Green and Sustainable Environments

Green and Sustainable Environments **1** (2026) 06
https://doi.org/10.62755/gsenv.2026.06
journal homepage: pubs.greensus.org/gsenv

Check for updates

REVIEW PAPERS

A Decade of Machine Learning for Adsorption of Emerging Contaminants: A Systematic Review of Algorithms, Performance, and Methodological Transparency

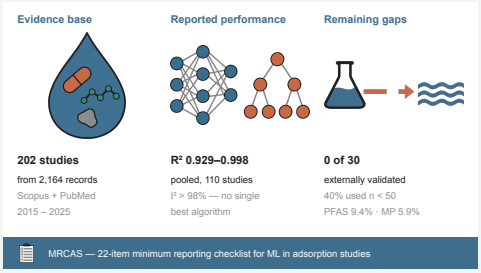
Tarmizi Taher[1,2,*], Sephia Amanda Muhtar[3], Dian Ahmad Hapidin[4], Aditya Rianjanu[2,5], Khairurrijal Khairurrijal[2,4,*]

[1] Department of Environmental Engineering, Faculty of Infrastructure and Regional Technology, Institut Teknologi Sumatera, Terusan Ryacudu, Way Hui, Jati Agung, Lampung Selatan 35365, Indonesia
[2] Center for Green and Sustainable Materials, Institut Teknologi Sumatera, Terusan Ryacudu, Way Hui, Jati Agung, Lampung Selatan 35365, Indonesia
[3] Department of Environmental Engineering, Faculty of Civil and Environmental Engineering, Institut Teknologi Bandung, Jalan Ganesa No. 10, Bandung 40132, Jawa Barat, Indonesia
[4] Research Group of Physics and Technology of Advanced Materials, Department of Physics, Faculty of Mathematics and Natural Sciences, Institut Teknologi Bandung, Jalan Ganesa No. 10, Bandung 40132, Jawa Barat, Indonesia
[5] Department of Materials Engineering, Faculty of Industrial Technology, Institut Teknologi Sumatera, Terusan Ryacudu, Way Hui, Jati Agung, Lampung Selatan 35365, Indonesia

* **Corresponding author**: tarmizi.taher@tl.itera.ac.id; krijal@itb.ac.id

ARTICLE HISTORY: Received: May 1, 2026 · Revised: September 12, 2026 · Accepted: September 23, 2026

ABSTRACT



Machine learning (ML) is increasingly applied to predict and optimize adsorption-based removal of emerging contaminants (ECs) from water, yet the literature remains fragmented across diverse algorithms, adsorbents, and pollutant classes, and existing reviews of the topic report neither a quantitative synthesis of performance nor a cross-study interpretability analysis. This systematic review, conducted following PRISMA 2020 guidelines, screened 2,164 records from Scopus and PubMed and analyzed 202 peer-reviewed studies published between 2015 and 2025. Among 59 unique algorithms, artificial neural networks (ANN) dominated (64.9%), though ensemble methods (particularly random forest and XGBoost) are rapidly gaining adoption. Inverse-variance weighted meta-analysis of 110 studies revealed pooled R^2 values ranging from 0.998 (ANFIS) to 0.929 (CatBoost), with extremely high heterogeneity ($I^2 > 98\%$) precluding universal algorithm recommendations; funnel plot analysis revealed patterns consistent with publication bias. A cross-study synthesis of SHAP-based feature importance from 15 explainable ML studies identified initial concentration, BET surface area, and pH as universally important predictors, consistent with classical adsorption theory, while revealing system-specific drivers (elemental ratios for biochar, molecular weight for PFAS) that extend beyond traditional frameworks. Antibiotics, pharmaceuticals, NSAIDs and analgesics together accounted for 81.7% of target contaminants, while PFAS (9.4%) and microplastics (5.9%) remained substantially underrepresented. Critical methodological concerns persist: among studies reporting a dataset size, 40% used fewer than 50 data points, and an audit of 30 randomly selected papers read in full found that although 90% describe a validation scheme, every one is an internal partition of a single experimental dataset and none validates against independently collected data. To address these gaps, we propose a 22-item Minimum Reporting Checklist for ML in Adsorption Studies (MRCAS), a technology readiness assessment for six frontier architectures (PINNs, GNNs, transfer learning), and five evidence-based priority directions to advance the field from prediction-focused modeling toward practical decision-support tools for water treatment.

KEYWORDS *machine learning; adsorption; emerging contaminants; systematic review; water treatment; artificial intelligence*

1. INTRODUCTION

The contamination of water resources by emerging pollutants has become one of the most pressing environmental challenges of the 21st century. Unlike conventional contaminants such as heavy metals and organic dyes, emerging contaminants (ECs: pharmaceuticals, antibiotics, per- and polyfluoroalkyl substances (PFAS), endocrine-disrupting compounds (EDCs), microplastics, and pesticides) are continuously released into aquatic environments through wastewater discharge, agricultural runoff, and industrial effluents [1, 2]. These compounds are detected at trace concentrations (ng/L to µg/L), yet even at such levels they can disrupt endocrine function, promote antimicrobial resistance, and exert chronic toxicity on aquatic organisms [3, 4]. Conventional wastewater treatment plants are not designed to remove these micropollutants, and reported removal efficiencies vary widely depending on the compound, treatment configuration, and operating conditions [4].

Among the available remediation strategies, adsorption has emerged as one of the most versatile and cost-effective technologies for EC removal from aqueous systems [5, 6]. Its advantages include operational simplicity, high removal efficiency across a broad range of pollutant classes, and the availability of diverse adsorbent materials, from low-cost agricultural waste and biochar to engineered nanomaterials such as metal–organic frameworks (MOFs) and graphene-based composites [7]. However, adsorption performance depends on a complex interplay of factors, including adsorbent properties (surface area, pore structure, functional groups), solution chemistry (pH, temperature, ionic strength), and contaminant characteristics (molecular size, polarity, charge). The highly non-linear and multivariable nature of these interactions makes it difficult to predict adsorption behavior using conventional empirical models such as Langmuir, Freundlich, or pseudo-kinetic equations alone [6].

Machine learning (ML) offers a fundamentally different approach to this modeling challenge. Rather than assuming a predefined functional relationship, ML algorithms learn complex, non-linear mappings directly from data, enabling them to capture interactions that elude classical models [8]. In the broader field of environmental engineering, ML has been successfully applied to water quality prediction, process optimization, and pollutant fate modeling [9]. For adsorption systems specifically, ML models have been trained to predict adsorption capacity, removal efficiency, and breakthrough curves from physicochemical descriptors, often achieving coefficients of determination (R^2) exceeding 0.95 [10, 11]. The rapid growth of computational resources, coupled with the increasing availability of experimental adsorption data, has accelerated the adoption of ML in this domain, particularly since 2020.

Despite this momentum, the literature on ML-assisted adsorption of ECs remains fragmented. Individual studies typically focus on a single

adsorbent–contaminant system, employ different algorithms and validation strategies, and report inconsistent performance metrics, making it difficult to draw generalizable conclusions. The intersection of ML, adsorption and ECs has not gone unreviewed, and it is worth being exact about what the existing reviews do and do not provide, because a claim of novelty is only as good as the comparison behind it. Table 1 sets thirteen reviews published between 2022 and 2026 against the present work. Several are defined by adsorbent rather than contaminant, covering biochar [12], carbonaceous materials [13] or metal–organic frameworks [14]; several address conventional pollutants, principally heavy metals [15–17]; and others treat ML in adsorption generally without separating ECs as a class [18–20]. Three reviews published in 2026, after this review's search cut-off of 24 February 2026 and therefore outside the studies it reviews, come closer. Fan et al. [21] address EC adsorption specifically, and Elngar et al. [22] report PRISMA adherence for adsorption-based treatment, while Yaghoobian et al. [23] cover PFAS across the whole management pipeline and compare interpretability strategies qualitatively. We therefore do not claim to be the first review of this subject.

What the comparison does show is a consistent difference in kind. Of the thirteen, eleven state neither the databases they searched nor the number of studies they included, and none reports pooling reported performance across studies quantitatively or synthesising feature-importance rankings across explainable-ML studies. Every DOI in Table 1 was checked against Crossref, and the verification records are provided in the Supplementary Information. The contribution of the present work is accordingly methodological rather than territorial: it is, to our knowledge, the first review of ML for EC adsorption to combine a stated and reproducible search and screening protocol with a quantitative meta-analysis of reported performance, a cross-study synthesis of SHAP-derived feature importance, and an assessment of reporting quality in the underlying literature. That this can be claimed at all is itself a comment on the field, since it is unusual for a body of reviews to leave the size and provenance of its evidence base unstated.

To address this gap, we conducted a systematic review following the PRISMA 2020 guidelines, screening 2,164 records from Scopus and PubMed and analyzing a final set of 202 peer-reviewed studies that first appeared online between 2015 and 2025. Against the reviews in Table 1, which describe individual studies qualitatively, this work adds three analytical contributions: (i) a quantitative meta-analysis of algorithm performance using inverse-variance weighted pooling to generate cross-study comparisons; (ii) a cross-study synthesis of SHAP-based feature importance to identify universal and system-specific drivers of adsorption prediction; and (iii) a Minimum Reporting Checklist for Machine Learning in Adsorption Studies (MRCAS) and a technology readiness assessment for frontier ML architectures. We also report audits of our own screening and extraction procedures (Sections 2.6.1 and 2.6.2), on the view that a review which assesses the reporting quality of others should document its own. The specific objectives are: (1) to map the bibliometric landscape of ML-assisted adsorption research for ECs; (2) to identify which ML algorithms are most frequently applied and conduct a meta-analysis of their predictive performance; (3) to characterize the emerging contaminants and adsorbent materials studied and assess coverage gaps; (4) to evaluate methodological rigor, including dataset sizes, validation protocols, and reporting practices; (5) to synthesize feature importance findings across explainable ML studies; and (6) to provide a frontier technology roadmap and evidence-based recommendations for future work.

We restricted the scope to studies in which (1) ML is applied as a core modeling methodology, not merely cited; (2) adsorption is the primary removal mechanism; and (3) the target pollutant belongs to a recognized EC category. Studies addressing only conventional pollutants (heavy metals, dyes) without an EC component, or employing non-adsorption processes (photocatalysis, biodegradation, membrane filtration) as the primary mechanism, were excluded. This focused scope enables a rigorous and comparable assessment of the current state of ML in EC adsorption, while the lessons learned are broadly applicable to ML-assisted environmental modeling.

2. METHODOLOGY

This systematic review was conducted following the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) 2020 guidelines. The completed checklist is provided in Table S1, and the subsections below are ordered to follow its Methods items so that each item maps to one subsection.

Parts of this review were automated. Screening was performed by rule-based scripts and one phase of data extraction by a language model. We report this at the level of detail required by PRISMA items 8 and 9, both of which ask for details of any automation tools used, and by the 2025 joint position statement on the use of artificial intelligence in evidence synthesis issued by Cochrane, the Campbell Collaboration, JBI and the Collaboration for Environmental Evidence [25], which asks that such use be disclosed, that human oversight be described, and that the authors demonstrate rather than assert that methodological rigour was not compromised. The justification for automating is scale: 1,745 unique records and 383 full texts exceed what this author team could screen and extract by hand, and automation was adopted to make the review feasible rather than to make a feasible review faster. Because that choice substitutes reproducible rules for human judgement, we did not assume the substitution was safe. Both automated stages were audited against re-reading of the source records, and those audits, including the errors they found, are reported in Section 2.6.

2.1 Eligibility Criteria

Throughout this review, "machine learning" (ML) denotes the class of methods under study, and "artificial intelligence" (AI) is used only where it is part of a term taken from the source literature or the screening criteria, or where it refers to the language model used as an extraction tool in Section 2.5; the two are not used interchangeably. Studies were assessed against three core pillars for inclusion: (1) ML or AI methods must be applied as a core methodology, not merely cited; (2) adsorption must be the primary removal mechanism; and (3) the target pollutant must be an emerging contaminant. Table 2 summarizes the full inclusion and exclusion criteria.

Review articles are excluded by criterion E5 as sources of primary data, but they were deliberately retained by the database filters described in Section 2.2, because the comparison of prior reviews in Table 1 and the backward citation checking both required them. Removal therefore happened downstream of retrieval rather than at the query, and Section 2.6.1 reports that this removal initially failed for two records.

2.2 Information Sources

Two electronic databases, Scopus and PubMed, were searched for articles published between January 2015 and December 2025. Results were restricted to peer-reviewed journal articles and reviews written in English. Both searches were executed on 24 February 2026, which is the cut-off date for this review: records indexed after that date were not retrieved, and no records were added subsequently. The publication window closes at 31 December 2025, so every article in the final set had been indexed for at least eight weeks before the search ran. A single dating convention is applied throughout: the year assigned to a study is the year it first appeared online, taken from the online publication date recorded in Crossref where one is available and from the date the DOI was registered where it is not. The issue label is not used. This matters for four records that a database's indexed year would place outside the window. Three appeared online and were indexed during 2025 but carry a 2026 issue; one appeared online in August 2022 and was assigned to a 2024 issue almost two years later. Under the convention they are 2025 and 2022 literature respectively, and every study in the final set therefore falls within 2015–2025 with no exceptions. The four records and their Crossref evidence are listed in the dataset changelog accompanying the Supplementary Data. Two consequences follow and should be borne in mind when reading the temporal trends in Section 3.1. First, coverage of 2025 is likely to be marginally less complete than for earlier years, because indexing of late-2025 articles was still accruing when the search ran; the direction of the upward trend is robust to this, but the 2025 count should be read as a lower bound. Second, because the review is restricted to indexed, peer-reviewed literature, it inherits the publication bias of that literature, in which models that perform well are more likely to be reported than models that do not. The funnel plot in Section 3.5 addresses this directly.

This review was not registered in PROSPERO or another prospective

Table 1. Prior reviews of machine learning for adsorption, compared with the present work. "Not stated" means the review does not report the item. † Published after the 24 February 2026 search cut-off.

Review	Pollutant scope	Databases searched	Studies included	Quant. meta-analysis	Cross-study XAI	Principal limitation relative to this work
Reynel-Avila et al. (2022) [18]	Organic and inorganic pollutants; ECs not a separate class	Web of Science	>250 searched; inclusion count not stated	No	No	ANN only; no protocol or eligibility criteria
Zhang et al. (2023) [19]	Water pollutants generally	Not stated	Not stated	No	No	No screening protocol; ECs not analysed separately
Zhang et al. (2023) [12]	Pollutants removed by biochar	Not stated	Not stated	No	No	Defined by adsorbent, not contaminant
Fiyadh et al. (2023) [15]	Heavy metals only	Not stated	Not stated	No	No	Single model family and single pollutant class
Yuan et al. (2023) [16]	Heavy metals, mainly on biochar	Not stated	Not stated	No	No	Conventional pollutants; no protocol
Wang et al. (2024) [13]	Organic pollutants on carbonaceous adsorbents	Not stated	39	No	No	Carbonaceous adsorbents only; cross-study statistics are algorithm tallies, not pooled effects
Sharmila et al. (2024) [14]	Organic pollutants on MOFs	Not stated	Not stated	No	No	Confined to MOFs; ML is one section of a materials review
El Mallahi et al. (2025) [24]	Not pollutant-specific (bibliometric)	Scopus	Not stated	No	No	Counts and clusters publications; extracts no performance data
Zhao et al. (2025) [17]	Heavy metals only	Not stated	Not stated	No	No	Conventional pollutants; feature importance not pooled
Yuan et al. (2025) [20]	Water-treatment adsorbents generally	Not stated	Not stated	No	No	Materials-design perspective; no defined set of studies
Fan et al. (2026)† [21]	Emerging contaminants	Not stated	Not stated	No	No	Same population as this work, but a decision-oriented framework rather than a synthesis over a defined set of studies; no protocol or study count
Yaghoobian et al. (2026)† [23]	PFAS only, whole management pipeline	Not stated	Not stated	No	Qualitative	Single EC class; interpretability compared narratively, not pooled
Elngar et al. (2026)† [22]	Dyes, heavy metals, organics	Not stated	Not stated	No	No	States PRISMA adherence but reports neither databases nor study count
This work	Emerging contaminants (PFAS, pharmaceuticals, EDCs, microplastics, ARGs, PCPs)	Scopus, PubMed (24 Feb 2026)	202	Yes	Yes (15 SHAP studies)	Two databases only; not prospectively registered

Table 2. Inclusion and exclusion criteria for study selection.

ID	Criterion
Inclusion criteria	
I1	At least one ML algorithm applied to model or predict adsorption
I2	Target contaminant belongs to an emerging category (PFAS, pharmaceuticals, EDCs, microplastics, ARGs, or PCPs)
I3	Adsorption occurs in aqueous or liquid phase
I4	Study reports quantitative performance metrics (e.g., R^2, RMSE, MAE)
I5	Published between 2015 and 2025
I6	Written in English
I7	Published in a peer-reviewed journal
Exclusion criteria	
E1	Only conventional pollutants (heavy metals or dyes) without emerging contaminants
E2	No ML/AI application (traditional statistics only)
E3	Gas-phase adsorption only
E4	Non-adsorption process as primary mechanism (photocatalysis, biodegradation, Fenton, etc.)
E5	Non-peer-reviewed source (conference paper, preprint, book chapter)
E6	ML mentioned only in future work, not applied
E7	Duplicate publication across databases
E8	Full text unavailable
E9	Adsorption of pollutants onto microplastics (without microplastic removal)

register, and no protocol was published in advance; this is recorded in item 24 of the PRISMA checklist (Table S1) and is stated here because registration, while not mandatory, is the usual safeguard against post hoc changes to eligibility criteria. Two features of the workflow limit that risk. The eligibility criteria and the full Boolean query were fixed on 23 February 2026, one day before the searches were run, and neither was altered afterwards. And because every screening and extraction step was executed by scripts rather than by ad hoc judgement, the decision rules are recoverable in full from the code and the logged per-record decisions, which are available from the corresponding author on request.

2.3 Search Strategy

The search strategy employed a three-block Boolean query combining: (1) machine learning and artificial intelligence terms, (2) adsorption and sorption-related terms, and (3) emerging contaminant categories. The ML/AI block included terms such as "machine learning," "artificial neural network," "random forest," "support vector machine," "deep learning," "gradient boosting," and related algorithm names. The adsorption block covered "adsorption," "biosorption," "sorption capacity," "isotherm," and specific adsorbent materials. The emerging contaminant block encompassed six sub-categories: per- and polyfluoroalkyl substances (PFAS), pharmaceuticals and antibiotics, endocrine-disrupting compounds (EDCs), microplastics, antibiotic resistance genes (ARGs), and personal care products (PCPs), each with specific compound names and synonyms. The three blocks were combined using the AND operator. The full query strings as executed against each database are reproduced in SI Section S2.1.

