

Multimodal Artificial Multimodal Artificial Intelligence in Companion Animal Healthcare

A Review of Vision, Vocalization, Behavioural, Clinical and Predictive Technologies *PawSense project*

PRESENTED BY

1. Bharti Pardhi • 2. Suhangini Umredkar • 3. Ankit Saha • 4. Yugant Pandey • 5. Piyush Bagde

ABSTRACT

Artificial intelligence (AI) is increasingly used in companion-animal healthcare to analyse images, vocalisations, behaviour, clinical text and physiological measurements. This review synthesises 25 studies relevant to dogs and cats, including breed recognition, visual disease screening, body-condition assessment, posture and activity monitoring, animal-vocalisation analysis, veterinary large language models (LLMs), estrous-stage identification, non-invasive temperature prediction, and food-safety knowledge. The literature shows that AI can achieve strong task-specific performance under controlled conditions, but its clinical readiness remains limited by small or unbalanced datasets, canine-dominated evidence, inconsistent ground truth, weak external validation, domain shift and uncertainty in real-world inputs. Several studies also report meaningful setbacks, including image-quality sensitivity, performance loss on real consultation data, subjectivity in emotion labels, hardware dependence and limited generalisability across breeds or clinics. The central research opportunity is therefore not another isolated classifier, but a multimodal, longitudinal and veterinarian-supervised platform that combines complementary evidence while communicating uncertainty and escalating high-risk cases to professional care.

Keywords: Veterinary AI; companion animals; deep learning; computer vision; multimodal AI; LLM; animal emotion; pet health monitoring.

REVIEW AT A GLANCE

25 SOURCES	5 MODALITIES	CORE SETBACK	MAIN GAP
Primary studies + reviews	Vision • Audio • Sensor • Text • Time	External validation remains limited	Integrated multimodal care

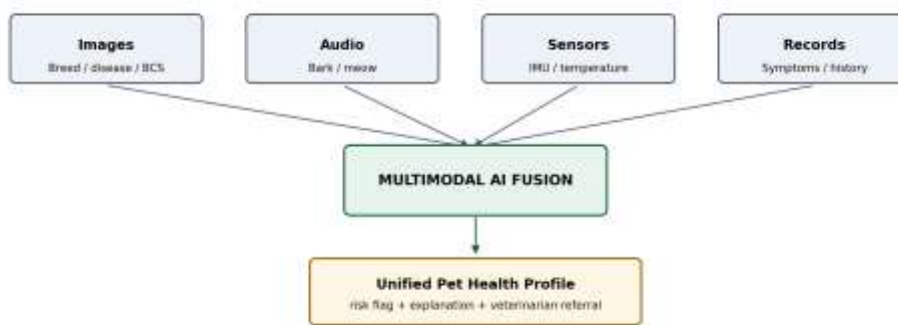


Figure 1. Multimodal companion-animal health-support workflow.

2

Introduction

Veterinary care is inherently multimodal: clinicians combine signalment, history, physical examination, imaging, laboratory findings and behavioural observations. Recent AI research mirrors this diversity, but most studies optimise only one task at a time. A 2026 scoping review identified 98 peer-reviewed diagnostic-AI studies in dogs and cats, with 81.6% focused only on dogs and only 14.3% reporting independent external validation [1]. This creates a central tension: technical performance is improving rapidly, while real-world clinical generalisability remains less mature.

For PawSense-type systems, the relevant question is therefore not whether AI can recognise a breed, classify a bark or analyse text. The more important question is whether several imperfect modules can be combined safely into a coherent health-support workflow that preserves context, uncertainty and veterinary oversight.

Why this review matters: The literature already supports many isolated AI capabilities. The research contribution now shifts toward integration, longitudinal personalisation, validation and safe escalation.

3

Literature Survey

The 25 reviewed sources can be grouped into five evidence streams. Together they cover identification, visual health assessment, behaviour, physiological monitoring, clinical decision support and reproductive care.

