Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Machine learning for detecting colony collapse disorder and colony losses in honeybees (*Apis mellifera*): a PRISMA 2020 systematic review

Fredrick Michael Ogore

fogore@daystar.ac.ke

Daystar University Nairobi School of Science Engineering and Health https://orcid.org/0000-0001-7099-6761

Systematic Review

Keywords: Colony collapse disorder, Colony losses, Machine learning, Deep learning, Precision beekeeping, Bioacoustics, Varroa destructor, Systematic review

Posted Date: November 3rd, 2026

DOI: https://doi.org/10.21203/rs.3.rs-11577679/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License. Read Full License

Additional Declarations: The authors declare no competing interests.

Abstract

Introduction: Globally, pollination services and beekeeping livelihoods are at risk due to colony collapse disorder (CCD) and increased colony losses. Although early warning is promised by machine learning (ML) applied to hive sensor, image, and survey data, the proof has not been synthesized with a concentration on collapse and loss.

Objectives: This systematic review mapped and assessed studies that employed machine learning (ML) to identify or predict CCD, colony loss, or the colony-health states that precede loss in Apis mellifera, commonly known as honeybees. The data, models, validation processes, and deployment levels were also examined.

Methods: Peer-reviewed studies published between 2015 and 2025 were found using citation searches of two recent reviews and organized academic web searching after PRISMA 2020. 58 studies were included after 110 records were screened and 69 full texts were evaluated. An eight-point standardized checklist tailored for machine learning research was used to evaluate methodological quality after task, modality, dataset, model, validation, performance, and deployment data were extracted.

Results: Colony collapse was explicitly modeled in only two studies. The remaining investigations identified the following precursors: swarming (8), multi-condition anomalies (12), parasites and diseases (16 studies, primarily Varroa destructor), queen status (13), colony loss or winter mortality (4), and colony strength (3). Deep learning was the most common model family (20 research), and acoustic monitoring predominated (27 studies). Although just ten research examined models on independent hives, where performance significantly declined, reported accuracies frequently approached 95%; prediction of loss models reached roughly 65–75% accuracy. 18 studies were good quality, 26 had moderate quality, and 14 had low quality.

Conclusions: Although there is little evidence to forecast collapse itself, machine learning (ML) consistently identifies a few proximate causes of colony failure. Multimodal fusion, hive-independent validation, open data and code, and longitudinal, multi-apiary datasets with confirmed loss consequences are all necessary for progress.

INTRODUCTION

The western honeybee (*Apis mellifera*) is the most important managed pollinator of crops and wild plants, and the global decline of pollinators has been linked to land-use change, pathogens, pesticides and climate (Klein et al., 2007; Potts et al., 2010). From 2006 onwards, beekeepers in the United States reported sudden, unexplained disappearances of adult workers from otherwise healthy-looking colonies, a syndrome named colony collapse disorder (CCD), characterised by rapid loss of the adult population, the presence of brood and food stores, and the absence of dead bees in or near the hive (Cox-Foster et al., 2007; vanEngelsdorp et al., 2009). Although classical CCD has become rarer, elevated overwintering and annual colony losses persist across Europe, North America and other regions and are now attributed to interacting stressors, including *Varroa destructor* and associated viruses, other pathogens, pesticide exposure, poor nutrition, queen failure and extreme weather (Neumann & Carreck, 2010; Goulson et al., 2015; Steinhauer et al., 2018; Gray et al., 2020).

Colony failure is usually recognised only after the event, because manual hive inspections are labour-intensive, intermittent and disruptive. Precision beekeeping seeks to replace or complement inspections with continuous monitoring of hive weight, temperature, humidity, sound, vibration, gases and images (Meikle & Holst, 2015; Zacepins et al., 2015; Hadjur et al., 2022). The resulting data streams are high-dimensional and noisy, and machine learning (ML) — from support vector machines and random forests to convolutional and recurrent neural

networks — has become the main tool for turning them into actionable alerts about queen loss, swarming, parasitism or colony decline (Bilik et al., 2024).

Existing reviews have catalogued precision-beekeeping hardware, sensing systems and computer-vision methods (Hadjur et al., 2022; Bilik et al., 2024; Šabić et al., 2025) and the use of tiny machine learning (TinyML) for hive monitoring (Sucipto et al., 2025), while broader surveys document rapid progress in running ML models on low-power microcontrollers and edge devices (Abadade et al., 2023; Heydari & Mahmoud, 2025; Somvanshi et al., 2025; Han et al., 2026). However, none has synthesised the evidence from the standpoint of the problem that motivates much of this work: detecting or predicting CCD and colony losses. It therefore remains unclear which loss-related outcomes have been modelled, how robustly models have been validated, whether they generalise beyond the hives used for training, and how close they are to field deployment.

Objectives

This systematic review addresses that gap. Following the PRISMA 2020 statement (Page et al., 2021), it maps and appraises peer-reviewed studies published between 2015 and 2025 that applied ML to detect CCD, colony loss, or the health states that precede loss in *A. mellifera*. The review answers four research questions (RQs):

- **RQ1.** Which collapse- and loss-related outcomes have been detected or predicted with ML?
- **RQ2.** Which data modalities, features and ML model families have been used?
- **RQ3.** How have models been validated, how well do they perform, and do they generalise across hives and apiaries?
- **RQ4.** To what extent have systems been deployed in the field, and what are the main research gaps?

METHODS

Protocol and reporting

PRISMA 2020 was followed in the conduct and reporting of the review (Page et al., 2021). The protocol was not pre-registered. The PRISMA counts and descriptive syntheses are produced using formulas from the record-level data in Supplementary File S1, an Excel spreadsheet that contains all records, screening choices, extracted data, quality ratings, and the computations underlying each count and figure.

Eligibility criteria

The PICOS framework determined eligibility (Table 1). Eligible research was published from 1 January 2015 to 31 December 2025, focused on A. mellifera colonies, trained or assessed at least one machine learning model (supervised, unsupervised, semi-supervised, or deep learning), examined CCD, colony loss or mortality, or a colony-level health condition preceding loss, and consisted of peer-reviewed full journal or conference articles or data descriptors with an ML evaluation. Lack of a queen, swarming, parasite or disease infestation, diminished colony strength, and unusual colony behavior are immediate causes of colony failure (Rosenkranz et al., 2010; Steinhauer et al., 2018).

Table 1. Eligibility criteria (PICOS) and exclusion codes used during screening.

Element	Inclusion	Exclusion code
Population	*Apis mellifera* colonies (any subspecies, any country)	EC2 Other *Apis* species (e.g., *A. cerana*)
Intervention	ML or deep-learning model trained and/or evaluated on colony data	EC3 Purely statistical, rule-based or descriptive analysis
Comparator	Ground truth from inspections or labels; alternative models (not required)	—
Outcomes	CCD or collapse; colony loss or winter mortality; queen status; swarming; parasites and diseases; colony strength; health anomalies	EC4 No colony-health or loss outcome (e.g., traffic counting, pollen, honey or weight only); EC7 predator (hornet) detection
Study design	Peer-reviewed journal or conference paper, or data descriptor with ML evaluation	EC5 Reviews; EC6 theses, abstracts, preprints, withdrawn papers; EC8 dataset-only papers; EC9 unverifiable records
Time and language	2015–2025; English (one Portuguese paper with an English abstract retained)	EC1 Outside 2015–2025

Information sources and search strategy

Structured scholarly web searches on IEEE Xplore, ScienceDirect, SpringerLink, MDPI, Nature Portfolio, the Brazilian Computer Society (SBC) digital library, SciTePress, IOS Press, and institutional repositories found records. The searches occurred in October 2026. The search strategy included three concept blocks: population (honey bee, Apis mellifera, beehive), outcome (colony collapse disorder, colony loss, winter mortality, queenless, swarming, Varroa, foulbrood, Nosema, colony strength, anomaly), and method. The complete search strings are in Table 2 and Supplementary File S1. The count excludes news articles, product pages, and blogs that were not scholarly records before screening. Backward citation searching was used to examine the reference lists of two recent reviews (Šabić et al., 2025; Sucipto et al., 2025).

Table 2. Search strings and approaches (searches run in October 2026).

NO	Search string or approach	Concept
1	"honey bee" OR "honeybee" "colony collapse disorder" "machine learning" OR "deep learning" detection	CCD + ML
2	"*Apis mellifera*" "colony loss" OR "winter mortality" "random forest" OR "machine learning" prediction	Colony loss
3	beehive audio "queenless" OR "queen absence" CNN OR SVM OR LSTM	Queen status
4	honey bee swarming prediction audio OR vibration "machine learning" OR "neural network"	Swarming
5	"*Varroa destructor*" detection "deep learning" OR YOLO OR "Faster R-CNN" images	*Varroa* (vision)
6	*Varroa* hyperspectral OR "gas sensor" OR "electronic nose" classification honey bee	*Varroa* (spectral/gas)
7	beehive "anomaly detection" autoencoder OR "isolation forest" sensor data	Anomalies
8	beehive monitoring TinyML OR "edge computing" OR IoT "deep learning" colony health	Deployment
9	honey bee colony strength OR population estimation audio "machine learning"	Colony strength
10	honey bee disease foulbrood OR *Nosema* detection "machine learning"	Other pathogens
11	beehive sound classification health MFCC spectrogram classifier	Acoustic health
12	Publisher-targeted repeats of strings 1–11 (IEEE Xplore, ScienceDirect, MDPI, SBC, SciTePress, Nature)	Verification
13	Backward citation searching of two reviews	Other methods

Study selection

Titles and abstracts were screened against the eligibility criteria, followed by assessment of full texts or full publisher records. Duplicates, including the same study published in two venues, were removed before screening. Selection was performed by a single reviewer; every decision and its exclusion code is recorded in Supplementary File S1 so that it can be audited and repeated.

Data extraction

A standardised extraction form captured: bibliographic details; country; detection task; target outcome; data modality; dataset or data source; sample size; ML family; models tested; best-performing model; validation strategy; key performance metrics; deployment level; limitations; and evidence basis (full text or abstract). Tasks were grouped into seven categories, modalities into seven groups and models into five families (deep learning; classical ML; hybrid classical ML with neural networks; unsupervised, anomaly-detection or semi-supervised; and other or unspecified). Values not reported in the accessible material were coded as not reported (NR).

Quality appraisal

There is no universally accepted risk-of-bias instrument for machine learning research in animal or agricultural surveillance. Consequently, we employed an eight-point checklist modified from PROBAST and CLAIM (Wolff et al., 2019; Mongan et al., 2020), which encompasses: Q1 a distinctly articulated objective and result; Q2 an elucidated data source and collection method; Q3 a sufficient sample (hives, recordings, or images); Q4 a

documented validation procedure; Q5 an autonomous or cross-hive assessment; Q6 numerical performance indicators; Q7 juxtaposition with a baseline or alternative models; and Q8 accessibility of data or code. Every item received a rating of yes (1), partial (0.5), or no (0), resulting in a maximum possible score of 8. Research scoring a minimum of 6 was classified as high quality, scores between 4 and 5.5 were deemed moderate, and those below 4 were considered low. Quality assessments were employed to influence the narrative synthesis, rather than to eliminate studies.

