

Review

Advancing Identification of Transformation Products and Predicting Their Environmental Fate: The Current State of Machine Learning and Artificial Intelligence in Antibiotic Photolysis

Sultan K. Alharbi 

Department of Chemistry, College of Science, Taibah University, Madinah P.O. Box 344, Saudi Arabia;
sbdrani@taibahu.edu.sa

Abstract

The environmental persistence of antibiotic residues in aquatic systems represents a critical global challenge, with photolysis serving as a primary abiotic degradation pathway. Traditional approaches to studying antibiotic photodegradation and transformation product (TP) identification face significant limitations, including complex reaction mechanisms, multiple concurrent pathways, and analytical challenges in characterizing unknown metabolites. The integration of artificial intelligence (AI) and machine learning (ML) technologies has begun to transform this field, offering new capabilities for predicting photodegradation kinetics, elucidating transformation pathways, and identifying novel metabolites. This comprehensive review examines current applications of AI/ML in antibiotic photolysis research, analyzing developments from 2020 to 2025. Key advances include quantitative structure–activity relationship (QSAR) models for photodegradation prediction, deep learning approaches for automated mass spectrometry interpretation, and hybrid computational–experimental frameworks. Machine learning algorithms, particularly Random Forests, support vector machines, and Neural Networks, have demonstrated capabilities in handling multi-dimensional environmental datasets across diverse antibiotic classes, including fluoroquinolones, β -lactams, tetracyclines, and sulfonamides. Despite progress in this field, challenges remain in model interpretability, standardization of datasets, validation protocols, and integration with regulatory frameworks. Future directions include machine-learning-enhanced quantum dynamics for improving mechanistic understanding, real-time AI-guided experimental design, and predictive tools for environmental risk assessment.

Keywords: antibiotic photolysis; transformation products; machine learning; artificial intelligence; environmental fate; mass spectrometry; QSAR; deep learning



Academic Editor: Antonio Miotello

Received: 3 December 2025

Revised: 21 December 2025

Accepted: 23 December 2025

Published: 26 December 2025

Copyright: © 2025 by the author.

Licensee MDPI, Basel, Switzerland.

This article is an open access article distributed under the terms and conditions of the [Creative Commons Attribution \(CC BY\)](https://creativecommons.org/licenses/by/4.0/) license.

1. Introduction

1.1. Global Challenge of Antibiotic Environmental Contamination

The pervasive presence of antibiotic residues in environmental waters represents a significant global challenge. Global antibiotic consumption also poses a substantial challenge. Estimated consumption in several countries increased by 16.3% from 29.5 to 34.3 billion defined daily doses (DDDs) between 2016 and 2023 [1]. Studies have also indicated large increases in the consumption of different classes of antibiotics, such as fluoroquinolones and third-generation cephalosporins, in North Africa, the Middle East, and South Asia [2]. A significant portion of pharmaceutically active compounds enter aquatic ecosystems through

multiple pathways, including incomplete removal in wastewater treatment plants, agricultural runoff, and direct discharge from pharmaceutical manufacturing facilities [1,3,4]. The persistence of antibiotics in the aquatic environment, such as surface waters, groundwaters, and seawater, can have detrimental effects on ecosystems and public health [3]. Antibiotics persist in these environments, and their breakdown into transformation products (TPs) can contribute to antimicrobial resistance (AMR), which the World Health Organization has identified as one of the top global public health and development threats [5].

1.2. Photolysis as a Primary Environmental Degradation Mechanism

Antibiotics released into the environment undergo various natural degradation processes, including oxidation–reduction reactions, abiotic degradation (hydrolysis and photolysis), biodegradation, and photodegradation [6]. Photolysis (photodecomposition, photodissociation, or photodegradation) is the degradation of inorganic or organic compounds caused by solar irradiation [7–9]. Photodegradation, particularly that initiated by sunlight ultraviolet (UV) radiation, is recognized as a critical pathway for the removal of antibiotics in aquatic environments due to its low cost, efficiency, and environmental friendliness [8]. In particular, solar-induced photodegradation offers consistent degradation pathways under appropriate light conditions unlike biodegradation (microbial degradation), for example, which can be inhibited by the antimicrobial properties of antibiotics [8]. Direct photodegradation proceeds through the absorption of UV radiation by antibiotic molecules, inducing electronic transitions that lead to their breakdown. Indirect photodegradation (advanced oxidation processes (AOPs)), however, relies on various natural photosensitizers within water bodies (such as humus or algal organic matter). These photosensitizers transform absorbed light energy into Reactive Oxygen Species (ROS) and promote the degradation of antibiotics [6].

Indirect photolysis, often referred to as AOPs when engineered, relies on the generation of highly reactive species. These include hydroxyl radicals ($\bullet\text{OH}$), which are powerful, nonselective, and highly effective oxidants capable of degrading a wide range of organic contaminants [9]. Also, sulfate radicals ($\text{SO}_4\bullet^-$), which are well-known for their strong oxidation potential, demonstrated prolonged half-life and electron transfer mechanisms that can be more effective in degraded contaminants such as tetracycline hydrochloride (TCH) [10,11]. In natural waters, dissolved organic matter (DOM) is a primary source of reactive intermediates essential for indirect photolysis [12]. As a natural photosensitizer, DOM converts light energy through a photochemical process into several reactive species (ROS) such as singlet oxygen ($^1\text{O}_2$), superoxide radical ($\text{O}_2\bullet^-$), and the triplet state of DOM ($^3\text{DOM}^*$) [13]. Research indicated that $^3\text{DOM}^*$ facilitates the photo-transformation and toxicity decrease of antibiotics and contributes to the natural degradation of organic pollutants through photochemical reactions [14].

Notably, it is important to clearly distinguish between direct photolysis, indirect photolysis, and AOPs, as these pathways differ in terms of both mechanisms and modeling requirements. For example, direct photolysis involves the absorption of photons by the antibiotic itself, while indirect photolysis relies on naturally occurring photosensitizers such as DOM to generate reactive intermediates (e.g., $^1\text{O}_2$, $\text{O}_2\bullet^-$, $^3\text{DOM}^*$). In contrast, engineered AOPs intentionally generate strong oxidants such as $\bullet\text{OH}$ or $\text{SO}_4\bullet^-$ through methods such as UV/ H_2O_2 , photocatalysis, or ozonation, and these systems differ from natural processes in terms of their scale, intensity, and level of control. Therefore, these mechanistic differences could influence the descriptors, reaction pathways, and environmental variables that ML models must learn, which highlights that explicit separation of these processes is essential for accurate prediction or interpretation. Throughout this

review, the term ‘photolysis’ is employed as an overarching descriptor for both direct and indirect natural processes unless a specific distinction is clarified.

1.3. Analytical Challenges in Photolysis Research

The complexity of photolytic processes is highlighted by the environmental factors influencing their degradation rates and chemical reactions. These include pH, catalyst concentration, initial antibiotic concentration, light intensity and irradiation time, temperature, and the specific nature of the catalyst employed [15]. For example, the pH of a solution impacts how antibiotic molecules are structured (their molecular speciation) and how many natural photosensitizers are available [16]. Similarly, DOM plays a dual role, as it can block light from reaching the antibiotic, but it can also create reactive substances that promote degradation [17]. Recent computational studies highlight the importance of electronic structure descriptors in governing photodegradation behavior [18,19].

A well-established advanced treatment strategy for degrading antibiotics in water involves AOPs; however, the identification of TPs is a complex analytical challenge. A significant concern is the risk of generating intermediate compounds with biological activity/chemical toxicity similar to or different from the parent compound, such as chlorinated organic byproducts, which may be harmful, more toxic, or more difficult to remove than the parent antibiotics [20]. Numerous studies have highlighted the limited comprehensive understanding of transformation processes [21–23]. HRMS-NTA (high-resolution mass spectrometry with non-target analysis) has proven successful in identifying various TPs of various pollutants in diverse environmental samples; however, the environmental relevance and potential toxicity of many newly identified TPs still require comprehensive investigation [24]. Liquid Chromatography-Mass Spectrometry (LC-MS/MS) is also widely employed for detecting and characterizing chemical intermediates, including antibiotics such as tetracycline and amoxicillin/ampicillin, and their degradation products [25,26]. However, balancing the high sensitivity required for low-concentration TPs with the broad scope needed to capture all potential intermediates remains a persistent analytical challenge [27–30]. Furthermore, the kinetic complexity of photolysis reactions which often involves multiple competing pathways and secondary transformations, significantly challenges conventional modelling approaches [31].

1.4. The AI/ML Revolution in Mass Spectrometry and Environmental Chemistry

The complexity of photolysis mechanisms, particularly when dealing with matrix-dependent reactivity and non-linear relationships between molecular structure and light-induced transformations, makes them especially challenging to model using traditional analytical or statistical approaches [32]. Machine Learning (ML) offers a solid alternative due to its ability to integrate diverse descriptors and map high-dimensional patterns, which allows for a level of predictive accuracy that conventional models simply cannot achieve in these complex chemical systems [33]. The rapid advancements in Artificial Intelligence (AI), particularly ML, are profoundly impacting chemical analysis, including the study of antibiotic photolysis and the environmental fate of their TPs. Recent AI-related techniques demonstrated advanced capabilities in managing high-dimensional datasets, uncovering non-linear relationships between chemical structures, and analyzing and predicting various chemical properties [34,35]. This surge in ML capabilities has revolutionized mass spectrometry (MS) applications; for example, it has changed how data is acquired, processed, and interpreted [36]. Furthermore, ML-enhanced quantum dynamics simulations are showing significant promise for predicting molecular photochemistry and offer novel approaches to understanding complex light-driven chemical transformations [37]. These computational

advances have created unprecedented opportunities for data-driven discovery in critical areas such as antibiotic photolysis research.

1.5. Conceptual Framework for AI/ML Integration

To contextualize the methodological and application-focused discussions that follow, an integrated conceptual framework (Figure 1) is introduced. This framework illustrates how AI and ML technologies are being leveraged to address the central challenges in antibiotic photolysis research. This framework is structured around three primary domains: predictive kinetic modeling, TP identification, and mechanistic elucidation. Each domain is supported by distinct yet complementary ML approaches.