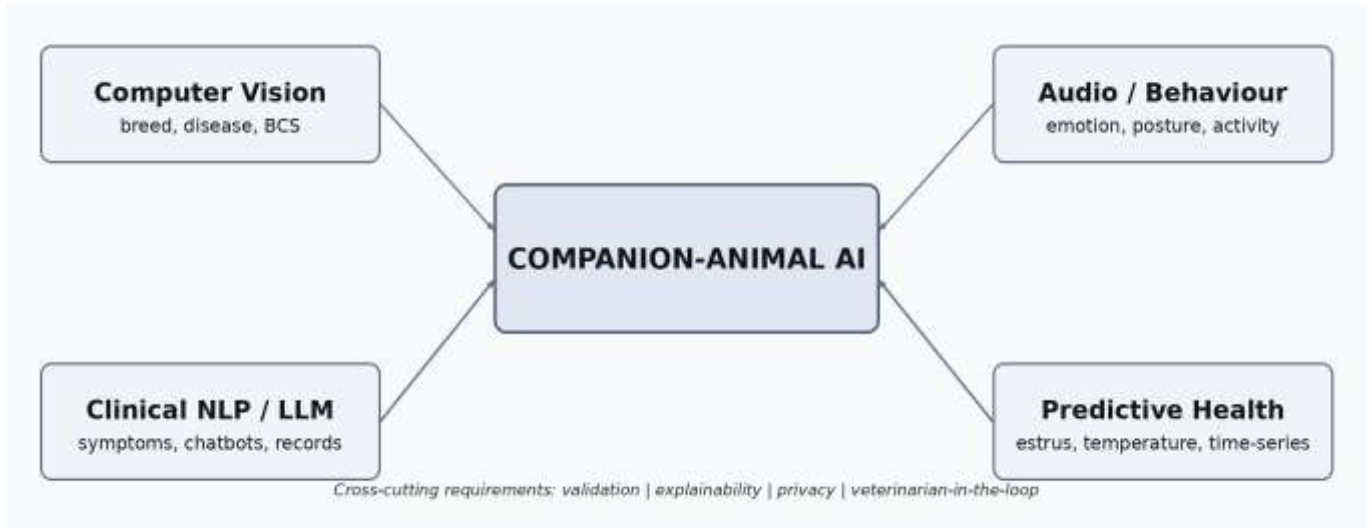


Figure 2. Taxonomy of the literature reviewed for a PawSense-type platform.

Proprietary visual intelligence	Current literature validates the efficacy of non-invasive visual profiling, encompassing automated phenotypic and breed classification, morphological disease screening, and algorithmic body condition scoring (BCS) (3)-(8), (18),”
Ethological and vocalization analysis	DeepLabCut posture analysis, IMU activity recognition, animal-vocal emotion recognition and behavioural clinical context [9]-[11], [19], [21].
Semantic clinical data extraction	Veterinary LLM decision support, chatbots and conversational interfaces [12], [13], [20].
Metabolic and fertility baselining	Estrous-stage identification, canine/feline reproductive evidence and non-invasive temperature prediction [15], [16], [22]-[24].
Deterministic clinical boundaries	Food toxicity, clinical pathology governance, time-series evidence and complex endocrine disease context [2], [14], [17], [25].

Algorithmic inference pathways

01 Pathological corroboration What has been demonstrated?	02 Validation Bottleneck Where does the evidence fail or weaken?	03 Market whitespace What remains unresolved for integrated care?
---	--	---