Synthesis

Because tasks, outcomes, datasets and metrics were highly heterogeneous, meta-analysis was not appropriate. Results were synthesised narratively and descriptively by task, modality, model family, validation strategy and deployment level (RQ1–RQ4). Counts and cross-tabulations were computed from the extraction sheet and are shown in Figures 2–7.

RESULTS

Study selection

Atotal of \92 records were found through web searching, and 18 records were found through citation searching (Figure 1). In total, 110 records were found. Following the elimination of two duplicates, 108 records were chosen for screening, and 27 were rejected based on their titles or abstracts. These 27 records were primarily reviews, theses, abstracts, or preprints, as well as studies that did not include a colony-health outcome and studies of predatory hornets. Twelve of the 81 reports that were requested could not be located. 11 of the 69 reports that were evaluated in their entirety were disregarded. This was primarily due to the fact that no machine learning model was trained or evaluated, the paper described a dataset that did not have an evaluation of machine learning, or the study concerned a different species of Apis. There were a total of 58 studies that were incorporated, with 49 coming from web searches and 9 coming from citation searches.

Characteristics of included studies

The 58 studies incorporated were published from 2017 to 2025, with a significant rise in the number of publications during this timeframe: 33 studies (57%) were released starting in 2023, peaking at 18 studies in 2024 (Figure 2). A total of 39 articles were disseminated in journals, 17 papers were delivered at conferences, and two data descriptors featuring benchmark analyzes completed the compilation. A considerable segment of the inquiry was conducted in Europe, particularly in the Czech Republic, the United Kingdom, Poland, Spain, Germany, Italy, and Greece. The nations of Brazil, Canada, the United States, and New Zealand each provided further contributions. None of the studies referenced sub-Saharan Africa. The synopsis of all studies incorporated is available in Table 3.

Table 3. Summary of the 58 included studies, grouped by detection task. NR = not reported in the accessible material.

Study	Country	Task	Data modality	Best model	Key performance
Kontogiannis (2024)	Greece	Collapse	Multimodal (temperature, humidity, audio)	NR	NR in retrieved excerpt
Zhukov and Horbenko (2025)	Ukraine	Collapse	Weight + temperature time series	LSTM	Accuracy 96% (point states); 92% (global states incl. collapse)
Calovi et al. (2021)	USA	Loss / mortality	Tabular (beekeeper survey + weather)	Random forest	Accuracy 73.3% (colonies), 65.7% (apiaries) where *Varroa* managed
Insolia et al. (2022)	USA	Loss / mortality	Tabular (survey, stressors, weather, land use)	NR (abstract)	NR (abstract); *Varroa*, pesticides, extreme temperature/precipitation significant
García-Vicente et al. (2024)	Spain	Loss / mortality	Tabular (inspections + pathogen loads)	Random forest	Test accuracy 0.70; sensitivity 0.46; specificity 0.93
Zhu et al. (2024)	Canada	Loss / mortality	Multimodal (audio, temperature, humidity)	Isolation forest (survival)	Winter survival AUC 0.730 (all modalities); population balanced accuracy 65.8%; queen cells AUC 0.646
Braga et al. (2019)	Brazil	Multi-condition / anomaly	Multimodal (in-hive/out-hive sensors + inspections)	NR	Accuracy 99.36%
Kaplan Berkaya et al. (2021)	Turkey	Multi-condition / anomaly	Images	NR (deep-feature SVM/DNN)	Accuracy up to 99.07%
Kim et al. (2021)	South Korea	Multi-condition / anomaly	Audio	VGG-13 (MFCC)	Accuracy 91.93%
Rodrigues et al. (2021)	Brazil	Multi-condition / anomaly	Multimodal (audio, temperature, humidity)	ELM (temp/RH); GMM (audio)	Mean accuracy 92.6% (ELM, temp/RH); 84.9% (GMM, queen absence)
Cejrowski and Szymański (2022)	Poland	Multi-condition / anomaly	Multimodal (audio, temperature)	CVAE + Isolation Forest	F1 0.16 → 0.8 → 1.0 on anomaly days
Senger et al. (2023)	Germany	Multi-condition / anomaly	Multimodal (temperature, humidity, weight + beekeeper logs)	NR	NR
Torky et al. (2023)	Egypt	Multi-condition	Images	MobileNet + Adam	Test accuracy 95%

Study	Country	Task	Data modality	Best model	Key performance
		/ anomaly			
Chen et al. (2024)	Taiwan	Multi-condition / anomaly	Audio (IoT)	NR	Accuracy 98.8%; F1 0.98
Faramarzi et al. (2024)	Iran	Multi-condition / anomaly	Audio (IoT)	InceptionV3	Overall 98.74%; queenless 98.62%; *Varroa* 98.59%; foulbrood 98.71%; *Nosema* 98.80%
Hammami and Abdulaziz (2024)	UAE	Multi-condition / anomaly	Multimodal (temperature, CO2, weight)	NR	NR
Zhukov and Horbenko (2024)	Ukraine	Multi-condition / anomaly	Audio	Pre-trained CNN (MFCC)	NR (qualitative)
Mustafa et al. (2025)	New Zealand	Multi-condition / anomaly	Audio	Transformer	Accuracy 99%; Normal/NoQueen P,R ≥0.99; Swarm 0.96
Schurischuster et al. (2018)	Austria / Spain	Parasite / disease	Video frames	NR	NR ("adequate")
Bjerge et al. (2019)	Denmark	Parasite / disease	Video (multispectral illumination)	Infestation Level Estimator	Infestation 5.80% vs 5.52% (ground truth); F1 bees 0.97; F1 mites 0.91
Szczurek et al. (2020a)	Poland	Parasite / disease	Gas / VOC sensors	SVM	Error $_{10\%\ (low\ class);}$ 20% (high class)
Szczurek et al. (2020b)	Poland	Parasite / disease	Gas / VOC sensors	SVM (balanced data)	TPR 0.93; TNR 0.95 (balanced)
Bilik et al. (2021)	Czech Republic	Parasite / disease	Images	Object detectors (YOLO/SSD)	F1 infected bee up to 0.874; F1 mite up to 0.727
Mrozek et al. (2021)	Poland	Parasite / disease	Video / images	NR	NR (abstract); effect of image resolution and time performance assessed
Picek et al. (2022)	Czech Republic	Parasite / disease	Images (smartphone)	VarroaNet	Mean infestation-level accuracy 96.0% (autumn), 93.8% (winter)
Voudiotis et al. (2022)	Greece	Parasite / disease	Images (in-hive cameras)	Faster R-CNN MobileNetV2	Overall *Varroa* detection accuracy ~68–70% (77% accuracy / 86% precision in mite step)
Divasón et al. (2023)	Spain	Parasite / disease	Images (smartphone)	Faster R-CNN ResNet50-FPN	mAP50 0.737; 163/192 mites detected (96 FP)

Study	Country	Task	Data modality	Best model	Key performance
Divasón et al. (2024)	Spain	Parasite / disease	Images (smartphone)	Faster R-CNN resnet18_fpn (recommended)	mAP50 0.9073
Duma et al. (2024)	Czech Republic / Finland	Parasite / disease	Hyperspectral images	KF-PLS	Four spectral bands sufficient for accurate identification (values NR)
Ghezal et al. (2024)	Germany	Parasite / disease	Hyperspectral images	PCA-kNN-SVM	Segmentation accuracy >99%; AP 0.9983; recall 0.9947
Sharma et al. (2024)	New Zealand	Parasite / disease	Images	NR	NR ("promising")
Ghezal and König (2025)	Germany	Parasite / disease	Spectral images	Hybrid ML (low-data regime)	NR (record); ML superior in low-data scenarios
Scutaru et al. (2025)	Italy	Parasite / disease	Images (macro scanner)	Two-stage CNN/ResNet	R^2 0.998 vs ground truth (visual 0.967); error <1% above 10 mites
Yániz et al. (2025)	Spain	Parasite / disease	Images (smartphone)	YOLOv11-nano	R^2 0.98–0.99 vs control
Robles-Guerrero et al. (2017)	Mexico	Queen status	Audio	Logistic regression (2 MFCC means)	Perfect ROC classification with two features (qualitative)
Nolasco et al. (2019)	UK / Italy	Queen status	Audio	NR	NR (abstract); generalisation to new hives challenging
Ruvinga et al. (2021)	UK	Queen status	Audio	LSTM (qualitative)	NR ("particularly encouraging")
Rodrigues et al. (2022)	Brazil / Sweden	Queen status	Audio	Naive Bayes	Accuracy 97% (NB, 10 × 1-s windows/day); response 0.93 s
Terenzi et al. (2022)	Italy / UK	Queen status	Audio	NR	NR (record)
Barbisan et al. (2023)	Italy	Queen status	Audio	MLP (STFT 2048)	Accuracy 0.9939; F1 0.9965
Ruvinga et al. (2023)	UK	Queen status	Audio	NR in retrieved text (LSTM reported strong)	NR in retrieved text ("extremely positive")
Bricout et al. (2024)	France	Queen status	Audio	NR	NR (record)
De Simone et al. (2024)	Italy	Queen status	Audio	Compact NN	Comparable to baseline with reduced precision (values NR)
Huang et al. (2024)	China	Queen status	Audio	SVM	Accuracy 99.30%

Study	Country	Task	Data modality	Best model	Key performance
Nguyen Ba et al. (2024)	Vietnam	Queen status	Audio (IoT)	CNN + SVM	Accuracy 98.61% (CNN alone 84.67%)
Carvalho et al. (2025)	Brazil	Queen status	Audio	CAE + HDBSCAN	Accuracy 98.78%
Sad et al. (2025)	Greece	Queen status	Audio	NN (edge); KNN (hive m12)	NN F1 0.85/0.92; KNN F1 0.84/0.97; transfer drop 10–14%
Davidson et al. (2020)	Germany	Swarming	Temperature (multi-sensor hives)	LSTM autoencoder	Detected all baseline-detected swarms + 1 additional; no quantitative metrics
Ramsey et al. (2020)	UK / France	Swarming	Vibration (accelerometers)	NR	Accuracy >90%; prediction up to 30 days ahead
Zgank (2021)	Slovenia	Swarming	Audio	DNN (4 hidden layers)	Accuracy 94.09%; F1 0.970 (HMM 88.02%)
Kiromitis et al. (2022)	Greece	Swarming	Audio	SVM (early); k-NN (late)	Early: SVM 90%, CNN 89%; late: k-NN 98%, SVM 97%, CNN 95%
Ruvinga et al. (2022)	UK	Swarming	Audio	CNN (Mel)	Accuracy 99.6–99.8% (3–7 days before); ~90% (28 days)
Várkonyi et al. (2023)	Hungary / UK	Swarming	Audio	NR (record)	NR (record); proposed filtering outperformed baselines
Iqbal et al. (2024)	Pakistan / Saudi Arabia / Korea	Swarming	Audio	CNN (MFCC)	Accuracy 99% (CNN-MFCC); 97% (SVM-Mel)
Libal and Biernacki (2024)	Poland	Swarming	Audio (entrance)	1-layer AE + Burg PSD	Accuracy 95.9%; recall 91.9%; F1 ~96.1%; false alarms 0%
Abdollahi et al. (2023)	Canada	Colony strength	Audio	CNN-MFCC	MAE 0.86 frames (random split); large errors hive-independent; rain +355% MAE
Luz et al. (2024)	Brazil	Colony strength	Audio (smartphone)	MFCC + MLP	Accuracy 95.53%; F1 0.955; kappa 0.908
Abdollahi et al. (2025)	Canada	Colony strength	Multimodal (audio, temperature, humidity)	NR	NR (abstract)

Detection tasks (RQ1)

Figure 3 shows the distribution of tasks. Parasite and disease detection was the largest category (16 studies, 28%), followed by queen-status detection (13), multi-condition health or anomaly classification (12) and swarming (8). Colony loss or mortality prediction (4), colony strength estimation (3) and explicit collapse detection (2) were rare.

