emotional states, a model that consistently predicts the majority class may performs poorly in identification of minority classes. Table 11 describes the appropriate applications and shortfalls of different performance metrics. True positives (TP), false positives (FP), true negatives (TN), and false negatives (FN) are the standard counts used in definitions. Although recall and precision can trade off as the decision threshold shifts, they are not intrinsically adversely correlated;

Table 9. Model interpretability and the validity of explanations

Question	How it is tested	Interpretation of a positive result	What it does not show
Where does the model attend?	Grad-CAM++, SHAP, attention maps	The input regions driving the output	Whether the regions are biologically meaningful
Is the explanation faithful to the model?	Deletion and insertion tests; randomization sanity checks [Adebayo et al. 2018]	The attribution map reflects the model's computation process	That the model is right for the right reasons
Is it biologically valid?	Overlap with expert-defined regions or coded action units; controlled manipulation of those regions	The model uses cues known to be linked to the state	Subjective experience

Table 10. Potential risk and mitigation strategies in CNN-based animal emotion recognition analysis

Risk	Examples	Mitigation
Stress during data collection	Restraint to obtain clear faces	Non-invasive recording; ethics approval
Anthropomorphic labels	"Sad" assigned from a drooping ear	Coding systems; expert agreement
False-positive detection	Normal behaviour flagged as pain	Tuned threshold; human confirmation
False-negative detection	Pain not detected in a stoic species	Species-specific validation; sensitivity targets
Overinterpretation of the model outputs	Prediction read as measured emotion	Explicit wording; limits in reports
Data provenance	Images collected without consent from online sources	Documented source and license

both can be improved by a better model. At least, studies should report macro-F1, balanced accuracy, per-class recall and precision, a confusion matrix, and confidence intervals estimated at the level of animal.

10. Research gaps and future directions

This section distinguishes between hypothetical future research possibilities from the research gaps identified from the reviewed evidence.

10.1 Identified gaps

Table 12 identifies the major research gaps.

10.2 Potential opportunities

Only improvements assessed on animal-level data splitting would provide convincing evidence of generalization, but self-supervised learning and domain adaptation could decrease the requirement for labels and bridge camera and location differences. Image diversity can be increased by using synthetic data, such as that obtained using Generative Adversarial Networks. It risks label bias and artificial facial or behavioral traits, and it does not produce biologically meaningful emotional expressions. Experts should verify synthetic data, and models should only be tested on actual animals. Although there may be significant anatomical and behavioral differences, cross-species transmission learning is appealing research direction. Cross-species claims should be avoided without proof, and transfer should be verified for each species. Although not specifically affecting recognition, edge deployment and model compression deal with feasibility rather than validity. Although multi-modal foundation models may incorporate text, audio, and visuals, their outputs may be poorly calibrated, lack proven animal-affect ground truth to evaluate them, and inherit human-centered priors from their training data. Currently, not enough evidence is there to establish their reliability in accurately identifying the animal affective states.

11. Conclusions

This review examined 44 studies that identified animal emotional states using CNN-based algorithms. The evidence supports a limited conclusion that CNN-based models can discriminate between defined behavioral or facial categories in different species, mostly under controlled conditions. However, reported macro-F1 / accuracy values cannot be directly compared across studies due to differences in datasets, labels, and validation protocols. Many significant questions remain unsolved. Among the 102 studies searched, only 44 reported validated labels with demonstrated inter-rater agreement, used animal-level data splitting, and evaluated model performance on external datasets. Whether multi-modal input, transfer learning, or more