The initial search yielded 1,337 records from Scopus and 827 from PubMed, totaling 2,164 records. After removing 419 duplicates (identified by DOI, PMID, and title matching), 1,745 unique records remained for screening.

2.4 Selection Process

The PRISMA flow diagram summarizing the study selection process is presented in Fig. 1.

2.4.1 Title and abstract screening.

Title and abstract screening was performed using an automated dual-reviewer approach. It is worth being precise about what the two screeners were, because the term "AI-assisted" would overstate them: both are deterministic, rule-based scripts that apply curated regular-expression dictionaries with contextual weighting to the title, abstract and keyword fields. No language model was involved at any point in screening. Reviewer A used a regex-contextual scoring method and Reviewer B a complementary pattern-matching strategy built on independently compiled term lists; each returned INCLUDE, EXCLUDE or UNCERTAIN per record. Because the rules are fixed and the inputs are text fields, both screeners are deterministic: re-running them on the same records reproduces every decision exactly.

Decisions were combined by an explicit rule rather than by discussion. Concordant decisions were carried through unchanged. Where one screener returned UNCERTAIN, the confident screener decided, in both directions (INCLUDE with UNCERTAIN yielded inclusion; EXCLUDE with UNCERTAIN yielded exclusion). Direct conflicts, in which one screener returned INCLUDE and the other EXCLUDE, were not resolved by this rule but passed to a third rule-based script ($n = 220$) that was deliberately biased toward retention, on the principle that full-text screening provides a further filter whereas an exclusion at this stage is final. Inter-screener agreement before merging was moderate (Cohen's $\kappa = 0.401$ for the binary include/exclude contrast). This value should be read for what it measures. It is agreement between two deliberately non-redundant rule sets, not between two trained human readers, and the two were built to be complementary precisely so that their union would be broader than either alone; a high κ would have indicated redundancy rather than reliability. What matters for the integrity of the review is not their agreement but whether the merged rule excluded studies that should have been retained, which we assessed directly by audit (Section 2.6.1). This initial screening excluded 1,057 records, retaining 688 studies.

A subsequent tightened screening phase applied stricter thresholds, requiring that ML/AI be demonstrably the *core* methodology, adsorption be the *primary* process, and the study be original research rather than a review. This phase excluded an additional 305 records (266 auto-removed, 39 borderline cases excluded after manual inspection), yielding 383 studies for full-text assessment.

2.4.2 Full-text screening.

Full-text screening employed a section-aware scoring algorithm that assigned different weights to matches found in different parts of each paper. Matches in the Methods section received the highest weight (3.0×), followed by Results (1.5×), with additional bonuses for first-person ML application verbs (e.g., "we trained," "we optimized"), quantitative ML results (e.g., reported R^2 or RMSE values), and title-level keyword presence. Each paper received three sub-scores (ML, Adsorption, Emerging Contaminant) on a 0–10 scale, with minimum thresholds required for inclusion.

Of the 383 candidate papers, full texts were obtained for 382 (one paper could not be retrieved). Of these 382, 173 were directly included, 143 were excluded, and 66 were classified as borderline. The 66 borderline cases were manually reviewed by one author, resulting in 31 additional inclusions and 35 exclusions. The final set comprised 204 studies for data extraction and synthesis. Two of these were identified during revision as review articles rather than primary studies and were removed, giving the 202 studies analysed here (Section 2.6.1).

2.5 Data Collection Process and Data Items

A structured data extraction framework was developed to capture over 30 fields from each included study, organized into five categories: (1) bibliographic metadata (DOI, title, authors, journal, year, country); (2) ML methodology (algorithms used, best-performing model, validation method, dataset size); (3) adsorption system (adsorbent material, target contaminant, EC category); (4) model performance (R^2, RMSE, MAE); and (5) adsorption outcomes (maximum adsorption capacity q_{max}, removal efficiency). The complete field list and definitions are given in SI Section S2.4.

Three properties of the process should be stated before the phases are described, because each bounds what the extracted dataset can support. Data were collected by a single automated pass over each report, reconciled between two extraction methods but not duplicated by an independent second extractor. Study authors were not contacted to supply missing or ambiguous values. And no values were digitised from figures: extraction operated on text recovered from each PDF, so a quantity reported only inside a plot is systematically absent from the dataset. The verification in Section 2.6.2 measures the consequences of all three.

Full texts were extracted from the 383 candidate PDFs using the pypdf library, achieving a 92.1% good-quality text extraction rate. Data extraction was then performed in two complementary phases.

2.5.1 Phase 1: Regex-based extraction.

An automated extraction pipeline applied 55 ML algorithm patterns, 25+ adsorbent material categories, and 80+ contaminant-specific regular expressions to the extracted full texts. This approach achieved high coverage for well-defined fields such as ML algorithms (98%), EC category (99%), and country (97%), but lower coverage for context-dependent fields requiring interpretive reading, such as best-performing model (51%), R^2 of the best model (20%), and dataset size (10%). The pipeline was written in Python and is deterministic: re-running it on the same texts reproduces every value.

2.5.2 Phase 2: Language-model re-extraction.

To address gaps in context-dependent fields, the same reports were re-extracted using a large language model. The model was Claude Opus 4.6 (Anthropic), accessed in February 2026 and operated through an interactive agent session rather than through a scripted API call. Three properties of that arrangement follow, and we state them rather than leave them to be inferred. The session used the provider's default sampling settings, so the model is *not* deterministic: unlike the rule-based screeners of Section 2.4, re-running it is not guaranteed to reproduce every value. The operative instruction was not preserved verbatim, so the input specification and the output schema can be published (SI Section S2.4) but the prompt itself cannot. And no second, independent language-model

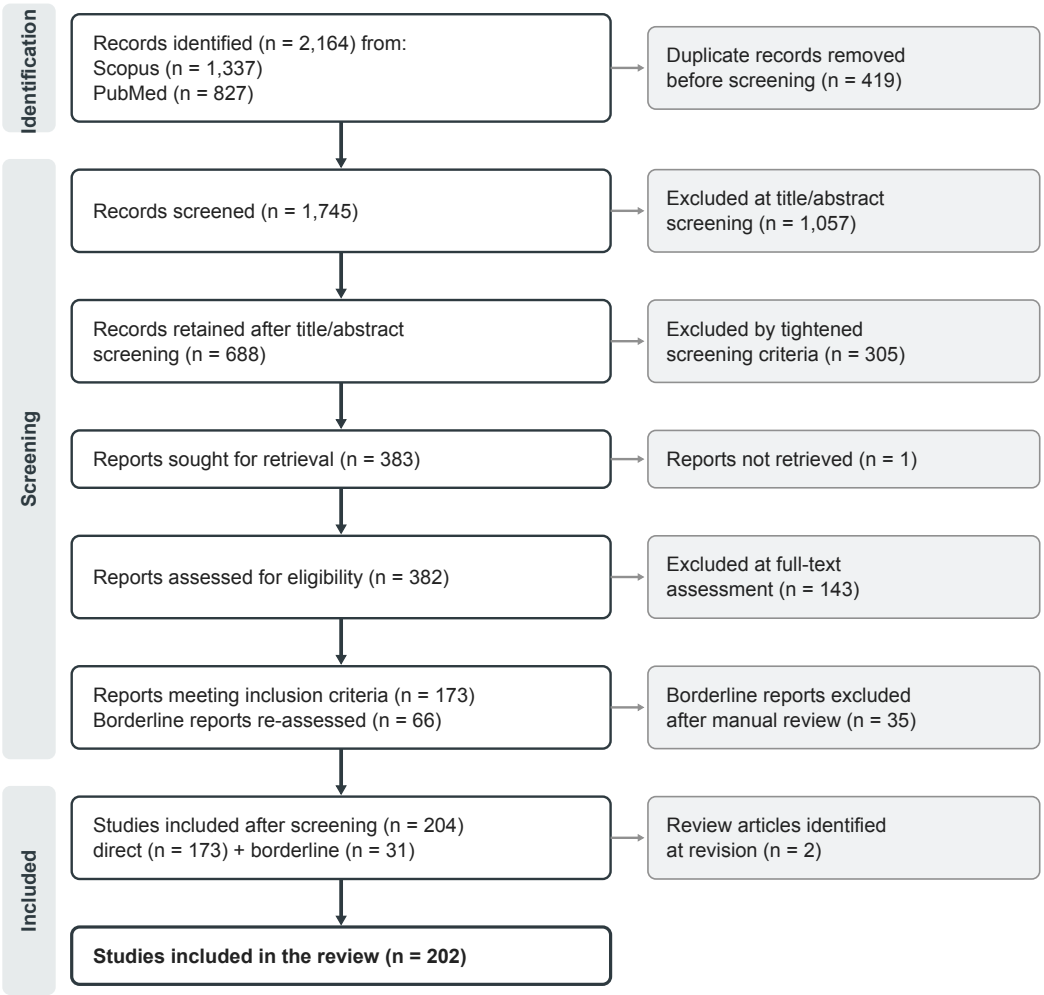


Figure 1. PRISMA flow diagram of the study selection process.

pass was run. These three properties are the reason the accuracy of the output was measured directly rather than argued from the design (Section 2.6.2).

What the model was given, and what it was permitted to return, are both recoverable and are reported in full. The 204 papers then in the included set were prepared in 41 batches of 5 papers each. For each paper, a structured text package was compiled from that paper's own PDF, containing the header (first 3,000 characters, for affiliation information), abstract (up to 2,000 characters), Methods section (up to 4,000 characters), Results and Discussion section (up to 4,000 characters), and Conclusion (up to 2,000 characters). The model was given no access to any other source and was not asked to recall the study from memory. Its output was confined to a fixed schema of 11 context-dependent fields: country, ML algorithms, best model, best R^2, RMSE, q_{max} (mg/g), removal percentage, dataset size, target contaminant, adsorbent material, and study type, with a null permitted in every field.

Results from both extraction phases were merged using a systematic reconciliation protocol: when both sources provided values, agreement was tracked; when only one source had a value, it was retained; disagreements ($n = 582$ field-level cases) were logged for review. This dual-method approach substantially improved coverage, with notable gains for best model (51% → 93%), R^2 (20% → 54%), dataset size (10% → 38%), and q_{max} (20% → 42%).

2.5.3 Standardization.

A final standardization step applied canonical mappings to seven key fields: country names, ML algorithm names, best model identifiers, target contaminants, and adsorbent materials. Algorithm abbreviations were unified (e.g., "LR," "MLR" → "Linear Regression"; "GBDT," "GBR" → "GBM"), compound names were normalized to base forms, and adsorbent materials were categorized into 25 canonical groups. All transformations were recorded in an audit log to ensure reproducibility. The final standardized dataset comprised 202 papers across 40 extracted fields.

2.6 Evaluation of Automation Tool Performance

Screening and extraction were both automated, so the reliability of this review rests on how well those two tools performed on this specific body of literature. Neither the agreement statistics of Section 2.4 nor the coverage gains of Section 2.5 answer that question: agreement measures whether two rule sets concur, and coverage measures how many cells were filled, while what matters is whether the decisions were right and the values correct. We therefore evaluated both tools against re-reading of the source records. Both evaluations were carried out by one of the authors, without an independent second assessor, which is a limitation we state plainly here and again in Section 3.10.

2.6.1 Audit of screening decisions.

Because the merged rule can exclude a record on a single confident vote, the reliability of the included set depends on whether that rule discarded studies it should have kept. Agreement statistics cannot answer this; only re-reading the discarded records can. We therefore audited them directly. The 993 records excluded at the title/abstract stage were stratified by how the exclusion was reached (503 where one screener returned EXCLUDE and the other UNCERTAIN, 405 where both returned EXCLUDE, and 85 removed by the prescreening filter), and a proportionally allocated random sample of 101 records (10.2%, random seed 20260912) was drawn. Each sampled record was then re-assessed by an author against the same inclusion and exclusion criteria, from title and abstract, blind to the screeners' stated reasons, and classified as one that should

have proceeded to full text, one correctly excluded, or one the abstract cannot settle.

No record in the sample was judged a clear false exclusion (0 of 101; 95% Wilson CI 0–3.7%). Four records (4.0%) could not be settled from the abstract and would, on the audit rule, have proceeded to full text; taking all four as potential losses gives a conservative upper bound of 9.7% on the false-exclusion rate, or roughly 97 of the 993 exclusions. The four are informative individually. Two are review articles, which the tightening rule excludes on separate grounds; one models removal across whole treatment trains without naming an adsorption unit; and one models powdered-activated-carbon dosing for taste-and-odour compounds, which are not among the emerging-contaminant categories this review covers. The stratification bore out its premise only weakly: the one-sided stratum contributed three of the four unsettled records and the concordant stratum one, a difference well inside sampling error. The audit is reported by exclusion route in Table S12, with per-record judgements in SI Section S11.

The audit of exclusions prompted a complementary check of the inclusions, and that check found a failure in the opposite direction. Screening the titles of the included set against the same review-article rule identified two records that are review articles: one surveying machine learning for organic pollutant adsorption on carbonaceous materials, which also appears in Table 1 as prior literature, and one reviewing predictive modelling of PFAS behaviour. Their routes differ and both are informative. The first was passed by both screeners as meeting the inclusion criteria. The second was auto-included at the prescreening stage on keyword strength alone, a route that bypasses the rule-based screeners and therefore never met the review-article test. Both were removed, which is why 204 studies leave screening and 202 enter the synthesis. The effect on the reported results is confined to descriptive counts: neither record carries an R^2, a dataset size or a best-performing model, so neither enters the meta-analysis or any performance comparison. The algorithm inventory is the exception and the reason the check mattered, since one of the two lists 21 algorithms that its authors surveyed rather than applied; removing it takes the inventory from 65 unique algorithms to 59 and the total mentions from 501 to 475. This check rests on titles, so it would not detect a review that does not describe itself as one.

Two further limitations of this audit should be stated. It samples the title/abstract stage only, so exclusions made at the tightening and full-text stages are not covered by it. And the assessor applied the same written criteria as the screeners, so it detects records wrongly removed under those criteria, not any narrowness in the criteria themselves.

2.6.2 Verification of extracted data.

Coverage is not accuracy, and the gains reported in Section 2.5 say nothing about whether the values recovered are correct. Three properties of the pipeline constrain how far a language model could invent content: it was shown only text taken from the paper's own PDF rather than being asked to recall the study, its output was confined to a fixed field schema, and every field was reconciled against the independent regex pass, with the 582 disagreements logged. None of these is a guarantee, so we measured the result rather than arguing from the design.

Thirty studies were drawn at random from the 198 whose full text was available (random seed 20260912), and six fields that carry the quantitative synthesis (best model, best-model R^2, dataset size, adsorbent material, target contaminant, and validation method) were checked cell by cell against the source PDF by an author, giving 180 audited cells. Each cell was classed as correct, incorrect, a wrongly empty cell where the paper does report the quantity, or correctly empty.

Where the pipeline recorded a value, it was right in 112 of 120 cells (93.3%, 95% CI 87.4–96.6%); active errors were 8 of all 180 cells (4.4%). Accuracy divides sharply by field type. The three descriptive fields (best model, adsorbent material and target contaminant) were correct in all 87 populated cells and were never wrongly left empty. Every error and every omission falls among the numeric and methodological fields, where accuracy when populated was 75.8% and where 35 of 90 cells were wrongly empty. Validation method is the weakest field by a wide margin: only half its populated cells were correct, and 15 of 18 empty cells should have carried a value. The errors are not random. Four of the eight are the same failure, in which a three-way 60/20/20 partition was compressed to "60:40 split"; two more attach a ratio from elsewhere in the paper to the data split, once from an HPLC mobile phase and once from an adsorbent activation recipe; and one imports the R^2 of an Elovich kinetic fit into the machine-learning column. Per-field results are given in Table S13, every identified error in Table S14, and the full cell-level record in SI Section S11.

Two consequences follow, and we have applied both. First, the descriptive syntheses of algorithms, adsorbents and contaminants in Sections 3.2–3.4 rest on the fields the audit found reliable. Second, no claim in this review is now made from the absence of a value in a sparsely populated field, because the audit shows such absences are substantially extraction failures; where the argument requires knowing what studies actually reported, as for validation practice in Section 3.5, we rely on the 30 papers read in full rather than on field-occupancy counts.

2.7 Study Quality and Risk of Bias

No validated instrument exists for appraising risk of bias in machine-learning adsorption studies. Instruments developed for clinical prediction models, such as PROBAST, presuppose a diagnostic or prognostic outcome and a defined patient population, and they do not transfer to a regression of adsorption capacity on physicochemical descriptors. We therefore did not attempt a formal risk-of-bias appraisal of the included studies, and none should be read into Section 3.7: the eight-indicator score reported there measures what could be recovered from each study by automated extraction, which Section 2.6.2 shows is a lower bound on what the studies actually report, and it is a measure of machine-readability rather than of quality. The absence of a study-level appraisal is a limitation of this review and is recorded as one in Section 3.10.

Two related assessments were made and should not be confused with the one that was not. Risk of bias arising from missing results across the synthesis, that is publication bias, was assessed by funnel plot and Egger's regression and is reported in Section 3.5. And the methodological property that most threatens the pooled estimates, namely that reported R^2 values are almost all obtained by partitioning a single experimental dataset, was established by reading the 30 audited papers in full rather than by scoring all 202, and is reported in the same section.

Certainty in the body of evidence was not graded with GRADE or an equivalent framework. Those frameworks are built around an intervention-and-outcome question with a directional effect estimate, which this review does not pose: the pooled quantity here is a reported goodness-of-fit statistic, not a treatment effect, and its heterogeneity is reported directly instead (Section 3.5).

