
*Review***Applicability domains of neural networks for toxicity prediction****Efrén Pérez-Santín¹, Luis de-la-Fuente-Valentín¹, Mariano González García¹, Kharla Andreina Segovia Bravo¹, Fernando Carlos López Hernández² and José Ignacio López Sánchez^{1,*}**

¹ Escuela Superior de Ingeniería y Tecnología (ESIT), Universidad Internacional de La Rioja (UNIR), Av. de la Paz 137, Logroño, La Rioja 26004, Spain

² Applied Mathematics Department, Faculty of Mathematics, Universidad Complutense de Madrid

* **Correspondence:** Email: joseignacio.lopez@unir.net; Tel: +34941210211.

Abstract: In this paper, the term “applicability domain” refers to the range of chemical compounds for which the statistical quantitative structure-activity relationship (QSAR) model can accurately predict their toxicity. This is a crucial concept in the development and practical use of these models. First, a multidisciplinary review is provided regarding the theory and practice of applicability domains in the context of toxicity problems using the classical QSAR model. Then, the advantages and improved performance of neural networks (NNs), which are the most promising machine learning algorithms, are reviewed. Within the domain of medicinal chemistry, nine different methods using NNs for toxicity prediction were compared utilizing 29 alternative artificial intelligence (AI) techniques. Similarly, seven NN-based toxicity prediction methodologies were compared to six other AI techniques within the realm of food safety, 11 NN-based methodologies were compared to 16 different AI approaches in the environmental sciences category and four specific NN-based toxicity prediction methodologies were compared to nine alternative AI techniques in the field of industrial hygiene. Within the reviewed approaches, given known toxic compound descriptors and behaviors, we observed a difficulty in being able to extrapolate and predict the effects with untested chemical compounds. Different methods can be used for unsupervised clustering, such as distance-based approaches and consensus-based decision methods. Additionally, the importance of model validation has been highlighted within a regulatory context according to the Organization for Economic Co-operation and Development (OECD) principles, to predict the toxicity of potential new drugs in medicinal chemistry, to determine the limits of detection for harmful substances in food to predict the toxicity limits of chemicals in the environment, and to predict the exposure limits to harmful substances in the workplace. Despite its importance, a thorough application of toxicity models is still restricted in the field of medicinal chemistry and is virtually overlooked in other scientific domains. Consequently, only a small

proportion of the toxicity studies conducted in medicinal chemistry consider the applicability domain in their mathematical models, thereby limiting their predictive power to untested drugs. Conversely, the applicability of these models is crucial; however, this has not been sufficiently assessed in toxicity prediction or in other related areas such as food science, environmental science, and industrial hygiene. Thus, this review sheds light on the prevalent use of Neural Networks in toxicity prediction, thereby serving as a valuable resource for researchers and practitioners across these multifaceted domains that could be extended to other fields in future research.

Keywords: applicability domain; OECD principles; quantitative structure-activity relationship (QSAR); toxicity; machine learning

Mathematics Subject Classification: 68T07

1. Introduction

Chemical toxicity is a matter of growing concern due to the harmful effects that millions of the chemical agents used by industry can have regarding human health and the environment. It can be tested through a variety of criteria (chronic, acute, specific to a certain organ such as eye or skin corrosion/irritation or sensitization, as well as potential carcinogen or genotoxic, etc.), and is measured by several quantitative and qualitative criteria (LD₅₀, low, moderate or high toxicity, to name a few). To minimize any potential damage, new chemicals within the industry must be approved by the authorities prior to production and commercialization, and is subjected to a prior behavioral analysis in contact with humans, animals, and the environment [1–4].

Traditional toxicity tests require the application of chemicals in animals, usually rodents. However, animal experimentation is becoming increasingly controversial due to ethical and practical issues [5,6]. Therefore, there has been a recent trend towards the substitution of animal (*in vivo*) tests with *in vitro* laboratory models and computational (*in silico*) methodologies [7]. Thus, alternatives to animal experimentation are required according to the following 3R principles: 1) replacement, which is the development of alternative methods to the use of experimental animals; 2) reduction, which is the minimization of animal testing if total substitution is not possible; and 3) refinement, which is providing all possible welfare measures during animal life. Against this backdrop, *in silico* methodologies have taken an important role in the estimation of the toxicological properties of chemical compounds. Currently, the main *in silico* toxicity prediction methodologies are defined in Table 1 [8].

