



What Can Artificial Intelligence Reliably Predict in Vector Biology? A Systematic Review of Analytical Tasks, Data Integration, Explainability, and Reproducibility

Michael Roberts¹, Sarah Thompson^{2*}, James Anderson³

¹*Department of Artificial Intelligence and Data Science, Faculty of Engineering, Imperial College London, London, United Kingdom.*

²*Department of Computational Biology and Bioinformatics, Faculty of Science, Nanyang Technological University, Singapore.*

³*Department of Medical Entomology and Spatial Epidemiology, Faculty of Agriculture, University of Queensland, Brisbane, Australia.*

ABSTRACT

Artificial intelligence is increasingly used to classify vectors, automate surveillance, forecast abundance and transmission risk, and inform intervention planning. These applications may strengthen predictive entomology by integrating images, acoustic and optical signals, environmental observations, entomological records, and epidemiological data. However, the reliability of such predictions remains difficult to judge because analytical tasks, biological outcomes, validation designs, and intended decisions are frequently treated as comparable when they represent distinct constructs. This systematic review examines what artificial intelligence can defensibly predict in vector biology and under which biological, spatial, temporal, technical, and operational conditions those predictions remain credible. The review uses prespecified questions and eligibility boundaries, database-specific search logic, staged screening, structured evidence extraction, ecological adaptation of prediction-model bias and applicability appraisal, and narrative synthesis organized by analytical task and evidence class. It integrates evidence on vector identification, automated surveillance, abundance and distribution forecasting, transmission prediction, intervention-relevant classification, decision support, explainability, reproducibility, and external validity. The strongest synthesis is that artificial intelligence can provide useful task-specific predictions within bounded data-generating environments, particularly when targets, sensors, species, scales, and validation contexts are explicitly defined. Reliability weakens when internal performance is generalized to untested sites, devices, seasons, taxa, or decisions. Prediction accuracy does not by itself demonstrate decision usefulness; post-hoc explanation does not establish biological plausibility; and reproducible code does not ensure reproducible performance under distribution shift. Progress therefore depends less on adding model complexity than on independent validation, biologically coherent outcome definitions, leakage-resistant evaluation, transparent uncertainty, interoperable multimodal data, and prospective testing of whether predictions improve surveillance or control decisions.

Keywords: Artificial intelligence in vector biology, Predictive entomology, Machine learning, Computer vision, Automated vector identification, Sensor surveillance.

HOW TO CITE THIS ARTICLE: Roberts M, Thompson S, Anderson J. What Can Artificial Intelligence Reliably Predict in Vector Biology? A Systematic Review of Analytical Tasks, Data Integration, Explainability, and Reproducibility. *Entomol Appl Sci Lett.* 2026;13(1):99-110. <https://doi.org/10.51847/ZXVYL2H9N>

Corresponding author: Sarah Thompson

E-mail ✉ sarah.thompson@ntu.edu.sg

Received: 27/11/2025

Accepted: 11/03/2026

INTRODUCTION

Artificial intelligence (AI) is becoming an analytical layer between vector observations and

biological or operational interpretation. Image classifiers can assign taxonomic labels, acoustic and optical systems can automate detection, and forecasting models can connect surveillance

records with environmental variation. Yet these applications do not constitute one uniform predictive problem. A model identifying an adult mosquito from a curated image addresses a different construct from a model estimating trap abundance, mapping habitat suitability, predicting arbovirus incidence, or comparing intervention scenarios. Machine-learning research in mosquito control spans geospatial, visual, and acoustic pipelines, but evaluation remains heterogeneous and incompletely standardized [1]. The central issue is therefore not whether AI can generate predictions, but what each prediction represents, how it was validated, and whether its stated interpretation exceeds the evidence produced by the underlying task.

This distinction is scientifically important because vector biology links several partially observed processes: vector identity, abundance, spatial distribution, age structure, host contact, infection, intervention exposure, and pathogen transmission. Each process can become an input, target, proxy, or latent assumption within an AI pipeline. AI-based prediction has expanded across vector-borne diseases and pathogens, although disease-, pathogen-, and vector-level outcomes are not interchangeable [2]. A model predicting reported disease incidence may perform well without accurately representing vector abundance, while a classifier detecting infection in mosquitoes does not automatically estimate transmission intensity. Reliability must consequently be evaluated against the biological level at which the output is defined, rather than inferred from a broad label such as “vector-borne disease prediction.”

Time and space further complicate interpretation. Vector and pathogen observations are structured by seasonality, surveillance effort, environmental dependence, human movement, intervention activity, and reporting systems. Randomly partitioned records may therefore place closely related observations in both development and test sets, producing strong internal performance without demonstrating transportability. Temporal and spatiotemporal arbovirus forecasts vary in outcome definition, horizon, and validation design, restricting direct performance comparison [3]. A forecast that reconstructs a familiar seasonal pattern is not equivalent to one

that anticipates an outbreak at a useful lead time, and a model calibrated within one surveillance programme cannot be assumed to retain calibration under altered trap designs, climates, reporting practices, or control regimes.