2.8 Effect Measures and Synthesis Methods

The effect measure carried into quantitative synthesis is the R^2 reported by each study for its best-performing model. Studies were eligible for the meta-analysis if they reported both an R^2 and the identity of the model that achieved it, which 110 of the 202 studies do. The remaining 92 contribute to the descriptive syntheses only. Because R^2 is bounded and its sampling distribution is skewed near unity, pooling was performed on Fisher's z-transform of $\sqrt{R^2}$ rather than on R^2 itself, with the pooled estimate back-transformed for reporting. Studies were weighted by inverse variance, with the standard error of each z value approximated as $1/\sqrt{n-3}$ from the reported dataset size n. Dataset size is itself sparsely reported, and where a study eligible for pooling did not report one, the median dataset size of the pooled set was imputed in its place. This is an analytical choice that affects weights rather than point estimates, and it is stated here because it cannot be recovered from the reported results. Heterogeneity was quantified by Cochran's Q and I^2. Publication bias was assessed by funnel plot and Egger's regression test.

Algorithm names were normalized to the canonical forms of Section 2.5 before pooling, so that a study reporting "GBR" and one reporting "GBDT" contribute to the same stratum. Pooled estimates are reported for the six most frequently reported best-performing algorithms.

Comparisons of R^2 distributions across algorithms used the Kruskal–Wallis rank sum test with ε^2 as the effect size, followed by Dunn's post hoc test with Holm correction across all pairs and Cliff's δ as a pairwise effect size. Two-group comparisons used the Mann–Whitney U test with Cliff's δ, proportions were compared with Fisher's exact test, and monotone associations were tested with Spearman's ρ. Binomial proportions

arising from the audits are reported with Wilson confidence intervals. All analyses were carried out in Python using pandas, NumPy, SciPy and scikit-posthocs. The analysis scripts, which pin the exact versions used, are available from the corresponding author on request.

3. RESULTS AND DISCUSSION

3.1 Bibliometric Overview

The 202 included studies, listed individually in Table S9, were published between 2015 and 2025, with a clear upward trajectory (Fig. 2a). Only eight appeared before 2019, the earliest of them in 2015 [26], whereas the period from 2021 onward accounted for 179 of the 202 studies (88.6%). The most productive year was 2025 (n = 64), followed by 2024 (n = 41) and 2023 (n = 32). This growth mirrors the broader expansion of ML applications in environmental engineering and suggests that the field is still in an accelerating phase.

Geographically, research output was dominated by Asian institutions (Fig. 2b; country assigned from the Scopus affiliation field of the first author). China contributed the largest share of studies (n = 52, 25.7%), followed by India (n = 26, 12.9%), Iran (n = 24, 11.9%), and South Korea (n = 16, 7.9%). Together, these four countries accounted for 58.4% of all included papers. Brazil (n = 11) and Japan (n = 10) were the next largest contributors. The concentration of research in a limited number of countries leaves the regions with the least treatment infrastructure largely unrepresented among the studies that define current model performance.

The 202 studies were distributed across more than 90 peer-reviewed journals, indicating a broad interdisciplinary reach (Fig. 2c). The most frequent outlets were the *Journal of Environmental Chemical Engineering* (n = 12), *Chemosphere* (n = 11), and the *Journal of Hazardous Materials* (n = 11), followed by *Environmental Research* (n = 10) and the *Journal of Water Process Engineering* (n = 9). The top five journals collectively published 26% of all studies, while the remaining 74% were spread across journals spanning environmental science, chemical engineering, and water treatment, reflecting the multidisciplinary nature of this research domain.

To characterize the broader research landscape beyond the 202 included studies, a bibliometric network analysis was performed on the full set of 1,337 Scopus-indexed records (Fig. 3). Country nodes were resolved from the last comma-separated element of each Scopus affiliation string. The keyword co-occurrence network (Fig. 3a) reveals distinct thematic clusters: a central core around “machine learning” and “adsorption” is tightly linked to “artificial neural network,” “random forest,” and “optimization,” while satellite clusters emerge around specific contaminant classes (“tetracycline,” “antibiotics,” “pharmaceuticals,” “microplastics,” “PFAS”) and adsorbent materials (“biochar,” “activated carbon”). The temporal overlay indicates that keywords such as “XGBoost,” “PFAS,” “microplastics,” “deep learning,” and “SHAP” (shown in lighter tones) represent the most recent research frontiers, consistent with the trends observed in the included studies. In contrast, earlier keywords such as “response surface methodology” and “genetic algorithm” appear in darker tones, reflecting their established but plateauing role.

The co-authorship country network (Fig. 3b) reveals that China and the United States serve as the two dominant hubs, with the strongest bilateral collaboration link between them. India, Iran, and South Korea form secondary hubs with extensive connections to both China and Western countries. European nations (United Kingdom, Germany, France, Spain) form a densely interconnected cluster, while countries in Southeast Asia, Africa, and Latin America appear as peripheral nodes with fewer collaboration links. This pattern underscores the geographic concentration noted earlier and highlights opportunities for expanding collaborative networks to underrepresented regions.

3.2 Machine Learning Algorithms

A total of 59 unique ML algorithms were identified across the 202 studies, with 475 algorithm mentions in total (Fig. 4a; the complete inventory, with the number of studies using each algorithm, is given in Table S3). Artificial neural networks (ANN) were by far the most widely adopted, appearing in 131 studies (64.9%). Random forest (RF) was the second most popular (n = 52, 25.7%), followed by response surface methodology (RSM; n = 47, 23.3%), extreme gradient boosting (XGBoost; n = 23, 11.4%), and support vector machines (SVM; n = 21, 10.4%). Other frequently used algorithms were the adaptive neuro-fuzzy inference system (ANFIS; n = 22), decision trees (n = 18), gradient boosting machines (GBM; n = 18), and linear regression (n = 18).

Regarding algorithm multiplicity, 45.5% of studies employed a single algorithm, 53.5% used two or more (mean = 2.35, median = 2), and the remaining two studies (1.0%) have no algorithm recorded at all, an extraction gap of the kind quantified in Section 2.6.2. Multi-algorithm comparison studies typically benchmarked three to five algorithms against each other to identify the best performer for their specific adsorption system (e.g. Refs. [11, 27, 28]). Regression was the dominant task type (84.2% of studies), with 41.1% also incorporating optimization components. Classification-based approaches were relatively rare (n = 14, 6.9%), suggesting that continuous prediction of adsorption capacity or removal efficiency remains the primary objective.

When evaluating reported best-performing models, ANN maintained its leading position, being selected as the top model in 88 studies (43.6%), followed by RF (n = 20, 9.9%), ANFIS (n = 14, 6.9%), and XGBoost (n = 12, 5.9%) (Fig. 4b). Other algorithms that achieved best-model status included GBM (n = 6), CatBoost (n = 5), and stacking ensembles (n = 3). ANN’s share as the best model (43.6%) is lower than its usage rate (64.9%), indicating that in head-to-head comparisons ensemble methods such as RF and XGBoost frequently outperform neural network approaches (e.g. Refs. [29–31]).

Temporal analysis of the top five algorithms revealed distinct adoption trajectories (Fig. 4c). ANN usage has remained consistently high since 2019, whereas ensemble methods have gained momentum since 2021. RF usage grew from one study in 2020 [32] to 18 in 2025, and XGBoost, virtually absent before 2022 [33, 34], rose to 15 studies in 2025 alone. This shift follows a broader trend in the ML community toward gradient boosting frameworks, which often achieve competitive performance with less hyperparameter tuning than deep neural networks. CatBoost and LightGBM have also appeared as alternatives in recent studies, while crystal graph convolutional neural networks (CGCNN) have so far been applied in a single study [35]. RSM usage has remained relatively stable, owing to its long-established role in experimental design optimization within chemical engineering.

Regarding software platforms, MATLAB was the most frequently reported tool (31.2% of all studies), followed by Python (19.8%) and Design-Expert (14.4%). A validation method could be recovered from the text of 45.5% of studies, cross-validation being the most common (n = 31); as Section 3.5 shows, that figure reflects what the extraction could capture rather than what the studies describe.

3.3 Emerging Contaminants

The reviewed studies addressed 12 distinct emerging contaminant categories encompassing 146 unique compounds (Fig. 5a; the 50 most frequently studied compounds are listed in Table S7). Antibiotics were the most frequently studied category, appearing in 105 studies (52.0%), followed by pharmaceuticals (n = 72, 35.6%), non-steroidal anti-inflammatory drugs (NSAIDs; n = 38, 18.8%), and other pharmaceuticals (n = 31, 15.3%). Combining the four pharmaceutical categories, antibiotics, pharmaceuticals, NSAIDs and analgesics, gives 165 studies (81.7%), which underscores the strong research focus on drug residues in water systems.

Beyond pharmaceuticals, endocrine-disrupting compounds (EDCs) were targeted in 24 studies (11.9%), per- and polyfluoroalkyl substances (PFAS) in 19 (9.4%), and microplastics in 12 (5.9%). Pesticides (n = 13, 6.4%) and personal care products (PCPs; n = 14, 6.9%) received comparatively less attention. Heavy metals appeared as co-contaminants alongside emerging pollutants in 22 studies (10.9%) and dye removal in 15 (7.4%), reflecting the multi-contaminant nature of real wastewater streams.

Regarding co-occurrence, 47.6% of studies focused on a single EC category, while 52.4% addressed contaminants spanning two or more categories (mean = 1.8 categories per paper). This indicates a growing recognition that real-world water matrices contain mixtures of emerging pollutants, necessitating ML models capable of handling multi-contaminant scenarios.

At the individual compound level, tetracycline was the most

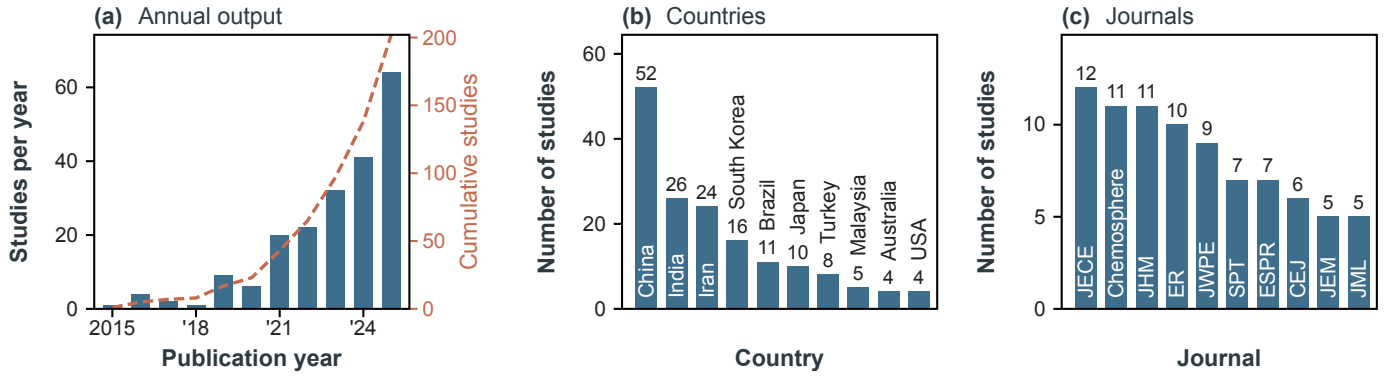


Figure 2. Bibliometric overview of the 202 included studies: (a) annual publication trend with cumulative count (dashed line), (b) top 10 contributing countries, and (c) top 10 publishing journals. Journal abbreviations: JECE, J. Environ. Chem. Eng.; JHM, J. Hazard. Mater.; ER, Environ. Res.; JWPE, J. Water Process Eng.; SPT, Sep. Purif. Technol.; ESPR, Environ. Sci. Pollut. Res.; CEJ, Chem. Eng. J.; JEM, J. Environ. Manage.; JML, J. Mol. Liq.

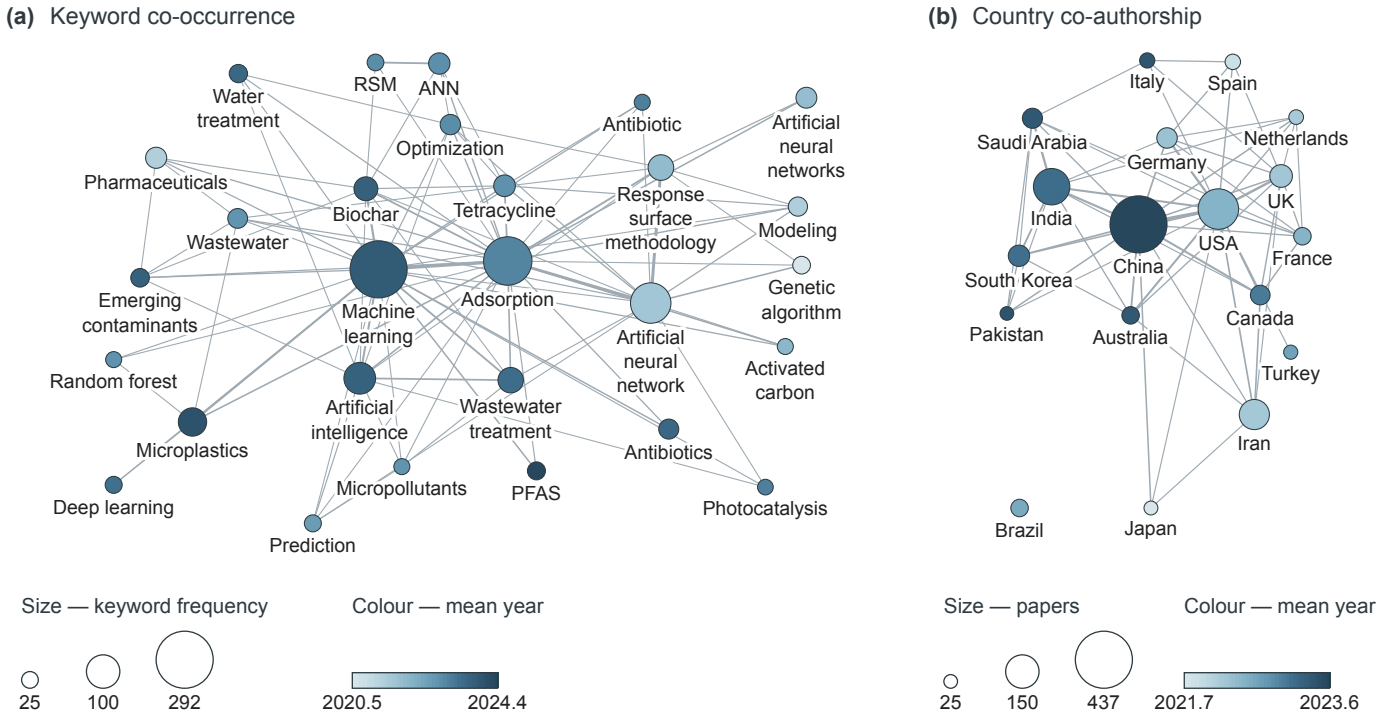


Figure 3. Bibliometric network analysis of 1,337 Scopus-indexed records. (a) Author keyword co-occurrence. (b) Co-authorship by country. Node size is keyword frequency or publication count; node colour is the mean publication year of the records concerned (light = earlier, dark = more recent); edge thickness is co-occurrence or collaboration strength.

frequently modeled contaminant (n = 36, 17.8%), followed by ciprofloxacin (n = 32, 15.8%), ibuprofen and diclofenac (each n = 16, 7.9%), and carbamazepine (n = 15, 7.4%) (Fig. 5b). Among non-pharmaceutical compounds, bisphenol A (n = 13, 6.4%) was the most studied EDC, while PFOA (n = 7) and PFOS (n = 5) were the most targeted PFAS compounds. The dominance of a small number of model compounds, with the top 10 contaminants appearing in over 80% of studies, suggests that future research should broaden the scope to include less-studied but environmentally relevant pollutants.

Temporal analysis of EC category trends (Fig. 5c) revealed that antibiotic adsorption modeling has grown consistently from 4 studies in 2019 to 30 in 2025, with pharmaceutical studies following a similar trajectory. The most striking temporal trend was the emergence of PFAS-related ML studies, absent before 2021 but rising to 10 studies by 2025, which reflects heightened global regulatory attention to "forever chemicals." Microplastic studies appeared only from 2023 onward (n = 12 in total), and EDC studies showed steady but modest growth. These trends collectively suggest that while the field remains anchored in pharmaceutical contaminants, it is diversifying toward PFAS and microplastics in response to evolving environmental priorities.

3.4 Adsorbent Materials

Adsorbent information was available for 201 of the 202 studies (99.5%), spanning 18 distinct material categories (Fig. 6a). Biochar was the most frequently studied adsorbent, appearing in 43 studies (21.3%), reflecting its low cost, tunable surface chemistry, and growing availability from agricultural and industrial waste streams. Activated carbon ranked second (n = 30, 14.9%), followed by carbon nanotubes (n = 14, 6.9%), microplastics as adsorbent substrates (n = 9, 4.5%), graphene oxide (n = 8, 4.0%), and nanocomposites (n = 8, 4.0%). Advanced materials such as metal–organic frameworks (MOFs; n = 8, 4.0%), iron oxide and magnetic particles (n = 6, 3.0%), and layered double hydroxides (LDHs; n = 4, 2.0%) constituted a smaller but growing share. Traditional adsorbents including clay and bentonite (n = 5), chitosan (n = 5), and zeolite

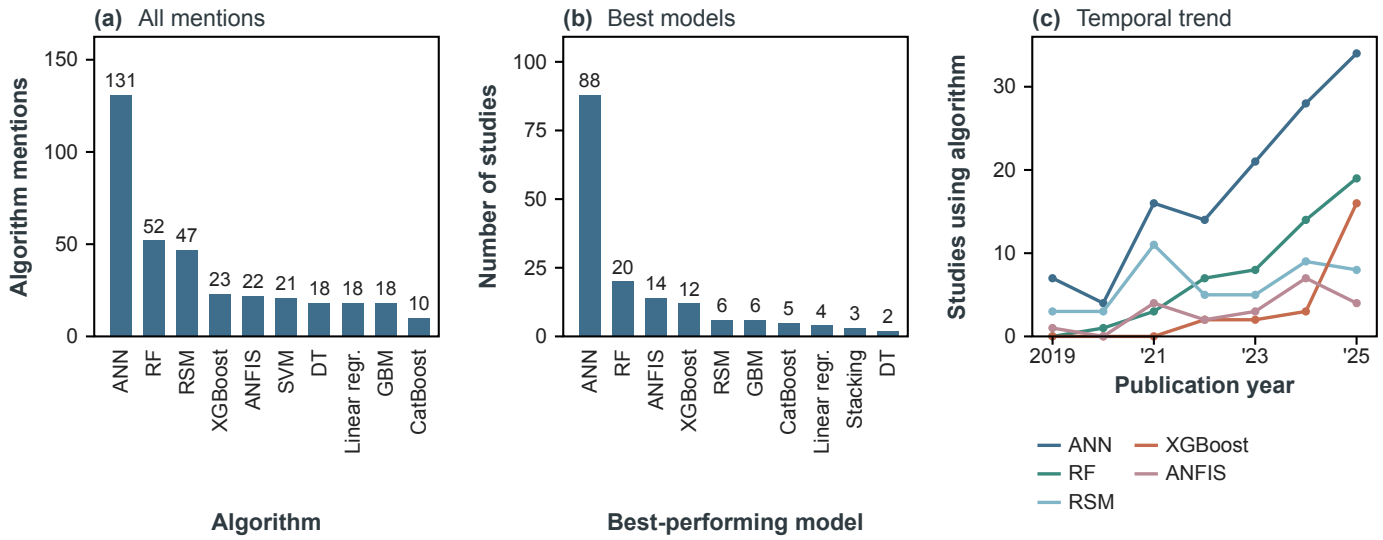


Figure 4. Machine learning algorithms in the reviewed studies: (a) top 10 most frequently used algorithms, (b) top 10 best-performing models, and (c) temporal trend of the five most popular algorithms (2019–2025).