The science behind the relationships between theoretical molecular representations through descriptors and molecular experimental properties is an interdisciplinary research area [9]. The Organization for Economic Co-operation and Development (OECD) guidelines, established with the assistance of thousands of experts from OECD member countries, comprise internationally accepted standard methods for safety testing and the assessment of chemicals (pesticides, personal care products, industrial chemicals, etc.), alongside guiding decision-making processes for emergency responses. Regarding the replacement of animal experimentation with computational methodologies for the assessment of chemical compounds, the OECD is at the forefront in the publication of beneficial practice guidelines. They have developed five principles for the use of computational techniques in a regulatory context, which are an internationally accepted reference, known as the “OECD Principles for the Validation, for Regulation Purposes, of (Quantitative) Structure-Activity Relationship Models” [10]. Specifically, these principles are as follows: (a) a defined endpoint; (b) an unambiguous algorithm; (c)

a defined applicability domain (AD); (d) appropriate measures of suitability vis-à-vis fit, robustness; and (e) predictivity, which is a mechanistic interpretation, if possible.

Table 1. Main *in silico* toxicity prediction methods.

Methodology	Characteristics	Advantages	Limitations
Statistical based	Chemical structures and toxicity responses are known. They are correlated using Quantitative Structure-Activity Relationship (QSAR) models	Predicts a response to new structures	There must be database of the similar chemical structure
Expert rules or alerts	Structural rules created by experts based on their expertise	Predicts the potential toxicity of a new molecule when a particular potentially toxic substructure is included	This must be expert rule-based or have alerts about substructures
Read-across or semi-manual approaches	Predictions from historical data by identifying specific structural categories and analogues based on mechanism.	Acceptable predictions	Semi-manual application
Other approaches	Use of quantum descriptors: EHOMO - ELUMO differences	Acceptable predictions	Need for expertise in quantum chemistry

Additionally, three significant guiding principles assist the Quantitative Structure Activity Relationship (QSAR) model development for toxicity predictions: 1) simplicity, which is keeping the models as simple as possible by using the fewest and simplest descriptors conceivable, avoiding overtraining, supporting interpretation and ensuring the broadest model's domain; 2) transparency, which is providing the structural basis for the prediction, the descriptor weighting, and the indication of biological significance, and 3) utility, which is providing information to support a hazard or risk assessment such as classification, specific toxicity, or dosage for the toxic effects to appear.

Specifically, in recent years, the European Union (EU) has been making important strides towards REACH/3R principles, fostering the use of computational and statistical QSAR methodologies for the prediction of chemical toxicity properties, which are critical for regulatory aspects in many industries such as the pharmaceutical, food and environmental sectors and for industrial hygiene promotion. General principles considered by the OECD have been formally declared fundamental tools in estimating data on chemicals using QSARs models [11]. Additionally, the Chemical Policy of the European Commission, known as REACH [12], obliges the registrant to include information from alternative sources (*e.g.*, from *in silico* studies), which can, in certain cases, replace animal tests [13] if reliable estimates from validated *in silico* models can be produced. Consequently, model validation is a subject of recent considerable debate in the scientific and regulatory communities [9].

2. Background and applicability domain

Historically, a series of statistical techniques have been developed aiming to predict the effects chemical products. These techniques study the active effect of a chemical and thus have been grouped under the name QSAR. In real terms, depending on the endpoint (that is, the biological activity

predicted), these techniques have received more specific names such as Quantitative Structure Toxicity Relationship (QSTR), Quantitative Structure Property Relationship (QSPR), which aims to find the toxicity/properties of a chemical, Quantitative Structure-Metabolism Relationship (QSMR), and Quantitative Structure-Reactivity Relationship (QSRR), to name a few. For the aims of this research, the Quantitative Structure *Toxicity* Relationship is especially relevant. QSAR approaches rely on a basic chemical principle, stating that the biological activities of compounds are associated with the arrangement of their atoms (encoded in terms of a series of parameters called molecular descriptors), and therefore, structurally related molecules should possess somewhat similar biological activities [14].