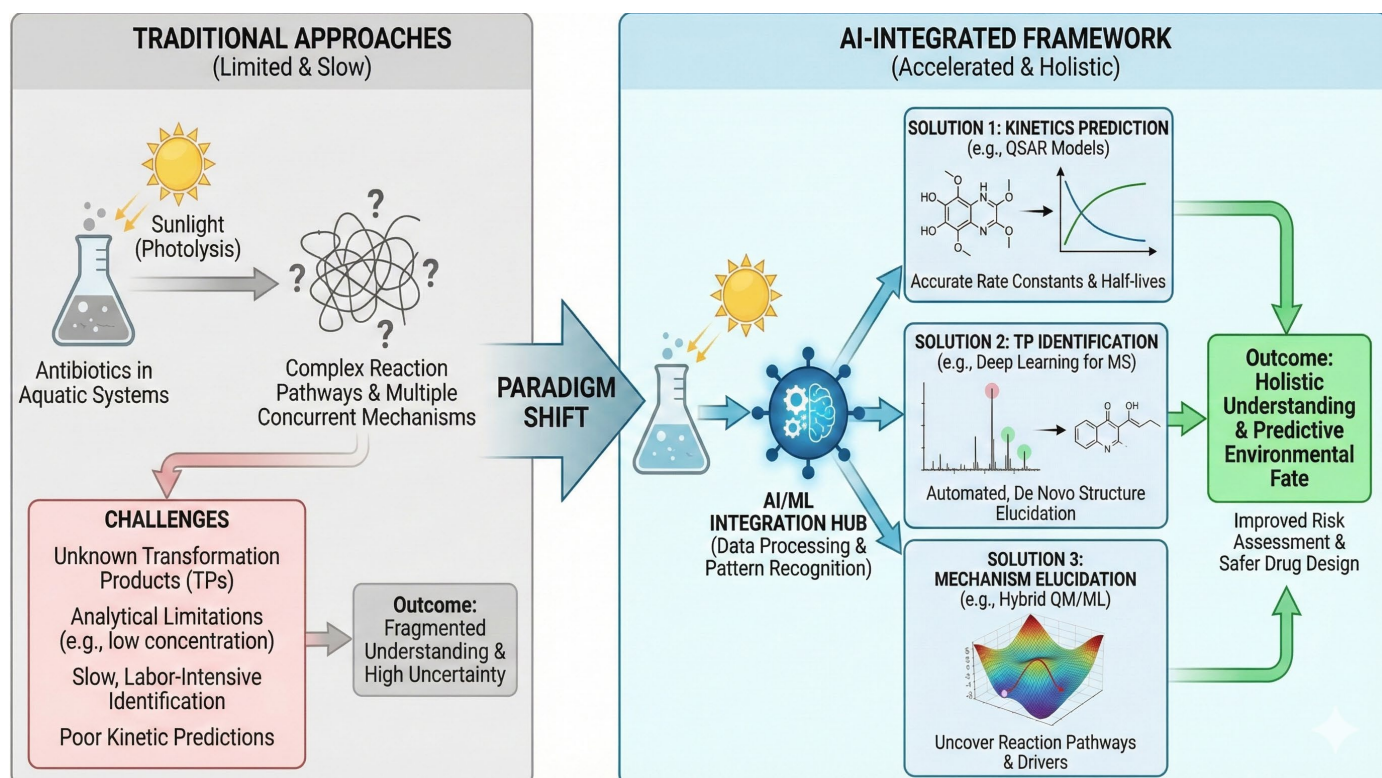


Figure 1. Conceptual workflow: AI/ML transforming antibiotic photolysis research. Note: Question marks (?) denote scientific uncertainties and unknown transformation products. Arrows indicate the transition from traditional data processing to AI-driven solutions. Color coding differentiates research challenges (red) from accelerated frameworks (blue) and predictive outcomes (green).

1.6. Scope and Objectives

This review comprehensively examines AI/ML integration in antibiotic photolysis research, analyzing developments from 2020 to 2025. This research integrates photolysis chemistry domains and ML applications in analytical chemistry, with specific emphasis on TP identification and characterization. This research also evaluates methodological approaches and assesses their applications across antibiotic classes, and also identifies emerging computational tools and highlights their transformative potential while acknowledging their current limitations. The review differentiates itself from existing literature through a novel approach. Specifically, it offers systematic categorization of various ML architectures tailored for TP identification in photolysis studies, supported by a comprehensive analysis of hybrid quantum mechanics frameworks to achieve a mechanistic understanding. Furthermore, the review provides a critical evaluation of validation protocols and applicability domains specific to environmental photochemistry and provides a

forward-looking analysis of emerging technologies, including Physics-Informed Neural Networks (PINNs) and foundation models tailored for photolysis applications.

2. Methodological Approaches in AI/ML Applications

2.1. AI/ML in Identifying and Characterizing Transformation Products

AI and ML are transforming the analytical challenges associated with identifying and characterizing complex TPs. These technologies are now central to modern research in cheminformatics and environmental data science, as they offer unprecedented solutions to problems that have historically limited a complete understanding of chemical fate [38]. Recent advances in MS data interpretation have moved beyond traditional algorithms to incorporate deep learning architectures tailored to specific analytical tasks (Table 1). Convolutional Neural Networks (CNNs), originally developed for image processing, are now applied to MS spectra as 1D or 2D images and excel in feature extraction, automated peak picking, and spectral deconvolution and enhancing TP detection in complex environmental matrices [39]. Recurrent Neural Networks (RNNs) and transformer models are used to interpret fragmentation trees and predict molecular structures such as SMILES strings, enabling de novo identification of unknown compounds [40]. Moreover, Autoencoders play a critical role in preprocessing by performing spectral denoising and dimensionality reduction, isolating relevant TP signals from background noise [41]. Meanwhile, Multilayer Perceptrons (MLPs) are employed for retention time prediction and compound classification based on physicochemical properties, helping to narrow down candidate structures [36,42].

Table 1. Machine learning architectures applied to mass spectrometry data interpretation in transformation product analysis.

ML Architecture	Primary Application	Key Capabilities	Performance Metrics	Limitations	References
Convolutional Neural Networks (CNNs)	Processing of MS spectra as 1D/2D images	Automated peak picking, spectral deconvolution, robust feature extraction	Improved TP detection accuracy; noise resilience	Requires large training datasets; computationally intensive	[39,43,44]
Recurrent Neural Networks (RNNs) and Transformers	Fragmentation tree interpretation, molecular structure prediction	SMILES prediction, sequential pattern learning, de novo compound identification	Enables structure elucidation from MS/MS spectra	Limited generalization to novel chemical scaffolds	[40,45]
Autoencoders	Spectral preprocessing	Denoising, dimensionality reduction, background signal isolation	Enhanced signal-to-noise ratio for TP detection	May lose subtle spectral features during compression	[36,41]
Multilayer Perceptrons (MLPs)	Retention time prediction, compound classification	Property-based classification, candidate narrowing	High retention-time-prediction accuracy	Limited interpretability of learned features	[36,42]
Graph Neural Networks (GNNs)	Molecular structure–activity relationships	Substructure identification, CID simulation via geometric deep learning	Generates virtual MS/MS libraries; identifies activity-linked features	Requires extensive molecular graph databases	[46,47]
Transformer-based Models (DreaMS, MS2DeepScore, Spec2Vec, CLERMS)	Spectral similarity, automated annotation	Self-supervised spectral clustering, annotation from unlabeled data	85–95% TP identification success; low reference library coverage	Data leakage risks; requires standardized benchmarks	[45,48,49]

Note: TP = Transformation Products; MS/MS = Tandem Mass Spectrometry; SMILES = Simplified Molecular Input Line Entry System; CID = Collision-Induced Dissociation. DreaMS (v1.0, released 2025, implemented in Python), MS2DeepScore (implemented within the matchms library v0.24.0, available via PyPI), Spec2Vec (part of the matchms ecosystem v0.18.0 and higher), CLERMS (2025 implementation).

2.2. Data Generation and Preprocessing Strategies

The reliability of utilizing AI/ML models in antibiotic photolysis research fundamentally depends on the quality and consistency of the data used to train, validate, and test those models. Recent advances in data preprocessing have addressed critical challenges in integrating heterogeneous experimental data from multiple sources, which significantly affect the quality, reliability, and validity of the results generated by AI/ML models. Reproducibility is considered a significant challenge across analytical chemistry in general, and particularly in MS, where complex workflows such as sample preparation and data analysis can cause high variability, false positives, and inconsistent results [50]. However, applying AI/ML normalization and standardization was proven to reduce noise and improve comparability across datasets [51,52].

Moreover, spectroscopic preprocessing workflows recently incorporated advanced, automated algorithms that perform traditional tasks like peak detection, baseline correction, and spectral alignment but also apply sophisticated methods such as cosmic ray removal, scattering correction, and filtering. Moreover, Modern preprocessing for spectroscopic data currently uses automated algorithms for peak detection, baseline correction, and spectral alignment, which reduce human bias and improve data consistency across laboratories [53].

3. Current Applications of AI and Machine Learning

3.1. Modelling Photodegradation

For decades, the primary computational tool for predicting the environmental fate of chemicals has been the quantitative structure–activity relationship (QSAR) model. QSAR is a foundational technique in cheminformatics that uses statistical or ML methods to build a mathematical model that predicts a chemical property based on a molecule's structure [54,55]. QSAR attempts to correlate the structural or physicochemical properties of a molecule “descriptors” with a specific activity or property, such as its degradation rate. However, various studies have reported a “poor correlation” between QSAR-predicted photolysis rates and experimentally observed values due to the complexity of environmental photodegradation, and highlighted that current molecular descriptors may not fully capture the reaction mechanisms [56,57]. A review analyzing 1533 QSAR articles found that only 56% of those focused on environmental fate were documented sufficiently for potential reproducibility, which reduces the transparency, availability, and reproducibility of the experimental methods [58]. This has prompted a shift towards leveraging advanced computational power to develop more robust, predictive, and reliable QSAR models for complex environmental fate processes [59].

QSAR models have experienced remarkable advancement through ML integration, moving beyond the limitations of traditional, non-reproducible methods [60]. A recent study successfully developed several 2D-QSAR models using the Lasso regression algorithm, which is a type of ML regression algorithm, to predict the degradation capabilities of 11 fluoroquinolone (FQ) molecules during advanced oxidation processes. The performance of these models, evaluated by the coefficient of determination (R^2), demonstrated good predictive capabilities [61]. Building on these advancements, the development and application of QSAR models have been further empowered by specialized cheminformatics software such as RDKit. This open-source toolkit provides a comprehensive suite of tools for molecular descriptor calculation, chemical structure manipulation, 2D/3D structure visualization, and seamless integration with ML workflows [62]. Its seamless integration with ML workflows enables researchers to efficiently construct and validate predictive models while also supporting the implementation of QSAR and structure–activity relationship analyses with minimal programming effort, which eventually make cheminformatics more accessible to a wider range of users in pharmaceutical and chemical research [63].

Table 2 presents representative ML-enhanced QSAR models demonstrating the evolution from traditional regression approaches to refined hybrid computational frameworks. The key roles of these models in effective photodegradation modeling are demonstrated by highlighting the unique advantages of different ML families. Linear models provide the interpretability needed for small-scale studies. In contrast, tree-based ensemble methods, such as Random Forest and XGBoost, are able to capture non-linear interactions and rank influential descriptors effectively, yet they may overfit noisy or imbalanced datasets. Deep learning architectures such as CNNs, GNNs, and Transformers, often perform well in learning complex spectral or structural patterns and enable de novo TP prediction. However, they require large training datasets and often operate as ‘black boxes’ with limited mechanistic interpretability.

Table 2. Evolution of machine learning-enhanced QSAR models for antibiotic photodegradation prediction.

Model Generation	ML Algorithm	Target Antibiotics	Critical Descriptors	Predictive Performance	Key Advances Over Previous Generation	References
Traditional Statistical QSAR (Pre-ML Era)	Multiple Linear Regression (MLR)	Various classes	Basic physico-chemical properties (log P, MW)	Poor correlation with experimental data	Baseline approach; limited mechanistic insight	[56,58]
First-Gen ML Regression Models	Lasso Regression	11 fluoro-quinolones (FQs)	Multiple physicochemical properties	Good R ² (specific values reported)	• Handles non-linear relationships • Improved reproducibility	[61]
Quantum-Informed ML Models	Random Forest (RF), XGBoost	77 antibiotics across 7 classes	E_{gap} (HOMO–LUMO energy gap) Frontier orbital energies Electron density Hydrophobicity	Strong negative correlation between E_{gap} and log k	• Identified quantum descriptor as primary driver • Improved mechanistic interpretability	[64–66]
Ensemble Ecotoxicity QSAR	RF + XGBoost (consensus)	Ionic liquids (ILs) affecting antibiotics	Cation hydrophobicity (primary) 14 molecular descriptors	Identified 14 critical molecular descriptors	• Multi-algorithm consensus • Improved reliability	[65]
Hybrid DFT/ML Workflows	DFT calculations + ML prediction	Multiple classes	Excited-state energies Reaction barriers Thermodynamic descriptors	Sensitive to training data scope; requires external validation	• Integrates ab initio calculations • Predicts TP formation • Limited by QM computational cost	[67,68]
Hybrid ANFIS-GMM for Material Optimization	Adaptive Neuro-Fuzzy Inference System + Gaussian Mixture Model	Photocatalytic systems (TiO ₂ -based)	Material composition Structural parameters	Lowest prediction error with Gaussian MF 98% photocatalytic efficiency	• Rapid material screening • Optimized H₂ production (2400 µmol/g·h)	[69]

Note: E_{gap} = Energy gap between HOMO (Highest Occupied Molecular Orbital) and LUMO (Lowest Unoccupied Molecular Orbital); QSAR = quantitative structure–activity relationship; ML = machine learning; DFT = density functional theory; QM = quantum mechanical; ANFIS = Adaptive Neuro-Fuzzy Inference System; GMM = Gaussian mixture model; log k = logarithm of rate constant; log P = partition coefficient; MW = molecular weight; TP = transformation product; MF = membership function; R² = coefficient of determination.