The same boundary applies to emerging multimodal and digital-twin-oriented representations. Combining drone imagery, remote sensing, sensor streams, entomological observations, and epidemiological data may improve system representation, but additional modalities do not automatically produce biologically valid or decision-ready models. Drone-enabled sensing can extend mosquito surveillance, but ecological ground truth determines whether mapped features are operationally meaningful [4]. This review therefore asks what AI can reliably predict across vector identification, automated surveillance, abundance, distribution, transmission, and intervention planning. Its central argument is that reliability is task-, scale-, species-, sensor-, and context-specific. Internal test performance is not external validity; prediction accuracy is not decision usefulness; post-hoc explainability is not biological plausibility; and computational reproducibility is not evidence that performance will recur under distribution shift.

Review questions and eligibility criteria

The review was structured around ten linked questions concerning what the selected evidence establishes, leaves uncertain, or contradicts across vector identification, automated surveillance, abundance and distribution forecasting, transmission prediction, intervention planning, decision support, explainability, reproducibility, and external validity. Additional questions examined which claims were directly demonstrated rather than conceptually inferred, which biological or methodological conditions explained divergent findings, and which apparently similar measurements represented non-equivalent constructs. Eligibility, searching, selection, and synthesis will be reported using the PRISMA 2020 framework [5]. The review questions were operationalized by specifying the entomological target, biological scale, input modality, prediction output, validation design, intended decision, and principal interpretation boundary for each study.

Eligible evidence comprised peer-reviewed journal articles within the prespecified recent-publication window that directly addressed AI in vector biology or supplied a methodological principle necessary for evaluating a named article domain. Primary prediction studies, field-validation studies, systematic reviews, methodological evaluations, and closely relevant analytical perspectives were eligible when they contained sufficient information to judge sampling, predictors, outcomes, validation, applicability, explanation, or reproducibility. Generic machine-learning studies without vector relevance, non-journal sources, conference-only papers, preprints, software descriptions, vendor material, and records without verifiable bibliographic metadata were excluded. The accompanying explanation and elaboration will guide explicit rationale for each eligibility boundary and any protocol deviation [6]. Evidence was not accepted merely because it contained title-related terminology; each source had to support a planned claim, comparison, limitation, table element, figure relation, or research implication.

Outcome eligibility preserved distinctions among identification, detection, abundance, distribution, transmission, infection or trait classification, intervention prediction, and decision utility. Habitat suitability was not treated as observed vector distribution, predicted abundance was not treated as disease transmission, and intervention-relevant traits were not treated as demonstrated intervention effectiveness. Extraction fields covered taxon and life stage, setting, data provenance, model family, output, temporal and spatial scale, validation partition, external testing, explanation method, reproducibility evidence, distribution-shift risk, decision relevance, and deployment boundary. Methodological appraisal examined sampling-unit dependence, predictor measurement, outcome definition, leakage, model tuning, calibration, uncertainty, external validation, and applicability. Missing information was retained as uncertainty rather than converted into a favourable appraisal, and no quantitative synthesis was planned where tasks, outcomes, scales, or validation designs were not meaningfully comparable.

Search strategy, screening, and study classification

Database searches were organized through intersecting concept blocks rather than disconnected repetitions of the article title. The biological block included mosquitoes, disease vectors, and vector-borne systems; the analytical block included artificial intelligence, machine learning, deep learning, neural networks, computer vision, and related predictive methods; and the task block covered identification, surveillance, abundance, distribution, forecasting, transmission, intervention, decision support, explainability, reproducibility, and external validity. Controlled vocabulary was combined with title, abstract, and keyword terms according to database syntax. Searches were restricted using the prespecified publication and source criteria, while supplementary citation searching was reserved for directly relevant journal articles that satisfied the same eligibility rules. Because outcomes and validation designs are non-comparable, synthesis will follow SWiM reporting rather than forced statistical pooling [7].

Records were to be merged through DOI-first deduplication, followed by normalized comparisons of title, publication year, and authorship where DOI data were incomplete. Title-and-abstract screening was followed by full-text eligibility assessment, with exclusions recorded under predefined categories such as absent vector relevance, ineligible source type, unverified article status, no usable predictive task, or inadequate validation information for the planned claim. Disagreements were resolved through discussion and, where necessary, additional adjudication. Included studies were classified by analytical task, biological target, data modality, model family, prediction scale, validation design, external-validity evidence, explanation, reproducibility, decision relevance, and deployment boundary. Prediction studies will be appraised across sampling units, predictors, outcomes, analysis, and applicability using an ecologically adapted PROBAST logic [8]. Classification was designed to prevent evidence from being aggregated merely because studies used similar algorithms. An image classifier, an abundance forecast, and a transmission-risk model could all use a convolutional or neural architecture while differing fundamentally in inputs, targets, sampling units, validation requirements, and decision consequences.

Primary studies were therefore separated from reviews and methodological critiques, while internal holdout tests, temporal holdouts, spatial holdouts, device transfer, geographic external validation, and field evaluation were recorded as distinct evidence classes. Explanation methods were documented separately from evidence of

biological plausibility, and code or data availability was recorded separately from performance transportability. The review questions, eligibility boundaries, search and screening logic, evidence-classification rules, and bias controls are specified in **Table 1**.