Colony collapse

Merely two investigations explicitly simulated the system's collapse. Kontogiannis (2024) created a hive device for edge computing that amalgamated temperature and humidity data streams. This information was examined utilizing a fuzzy-stranded neural network to discern essential factors associated with Colony Collapse Disorder (CCD), in conjunction with audio convolutional neural networks (CNNs) to observe swarming activity and the loss of queens. The system issued notifications to a cloud dashboard; nevertheless, quantitative performance metrics for CCD were absent from the available text. Zhukov and Horbenko (2025) employed long short-term memory (LSTM) networks on time series data pertaining to hive weight and temperature, attaining a classification accuracy of 92% for global hive states, encompassing cases of colony collapse. Furthermore, a later investigation inspired by CCD detected seasonal irregularities in acoustic, temperature, and humidity trends (Rodrigues et al., 2021), yet did not classify collapse as a result in this study.

Colony loss and winter mortality

Four studies predicted the decline of colonies. Three utilized tabular datasets: state-specific stressor, meteorological, and land-use information throughout the USA (Insolia et al., 2022); a beekeeper survey alongside weather data from Pennsylvania, where random forests attained 73.3% accuracy for colonies and 65.7% for apiaries (Calovi et al., 2021); and inspection and pathogen datasets from 179 hives in Spain, where a SHAP-explained random forest reached 0.70 accuracy but only 0.46 sensitivity regarding loss (García-Vicente et al., 2024). The MSPB dataset, comprising 53 colonies observed through audio, temperature, and humidity over a year, established the sole sensor-based standard for winter survival (AUC 0.730) (Zhu et al., 2024). In these investigations, Varroa infestation levels, brood conditions, food supplies, and severe temperature or rainfall were identified as the primary indicators.

Parasites and diseases

The majority of the 16 research studies in this classification focused on V. destructor. Computer-vision systems identified mites on bees at the hive's entrance (Bjerge et al., 2019; Bilik et al., 2021; Voudiotis et al., 2022; Schurischuster et al., 2018) or quantified fallen mites on adhesive boards and debris captured via smartphones or scanners (Picek et al., 2022; Divasón et al., 2023; Divasón et al., 2024; Scutaru et al., 2025; Yániz et al., 2025). Board-counting methodologies exhibited a strong correlation with manual tallies (R^2 0.98–0.998) (Scutaru et al., 2025; Yániz et al., 2025), while the detection of mites on bees proved more challenging, yielding F1 scores between 0.73 and 0.91 (Bjerge et al., 2019; Bilik et al., 2021) and approximately 70% overall accuracy for an in-hive imaging system (Voudiotis et al., 2022). Hyperspectral and spectral imaging utilizing traditional machine learning effectively distinguished mites from bees with exceptional accuracy on limited datasets (Ghezal et al., 2024; Ghezal & König, 2025; Duma et al., 2024), while gas-sensor arrays employing support vector machines differentiated infested colonies from healthy ones (Szczurek et al., 2020a; Szczurek et al., 2020b). Conditions aside from varroosis, including foulbrood and Nosema, were only discussed within the context of more general acoustic classifiers (Faramarzi et al., 2024).

Queen status

The absence of a queen was identified primarily through hive acoustics (Table 3). Pipelines utilizing mel-frequency cepstral coefficients (MFCC), short-time Fourier transform (STFT), or Hilbert–Huang transform features in conjunction with SVM, MLP, CNN, or LSTM classifiers have demonstrated accuracies ranging from 97% to 99.9% across various datasets (Nguyen Ba et al., 2024; Huang et al., 2024; Barbisan et al., 2023; Rodrigues et al., 2022). Nevertheless, research assessing cross-hive generalization revealed significantly diminished and less consistent performance (Nolasco et al., 2019; Bricout et al., 2024; Sad et al., 2025); for instance, Sad et al. (2025) noted a 10–14% decline when transferring to a different hive. Unsupervised methodologies (convolutional autoencoder combined with HDBSCAN) attained 98.8% accuracy devoid of labels (Carvalho et al., 2025), while TinyML applications exhibited real-time inference capabilities on microcontrollers (De Simone et al., 2024; Sad et al., 2025).

Swarming

Swarming predictions relied on auditory signals (Iqbal et al., 2024; Ruvinga et al., 2022; Zgank, 2021; Kiromitis et al., 2022; Libal & Biernacki, 2024; Várkonyi et al., 2023), vibrational cues (Ramsey et al., 2020), or thermal anomalies (Davidson et al., 2020). Accelerometer spectra predicted swarming with over 90% accuracy up to 30 days ahead (Ramsey et al., 2020), whereas Mel-spectrogram CNNs attained an accuracy of 99.6–99.8% three to seven days before swarming, declining to around 90% at 28 days (Ruvinga et al., 2022). In the context of late-stage detection, conventional classifiers such as k-NN and SVM yielded comparable or superior outcomes to a U-Net CNN (Kiromitis et al., 2022).

Colony strength and multi-condition anomalies

Colony strength was estimated from acoustics by either MFCC-based CNN regression or MLP classification (Luz et al., 2024; Abdollahi et al., 2023; Abdollahi et al., 2025). Abdollahi et al. (2023) pointed out the significance of context; errors peaked in hive-independent testing and surged by 355% in rain. Studies of multi-condition and anomaly (Kaplan Berkaya et al., 2021; Kim et al., 2021; Mustafa et al., 2025; Torky et al., 2023; Faramarzi et al., 2024; Braga et al., 2019; Cejrowski & Szymański, 2022; Senger et al., 2023; Hammami & Abdulaziz, 2024; Chen et al., 2024) evaluated multiple states at once, e.g., identifying queenless, virus infected, and pesticide contaminated colonies with 98.8% accuracy in field tests (Chen et al., 2024), or detected deviations from learned normal behaviour using autoencoders and isolation forests (Cejrowski & Szymański, 2022).

Data modalities and features (RQ2)

Audio and vibration were employed in 27 investigations, representing 47% of the overall total. Thirteen studies utilized images or videos, whereas eight employed multimodal in-hive sensors. Furthermore, three studies utilized survey or tabular data, another three employed spectral or hyperspectral imaging, two incorporated single physical sensors, and two featured gas sensors (refer to Figure 4). Acoustic research primarily utilized Mel-frequency cepstral coefficients (MFCC), Mel spectrograms, and short-time Fourier transform (STFT), with only a limited number of studies, including Terenzi et al. (2022), examining wavelet, Hilbert–Huang, and constant-Q representations. The task-by-modality matrix (Figure 6) reveals a significant relationship between modalities: auditory methods were chiefly utilized for queen status and swarming analysis, visual techniques were mainly employed for parasite detection, and tabular data was essential for loss prediction. A mere eight studies incorporated various modalities.

Machine-learning approaches

The most widespread category was deep learning with 20 studies, followed by classical machine learning (13 studies), hybrid pipelines combining deep features or neural networks with classical classifiers (13 studies) and unsupervised, anomaly-detection or semi-supervised methods (9 studies). Deep learning went up from one study before 2020 to 13 studies from 2023 to 2025, whereas classical machine learning stayed stable (Figure 5). Convolutional Neural Networks (custom, VGG, ResNet, MobileNet, Inception, DenseNet, EfficientNet) and object detection algorithms (Faster R-CNN, YOLO, SSD) have been the best in image processing tasks; Convolutional Neural Networks, Long Short-Term Memory networks, and more recently, Transformers have been the best in acoustic processing tasks; while random forests, Support Vector Machines, and k-Nearest Neighbors have remained competitive, especially in scenarios with limited data (Ghezal & König, 2025; Kiromitis et al., 2022). Explainability was not much discussed, with SHAP (García-Vicente et al., 2024) and Grad-CAM (Kim et al., 2021) being notable exceptions.

Validation and performance (RQ3)

The majority of the studies reported accuracy or F1 scores above 90%, and many above 95% (Table 3). However, these figures were usually obtained using random train-test splits or k-fold cross-validation on clips or images taken from the same few hives. The recordings from the adjacent hives are strongly correlated and hence this design is prone to information leakage and optimistic estimates (Kapoor & Narayanan, 2023). Only 10 studies (17%) used an independent temporal or cross-hive test. Where reported, performance was consistently lower (Nolasco et al., 2019; Abdollahi et al., 2023; Sad et al., 2025), with one proposed solution being the harmonization of multiple public datasets (Bricout et al., 2024). Many acoustic studies had small sample sizes (often <10 hives) and 18 studies (31%) did not provide quantitative metrics in the available material. By contrast, loss-prediction models trained on survey and inspection data reached more modest accuracies (around 65-75%) and low sensitivity for the minority loss class (Calovi et al., 2021; García-Vicente et al., 2024), arguably a more realistic indication of the difficulty of forecasting colony failure.

Deployment (RQ4)

Of these 28 (48%) were offline analyzes. Ten reported edge or embedded deployment, including Edge TPU, Arduino-class microcontrollers and TinyML (Mrozek et al., 2021; De Simone et al., 2024; Sad et al., 2025). Seven described IoT or cloud monitoring systems. Five released smartphone applications or open software for Varroa counting (Picek et al., 2022; Divasón et al., 2024; Yániz et al., 2025). Six reported field prototypes. Two studies have released longitudinal multi-sensor datasets with phenotypic and survival labels (Zhu et al., 2024; Abdollahi et al., 2025) providing much-needed benchmarks. None of the studies reported long-term field evaluation with beekeepers, alarm-threshold calibration and cost-benefit analysis.