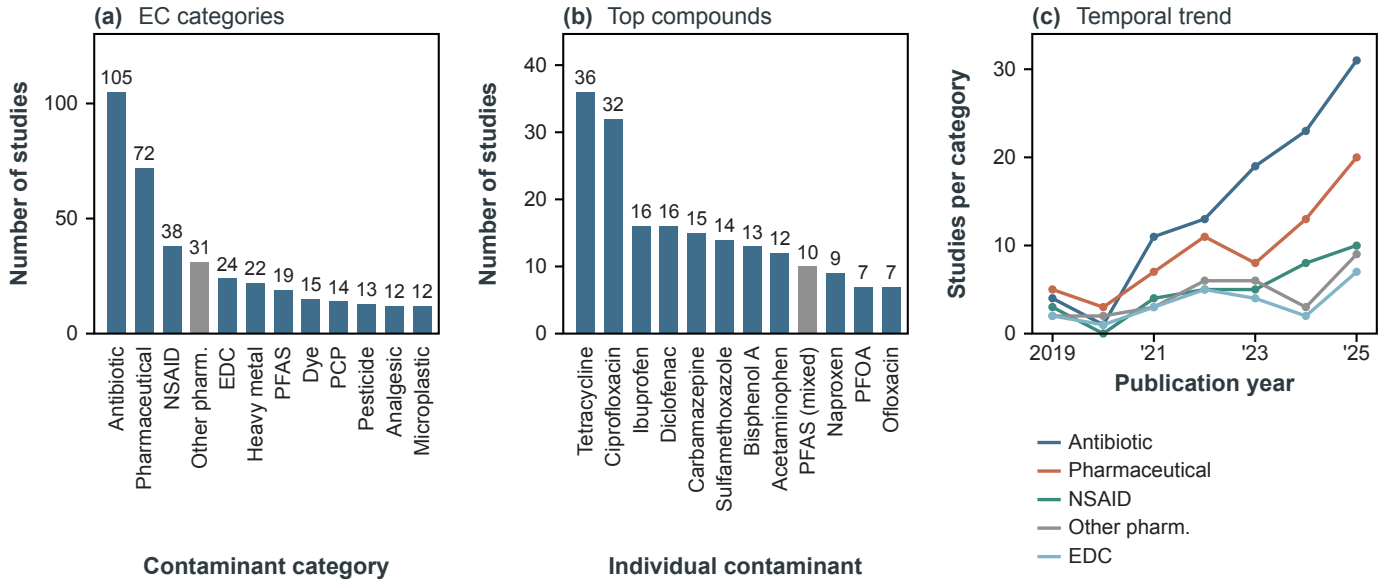


Figure 5. Emerging contaminant analysis: (a) frequency of EC categories across the 202 studies, (b) top 12 most frequently studied individual contaminants, and (c) temporal trend of the five major EC categories (2019–2025).

($n = 2$) were comparatively underrepresented, likely because their well-characterized adsorption behavior leaves less scope for ML-driven discovery. A further 33 studies employed adsorbents classified as "other," encompassing diverse or less-common materials.

Carbon-based materials collectively dominated the dataset. The four carbon categories account for 94 studies (46.5%) when each study is assigned to a single category, as in the counts above, and for 104 studies (51.5%) if studies combining a carbon material with another class are also counted; thirteen studies carry more than one adsorbent category and are the difference between the two figures. This dominance aligns with their high surface area, chemical modifiability, and established effectiveness for organic micropollutant removal. The remaining studies featured diverse materials spanning biosorbents, agricultural waste, polymers and resins, silica, and soil or sediment, indicating that ML modeling is being applied across the full spectrum of adsorbent technologies.

Temporal analysis revealed notable shifts in material preferences (Fig. 6b). Biochar usage grew sharply from 2022 onward, with 12 studies in 2024 and 13 in 2025, together 25 of the 43 biochar papers (58.1%). Activated carbon showed a similar recent surge, with 13 studies in 2025. The most striking trends were the emergence of MOF-based studies, which appeared only from 2023 [36] and reached 5 studies in 2025, and microplastic adsorption studies, which followed a parallel trajectory from 2023 through 2025. These trends mirror the broader materials science literature, where MOFs and microplastic–contaminant interactions are active research frontiers.

The cross-tabulation of adsorbent categories with EC categories (Fig. 6c) revealed clear material–contaminant pairing patterns. Biochar was predominantly applied to antibiotic adsorption (12 studies) and pharmaceutical removal (6 studies), consistent with its proven affinity for aromatic organic compounds. Activated carbon showed a similar antibiotic focus (7 studies) but was more evenly distributed across contaminant classes, including PFAS, EDCs, and heavy metals. Carbon nanotubes were applied exclusively to antibiotics and pharmaceuticals. PFAS removal studies showed no strong preference for a single adsorbent type, with papers distributed across biochar, activated carbon, MOFs, LDH, graphene oxide, polymer and resin materials, soil and sediment, and other sorbents,

which reflects ongoing efforts to identify optimal sorbents for these recalcitrant compounds. EDC studies were concentrated in the biochar and graphene oxide categories, while pesticide adsorption was primarily modeled using activated carbon and biochar.

To provide a quantitative perspective on adsorption performance across material types, the reported maximum adsorption capacity (q_{max}) values were compared across adsorbent and contaminant categories (Fig. 7). Among the 86 studies reporting q_{max} (Table S6), values spanned over four orders of magnitude, from 0.69 mg/g for heavy metal uptake on low-cost biosorbents to 14,499 mg/g for high-capacity engineered materials, with an overall median of 115 mg/g. When stratified by adsorbent category (Fig. 7a), MOFs and nanocomposites exhibited the highest median capacities, consistent with their high surface areas and designed pore architectures, while traditional materials such as clay/bentonite and iron oxide/magnetic particles showed lower but more consistent values. Biochar and activated carbon displayed wide ranges reflecting the substantial variability in feedstock, preparation conditions, and target contaminants. When grouped by EC category (Fig. 7b), antibiotic and pharmaceutical studies reported the highest median q_{max} values, whereas PFAS studies, despite employing high-performance adsorbents, showed lower capacities, likely reflecting the lower initial concentrations and higher recalcitrance of fluorinated compounds. These cross-categorical comparisons highlight that ML models in this field must accommodate adsorption capacities varying by several orders of magnitude, underscoring the importance of appropriate data normalization and log-transformation in model training.

3.5 Model Performance

Model performance was assessed primarily through the coefficient of determination (R^2), which was reported in 110 of the 202 studies (54.5%). Other metrics were reported less frequently: RMSE in 52 studies (25.5%), removal efficiency in 116 studies (56.9%), maximum adsorption capacity (q_{max}) in 86 studies (42.2%), and MAE in only 16 studies (7.8%) [37–39]. The inconsistent reporting of performance metrics across studies represents a significant barrier to systematic comparison and meta-analysis.

Among the 110 studies reporting an R^2 value, model performance was generally high (Fig. 8a). The median R^2 was 0.984 (mean = 0.969): 86 of the 110 studies (78.2%) reached $R^2 \geq 0.95$, 46 (41.8%) reached $R^2 \geq 0.99$, and only 10 (9.1%) reported a value below 0.90. These values indicate strong reported predictive capability across the literature, but such uniformly high figures may partly reflect publication bias toward favorable outcomes and the tendency to report only the best-performing model configuration.

Comparison of R^2 values across the five most frequently reported best-performing models (Fig. 8b) revealed that ANFIS achieved the highest median R^2 (0.997, $n = 8$), followed by ANN (median = 0.991, $n = 48$) and GBM (median = 0.990, $n = 6$). XGBoost showed slightly lower but competitive performance (median = 0.970, $n = 9$), while RF exhibited the most variable results (median = 0.947, $n = 11$) with the widest interquartile range. A Kruskal–Wallis rank sum test showed that these differences were statistically significant ($\chi^2 = 19.39$, df = 4, $p = 6.6 \times 10^{-4}$, $\varepsilon^2 = 0.20$; Table S4), although the unequal sample sizes and potential confounding by study-specific factors (dataset size, contaminant type, adsorbent material) warrant cautious interpretation. A global test of this kind establishes only that the five distributions are not interchangeable. It does not identify which algorithms differ. We therefore added Dunn's post hoc test with Holm correction across all ten pairs, reported in full with Cliff's δ effect sizes in Table S10. Two comparisons survive correction: ANN against RF ($p = 0.005$, $\delta = 0.64$) and ANFIS against RF ($p = 0.012$, $\delta = -0.77$). The remaining eight pairs, including ANN against ANFIS and ANN against XGBoost, are not distinguishable at the 5% level. The significant global result is thus driven almost entirely by RF sitting below the others, and the apparent ordering among ANFIS, ANN, GBM and XGBoost in Fig. 8b is not statistically supported. The sample sizes bound this conclusion as well: GBM ($n = 6$) and ANFIS ($n = 8$) contribute few studies, so the power to detect anything but a large difference involving them is low, and a non-significant pair here indicates insufficient data rather than equivalence. Less common algorithms also performed strongly in individual studies: CatBoost achieved top results in multi-algorithm benchmarks (e.g. Refs. [11, 27, 40]), while Cubist [41], LS-SVM [42, 43], and gene expression programming (GEP) [44, 45] were reported as best-performing in specific adsorption systems.

Dataset size was reported in 78 studies (38.2%), revealing a concerning distribution (Fig. 8c). Nearly 40% of those studies ($n = 31$) worked with fewer than 50 data points, and the median dataset size was only 131 samples. Studies using 200 or more samples represented 38.5% of those reporting dataset size, while only 15.4% exceeded 1,000 samples (e.g. Refs. [46, 11, 30]). The prevalence of small datasets raises questions about model generalizability and the risk of overfitting, particularly for complex models such as deep neural networks that require substantially more training data [47, 48]. The reported removal efficiencies were consistently high (median = 95.3%, $n = 116$), and q_{max} values spanned a wide range from 0.69 to 14,499 mg/g (median = 115 mg/g, $n = 86$), reflecting the diversity of adsorbent–contaminant systems studied.

Small training sets are the most frequently voiced concern about this literature, and the 202 studies allow it to be examined rather than assumed. The relationship between R^2 and dataset size was tested on the 51 studies that report both quantities, 46% of the 110 with a usable R^2. The analysis therefore rests on under half the performance data, and every statement below is conditional on that subset (Fig. 9). The result is not the one usually expected. The rank correlation between dataset size and reported R^2 is weak and not significant (Spearman $\rho = -0.147$, $p = 0.304$), and a direct comparison of the two strata contradicts the inflation hypothesis outright: studies with fewer than 50 samples report a median R^2 of 0.964, slightly *below* the 0.977 of studies with 50 or more (Mann–Whitney $p = 0.70$, Cliff's $\delta = -0.07$, negligible). Nor are implausibly perfect fits concentrated among the small studies: R^2 exceeds 0.99 in 3 of 16 small-sample studies (19%) and 6 of 35 larger ones (17%), a difference indistinguishable from chance (Fisher exact $p = 1.0$). Dispersion runs the same way, with the smallest stratum no tighter than the largest (SD 0.044 against 0.044). On the evidence available, then, reported performance in this literature does not degrade with sample size, and we can find no statistical signature of small-sample overfitting in the values that authors report.

The correct inference from this is not reassurance. It is that reported R^2 is not a quantity in which overfitting would be visible, because in most of these studies nothing was held back that could reveal it. This is where the extraction audit of Section 2.6.2 changes what can responsibly be claimed. The extracted validation-method field is populated for only 92 of 202 studies (46%), which invites the conclusion that half the literature reports no validation at all; the audit shows that conclusion would be wrong. Reading 30 randomly selected papers in full, 27 (90%, 95% CI 74–97%) do state a validation scheme, so most of the empty cells are extraction failures rather than silent authors. What those 27 schemes have in common is the substantive point. Every one of them is an internal partition of the same experimental dataset: a random train/test or train/validation/test split, k-fold cross-validation, or both. Exactly one study describes any check against separately collected data, and that check comprises four additional experimental conditions run in duplicate. None validates against an independent dataset from another laboratory or study. The label is unreliable as well: of the four studies the extracted dataset records as performing "external validation," the one that fell into the audit sample turns out to mean a held-out test set from a 7:3 split of its own data. A model evaluated on data it was fitted to will report a high R^2 whatever its sample size, and that is what the flat relationship in Fig. 9 most plausibly reflects: not an absence of overfitting, but an absence of the reporting that would detect it. The same reasoning limits what the algorithm comparison above can mean, since the same undisclosed validation practices underlie every R^2 entering the Kruskal–Wallis and Dunn tests, and they need not be uniform across algorithm families. The stratified analysis makes the point concretely. Splitting the 51 studies by sample size leaves 9 and 23 studies respectively in the five-algorithm subset, several algorithms represented by a single study, and no interpretable test of whether algorithm ranking depends on dataset size; the honest answer to whether small datasets distort the comparison between algorithms is that these studies cannot say. What can be said is that 77% of reported R^2 values exceed 0.95, and that essentially all of them were obtained by splitting a single experimental dataset rather than by confronting a model with data gathered independently. That, and not sample size in itself, is the substantive threat to the reliability

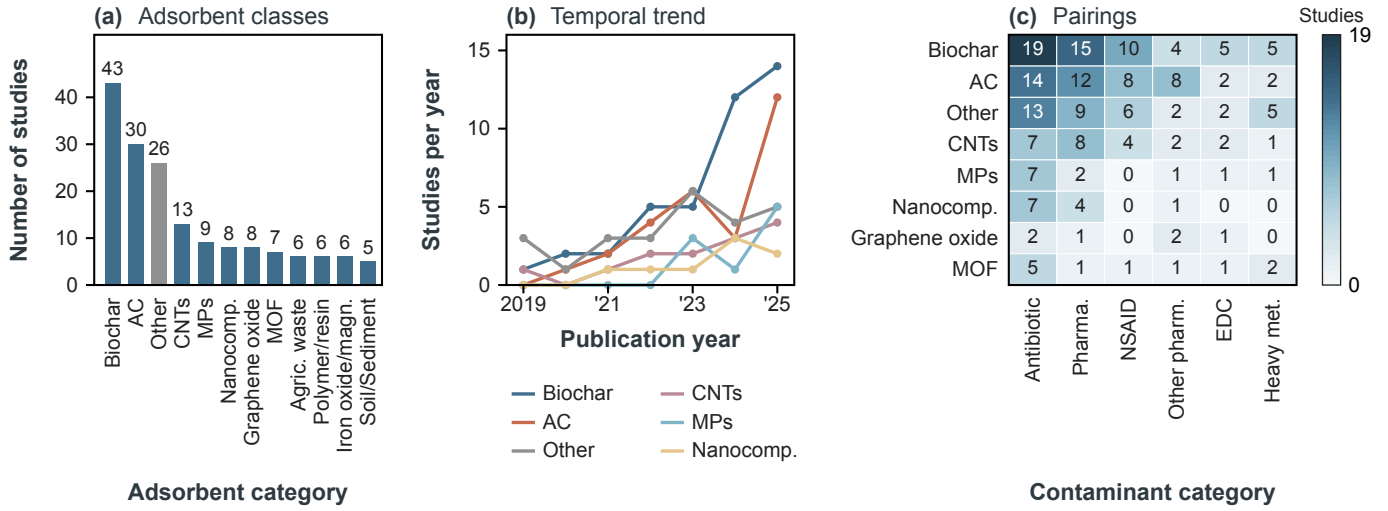


Figure 6. Adsorbent material analysis: (a) frequency of the top 12 adsorbent categories, (b) temporal trend of six selected categories (2019–2025), and (c) heatmap of adsorbent–contaminant pairings showing the number of studies at each intersection. Abbreviations: AC, activated carbon; CNTs, carbon nanotubes; MPs, microplastics; MOF, metal–organic framework; Nanocomp., nanocomposite.

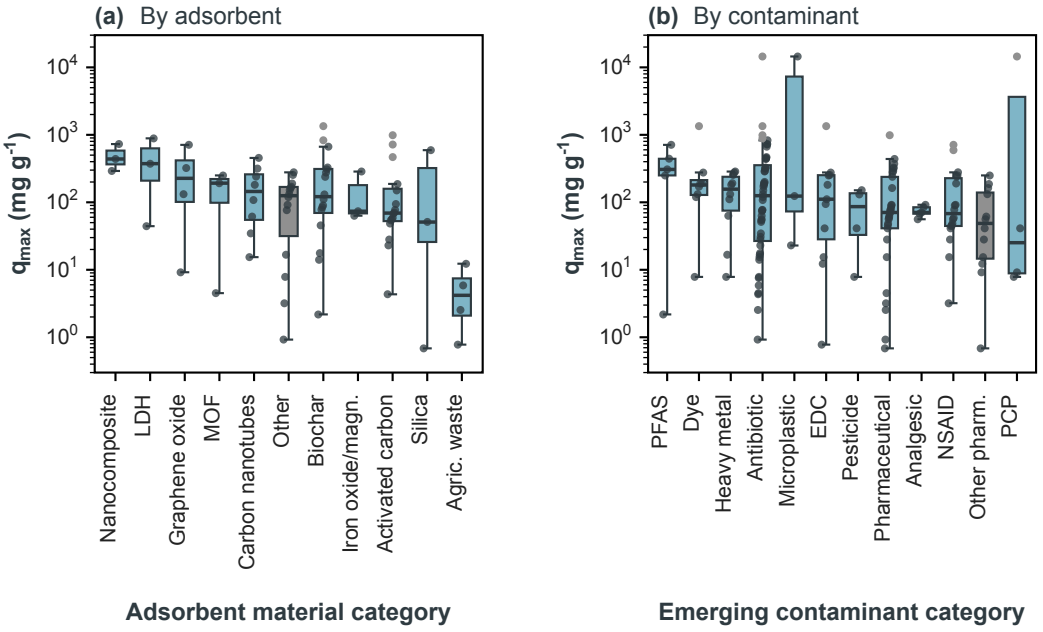


Figure 7. Distribution of reported maximum adsorption capacity (q_{max}, log scale) across (a) adsorbent material categories and (b) emerging contaminant categories. Box plots show the median (center line), interquartile range (box), and individual data points (jittered dots). Only categories with ≥ 3 studies are shown.

of the pooled estimates reported here, and it is the practice this review would most like to see change. This underscores the urgent need for standardized external validation, reporting of confidence intervals, and the use of held-out test sets that are independent of the training data.