Table 11. Performance metrics for animal emotional-state recognition models

Metric	Used when:	Shortcomings
Accuracy	Classes are balanced and errors consequences are comparable	Hides failure on minority classes
Precision	False alarms are costly (for example unnecessary welfare intervention)	Ignores missed cases
Recall or sensitivity	Failing to identify a distressed animal has serious consequences	Ignores false alarms
Specificity	Correctly identifying non-distressed animals is important	Not informative alone
F1 and macro-F1	Imbalanced classes; macro-F1 weights each class equally	Does not use true negatives; hides which class fails
Balanced accuracy	Imbalanced classes, a single summary	Hides the precision cost
Matthews correlation coefficient and Cohen's kappa	One chance-corrected score for imbalanced data is provided (Chicco and Jurman 2020)	Harder to interpret across datasets
ROC-AUC	Ranking quality, threshold-free	Can look good on imbalanced data; PR-AUC is more informative than ROC-AUC in highly imbalanced datasets (Saito and Rehmsmeier, 2015)
Confusion matrix and per-class scores	Always, to show which states are confused	Difficult to compare across studies
Calibration (reliability curve, expected calibration error)	Predicted probabilities drive decisions (Guo et al. 2017)	Rarely reported
Confidence intervals	Always, over animals and not over frames	Frame-level intervals may understate uncertainty when observation from same animal are correlated

Table 12. Identified research gaps and proposed future scope in animal affective-state recognition

Gap	Proposed work
Few validated labels	Triangulated labels with reported agreement
Little subject-independent testing	Mandatory animal-level and external validation
Few public datasets with animal identity	Open benchmarks with animal IDs and licenses
Scarce multimodal comparisons	Unimodal baselines in every study
Little field validation	Prospective field studies
Unvalidated explanations	Biological validation of attribution maps

modern architectures regularly produce gains is still up for debate. Most results cannot be construed as evidence that a model detects an internal emotional state because categorization agreement with human labels does not prove validity. For reliable animal emotion recognition, future studies need independently validated labels with reported agreement, animal-level and external validation, imbalance-robust metrics with confidence intervals, a unimodal baseline for every multimodal comparison, expert validation of explanations, and human oversight in welfare use. Until these methodological conditions are met, CNN-based systems should be best regarded as decision-support tools for trained observers and not definitive measures of animal emotional states.

Declarations

Funding: This research received no specific grant from any funding agency in the public, commercial, or not-for-profit sectors

Conflict of interest: The authors do not have any conflicts of interest

Acknowledgements: The authors wish to acknowledge the financial support provided by the Ministry of Education (India) through National Institute of Technology Jamshedpur for conducting the research work

Ethics approval: This article does not contain any original studies with human participants or animal subjects performed by any of the authors. Ethical approval was not required as this is a review paper based on previously published literature

Data availability statement: Data sharing is not applicable to this article as no new datasets were generated or analyzed during the current study.

References

Adebayo J, Gilmer J, Muelly M, Goodfellow I, Hardt M, Kim B. (2018). Sanity checks for saliency maps. Advances in Neural Information Processing Systems 31: 9505–9515. https://doi.org/10.48550/arXiv.1810.03292

Ahmad M, Zhang W, Smith M, Brilot B, Bell M. (2023). Real-time livestock activity monitoring via fine-tuned Faster R-CNN for multiclass cattle behaviour detection. 2023 IEEE 14th Annual Ubiquitous Computing, Electronics & Mobile Communication Conference (UEMCON), New York, USA. Pp. 805–811. https://doi.org/10.1109/UEMCON59035.2023.10316066

Bhave A, Hafner A, Bhave A, Gloor PA. (2024). Unsupervised canine emotion recognition using momentum contrast. Sensors 24(22): 7324. https://doi.org/10.3390/s24227324

Bhave A, Kieson E, Hafner A, Gloor PA. (2025). Identifying novel emotions and wellbeing of horses from videos through unsupervised learning. Sensors 25(3): 859. https://doi.org/10.3390/s25030859

Boneh-Shitrit T, Feighelstein M, Bremhorst A, Amir S, Distelfeld T, Dassa Y, Yaroshetsky S, Riemer S, Shimshoni I, Mills DS, Zamansky A. (2022). Explainable automated recognition of emotional states from canine facial expressions: the case of positive anticipation and frustration. Scientific Reports 12(1): 22611. https://doi.org/10.1038/s41598-022-27079-w