Table 1. Search Strategy, Screening, and Study Classification: Review Questions, Eligibility Boundaries, Search Logic, Screening Rules, Evidence Classification, and Bias Controls

Review-method element	Operational definition	Inclusion rule	Exclusion rule	Search or screening implementation	Evidence-classification rule	Bias-control measure	Reporting requirement
Review question and construct definition	Each question specifies the vector-related task, biological target, output, validation context, and intended use.	Record maps directly to a named review question.	Generic AI evidence without an explicit vector or methodological link.	Assign one primary review-question code and any justified secondary code.	Separate identification, surveillance, abundance, distribution, transmission, intervention, and decision outcomes.	Resolve ambiguous constructs before synthesis.	Report the construct evaluated rather than a generic AI category.
Biological-system boundary	Vector species, life stage, vector–pathogen system, or surveillance system is explicitly identified.	Vector relevance is demonstrable from the article.	Human-only prediction without a vector-related analytical pathway.	Search taxonomic, functional, and vector-borne-system terms.	Record taxon, life stage, pathogen, host, geography, and season.	Avoid transferring findings across untested taxa or ecological settings.	State the biological and geographical scope of every claim.
Outcome boundary	The output is classified by the construct actually predicted.	Identification, detection, abundance, distribution, transmission, trait, intervention, or decision output is evaluated.	No validated predictive output or outcome definition.	Extract metric, scale, forecast horizon, and uncertainty representation.	Do not pool non-equivalent outcomes.	Prevent proxy substitution and outcome conflation.	Report outcome definitions and relevant interpretation limits.
Publication and evidence-source boundary	Peer-reviewed DOI-verifiable journal evidence within the prespecified source window.	Eligible primary, review, or methodological journal article.	Preprint, conference-only record, software page, vendor source, report, or unverifiable article.	Apply source filters and verify bibliographic metadata independently.	Classify article type before using its claims.	Retain a provisional flag until journal and metadata checks are complete.	Document source type and eligibility rationale.
Search concept blocks	Vector terms are intersected with AI terms, task terms, and validity terms.	Search string contains a biological block and an analytical or validity block.	Disconnected title fragments without construct logic.	Adapt controlled vocabulary and field codes to each database.	Tag records by the concept block that retrieved them.	Peer-check syntax and document all amendments.	Report exact database-specific strings, dates, and filters.
Deduplication and screening	Unique records undergo title-and-abstract screening followed by full-text assessment.	Eligibility can be assessed using predefined criteria.	Duplicate or clearly ineligible record.	Resolve DOI matches first, then normalized title, year, and authorship.	Record screening stage and exclusion category.	Preserve reviewer decisions and adjudication outcomes.	Report actual counts only after completion.
Evidence classification	Studies are grouped by task, data, scale, model, validation, explanation, reproducibility,	Sufficient information exists for stable classification.	Evidence cannot support a planned claim or classification.	Use structured extraction fields and one primary analytical task.	Permit secondary tags without double counting the study.	Synthesize heterogeneous evidence without forced pooling.	State grouping logic and metric-handling rules.

	and intended decision.						
Risk of bias and applicability	Appraisal addresses sampling units, predictors, outcomes, analysis, leakage, and transferability.	Methods permit at least an explicit low, high, or unclear domain judgment.	Information is too limited for the intended claim.	Apply an ecologically adapted prediction-model appraisal.	Separate risk of bias from applicability.	Record rationale and retain “unclear” where evidence is absent.	Report domain-level concerns rather than an invented summary score.
Interpretation boundary	Performance, explanation, reproducibility, and usefulness are evaluated as separate properties.	The article supports the specific property being cited.	One property is inferred solely from another.	Extract each validity property in a separate field.	Internal, external, explanatory, reproducibility, and decision evidence remain distinct.	Audit every synthesis statement against its supporting evidence class.	State limitations at the sentence where interpretation changes.

Vector identification and automated surveillance

Vector identification is among the most developed AI tasks in vector biology because images can be standardized, labelled, and processed using established computer-vision architectures. Nevertheless, the apparent clarity of the classification task can conceal important biological and operational assumptions. Convolutional neural networks can classify selected adult mosquito species from images, but closed-class performance does not establish open-set field validity [9]. Curated image sets may underrepresent damaged specimens, unusual orientations, cryptic taxa, non-target insects, immature stages, and field backgrounds. A multitiered ensemble can add novelty detection to mosquito image classification, yet detecting an unknown class is not equivalent to resolving its taxonomy [10]. Novelty scores can support triage by indicating that a specimen differs from known training classes, but they do not establish the identity, epidemiological importance, or taxonomic relationship of the unrecognized specimen.

Acoustic and optical systems extend automated surveillance beyond static morphology by representing wingbeat-related or flight-dependent signals. Budget smartphones can support distributed acoustic monitoring, but microphone variation, ambient noise, and overlapping flight tones constrain taxonomic certainty [11]. Acoustic frequency can vary with temperature, sex, physiological condition, flight behaviour, recording geometry, and hardware, while unrelated insects may generate overlapping signal structures. Backscattered optical signals are machine-learnable for mosquito identification, although algorithm

rankings from one sensing configuration may not transfer to another [12]. Such systems may reduce manual processing and increase observation frequency, but sensor calibration, non-target detection, missingness, maintenance, and changing field conditions become components of the prediction problem. Automated surveillance therefore requires evaluation at both the model level and the complete sensing-system level.