3.5.1 Meta-analysis of algorithm performance.

To move beyond descriptive comparison and provide a statistically rigorous cross-study synthesis, we pooled the reported R^2 values by the method set out in Section 2.8, across the 110 studies reporting both an R^2 and the identity of the model that achieved it (Fig. 10a). Among the six most frequently reported best-performing algorithms, ANFIS achieved the highest pooled R^2 (0.998, 95% CI [0.998, 0.998], $n = 8$), followed by ANN (0.995 [0.995, 0.995], $n = 50$), XGBoost (0.979 [0.978, 0.980], $n = 9$), GBM (0.978 [0.977, 0.978], $n = 7$), RF (0.955 [0.951, 0.958], $n = 11$), and CatBoost (0.929 [0.926, 0.932], $n = 5$). However, heterogeneity was extremely high across all algorithms ($I^2 > 98\%$, Cochran's Q $p < 0.001$ for all), indicating that the substantial between-study variability (driven by differences in adsorbent–contaminant systems, dataset characteristics, and feature engineering) precludes a simple "universal best algorithm" conclusion. Notably, the apparently lower pooled R^2 for CatBoost likely reflects its preferential application to more challenging multi-contaminant or multi-adsorbent datasets [11, 27, 40], rather than inferior algorithmic capability.

A funnel plot analysis was performed to assess potential publication bias (Fig. 10b). Visual inspection revealed notable asymmetry, with a concentration of studies reporting high R^2 values (> 0.95) regardless of sample size, and a conspicuous absence of studies in the lower-left region (low precision, low R^2). While Egger's regression test did not reach statistical significance ($p = 0.97$), the visual pattern strongly suggests that studies reporting low predictive performance are either not submitted or not published, consistent with publication bias. The funnel plot further reveals that small-sample studies (high standard error) cluster near the

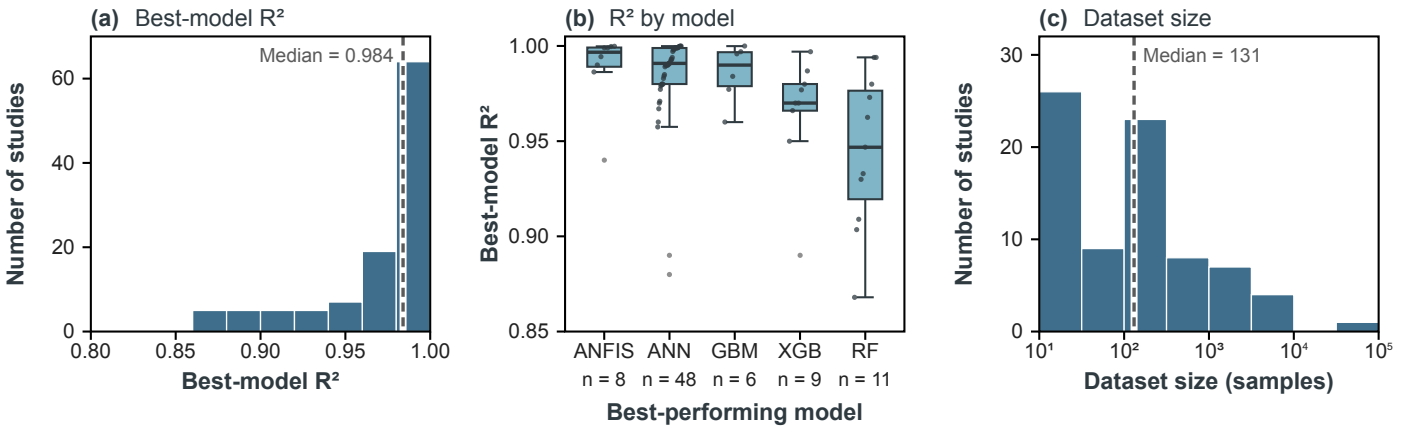


Figure 8. Model performance analysis: (a) distribution of best-model R^2 values ($n = 110$; dashed line indicates the median), (b) R^2 distribution by the five most frequently reported best-performing models, and (c) dataset size distribution ($n = 78$).

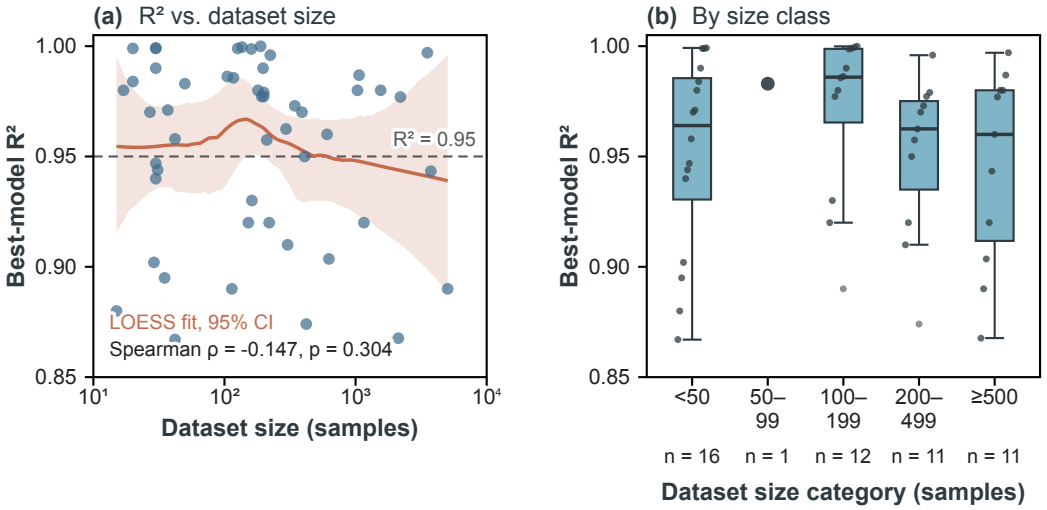


Figure 9. Model performance against training data availability ($n = 51$ studies reporting both). (a) Best-model R^2 versus dataset size (log scale), with LOESS curve and 95% band; dashed line marks $R^2 = 0.95$; Spearman statistics inset. (b) R^2 by dataset size category.

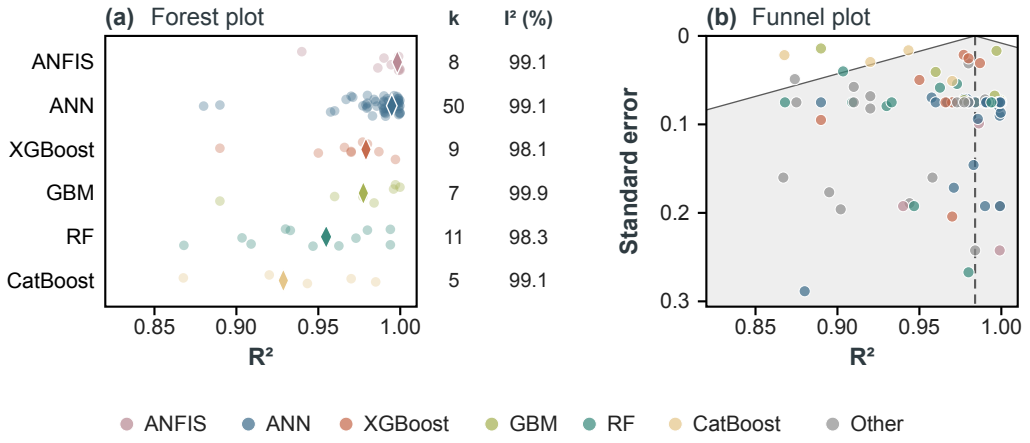


Figure 10. Meta-analysis of algorithm performance. (a) Forest plot of pooled R^2 (inverse-variance weighted, Fisher's z) for the six most reported best models; diamonds are pooled estimates with 95% CI, translucent points individual studies. (b) Funnel plot, one point per study coloured by algorithm; dashed line is the overall median R^2, shaded region the pseudo-95% interval.

upper boundary of $R^2 = 1.0$, reinforcing the overfitting concern identified earlier. These meta-analytic findings collectively argue that the field's reported predictive performance is likely inflated, and that future studies should pre-register analysis plans, report negative results, and provide external validation on independent datasets.

Subgroup analysis by adsorbent and contaminant categories (Fig. 11) revealed that model performance varies substantially across adsorption systems. By contaminant category, pesticide and antibiotic studies achieved the highest median R^2 ($\geq$ 0.98), likely reflecting the more extensive datasets available for these well-studied compound classes. PFAS studies, despite employing advanced algorithms (XGBoost, CatBoost), showed slightly lower and more variable performance (median $R^2 \approx 0.96$), consistent with the greater chemical diversity and lower concentrations characteristic of PFAS mixtures. By adsorbent category, carbon nanotube-based studies reported the highest median R^2, followed by activated carbon and biochar, while MOF-based studies, being relatively new, showed wider performance variability.

3.6 Feature Importance Synthesis Across Studies

While most ML models in the adsorption literature function as "black boxes," a growing subset of studies ($n = 15$, 7.4%) have employed model-agnostic interpretability tools, primarily SHAP (SHapley Additive exPlanations), to identify the key drivers of predicted adsorption performance. To date, these feature importance analyses have been reported in isolation, with each study interpreting its SHAP results within its own adsorbent–contaminant context. Here, we synthesize feature importance rankings across those 15 studies to identify universally important predictors and evaluate their alignment with classical adsorption theory (Fig. 12).

Across the 15 SHAP-enabled studies, initial contaminant concentration and BET surface area emerged as the two most universally important features, each appearing as a top-ranked predictor in all 15 studies with mean relative importance scores of 0.81 and 0.81, respectively (Fig. 13a). This finding is directly consistent with classical adsorption theory: the Langmuir and Freundlich isotherms both predict strong dependence on equilibrium concentration (C_e), while maximum adsorption capacity (q_{max}) scales with available surface area. Solution pH ranked third (mean importance = 0.56, 15/15 studies), reflecting its well-established role in governing adsorbent surface charge (relative to pH_{pzc}) and contaminant speciation. Adsorbent dosage (0.55, 12/15 studies), pore volume (0.47, 12/15 studies), and temperature (0.31, 14/15 studies) constituted the remaining universally important features.

Beyond universal features, the synthesis revealed system-specific predictors that carry mechanistic information (Fig. 13). For biochar-based adsorption, elemental ratios (H/C, O/C, O + N/C) and pyrolysis temperature emerged as important predictors: features not captured by classical isotherm models but directly related to surface aromaticity and functional group density. For PFAS adsorption, contaminant molecular weight and perfluoroalkyl chain length were highly ranked [40, 27, 33], reflecting the dominance of hydrophobic interactions and size-dependent partitioning. For soil/sediment systems, organic matter content and clay content outranked surface area [49], consistent with the known role of soil organic matter in contaminant sorption. When categorized by feature type, contaminant properties (molecular weight, log K_{ow}) showed the highest mean importance among system-specific features, suggesting that future ML models should systematically incorporate molecular descriptors alongside traditional physicochemical parameters.

The strong alignment between ML-derived feature importance and classical adsorption theory validates the physical plausibility of these data-driven models. However, the synthesis also highlights that ML can extract insights beyond classical frameworks (particularly the importance of elemental ratios and molecular descriptors), suggesting a complementary role for explainable ML in hypothesis generation for adsorption mechanism research.

3.7 Machine-Extractable Reporting

To characterize how readily methodological detail can be recovered from this literature at scale, each study was scored on eight binary indicators (Table S5): whether a value could be extracted for (1) R^2, (2) RMSE, (3) MAE, (4) dataset size, (5) validation methodology, (6) number of input features, (7) software or tool used, and (8) multi-algorithm comparison (Fig. 14). The score measures what the extraction recovered, which the audit in Section 2.6.2 shows is not the same as what the studies contain: populated cells were 93.3% accurate, but 35 of 90 empty numeric and methodological cells in the audited sample should have carried a value. Each rate below is therefore a lower bound on reporting. The mean score across the 202 studies was 2.97 of 8.

Recovery rates differed sharply between indicators (Fig. 14a). Software disclosure was the most readily extracted (61.9%), followed by R^2, recovered for 110 of 202 studies (54.5%), and multi-algorithm comparison (108 studies, 53.5%). Methodological detail was far less accessible: validation methodology was recovered for 45.5% of studies, dataset size for 38.6%, RMSE for 25.7%, number of input features for 9.4%, and MAE for 7.9%. The gradient itself is the finding. Fields that a study reports in a fixed, labelled place, such as software and headline R^2, are recoverable; those reported in prose, in figure captions, or not at all are largely lost. Validation methodology is the clearest case: the audit found that 90% of studies do describe a scheme, against the 45.5% recovered here. Whatever these studies contain, it cannot at present be read out of them at the scale a synthesis requires.

The score improved modestly over time (Fig. 14b; Table S8; Spearman $\rho = 0.202$, $p = 0.004$, $n = 202$). Restricted to the 2019–2025 window shown in Fig. 14b, where all but eight of the studies lie, the same test gives $\rho = 0.157$ and $p = 0.029$. The trend is positive and reaches conventional significance in both windows, but the effect is small: a rank correlation of this size accounts for well under a tenth of the variance in the score, and the practical gain over six years is less than one indicator of eight. The mean score rose from 2.69 in 2019–2021 to 3.30 in 2024–2025, driven by the adoption of multi-algorithm comparison and more consistent R^2 reporting. Even in 2025 it remains below 4 of 8. Standardized placement, rather than more disclosure alone, is what would raise it, and that is what the Minimum Reporting Checklist for Machine Learning in Adsorption Studies (MRCAS) proposed here is designed to supply: a 22-item checklist organized into five categories covering data description, model specification, validation protocol, performance metrics, and interpretability or reproducibility (Table S15). Analogous to the PRISMA guidelines for systematic reviews and the TRIPOD statement for clinical prediction models, MRCAS gives each item a fixed reporting slot, so that a future synthesis can recover it without reading every paper in full.

3.8 Explainable AI and the Distance to Deployment

The preceding sections establish which algorithms are used and how well they are reported to perform. Neither question is the one a utility engineer would ask, which is whether any of these models could be relied upon outside the laboratory that produced it. Two conditions bear on that: the model must be interpretable enough to be trusted when it is wrong, and it must have been shown to work on water resembling the water it would treat. The reviewed studies speak to both, and in each case the answer is that the field is earlier in its development than the reported R^2 values suggest.

Interpretability is advancing but remains a minority practice. Fifteen studies (7.4%) apply SHAP, and their cross-study synthesis in Section 3.6 is encouraging: the features that emerge as most influential, initial contaminant concentration and BET surface area, are the ones classical adsorption theory would nominate, which is a meaningful, if modest, validation that these models are learning adsorption rather than the idiosyncrasies of a particular experiment. Its limits should be equally clear. A consensus drawn from 15 studies is a consensus among the small subset of authors who chose to look, and those authors were disproportionately working with tree ensembles, for which SHAP is computationally convenient. Whether the same features dominate in the 48 studies whose best model was a neural network is not established by these studies, and the agreement with theory does not by itself license using any of these models predictively.

Deployment is the larger gap. The water matrix was recorded for 132 studies (65%), and among these at least 46 (23% of all studies) report working with a real or non-synthetic matrix (wastewater, groundwater, drinking water, surface water, urine or tap water), while 86 (42%) report synthetic solutions alone. These counts should be read as lower bounds on real-matrix work, given the extraction audit's finding that empty cells

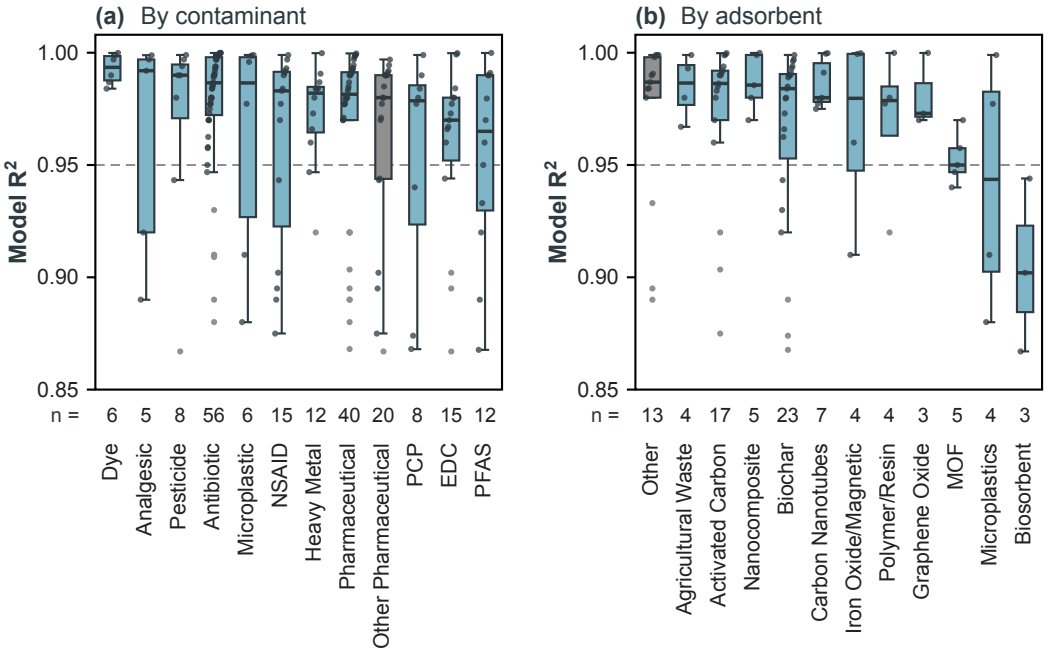


Figure 11. Subgroup analysis of model performance (R^2) across (a) emerging contaminant categories and (b) adsorbent material categories. Box plots show medians, interquartile ranges, and individual data points. Only categories with ≥ 3 studies are shown. The dashed line indicates $R^2 = 0.95$.