Traditionally, QSAR has emerged from the analysis of correlations of similar compounds, aiming to identify activity/toxicity of an untested compound. In this analysis, the so-called *descriptors* are the chemical attributes [15]. Usually, these descriptors are compiled in tables as quantitative values. Ideally, the descriptor should be relevant to a broad class of compounds and must correlate with the human body response. The term *features* is frequently used in the literature to refer to the subset of descriptors that allow the response to be predicted. Frequently, these features are chosen by experts based on their previous experience. The endpoints (*i.e.*, the responses) obtained from the descriptors are mathematically expressed as a function of the descriptors, whereby:

$$\text{Biological response or activity} = f(\text{molecular descriptors}).$$

According to the dimensions of the descriptors involved, models are referred to as 1D, 2D, 3D, 4D, and so forth. The idea of correlation evolved to the use of regressions models, which, depending on the dose, allows for the inclusion of toxicity. In regression models, the predictors are usually quantitative, and the response is the toxicity of tested or untested chemicals.

QSAR models can be grouped as correlation-linear and non-linear, among other possible forms of classification. The classical QSAR methods have evolved to include machine learning (ML) techniques, which entails superior extrapolation capabilities. ML techniques can be divided into the following: i) regression based methods, where multiple linear regression (MLR) methods allow for a correlation between the independent (molecular descriptors) and dependent variables or endpoints (biological or physiochemical properties); ii) clustering based methods, in which data is placed into *ad-hoc* groups according to metrics such as similarity or Euclidean distance for clustering (maximizing similarity and dissimilarity within groups), and iii) classification methods, where data is assigned to a pre-defined set of categories, usually with ML techniques, such as the use of NNs, which mimic the behavior of biological neurons through the use of an input layer, several hidden layers and an output layer, support vector machine (SVM) or gene expression programming (GEP) and, more recently, convolutional neural networks (CNN) and transformer-based models (TBM), to name a few [14].

Thus, the application of *in silico* tools to predict toxicity and making pre-tests for their regulatory acceptance as part of product development has increased exponentially in recent years thanks to enhancements in model performance and simplicity, with the number of guidelines growing in order to support interpretations and to gain acceptance [8]. In this context, new ML prediction tools have been a disruptive tool that is being extensively applied to the problem of chemical toxicity with regulatory purposes. This subject has been reviewed in greater depth in [7].

AI encompasses a vast set of computational techniques capable of simulating the human process of thinking [16,17]. One prominent AI technique is ML, which allows for the construction of systems that can learn from the data and can be trained to predict a wide range of different outcomes. Among the ML approaches, NNs stand out as one of the most promising methods. In a typical configuration,

a NN starts with a random set of parameters and receives a large set of examples as an input. As the NN processes each sample, the parameters are adjusted so that they adapt their output to what is expected. At the end of the process, the NN performs either prediction classification or decision-making of unseen data with great accuracy.

Depending on the nature of the learning method, NNs can be divided into supervised, unsupervised and reinforcement learning. In supervised learning, each sample needs to be properly tagged prior to the training process. Thus, the system is fed with the sample (without the tag) and then provides an estimated tag as an output. A comparison between the expected and the estimated tags serves as feedback for the system, which iteratively adjusts its parameters until all the samples have each been processed several times (*i.e.*, several epochs). At the end of this training process, the system is able to assign tags to samples that do not appear in the training set. On the contrary, unsupervised learning does not require postulated tagged datasets. These types of systems are useful to identify features of a dataset, and they are good approaches for clustering tasks, for detecting characteristics that tend to appear grouped in a dataset (*i.e.*, underlying patterns), and for anomaly detection. Finally, in reinforcement learning, the goal of the system is to streamline cost-functions, which are calculated at each interaction with the environment. When a properly defined cost-function and a large enough number of iterations are in place, the system learns (and applies) behavioral aspects that minimizes costs. This technique is mainly used in robotics, where the robot needs to learn the best performing sequence of actions for a given task, without the need for a human to provide any previous instruction.

NNs base their operation on the connection between the so-called neurons, which are tiny processing units that typically receive a set of inputs and return a weighted sum of the inputs as an output. Those weights are the aforementioned parameters, which are randomly initialized and iteratively modify the weights of the network, with the network being structured in layers of neurons. When the number of layers is greater than two, the network is called a deep neural network (DNN), thus leading to the concept of deep learning (DL). Some examples of this type of learning are CNNs, which are primarily used for image processing or transformer-based networks, which are designed for textual input. One relevant characteristic of DNNs is that they do not need feature extraction as a preprocessing step for the input, that is, they can process an image pixel by pixel (or a voice from the very raw signal), while non-DL systems require a feature extraction process that translates the input into a numerical vector. This enables ML methods with input types where feature extraction is difficult to achieve.