The strongest evidence for generalization comes from studies that move beyond laboratory partitions and test trained systems under field conditions. A laboratory-trained optical system retained useful performance in two field trials for *Aedes* and *Culex* classification, providing bounded external-validation evidence [13]. This result is more informative than an additional random internal split because the field trials introduce environmental noise and operational variation. It nevertheless remains bounded by the tested genera, locations, sensor configuration, and sampling conditions. Automated counts may support downstream abundance forecasting, but only if detection error, taxonomic uncertainty, sensor downtime, and surveillance changes are propagated rather than treated as error-free observations. Internal test performance is therefore evidence about a particular evaluation set, not a general licence for autonomous field identification or surveillance replacement.

Abundance, distribution, and transmission forecasting

Abundance forecasting asks whether environmental observations, recent surveillance records, and temporal structure can anticipate

changes in vector populations at a defined site and horizon. Neural networks can represent nonlinear *Aedes*-abundance dynamics, but forecast value remains conditional on horizon, calibration, and surveillance continuity [14]. This evidence supports the feasibility of learning recurrent seasonal and environmental relationships; it does not show that the same relationship will persist when trap design, control activity, climate, urban form, or reporting practices change. Earth-observation variables can support mosquito-abundance prediction, while their ecological meaning remains dependent on spatial scale and local calibration [15]. Vegetation, surface water, temperature, and built-environment measurements may represent vector habitat indirectly, but remote-sensing variables are not themselves mosquito observations. Their predictive contribution may arise through genuine ecological relationships, through correlation with unmeasured social or surveillance processes, or through both.

Distribution forecasting introduces a related but distinct representational problem. A habitat-suitability surface, a probability of occurrence, an estimated population density, and a forecast of trap abundance describe different outcomes. Remote-sensing and weather variables can model dengue-vector populations, but geographic transfer cannot be assumed from retrospective fit [16]. Spatial autocorrelation can make neighbouring development and test locations appear independent while preserving nearly identical environmental conditions. A model may therefore interpolate effectively within a sampled landscape yet fail when transferred to a different ecological zone. Near-term relative-abundance forecasts may improve operational awareness, although relative abundance is neither absolute burden nor an intervention outcome [17]. Operational relevance depends on whether the forecast is calibrated, timely, available at a controllable spatial scale, and demonstrably more useful than seasonal expectations or conventional surveillance alone.

Transmission forecasting is more complex because disease outcomes emerge from interacting vector, pathogen, host, environmental, behavioural, intervention, and reporting processes. A model can predict cases without representing vector mechanisms

correctly, particularly when recent incidence dominates the forecast. Conversely, an informative abundance forecast may add little to disease prediction where pathogen introduction, immunity, human mobility, or reporting delay is the principal constraint. Hybrid dynamic and neural models can generate mechanistic hypotheses about meteorological effects, but model-derived mechanisms remain inferential [18]. Mechanistic structure can improve transparency by forcing assumptions about population processes to be stated, yet it does not transform observational associations into experimental evidence. Reliable transmission forecasting therefore requires explicit outcome definitions, prespecified horizons, temporally and spatially appropriate validation, uncertainty calibration, simple comparator models, and assessment of the incremental information contributed by vector data.

Intervention prediction and decision support

Intervention-relevant prediction often begins with estimating biological traits that conventional surveillance measures slowly or incompletely. Spectroscopy and supervised learning can classify vertebrate blood meals, but host-source prediction does not directly quantify transmission contribution [19]. Blood-meal classification can describe contact patterns and support investigation of residual transmission, yet its decision relevance depends on sampling design, digestion stage, host availability, and whether identified feeding patterns indicate an actionable intervention pathway. Diffuse reflectance spectroscopy may predict mosquito age, species, and cuticular resistance under laboratory conditions, but field operational validity remains unestablished [20]. In particular, a spectral distinction associated with a resistant strain may reflect cuticle composition, strain background, rearing conditions, or correlated physiology. It should not be interpreted as a universally validated resistance mechanism without independent populations and field testing.

Prediction of mosquito age and infection status could improve estimation of epidemiologically important population structure. Mass spectrometry with deep learning can estimate the age of field *Anopheles* mosquitoes, while transfer across sites, populations, and

instruments requires separate validation [21]. An age estimate may contribute to surveillance because older mosquitoes have had more opportunity to acquire and transmit pathogens, but predicted age is not a direct measurement of infectiousness or intervention impact. Mid-infrared spectroscopy can predict malaria infection status in field-collected mosquitoes, but infection classification is not equivalent to intervention-effect prediction [22]. The evidence supports a potentially faster surveillance measurement under the evaluated species, location, preparation, instrument, and reference assays. It does not establish that the same model will remain calibrated across parasite prevalence, vector species, instruments, collection methods, or national control programmes.

Decision support extends beyond predicting a state to selecting or prioritizing an action. Spatial models can compare population-replacement

scenarios at continental scale, although scenario success is not observed operational effectiveness [23]. Such models are valuable for exposing assumptions about dispersal, fitness, release location, landscape connectivity, and programme coverage, but their outputs remain conditional on those assumptions. Across the evidence, an end-to-end decision pathway can be expressed as multimodal observation, task-specific biological representation, prediction with uncertainty, contextual qualification, and a bounded surveillance or intervention decision. A component should enter this pathway only when its target, validation context, failure modes, uncertainty, and expected contribution to a decision are explicit. **Figure 1** classifies artificial-intelligence applications across vector surveillance, identification, forecasting, and intervention planning within the analytical logic developed in this section.