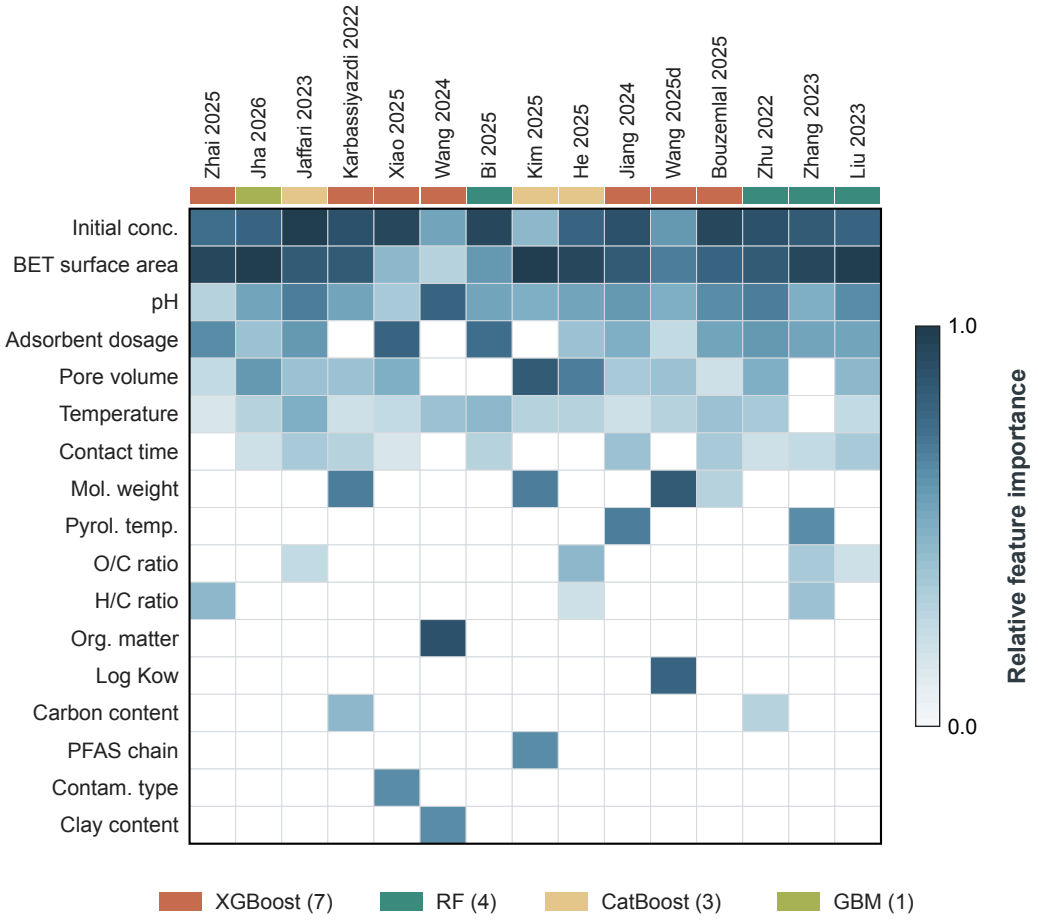


Figure 12. Cross-study synthesis of SHAP-based feature importance from 15 explainable ML studies. Columns are studies (first author, year); the strip above each column marks the algorithm. Rows are input features. Colour intensity is relative importance (0–1); white cells mark features not used. Seven features used by a single study with importance below 0.5 are omitted. Abbreviations: conc., concentration; contam., contaminant; mol., molecular; org., organic; pyrol. temp., pyrolysis temperature; PFAS chain, perfluoroalkyl chain length.

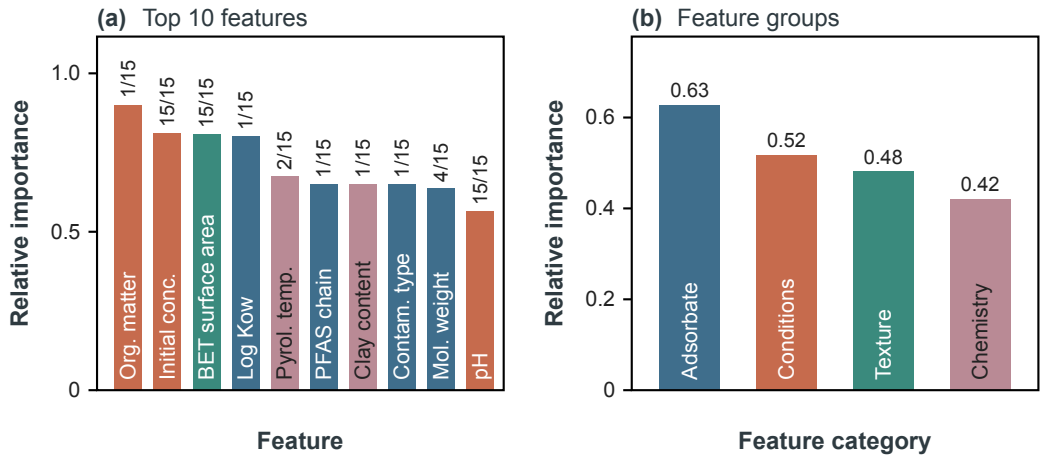


Figure 13. Consensus feature importance ranking from 15 SHAP-enabled studies: (a) mean relative importance of the top 10 features, with the number of studies in which each feature appeared shown as fractions above the bars and bar colour marking the feature category in (b); (b) mean importance aggregated by feature category (adsorbate properties, operating conditions, adsorbent texture, adsorbent chemistry). Features are ranked by their cross-study mean relative importance score. Abbreviations as in Fig. 12.

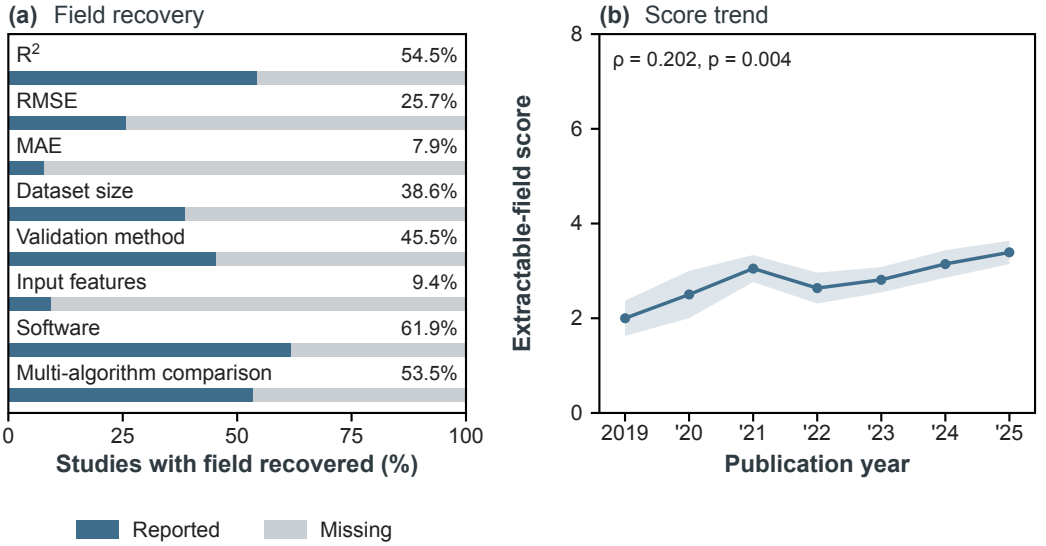


Figure 14. Machine-extractable reporting across the 202 reviewed studies. (a) Share of studies for which each of eight methodological fields could be extracted (blue = recovered, grey = not). (b) Mean extractable-field score (of 8) by year, with standard-error bands; the number of studies per year is n = 9, 6, 20, 22, 32, 41 and 64 for 2019–2025. Rates are lower bounds on reporting (Section 2.6.2).

substantially reflect extraction failure rather than silence. Even read generously, the picture is one in which most models are trained and tested on single-contaminant solutions in deionised water, whereas the competitive adsorption, natural organic matter and variable ionic strength of a real influent are precisely what would degrade their predictions. That concern compounds with the validation practice documented in Section 3.5: a model fitted and evaluated on random splits of one synthetic dataset has not been tested against any of the conditions that deployment would impose on it.

Three further reporting deficits stand between this literature and reuse of its models by others. Software is identified in 125 studies (61.9%), with MATLAB the most common environment and Python-based stacks a growing minority, but the extracted dataset contains no field for code or data availability, so the number of studies releasing either cannot be stated here. Establishing it would require a separate full-text pass. Only 11 studies (5.4%) report both a dataset size and a number of input features, so the sample-to-feature ratio, the single most direct indicator of whether a model had enough data to support its complexity, is computable for one study in twenty. Among those 11 the median ratio is 47:1, with two studies below 10:1. And no reviewed study reports a prospective test, in which a model trained on one adsorbent–contaminant system predicts an experiment not yet performed. Until such tests become routine, the performance figures synthesised here should be read as descriptions of model fit, not as evidence of predictive capability in service.

3.9 Research Gaps and Future Directions

The synthesis of the 202 reviewed studies reveals several critical gaps that should guide future research in ML-assisted adsorption of emerging contaminants (Fig. 15).

3.9.1 Contaminant and material diversity.

The current literature is heavily skewed toward pharmaceutical compounds, which account for 81.7% of all studies, while PFAS (9.4%), microplastics (5.9%), and pesticides (6.4%) remain substantially underrepresented. The top 10 individual contaminants appear in over 80% of studies, leaving hundreds of environmentally relevant emerging pollutants unstudied by ML approaches. On the adsorbent side, carbon-

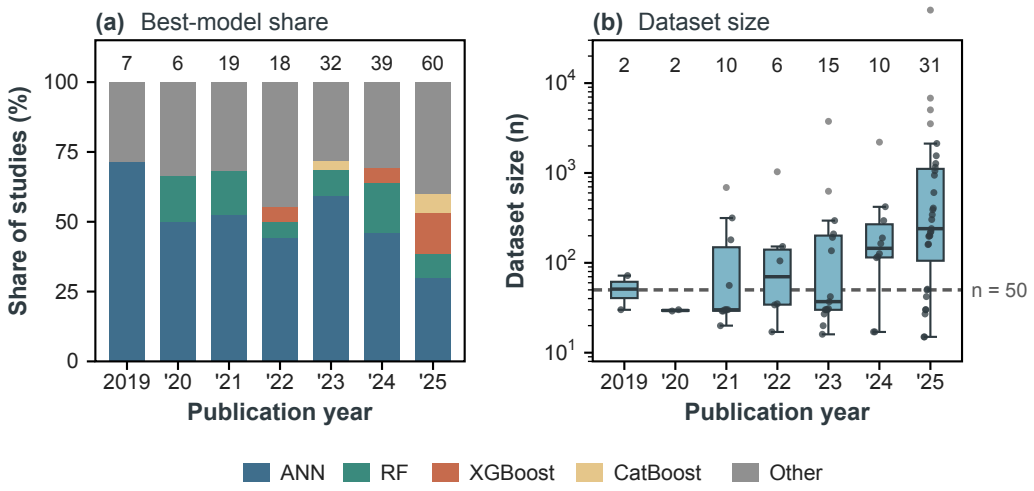


Figure 15. Research gaps across the 202 reviewed studies. (a) Best-performing model composition by year. (b) Dataset size by year (log scale); dashed line marks $n = 50$. Numbers above each bar or box give the number of studies reporting that metric.

based materials (biochar, activated carbon, CNTs, graphene oxide) dominate at 46.5%, whereas high-performance materials such as MOFs (e.g. Refs. [36, 50, 40]), LDHs (e.g. Refs. [51, 44, 52]), and engineered nanocomposites (e.g. Refs. [42, 53, 54]) have received limited ML attention despite their growing prominence in adsorption research. Future studies should prioritize ML modeling for PFAS, microplastics, and multi-contaminant systems using a broader range of advanced adsorbents.

3.9.2 Data quantity and quality.

Perhaps the most pressing concern is the prevalence of small datasets (Fig. 15b): 40% of studies that reported dataset size worked with fewer than 50 samples, and the median was only 131. Such limited training data raises questions about overfitting and generalizability, particularly for complex models like deep neural networks [47, 48]. Exceptions exist among studies that compiled large multi-source datasets (e.g. Refs. [11, 30, 35]), showing that aggregating published adsorption data across studies can yield training sets above 1,000 samples. Shared, open-access adsorption databases, analogous to those available in drug discovery or materials science, would accelerate that aggregation. A further 9.4% of studies reported the number of input features, which limits assessment of model complexity relative to dataset size.

3.9.3 Methodological rigor.

Standardized reporting of performance metrics remains inconsistent: R^2 was recovered for 54.5% of studies, RMSE for 25.7%, and MAE for only 7.9%. The uniformly high reported R^2 values (78% $\geq$ 0.95) likely reflect publication bias and selective reporting of best-case results rather than the true predictive capability of these models on unseen data. Multi-algorithm comparison studies that benchmark several models on the same dataset (e.g. Refs. [11, 27, 55]) provide more reliable performance assessments, whereas single-algorithm studies, which constitute 45.5% of the reviewed set, offer limited insight into relative model suitability. Future studies should adopt rigorous validation protocols, including k-fold cross-validation, external test sets, and temporal validation, and report multiple metrics to enable meaningful cross-study comparison. Standardized reporting frameworks, such as those proposed for ML in chemistry, would make that comparison possible at the scale of a review.

3.9.4 Algorithm innovation and frontier roadmap.

While 59 unique algorithms were identified, ANN alone accounted for 64.9% of studies, although its share as the best-performing model has declined from 71% in 2019 to 32% in 2025. That comparison should be read with its base in view: 2019 contributed seven studies with a named best model, so 71% is five of seven, whereas the 2025 figure rests on 57 (Fig. 15a). More recent ML architectures, including graph neural networks (GNNs), transformers, and physics-informed neural networks (PINNs), were virtually absent from the reviewed literature, despite their demonstrated potential in related domains such as molecular property prediction and process modeling. Among deep learning approaches, only scattered examples of CNN [56, 57, 31], RNN [58], LSTM [59, 28], and DNN [47] were identified, while a crystal graph convolutional neural network (CGCNN) was applied in a single PFAS study [35]. The rapid adoption of XGBoost (from 0 studies before 2022 to 15 in 2025) and CatBoost (e.g. Refs. [11, 27, 40]) shows that the field is receptive to new methods, motivating a systematic assessment of emerging architectures.

Table 3 presents a technology readiness assessment for six frontier ML paradigms with high potential for adsorption modeling. Physics-informed neural networks (PINNs) can embed thermodynamic constraints (e.g., Gibbs free energy, mass balance) directly into the loss function, ensuring physically plausible predictions even with limited data, a critical advantage given that 40% of current studies use fewer than 50 samples. Graph neural networks (GNNs) can represent molecular structures and porous material topologies as graphs, enabling structure–property predictions for novel adsorbents without requiring hand-crafted descriptors; the sole CGCNN application to PFAS sorption [35] demonstrated this potential. Transfer learning offers a practical solution to the small-dataset problem by pre-training models on large related datasets (e.g., heavy metal adsorption, $n > 10{,}000$ in existing databases) and fine-tuning on target EC systems. Generative models (VAE, diffusion models) could accelerate adsorbent design by proposing novel material compositions optimized for specific contaminants, while Bayesian approaches provide uncertainty quantification essential for risk-informed decision-making in water treatment. Each paradigm addresses a specific limitation identified in this review, and their systematic evaluation should be a priority for future research.

3.9.5 Interpretability and mechanistic insight.

Only 7.4% of studies employed model interpretability tools such as SHAP (SHapley Additive exPlanations), which leaves most ML models in this field functioning as "black boxes." Studies employing ensemble methods with built-in feature importance analysis (e.g. Refs. [60–62]) provide partial interpretability, but systematic application of post-hoc explainability methods remains uncommon. Integrating explainability methods, including SHAP, LIME, partial dependence plots, and attention mechanisms, would increase trust in model predictions and generate mechanistic hypotheses about adsorption behavior. Coupling ML predictions with physicochemical understanding (e.g., surface functional groups, pore size distributions, electrostatic interactions) points toward physics-informed ML models, as exemplified by the QSAR/LFER-based approaches of Cho et al. [63, 37] and Lee et al. [38], which embed molecular descriptors into predictive frameworks.

Table 3. Frontier ML technologies for adsorption modeling. TRL = Technology Readiness Level adapted for ML research (1 = concept only, 5 = validated in target domain).

Technology	TRL	Impact	Adsorption Application	Key Challenge
Physics-Informed NN (PINN)	2	High	Embed isotherm/kinetic equations as constraints; enforce thermodynamic consistency	Formulating differentiable physics losses for heterogeneous systems
Graph Neural Networks (GNN)	2	High	Structure–property prediction for MOFs, polymers; molecular fingerprinting of ECs	Representing amorphous materials (biochar, AC); limited training data
Transfer Learning	1	High	Pre-train on heavy metal/dye datasets ($n > 10{,}000$); fine-tune on EC systems	Domain shift between pollutant classes; negative transfer risk
Generative Models (VAE/Diffusion)	1	Medium	Inverse design of optimal adsorbent compositions for target contaminants	Synthesizability constraints; experimental validation loop
Bayesian ML	3	Medium	Uncertainty-aware predictions; active learning for experimental design	Computational cost; scalability to large feature spaces
Transformer / LLM	1	Medium	Multi-modal input (text + numeric); literature mining for automated data extraction	Hallucination risk; interpretability; data formatting

3.9.6 Geographic and collaborative gaps.

Research output is concentrated in a small number of countries, with China, India, Iran, and South Korea contributing 58.3% of all studies. Regions facing acute emerging contaminant challenges (including Sub-Saharan Africa [64–66], Southeast Asia [67], and Latin America [68–71]) are underrepresented despite producing valuable individual studies. International collaborative networks and capacity-building initiatives could help democratize access to ML tools for water treatment research. Furthermore, the development of region-specific models trained on local water matrices and indigenous adsorbent materials would enhance practical applicability.