3

Literature Survey - Key Findings by Modality

A. Computer vision based phenotypic classification Dog-breed work reports up to 92.4% validation accuracy using a hybrid Inception-v3/Xception model on 120 breeds [4], while an earlier Android study reported 94% testing accuracy [18]. Cat recognition remains less mature: ResNet50 achieved 84.89% validation accuracy on only five breeds [5]. Animal detection work using deformable YOLOv2 reached 85.5% mAP [3].	B. Clinical anomaly and Biometric evaluation While PAWMATE reported on 89.2% baseline accuracy for targeted phenotypic condition(6), the architecture exhibited a significant generalization shortfall, proving highly vulnerable to ambient lighting and varying optical fidelity. Furthermore, despite clinical evidence indicating a 55% prevalence of overweight morbidity in specific cohorts(7), current automated body condition scoring (BCS) solutions remain diagnostically siloed. For instance, a 2026 preprint successfully validated feline BCS prediction exclusively via post-mortem radiography, highlighting a critical translational void in deploying non-invasive, smartphone-derived visual diagnostic for live patients(8).”
C. Ethological ,postural and acoustic assessment Current approaches to non-invasive ethological assessment demonstrate that unimodal architecture are clinically informative but diagnostically incomplete. For example, DeepLabcut-based automated pose estimation achieved only 67.5% accuracy cross four canine affective states (9), exposing a significant diagnostic blind spot when postural analysis is isolated from acoustic context. While wearable IMU-based biomechanical tracking provides a strong quantitative baseline (F1 = 0.90) [10], it remains physically invasive. Conversely, acoustic vocalization analysis shows a strong technical precedent : Multi-QuadEmoNet achieved high-fidelity ethological screening (95% canine, 90% feline accuracy) [11], successfully overcoming the weak external validation observed in earlier DNN architectures [19]. Together, these findings highlight a critical translational void that can only be resolved through integrated audiovisual behavioral profiling.”	D. biomedical and therioegenological AI Recent model highlight a significant translational void between theoretical performance and clinical reality. For example, while AVA achieved a 91.4% top 3 diagnostic accuracy on synthetic cohorts, the system suffered a severe generalization shortfall, plummeting to 38.6% accuracy during real-world clinical chart audits(120). Similarly, while faster R-CNN architectures demonstrated 91.6% accuracy for therioegenological (estrous-stage) identification, the finding are heavily constrained by an underpowered, low-N patient population of only 12 cases (15). Conversely, non-invasive physiological profiling demonstrates a more robust clinical premise, with surface-temperature models achieving a prospective RMSE of 0.15 degree Calcus in cats and 0.14 degree calcus in dogs, establishing a strong precedent for automated biometric monitoring (16).”

Clinical data synthesis

The literature does not support treating these percentages as a single leaderboard. Each value arises from a different task, dataset, unit of analysis and reference standard. A 95% audio classifier is not automatically more clinically useful than an 89% disease-image model; robustness depends on label quality, external validation, uncertainty and the consequence of error.

De-risked technological foundation Breed, audio, IMU	Niche technological viability Disease, BCS, estrus	Diagnostically siloed efficacy Food, reproduction	Theoretically robust but clinically unproven OCR + multimodal fusion
--	--	---	--

What the literature teaches across modalities

Accuracy must be contextualised A result is meaningful only when the sample unit, label source, class balance and validation design are known.	The superiority of linear or shallow models in data-constrained environments The BCS study [8] found regularised regressors stronger than more complex nonlinear alternatives on a small high-dimensional dataset.
Real-world data are harder AVA [12] and image/audio studies show that informal language, poor capture conditions and distribution shift can sharply reduce reliability.	Multimodal evidence is complementary Posture, vocalisation, physiology and clinical text describe different aspects of an animal; combining them may reduce ambiguity.
Cats remain underrepresented Dog-only studies dominate the veterinary-AI literature [1], so feline generalisation requires dedicated evaluation.	Veterinary oversight remains essential The review and clinical-pathology literature [2], [13] consistently support AI as an assistive rather than autonomous clinical tool.
Evidence-quality checklist: Before trusting a result, ask: Was the split animal-level? Was the reference label clinically credible? Was performance tested outside the development site? Were failure cases reported? Is uncertainty available?	

4

Comparison & Analysis

Table 1 compares representative studies selected for their relevance to major PawSense modules. The setback column is intentionally prominent because technical accuracy alone does not establish clinical readiness.

Area	Method	Reported result	Evidence	Important setback
Dog breed	Hybrid CNN [4]	92.4% val.	10,222 images / 120 breeds	Dataset-specific; no independent clinical validation
Cat breed	ResNet50 [5]	84.89% val.	953 images / 5 breeds	Small class coverage; limited generalisation
Disease image	ResNet50 [6]	89.2%	>5,000 images	Sensitive to lighting and image quality
BCS	EfficientNetB3 + regression [8]	R ² 0.72	257 post-mortem cats	Preprint; not live smartphone photography
Posture	RF / Isolation Forest [10]	F1-w 0.90	42 dogs + IMUs	Wearable hardware; working-dog bias
Emotion audio	LSTM-GRU [11]	95% dog / 90% cat	Augmented audio	Emotion-label validity and real-world noise
Veterinary LLM	GPT-4o + expert DB [12]	91.4% → 38.6% Top-3	Constructed vs real records	Large real-world performance drop
Estrus stage	Faster R-CNN [15]	91.6%	250 images; 12 final cases	Very small final validation sample
Temperature	AdaBoost / SVR [16]	RMSE 0.15 / 0.14°C	Prospective cats/dogs	Single-centre and missing covariates