DL methods are first-rate grading systems, since they can analyze a complex input and produce a simple classification output. In a reverse way, NNs can be trained to generate the complex values that would generate a given classification. In other words, they can devise images (or any type of signal). This is the goal of generative adversarial networks (GANs), which is a type of system that has been used to generate new molecular structures [18]. In brief, NNs learn from data through information abstraction in layers with non-linear processing units, such as DNNs. Thus, DL has been successfully transferred to *in silico* toxicology, allowing for reliable chemical toxicity predictions. These include deep convolutional neural networks (DCNNs) and transformer-based networks, to name but a few. Nowadays, ML-based computational *in-silico* toxicology has many advantages, not just because of ethical aspects, but also for frequently providing more reliable results than *in vivo* tests. Therefore, QSAR models have evolved with the arrival of ML and DL, and nowadays, they are intensively used for the prediction of toxicity endpoints for regulatory purposes [19].

Morger et al. (2021) [20] unveiled a study assessing the calibration of ML models for toxicity prediction, using the freely available Tox21 datasets. Conformal prediction (CP) was used to assess

the calibration of the models and to diagnose data drifts and other issues related to model calibration. While internally valid models could be trained using cross-validation on Tox21Train, predictions on the external Tox21Score data resulted in higher error rates than expected, concluding that conformal prediction could be used to diagnose data drifts and other issues related to model calibration. The applicability domain of the models used in this work was determined through traditional metrics.

Norinder (2022) [21] discussed the use of molecular structure property modelling as a tool for predicting compounds with desired properties, comparing a traditional physico-chemical descriptor and ML-based approaches with DL architectures, showing that for the binary Collaborative Acute Toxicity Modeling Suite (CATMoS) non-toxic dataset, all methods performed equally well; alternatively, for the binary CATMoS very-toxic dataset, the neural framework for molecular property prediction based on Bidirectional Encoder Representations from Transformers (Mol-BERT model) performed somewhat better compared to the rest, concluding that descriptor-free, Simplified Molecular Input Line Entry Specification (SMILES)-based, deep learning BERT architectures seem capable of producing well-balanced predictive models with defined applicability domains. Thus, through an intermediate Random Forest/auto encoder representation to deep learning BERT/molecular-graph-based approaches, the study, alongside results from the traditionally used Random Forest/physico-chemical descriptor approach, was capable of producing models with a defined applicability domain.

Nascimben & Rimondini (2023) [22] presented a study on the application of spiking neural networks (SNNs), which are a type of energy-efficient biologically inspired ML algorithms, for virtual screening of molecule databases targeting toxicity. Using structural information derived from molecular fingerprints applied to several public-domain toxicological datasets, including TOXCAST, Tox21, BBBP, SIDER, and Clintox, their work showed that SNNs obtained remarkable performance compared to previous models, advantageously and directly handling molecular fingerprints as binary inputs. Thus, the authors explored the potential use of neuromorphic computation solutions as an alternative to tackle the von Neumann bottleneck problem, suggesting that technological progress in neuromorphic computing could provide further applications to ‘chemoinformatics’, employing systems that reflect the mechanisms of brain activity, which is more understandable than standard “black box” approaches derived from NNs. The models’ applicability domain used in this work was not explicitly stated. The first principle of the OECD guidelines, which is the need to designate “a defined endpoint”, is related to the need to describe the goal of the study. Thus, toxicity endpoints to be predicted require special consideration. First, endpoints must be toxicologically relevant. Similarly, there should be enough structural and biological information to create and select a related dataset, which must be later cleaned, pre-processed, reduced and projected (*i.e.*, explored to select the most relevant descriptors by data reduction methods), while a mechanistic interpretation must similarly be provided. The endpoints that represent good candidates for predictive models according to regulatory agencies have been summarized above (*e.g.*, acute oral, dermal or inhalation toxicities, skin sensitization, repeated dose 28-day oral toxicity study in rodents, developmental neurotoxicity, carcinogenicity studies, chronic toxicity studies, genetic toxicology, to name a few).