A recent ML-developed QSAR model revealed that the energy gap (E_{gap}) between E_{LUMO} (Lowest Unoccupied Molecular Orbital) and E_{HOMO} (Highest Occupied Molecular Orbital) serves as the main chemical descriptor determining photodegradation behavior, showing a negative correlation with logarithmic rate constants [64]. This predictive capability is essential for forecasting environmental behavior and assessing ecological risk. Advanced QSAR approaches now incorporate ensemble methods combining multiple algorithms to improve prediction reliability. For example, in ecotoxicity studies, QSAR models built using Random Forest (RF) and (XGBoost) ML models successfully identified

the hydrophobicity of the cation as the main influence determining the inhibitory effect of ionic liquids (ILs) on the Acetylcholinesterase (AChE) enzyme, leading to the identification of 14 critical molecular descriptors for accurate prediction [65]. Furthermore, a recent study reported that the combined use of ^{19}F -NMR and MS enabled precise tracking of fluorinated intermediates and final products [70]. This detailed experimental evidence is critical for developing mechanistically accurate QSAR and ML models, as it provides the validated data needed to ensure the models correctly recognize and prioritize the influence of specific fluorine substituents.

The evolution of QSAR methodologies for predicting antibiotic photodegradation reflects the broader transformation of computational chemistry through ML integration. Traditional QSAR approaches, while foundational, demonstrated significant limitations in capturing the mechanistic complexity of environmental photolysis processes and capturing photochemical mechanisms. Table 3 shows this methodological evolution from conventional statistical correlations to refined hybrid computational frameworks that combine experimental validation with ML algorithms. Traditional QSAR models relied on statistical associations between simply measured descriptors and degradation rates, which offers predictive but limited insights. In contrast, modern ML approaches can uncover quantum mechanical features that directly govern photochemical reactivity. This evolution marks the development of the field and demonstrates moving beyond prediction to generating mechanistic explanations. As a result, ML-enhanced QSAR is currently used as a forecasting tool and as a powerful framework.

Table 3. Deep learning architectures for de novo transformation product structure elucidation from mass spectrometry data.

Model/Tool Name	Architecture Type	Training Data Source	Primary Function	Key Innovation	Performance Metrics	Current Limitations	Reference
Spec2Mol	Deep learning encoder–decoder	Mass spectra (database-independent)	Molecular structure reconstruction from MS	Database-independent de novo approach	Successfully reconstructs structures from spectra alone	Requires high-quality, diverse training data	[71]
MASSISTANT	Deep learning with SELFIES encoding	Low-resolution EI spectra	Structure and functional group prediction from GC-EI-MS	Operates on low-resolution data; SELFIES encoding	High accuracy with focused, homogeneous training sets	Performance degrades with broad databases	[72]
DreaMS	Transformer, self-supervised	Millions of unannotated tandem MS spectra (GeMS/GNPS)	Automated annotation via masked peak prediction	Self-supervised pre-training on unannotated data	Exceeds 2% reference library limit; rich molecular representations	Only ~10% of spectra currently annotatable; data leakage risks	[49]
Geometric Deep Learning for CID	Graph Neural Networks (GNNs)	Candidate structure libraries	CID simulation for MS/MS generation	Generates high-quality virtual MS/MS data	Significantly accelerates structure elucidation in vast chemical space	Requires extensive molecular graph databases	[46]
MS2DeepScore, Spec2Vec, CLERMS	Transformer architectures	Large spectral databases	Spectral clustering and compound identification	Highly accurate spectral similarity prediction	85–95% success rate for TP identification	Requires standard benchmarks; train/test leakage issues	[45,48]
SynFormer	Generative transformer	Synthesizable molecule databases	Navigates chemical space for novel scaffold discovery	Explores synthesizable molecules beyond training analogs			[73]

Note: MS = Mass Spectrometry; GC-EI-MS = Gas Chromatography–Electron Impact Mass Spectrometry; SELFIES = Self-Referencing Embedded Strings; GeMS = GNPS Experimental Mass Spectra; GNPS = Global Natural Products Social Molecular Networking; CID = Collision-Induced Dissociation; TP = Transformation Product.

It is important to note that ML-enhanced QSAR models typically achieve higher predictive accuracy when trained and validated on controlled laboratory datasets, where experimental conditions such as pH, light intensity, ionic strength, and DOM composition are tightly regulated [74,75]. Under these conditions, molecular descriptors show strong and consistent correlations with photodegradation kinetics. However, model performance often declines when applied to real environmental matrices, which contain variable and interacting factors such as natural organic matter, competing pollutants, turbidity, and fluctuating irradiation conditions [76]. These matrix effects introduce noise and non-linear interactions that are not fully captured in lab-derived datasets, reducing generalizability. Therefore, distinguishing between laboratory and environmental performance is essential for evaluating the robustness and ecological relevance of ML-based photodegradation models.

3.2. Hybrid Computational/Experimental Approaches

The integration of quantum mechanical (QM) calculations with ML is increasingly recognized as a powerful strategy for modeling and understanding photolysis mechanisms, especially in complex systems such as photocatalysis and environmental degradation. Through hybrid ML/quantum dynamics approaches, researchers can currently leverage the efficiency of statistical learning with the accuracy of first-principles calculations to predict molecular photochemistry. These methods are critical for systematically exploring reaction pathways and supporting foundational technologies, including photocatalysis and the design of fluorescent probes [77]. A study focused on optimizing photocatalytic hydrogen production using nanostructured composites through a hybrid model that combines Adaptive Neuro-Fuzzy Inference System (ANFIS) and Gaussian Mixture Model (GMM) methodologies found that the Gaussian membership function yielded the lowest prediction error, and identified graphene/TiO₂ (P25-Acetylene) as the most efficient material, achieving a 2400 $\mu\text{mol/g}\cdot\text{h}$ hydrogen output with 98% photocatalytic efficiency, 56-h stability, and a production cost of USD 14/g [69]. These findings highlight the advantages of hybrid ML approaches in rapidly predicting and guiding the design of advanced materials for efficient solar-based hydrogen production.

3.3. Transformation Product Identification

The identification of TP represents the most transformative application of AI/ML in antibiotic photolysis research. In particular, ML is reshaping small-molecule MS for TP elucidation through three key advancements [45]. First, ML models can predict MS/MS spectra and related physicochemical properties, enabling the construction of extensive virtual libraries that include novel or previously uncatalogued TPs. Second, ML enhances spectral matching accuracy by automatically extracting subtle patterns, which is crucial for confirming TP identities in complex environmental matrices. Third, advanced ML models are being developed for direct structure prediction, also known as *de novo* elucidation, which allows researchers to infer the molecular structures of unknown TPs directly from their MS/MS spectra. These capabilities are driven by the evolution of deep learning approaches for spectral representation, such as DeepMASS and innovative tools such as MS2DeepScore, Spec2Vec, and CLERMS, which leverage transformer architectures for highly accurate spectral clustering and compound identification [45]. Furthermore, miniature mass spectrometers (compact, field-portable versions of traditional lab-scale MS) are gaining traction across analytical fields due to their portability and cost-effectiveness. The integration of AI into these devices has significantly enhanced their performance by enabling intelligent sample identification and expanding their applications to critical areas such as food safety, agricultural disease detection, and environmental monitoring [46].

ML approaches have addressed HRMS-NTA challenges, such as low concentrations of compounds of interest or lack of reference standards, through automated peak annotation, spectral interpretation, and structure prediction algorithms [78]. For instance, Litsa et al. (2023) [79] introduced a deep learning framework named Spec2Mol, (matchms v0.18.0+), that reconstructs molecular structures directly from mass spectra without relying on existing spectral databases, which demonstrates how AI can overcome the limitations of reference-based methods. This *de novo* approach is also being successfully applied to other common analytical techniques, such as gas chromatography–electron impact mass spectrometry (GC-EI-MS). The recently developed deep learning model MASSISTANT, for example, was trained to predict molecular structures and functional groups directly from low-resolution EI spectra [72]. This work further highlights that the success of such models can be highly dependent on dataset curation, with focused, chemically homogeneous training data yielding significantly higher prediction accuracy than broad-scope databases.

Today, ML models can simulate MS fragmentation patterns for hypothetical or uncatalogued compounds. This capability allows researchers to generate vast *in silico* (computer-generated) spectral libraries by expanding searchable databases far beyond what is available from experimental standards. These virtual libraries are critical for detecting and annotating novel environmental contaminants and their TPs [80]. A recent study indicated that advanced methods, such as geometric deep learning, can simulate collision-induced dissociation (CID) to generate high-quality MS/MS data for plausible candidate structures, enabling rapid virtual screening and significantly accelerating the elucidation of unknown structures within a vast chemical space [79]. Notably, AI is fundamentally reshaping the landscape of antibiotic discovery. For decades, the conventional drug development pipeline has been hindered by high costs, slow progress, and limited success [81]. Deep learning approaches have emerged as a powerful solution to overcome these longstanding barriers. For instance, explainable graph neural networks (GNNs) have been used to identify chemical substructures linked to antibiotic activity, enabling the prediction of novel structural classes with low cytotoxicity. A landmark example is the discovery of Halicin, a structurally unique antibiotic identified by a deep neural network trained to recognize molecular features associated with antibacterial efficacy [47]. This AI model screened over 100 million compounds *in silico*, and Halicin was later validated in laboratory and animal studies as effective against multidrug-resistant pathogens, including *Acinetobacter baumannii* [82]. Moreover, Convolutional Neural Networks (CNNs), which are a class of deep learning models designed for visual pattern recognition, are increasingly applied to mass spectrometry data to identify characteristic fragmentation patterns. Notably, CNN models have been developed to process GC-MS data matrices, learning target compound signatures by recognizing peak shapes and enhancing robustness against noise [43]. These convolutional filters effectively capture spectral features, further expanding the searchable space of compound databases and improving the reliability of TP detection in complex environmental samples [44]. Therefore, deep learning models for *de novo* structure elucidation have been demonstrated to represent transformative advances in TP identification. Table 3 summarizes representative deep learning architectures enabling direct structure prediction from MS data.

4. Challenges and Limitations

The transformative potential of AI and ML in analytical chemistry is predicted to fundamentally reshape the field, moving it from a primarily data-generation discipline to a data-interpretation and predictive science. However, the path from research to real-world application is challenging because of several issues, such as fundamental data limitations,

computational difficulties in model explainability, and a substantial gap between in silico modeling and practical implementation [83–85].

Data limitations pose a significant challenge for advancing AI and ML applications in antibiotic photolysis research. Available datasets for antibiotic photolysis and TP characterization are often small, fragmented, and generated using non-standardized methods, which constrains the training of robust and generalizable models [86]. To illustrate, the lack of harmonized protocols in experimental setups, such as differences in light source, water matrix, and compound concentration, results in datasets with limited interoperability [87]. Efforts to enhance model accuracy and generalizability are further restricted by the absence of large, well-annotated, and standardized repositories documenting both degradation kinetics and detailed TP profiles. Environmental heterogeneity presents substantial challenges for ML model development (Table 4). Moreover, interlaboratory variability in analytical methods, especially regarding TP identification, introduces noise that can substantially reduce model performance in AI-driven antibiotic photolysis research [88]. These limitations are acknowledged in recent AI-driven antimicrobial research, which highlighted the urgent need for collaborative dataset generation, harmonized reporting standards, and improved data-sharing protocols to fully realize the potential of ML in environmental and antibiotic fate modelling [86,89].

Table 4. Environmental factors influencing antibiotic photolytic degradation and implications for machine learning model development.