Research-paper setbacks that repeatedly appear

Low volume training sets Several studies use a few hundred samples or limited breeds/classes, increasing overfitting risk.	Poor model generalization Only a minority of veterinary-AI studies are tested on independent hospitals, devices or populations [1].	Training deployment Divergence Household images, audio and owner descriptions differ from curated training data.
Expert-Annotated Labels Emotion and behaviour labels may depend on human interpretation rather than objective biomarkers.	Algorithmic risk discrepancy A high classification score does not prove safe triage, diagnosis or treatment guidance.	Species-class data Skew Dog-focused evidence dominates; cat-specific generalisation is weaker [1].

Cross-study interpretation: The most credible evidence is not necessarily the study with the largest headline percentage. Stronger evidence combines a clear clinical target, representative sampling, animal-level splitting, external or prospective validation, transparent ground truth and error analysis.

Model accuracy metrics Accuracy, F1, mAP, RMSE	Clinical Confidence level External validation, sample diversity	Translational Efficacy Safety, explainability, escalation
--	---	---

The literature demonstrates that most individual capabilities already have technical precedents. The unresolved problem is not the absence of algorithms, but the lack of clinically robust integration across modalities, How and when you care for a pet.

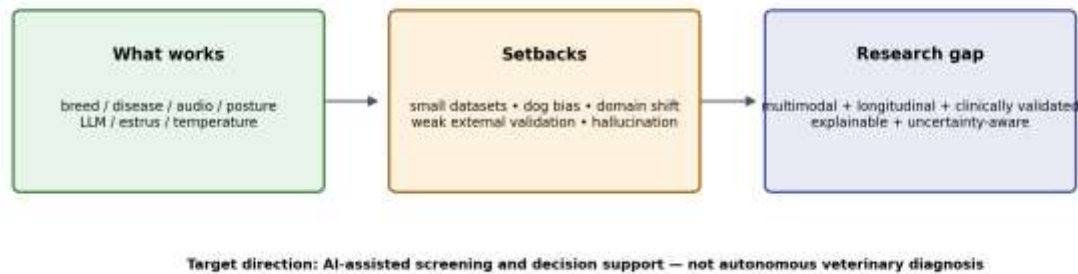


Figure 3. Transition from isolated model performance to the main research gap.

Model Ensembling Most studies analyse one input type. Evidence is limited for systems combining image, audio, posture, physiology and clinical text for the same pet.	Long Term User Adaption Most systems make one-time predictions rather than learning an individual animal's baseline and detecting meaningful change over time.
Cross-Institutional Validation Performance across clinics, breeds, devices, countries and owner-generated data remains insufficiently demonstrated.	Interpretability and Confidence Calibration Healthcare outputs require evidence, confidence and escalation rules. A probability score alone is inadequate when consequences are serious.
Veterinary Clinical Text Mining The reviewed literature supports EHR/text analysis, but direct OCR of photographed reports and prescriptions remains insufficiently validated.	Estrus Prediction Current evidence supports stage identification and biological cycle knowledge, but not robust personalised prediction of the next heat date from owner history alone.

Bridging the gap: Lab to Clinic

Previously Established Breed detection, audio classification, posture/activity recognition, estrous-stage identification.	Warrants More Compelling Data Cross-clinic validation, feline datasets, uncertainty calibration, owner-generated input robustness.	Gap in the Literature Unified multimodal fusion + longitudinal profile + veterinarian-supervised escalation.
Principal Area of Unexplored Enquiry: A unified, multimodal and veterinarian-supervised companion-animal platform that integrates complementary evidence while tracking uncertainty and longitudinal change.		

The Strategic Opportunities for Pawsense.

A Veternity platform becomes more useful when it can connect signals that are weak in isolation. For example, a Voice detection classified as distress may be more meaningful if posture, activity and temperature also deviate from the animal's normal baseline. This does not create a diagnosis automatically; it creates a stronger evidence bundle that can support safer prioritisation and veterinary referral.