Methodological quality

The mean curricular quality score was 4.9 out of 8. Eighteen studies (31%) were classified as high quality, 26 (45%) as moderate quality, and 14 (24%) as low quality. All research articulated a distinct objective, yet merely 14 (24%) possessed a sufficiently robust sample, 10 (17%) evaluated models using independent datasets, and 16 (28%) disseminated data or code (Figure 7). Deficiencies were particularly evident in conference papers and in research evaluated solely from abstracts.

DISCUSSION

Principal findings

This analysis recognized 58 investigations that utilized machine learning to assess colony health results associated with colony collapse disorder and colony decline. The domain has expanded swiftly since 2020 and has transitioned toward deep learning and edge implementation. Three discoveries are particularly noteworthy. Initially, nearly all research does not forecast collapse or failure directly from sensor data; rather, it focuses on immediate, observable indicators. Secondly, the accuracy within the dataset is elevated, yet the proof of generalization is feeble. Thirdly, the foundation of evidence is primarily comprised of limited, single-location acoustic and imaging datasets originating from Europe.

From precursor detection to collapse prediction

Detecting queenlessness, swarming, or Varroa is crucial as each represents a known pathway to colony collapse and is manageable by beekeepers (Rosenkranz et al., 2010; Gray et al., 2020).However, winter mortality and CCD are consequences of the interaction of stressors building up over weeks to months (Goulson et al., 2015; Steinhauer et al., 2018). This requires longitudinal data linking continuous sensor streams to validated outcomes, a criterion that was missing in almost all the included studies. The MSPB and UrBAN datasets (Zhu et al., 2024; Abdollahi et al., 2025) prove the feasibility of such resources, and their moderate survival-prediction performance (AUC 0.73) provides a pragmatic benchmark. Promising directions include the concept of survival analysis, multi-task models integrating precursor detectors, and the fusion of sensor data with inspection, terrain and meteorological data, as shown in the tabular loss research (Insolia et al., 2022; Calovi et al., 2021; García-Vicente et al., 2024).

Methodological challenges

The common assessment methodology, which involves random split of clips or images from a limited number of hives, exaggerates performance due to the shared background noise, microphone positioning, and unique acoustic signatures of the same colony. This type of leakage poses a recognized risk to the reproducibility of machine learning-driven research (Kapoor & Narayanan, 2023). Subsequent research ought to implement hive-level, apiary-level, and temporal divisions, disclose class equilibrium and confidence intervals, and compare results against basic benchmarks. Environmental variables such as precipitation, wind, vehicular noise, and seasonal changes (Abdollahi et al., 2023) need to be accounted for or regulated. Precision alone is inadequate as a measure for infrequent, high-expense occurrences because of its limited sensitivity toward minority loss categories (García-Vicente et al., 2024).

Deployment and practice

Ongoing low-energy acoustic monitoring is technically feasible due to TinyML and edge inference (De Simone et al., 2024; Sad et al., 2025), while smartphone-aided Varroa counting is nearing deployment (Picek et al., 2022; Yániz et al., 2025). The reviewed advancements in model compression, quantization, and on-device inference engines (Abadade et al., 2023; Heydari & Mahmoud, 2025; Somvanshi et al., 2025; Han et al., 2026), as reviewed in extensive TinyML surveys, indicate that the analyzed acoustic and image models may increasingly run in the hive, thus reducing bandwidth usage, latency and dependency on cloud connectivity. For the implementation of this technology, it will be necessary to have strong alarms with low false-positive rates, adaptable to different hive

types, climates and subspecies, affordable and integrated into beekeeping decision-making frameworks. None of these factors were assessed in the included studies. Research is scarce in Africa, South Asia and much of Latin America, which poses an additional challenge, as honeybee subspecies, hive structures, pests and management practices differ markedly in these regions.

Research agenda

For the establishment of solid groundwork for precision beekeeping, subsequent studies must carefully solve various data, modeling, and implementation hurdles. The discipline must first focus on establishing open, longitudinal, multi-apiary datasets that incorporate standardized sensors, stringent inspection protocols, and validated labels for colony loss and collapse. To guarantee generalizability, machine learning frameworks ought to implement hive-independent and cross-region validation as the standard for benchmarking, thoroughly documenting sensitivity, specificity, precision-recall, and calibration metrics. These uniform data streams will enable the creation of multimodal fusion models that can integrate sound, temperature, humidity, weight, visual data, and inspection records to enhance colony survival and early-warning systems. Importantly, the incorporation of explainable machine learning enables researchers to directly correlate predictive signal characteristics with fundamental biological processes, including Varroa–virus loads, brood disruptions, and thermoregulation deficiencies. Ultimately, technological progress must be substantiated through practical applications that assess the utility for practitioners—particularly gaging the effectiveness of alarms, the burden of false alarms, and economic feasibility—while enhancing geographic and genetic diversity by incorporating underrepresented areas and Apis mellifera subspecies, including African honeybees.

Strengths and limitations

The present review is, to our understanding, the inaugural synthesis of machine learning research concerning colony collapse and loss. It implements a transparent PRISMA 2020 methodology and offers completely verifiable record-level information. Numerous constraints are in effect. The identification process depended on systematic academic web searches and citation analysis instead of relying on exports from bibliographic databases like Scopus or Web of Science, which may have resulted in the omission of some qualifying studies, and the PRISMA counts represent scholarly records preserved rather than the total database hits. The processes of screening, extraction, and evaluation were conducted by a sole reviewer, lacking inter-rater consensus. Twelve reports were inaccessible, and a significant portion of the studies was evaluated based on abstracts or incomplete full texts, rendering their quality assessments tentative. The quality checklist was modified instead of being officially validated. Ultimately, the diversity of tasks and metrics hindered meta-analysis.

CONCLUSION

Machine learning can accurately detect multiple immediate factors contributing to honeybee colony collapse—such as Varroa mites, queen death, swarming tendencies, and atypical colony environments—by leveraging acoustic and visual data in conjunction with increasingly cost-effective edge computing technology. The evidence for predicting colony collapse or loss is scarce and moderately effective, with claims of near-perfect accuracy frequently falling short when tested on new hives. The impending advancements will stem from extensive, multi-apiary, multimodal datasets that include verified loss results, rigorous hive-independent validation, clarified models, and empirical trials with beekeepers. These initiatives could convert machine

learning from a simple experimental demonstration into an effective early-warning system for averting colony losses.

Declarations

CRediT authorship contribution statement

Fredrick Michael Ogore: Conceptualization, Methodology, Investigation, Data curation, Formal analysis, Visualization, Writing – original draft, Writing – review & editing.

Declaration of competing interest

The author declares no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Funding

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

Supplementary material

File S1: https://drive.google.com/drive/folders/1gy4J7KFb_yAud6Yj2Zl07CuIQi5PHOcR?usp=drive_link Excel workbook containing the protocol, search log, records and screening log, data-extraction sheet, quality appraisal, PRISMA counts, descriptive synthesis and the background literature cited for context.

References

1. Abadade, Y., Temouden, A., Bamoumen, H., Benamar, N., Chtouki, Y., & Hafid, A. S. (2023). A comprehensive survey on TinyML. *IEEE Access, 11*, 96892–96922. https://doi.org/10.1109/ACCESS.2023.3294111
2. Abdollahi, M., Henry, E., Giovenazzo, P., & Falk, T. H. (2023). The importance of context awareness in acoustics-based automated beehive monitoring. *Applied Sciences, 13*(1), 195. https://doi.org/10.3390/app13010195
3. Abdollahi, M., Zhu, Y., Guimarães, H. R., Coallier, N., Maucourt, S., Giovenazzo, P., & Falk, T. H. (2025). UrBAN: Urban Beehive Acoustics and PheNotyping Dataset. *Scientific Data, 12*. https://doi.org/10.1038/s41597-025-04869-1
4. Barbisan, L., Turvani, G., & Riente, F. (2023). Audio-based identification of queen bee presence inside beehives. In *IEEE CAFE 2023* (pp. 70–74). https://doi.org/10.1109/CAFE58535.2023.10291679
5. Bilik, S., Kratochvila, L., Ligocki, A., Bostik, O., Zemcik, T., Hybl, M., Horak, K., & Zalud, L. (2021). Visual diagnosis of the *Varroa destructor* parasitic mite in honeybees using object detector techniques. *Sensors, 21*(8), 2764. https://doi.org/10.3390/s21082764
6. Bilik, S., Zemcik, T., Kratochvila, L., Ricanek, D., Richter, M., Zambanini, S., & Horak, K. (2024). Machine learning and computer vision techniques in continuous beehive monitoring applications: A survey. *Computers and Electronics in Agriculture, 217*, 108560. https://doi.org/10.1016/j.compag.2023.108560