Environmental Factor	Effect on Photodegradation	Mechanistic Role	ML Modeling Implications	References
pH	Alters degradation rates and pathways	- Controls molecular speciation - Affects photosensitizer availability	- Critical feature in QSAR models - Non-linear relationships require advanced algorithms	[15,16]
Dissolved Organic Matter (DOM)	Dual role: inhibition and promotion	- Photosensitizer generating ROS ($^1\text{O}_2$, $\text{O}_2^{\bullet-}$, $^3\text{DOM}^*$) - UV light screening (inner filter effect)	- Requires modeling of competing pathways - Increases dataset complexity - Non-monotonic relationships	[12,14,17]
Light Intensity and Wavelength	Directly proportional to degradation rate	- Determines quantum yield - Controls energy available for photochemical reactions	- Standardization challenges across datasets - Requires wavelength-specific models	[8,15]
Initial Antibiotic Concentration	Affects kinetics and TP formation	- Higher concentrations may alter pathways - Competing reactions at high concentrations	- Non-standardized conditions limit generalizability - Requires concentration-dependent models	[15]
Temperature	Influences reaction rates and ROS generation	- Affects molecular mobility - Impacts reaction kinetics	- Requires temperature-dependent correction factors - Seasonal variation considerations	[15]
Reactive Oxygen Species (ROS)	Primary degradation agents in indirect photolysis	$\bullet\text{OH}$: powerful, non-selective oxidant $\text{SO}_4^{\bullet-}$: strong oxidation, longer half-life $^1\text{O}_2$, $\text{O}_2^{\bullet-}$: DOM-derived	- Essential features for mechanistic ML models - Requires ROS quantification methods - Species-specific reaction pathways	[9–11]
Catalyst Type and Concentration	Determines indirect photolysis efficiency	- Influences ROS generation rates - Affects reaction selectivity	- Catalyst-specific models needed - Material properties must be encoded - Surface chemistry considerations	[15,69]

Note: ROS = Reactive Oxygen Species; $^1\text{O}_2$ = singlet oxygen; $\text{O}_2^{\bullet-}$ = superoxide radical; $^3\text{DOM}^*$ = triplet state of dissolved organic matter; $\bullet\text{OH}$ = hydroxyl radical; $\text{SO}_4^{\bullet-}$ = sulfate radical; TP = transformation product.

Despite the various advances, deep learning models encounter several limitations when applied to complex environmental samples. Many architectures struggle to reliably distinguish structural isomers and stereochemical variants because MS/MS spectra often lack sufficiently unique fragmentation features, which may lead to ambiguous or incorrect structural assignments [90]. Also, low-abundance TPs present an additional challenge, as their weak or noisy spectral signatures are frequently underrepresented in training datasets, which reduces model sensitivity and increases false-negative rates [91]. Furthermore, in natural waters, matrix interferences such as DOM, competing pollutants, and variable ionic composition may mask or alter subtle spectral differences that deep learning models rely on, thus limiting their generalizability beyond controlled laboratory conditions [92].

Moreover, recent work has highlighted the lack of standardized benchmarks for evaluating ML-based spectral similarity methods, as well as the risk of data leakage between training and test sets, which compromises model generalizability [48]. The introduction of domain-inspired evaluation metrics and structurally stratified train/test splits represents a critical step toward more reliable and reproducible ML workflows in untargeted MS.

One recent advancement in MS interpretation is the development of DreaMS (Deep Representations Empowering Annotation of Mass Spectra), a transformer-based model pre-trained in a self-supervised manner on millions of unannotated tandem mass spectra from the GNPS Experimental Mass Spectra (GeMS) dataset [49]. By predicting masked spectral peaks and chromatographic retention orders, DreaMS learns rich molecular representations that support automated annotation. However, despite such progress, significant challenges persist. Notably, only 2% of MS/MS spectra can be annotated using reference libraries, and fewer than 10% using current state-of-the-art ML tools, which highlights the scale of the identification problem [93]. This limitation is especially pronounced for TPs, which often exhibit subtle deviations in fragmentation patterns compared to known compounds and require models to distinguish between closely related structures with high precision.

Beyond data limitations, significant computational challenges persist in applying advanced ML methods to antibiotic photolysis research. Deep learning models with large architectures, such as GNNs and transformer-based models, often require substantial computational resources, including access to GPU clusters for efficient training. For example, generating accurate molecular descriptors through QM calculations (e.g., density functional theory) adds further computational complexity and runtime demands, restricting the feasibility for smaller research groups or those with limited infrastructure. These descriptors are crucial for accurately modeling antibiotic photolysis mechanisms and predicting transformation behaviors [66]. Training these models often involves prolonged computation times, which limits experimentation, hyperparameter tuning, and timely model development [67]. Generating accurate ground truth quantum chemical data requires running complex simulations, often at levels like those of time-dependent density functional theory (TD-DFT) or multireference methods that are computationally prohibitive for large molecular libraries or diverse chemical spaces. Recent studies highlighted the high computational cost of TD-DFT for large systems or when modelling one- and two-photon absorption with microhydration [49]. Even simplified TD-DFT methods reduced CPU time only partially, as accurate benchmarks still depended on full TD-DFT or coupled cluster calculations [94].

A key limitation in many ML applications is the tendency toward overfitting, where the model adapts too closely to the training data, which affects model generalization and performance and often leads to poor predictive performance on new data [95,96]. Although such models may excel on training data, they often fail to generalize to new, unseen inputs, which leads to overly optimistic performance claims [97]. Mitigating overfitting requires rigorous validation, including cross-validation and, critically, testing on external datasets not used during training. However, this essential step is frequently omitted or under-reported, which may limit the reliability and real-world relevance of many published models [98]. For instance, Zhao et al. (2025) developed a DFT-informed ML workflow to predict antibiotic degradation and TP formation; however, they acknowledged that model performance was sensitive to the scope of training data and lacked external benchmarking, which constrained its applicability across diverse water matrices [99]. Similarly, Löffler et al. (2025) emphasized that many ML models in photolysis research are trained on narrow datasets and fail to generalize to natural environmental conditions [100]. These findings highlight the need for standardized validation protocols and broader dataset diversity to ensure robust, transferable predictions in analytical and environmental ML applications.

Recent advances in generative AI, which is type of AI that creates new content rather than only classifying or analyzing existing data, have enabled the design of novel molecules for drug discovery, yet these models often remain confined to the chemical structures present in their training data [101]. This bias restricts exploration to a narrow chemical space, resulting in molecules that closely resemble known compounds [99]. While such behavior may support lead optimization, it limits the discovery of new scaffolds and mechanisms of action, an issue particularly relevant in the context of antibiotic photolysis and TP prediction. Generative models still face challenges in producing structurally diverse and functionally novel outputs, especially when trained on limited or homogeneous datasets [102]. For example, a 2025 review noted that ligand-based generative models usually yield molecular candidates with “relatively low novelty” because their outputs mostly reflect the scope and diversity of molecules they have been exposed to during training [99]. A new model named “SynFormer” was introduced to better navigate chemical spaces containing synthesizable molecules. Nevertheless, the model demonstrated a tendency toward local analog generation unless explicitly guided toward broader exploration, which indicates that manual guidance is necessary to encourage the discovery of fundamentally new scaffolds [73].

The integration of AI and ML into antibiotic photolysis and TP prediction faces significant regulatory, ethical, and technical challenges. Formal chemical risk assessment bodies remain cautious and therefore still prefer validated experimental methods over novel *in silico* approaches, particularly when models are complex or opaque [103]. For AI/ML predictions to inform regulatory decisions, they must meet rigorous standards for validation, transparency, and uncertainty quantification. Yet, the black-box nature of many ML models complicates regulatory acceptance, as agencies require not only accurate predictions but also interpretable justifications [104]. Ethical concerns further complicate implementation, including the risk of amplifying biases from underrepresented chemical classes or geographic regions, which could lead to inequitable environmental protection [105]. The potential for dual-use applications, such as repurposing generative models to design harmful compounds, also demands careful governance [106]. Although frameworks such as those developed for QSAR offer a starting point, they must be adapted to the unique complexities of photolysis modelling [107]. Progress is further hindered by limited stakeholder consensus and barriers to data sharing, as proprietary restrictions on antibiotic structures and environmental fate data constrain the development of robust training datasets [108,109]. Federated learning offers a promising technical solution by enabling distributed model training without centralizing sensitive data, but its implementation requires institutional coordination and trust-building across sectors [110]. These regulatory and ethical challenges highlight that technical performance is insufficient for implementation, and explainability and standardized validation will be essential for integrating AI and ML into environmental decision-making frameworks. Table 5 systematically categorizes the major challenges hindering the integration of AI/ML into photolysis research. Notably, the spectral annotation gap remains substantial, with current ML tools enabling annotation of fewer than 10% of MS/MS spectra [93].

Table 5. Major challenges and limitations in AI/ML applications for antibiotic photolysis research.

Challenge Category	Specific Issues	Impact on Research	Current Mitigation Approaches	References
Data Limitations	• Small, fragmented datasets • Non-standardized protocols • Limited interoperability	• Constrains training of robust, generalizable models • Reduces prediction accuracy across diverse conditions	• Collaborative data generation • Harmonized reporting standards • Improved data-sharing protocols	[86–89]
Spectral Annotation Gap	• Only 2% annotated with libraries • <10% annotated with ML tools • Isobaric compounds indistinguishable	• Severely limits TP identification • Especially challenging for subtle structural variations	• Self-supervised learning (e.g., DreaMS) • Virtual spectral libraries • Transformer architectures	[45,48,93]
Overfitting and Generalization	• Excessive adaptation to training data • Poor external validation • Data leakage risks	• Overly optimistic performance claims • Failure on new datasets	• Rigorous cross-validation • External dataset testing • Regularization techniques	[68,95–99]
Computational Complexity	• GPU requirements • Expensive QM calculations (TD-DFT) • Needs large amounts of memory	• Limits accessibility for smaller research groups • Restricts experimentation and hyperparameter tuning	• Simplified TD-DFT methods • Cloud computing resources • Transfer learning	[49,66,67,94]
Model Interpretability	• Black-box deep learning architectures • Difficult to extract mechanistic insight	• Limits regulatory acceptance • Hinders mechanistic understanding	• Explainable AI (XAI) frameworks • Attention mechanisms • Feature-importance analysis	[103,104,111,112]
Limited Chemical Space Exploration	• Models confined to training data • Low novelty in generated structures	• Restricts discovery of new scaffolds and TP types • Bias toward known compounds	• Explicit exploration guidance • Diverse training datasets • Generative model improvements	[73,99–101]
Regulatory and Ethical Barriers	• Cautious agency acceptance • Bias amplification risks • Dual-use concerns • Data-sharing restrictions	• Slow adoption in regulatory decision-making • Inequitable environmental protection • Limited training data access	• Adapted QSAR frameworks • Federated learning • Stakeholder engagement	[103,105–108,110]

Note: QM = Quantum Mechanical; XAI = Explainable Artificial Intelligence; TP = Transformation Product; GPU = Graphics Processing Unit.

Future Perspectives and Emerging Directions

The integration of AI with analytical chemistry represents a paradigm shift in characterizing and forecasting the environmental trajectory of antibiotics and their TPs. Future advancements will depend on the development of foundation models pre-trained on large, harmonized datasets that integrate molecular structures, quantum descriptors, spectroscopic data, and environmental monitoring results to enable fine-tuning for photolysis and TP prediction with minimal task-specific data [113].

To enhance regulatory trust, PINNs and explainable AI (XAI) can embed photochemical principles and thermodynamic constraints directly into the model architecture and ensure that predictions of TPs remain physically plausible [111,112]. For photolysis, PINNs could encode conservation laws, thermodynamic constraints, and fundamental photochemical principles as constraints or regularization terms during training [114,115]. Furthermore, generative AI enables the “de novo” design of environmentally benign antibiotics which prioritize molecular structures that degrade into non-toxic TPs [116,117]. This discovery process can be accelerated by autonomous laboratories, where AI-driven hypotheses and robotic photolysis experiments form a closed-loop cycle, which reduces human bias and the time required to characterize complex antibiotic degradation pathways [118]. However, realizing this future requires more than technical breakthroughs, as it demands interdisciplinary collaboration, standardized validation protocols, and open-access data infrastructures [119,120]. Autonomous laboratories integrating AI-driven hypothesis generation with robotic experimentation enable closed-loop discovery cycles that accelerate

research while reducing human bias [118]. In addition, the development of foundation models pre-trained on large, harmonized datasets will facilitate rapid fine-tuning for specific photolysis applications with minimal task-specific data [119,120]. Finally, ensuring the sustainability of these computational solutions requires a commitment to ‘Green AI’ practices that aims to reduce the carbon footprint of ML and minimize the environmental footprint of model training itself [121–125]. These advancements outline a clear trajectory toward a robust, transparent, and environmentally conscious framework for understanding antibiotic photolysis.