Multi-Factor Clinical Risk Prediction Fuse visual, acoustic, behavioural, physiological and clinical evidence rather than treating modules independently.	Feline and Canine-specific longitudinal models Track weight/BCS, activity, temperature, symptoms, reproductive events and medications as time-stamped trends.
Quality-Gated data Ingestion Automatically detect blur, poor framing, occlusion or noisy audio before inference.	Factually anchored NLP Use retrieval-augmented or knowledge-base-grounded answers with sources, uncertainty statements and emergency escalation.
Clinical Supervised verification Validate outputs prospectively with veterinarians across multiple clinics, species and breeds.	On-Premise AI Integration Use consent, minimal data collection, secure storage and potentially edge/federated learning for longitudinal records.

Instructed development roadmap

01	02	03	04
Validate single modules	Add quality + uncertainty	Fuse multimodal evidence	Prospective vet evaluation

Minimum safeguards before real-world deployment

Input gate	Confidence	Facts	Escalation
Reject unusable image/audio	Calibrate uncertainty	Show source/feature basis	Route serious cases to vet
7		Conclusion	

Artificial intelligence has clear potential to support companion-animal healthcare, and the reviewed literature provides evidence for breed recognition, visual disease screening, body-condition assessment, posture/activity monitoring, vocal-emotion classification, veterinary decision support, reproductive-stage identification and non-invasive physiological monitoring. However, the strongest papers also reveal recurring weaknesses: restricted datasets, uncertain labels, hardware or image-quality dependence, limited cat representation, and substantial performance loss when systems move from controlled to real-world conditions.

The real magic happens when the app connects the dots—taking small everyday clues, tracking them over time, and honestly waving a flag when it's not sure what's wrong. When things look serious or confusing, it immediately hands the reins over to a licensed vet. We aren't trying to replace vets here; we're building the ultimate tool to back them up.

Key Objectives for the future system

Early clinical detection	Comprehensive clinical documentation	Predictive grade clinical data	Diagnostic Augmentation
Detect meaningful change sooner	Unify longitudinal observations	Communicate uncertainty + escalation	Provide organised evidence, not replacement