7. Bjerge, K., Frigaard, C. E., Mikkelsen, P. H., Nielsen, T. H., Misbih, M., & Kryger, P. (2019). A computer vision system to monitor the infestation level of *Varroa destructor* in a honeybee colony. *Computers and Electronics in Agriculture, 164*, 104898. https://doi.org/10.1016/j.compag.2019.104898
8. Braga, A. R., Silva, D. A., Nobre, J. S., Freitas, B. M., & Gomes, D. G. (2019). Defining and predicting honeybee colony health levels through clustering and classification (Definindo e predizendo níveis de saúde de colônias de abelhas via clusterização e classificação). In *SBBD 2019* (pp. 241–246). https://doi.org/10.5753/sbbd.2019.8830
9. Bricout, A., Leleux, P., Acco, P., Escriba, C., Fourniols, J. Y., Soto-Romero, G., & Floquet, R. (2024). Bee Together: joining bee audio datasets for hive extrapolation in AI-based monitoring. *Sensors, 24*(18), 6067. https://doi.org/10.3390/s24186067
10. Calovi, M., Grozinger, C. M., Miller, D. A., & Goslee, S. C. (2021). Summer weather conditions influence winter survival of honey bees (*Apis mellifera*) in the northeastern United States. *Scientific Reports, 11*. https://doi.org/10.1038/s41598-021-81051-8
11. Carvalho, C. M., Jr., Rodrigues, I. L., & Gomes, D. G. (2025). Unsupervised acoustic detection of queenless hives in honeybees (*Apis mellifera ligustica*). In *Brazilian e-Science Workshop (BreSci 2025)* (pp. 49–56). https://doi.org/10.5753/bresci.2025.248141
12. Cejrowski, T., & Szymański, J. (2022). Detection of anomalies in bee colony using transitioning state and contrastive autoencoders. *Computers and Electronics in Agriculture, 200*, 107207. https://doi.org/10.1016/j.compag.2022.107207
13. Chen, S. H., Wang, J. C., Lin, H. J., Lee, M. H., Liu, A. C., Wu, Y. L., Hsu, P. S., Yang, E. C., & Jiang, J. A. (2024). A machine learning-based multiclass classification model for bee colony anomaly identification using an IoT-based audio monitoring system with an edge computing framework. *Expert Systems with Applications, 255*, 124898. https://doi.org/10.1016/j.eswa.2024.124898
14. Cox-Foster, D. L., Conlan, S., Holmes, E. C., Palacios, G., Evans, J. D., Moran, N. A., Quan, P.-L., Briese, T., Hornig, M., Geiser, D. M., Martinson, V., vanEngelsdorp, D., Kalkstein, A. L., Drysdale, A., Hui, J., Zhai, J., Cui, L., Hutchison, S. K., Simons, J. F., … Lipkin, W. I. (2007). A metagenomic survey of microbes in honey bee colony collapse disorder. *Science, 318*(5848), 283–287. https://doi.org/10.1126/science.1146498
15. Davidson, P., Steininger, M., Lautenschlager, F., Kobs, K., Krause, A., & Hotho, A. (2020). Anomaly detection in beehives using deep recurrent autoencoders. In *SENSORNETS 2020* (pp. 142–149). https://doi.org/10.5220/0009161201420149
16. De Simone, A., Barbisan, L., Turvani, G., & Riente, F. (2024). Advancing beekeeping: IoT and TinyML for queen bee monitoring using audio signals. *IEEE Transactions on Instrumentation and Measurement, 73*, 1–9. https://doi.org/10.1109/TIM.2024.3449981
17. Divasón, J., Martinez-de-Pison, F. J., Romero, A., Santolaria, P., & Yániz, J. L. (2023). *Varroa* mite detection using deep learning techniques. In *Hybrid Artificial Intelligent Systems (HAIS 2023), LNCS*.
18. Divasón, J., Romero, A., Martinez-de-Pison, F. J., Casalongue, M., Silvestre, M. A., Santolaria, P., & Yániz, J. L. (2024). Analysis of *Varroa* mite colony infestation level using new open software based on deep learning techniques. *Sensors, 24*(12), 3828. https://doi.org/10.3390/s24123828
19. Duma, Z. S., Zemcik, T., Bilik, S., Sihvonen, T., Honec, P., Reinikainen, S. P., & Horak, K. (2024). *Varroa destructor* detection on honey bees using hyperspectral imagery. *Computers and Electronics in Agriculture, 224*, 109219. https://doi.org/10.1016/j.compag.2024.109219

20. Faramarzi, P., Alimardani, R., Rabbani, H., & Mousazadeh, H. (2024). Acoustic analysis of beehives for precision beekeeping based on the Internet of Things: a case study on improving hive health and productivity. *Iranian Journal of Biosystem Engineering, 55*(3), 49–71. https://doi.org/10.22059/ijbse.2025.383049.665567
21. García-Vicente, E. J., Benito-Murcia, M., Martín Domínguez, M., Pérez Pérez, A., González Sánchez, M., Rey-Casero, I., Alonso Rodríguez, J. M., Barquero-Pérez, Ó., & Risco Pérez, D. (2024). Main causes of producing honey bee colony losses in southwestern Spain: a novel machine learning-based approach. *Apidologie, 55*(5), 67. https://doi.org/10.1007/s13592-024-01108-1
22. Ghezal, A., & König, A. (2025). A comparative study of hybrid machine-learning vs. deep-learning approaches for *Varroa* mite detection and counting. *Sensors, 25*(16), 5075. https://doi.org/10.3390/s25165075
23. Ghezal, A., Peña, C. J. L., & König, A. (2024). *Varroa* mite counting based on hyperspectral imaging. *Sensors, 24*(14), 4437. https://doi.org/10.3390/s24144437
24. Goulson, D., Nicholls, E., Botías, C., & Rotheray, E. L. (2015). Bee declines driven by combined stress from parasites, pesticides, and lack of flowers. *Science, 347*(6229), 1255957. https://doi.org/10.1126/science.1255957
25. Gray, A., Adjlane, N., Arab, A., Ballis, A., Brusbardis, V., Charrière, J.-D., Chlebo, R., Coffey, M. F., Cornelissen, B., Amaro da Costa, C., Dahle, B., Danihlík, J., Dražić, M. M., Evans, G., Fedoriak, M., Forsythe, I., Gajda, A., de Graaf, D. C., Gregorc, A., … Brodschneider, R. (2020). Honey bee colony winter loss rates for 35 countries participating in the COLOSS survey for winter 2018–2019, and the effects of a new queen on the risk of colony winter loss. *Journal of Apicultural Research, 59*(5), 744–751. https://doi.org/10.1080/00218839.2020.1797272
26. Hadjur, H., Ammar, D., & Lefèvre, L. (2022). Toward an intelligent and efficient beehive: A survey of precision beekeeping systems and services. *Computers and Electronics in Agriculture, 192*, 106604. https://doi.org/10.1016/j.compag.2021.106604
27. Hammami, H., & Abdulaziz, N. (2024). BeeBetter: a multi-modal beehive system for honeybee health monitoring and hazard detection. In *ICSPIS 2024, IEEE*. https://doi.org/10.1109/ICSPIS63676.2024.10812634
28. Han, H., Trimi, S., & Lee, S. M. (2026). Tiny machine learning (TinyML): Research trends and future application opportunities. *Array, 29*, 100674. https://doi.org/10.1016/j.array.2025.100674
29. Heydari, S., & Mahmoud, Q. H. (2025). Tiny machine learning and on-device inference: A survey of applications, challenges, and future directions. *Sensors, 25*(10), 3191. https://doi.org/10.3390/s25103191
30. Huang, W., Yang, W., Luo, Z., Qi, J., Zhang, T., & Kong, X. (2024). Sound-based bee colony state analysis using compact MFCC patterns. In *IEEE ISPA 2024* (pp. 2164–2169). https://doi.org/10.1109/ISPA63168.2024.00295
31. Insolia, L., Molinari, R., Rogers, S. R., Williams, G. R., Chiaromonte, F., & Calovi, M. (2022). Honey bee colony loss linked to parasites, pesticides and extreme weather across the United States. *Scientific Reports, 12*. https://doi.org/10.1038/s41598-022-24946-4
32. Iqbal, K., Alabdullah, B., Al Mudawi, N., Algarni, A., Jalal, A., & Park, J. (2024). Empirical analysis of honeybees acoustics as biosensors signals for swarm prediction in beehives. *IEEE Access, 12*, 148405–148421. https://doi.org/10.1109/ACCESS.2024.3471895
33. Kaplan Berkaya, S., Sora Gunal, E., & Gunal, S. (2021). Deep learning-based classification models for beehive monitoring. *Ecological Informatics, 64*, 101353. https://doi.org/10.1016/j.ecoinf.2021.101353

34. 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

35. Kim, J., Oh, J., & Heo, T. Y. (2021). Acoustic scene classification and visualization of beehive sounds using machine learning algorithms and Grad-CAM. *Mathematical Problems in Engineering, 2021*, 5594498. https://doi.org/10.1155/2021/5594498

36. Kiromitis, D. I., Bellos, C. V., Stefanou, K. A., Stergios, G. S., Andrikos, I., Katsantas, T., & Kontogiannis, S. (2022). Performance evaluation of classification algorithms to detect bee swarming events using sound. *Signals, 3*(4), 807–822. https://doi.org/10.3390/signals3040048

37. Klein, A.-M., Vaissière, B. E., Cane, J. H., Steffan-Dewenter, I., Cunningham, S. A., Kremen, C., & Tscharntke, T. (2007). Importance of pollinators in changing landscapes for world crops. *Proceedings of the Royal Society B: Biological Sciences, 274*(1608), 303–313. https://doi.org/10.1098/rspb.2006.3721

38. Kontogiannis, S. (2024). Beehive smart detector device for the detection of critical conditions that utilize edge device computations and deep learning inferences. *Sensors, 24*(16), 5444. https://doi.org/10.3390/s24165444

39. Libal, U., & Biernacki, P. (2024). Drone flight detection at an entrance to a beehive based on audio signals. *Archives of Acoustics, 49*(3), 459–468. https://doi.org/10.24425/aoa.2024.148796

40. Luz, J. S., de Oliveira, M. C., Pereira, F. M., de Araújo, F. H. D., & Magalhães, D. M. V. (2024). Cepstral and deep features for *Apis mellifera* hive strength classification. *Journal of Internet Services and Applications, 15*(1). https://doi.org/10.5753/jisa.2024.4015

41. Meikle, W. G., & Holst, N. (2015). Application of continuous monitoring of honeybee colonies. *Apidologie, 46*(1), 10–22. https://doi.org/10.1007/s13592-014-0298-x

42. Mongan, J., Moy, L., & Kahn, C. E., Jr. (2020). Checklist for Artificial Intelligence in Medical Imaging (CLAIM): A guide for authors and reviewers. *Radiology: Artificial Intelligence, 2*(2), e200029. https://doi.org/10.1148/ryai.2020200029

43. Mrozek, D., Górny, R., Wachowicz, A., & Małysiak-Mrozek, B. (2021). Edge-based detection of varroosis in beehives with IoT devices with embedded and TPU-accelerated machine learning. *Applied Sciences, 11*(22), 11078. https://doi.org/10.3390/app112211078

44. Mustafa, S., Mohaghegh, M., Ardekani, I., & Sarrafzadeh, A. (2025). Leveraging machine learning approaches to decode hive sounds for stress prediction. *IEEE Access*. https://doi.org/10.1109/ACCESS.2025.3599330

45. Neumann, P., & Carreck, N. L. (2010). Honey bee colony losses. *Journal of Apicultural Research, 49*(1), 1–6. https://doi.org/10.3896/IBRA.1.49.1.01

46. Nguyen Ba, N., Pham Thi Kim, P., Tran Thanh, H., Le An, T., Doan Van, T., & Phan, T. T. H. (2024). Monitoring the absence of queen bee in the hive using deep learning and Hilbert Huang Transform. *ASEAN Engineering Journal, 14*(1), 113–120. https://doi.org/10.11113/aej.V14.20163

47. Nolasco, I., Terenzi, A., Cecchi, S., Orcioni, S., Bear, H. L., & Benetos, E. (2019). Audio-based identification of beehive states. In *IEEE ICASSP 2019*. https://doi.org/10.1109/ICASSP.2019.8682981

48. Page, M. J., McKenzie, J. E., Bossuyt, P. M., Boutron, I., Hoffmann, T. C., Mulrow, C. D., Shamseer, L., Tetzlaff, J. M., Akl, E. A., Brennan, S. E., Chou, R., Glanville, J., Grimshaw, J. M., Hróbjartsson, A., Lalu, M. M., Li, T., Loder, E. W., Mayo-Wilson, E., McDonald, S., … Moher, D. (2021). The PRISMA 2020 statement: An updated guideline for reporting systematic reviews. *BMJ, 372*, n71. https://doi.org/10.1136/bmj.n71