Bremhorst A, Sutter NA, Würbel H, Mills DS, Riemer S. (2019). Differences in facial expressions during positive anticipation and frustration in dogs awaiting a reward. Scientific Reports 9: 19312. https://doi.org/10.1038/s41598-019-55714-6

Briefer EF, Sypherd CCR, Linhart P, Leliveld LM, Padilla de la Torre M, Read ER, Guérin C, Deiss V, Monestier C, Rasmussen JH, Špinka M, Düpjan S, Boissy A, Janczak AM, Hillmann E, Tallet C. (2022). Classification of pig calls produced from birth to slaughter according to their emotional valence and context of production. Scientific Reports 12(1): 3409. https://doi.org/10.1038/s41598-022-07174-8

Chattopadhyay A, Sarkar A, Howlader P, Balasubramanian VN. (2018). Grad-CAM++: generalized gradient-based visual explanations for deep convolutional networks. 2018 IEEE Winter Conference on Applications of Computer Vision (WACV), Lake Tahoe, USA. Pp. 839–847. https://doi.org/10.1109/WACV.2018.00097

Chen HY, Lin CH, Lai JW, Chan YK. (2023). Convolutional neural network-based automated system for dog tracking and emotion recognition in video surveillance. Applied Sciences 13(7): 4596. https://doi.org/10.3390/app13074596

Cheng WK, Leong WC, Tan JS, Hong ZW, Chen YL. (2022). Affective recommender system for pet social network. Sensors 22(18): 6759. https://doi.org/10.3390/s22186759

Chicco D, Jurman G. (2020). The advantages of the Matthews correlation coefficient (MCC) over F1 score and accuracy in binary classification evaluation. BMC Genomics 21: 6. https://doi.org/10.1186/s12864-019-6413-7

Cohen J. (1960). A coefficient of agreement for nominal scales. Educational and Psychological Measurement 20(1): 37–46. https://doi.org/10.1177/001316446002000104

Corujo LA, Kieson E, Schloesser T, Gloor PA. (2021). Emotion recognition in horses with convolutional neural networks. Future Internet 13(10): 250. https://doi.org/10.3390/fi13100250

Coutant M, Villain AS, Briefer EF. (2024). A scoping review of the use of bioacoustics to assess various components of farm animal welfare. Applied Animal Behaviour Science 275: 106286. https://doi.org/10.1016/j.applanim.2024.106286

Descovich KA, Wathan J, Leach MC, Buchanan-Smith HM, Flecknell P, Farningham D, Vick SJ. (2017). Facial expression: an under-utilised tool for the assessment of welfare in mammals. ALTEX 34(3): 409–429. https://doi.org/10.14573/altex.1607161

Dhenia RNK, Sridhar R, Sharma J, Kanani IJ, Banerjee D, Sharma G. (2025). Using MobileNetV2 for real-world applications: effective animal image classification. Global Conference on Information Technology and Communication Networks (GITCON), Belagavi, India. Pp. 1–5. https://doi.org/10.1109/GITCON65266.2025.11376889

Du X, Shan W, Wang Y, Ye T, Zhuo J. (2023). Dog emotion classification with transfer learning. Research project report available on Julia(Xi) Du Personal Website. https://superjuliadu.com/report/Dog-Image-Classification-Final-Report.pdf

Feighelstein M, Ricci-Bonot C, Hasan H, Weinberg H, Rettig T, Segal M, Distelfeld T, Shimshoni I, Mills DS, Zamansky A. (2024). Automated recognition of emotional states of horses from facial expressions. PLOS One 19(7): e0302893. https://doi.org/10.1371/journal.pone.0302893

Feighelstein M, Shimshoni I, Finka LR, Luna SP, Mills DS, Zamansky A. (2022). Automated recognition of pain in cats. Scientific Reports 12(1): 9575. https://doi.org/10.1038/s41598-022-13348-1

Gähwiler S, Bremhorst A, Tóth K, Riemer S. (2020). Fear expressions of dogs during New Year fireworks: A video analysis. Scientific Reports 10: 16035. https://doi.org/10.1038/s41598-020-72841-7