8	References
---	------------

- [1] L. Alaman Trigo et al., "Data-driven artificial intelligence for current diagnostic support in dogs and cats: a scoping review," Research Square, preprint, 2026, doi: 10.21203/rs.3.rs-9963688/v1.
- [2] S. V. Neal, D. G. Rudmann, and K. N. Corps, "Artificial Intelligence in Veterinary Clinical Pathology-An Introduction and Review," Veterinary Clinical Pathology, vol. 54, suppl. 2, pp. S13-S29, 2025, doi: 10.1111/vcp.70012.
- [3] M. Ibraheam, K. F. Li, and F. Gebali, "An Accurate and Fast Animal Species Detection System for Embedded Devices," IEEE Access, vol. 11, 2023, doi: 10.1109/ACCESS.2023.3252499.
- [4] B. Valarmathi, N. Srinivasa Gupta, G. Prakash, R. Hemadri Reddy, S. Saravanan, and P. Shanmugasundaram, "Hybrid Deep Learning Algorithms for Dog Breed Identification-A Comparative Analysis," IEEE Access, vol. 11, 2023, doi: 10.1109/ACCESS.2023.3297440.
- [5] C. H. Patil, S. Bhoite, H. Patil, and M. Jabde, "Deep Learning based Cat Breed Recognition," Grenze International Journal of Engineering and Technology, Jan. 2025.
- [6] M. Dhotay, S. Kirve, A. Walke, P. Mulmule, and M. Polisetty, "AI-Driven Disease Detection and Intelligent CHATBOT for Pet Healthcare Management," Journal of Scientific Advances, vol. 2, no. 2, pp. 50-57, 2025.
- [7] S. A. Sapowicz, D. E. Linder, and L. M. Freeman, "Body Condition Scores and Evaluation of Feeding Habits of Dogs and Cats at a Low Cost Veterinary Clinic and a General Practice," The Scientific World Journal, vol. 2016, Article ID 1901679, 2016, doi: 10.1155/2016/1901679.
- [8] H. Jo et al., "Artificial intelligence-assisted prediction of body condition scores from radiographs of post-mortem free-roaming cats," Research Square, preprint, 2026, doi: 10.21203/rs.3.rs-9458448/v1.
- [9] K. Ferres, T. Schloesser, and P. A. Gloor, "Predicting Dog Emotions Based on Posture Analysis Using DeepLabCut," Future Internet, vol. 14, no. 4, art. 97, 2022, doi: 10.3390/fi14040097.
- [10] M. Marcato, S. Tedesco, C. O'Mahony, B. O'Flynn, and P. Galvin, "Machine learning based canine posture estimation using inertial data," PLOS ONE, vol. 18, no. 6, e0286311, 2023, doi: 10.1371/journal.pone.0286311.
- [11] S. S. R. A. Pearline, G. Subbiah, A. Sanu, and K. Charumathi Venkatesan, "Multi-QuadEmoNet: cat and dog emotion classification model from animal vocalization using multi-stage LSTM-GRU paradigm," Frontiers in Veterinary Science, vol. 13, art. 1799281, 2026, doi: 10.3389/fvets.2026.1799281.
- [12] Y.-G. Jin et al., "AI Veterinary Assistance: Enhancing Clinical Decision-Making in Animal Healthcare," IEEE Access, vol. 13, 2025, doi: 10.1109/ACCESS.2025.3587787.
- [13] M. Jokar, A. Abdous, and V. Rahmian, "AI chatbots in pet health care: Opportunities and challenges for owners," Veterinary Medicine and Science, vol. 10, e1464, 2024, doi: 10.1002/vms3.1464.
- [14] C. Cortinovis and F. Caloni, "Household Food Items Toxic to Dogs and Cats," Frontiers in Veterinary Science, vol. 3, art. 26, 2016, doi: 10.3389/fvets.2016.00026.
- [15] G. Calderon, C. Carrillo, M. Nakano, J. Acevedo, and J. E. Hernandez, "Automatic Estrus Cycle Identification System on Female Dogs Based on Deep Learning," in Pattern Recognition, Lecture Notes in Computer Science, vol. 12088, pp. 261-268, Springer, 2020, doi: 10.1007/978-3-030-49076-8_25.
- [16] Z. Zhao et al., "A non-invasive method to determine core temperature for cats and dogs using surface temperatures based on machine learning," BMC Veterinary Research, vol. 20, art. 199, 2024, doi: 10.1186/s12917-024-04063-2.
- [17] V. Punyapornwithaya et al., "Time series analysis and forecasting of the number of canine rabies confirmed cases in Thailand based on national-level surveillance data," Frontiers in Veterinary Science, vol. 10, art. 1294049, 2023, doi: 10.3389/fvets.2023.1294049.
- [18] D. D. Bhavani, M. H. S. Quadri, and Y. Ram Reddy, "Dog Breed Identification Using Convolutional Neural Networks on Android," CVR Journal of Science and Technology, vol. 17, Dec. 2019, doi: 10.32377/cvrjst1711.
- [19] S. Gupta, A. S. Chavan, A. Deepak, A. Kumar C., S. Pundir, R. Bajaj, and A. Shrivastava, "Speech Emotion Recognition of Animal Vocals Using Deep Learning," International Journal of Intelligent Systems and Applications in Engineering, vol. 12, no. 13s, pp. 129-136, 2024.
- [20] Sanjana M., Shreya P., R. G. K., and Chaitra H. V., "Petzy: AI-powered interactive pet companion," ITM Web of Conferences, vol. 81, art. 01010, 2026, doi: 10.1051/itmconf/20268101010.
- [21] A. U. Haq, "Behavioral disorders of dogs and cats," in Introduction to Diseases, Diagnosis, and Management of Dogs and Cats, T. Rana, Ed. Academic Press/Elsevier, 2024, ch. 8.
- [22] L. Lindh et al., "Ovarian and uterine changes during the oestrous cycle in female dogs," Reproduction, Fertility and Development, 2022, doi: 10.1071/RD22177.
- [23] G. McRae, E. Runcan, and A. Johnson, "Estrous cycle manipulation in dogs," Clinical Theriogenology, vol. 17, art. 11647, 2025, doi: 10.58292/CT.v17.11647.
- [24] S. E. Little, "Female Reproduction," in The Cat: Clinical Medicine and Management. St. Louis, MO, USA: Saunders/Elsevier, 2012, ch. 40.
- [25] L. van Stee, F. van der Steen, S. Galac, H. Kooistra, and B. Meij, "Rare pituitary diseases in dogs and cats and their human companions," Rare Disease and Orphan Drugs Journal, vol. 5, art. 19, 2026, doi: 10.20517/rdodj.2025.78.