To provide a structured overview of these emerging directions, the major AI/ML technologies expected to shape the next decade of antibiotic photolysis research are presented in Table 6. The table organizes each technology according to its main technical contribution, expected applications in photolytic degradation studies, expected benefits, current deployment stage, and estimated implementation timeline. The table provides a structured view of the field’s transition toward automated and interpretable modeling by mapping the technologies across near-term, medium-term, and long-term scopes. This roadmap aims to highlight how foundational models such as PINNs, XAI frameworks, autonomous laboratories, generative design tools, and sustainable Green AI practices facilitate an integrated ecosystem for mechanistic prediction, experimental automation, and environmentally responsible model development. This structured perspective clarifies the field’s evolution and the technological shifts needed to achieve a scalable and data-driven framework for investigating antibiotic degradation and antibiotic photolysis research.

Table 6. Emerging AI/ML technologies and future directions for antibiotic photolysis research.

Technology/Approach	Core Innovation	Applications in Photolysis	Expected Benefits	Development Stage	Implementation Timeline	References
NEAR-TERM IMPLEMENTATION						
Explainable AI (XAI)	Transparent architectures with mechanistic insights	- Regulatory acceptance - Mechanistic understanding - Feature importance analysis	- Improved trust - Actionable insights - Knowledge discovery	Advancing—multiple frameworks available	Near commercial deployment	[47,112]
Federated Learning	Distributed training without data centralization	- Privacy-preserving collaboration - Multi-institutional model building	- Overcomes data-sharing barriers - Enables large-scale collaboration	Available—technical solutions ready	Requires institutional coordination	[108,110]
Active Learning Frameworks	Strategic experimental design guided by model uncertainty	- Adaptive experimental design - High-throughput screening optimization	- Reduced experimental costs - Faster convergence	Advancing—successful demonstrations in drug discovery	Early adoption in some labs	[37]
MEDIUM-TERM DEVELOPMENT						
Physics-Informed Neural Networks (PINNs)	Neural networks embedding physical laws, thermodynamics, photochemistry	- Mechanistic modeling - Constraint-guided predictions	- Enhanced interpretability - Improved data efficiency - Physically consistent predictions	Active research—successful in related fields	Proof-of-concept in photolysis	[111,112, 114,115]
Foundation Models	Large models pre-trained on harmonized chemical datasets	- Fine-tuning for photolysis and TP prediction - Transfer learning	- Improved generalizability - Reduced data requirements - Cross-domain capabilities	Early development—proof-of-concept stage	Requires large data infrastructure	[113]
Autonomous Laboratories	AI-driven hypothesis generation + robotic execution	- Closed-loop discovery - Adaptive screening	- Accelerated discovery - Reduced bias 24/7 operation	Emerging—pilot implementations	High capital investment required	[118]
Generative AI for Molecular Design	Design of degradation-optimized antibiotics	- Environmentally benign antibiotics - Predictable TP profiles	- Proactive environmental protection - Sustainable development	Proof-of-concept—early successes	Moving toward practical pharma integration	[101,116, 117]

Table 6. Cont.

Technology/Approach	Core Innovation	Applications in Photolysis	Expected Benefits	Development Stage	Implementation Timeline	References
LONG-TERM VISION						
Multi-Scale Integration	Integration from QM to ecosystem modeling	• Molecular-to-ecosystem modeling • Risk assessment across scales	• Holistic understanding • Comprehensive predictions	Conceptual framework	Requires major integration effort	Proposed in review synthesis
Green AI Practices	Energy-efficient algorithms AND training	• Sustainable computational approaches	• Reduced carbon footprint • Responsible AI • Cost reduction	Framework development—growing awareness	Continuous improvement	[120–122]

Note: QM = Quantum Mechanical; XAI = Explainable Artificial Intelligence; PINNs = Physics-Informed Neural Networks; TP = Transformation Product.

5. Conclusions

The integration of AI and ML into antibiotic photolysis research represents a transformative advancement in environmental analytical chemistry, addressing critical challenges in predicting photodegradation kinetics, identifying TPs, and interpreting reaction mechanisms. This review demonstrates that AI/ML technologies have achieved significant progress through QSAR models, which incorporate molecular descriptors such as frontier orbital energy gaps; deep learning architectures, which enable automated MS interpretation with 85–95% success rates for TP identification; and ML-enhanced quantum dynamics simulations, which provide mechanistic insights into photochemical processes. Despite these achievements, significant challenges remain, including limited standardized databases, model interpretability concerns, validation framework gaps, and regulatory integration barriers. The field urgently requires coordinated efforts to build comprehensive open-access databases, develop Physics-Informed Neural Networks architectures that incorporate photochemical principles, establish rigorous validation protocols with explicit applicability domain assessments, and create high-throughput experimental platforms generating ML-ready datasets. Future directions include the establishment of foundation models pre-trained on diverse chemical data, active learning frameworks guiding adaptive experimental design, multi-scale integration linking molecular photochemistry to ecosystem-level impacts, and proactive integration with pharmaceutical development pipelines. Success requires sustained interdisciplinary collaboration among environmental chemists, analytical chemists, computational scientists, and regulatory agencies, combined with commitment to open science practices, including data sharing, code repositories, and reproducible workflows. The rapid advancement of AI/ML methodologies, growing environmental datasets, and increasing computational power positions the field at an inflection point where thoughtful attention to identified challenges, rigorous validation, and community-wide cooperation can fundamentally transform our ability to understand and predict antibiotic environmental fate, therefore contributing to effective water quality protection, antimicrobial resistance mitigation, and sustainable pharmaceutical development in support of global environmental and public health goals.

Funding: This research received no external funding. The APC was funded by the author.

Data Availability Statement: The data presented in this study are available on request from the corresponding author.

Acknowledgments: Generative artificial intelligence (GenAI Version 1.5 Pro) was utilized in the creation of the graphics and visual elements presented in this paper. All final figures were reviewed and validated by the author for technical accuracy.

Conflicts of Interest: The author declare no conflict of interest.

References

- Klein, E.Y.; Impalli, I.; Poleon, S.; Denoel, P.; Cipriano, M.; Van Boeckel, T.P.; Pecetta, S.; Bloom, D.E.; Nandi, A. Global Trends in Antibiotic Consumption during 2016–2023 and Future Projections through 2030. *Proc. Natl. Acad. Sci. USA* **2024**, *121*, e2411919121. [\[CrossRef\]](#) [\[PubMed\]](#)
- Browne, A.J.; Chipeta, M.G.; Haines-Woodhouse, G.; Kumaran, E.P.A.; Hamadani, B.H.K.; Zarea, S.; Henry, N.J.; Deshpande, A.; Reiner, R.C.; Day, N.P.J.; et al. Global Antibiotic Consumption and Usage in Humans, 2000–18: A Spatial Modelling Study. *Lancet Planet Health* **2021**, *5*, e893–e904. [\[CrossRef\]](#) [\[PubMed\]](#)
- Aib, H.; Parvez, M.S.; Czédli, H.M. Pharmaceuticals and Microplastics in Aquatic Environments: A Comprehensive Review of Pathways and Distribution, Toxicological and Ecological Effects. *Int. J. Environ. Res. Public Health* **2025**, *22*, 799. [\[CrossRef\]](#) [\[PubMed\]](#)
- Ortúzar, M.; Esterhuizen, M.; Olicón-Hernández, D.R.; González-López, J.; Aranda, E. Pharmaceutical Pollution in Aquatic Environments: A Concise Review of Environmental Impacts and Bioremediation Systems. *Front. Microbiol.* **2022**, *13*, 869332. [\[CrossRef\]](#)
- Bertagnolli, S.; Dobrev, Z.; Centner, C.M.; Olaru, I.D.; Donà, D.; Burzo, S.; Huttner, B.D.; Chaillon, A.; Gebreselassie, N.; Wi, T.; et al. WHO Global Research Priorities for Antimicrobial Resistance in Human Health. *Lancet Microbe* **2024**, *5*, 100902. [\[CrossRef\]](#)
- Alharbi, S.K.; Price, W.E. Degradation and Fate of Pharmaceutically Active Contaminants by Advanced Oxidation Processes. *Curr. Pollut. Rep.* **2017**, *3*, 268–280. [\[CrossRef\]](#)
- Zhang, Y.; Jiang, J.; Feng, M.; Ye, C. Transformation Products of Antibiotics: Overlooked Drivers for Enhancing the Environmental Spread of Antibiotic Resistance. *J. Environ. Expo. Assess.* **2025**, *4*, 10. [\[CrossRef\]](#)
- Alharbi, S.K.; Kang, J.; Nghiem, L.D.; van de Merwe, J.P.; Leusch, F.D.L.; Price, W.E. Photolysis and UV/H₂O₂ of Diclofenac, Sulfamethoxazole, Carbamazepine, and Trimethoprim: Identification of Their Major Degradation Products by ESI–LC–MS and Assessment of the Toxicity of Reaction Mixtures. *Process Saf. Environ. Prot.* **2017**, *112*, 222–234. [\[CrossRef\]](#)
- Alharbi, S.K.; Ansari, A.J.; Nghiem, L.D.; Price, W.E. New Transformation Products from Ozonation and Photolysis of Diclofenac in the Aqueous Phase. *Process Saf. Environ. Prot.* **2022**, *157*, 106–114. [\[CrossRef\]](#)
- Fiaz, A.; Zhu, D.; Sun, J. Environmental Fate of Tetracycline Antibiotics: Degradation Pathway Mechanisms, Challenges, and Perspectives. *Environ. Sci. Eur.* **2021**, *33*, 64. [\[CrossRef\]](#)
- Bai, X.; Chen, W.; Wang, B.; Sun, T.; Wu, B.; Wang, Y. Photocatalytic Degradation of Some Typical Antibiotics: Recent Advances and Future Outlooks. *Int. J. Mol. Sci.* **2022**, *23*, 8130. [\[CrossRef\]](#) [\[PubMed\]](#)
- Gautam, R.K.; Chattopadhyaya, M.C. Remediation Technologies for Water Cleanup: New Trends. In *Nanomaterials for Wastewater Remediation*; Gautam, R.K., Chattopadhyaya, M.C., Eds.; Butterworth-Heinemann: Boston, MA, USA, 2016; pp. 19–32. [\[CrossRef\]](#)
- Guerra-Rodríguez, S.; Rodríguez, E.; Singh, D.N.; Rodríguez-Chueca, J. Assessment of Sulfate Radical-Based Advanced Oxidation Processes for Water and Wastewater Treatment: A Review. *Water* **2018**, *10*, 1828. [\[CrossRef\]](#)
- Li, X.; Fan, S.; Jin, C.; Gao, M.; Zhao, Y.; Guo, L.; Ji, J.; She, Z. Electrochemical degradation of tetracycline hydrochloride in sulfate solutions on boron-doped diamond electrode: The accumulation and transformation of persulfate. *Chemosphere* **2022**, *305*, 135448. [\[CrossRef\]](#) [\[PubMed\]](#)
- Cheng, K.; Li, H.; Laszakovits, J.R.; Sharpless, C.M.; Rosario-Ortiz, F.; McKay, G. Probing the Photochemical Formation of Hydroxyl Radical from Dissolved Organic Matter: Insights into the H₂O₂-Dependent Pathway. *Environ. Sci. Technol.* **2025**, *59*, 2245–2256. [\[CrossRef\]](#)
- Dutta Majumdar, R.; Blumkin, L.; Lane, D.; Soong, R.; Simpson, M.; Simpson, A.J. Analysis of DOM Phototransformation Using a Looped NMR System Integrated with a Sunlight Simulator. *Water Res.* **2017**, *120*, 64–76. [\[CrossRef\]](#)
- Guo, Y.; Peng, B.; Liao, J.; Cao, W.; Liu, Y.; Nie, X.; Li, Z.; Ouyang, R. Recent Advances in the Role of Dissolved Organic Matter during Antibiotics Photodegradation in the Aquatic Environment. *Sci. Total Environ.* **2024**, *916*, 170101. [\[CrossRef\]](#)
- Bizirtsakis, P.A.; Anagnostopoulou, K.; Sarasidis, V.C.; Petsi, P.; Moschona, A.; Plakas, K.V.; Lambropoulou, D.A. A Comparative Study of Advanced Oxidation Processes for the Removal of the Antibiotic Sulfadoxine from Water—Transformation Products and Toxicity. *Appl. Sci.* **2025**, *15*, 793. [\[CrossRef\]](#)
- Yuan, S.; Wang, Z.; Yuan, S. Insights into the PH-Dependent Interactions of Sulfadiazine Antibiotic with Soil Particle Models. *Sci. Total Environ.* **2024**, *917*, 170537. [\[CrossRef\]](#)
- Wang, Y.; Li, N.; Fu, Q.; Cheng, Z.; Song, Y.; Yan, B.; Chen, G.; Hou, L.; Wang, S. Conversion and Impact of Dissolved Organic Matters in a Heterogeneous Catalytic Peroxymonosulfate System for Pollutant Degradation. *Water Res.* **2023**, *241*, 120166. [\[CrossRef\]](#)
- Magaji, B.; Singh, P.; Skelton, A.A.; Martincigh, B.S. A Density Functional Theory Study of a Series of Symmetric Dibenzylideneacetone Analogues as Potential Chemical UV-Filters. *Heliyon* **2024**, *10*, e39910. [\[CrossRef\]](#)
- Lian, L.; Jiang, B.; Xing, Y.; Zhang, N. Identification of Photodegradation Product of Organophosphorus Pesticides and Elucidation of Transformation Mechanism under Simulated Sunlight Irradiation. *Ecotoxicol. Environ. Saf.* **2021**, *224*, 112655. [\[CrossRef\]](#) [\[PubMed\]](#)