Gilhotra K, Yadav V, Diwakar M, Lohumi Y, Singh P, Roka S. (2025). Few shot pet emotion recognition using prototypical networks with ResNet34 backbone. In: Goar V, Gajrani J, Shin J, Stowik A, editors, Deep Learning and Visual Artificial Intelligence. ICDLAI 2025. Lecture Notes in Networks and Systems 2097. Springer Nature Switzerland. Pp. 368–378. https://doi.org/10.1007/978-3-032-32476-4_27

Guo C, Pleiss G, Sun Y, Weinberger KQ. (2017). On calibration of modern neural networks. Proceedings of the 34th International Conference on Machine Learning 70: 1321–1330. https://doi.org/10.48550/arXiv.1706.04599

Hernández-Vanegas LA, Rodríguez-Romo S. (2026). Exploring canine emotions: A transfer learning and 3DCNN-based study for small databases. IEEE Transactions on Affective Computing 17(3): 3824-3831. https://doi.org/10.1109/TAFFC.2026.3652076

Jamalpur B, Kothandaraman D, Sathwika V, Pranathi V. (2024). Pet facial emotion recognition using convolutional neural networks. In: Reddy VS, Wang J, Chetti P, Reddy KTV, editors, Soft Computing and Signal Processing, ICSCSP 2024. Lecture notes in Network Systems 1221. Springer Nature Singapore. Pp. 207–216. https://doi.org/10.1007/978-981-96-0924-6_17

Kapoor S, Narayanan A. (2023). Leakage and the reproducibility crisis in machine-learning-based science. Patterns 4(9): 100804. https://doi.org/10.1016/j.patter.2023.100804

Kaufman S, Rosset S, Perlich C, Stitelman O. (2012). Leakage in data mining: formulation, detection, and avoidance. ACM Transactions on Knowledge Discovery from Data 6(4): 15. https://doi.org/10.1145/2382577.2382579

Kim JH, Colaco SJ, Poulose A, Neethirajan S, Han DS. (2023). Advancing pig welfare assessment: Introducing the SSPD-PER method for objective and reliable pig emotion recognition. Available on Preprints. https://doi.org/10.20944/preprints202307.1510.v1

Koo TK, Li MY. (2016). A guideline of selecting and reporting intraclass correlation coefficients for reliability research. Journal of Chiropractic Medicine 15(2): 155–163. https://doi.org/10.1016/j.jcm.2016.02.012

Kopczyński M, Czubenko M. (2026). Enhancing human-dog interaction through deep learning and explainable AI. Scientific Reports 16: 30258. https://doi.org/10.1038/s41598-026-51009-9

Landis JR, Koch GG. (1977). The measurement of observer agreement for categorical data. Biometrics 33(1): 159–174. https://pubmed.ncbi.nlm.nih.gov/843571/

Lencioni GC, de Sousa RV, de Souza Sardinha EJ, Corrêa RR, Zanella AJ. (2021). Pain assessment in horses using automatic facial expression recognition through deep learning-based modeling. PLOS One 16(10): e0258672. https://doi.org/10.1371/journal.pone.0258672

Li DL, Lee SK, Tsai YC. (2026). Deep learning-based dog expression recognition. IEEE Access 14: 5675–5683. https://doi.org/10.1109/ACCESS.2025.3650550

Li S, Zhang H, Lv Y, Wang Z, Yang H, Yu Q. (2025). Intelligent canine emotion recognition system utilizing EfficientNet-B0 architecture. 2025 IEEE 3rd International Conference on Sensors, Electronics and Computer Engineering (ICSECE), Jinzhou, China. Pp. 325–330. https://doi.org/10.1109/ICSECE65727.2025.11257013

Lundberg SM, Lee SI. (2017). A unified approach to interpreting model predictions. In: Proceedings of the 31th International Conference on Neural Information Processing Systems. Pp. 4765–4774. Curran Associates Inc., Long Beach, California, USA. https://doi.org/10.48550/arXiv.1705.07874

Martvel G, Riemer S. (2025). Automated analysis of emotional expressions in dogs based on geometric morphometrics. Scientific Reports 15: 32331. https://doi.org/10.1038/s41598-025-15741-y