Recently, Jingshan et al. (2021) [23] proposed a representative feature selection (RFS) method to select representative molecular descriptors for QSAR modelling by calculating the Euclidean distances and Pearson correlation coefficients, revealing that RFS effectively selects representative features from the feature space with information redundancy, thereby enhancing the performances of the QSAR model. The applicability domain of the QSAR model was determined using a distance-based approach, concluding that RFS with a proper clustering algorithm can effectively and automatically build a

multiple dimensional feature space for subsequent QSAR modeling. ML methodologies used in this work were the gradient boosting decision tree (GBDT) classifier, which is a classic ensemble learning algorithm that integrates multiple decision trees (DTs) to improve the prediction performance. Additionally, the authors used six clustering algorithms for the preliminary screening of molecular descriptors before performing the RFS method, including affinity propagation (AP), balanced iterative reducing and clustering using hierarchies (BIRCH), density-based spatial clustering of applications with noise (DBSCAN), k-means, mean shift, and ordering points to identify the clustering structure (OPTICS). The applicability domain of the QSAR model was determined through a distance-based approach to judge the reliability of the prediction for new molecules.

The second principle of the OECD guidelines, which is related to the statement of the hypothesis in the selection of the model and the confidence and consensus among models, is the need to “define the applicability domain”, which is the purpose of discussion in this work. Thus, the response and chemical structures for which the model is capable to obtain predictions with a certain reliability is known as the *applicability domain*, which should be capable of answering if a model is adequate to be applied to a chemical compound, though with a variety of possible methods such as those based on distances within the domains. It is unreasonable to expect a model to predict toxicity in any chemical, thereby deciding whether a model is suitable is a process of clustering compounds by similarity. Each application model is built using a training set with the descriptors and response. These models can be used to predict untested chemicals or to develop wholly new chemicals.

Considering that there is no unique way to define the applicability domain, methods for its estimation will be briefly reviewed hereafter, though this is not exactly the main topic of discussion. Moreover, the onus will be on how they are being applied in practice, if they are indeed, and scientific publications related to the development of toxicity prediction ML-based models in different areas, namely medicinal, food and environmental chemistry, and industrial hygiene. Even though proposals of harmonization are progressively more widespread [10], there is still a lack of consensus regarding how to define the applicability domain. Other OECD basics such as the need for an unambiguous algorithm, the appropriate measures of goodness of fit, robustness, and predictivity, as well as the mechanistic interpretation, are out of the scope of this review and are further discussed in depth in [9].

Logically, there is a difficulty in extrapolating other chemicals. The idea of the applicability domain comes from the statistical QSAR models and is related to the quality of predictions and the prevention of spurious extrapolations of results of the model. For this purpose, two general strategies have been proposed involving the statistical analysis of the training, adhering to the assumption that interpolated prediction results are more reliable than extrapolated, and the evaluation based on the similarity/diversity of the model descriptors space, considering the compound with respect to the training set and assuming that predictions should be more reliable if the problem compound is more similar to the ones in the training set, which is basically another means of extrapolation.

Specifically, the main strategies to find the applicability domain in QSAR models are the extent of extrapolation, the effective prediction domain, error estimation and residual standard deviation, and the similarity distance [24]. The extrapolation method's scope is based on the degree of extrapolation from the model when interpreting the descriptors of a problem compound, on relation to a maximum extrapolation that is defined as the reliability limit, to the extent that the prediction is considered unreliable if the response is the result of the significant extrapolation of the model. On the other hand, an effective prediction domain is a methodology applied to regression-like models, which is especially useful for those models with significantly correlated descriptors, according to which the estimations

of the regression model should be considered reliable only inside or near the periphery of the effective prediction domain (EPD). Moreover, the residual standard deviation approach can be used to evaluate the applicability domain through the calculation of the residual standard deviation (RSD) of the descriptor values produced for a test compound. On the contrary, similarity distance methods determine the applicability domain as a function of the chemical similarity in such a way that major similarity distances generally mean that the query compounds are overly different to the training set compounds for reliable predictions to be performed. In this case, numerical approaches can be taken, such as calculating the Euclidian distance (*i.e.*, geometric distance), which assumes that the model space is spherically distributed, or the Mahalanobis distance (*i.e.*, probability distance), which accepts that the shape of the model space is more like an “ellipsoid” [25].

Thus, the most widely used idea in addressing the applicability domain issue is that of similarity, assessing how similar the new surveyed problem compound is to the population of currently available training compounds. Even though the concept of an applicability domain arose at the beginning of the development of QSAR models, there are still many studies which either confuse or ignore this concept [26].