23. Ike, I.A.; Karanfil, T.; Cho, J.; Hur, J. Oxidation Byproducts from the Degradation of Dissolved Organic Matter by Advanced Oxidation Processes—A Critical Review. *Water Res.* **2019**, *164*, 114929. [[CrossRef](#)] [[PubMed](#)]
24. Sharma, A.; Ahmad, J.; Flora, S.J.S. Application of Advanced Oxidation Processes and Toxicity Assessment of Transformation Products. *Environ. Res.* **2018**, *167*, 223–233. [[CrossRef](#)] [[PubMed](#)]
25. Zahn, D.; Arp, H.P.H.; Fenner, K.; Georgi, A.; Hafner, J.; Hale, S.E.; Hollender, J.; Letzel, T.; Schymanski, E.L.; Sigmund, G.; et al. Should Transformation Products Change the Way We Manage Chemicals? *Environ. Sci. Technol.* **2024**, *58*, 7710–7718. [[CrossRef](#)]
26. McDonough, C.A.; Joudan, S.; Quinete, N.S.; Wang, X. Transformation of Environmental Contaminants: Uncovering Reaction Mechanisms, Identifying Novel Products, and Understanding Environmental Implications. *Environ. Toxicol. Chem.* **2024**, *43*, 2249–2251. [[CrossRef](#)]
27. Panagopoulos Abrahamsson, D.; Wang, A.; Jiang, T.; Wang, M.; Siddharth, A.; Morello-Frosch, R.; Park, J.S.; Sirota, M.; Woodruff, T.J. A Comprehensive Non-Targeted Analysis Study of the Prenatal Exposome. *Environ. Sci. Technol.* **2021**, *55*, 10542–10557. [[CrossRef](#)]
28. Hu, M.; Ben, Y.; Wong, M.H.; Zheng, C. Trace Analysis of Multiclass Antibiotics in Food Products by Liquid Chromatography-Tandem Mass Spectrometry: Method Development. *J. Agric. Food Chem.* **2021**, *69*, 1656–1666. [[CrossRef](#)]
29. Sun, L.; Jia, L.; Xie, X.; Xie, K.; Wang, J.; Liu, J.; Cui, L.; Zhang, G.; Dai, G.; Wang, J. Quantitative Analysis of Amoxicillin, Its Major Metabolites and Ampicillin in Eggs by Liquid Chromatography Combined with Electrospray Ionization Tandem Mass Spectrometry. *Food Chem.* **2016**, *192*, 313–318. [[CrossRef](#)]
30. De Girolamo, A.; Lippolis, V.; Pascale, M. Overview of Recent Liquid Chromatography Mass Spectrometry-Based Methods for Natural Toxins Detection in Food Products. *Toxins* **2022**, *14*, 328. [[CrossRef](#)]
31. Damiani, T.; Bonciarelli, S.; Thallinger, G.G.; Koehler, N.; Krettl, C.A.; Salihoğlu, A.K.; Korf, A.; Pauling, J.K.; Pluskal, T.; Ni, Z.; et al. Erratum: Software and Computational Tools for LC-MS-Based Lipidomics: Challenges and Solutions. *Anal. Chem.* **2023**, *95*, 287–303, Correction in *Anal. Chem.* **2023**, *95*, 3550. <https://doi.org/10.1021/acs.analchem.3c00305>. [[CrossRef](#)]
32. Vitale, D.; Andreu, V.; Picó, Y. Identifying Emerging Pollutants Using Non-Target or Wide-Screening Liquid Chromatography-Mass Spectrometry. In *Handbook of Environmental Chemistry*; Núñez-Delgado, A., Arias-Estévez, M., Eds.; Springer International Publishing: Cham, Switzerland, 2023; Volume 114, pp. 209–223. ISBN 978-3-031-07609-1.
33. Pan, H.; Zhao, G.; Tan, Y.; Miao, S.; Mao, X.; Ji, S.; Zhou, H.; Hu, Q. Liquid Chromatography Coupled to High-Resolution Mass Spectrometry for High-Throughput Pesticide Residue Detection: Methodological Advances and Insights for Herbal Medicine Analysis. *Crit. Rev. Anal. Chem.* **2025**, 1–15. [[CrossRef](#)] [[PubMed](#)]
34. Ding, Y.; Zhu, Y.; Yu, X.; Lyu, S.; Wu, Y.; Zhou, L.; Sui, Q. Mechanistic and Kinetic Aspects of Florfenicol Degradation by [Rad]OH: Chloride Moiety Resistance. *Chem. Eng. J.* **2024**, *479*, 147696. [[CrossRef](#)]
35. Qudsieh, I.Y.; Ali, M.A.; Maafa, I.M. Effect of Water Matrix on Photocatalytic Degradation of Organic Pollutants in Water: A Literature Review. *Rev. Chem. Eng.* **2025**, *41*, 539–573. [[CrossRef](#)]
36. Soltani, A.; Djani, F.; Abdesslam, Y. Machine Learning Predictive Model to Estimate the Photo-Degradation Performance of Stannates and Hydroxystannates Photocatalysts on a Variety of Waterborne Contaminants. *Comput. Theor. Chem.* **2025**, *1244*, 115003. [[CrossRef](#)]
37. Vora, L.K.; Gholap, A.D.; Jetha, K.; Thakur, R.R.S.; Solanki, H.K.; Chavda, V.P. Artificial Intelligence in Pharmaceutical Technology and Drug Delivery Design. *Pharmaceutics* **2023**, *15*, 1916. [[CrossRef](#)]
38. Huang, Z.; Yu, J.; He, W.; Yu, J.; Deng, S.; Yang, C.; Zhu, W.; Shao, X. AI-Enhanced Chemical Paradigm: From Molecular Graphs to Accurate Prediction and Mechanism. *J. Hazard. Mater.* **2024**, *465*, 133355. [[CrossRef](#)]
39. Beck, A.G.; Muhoberac, M.; Randolph, C.E.; Beveridge, C.H.; Wijewardhane, P.R.; Kenttämää, H.I.; Chopra, G. Recent Developments in Machine Learning for Mass Spectrometry. *ACS Meas. Sci. Au* **2024**, *4*, 233–246. [[CrossRef](#)]
40. Martyka, M.; Zhang, L.; Ge, F.; Hou, Y.F.; Jankowska, J.; Barbatti, M.; Dral, P.O. Charting Electronic-State Manifolds across Molecules with Multi-State Learning and Gap-Driven Dynamics via Efficient and Robust Active Learning. *NPJ Comput. Mater.* **2025**, *11*, 132. [[CrossRef](#)]
41. Shizo, A. Cheminformatics in Toxicology and Environmental Chemistry The Role of Cheminformatics in Toxicology. *J. Adv. Chem. Eng.* **2023**, *13*, 299.
42. Mottershead, K.; Miller, T.H. Application of Deep Learning to Support Peak Picking during Non-Target High Resolution Mass Spectrometry Workflows in Environmental Research. *Environ. Sci. Adv.* **2023**, *2*, 877–885. [[CrossRef](#)]
43. Stokes, J.M.; Yang, K.; Swanson, K.; Jin, W.; Cubillos-Ruiz, A.; Donghia, N.M.; MacNair, C.R.; French, S.; Carfrae, L.A.; Bloom-Ackerman, Z.; et al. A Deep Learning Approach to Antibiotic Discovery. *Cell* **2020**, *180*, 688–702.e13. [[CrossRef](#)] [[PubMed](#)]
44. Wong, F.; Zheng, E.J.; Valeri, J.A.; Donghia, N.M.; Anahtar, M.N.; Omori, S.; Li, A.; Cubillos-Ruiz, A.; Krishnan, A.; Jin, W.; et al. Discovery of a Structural Class of Antibiotics with Explainable Deep Learning. *Nature* **2024**, *626*, 177–185. [[CrossRef](#)] [[PubMed](#)]
45. Nunes, M.J.; Lopes, A.; Pacheco, M.J.; Fiadeiro, P.T.; Inacio, G.J.; Silveira, J.E.; Ribeiro, A.R.; Paz, W.S.; Ciriaco, L. Photocatalytic Degradation of Oxytetracycline and Imidacloprid Under Visible Light with Sr0.95Bi0.05TiO3: Influence of Aqueous Matrix. *Water* **2025**, *17*, 2177. [[CrossRef](#)]