Martvel G, Zamansky A, Pedretti G, Canori C, Shimshoni I, Bremhorst A. (2025). Dog facial landmarks detection and its applications for facial analysis. Scientific Reports 15: 21886. https://doi.org/10.1038/s41598-025-07040-3

Mendl M, Burman OHP, Paul ES. (2010). An integrative and functional framework for the study of animal emotion and mood. Proceedings of the Royal Society B: Biological Sciences 277(1696): 2895–2904. https://doi.org/10.1098/rspb.2010.0303

Neethirajan S. (2021). Happy cow or thinking pig? WUR Wolf—facial coding platform for measuring emotions in farm animals. AI 2(3): 342–354. https://doi.org/10.3390/ai2030021

Nie L, Li B, Du Y, Jiao F, Song X, Liu Z. (2024). Deep learning strategies with CReToNeXt-YOLOv5 for advanced pig face emotion detection. Scientific Reports 14(1): 1679. https://doi.org/10.1038/s41598-024-51755-8

Nithya R, Mariadas AEP, Anbarasu MM, Ravindran S. (2026). Intelligent NB-IoT and emotion-aware monitoring system for proactive wildlife security. Tehnički vjesnik 33(3): 1139–1147. https://doi.org/10.17559/TV-20250827002927

Ntalampiras S, Pesando Gamacchio G. (2025). Explainable classification of goat vocalizations using convolutional neural networks. PLOS One 20(4): e0318543. https://doi.org/10.1371/journal.pone.0318543

Page MJ, McKenzie JE, Bossuyt PM, Boutron I, Hoffmann TC, Mulrow CD, Shamseer L, Tetzlaff JM, Akl EA, Brennan SE, Chou R, Glanville J, Grimshaw JM, Hróbjartsson A, Lalu MM, Li T, Loder EW, Mayo-Wilson E, McDonald S, McGuinness LA, Stewart LA, Thomas J, Tricco AC, Welch VA, Whiting P, Moher D. (2021). The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. BMJ 372: n71. https://doi.org/10.1136/bmj.n71

Paul ES, Harding EJ, Mendl M. (2005). Measuring emotional processes in animals: the utility of a cognitive approach. Neuroscience & Biobehavioral Reviews 29(3): 469–491. https://doi.org/10.1016/j.neubiorev.2005.01.002

Pearline SSRA, Subbiah G, Sanu A, Charumathi Venkatesan K. (2026). Multi-QuadEmoNet: cat and dog emotion classification model from animal vocalization using multi-stage LSTM-GRU paradigm. Frontiers in Veterinary Science 13: 1799281. https://doi.org/10.3389/fvets.2026.1799281

Priya HD, Vimal S, Vignesh M, Devaguru S. (2026). Interpreting subtle facial cues in elephants: a deep learning approach to emotional state recognition. 2026 Second International Conference on Intelligent Systems for Communication, IoT and Security (ICISCoIS), Coimbatore, India. Pp. 1–6. https://doi.org/10.1109/ICISCoIS62701.2026.11447516

Saito T, Rehmsmeier M. (2015). The precision-recall plot is more informative than the ROC plot when evaluating binary classifiers on imbalanced datasets. PLOS One 10(3): e0118432. https://doi.org/10.1371/journal.pone.0118432

Schütz AK, Krause ET, Fischer M, Müller T, Freuling CM, Conraths FJ, et al. (2022). Computer vision for detection of body posture and behavior of red foxes. Animals 12(3): 233. https://doi.org/10.3390/ani12030233

Selvaraju RR, Cogswell M, Das A, Vedantam R, Parikh D, Batra D. (2017). Grad-CAM: visual explanations from deep networks via gradient-based localization. IEEE International Conference on Computer Vision (ICCV), Venice, Italy. Pp. 618–626. https://doi.org/10.1109/ICCV.2017.74