49. Picek, L., Novozamsky, A., Frydrychová, R. C., Zitová, B., & Mach, P. (2022). Monitoring of *Varroa* infestation rate in beehives: a simple AI approach. In *IEEE ICIP 2022* (pp. 3341–3345). https://doi.org/10.1109/ICIP46576.2022.9897809
50. Potts, S. G., Biesmeijer, J. C., Kremen, C., Neumann, P., Schweiger, O., & Kunin, W. E. (2010). Global pollinator declines: Trends, impacts and drivers. *Trends in Ecology & Evolution, 25*(6), 345–353. https://doi.org/10.1016/j.tree.2010.01.007
51. Ramsey, M. T., Bencsik, M., Newton, M. I., Reyes, M., Pioz, M., Crauser, D., Delso, N. S., & Le Conte, Y. (2020). The prediction of swarming in honeybee colonies using vibrational spectra. *Scientific Reports, 10*, 9798. https://doi.org/10.1038/s41598-020-66115-5
52. Robles-Guerrero, A., Saucedo-Anaya, T., González-Ramírez, E., & Galván-Tejada, C. E. (2017). Frequency analysis of honey bee buzz for automatic recognition of health status: a preliminary study. *Research in Computing Science, 142*, 89–98.
53. Rodrigues, I. L., de Melo, D. B., da Silva, D. A., Rybarczyk, Y., & Gomes, D. G. (2022). Padrões bioacústicos como identificadores precisos da presença de rainha em colmeias de abelhas melíferas. In *WCAMA 2022* (pp. 11–20). https://doi.org/10.5753/wcama.2022.222913
54. Rodrigues, I. L., de Melo, D. B., Freitas, B. M., & Gomes, D. G. (2021). Detecção de anomalias em padrões acústicos, de temperatura e umidade sazonais para abelhas melíferas (*Apis mellifera* L.). In *Anais do XII Workshop de Computação Aplicada à Gestão do Meio Ambiente e Recursos Naturais (WCAMA 2021)* (pp. 69–78). https://doi.org/10.5753/wcama.2021.15738
55. Rosenkranz, P., Aumeier, P., & Ziegelmann, B. (2010). Biology and control of *Varroa destructor*. *Journal of Invertebrate Pathology, 103*(Suppl. 1), S96–S119. https://doi.org/10.1016/j.jip.2009.07.016
56. Ruvinga, S., Hunter, G. J. A., Duran, O., & Nebel, J. C. (2021). Use of LSTM neural networks to identify 'queenlessness' in honeybee hives from audio signals. In *17th International Conference on Intelligent Environments (IE 2021)*.
57. Ruvinga, S., Hunter, G., Duran, O., & Nebel, J. C. (2023). Identifying queenlessness in honeybee hives from audio signals using machine learning. *Electronics, 12*(7), 1627. https://doi.org/10.3390/electronics12071627
58. Ruvinga, S., Hunter, G., Nebel, J. C., & Duran, O. (2022). Prediction of honeybee swarms using audio signals and convolutional neural networks. In *Workshops at IE2022, IOS Press (AISE)* (pp. 146–154). https://doi.org/10.3233/AISE220032
59. Šabić, J., Perković, T., Šolić, P., & Šerić, L. (2025). Buzzing with intelligence: A systematic review of smart beehive technologies. *Sensors, 25*(17), 5359. https://doi.org/10.3390/s25175359
60. Sad, C., Kampelopoulos, D., Sofianidis, I., Kanelis, D., Nikolaidis, S., Tananaki, C., & Siozios, K. (2025). Deep edge IoT for acoustic detection of queenless beehives. *Electronics, 14*(15), 2959. https://doi.org/10.3390/electronics14152959
61. Schurischuster, S., Remeseiro, B., Radeva, P., & Kampel, M. (2018). A preliminary study of image analysis for parasite detection on honey bees. In *ICIAR 2018, LNCS 10882*. https://doi.org/10.1007/978-3-319-93000-8_52
62. Scutaru, D., Bergonzoli, S., Costa, C., Violino, S., Costa, C., Albertazzi, S., Capano, V., Kostić, M. M., & Scarfone, A. (2025). An AI-based digital scanner for *Varroa destructor* detection in beekeeping. *Insects, 16*(1), 75. https://doi.org/10.3390/insects16010075
63. Senger, D., Kluss, T., & Förster, A. (2023). Towards a warning system for beekeepers: detecting anomalous changes in sensor data from honey bee colonies. In *EnviroInfo 2023, LNI P-342* (pp. 15–24).

https://doi.org/10.18420/env2023-001

64. Sharma, A., Varastehpour, S., Ardekani, I., & Sharifzadeh, H. (2024). Bee disease *Varroa* prediction: utilizing convolutional neural networks with augmentation for robust detection and identification of honeybee infection. In *ICIESTR 2024, IEEE*. https://doi.org/10.1109/ICIESTR60916.2024.10798343
65. Somvanshi, S., Islam, M. M., Chhetri, G., Chakraborty, R., Mimi, M. S., Shuvo, S. A., Islam, K. S., Javed, S. A., Rafat, S. A., Dutta, A., & Das, S. (2025). *From tiny machine learning to tiny deep learning: A survey* (arXiv:2506.18927). arXiv. https://doi.org/10.48550/arXiv.2506.18927
66. Steinhauer, N., Kulhanek, K., Antúnez, K., Human, H., Chantawannakul, P., Chauzat, M.-P., & vanEngelsdorp, D. (2018). Drivers of colony losses. *Current Opinion in Insect Science, 26*, 142–148. https://doi.org/10.1016/j.cois.2018.02.004
67. Sucipto, W., Zhou, J., Kwon, R. S. M., & Chen, F. (2025). *A survey of TinyML applications in beekeeping for hive monitoring and management* (arXiv:2509.08822). arXiv. https://doi.org/10.48550/arXiv.2509.08822
68. Szczurek, A., Maciejewska, M., Bąk, B., Wilk, J., Wilde, J., & Siuda, M. (2020a). Classification of honeybee infestation by *Varroa destructor* using gas sensor array. In *SENSORNETS 2020* (pp. 61–68). https://doi.org/10.5220/0009171100610068
69. Szczurek, A., Maciejewska, M., Bąk, B., Wilk, J., Wilde, J., & Siuda, M. (2020b). Gas sensor array and classifiers as a means of varroosis detection. *Sensors, 20*(1), 117. https://doi.org/10.3390/s20010117
70. Terenzi, A., Ortolani, N., Nolasco, I., Benetos, E., & Cecchi, S. (2022). Comparison of feature extraction methods for sound-based classification of honey bee activity. *IEEE/ACM Transactions on Audio, Speech, and Language Processing, 30*, 112–122. https://doi.org/10.1109/TASLP.2021.3133194
71. Torky, M., Nasr, A. A., & Hassanien, A. E. (2023). Recognizing beehives' health abnormalities based on MobileNet deep learning model. *International Journal of Computational Intelligence Systems, 16*, 135. https://doi.org/10.1007/s44196-023-00311-9
72. vanEngelsdorp, D., Evans, J. D., Saegerman, C., Mullin, C., Haubruge, E., Nguyen, B. K., Frazier, M., Frazier, J., Cox-Foster, D., Chen, Y., Underwood, R., Tarpy, D. R., & Pettis, J. S. (2009). Colony collapse disorder: A descriptive study. *PLoS ONE, 4*(8), e6481. https://doi.org/10.1371/journal.pone.0006481
73. Várkonyi, D. T., Seixas Junior, J. L., & Horváth, T. (2023). Dynamic noise filtering for multi-class classification of beehive audio data. *Expert Systems with Applications, 213*, 118850. https://doi.org/10.1016/j.eswa.2022.118850
74. Voudiotis, G., Moraiti, A., & Kontogiannis, S. (2022). Deep learning beehive monitoring system for early detection of the *Varroa* mite. *Signals, 3*(3), 506–523. https://doi.org/10.3390/signals3030030
75. Wolff, R. F., Moons, K. G. M., Riley, R. D., Whiting, P. F., Westwood, M., Collins, G. S., Reitsma, J. B., Kleijnen, J., & Mallett, S. (2019). PROBAST: A tool to assess the risk of bias and applicability of prediction model studies. *Annals of Internal Medicine, 170*(1), 51–58. https://doi.org/10.7326/M18-1376
76. Yániz, J., Casalongue, M., Martinez-de-Pison, F. J., Silvestre, M. A., Beeguards Consortium, Santolaria, P., & Divasón, J. (2025). An AI-based open-source software for *Varroa* mite fall analysis in honeybee colonies. *Agriculture, 15*(9), 969. https://doi.org/10.3390/agriculture15090969
77. Zacepins, A., Brusbardis, V., Meitalovs, J., & Stalidzans, E. (2015). Challenges in the development of precision beekeeping. *Biosystems Engineering, 130*, 60–71. https://doi.org/10.1016/j.biosystemseng.2014.12.001
78. Zgank, A. (2021). IoT-based bee swarm activity acoustic classification using deep neural networks. *Sensors, 21*(3), 676. https://doi.org/10.3390/s21030676

79. Zhu, Y., Abdollahi, M., Maucourt, S., Coallier, N., Guimarães, H. R., Giovenazzo, P., & Falk, T. H. (2024). MSPB: a longitudinal multi-sensor dataset with phenotypic trait measurements from honey bees. *Scientific Data, 11*, 860. https://doi.org/10.1038/s41597-024-03695-1
80. Zhukov, O. O., & Horbenko, V. I. (2025). Using time series and LSTM neural networks for identification of hive states. *Information Technologies and Mathematical Modelling (ITMM), 2025*(1). https://doi.org/10.34185/1991-7848.itmm.2025.01.098
81. Zhukov, O., & Horbenko, V. (2024). An adaptive approach to the bee hive state identification using neural networks and audio analysis. *System Technologies, 4*(153). https://doi.org/10.34185/1562-9945-4-153-2024-01