46. Richings, G.W.; Habershon, S. Predicting Molecular Photochemistry Using Machine-Learning-Enhanced Quantum Dynamics Simulations. *Acc. Chem. Res.* **2022**, *55*, 209–220. [\[CrossRef\]](#)
47. Liu, Y.; De Vijlder, T.; Bittremieux, W.; Laukens, K.; Heyndrickx, W. Current and Future Deep Learning Algorithms for Tandem Mass Spectrometry (MS/MS)-Based Small Molecule Structure Elucidation. *Rapid Commun. Mass Spectrom.* **2025**, *39*, e9120. [\[CrossRef\]](#)
48. Tahıl, G.; Delorme, F.; Le Berre, D.; Monflier, É.; Sayede, A.; Tilloy, S. Stereoisomers Are Not Machine Learning's Best Friends. *J. Chem. Inf. Model.* **2024**, *64*, 5451–5469. [\[CrossRef\]](#)
49. Bushuiev, R.; Bushuiev, A.; Samusevich, R.; Brungs, C.; Sivic, J.; Pluskal, T. Self-Supervised Learning of Molecular Representations from Millions of Tandem Mass Spectra Using DreaMS. *Nat. Biotechnol.* **2025**, 1–11. [\[CrossRef\]](#)
50. Yang, Y.; Sun, S.; Yang, S.; Yang, Q.; Lu, X.; Wang, X.; Yu, Q.; Huo, X.; Qian, X. Structural Annotation of Unknown Molecules in a Miniaturized Mass Spectrometer Based on a Transformer Enabled Fragment Tree Method. *Commun. Chem.* **2024**, *7*, 109. [\[CrossRef\]](#)
51. Berahmand, K.; Daneshfar, F.; Salehi, E.S.; Li, Y.; Xu, Y. Autoencoders and Their Applications in Machine Learning: A Survey. *Artif. Intell. Rev.* **2024**, *57*, 28. [\[CrossRef\]](#)
52. Liu, Y.; Yoshizawa, A.C.; Ling, Y.; Okuda, S. Insights into Predicting Small Molecule Retention Times in Liquid Chromatography Using Deep Learning. *J. Cheminform.* **2024**, *16*, 113. [\[CrossRef\]](#)
53. Sakagianni, A.; Koufopoulou, C.; Koufopoulos, P.; Kalantzi, S.; Theodorakis, N.; Nikolaou, M.; Paxinou, E.; Kalles, D.; Verykios, V.S.; Myrianthefs, P.; et al. Data-Driven Approaches in Antimicrobial Resistance: Machine Learning Solutions. *Antibiotics* **2024**, *13*, 1052. [\[CrossRef\]](#) [\[PubMed\]](#)
54. Boiko, D.A.; Ananikov, V.P. Machine Learning Meets Mass Spectrometry: A Focused Perspective. In *Artificial Intelligence in Catalysis*; Wiley: New York, NY, USA, 2025; pp. 35–54. [\[CrossRef\]](#)
55. Ebbels, T.M.D.; van der Hooft, J.J.J.; Chatelaine, H.; Broeckling, C.; Zamboni, N.; Hassoun, S.; Mathé, E.A. Recent advances in mass spectrometry-based computational metabolomics. *Curr. Opin. Chem. Biol.* **2023**, *74*, 102288. [\[CrossRef\]](#) [\[PubMed\]](#) [\[PubMed Central\]](#)
56. Yan, C. A Review on Spectral Data Preprocessing Techniques for Machine Learning and Quantitative Analysis. *iScience* **2025**, *28*, 112759. [\[CrossRef\]](#) [\[PubMed\]](#)
57. Odugbemi, A.I.; Nyirenda, C.; Christoffels, A.; Egieyeh, S.A. Artificial Intelligence in Antidiabetic Drug Discovery: The Advances in QSAR and the Prediction of α -Glucosidase Inhibitors. *Comput. Struct. Biotechnol. J.* **2024**, *23*, 2964–2977. [\[CrossRef\]](#)
58. Thareja, R.; Singh, J.; Bansal, P. Computational Tools in Cheminformatics. In *Cheminformatics and Bioinformatics in the Pharmaceutical Sciences*; Academic Press: New York, NY, USA, 2021; pp. 105–137. [\[CrossRef\]](#)
59. Meng, T.; Bai, Y.; Sun, P.; Zhang, H.; Zhang, R. Phototransformation Kinetics and Changes in the Antibacterial Activity of Antimicrobial Agents. *ACS ES&T Water* **2025**, *5*, 913–919. [\[CrossRef\]](#)
60. Sulmona, K.; Louazzani, M.; Ginestar, J.; Lancia, P.; Baleyrier, A. Comparative Study of (Q)SAR Models for Predicting the Environmental Fate of Cosmetic Ingredients: Persistence, Bioaccumulation, and Mobility. *NAM J.* **2025**, *1*, 100054. [\[CrossRef\]](#)
61. Piir, G.; Kahn, I.; García-Sosa, A.T.; Sild, S.; Ahte, P.; Maran, U. Best Practices for QSAR Model Reporting: Physical and Chemical Properties, Ecotoxicity, Environmental Fate, Human Health, and Toxicokinetics Endpoints. *Environ. Health Perspect.* **2018**, *126*, 126001. [\[CrossRef\]](#)
62. Du, Y.; Tang, T.; Song, D.; Wang, R.; Liu, H.; Du, X.; Dang, Z.; Lu, G. Prediction of Chlorination Degradation Rate of Emerging Contaminants Based on Machine Learning Models. *Environ. Pollut.* **2025**, *372*, 125976. [\[CrossRef\]](#)
63. Barros, R.P.C.; Sousa, N.F.; Scotti, L.; Scotti, M.T. Use of Machine Learning and Classical QSAR Methods in Computational Ecotoxicology. In *Ecotoxicological QSARs; Methods in Pharmacology and Toxicology*; Humana: New York, NY, USA, 2020; pp. 151–175. [\[CrossRef\]](#)
64. Sun, S.; Wang, Z.; Pu, Q.; Li, X.; Cui, Y.; Yang, H.; Li, Y. Identification and Mechanistic Analysis of Toxic Degradation Products in the Advanced Oxidation Pathways of Fluoroquinolone Antibiotics. *Toxics* **2024**, *12*, 203. [\[CrossRef\]](#)
65. Chetry, A.B.; Ohto, K. From Molecules to Data: The Emerging Impact of Chemoinformatics in Chemistry. *J. Cheminform.* **2025**, *17*, 121. [\[CrossRef\]](#)
66. Ali, A.; Zhou, G.; Pablo Antezana Lopez, F.; Xu, C.; Jing, G.; Tan, Y. Deep Learning for Water Quality Multivariate Assessment in Inland Water across China. *Int. J. Appl. Earth Obs. Geoinf.* **2024**, *133*, 104078. [\[CrossRef\]](#)
67. Strobel, M.; Gil-de-la-Fuente, A.; Zare Shahneh, M.R.; El Abiead, Y.; Bushuiev, R.; Bushuiev, A.; Pluskal, T.; Wang, M. An Evaluation Methodology for Machine Learning-Based Tandem Mass Spectra Similarity Prediction. *BMC Bioinform.* **2025**, *26*, 174. [\[CrossRef\]](#) [\[PubMed\]](#)
68. Fan, Y.; Huang, H.; Han, H. Quantifying Overfitting in Deep Learning. *Anal. Appl.* **2025**, *23*, 705–729. [\[CrossRef\]](#)
69. Dai, Z.; Xu, J.; Guan, J.; Feng, M.; Liu, Y.; Xing, C.; Cai, X.; Wang, S.; Lian, L.; Dong, H.; et al. TP-Transformer: An Interpretable Model for Predicting the Transformation Pathways of Organic Pollutants in Chemical Oxidation Processes. *Environ. Sci. Technol.* **2025**, *59*, 15853–15864. [\[CrossRef\]](#)

70. Alcázar, J.J.; Sánchez, I.; Merino, C.; Monasterio, B.; Sajuria, G.; Miranda, D.; Díaz, F.; Campodónico, P.R. A Simple Machine Learning-Based Quantitative Structure–Activity Relationship Model for Predicting PIC50 Inhibition Values of FLT3 Tyrosine Kinase. *Pharmaceuticals* **2025**, *18*, 96. [\[CrossRef\]](#)
71. Hong, Y.; Ye, Y.; Tang, H. Machine Learning in Small-Molecule Mass Spectrometry. *Annu. Rev. Anal. Chem.* **2025**, *18*, 193–215. [\[CrossRef\]](#)
72. Wang, J.; Liu, L.; Jiang, T. Integrating Artificial Intelligence with Miniature Mass Spectrometry. *Green Anal. Chem.* **2025**, *13*, 100281. [\[CrossRef\]](#)
73. Das, U. Generative AI for Drug Discovery and Protein Design: The next Frontier in AI-Driven Molecular Science. *Med. Drug Discov.* **2025**, *27*, 100213. [\[CrossRef\]](#)
74. Lyu, Y.; Huang, T.; Liu, W.; Sun, W. Unveil the Quantum Chemical Descriptors Determining Direct Photodegradation of Antibiotics under Simulated Sunlight: Batch Experiments and Model Development. *J. Environ. Chem. Eng.* **2022**, *10*, 108086. [\[CrossRef\]](#)
75. Wu, X.; Gong, J.; Ren, S.; Tan, F.; Wang, Y.; Zhao, H. A Machine Learning-Based QSAR Model Reveals Important Molecular Features for Understanding the Potential Inhibition Mechanism of Ionic Liquids to Acetylcholinesterase. *Sci. Total Environ.* **2024**, *915*, 169974. [\[CrossRef\]](#)
76. Mundhenke, T.F.; Bhat, A.P.; Pomerantz, W.C.K.; Arnold, W.A. Photolysis Products of Fluorinated Pharmaceuticals: A Combined Fluorine Nuclear Magnetic Resonance Spectroscopy and Mass Spectrometry Approach. *Environ. Toxicol. Chem.* **2024**, *43*, 2285–2296. [\[CrossRef\]](#) [\[PubMed\]](#)
77. Khondkaryan, L.; Tevosyan, A.; Navasardyan, H.; Khachatryan, H.; Tadevosyan, G.; Apresyan, L.; Chilingaryan, G.; Navoyan, Z.; Stopper, H.; Babayan, N. Datasets Construction and Development of QSAR Models for Predicting Micronucleus In Vitro and In Vivo Assay Outcomes. *Toxics* **2023**, *11*, 785. [\[CrossRef\]](#) [\[PubMed\]](#)
78. Khan, O.; Alsaduni, I.; Parvez, M.; Yadav, A.K. Enhancing Hydrogen Production Using Solar-Driven Photocatalysis with Biosynthesized Nanocomposites: A Hybrid Machine Learning Approach towards Enhanced Performance and Sustainable Environment. *Int. J. Hydrog. Energy* **2025**, *102*, 609–625. [\[CrossRef\]](#)
79. Litsa, E.E.; Chenthamarakshan, V.; Das, P.; Kaviraki, L.E. An End-to-End Deep Learning Framework for Translating Mass Spectra to de-Novo Molecules. *Commun. Chem.* **2023**, *6*, 132. [\[CrossRef\]](#)
80. Jin, Z.L.; Chen, L.; Wang, Y.; Shi, C.T.; Zhou, Y.; Xia, B. Application of Machine Learning in LC-MS-Based Non-Targeted Analysis. *TrAC—Trends Anal. Chem.* **2025**, *189*, 118243. [\[CrossRef\]](#)
81. Mommers, J.; Barta, L.; Pietrasik, M.; Wilbik, A. MASSISTANT: A Deep Learning Model for De Novo Molecular Structure Prediction from EI-MS Spectra via SELFIES Encoding. *J. Chromatogr. A* **2025**, *1759*, 466216. [\[CrossRef\]](#)
82. Wang, R.; Manjrekar, M.; Mahjour, B.; Avila-Pacheco, J.; Provenzano, J.; Reynolds, E.; Lederbauer, M.; Mashin, E.; Goldman, S.; Wang, M.; et al. Neural Spectral Prediction for Structure Elucidation with Tandem Mass Spectrometry. *bioRxiv* **2025**. [\[CrossRef\]](#)
83. Iovleva, A.; Fowler, V.G.; Doi, Y. Treatment Approaches for Carbapenem-Resistant *Acinetobacter Baumannii* Infections. *Drugs* **2025**, *85*, 21–40. [\[CrossRef\]](#)
84. Skarysz, A.; Alkhalifah, Y.; Darnley, K.; Eddleston, M.; Hu, Y.; McLaren, D.B.; Nailon, W.H.; Salman, D.; Sykora, M.; Thomas, C.L.P.; et al. Convolutional Neural Networks for Automated Targeted Analysis of Raw Gas Chromatography-Mass Spectrometry Data. In Proceedings of the International Joint Conference on Neural Networks 2018, Rio de Janeiro, Brazil, 8–13 July 2018. [\[CrossRef\]](#)
85. Fan, Y.; Yu, C.; Lu, H.; Chen, Y.; Hu, B.; Zhang, X.; Su, J.; Zhang, Z. Deep Learning-Based Method for Automatic Resolution of Gas Chromatography-Mass Spectrometry Data from Complex Samples. *J. Chromatogr. A* **2023**, *1690*, 463768. [\[CrossRef\]](#)
86. Guo, Z.; Huang, L.; Yan, J.; Zhang, H.; Jia, X.; Li, M.; Li, H. Artificial Intelligence Technology in Environmental Research and Health: Development and Prospects. *Environ. Int.* **2025**, *203*, 109788. [\[CrossRef\]](#)
87. de Oliveira, A.F.B.; de Melo Vieira, A.; Santos, J.M. Trends and Challenges in Analytical Chemistry for Multi-Analysis of Illicit Drugs Employing Wastewater-Based Epidemiology. *Anal. Bioanal. Chem.* **2023**, *415*, 3749–3758. [\[CrossRef\]](#) [\[PubMed\]](#)
88. Esaki, T.; Ikeda, K. Difficulties and Prospects of Data Curation for ADME in Silico Modeling. *Chem-Bio Inform. J.* **2023**, *23*, 1–6. [\[CrossRef\]](#)
89. Cesaro, A.; Hoffman, S.C.; Das, P.; de la Fuente-Nunez, C. Challenges and Applications of Artificial Intelligence in Infectious Diseases and Antimicrobial Resistance. *Npj Antimicrob. Resist.* **2025**, *3*, 2. [\[CrossRef\]](#) [\[PubMed\]](#)
90. Gadiya, Y.; Genilloud, O.; Bilitewski, U.; Brönstrup, M.; von Berlin, L.; Attwood, M.; Gribbon, P.; Zaliani, A. Predicting Antimicrobial Class Specificity of Small Molecules Using Machine Learning. *J. Chem. Inf. Model.* **2025**, *65*, 2416–2431. [\[CrossRef\]](#)
91. Ciornii, D.; Hodoroba, V.D.; Benismail, N.; Maltseva, A.; Ferrer, J.F.; Wang, J.; Parra, R.; Jézéquel, R.; Receveur, J.; Gabriel, D.; et al. Interlaboratory Comparison Reveals State of the Art in Microplastic Detection and Quantification Methods. *Anal. Chem.* **2025**, *97*, 8719–8728. [\[CrossRef\]](#)
92. Odone, A.; Barbati, C.; Amadasi, S.; Schultz, T.; Resnik, D.B. Artificial Intelligence and Infectious Diseases: An Evidence-Driven Conceptual Framework for Research, Public Health, and Clinical Practice. *Lancet Infect. Dis.* **2025**. [\[CrossRef\]](#)