3.9.7 From prediction to application.

The vast majority of reviewed studies focus on prediction accuracy as the primary objective, with limited attention to practical deployment. Studies that couple ML with response surface methodology for process optimization (e.g. Refs. [72–74]) or employ multi-objective optimization (e.g. Refs. [75–77]) take steps toward application-oriented modeling. Stacking and ensemble strategies designed for generalization [78–80] and Bayesian approaches that quantify prediction uncertainty [81, 82] offer pathways beyond point-estimate predictions. Future research should bridge this gap by developing ML-assisted decision-support tools for adsorbent selection (e.g. Refs. [83, 84]), real-time process optimization, and multi-objective optimization that balances removal efficiency, cost, and environmental impact. Integrating ML models with life cycle assessment (LCA) and techno-economic analysis would further support the translation of laboratory findings into scalable water treatment solutions.

3.10 Limitations of This Review

Five limitations bound what can be concluded from this synthesis, and one of them we regard as material.

Database coverage. The search drew on Scopus and PubMed only. Web of Science, IEEE Xplore and Engineering Village were not searched. The direction of the resulting bias can be reasoned about, though its size cannot be measured. Scopus and PubMed between them index the environmental-engineering and water-treatment journals in which adsorption work overwhelmingly appears, so studies motivated by a contaminant and an adsorbent are unlikely to have been missed systematically. The exposure is at the other end of the subject, among ML-motivated work published in computer-science and electrical-engineering venues, where IEEE Xplore has coverage the two databases used here do not. A study of that kind (an algorithmic contribution demonstrated on an adsorption dataset, published at an engineering venue) is precisely the kind this review would have failed to retrieve. No supplementary search was run to bound this, so the number of such studies is unknown. The literature reviewed here should be read as representative of the environmental-science work on ML for EC adsorption rather than of all literature in which such models appear. The consequence bears asymmetrically on the findings: the bibliometric and contaminant–adsorbent mapping would be little changed by the omission, whereas the algorithm-frequency picture, and in particular the dominance of ANN over more recent architectures, may understate methods favoured in computational venues.

Registration. The review was not prospectively registered and no protocol was published in advance, as set out in Section 2.2. The eligibility criteria and search query were fixed before the searches ran and were not amended, but this is a claim about our records rather than an independently verifiable commitment.

No study-level quality appraisal. The included studies were not appraised for risk of bias, because no validated instrument exists for this study design and the instruments built for clinical prediction models do not transfer to it (Section 2.7). The consequence is that a study whose reported R^2 rests on a weak validation design contributes to the pooled estimates on the same footing as one whose design is sound. Reading 30 studies in full establishes that the weakness is close to universal rather than confined to a subset, which is why we report it as a property of the literature in Section 3.5 rather than as a score that discriminates between studies, but a synthesis that could weight studies by appraised quality would be stronger than this one.

Automated screening and extraction. Screening was performed by deterministic rule-based scripts and extraction by a regex pass reconciled against a language-model pass, with the accuracy of both measured by audit (Sections 2.6.1 and 2.6.2). The audits found no false exclusions in 101 sampled records and 93.3% accuracy in populated extraction cells, but they were conducted by the authors rather than by an independent assessor, they sample one screening stage rather than all three, and they establish that sparsely populated fields are unreliable as evidence of what studies do not report. Claims in this review that depend on such fields have been confined to the audited subset accordingly.

Search cut-off and the moving target. The searches ran on 24 February 2026 against a window closing at the end of 2025. Coverage of 2025 is therefore a lower bound, and work published in 2026 (including at least three reviews of adjacent or identical scope, as Table 1 records) falls outside the reviewed period by construction. This is inherent to any systematic review with a fixed cut-off, but in a field growing at the rate documented in Section 3.1 the interval between cut-off and publication carries more consequence than it would elsewhere.

4. CONCLUSIONS

This systematic review analyzed 202 peer-reviewed studies (2015–2025) on machine learning for adsorption-based removal of emerging contaminants from water, and is, to our knowledge, the first review in this domain to combine a stated search and screening protocol with a quantitative meta-analysis of reported performance and a cross-study synthesis of feature importance. The field has grown rapidly, with 179 of the 202 studies (88.6%) published from 2021 onward. Among 59 unique algorithms, ANN dominated (64.9%), though ensemble methods are rapidly gaining adoption; inverse-variance weighted meta-analysis revealed pooled R^2 values of 0.998 (ANFIS), 0.995 (ANN), 0.979 (XGBoost), and 0.955 (RF), with extremely high heterogeneity ($I^2 > 98\%$) across all algorithms, precluding universal algorithm recommendations. A funnel plot analysis revealed patterns consistent with publication bias toward favorable results.

Cross-study synthesis of SHAP-based feature importance from 15 explainable ML studies identified initial concentration, BET surface area,

and pH as universally important predictors, findings that align with classical Langmuir/Freundlich theory, while also revealing system-specific drivers (elemental ratios for biochar, molecular weight for PFAS) that extend beyond traditional frameworks. However, critical gaps persist: antibiotics, pharmaceuticals, NSAIDs and analgesics account for 81.7% of target contaminants while PFAS (9.4%) and microplastics (5.9%) remain underrepresented; 40% of the studies that report a dataset size used fewer than 50 data points; and validation, though nearly always performed, is nearly always an internal split of a single dataset rather than a test against independent data. The proposed MRCAS checklist (22 items across 5 categories) provides a practical tool to standardize future reporting.

To advance from prediction-focused modeling toward actionable decision-support tools, we recommend five priority directions: (1) diversifying contaminant coverage toward PFAS, microplastics, and multi-contaminant mixtures; (2) developing shared open-access adsorption databases to address the small-dataset bottleneck; (3) adopting the MRCAS reporting framework with rigorous external validation; (4) systematically evaluating frontier architectures (particularly physics-informed neural networks and graph neural networks) that can embed domain knowledge and handle molecular structures; and (5) expanding the use of explainable ML to bridge data-driven predictions with mechanistic understanding of adsorption processes.

DECLARATION OF AI-ASSISTED TECHNOLOGIES

A large language model was used as a research instrument in this review, not as a writing aid. Claude Opus 4.6 (Anthropic) performed the second data-extraction pass described in Section 2.5, operating only on text taken from each included paper's own PDF and returning values into a fixed field schema. No language model was involved in screening, in statistical analysis, or in deciding which studies to include. The accuracy of the model's output was measured against the source PDFs and is reported, together with the errors it made, in Section 2.6.2. The authors reviewed that output, take full responsibility for the content of this article, and report this use in line with the 2025 joint position statement on AI use in evidence synthesis [25].

CRediT AUTHORSHIP CONTRIBUTION STATEMENT

Tarmizi Taher: Conceptualization, Data curation, Formal analysis, Funding acquisition, Methodology, Project administration, Software, Writing – original draft. **Sephia Amanda Muhtar:** Data curation, Investigation. **Dian Ahmad Hapidin:** Supervision, Writing – review & editing. **Aditya Rianjanu:** Data curation, Formal analysis, Methodology, Writing – original draft. **Khairurrijal Khairurrijal:** Data curation, Writing – review & editing.

DECLARATION OF COMPETING INTEREST

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

ACKNOWLEDGEMENTS

This research was supported by the Hibah Penugasan Institut Teknologi Sumatera (ITERA) 2024.

DATA AVAILABILITY

The complete extracted dataset containing all 202 studies and 40 extracted fields is provided as Supplementary Data. Additional details on the search strategy, screening protocol, and data extraction are available in the Supplementary Information. The screening and extraction scripts, the per-record decision logs from both screeners and the adjudication step, and the complete cell-level records of the screening and extraction audits reported in Sections 2.6.1 and 2.6.2 are available from the corresponding author on request.

REFERENCES

[1] R. P. Schwarzenbach, B. I. Escher, K. Fenner, T. B. Hofstetter, C. A. Johnson, U. von Gunten, B. Wehrli, The challenge of micropollutants in aquatic systems, Science 313 (5790) (2006) 1072–1077. doi:10.1126/science.1127291.

[2] R. Loos, R. Carvalho, D. C. António, S. Comero, G. Locoro, S. Tavazzi, B. Paracchini, M. Ghiani, T. Lettieri, L. Blaha, B. Jarosova, S. Voorspoels, K. Servaes, P. Haglund, J. Fick, R. H. Lindberg, D. Schwesig, B. M. Gawlik, EU-wide monitoring survey on emerging polar organic contaminants in wastewater treatment plant effluents, Water Research 47 (17) (2013) 6475–6487. doi:10.1016/j.watres.2013.08.024.

[3] S. D. Richardson, T. A. Ternes, Water analysis: Emerging contaminants and current issues, Analytical Chemistry 94 (1) (2022) 382–416. doi:10.1021/acs.analchem.1c04640.

[4] Y. Luo, W. Guo, H. H. Ngo, L. D. Nghiem, F. I. Hai, J. Zhang, S. Liang, X. C. Wang, A review on the occurrence of micropollutants in the aquatic environment and their fate and removal during wastewater treatment, Science of The Total Environment 473-474 (2014) 619–641. doi:10.1016/j.scitotenv.2013.12.065.

[5] M. B. Ahmed, J. L. Zhou, H. H. Ngo, W. Guo, Adsorptive removal of antibiotics from water and wastewater: Progress and challenges, Science of The Total Environment 532 (2015) 112–126. doi:10.1016/j.scitotenv.2015.05.130.

[6] S. De Gisi, G. Lofrano, M. Grassi, M. Notarnicola, Characteristics and adsorption capacities of low-cost sorbents for wastewater treatment: A review, Sustainable Materials and Technologies 9 (2016) 10–40. doi:10.1016/j.susmat.2016.06.002.

[7] J. O. Ighalo, S. Rangabhashiyam, K. Dulta, C. T. Umeh, K. O. Iwuozor, C. O. Aniagor, S. O. Eshiemogie, F. U. Iwuchukwu, C. A. Igwegbe, Recent advances in hydrochar application for the adsorptive removal of wastewater pollutants, Chemical Engineering Research and Design 184 (2022) 419–456. doi:10.1016/j.cherd.2022.06.028.

[8] M. I. Jordan, T. M. Mitchell, Machine learning: Trends, perspectives, and prospects, Science 349 (6245) (2015) 255–260. doi:10.1126/science.aaa8415.

[9] S. Zhong, K. Zhang, M. Bagheri, J. G. Burken, A. Gu, B. Li, X. Ma, B. L. Marrone, Z. J. Ren, J. Schrier, W. Shi, H. Tan, T. Wang, X. Wang, B. M. Wong, X. Xiao, X. Yu, J.-J. Zhu, H. Zhang, Machine learning: New ideas and tools in environmental science and engineering, Environmental Science & Technology 55 (2021) 12741–12754. doi:10.1021/acs.est.1c01339.

[10] X. Zhu, X. Wang, Y. S. Ok, The application of machine learning methods for prediction of metal sorption onto biochars, Journal of Hazardous Materials 378 (2019) 120727. doi:10.1016/j.jhazmat.2019.06.004.

[11] Z. Haider Jaffari, H. Jeong, J. Shin, J. Kwak, C. Son, Y.-G. Lee, S. Kim, K. Chon, K. Hwa Cho, Machine-learning-based prediction and optimization of emerging contaminants' adsorption capacity on biochar materials, Chemical Engineering Journal 466 (2023) 143073. doi:10.1016/j.cej.2023.143073.

[12] W. Zhang, R. Chen, J. Li, T. Huang, B. Wu, J. Ma, Q. Wen, J. Tan, W. Huang, Synthesis optimization and adsorption modeling of biochar for pollutant removal via machine learning, Biochar 5 (1) (2023). doi:10.1007/s42773-023-00225-x.

[13] Z. Wang, Q. Wang, F. Yang, C. Wang, M. Yang, J. Yu, How machine learning boosts the understanding of organic pollutant adsorption on carbonaceous materials: A comprehensive review with statistical insights, Separation and Purification Technology 350 (127790) (2024) 127790. doi:10.1016/j.seppur.2024.127790.

[14] V. G. Sharmila, M. D. Kumar, K. Tamilarasan, Machine learning-driven advances in metal-organic framework nanomaterials for wastewater treatment: Developments and challenges, Separation & Purification Reviews 55 (1) (2024) 35–55. doi:10.1080/15422119.2024.2437408.

[15] S. S. Fiyadh, S. M. Alardhi, M. Al Omar, M. M. Aljumaily, M. A. Al Saadi, S. S. Fayaed, S. N. Ahmed, A. D. Salman, A. H. Abdalsalm, N. M. Jabbar, A. El-Shafi, A comprehensive review on modelling the adsorption process for heavy metal removal from waste water using artificial neural network technique, Heliyon 9 (4) (2023) e15455. doi:10.1016/j.heliyon.2023.e15455.

[16] X. Yuan, J. Li, J. Y. Lim, A. Zolfaghari, D. S. Alessi, Y. Wang, X. Wang, Y. S. Ok, Machine learning for heavy metal removal from water: Recent advances and challenges, ACS ES&T Water 4 (3) (2023) 820–836. doi:10.1021/acsestwater.3c00215.

[17] S. Zhao, J. Guo, Y. Tang, Y. Zhou, Applications of machine learning in heavy metal adsorption modeling: A review, Separation and Purification Technology 377 (134168) (2025) 134168. doi:10.1016/j.seppur.2025.134168.

[18] H. E. Reynel-Ávila, I. A. Aguayo-Villarreal, L. L. Diaz-Muñoz, J. Moreno-Pérez, F. J. Sánchez-Ruiz, C. K. Rojas-Mayorga, D. I. Mendoza-Castillo, A. Bonilla-Petriciolet, A review of the modeling of adsorption of organic and inorganic pollutants from water using artificial neural networks, Adsorption Science & Technology 2022 (9384871) (2022). doi:10.1155/2022/9384871.

[19] W. Zhang, W. Huang, J. Tan, D. Huang, J. Ma, B. Wu, Modeling, optimization and understanding of adsorption process for pollutant removal via machine learning: Recent progress and future perspectives, Chemosphere 311 (137044) (2023) 137044. doi:10.1016/j.chemosphere.2022.137044.

[20] L. Yuan, S. Li, Y. Ma, H. Zhang, C. Chen, W. Zhang, Y. Zhang, M. Hua, L. Lv, B. Pan, Machine learning-aided design of adapted water treatment adsorbents: Advances, challenges, and opportunities, Chemical Engineering Journal 526 (171262) (2025) 171262. doi:10.1016/j.cej.2025.171262.

[21] Y. Fan, J. Li, Y. Ren, C. Liu, W. Zhang, Where machine learning fails in predicting emerging contaminant adsorption: A decision-oriented framework for model credibility and transferability, Environmental Science & Technology 60 (27) (2026) 18954–18974. doi:10.1021/acs.est.6c00882.

[22] A. A. Elngar, A. M. Ahmed, A. Saad, Artificial intelligence models for adsorption-based wastewater treatment: a critical review toward environmental sustainability, International Journal of Environmental Science and Technology 23 (6) (2026). doi:10.1007/s13762-026-07282-2.

[23] S. Yaghoobian, J. An, D.-W. Jeong, J.-H. Hwang, Towards smart PFAS management: Integrating artificial intelligence in water and wastewater systems, Journal of Hazardous Materials 502 (140934) (2026) 140934. doi:10.1016/j.jhazmat.2025.140934.

[24] M. E. Mallahi, E. Akichouh, A. Elyoussfi, A. Salhi, M. Ahari, H. Amhamdi, M. Ghalit, F. Mourabit, Artificial intelligence in adsorption process optimization: A 2014–2024 bibliometric analysis of research trends and developments, Results in Engineering 28 (107777) (2025) 107777. doi:10.1016/j.rineng.2025.107777.

[25] E. Flemyng, A. Noel-Storr, B. Macura, G. Gartlehner, J. Thomas, J. J. Meerpohl, Z. Jordan, J. Minx, A. Eisele-Metzger, C. Hamel, P. Jemioło, K. Porritt, M. Grainger, Position statement on artificial intelligence (AI) use in evidence synthesis across Cochrane, the Campbell Collaboration, JBI and the Collaboration for Environmental Evidence 2025, Environmental Evidence 14 (2025) 20. doi:10.1186/s13750-025-00374-5.

[26] H. E. Reynel-Avila, D. I. Mendoza-Castillo, A. Bonilla-Petriciolet, J. Silvestre-Albero, Assessment of naproxen adsorption on bone char in aqueous solutions using batch and fixed-bed processes, Journal of Molecular Liquids 209 (2015) 187–195. doi:10.1016/j.molliq.2015.05.013.

[27] J. He, B. Xiong, M. Ren, Z. Dang, C. Guo, Prediction of removal performance of per- and polyfluoroalkyl substances by biochar based on machine learning models, Journal of Environmental Chemical Engineering 13 (6) (2025) 119989. doi:10.1016/j.jece.2025.119989.

[28] P. E. Ovuoraye, V. I. Ugonabo, E. Fetahi, A. Chowdhury, M. A. Tahir, C. A. Igwegbe, M. H. Dehghani, Machine learning algorithm and neural network architecture for optimization of pharmaceutical and drug manufacturing industrial effluent treatment using activated carbon derived from breadfruit (Treculia africana), Journal of Engineering and Applied Science 70 (1) (2023). doi:10.1186/s44147-023-00307-4.

[29] G. M. Kubar, A. A. Kubar, K. A. Kubar, K. Shahzad, Biochar-based water remediation: A machine learning approach to antibiotic adsorption prediction, Water, Air, & Soil Pollution 236 (14) (2025). doi:10.1007/s11270-025-08651-2.

[30] Y. Xiao, Z. Zhang, J. Lin, W. Chen, J. Huang, Z. Chen, Machine learning predicts selectivity of green synthesized iron nanoparticles toward typical contaminants: critical factors in synthesis conditions, material properties, and reaction process, Environmental Research 277 (2025) 121605. doi:10.1016/j.envres.2025.121605.

[31] K. Tian, C. Li, L. Wang, SAP nanomicelle-functionalized biochar: A multifunctional and sustainable adsorbent for efficient removal of dyes and heavy metals, ACS Omega 10 (50) (2025) 61452–61470. doi:10.1021/acsomega.5c06590.