Shao D, He Z, Fan H, Sun K. (2023). Detection of cattle key parts based on the improved YOLOv5 algorithm. Agriculture 13(6): 1110. https://doi.org/10.3390/agriculture13061110

Sinnott RO, Aickelin U, Jia Y, Sinnott ER, Sun PY, Susanto R. (2021). Run or pat: Using deep learning to classify the species type and emotion of pets. 2021 IEEE Asia-Pacific Conference on Computer Science and Data Engineering (CSDE), Brisbane, Australia. Pp. 1–6. https://doi.org/10.1109/CSDE53843.2021.9718465

Song H, Brady S, Ward T, Smeaton AF. (2025). A dimensional approach to canine bark analysis for assistance dog seizure signaling. arXiv preprint. https://doi.org/10.48550/arXiv.2509.18375

Subair R, Benedict S. (2024). Enhancing elephant emotion detection using YOLOv5, YOLOv8, and YOLOv9: a study on performance and explainability. 2024 IEEE Silchar Subsection Conference (SILCON 2024),

Agartala, India. Pp. 1–6. https://doi.org/10.1109/SILCON63976.2024.10910543

Sui Z, Hong H, Shimtoku D, Kobayashi HH. (2024). catAction: deep learning for enhancing emotional cat–human interactions through the posture-based determination of the degrees of kittens' defensive and offensive aggressions. International Conference on Animal-Computer Interaction, Glasgow, UK. Pp. 1–9. https://doi.org/10.1145/3702336.3702341

Sumon RI, Ali H, Akter S, Uddin SMI, Mozumder MAI, Kim HC. (2025). A deep learning-based approach for precise emotion recognition in domestic animals using EfficientNetB5 architecture. Engineering 6(1): 9. https://doi.org/10.3390/eng6010009

Suneetha C, Anitha R. (2025). Speech-based emotion recognition by using a faster region-based convolutional neural network. Multimedia Tools and Applications 84(8): 5205–5237. https://doi.org/10.1007/s11042-024-19004-2

Tanwar V. (2024). Understanding dog emotions through deep learning: A CNN-based classification framework. 2024 5th International Conference on Smart Electronics and Communication (ICOSEC), Trichy, India. Pp. 964–969. https://doi.org/10.1109/ICOSEC61587.2024.10722523

Tribethran S, Hasan NJP, Rahman A. (2024). Classification of dog emotion using transfer learning on convolutional neural network algorithm. Paradigma-Jurnal Komputer dan Informatika 26(2): 102–108. https://doi.org/10.31294/p.v26i2.5295

Vikramathithan AC, Manjunath TG, Lakshya HE, Rakshitha R. (2025). Emotion detection for cross-species using VGG16 model with Ada-Sig-Max optimizer. 2025 International Conference on Data, Energy and Communication Networks (DECoN), Bhopal, India. Pp. 1–6. https://doi.org/10.1109/DECoN67170.2025.11447459

Wang H, Shi Z, Hu R, Wang X, Chen J, Che H. (2025). Real-time fear emotion recognition in mice based on multimodal data fusion. Scientific Reports 15(1): 11797. https://doi.org/10.1038/s41598-025-95483-z

Wu Y. (2023). Emotion detection of dogs and cats using classification models and object detection model (Research project). Yeshiva University. ResearchGate. Research Project Link

Xie Y. (2023). Dogs emotion recognition and parameter analysis based on EfficientNet with transfer learning. 1st International Conference on Data Analysis and Machine Learning (DAML), Kuala Lumpur, Malaysia. Pp. 240–245. https://doi.org/10.5220/0012799600003885

Yunas SU, Shahbaz A, Baxter EM, Hansen MF, Smith ML, Smith LN. (2025). Enhanced deep learning for robust stress classification in sows from facial images. Agriculture 15(15): 1675. https://doi.org/10.3390/agriculture15151675

Citation

Mandal B, Kumar D. (2026). Animal affective-state analysis using Convolutional Neural Networks: Methods, datasets, applications, challenges, and future directions. Letters in Animal Biology 06(1): 193 – 206. *https://doi.org/10.62310/liab.v6i1.454*