Figures

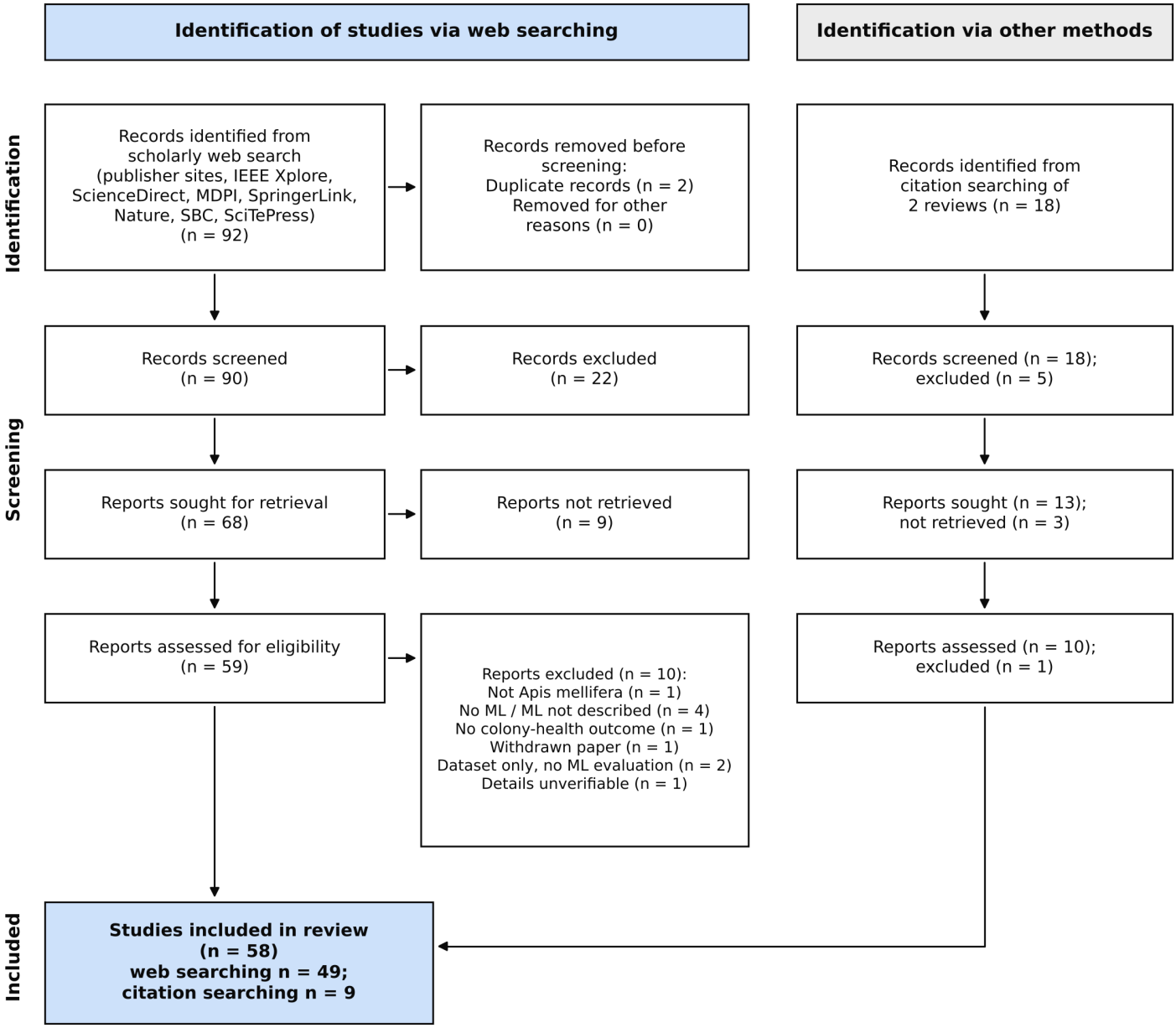


Figure 1

PRISMA 2020 flow diagram of study identification, screening and inclusion. Exclusion reasons at full-text stage are shown for the web-searching stream; the reason for the single full-text exclusion from citation searching is given in Supplementary File S1.

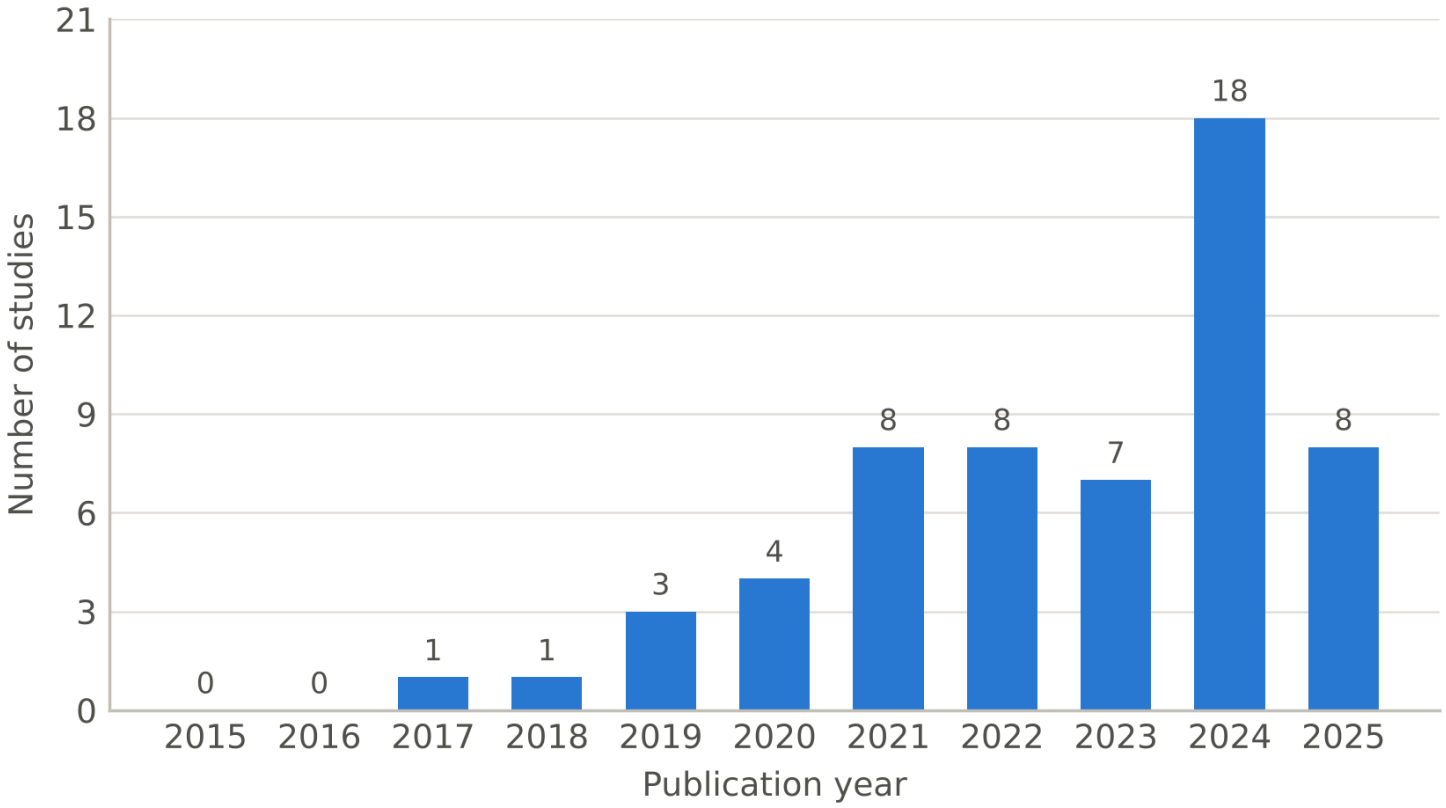


Figure 2

Number of included studies by publication year (n = 58).

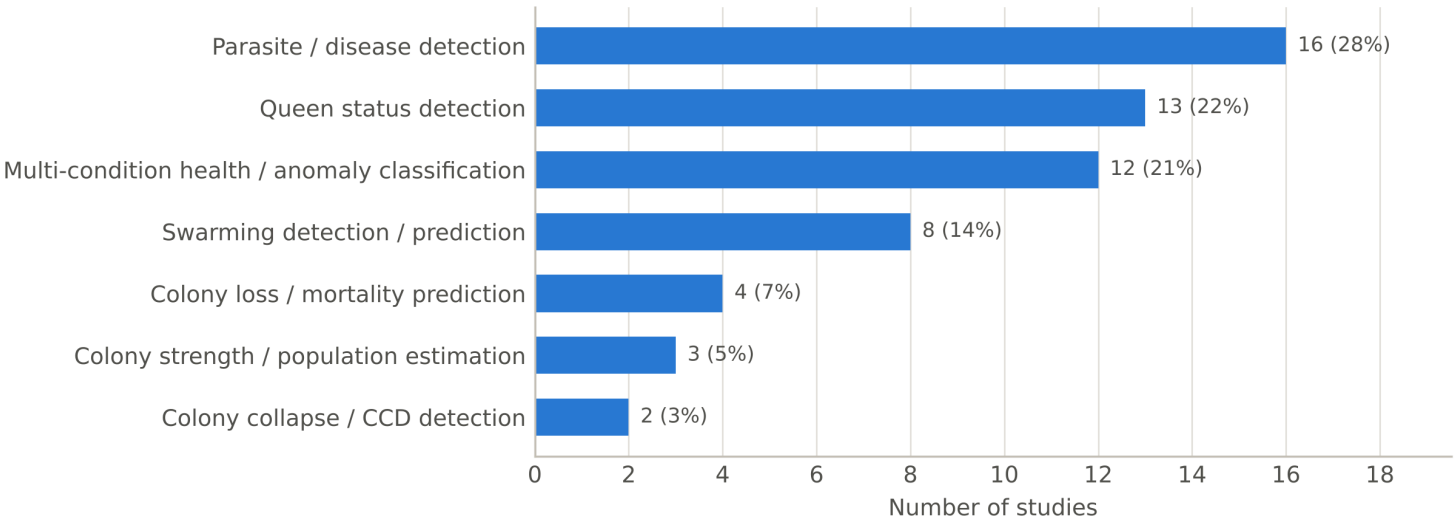


Figure 3

Included studies by detection task. Labels give the number and percentage of studies.