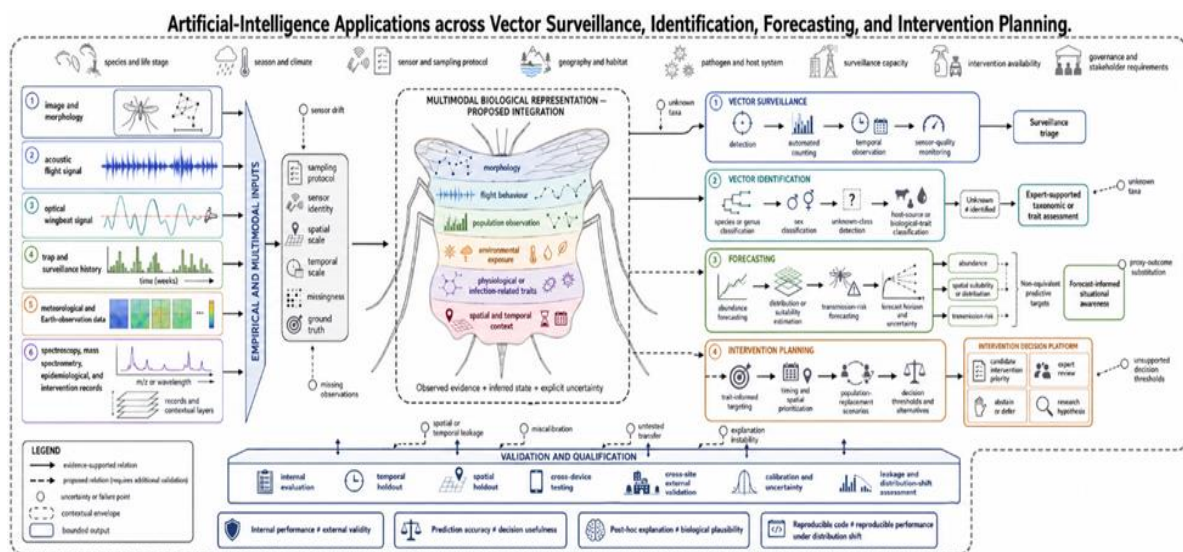


Figure 1. Artificial-intelligence applications across vector surveillance, identification, forecasting, and intervention planning

Alt text

A structured conceptual diagram that classifies artificial-intelligence applications across vector surveillance, identification, forecasting, and intervention planning, with labelled components, directional relations, contextual modifiers, uncertainty points, and a clear boundary between observed evidence and proposed synthesis.

Explainability, reproducibility, and external validity

Explainability should clarify how a model generates an output, which information it relies upon, and where its interpretation may fail. For high-stakes vector-control decisions, post-hoc explanation should not automatically substitute for an intrinsically interpretable model [24]. Feature attribution, saliency, or variable importance may reveal model sensitivity, but these methods do not establish that the represented feature is biologically causal, stable across datasets, or appropriate for intervention. An explanation can also be persuasive while the

underlying performance estimate is biased. Data leakage can reproduce apparently strong machine-learning results while invalidating the claimed generalization [25]. Leakage may occur when information from the evaluation set influences preprocessing, feature selection, imputation, model tuning, or representation learning, or when dependent observations are separated across nominally independent partitions.

External validity requires testing whether performance survives a change in the data-generating environment relevant to intended use. Ecological dependence in space and time makes random data splitting particularly vulnerable to leakage and optimistic performance estimates [26]. Useful external tests should specify what changed: site, season, climate, sensor, taxonomic composition, collection protocol, prevalence, intervention regime, or surveillance system. Ecological machine learning requires scale-aware validation and a clear separation between predictive performance, interpretation, and causal inference [27]. A model may discriminate among classes after transfer but become poorly calibrated; another may preserve average error while failing for a rare species or high-risk season. External validity is therefore multidimensional rather than a binary property, and the claimed generalization envelope should be no broader than the shifts actually evaluated. Reproducibility concerns both whether an analysis can be repeated and whether its performance recurs under relevant change. Data, code, preprocessing instructions, random seeds, software environments, trained parameters, and evaluation definitions can make a computational pipeline auditable. They cannot ensure that a model remains reliable when new observations differ from the released dataset. Post-hoc model explanation becomes biological insight only when it is linked to the data-generating process and independently testable evidence [28]. The same principle applies to reproducibility: repeating a pipeline may faithfully reproduce leakage, confounding, or an overly narrow validation design. Explainability, computational reproducibility, performance reproducibility, external validity, and decision usefulness must therefore be reported as distinct evidence dimensions.