The methodology adopted for conducting the literature review aimed to identify and analyze relevant research articles related to the AD, application of computational models based on AI and NNs in the field of toxicity prediction. The review primarily focused on the specialties of medicinal chemistry, environmental science, industrial hygiene, and food safety. The following steps were followed to perform the literature review:

1. **Keyword Search Strategy:** A systematic search was initiated using a set of specific keyword combinations. The initial search utilized the terms "toxicity prediction" AND the specialties "medicinal chemistry," "environment," "industrial hygiene," and "food". Additional searches were performed using the terms "applicability domain" AND the same set of specialties. The keywords "artificial intelligence" AND "toxicity prediction" were also used to retrieve relevant literature. Similarly, the combination "neural networks" AND "toxicity prediction" was used to complement the literature retrieval.
2. **Primary Search and Filtering:** The primary search yielded a collection of research articles that was further refined. Publications within the last 15 years were considered for inclusion in the review. The focus was placed on articles discussing "computational models based on AI" and "computational models based on neural networks" applied to toxicity prediction, in which information was provided about the applicability domain of the models. The results were categorized based on the following specialties: medicinal chemistry, environment, industrial hygiene, and food safety.
3. **Secondary Searches:** The articles obtained from the primary search were subjected to secondary searches. Relevant references cited within the selected articles were reviewed to ensure a comprehensive coverage of the topic. Moreover, articles that cited the initially selected papers were examined to capture the latest developments in the field.
4. **Inclusion and Exclusion Criteria:** The inclusion criteria for medicinal chemistry were papers explicitly mentioning the method for determining the model applicability domain or the applicability of the model to specific substances or substance groups. In the case of food safety, articles were included if they provided information on the applicability of the constructed models to specific substances, substance groups, or individual foods. For environmental sciences, articles mentioning the method for determining the model applicability domain or

the applicability of the model to specific substances or substance groups were considered. Industrial hygiene articles were included if they provided information on the applicability of AI and NN models to specific substances, substance groups, or specific areas within occupational hygiene.

5. Review of Collections/Book Chapters: To supplement the review, important collections and book chapters pertaining to the field were consulted to provide a broader perspective and context for the literature review process. No previous bibliographic reviews focusing on the specific topic of this review were found, that is, the application of techniques to determine the domains of applicability of the computational methods based on AI and NNs used in the determination of toxicity in a multidisciplinary context.
6. Data Extraction: Data from the selected articles were extracted and organized based on the specialties of interest.

The final selection of relevant articles was made based on the comprehensive analysis of their content with respect to the applicability of computational models in toxicity prediction in the specialties specified. This methodology ensured a rigorous and systematic approach to identifying and analyzing the most pertinent research articles in the field.

3. Applicability domain in medicinal chemistry

Hereinafter, the most significant results are presented from the review of a representative sample of papers published around toxicity prediction through QSAR methods applied in different computational medicinal chemistry studies. Thus, in this section, 70 relevant papers were reviewed covering the years from 2010 to 2023, of which approximately half of them made an explicit mention of the applicability domain; however, only a few reported an explicit methodology for the calculation. Only those deemed relevant will be discussed (see also Table 2).

Merely 10% of the new molecules developed as potential medicines that are tested in Phase 1 are finally approved by the FDA, mainly due to unacceptable preclinical toxicities [7]. Considering that toxicity predictions based on animal models fail in a very high percentage of cases when extrapolated to humans, thereby making traditional toxicological safety studies controversial, a major potential has been unearthed in the use of *in silico* models based on ML techniques to foresee the toxicological properties of new molecules.

More specifically, drug-induced liver injury (DILI) is the leading cause of post-marketing withdrawals of approved drugs, owing to reasons such as hepatotoxicity or liver necrosis, to name a few. To avoid discrepancies between *in vitro* and *in vivo* results, several models have been developed based on ML-computational approaches. Kotsampasakou & Ecker (2017) [27] applied a k-nearest neighbors' strategy to predict cholestasis by means of a set of 93 two-dimensional (2D) physicochemical descriptors, thereby predicting several selected hepatic transporters' inhibition. The AD of the models was verified on the basis of the Euclidean distances on the KNIME software with the Enalos nodes, though the exact number of reliable predictions was not provided since the inhibition model was generated with confidential training data. Additionally, Minerali *et al.* (2020) [28] generated and compared ML algorithms to predict DILI with the Assay Central software [29], thereby obtaining a Bayesian probability-like and applicability score for individual chemicals in which the applicability of the model could be inferred. Moreover, Mora *et al.* (2020) [30] developed and distributed a freely available software package based on ML models [31] for DILI prediction, thereby obtaining a broad