93. Wang, F.; Pasin, D.; Skinnider, M.A.; Liigand, J.; Kleis, J.N.; Brown, D.; Oler, E.; Sajed, T.; Gautam, V.; Harrison, S.; et al. Deep Learning-Enabled MS/MS Spectrum Prediction Facilitates Automated Identification Of Novel Psychoactive Substances. *Anal. Chem.* **2023**, *95*, 18326–18334. [\[CrossRef\]](#)
94. Ghasemi, J.B. Predicting photodegradation rate constants of water pollutants on TiO₂ using graph neural network and combined experimental-graph features. *Sci. Rep.* **2025**, *15*, 19156. [\[CrossRef\]](#)
95. Caldas, A.H.; Kokorin, A.; Tkatchenko, A.; Sandonas, L.M. Assessing the Performance of Quantum-Mechanical Descriptors in Physicochemical and Biological Property Prediction. *ChemRxiv* **2025**. [\[CrossRef\]](#)
96. Elayan, I.A.; Johansson, R.; Brown, A. Evaluating the Tamm-Dancoff Approximation for Two-Photon Absorption Properties. *ChemRxiv* **2025**. [\[CrossRef\]](#)
97. Maraldi, M.G.; de Wergifosse, M. Evaluating the Performance of the Exact Integral Simplified Time-Dependent Density Functional Theory (XsTD-DFT) to Compute One- and Two-Photon Absorption. *J. Phys. Chem. A* **2025**, *129*, 8178–8203. [\[CrossRef\]](#)
98. Aliferis, C.; Simon, G. Overfitting, Underfitting and General Model Overconfidence and Under-Performance Pitfalls and Best Practices in Machine Learning and AI. In *Artificial Intelligence and Machine Learning in Health Care and Medical Sciences*; Health Informatics; Springer: Cham, Switzerland, 2024; pp. 477–524. [\[CrossRef\]](#)
99. Zhao, C.-C.; Xing, S.; Fu, C.; Zheng, L.; Zhang, S.; Ma, J. Mapping Antibiotic Photocatalytic Transformation and Resistance Risks with a DFT-Informed Machine Learning Workflow. *ChemRxiv* **2025**. [\[CrossRef\]](#)
100. Löffler, P.; Henschel, H.; Ugolini, V.; Flores Quintana, H.; Wiberg, K.; Lai, F.Y. Exploring the Role of Photolysis in the Aquatic Fate of Antimicrobial Transformation Products: Implications for One Health. *ACS ES&T Water* **2025**, *5*, 4112–4119. [\[CrossRef\]](#)
101. Egwu, L.S.; Enayaba, O.F.; Ajiboye, A.A.; Damoye, T.; Ogundeji, I.S.; Agbams, P. A Review of Machine Learning Techniques Applications in Environmental Science. *Int. J. Sci. Res. Technol.* **2024**, *6*. [\[CrossRef\]](#)
102. Gallitto, G.; Englert, R.; Kincses, B.; Kotikalapudi, R.; Li, J.; Hoffschlag, K.; Bingel, U.; Spisak, T. External Validation of Machine Learning Models—Registered Models and Adaptive Sample Splitting. *Gigascience* **2025**, *14*, giaf036. [\[CrossRef\]](#)
103. Xie, S.; Zhu, H.; Huang, N. AI-Designed Molecules in Drug Discovery, Structural Novelty Evaluation, and Implications. *J. Chem. Inf. Model.* **2025**, *65*, 8924–8933. [\[CrossRef\]](#)
104. Khater, T.; Alkhatib, S.A.; AlShehhi, A.; Pitsalidis, C.; Pappa, A.M.; Ngo, S.T.; Chan, V.; Truong, V.K. Generative Artificial Intelligence Based Models Optimization towards Molecule Design Enhancement. *J. Cheminform.* **2025**, *17*, 116. [\[CrossRef\]](#)
105. Gao, W.; Luo, S.; Coley, C.W. Generative Artificial Intelligence for Navigating Synthesizable Chemical Space. *arXiv* **2024**, arXiv:2410.03494. [\[CrossRef\]](#)
106. Blümmel, T.; Rehn, J.; Mereu, C.; Graf, F.; Bazing, F.; Kneuer, C.; Sonnenburg, A.; Wittkowski, P.; Padberg, F.; Bech, K.; et al. Exploring the Use of Artificial Intelligence (AI) for Extracting and Integrating Data Obtained through New Approach Methodologies (NAMs) for Chemical Risk Assessment. *EFSA Support. Publ.* **2024**, *21*, 8567E. [\[CrossRef\]](#)
107. Kamal Samanta, S. Recent Developments in Artificial Intelligence (AI) in Chemical Research. *Int. J. Innov. Res. Technol.* **2025**, *12*, 1476–1479.
108. Oluwole Adegun, A.; Akinyele Akinnifesi, T.; Omotunde Adejumo, M.; Moriliat Hashimi, A.; Olufemi Adewuyi, G.; Adeniyi Aseperi, K.; James, B. Integrating AI into Environmental Chemistry: Predicting and Preventing Toxicity in Chemical Products. In Proceedings of the 10th Annual Symposium of the American Chemical Society (ACS) Nigeria International Chemical Sciences, Abuja, Nigeria, 4–7 May 2025.
109. Pasas-Farmer, S.; Jain, R. From Discovery to Delivery: Governance of AI in the Pharmaceutical Industry. *Green Anal. Chem.* **2025**, *13*, 100268. [\[CrossRef\]](#)
110. OECD. (Q)SAR Assessment Framework: Guidance for the Regulatory Assessment of (Quantitative) Structure Activity Relationship Models and Predictions, 2nd ed.; Series on Testing and Assessment No. 405; OECD Publishing: Paris, France, 2024.
111. Alem, K.; Dagnew, M.; Gizachew, M.; Gelaw, B.; Moges, F. Environmental Antimicrobial Resistance: Key Drivers, Hotspots, Innovative Strategies, and Challenges in the Fight Against Superbugs. *Microbiologyopen* **2025**, *14*, e70067. [\[CrossRef\]](#) [\[PubMed\]](#)
112. Hanser, T.; Ahlberg, E.; Amberg, A.; Anger, L.T.; Barber, C.; Brennan, R.J.; Brigo, A.; Delaunois, A.; Glowienke, S.; Greene, N.; et al. Data-Driven Federated Learning in Drug Discovery with Knowledge Distillation. *Nat. Mach. Intell.* **2025**, *7*, 423–436. [\[CrossRef\]](#)
113. Waldock, W.J.; Thould, H.; Chindelevitch, L.; Croucher, N.J.; de la Fuente, C.; Collins, J.J.; Ashrafian, H.; Darzi, A. Mitigating Antimicrobial Resistance by Innovative Solutions in AI (MARISA): A Modified James Lind Alliance Analysis. *Npj Antimicrob. Resist.* **2025**, *3*, 75. [\[CrossRef\]](#)
114. Zhao, Z.; Ma, D.; Chen, L.; Sun, L.; Li, Z.; Xia, Y.; Chen, B.; Xu, H.; Zhu, Z.; Zhu, S.; et al. Developing ChemDFM as a Large Language Foundation Model for Chemistry. *Cell Rep. Phys. Sci.* **2025**, *6*, 102523. [\[CrossRef\]](#)
115. Lee, J.; Shin, S.; Kim, T.; Park, B.; Choi, H.; Lee, A.; Choi, M.; Lee, S. Physics Informed Neural Networks for Fluid Flow Analysis with Repetitive Parameter Initialization. *Sci. Rep.* **2025**, *15*, 16740. [\[CrossRef\]](#)
116. Stanic, B.; Andric, N. Mapping the Rise in Machine Learning in Environmental Chemical Research: A Bibliometric Analysis. *Toxics* **2025**, *13*, 817. [\[CrossRef\]](#)

117. Ren, Z.; Zhou, S.; Liu, D.; Liu, Q. Physics-Informed Neural Networks: A Review of Methodological Evolution, Theoretical Foundations, and Interdisciplinary Frontiers Toward Next-Generation Scientific Computing. *Appl. Sci.* **2025**, *15*, 8092. [[CrossRef](#)]
118. Sturm, P.O.; Wexler, A.S. Conservation Laws in a Neural Network Architecture: Enforcing the Atom Balance of a Julia-Based Photochemical Model (v0.2.0). *Geosci. Model. Dev.* **2022**, *15*, 3417–3431. [[CrossRef](#)]
119. Ruiz-Gonzalez, A. AI-Driven Chemical Design: Transforming the Sustainability of the Pharmaceutical Industry. *Future Pharmacol.* **2025**, *5*, 24. [[CrossRef](#)]
120. Alam, F.; Mohammed Alnazzawi, T.S.; Mehmood, R.; Al-magthawi, A. A Review of the Applications, Benefits, and Challenges of Generative AI for Sustainable Toxicology. *Curr. Res. Toxicol.* **2025**, *8*, 100232. [[CrossRef](#)]
121. Hartung, T. AI, Agentic Models and Lab Automation for Scientific Discovery—the Beginning of ScAInce. *Front. Artif. Intell.* **2025**, *8*, 1649155. [[CrossRef](#)]
122. Yang, M.; Zhang, P. The Data-Centric Paradigm in Synthetic Chemistry: AI Reaction Modeling Needs More than Reactant-Product Pairs. *Commun. Comput. Chem.* **2025**, *7*, 281–288. [[CrossRef](#)]
123. Cao, H.; Oghenemaro, E.F.; Latypova, A.; Abosaoda, M.K.; Zaman, G.S.; Devi, A. Advancing Clinical Biochemistry: Addressing Gaps and Driving Future Innovations. *Front. Med.* **2025**, *12*, 1521126. [[CrossRef](#)]
124. Abosede, O.O.; Obasi, E.C.M. Artificial Intelligence (AI) in Green Chemistry: A Transformative Approach. In *AI and Curriculum Development for the Future*; National Universities Commission: Abuja, Nigeria, 2025.
125. Nagar, V. Artificial Intelligence and Machine Learning in Green Chemistry: A Future Perspective. *Int. J. Emerg. Technol. Innov. Res.* **2025**, *12*, 646–655.

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.