Systematic evidence synthesis

Across vector identification and automated surveillance, the evidence converges on a bounded conclusion: AI can extract useful information from images, acoustic recordings, optical signals, and other sensor data when the target classes and acquisition conditions are sufficiently represented. Deep learning has enabled ecological classification across images, sounds, and remote-sensing observations, while data representativeness and transfer remain persistent constraints [29]. Context-dependent findings arise because curated specimen datasets and controlled sensors simplify variability that becomes consequential in field settings. Identification performance is therefore most defensible for specified taxa, life stages, instruments, and environments. Evidence remains weaker for open-set recognition, cryptic taxa, damaged specimens, new devices, and autonomous replacement of expert surveillance. The forecasting evidence also converges on the ability of machine-learning and hybrid models to represent nonlinear associations across surveillance, environmental, and temporal data. It diverges in the meaning of the predicted outcome and in the strength of validation. Wildlife-monitoring experience shows that robust ecological AI depends on multimodal data, human oversight, deployment engineering, and external evaluation [30]. Digital surveillance can connect climate information with outbreak prediction and control, but technological novelty alone does not establish improved public-health outcomes [31]. Abundance forecasts, distribution maps, and transmission-risk estimates should therefore not be interpreted as interchangeable evidence. Forecast value depends on scale, horizon, calibration, observation continuity, comparator performance, and whether the output is delivered early and locally enough to influence an available action.

Intervention-related evidence is more heterogeneous. Spectroscopic and mass-spectrometric models can estimate traits such as feeding source, age, infection, species, or resistance-associated phenotype, while spatial models can compare intervention scenarios. These outputs may reduce measurement burden or structure strategic reasoning, but few studies demonstrate that model-guided action improves entomological or epidemiological outcomes

relative to current practice. Harmonizing entomological, environmental, epidemiological, and intervention data is necessary for risk management, yet fusion cannot erase incompatible measurement scales [32]. Multimodal integration is therefore valuable when provenance, timing, resolution, missingness, and uncertainty remain visible. It becomes misleading when distinct proxies are fused into an apparently comprehensive representation without evidence that they measure compatible biological processes.

The strongest cross-domain inference is that reliability is not an attribute of an algorithm in isolation. It is a property of a defined task, data-generating system, validation design, interpretation, and intended decision. Horizon-scanning evidence identifies foundation models and digital twins as promising ecological tools, but these remain prospective rather than validated vector-biology frameworks [33].

Research gaps and future directions

The first priority is to move evaluation from isolated model performance toward complete prediction-to-decision pathways. Decision usefulness must ultimately be judged against vector-control and disease-reduction objectives rather than predictive accuracy alone [34]. Progress would be demonstrated by prospective studies that predefine the intended action, comparator practice, forecast horizon, decision threshold, uncertainty response, implementation constraints, and outcome of interest. External validation should deliberately vary site, season, device, taxonomic composition, surveillance protocol, and intervention context instead of treating one independent sample as universal evidence. Satellite observations can expand environmental coverage for malaria risk assessment, but remotely sensed proxies require ecological calibration and scale matching [35]. Future forecasting studies should therefore report which biological construct each remote variable represents and whether it adds information beyond recent surveillance and simple baselines.

The second priority is biologically grounded multimodal representation. Future multimodal systems should link remote sensing to ecological and evolutionary processes rather than treating additional modalities as automatically

informative [36]. A credible vector digital twin would need explicit observation models for sensor error, state variables for biologically meaningful processes, temporal updating, uncertainty propagation, and validation of simulated responses to environmental or intervention change. Explanations should be designed for distinct stakeholders—including entomologists, modellers, control operators, affected communities, and auditors—and should identify what the system observed, inferred, assumed, and could not resolve. Stakeholder-specific explainable AI can improve transparency, although an explanation output is not evidence of biological plausibility [37]. Progress would be shown by stable explanations that generate testable hypotheses, survive external datasets, and improve appropriate human review without producing automation bias.

The third priority is governance of data, models, and failure. Multimodal language models may support integrated biodiversity representation, but their use as vector digital twins requires benchmarks, grounding, and external validation that do not yet exist [38]. Research should establish expert-verified open benchmarks that include difficult field specimens, non-target insects, unknown classes, multiple sensors, missing modalities, and explicit distribution shifts. Dataset and model documentation should record provenance, taxonomic authority, sampling effort, preprocessing, exclusions, licensing, uncertainty, and known failure conditions. Systems should support abstention, expert referral, version control, recalibration, monitoring, and withdrawal when the operating environment exceeds the validated domain. Advancement should be measured not by larger architectures alone, but by reproducible external performance, transparent limitations, biologically meaningful representations, equitable access, and demonstrated improvement in surveillance or intervention decisions.

CONCLUSION

Artificial intelligence can reliably predict selected outcomes in vector biology when the analytical task, biological target, input data, temporal and spatial scale, validation context,

and intended use are tightly specified. The evidence is strongest for bounded identification, automated measurement, and forecasting tasks conducted within represented taxa, sensors, locations, and surveillance systems. It is weaker for transfer to new ecological or technical environments and weakest where prediction is assumed to establish biological mechanism, intervention effectiveness, or decision benefit without direct evaluation. Reliability must therefore be treated as a multidimensional claim comprising construct validity, leakage-resistant internal evaluation, external validity, calibration, interpretability, computational reproducibility, performance under distribution shift, and decision relevance. The highest-priority implication is to replace model-centred evaluation with prospective, biologically grounded, and decision-centred validation of the complete sensing–prediction–action pathway. Under this standard, AI is neither dismissed as unreliable nor accepted on the basis of internal accuracy alone; its claims are restricted to the evidence actually demonstrated and expanded only through transparent, independent, and context-specific testing.