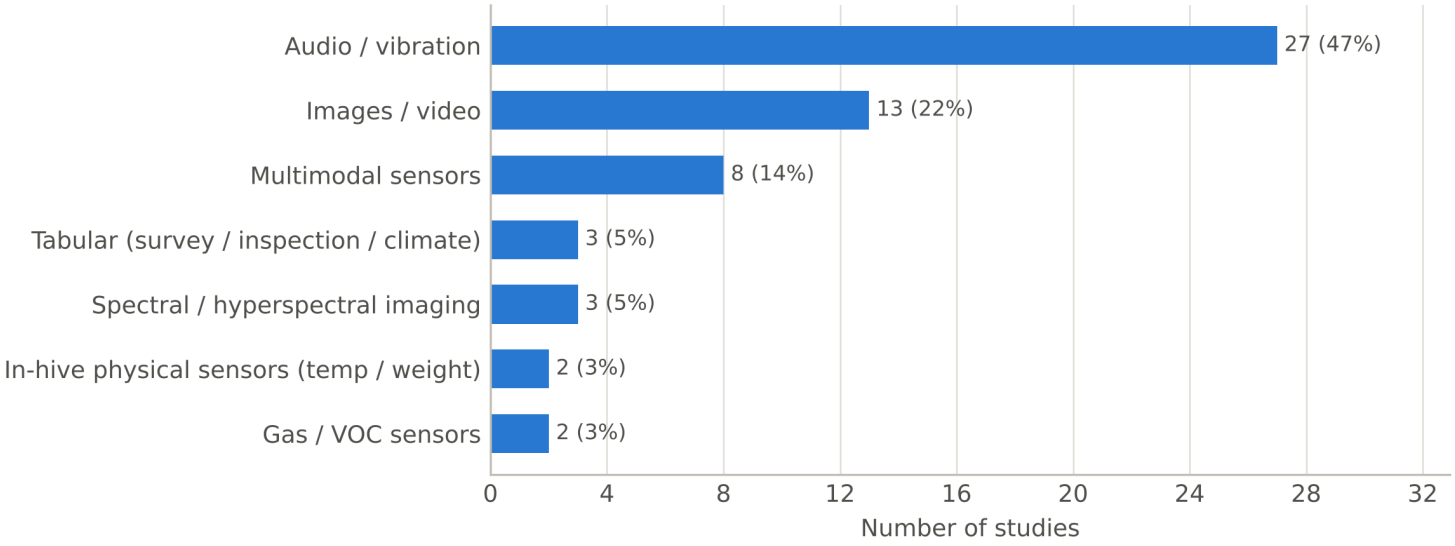


Figure 4

Included studies by data-modality group.

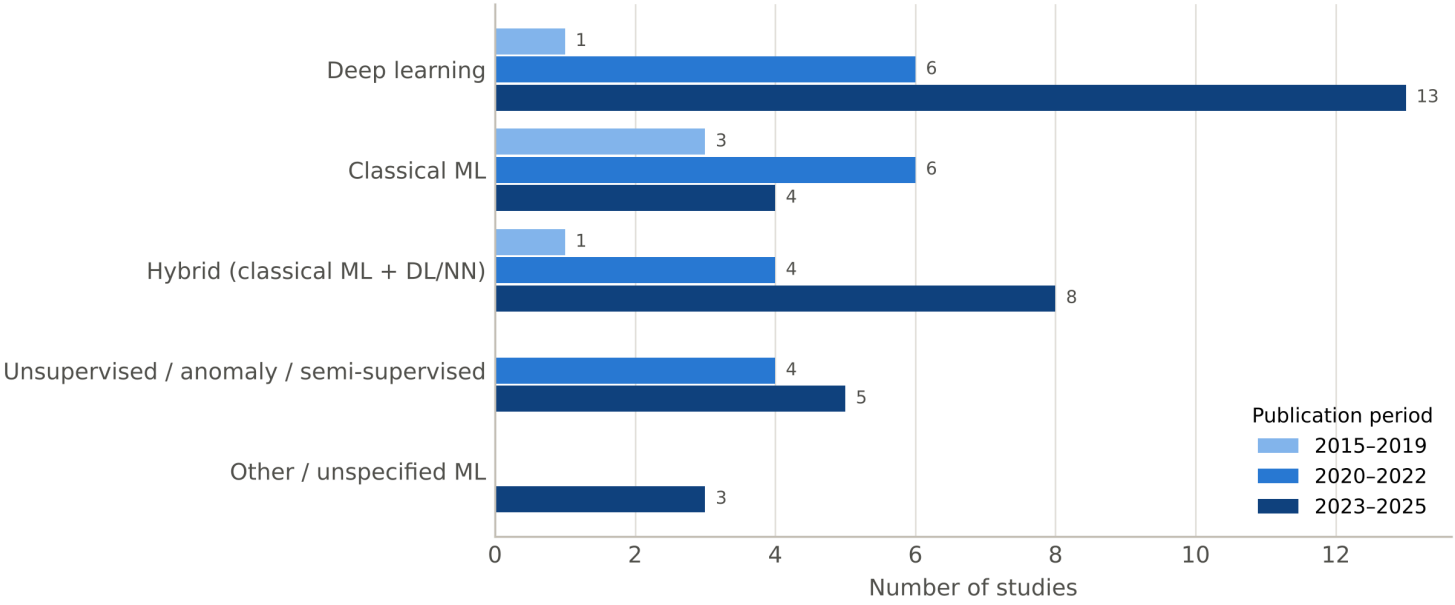


Figure 5

ML model families by publication period.

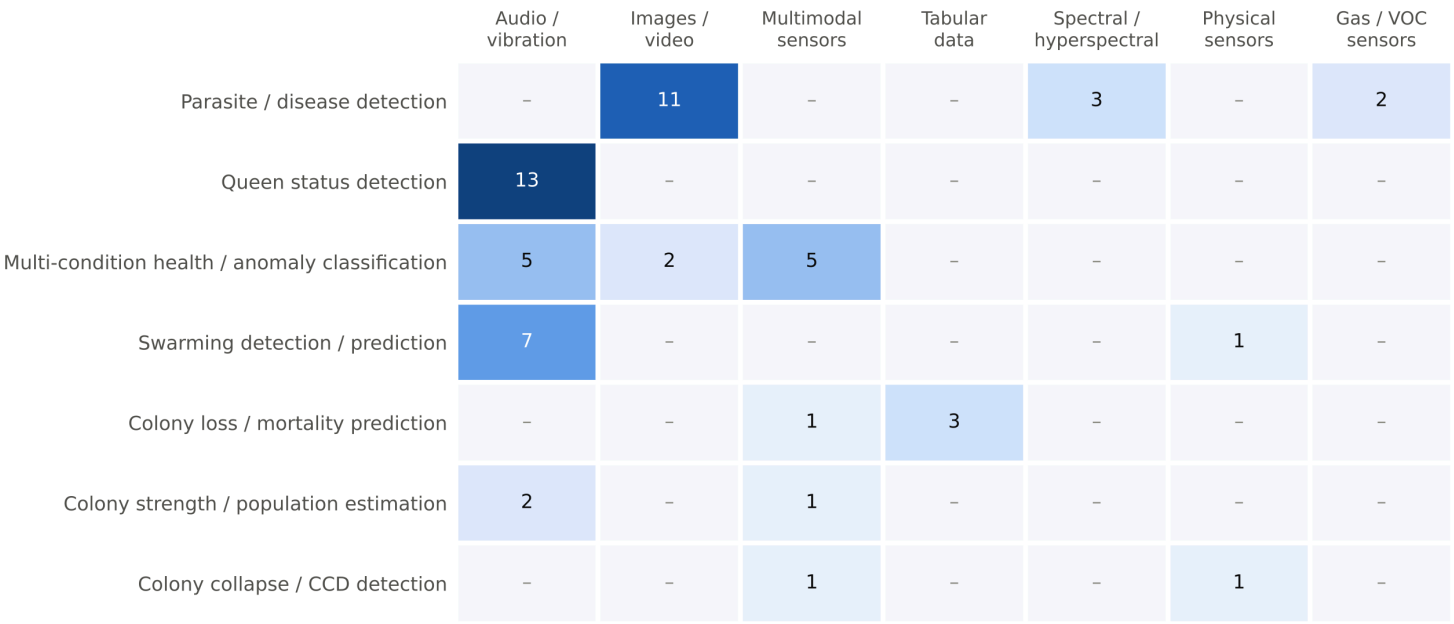


Figure 6

Cross-tabulation of detection task and data modality (number of studies). Darker cells indicate more studies; dashes indicate no studies.

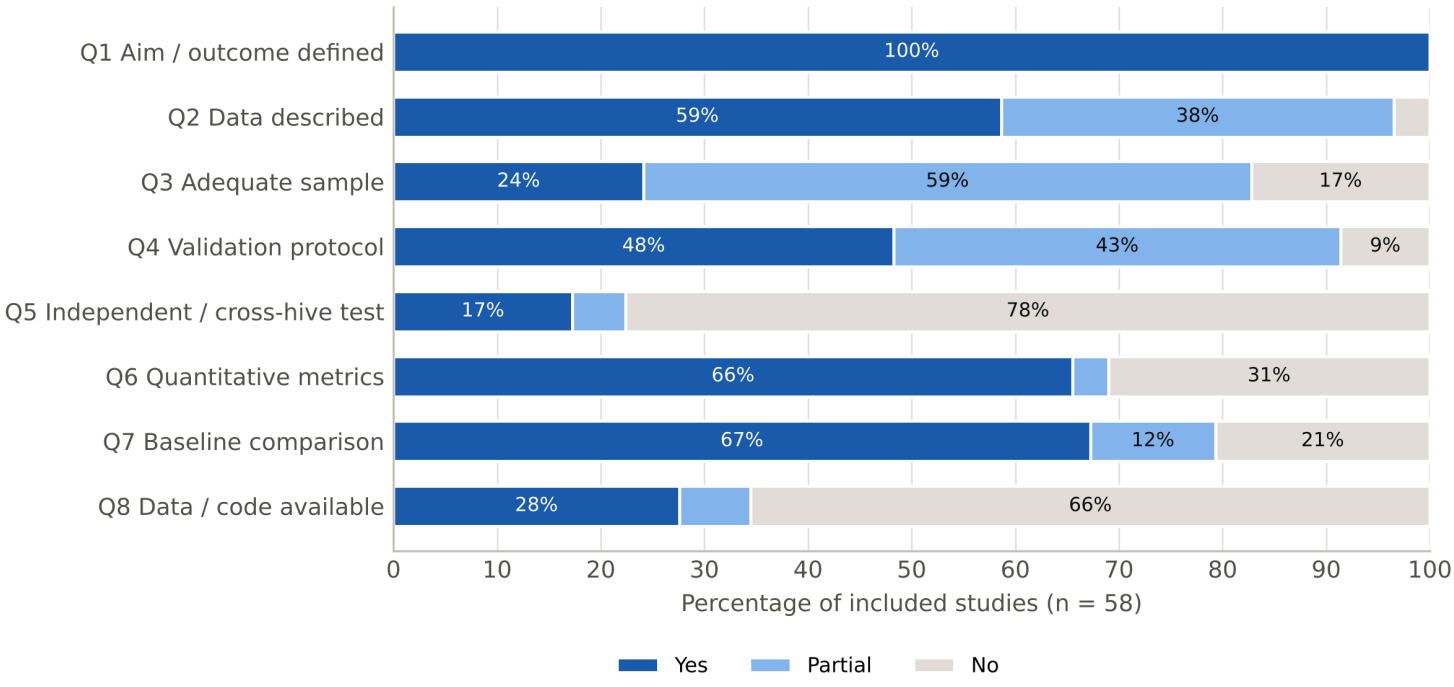


Figure 7

Methodological quality appraisal across eight checklist items (n = 58). Segments give the percentage of studies rated yes, partial or no.