[32] B. Beigzadeh, M. Bahrami, M. J. Amiri, M. R. Mahmoudi, A new approach in adsorption modeling using random forest regression, Bayesian multiple linear regression, and multiple linear regression: 2,4-D adsorption by a green adsorbent, Water Science and Technology 82 (8) (2020) 1586–1602. doi:10.2166/wst.2020.440.

[33] E. Karbassiyazdi, F. Fattahi, N. Yousefi, A. Tahmassebi, A. A. Taromi, J. Z. Manzari, A. H. Gandomi, A. Altaee, A. Razmjou, XGBoost model as an efficient machine learning approach for PFAS removal: Effects of material characteristics and operation conditions, Environmental Research 215 (2022) 114286. doi:10.1016/j.envres.2022.114286.

[34] N. Huang, K. Gao, W. Yang, H. Pang, G. Yang, J. Wu, S. Zhang, C. Chen, L. Long, Assessing sediment organic pollution via machine learning models and resource performance, Bioresource Technology 361 (2022) 127710. doi:10.1016/j.biortech.2022.127710.

[35] M. Zhang, T. Ma, M. Ozlek, A. O. Yazaydin, Machine learning-assisted exploration of covalent organic frameworks for short-chain per- and polyfluoroalkyl substances (PFAS) removal from water, Journal of Colloid and Interface Science 702 (2025) 138970. doi:10.1016/j.jcis.2025.138970.

[36] S. Edebali, Synthesis and characterization of MIL-101 (Fe) as efficient catalyst for tetracycline degradation by using NaBH4: Artificial neural network modeling, Applied Surface Science Advances 18 (2023) 100496. doi:10.1016/j.apsadv.2023.100496.

[37] B.-G. Cho, S.-B. Mun, C.-R. Lim, S. B. Kang, C.-W. Cho, Y.-S. Yun, Adsorption modeling of microcrystalline cellulose for pharmaceutical-based micropollutants, Journal of Hazardous Materials 426 (2022) 128087. doi:10.1016/j.jhazmat.2021.128087.

[38] K.-Y. Lee, B.-G. Cho, S.-R. Jin, S.-H. Park, J.-M. Cheon, C.-W. Cho, Experimental characterization and QSAR modeling of maximum uptake and distribution factor for neutral and ionic micropollutants on granular activated carbon in artificial and real wastewaters, Separation and Purification Technology 373 (2025) 133549. doi:10.1016/j.seppur.2025.133549.

[39] S.-B. Mun, B.-G. Cho, S.-R. Jin, C.-R. Lim, Y.-S. Yun, C.-W. Cho, Adsorption of organic micropollutants on yeast: Batch experiment and modeling, Journal of Environmental Management 334 (2023) 117507. doi:10.1016/j.jenvman.2023.117507.

[40] H. G. Kim, B.-M. Jun, H. Jeong, Y. Yoon, K. H. Cho, Machine learning-based optimization and interpretation of the adsorption capacities of metal-organic frameworks for endocrine-disrupting compounds, Journal of Water Process Engineering 80 (2025) 109105. doi:10.1016/j.jwpe.2025.109105.

[41] X. C. Nguyen, Q. V. Ly, T. T. H. Nguyen, H. T. T. Ngo, Y. Hu, Z. Zhang, Potential application of machine learning for exploring adsorption mechanisms of pharmaceuticals onto biochars, Chemosphere 287 (2022) 132203. doi:10.1016/j.chemosphere.2021.132203.

[42] Z. Gholami, M. Ahmadi Azqhandi, M. Hosseini Sabzevari, F. Khazali, Evaluation of least square support vector machine, generalized regression neural network and response surface methodology in modeling the removal of levofloxacin and ciprofloxacin from aqueous solutions using ionic liquid @graphene oxide@ ionic liquid NC, Alexandria Engineering Journal 73 (2023) 593–606. doi:10.1016/j.aej.2023.04.041.

[43] M. Yousefi, M. Gholami, V. Oskoei, A. A. Mohammadi, M. Baziar, A. Esrafili, Comparison of LSSVM and RSM in simulating the removal of ciprofloxacin from aqueous solutions using magnetization of functionalized multi-walled carbon nanotubes: Process optimization using GA and RSM techniques, Journal of Environmental Chemical Engineering 9 (4) (2021) 105677. doi:10.1016/j.jece.2021.105677.

[44] S. Tavassoli, A. Mollahosseini, S. Damiri, M. Samadi, Luffa–Ni/Al layered double hydroxide bio-nanocomposite for efficient ibuprofen removal from aqueous solution: Kinetic, equilibrium, thermodynamic studies and GEP modeling, Heliyon 11 (1) (2025) e40783. doi:10.1016/j.heliyon.2024.e40783.

[45] S. Hafeez, A. Ishaq, A. Intisar, T. Mahmood, M. I. Din, E. Ahmed, M. R. Tariq, M. A. Abid, Predictive modeling for the adsorptive and photocatalytic removal of phenolic contaminants from water using artificial neural networks, Heliyon 10 (19) (2024) e37951. doi:10.1016/j.heliyon.2024.e37951.

[46] D. K. Jha, Unveiling sulfonamide adsorption on biochar using explainable machine learning, Bioresource Technology 443 (2025) 133867. doi:10.1016/j.biortech.2025.133867.

[47] S. Nasrollahpour, A. Tanhadoust, R. Pulicharla, S. K. Brar, Long-chain perfluoroalkyl carboxylic acids removal by biochar: Experimental study and uncertainty based data-driven predictive model, iScience 27 (11) (2024) 111140. doi:10.1016/j.isci.2024.111140.

[48] S. Zhang, Y. Jin, W. Chen, J. Wang, Y. Wang, H. Ren, Artificial intelligence in wastewater treatment: A data-driven analysis of status and trends, Chemosphere 336 (2023) 139163. doi:10.1016/j.chemosphere.2023.139163.

[49] J. Wang, R. Huang, Y. Liang, X. Long, S. Wu, Z. Han, H. Liu, X. Huangfu, Prediction of antibiotic sorption in soil with machine learning and analysis of global antibiotic resistance risk, Journal of Hazardous Materials 466 (2024) 133563. doi:10.1016/j.jhazmat.2024.133563.

[50] N. Saeidi, A. Lai, F. Harnisch, G. Sigmund, A FAIR comparison of activated carbon, biochar, cyclodextrins, polymers, resins, and metal organic frameworks for the adsorption of per- and polyfluorinated substances, Chemical Engineering Journal 498 (2024) 155456. doi:10.1016/j.cej.2024.155456.

[51] S. Radmehr, M. Hosseini Sabzevari, M. Ghaedi, M. H. Ahmadi Azqhandi, F. Marahel, Adsorption of nalidixic acid antibiotic using a renewable adsorbent based on graphene oxide from simulated wastewater, Journal of Environmental Chemical Engineering 9 (5) (2021) 105975. doi:10.1016/j.jece.2021.105975.

[52] M. Omidi, M. A. Azqhandi, B. Ghalami-Choobar, Synthesis, characterization, and application of graphene oxide/layered double hydroxide /poly acrylic acid nanocomposite (LDH-rGO-PAA NC) for tetracycline removal: A comprehensive chemometric study, Chemosphere 308 (2022) 136007. doi:10.1016/j.chemosphere.2022.136007.

[53] O. Farzinmanesh, M. Hosseini Sabzevari, M. R. Asghariganjeh, Efficient removal of ciprofloxacin and ofloxacin from aqueous solutions using a novel nano-scale adsorbent: Modeling, optimization, and characterization, Chemosphere 354 (2024) 141640. doi:10.1016/j.chemosphere.2024.

141640.

[54] Q. Chen, Y. Liu, L. Yang, Q. Zeng, S. Ning, X. Wang, Y. Wei, D. Zeng, Hierarchical sheet-on-sheet Gd2O3/Bi2WxMo1-xO6 S-scheme photocatalyst for improved photocatalytic tetracycline degradation: Toxicity and machine learning insights, Journal of Cleaner Production 520 (2025) 146099. doi:10.1016/j.jclepro.2025.146099.

[55] Y. Ahmed, A. A. Siddiqua Maya, P. Akhtar, M. S. Alam, H. AlMohamadi, M. N. Islam, O. A. Alharbi, S. M. Rahman, A novel interpretable machine learning and metaheuristic-based protocol to predict and optimize ciprofloxacin antibiotic adsorption with nano-adsorbent, Journal of Environmental Management 370 (2024) 122614. doi:10.1016/j.jenvman.2024.122614.

[56] C. Probst, A. Zayats, V. Venkatachalam, B. Davidson, Advanced characterization of silicone oil droplets in protein therapeutics using artificial intelligence analysis of imaging flow cytometry data, Journal of Pharmaceutical Sciences 109 (10) (2020) 2996–3005. doi:10.1016/j.xphs.2020.07.008.

[57] M. S.A., S. Maniraj, T. Kavitha, S. Balraj, C. P. Reddy, A. Prabhakar, Fast iterative neural network-combined carbon composite electrodes for selective detection of emerging waterborne pollutants, Microchemical Journal 218 (2025) 115431. doi:10.1016/j.microc.2025.115431.

[58] A.-A. S. Mijwel, N. I. M. Pauzi, H. M. Alayan, H. A. Afan, A. N. Ahmed, M. M. Aljumaily, M. A. Al-Saadi, A. El-Shafie, Artificial intelligence - driven insights into bisphenol A removal using synthesized carbon nanotubes, Microporous and Mesoporous Materials 383 (2025) 113411. doi:10.1016/j.micromeso.2024.113411.

[59] H. Jin, K. Hong, J. Liu, C. Qiu, M. Zhu, C. Li, Q. Wang, Structural and functional design of CTAB-geopolymer adsorbents for rapid removal of tetracycline: A comparative study, Separation and Purification Technology 356 (2025) 129872. doi:10.1016/j.seppur.2024.129872.

[60] P. Zhang, C. Liu, D. Lao, X. C. Nguyen, B. Paramasivan, X. Qian, A. A. Inyinbor, X. Hu, Y. You, F. Li, Unveiling the drives behind tetracycline adsorption capacity with biochar through machine learning, Scientific Reports 13 (1) (2023). doi:10.1038/s41598-023-38579-8.

[61] X. Liu, Z. Shao, Y. Wang, Y. Liu, S. Wang, F. Gao, Y. Dai, New use for Lentinus edodes bran biochar for tetracycline removal, Environmental Research 216 (2023) 114651. doi:10.1016/j.envres.2022.114651.

[62] Y. Wang, C. Wang, X. An, R. Wang, Y. Li, Y. Xu, X. Cheng, Effectively removal of PPCPs by catalytic activated biochar derived from hazelnut shell: Modeled and predicted by machine learning, Colloids and Surfaces A: Physicochemical and Engineering Aspects 702 (2024) 135059. doi:10.1016/j.colsurfa.2024.135059.

[63] C.-W. Cho, Y. Zhao, J.-W. Choi, J.-A. Kim, J. K. Bediako, S. Lin, M.-H. Song, Y.-S. Yun, Prediction of organic pollutant removal using Corynebacterium glutamicum fermentation waste, Environmental Research 192 (2021) 110271. doi:10.1016/j.envres.2020.110271.

[64] A. O. Akinola, E. Prabakaran, K. Govender, K. Pillay, Magnetically-derived pecan nut shells for the adsorptive removal of cadmium: Artificial neural network modelling and photodegradation of sulfamethoxazole using the spent sorbent, Journal of Environmental Chemical Engineering 13 (5) (2025) 118057. doi:10.1016/j.jece.2025.118057.

[65] A. A. Inyinbor, D. T. Bankole, A. P. Oluyori, *Blighia sapida* waste biochar in batch and fixed-bed adsorption of chloroquine phosphate: Efficacy validation using artificial neural networks, ACS Omega (2024). doi:10.1021/acsomega.3c05008.

[66] S. N. Fayyadh, N. A. Tahrim, W. N. A. W. Mokhtar, ANN-based prediction of 17α-ethinylestradiol removal using green magnetic nanoparticles synthesized from celery extract, Results in Chemistry 17 (2025) 102591. doi:10.1016/j.rechem.2025.102591.

[67] P. Balasubramanian, M. R. Prabhakar, C. Liu, P. Zhang, F. Li, Predictive capability of rough set machine learning in tetracycline adsorption using biochar, Carbon Research 3 (1) (2024). doi:10.1007/s44246-024-00129-w.

[68] D. Polanco-Gamboa, M. Abatal, R. Tariq, F. Anguebes-Franseschi, E. C. Lima, A. A. Santiago, R. d. J. Palí-Casanova, F. Tamayo-Ordoñez, Unsupervised machine learning for sensitivity interpretation in the application of biochar derived from Haematoxylum campechianum to remove acetaminophen from aqueous solutions, Results in Engineering 27 (2025) 107008. doi:10.1016/j.rineng.2025.107008.

[69] M. G. Oliveira, M. P. Spaolonzi, E. D. Duarte, H. P. Costa, M. G. da Silva, M. G. Vieira, Adsorption kinetics of ciprofloxacin and ofloxacin by green-modified carbon nanotubes, Environmental Research 233 (2023) 116503. doi:10.1016/j.envres.2023.116503.

[70] H. P. Costa, E. D. Duarte, F. V. da Silva, M. G. da Silva, M. G. Vieira, Green synthesis of carbon nanotubes functionalized with iron nanoparticles and coffee husk biomass for efficient removal of losartan and diclofenac: Adsorption kinetics and ANN modeling studies, Environmental Research 251 (2024) 118733. doi:10.1016/j.envres.2024.118733.

[71] D. C. Henrique, D. U. Quitela, A. H. Ide, P. V. Lins, M. T. Perazzini, H. Perazzini, L. M. Oliveira, J. L. Duarte, L. Meili, Mollusk shells as adsorbent for removal of endocrine disruptor in different water matrix, Journal of Environmental Chemical Engineering 9 (4) (2021) 105704. doi:10.1016/j.jece.2021.105704.

[72] M. O. Anusi, M. C. Menkiti, A. I. Ikeuba, C. N. Njoku, C. E. Iloegbunam, C. J. Nnamani, A. C. Orga, Comparative analysis of the response surface methodology (RSM) and artificial neural network (ANN) modelling for the removal of diclofenac potassium from synthesized pharmaceutical wastewater using a palm sheath fiber nano-filtration membrane and optimization, Current Research in Green and Sustainable Chemistry 11 (2025) 100466. doi:10.1016/j.crgsc.2025.100466.

[73] D. Bulanga, B. Orero, S. Sibiya, T. Paepae, T. Mashifana, Water hyacinth-derived adsorbent for removal of diclofenac sodium from aqueous solution: Modeling and optimization via RSM-ANN/ANFIS hybrid, Next Materials 9 (2025) 100997. doi:10.1016/j.nxmate.2025.100997.

[74] A. E. D. Mahmoud, R. Ali, M. Fawzy, Insights into levofloxacin adsorption with machine learning models using nano-composite hydrochars, Chemosphere 355 (2024) 141746. doi:10.1016/j.chemosphere.2024.141746.

[75] C. Devatha, N. Pavithra, Isolation and identification of Pseudomonas from wastewater, its immobilization in cellulose biopolymer and performance in degrading triclosan, Journal of Environmental Management 232 (2019) 584–591. doi:10.1016/j.jenvman.2018.11.083.

[76] S. Bhattacharya, P. Das, A. Bhowal, A. Saha, Thermal, chemical and ultrasonic assisted synthesis of carbonized biochar and its application for reducing naproxen: Batch and fixed bed study and subsequent optimization with response surface methodology (RSM) and artificial neural network (ANN), Surfaces and Interfaces 26 (2021) 101378. doi:10.1016/j.surfin.2021.101378.

[77] S. Bhattacharya, P. Banerjee, P. Das, A. Bhowal, S. K. Majumder, P. Ghosh, Removal of aqueous carbamazepine using graphene oxide nanoplatelets: process modelling and optimization, Sustainable Environment Research 30 (1) (2020). doi:10.1186/s42834-020-00062-8.

[78] Z. Gao, L. Kong, D. Han, M. Kuang, L. Li, X. Song, N. Li, Q. Shi, X. Qin, Y. Wu, D. Wu, Z. Xu, Deciphering the adsorption mechanisms between microplastics and antibiotics: A tree-based stacking machine learning approach, Journal of Cleaner Production 486 (2025) 144589. doi:10.1016/j.jclepro.2024.144589.

[79] P. Liu, Y. Dong, X. Li, Y. Zhang, Z. Liu, Y. Lu, X. Peng, R. Zhai, Y. Chen, Multilayered Fe3O4@(ZIF-8)3 combined with a computer-vision-enhanced immunosensor for chloramphenicol enrichment and detection, Journal of Hazardous Materials 470 (2024) 134150. doi:10.1016/j.jhazmat.2024.134150.

[80] J. Fabregat-Palau, A. Ershadi, M. Finkel, A. Rigol, M. Vidal, P. Grathwohl, Modeling PFAS sorption in soils using machine learning, Environmental Science & Technology 59 (15) (2025) 7678–7687. doi:10.1021/acs.est.4c13284.

[81] N. T. P. Thao, N. L. T. Quang, P. T. L. Na, B.-T. Dang, Bayesian and unsupervised learning insights into pH- and temperature-driven sorption of fluoroquinolones and sulfonamides on marine algal biochar, Chemosphere 392 (2025) 144753. doi:10.1016/j.chemosphere.2025.144753.

[82] M. A. Hamza, M. M. Althobaiti, F. N. Al-Wesabi, R. Alabdan, H. Mahgoub, A. M. Hilal, A. Motwakel, M. Al Duhayyim, Gaussian process regression and machine learning methods for carbon-based material adsorption, Adsorption Science & Technology 2022 (2022). doi:10.1155/2022/3901608.

[83] O. A. Salawu, Z. Han, A. S. Adeleye, Shrimp waste-derived porous carbon adsorbent: Performance, mechanism, and application of machine learning, Journal of Hazardous Materials 437 (2022) 129266. doi:10.1016/j.jhazmat.2022.129266.

[84] S. Yuan, X. Wang, Z. Jiang, H. Zhang, S. Yuan, Contribution of air-water interface in removing PFAS from drinking water: Adsorption, stability, interaction and machine learning studies, Water Research 236 (2023) 119947. doi:10.1016/j.watres.2023.119947.