ACKNOWLEDGMENTS: None

CONFLICT OF INTEREST: None

FINANCIAL SUPPORT: None

ETHICS STATEMENT: None

REFERENCES

- Joshi A, Miller C. Review of machine learning techniques for mosquito control in urban environments. *Ecol Inform.* 2021;61:101241. doi:10.1016/j.ecoinf.2021.101241
- Kaur I, Sandhu AK, Kumar Y. Artificial intelligence techniques for predictive modeling of vector-borne diseases and its pathogens: A systematic review. *Arch Comput Methods Eng.* 2022;29(6):3741-71. doi:10.1007/s11831-022-09724-9
- de Lima CL, da Silva ACG, Moreno GMM, Cordeiro da Silva C, Musah A, Aldosery A, et al. Temporal and spatiotemporal arboviruses forecasting by machine learning: A systematic review. *Front Public Health.* 2022;10:900077. doi:10.3389/fpubh.2022.900077
- Carrasco-Escobar G, Moreno M, Fornace K, Herrera-Varela M, Manrique E, Conn JE. The use of drones for mosquito surveillance and control. *Parasit Vectors.* 2022;15(1):473. doi:10.1186/s13071-022-05580-5
- Page MJ, McKenzie JE, Bossuyt PM, Boutron I, Hoffmann TC, Mulrow CD, et al. The PRISMA 2020 statement: An updated guideline for reporting systematic reviews. *BMJ.* 2021;372. doi:10.1136/bmj.n71
- Page MJ, Moher D, Bossuyt PM, Boutron I, Hoffmann TC, Mulrow CD, et al. PRISMA 2020 explanation and elaboration: Updated guidance and exemplars for reporting systematic reviews. *BMJ.* 2021;372. doi:10.1136/bmj.n160
- Campbell M, McKenzie JE, Sowden A, Katikireddi SV, Brennan SE, Ellis S, et al. Synthesis without meta-analysis (SWiM) in systematic reviews: Reporting guideline. *BMJ.* 2020;368. doi:10.1136/bmj.l6890
- Wolff RF, Moons KGM, Riley RD, Whiting PF, Westwood M, Collins GS, et al. PROBAST: A tool to assess the risk of bias and applicability of prediction model studies. *Ann Intern Med.* 2019;170(1):51-8. doi:10.7326/M18-1376
- Motta D, Santos AAB, Winkler I, Machado BAS, Pereira DADI, Cavalcanti AM, et al. Application of convolutional neural networks for classification of adult mosquitoes in the field. *PLoS One.* 2019;14(1). doi:10.1371/journal.pone.0210829
- Goodwin A, Padmanabhan S, Hira S, Glancey M, Slinowsky M, Immidiseti R, et al. Mosquito species identification using convolutional neural networks with a multitiered ensemble model for novel species detection. *Sci Rep.* 2021;11(1):13656. doi:10.1038/s41598-021-92891-9
- Sinka ME, Zilli D, Li Y, Kiskin I, Kirkham D, Rafique W, et al. HumBug—An acoustic mosquito monitoring tool for use on budget smartphones. *Methods Ecol Evol.* 2021;12(10):1848-59. doi:10.1111/2041-210X.13663
- Genoud AP, Gao Y, Williams GM, Thomas BP. A comparison of supervised machine

- learning algorithms for mosquito identification from backscattered optical signals. *Ecol Inform.* 2020;58:101090. doi:10.1016/j.ecoinf.2020.101090
13. González-Pérez MI, Faulhaber B, Aranda C, Williams M, Villalonga P, Silva M, et al. Field evaluation of an automated mosquito surveillance system which classifies *Aedes* and *Culex* mosquitoes by genus and sex. *Parasit Vectors.* 2024;17(1):97. doi:10.1186/s13071-024-06177-w
 14. Kinney AC, Current S, Lega J. *Aedes*-AI: Neural network models of mosquito abundance. *PLoS Comput Biol.* 2021;17(11). doi:10.1371/journal.pcbi.1009467
 15. Tsantalidou A, Parselia E, Arvanitakis G, Kyratzi K, Gewehr S, Vakali A, et al. MAMOTH: An Earth observational data-driven model for mosquitoes abundance prediction. *Remote Sens (Basel).* 2021;13(13):2557. doi:10.3390/rs13132557
 16. Scavuzzo JM, Trucco F, Espinosa M, Tauro CB, Abril M, Scavuzzo CM, et al. Modeling dengue vector population using remotely sensed data and machine learning. *Acta Trop.* 2018;185:167-75. doi:10.1016/j.actatropica.2018.05.003
 17. Ventura PC, Kummer AG, Wilke ABB, Chitturi J, Hill MD, Vasquez C, et al. Forecasting the relative abundance of *Aedes* vector populations to enhance situational awareness for mosquito control operations. *PLoS Negl Trop Dis.* 2024;18(11). doi:10.1371/journal.pntd.0012671
 18. Zhang M, Wang X, Tang S. Integrating dynamic models and neural networks to discover the mechanism of meteorological factors on *Aedes* population. *PLoS Comput Biol.* 2024;20(9). doi:10.1371/journal.pcbi.1012499
 19. Mwanga EP, Mapua SA, Siria DJ, Ngowo HS, Nangacha F, Mgando J, et al. Using mid-infrared spectroscopy and supervised machine-learning to identify vertebrate blood meals in the malaria vector, *Anopheles arabiensis*. *Malar J.* 2019;18(1):187. doi:10.1186/s12936-019-2822-y
 20. Pazmiño-Betancourth M, Ochoa-Gutiérrez V, Ferguson HM, González-Jiménez M, Wynne K, Baldini F, et al. Evaluation of diffuse reflectance spectroscopy for predicting age, species, and cuticular resistance of *Anopheles gambiae sensu lato* under laboratory conditions. *Sci Rep.* 2023;13(1):18499. doi:10.1038/s41598-023-45696-x
 21. Mohammad N, Naudion P, Dia AK, Boëlle PY, Konaté A, Konaté L, et al. Predicting the age of field *Anopheles* mosquitoes using mass spectrometry and deep learning. *Sci Adv.* 2024;10(19). doi:10.1126/sciadv.adj6990
 22. Mwanga EP, Kweyamba PA, Siria DJ, Mshani IH, Mchola IS, Makala FE, et al. Reagent-free detection of *Plasmodium falciparum* malaria infections in field-collected mosquitoes using mid-infrared spectroscopy and machine learning. *Sci Rep.* 2024;14(1):12100. doi:10.1038/s41598-024-63082-z
 23. Beeton NJ, Wilkins A, Ickowicz A, Hayes KR, Hosack GR. Spatial modelling for population replacement of mosquito vectors at continental scale. *PLoS Comput Biol.* 2022;18(6). doi:10.1371/journal.pcbi.1009526
 24. Rudin C. Stop explaining black box machine learning models for high stakes decisions and use interpretable models instead. *Nat Mach Intell.* 2019;1(5):206-15. doi:10.1038/s42256-019-0048-x
 25. Kapoor S, Narayanan A. Leakage and the reproducibility crisis in machine-learning-based science. *Patterns (N Y).* 2023;4(9):100804. doi:10.1016/j.patter.2023.100804
 26. Stock A, Gregor EJ, Chan KMA. Data leakage jeopardizes ecological applications of machine learning. *Nat Ecol Evol.* 2023;7(11):1743-5. doi:10.1038/s41559-023-02162-1
 27. Pichler M, Hartig F. Machine learning and deep learning—A review for ecologists. *Methods Ecol Evol.* 2023;14(4):994-1016. doi:10.1111/2041-210X.14061
 28. Yu Q, Ji W, Prihodko L, Ross CW, Anchang JY, Hanan NP. Study becomes insight: Ecological learning from machine learning. *Methods Ecol Evol.* 2021;12(11):2117-28. doi:10.1111/2041-210X.13686
 29. Christin S, Hervet E, Lecomte N. Applications for deep learning in ecology. *Methods Ecol Evol.* 2019;10(10):1632-44. doi:10.1111/2041-210X.13256

30. Tuia D, Kellenberger B, Beery S, Costelloe BR, Zuffi S, Risse B, et al. Perspectives in machine learning for wildlife conservation. *Nat Commun.* 2022;13(1):792. doi:10.1038/s41467-022-27980-y
31. Pley C, Evans M, Lowe R, Montgomery H, Yacoub S. Digital and technological innovation in vector-borne disease surveillance to predict, detect, and control climate-driven outbreaks. *Lancet Planet Health.* 2021;5(10). doi:10.1016/S2542-5196(21)00141-8
32. Lowe R, Torres Codeço C. Harmonizing multisource data to inform vector-borne disease risk management strategies. *Annu Rev Entomol.* 2025;70:337-58. doi:10.1146/annurev-ento-040124-015101
33. Reynolds SA, Beery S, Burgess N, Burgman M, Butchart SHM, Cooke SJ, et al. The potential for AI to revolutionize conservation: A horizon scan. *Trends Ecol Evol.* 2025;40(2):191-207. doi:10.1016/j.tree.2024.11.013
34. Wilson AL, Courtenay O, Kelly-Hope LA, Scott TW, Takken W, Torr SJ, et al. The importance of vector control for the control and elimination of vector-borne diseases. *PLoS Negl Trop Dis.* 2020;14(1). doi:10.1371/journal.pntd.0007831
35. Wimberly MC, de Beurs KM, Loboda TV, Pan WK. Satellite observations and malaria: New opportunities for research and applications. *Trends Parasitol.* 2021;37(6):525-37. doi:10.1016/j.pt.2021.03.003
36. Cavender-Bares J, Schneider FD, Santos MJ, Armstrong A, Carnaval AC, Dahlin KM, et al. Integrating remote sensing with ecology and evolution to advance biodiversity conservation. *Nat Ecol Evol.* 2022;6(5):506-19. doi:10.1038/s41559-022-01702-5
37. Barredo Arrieta A, Díaz-Rodríguez N, Del Ser J, Bannetot A, Tabik S, Barbado A, et al. Explainable artificial intelligence (XAI): Concepts, taxonomies, opportunities and challenges toward responsible AI. *Inf Fusion.* 2020;58:82-115. doi:10.1016/j.inffus.2019.12.012
38. Miao Z, Zhang Y, Fabian Z, Hernandez Celis A, Beery S, Li C, et al. New frontiers in artificial intelligence for biodiversity research and conservation with multimodal language models. *Methods Ecol Evol.* 2026;17(2):238-56. doi:10.1111/2041-210X.70120