applicability domain with a coverage of 98% for the molecules in two test sets and five external test sets. The AD analysis was carried out with a consensus-based decision from five different approaches within the Ambit Discovery software (city-block, Euclidean, Mahalanobis, range, and density). If the prediction for the target molecule lies outside of the bounds in at least three of the AD methods (consensus-3), then it is considered unreliable. Additionally, they analyzed the AD on an external dataset to determine the reliability of the predictions with five different AD methods (PCARange, Euclidean distance, city-block distance, Mahalanobis distance, and probability density).

Wu et al. (2019) [32] combined several ML methods (NN, SVM, RF and kNN) to study hepatotoxicity cases inspired by traditional Chinese medicine, reporting an AD covering 90.6% of the traditional Chinese medicine systems pharmacology database of compounds (TCMSP). However, they did not provide information on the methodology utilized to calculate the AD. Moreover, Hussain et al. (2020) [33] developed multiple feature models combining *in vivo* and ML methodologies to predict hepatocyte toxicity in humans. Even though they reported the use of a diverse set of xenobiotics to ensure the model's wider applicability, no specific method for the determination of the AD was provided.

On the other hand, skin sensitization is a significant endpoint in the field of drug discovery and cosmetics, as it is one of the most frequent forms of human immune toxicity. Thus, Peiwen Di et al. (2019) [34] developed and compared a variety of ML models [35] determining the ADs to rationalize the results. ADs were assessed using the Enalos Domain-Similarity node, which is a freely available software available via the KNIME Community [36] and the website of NovaMechanics [37], based on the Euclidean distances among training compounds and test compounds. APD values and Euclidean distances were calculated for the most suitable models, considering the compound within the applicability domain when the Euclidean distance was lower than APD value, thereby demonstrating the wide applicability domains of models.

Furthermore, considering that the main adverse drug reaction (ADR) known is cardiotoxicity caused by the blockade of the hERG potassium channel, Ogura et al. (2019) [38] applied ML techniques to predict hERG inhibition by integrating multiple databases with nearly 300,000 total compounds. The applicability domain of the prediction model was assessed based on the molecular similarity between the training set and test set compounds, which was also based on the molecular similarity to the training compounds. Thus, the relationship between the structural similarity and the prediction accuracy was assessed, assuming that the prediction of a compound, similar to those in the training set, could be reliable. Authors argued that a similarity-based approach was applied for the analysis of the applicability domain instead of a probabilistic approach, since the data distribution was too high-dimensional and sparse to estimate the data distribution. Compounds with a high similarity to the training set showed an increased prediction accuracy, whilst a decrease of sensitivity and an increase in false negative compounds was found in general for the compounds featuring lower similarities. The model was released publicly alongside the integrated database [39].

Considering the increasing importance of carcinogenesis in drug discovery, Fjodorova et al. (2010) [40] implemented a methodology based on counter propagation neural networks (CPNN) to develop models for the prediction of carcinogenic potency according to specific requirements of chemical regulatory agencies. A tool for the general evaluation of the AD was implemented based on the descriptor range in the dataset, assuming that the predicted values for chemicals outside the descriptor range would be less reliable. Considering that the chemical space characterized by the descriptor range did not reflect the density of compounds distribution, and to avoid erroneous interpretations when the

target chemical is in a poorly represented area in the training set, as the AD was based on the chemical descriptors alone, the authors developed a tool for a further AD assessment. This was grounded on a similarity score of the six most similar chemicals in the training set, which was worthwhile for appraising whether these compounds are truly representative for the unknown compound, and thereby offering a visualization to be independently used to evaluate the compounds. These models can be accessed through a java-based web application “CAESAR Application”.

Additionally, Singh et al. (2013) [41] designed a probabilistic neural network (PNN) prediction model with five descriptors of more than 800 items of structural data to anticipate the carcinogenicity of diverse chemicals, taking into account the AD, represented as an optimum prediction space that is a function of the ranges of molecular descriptors in the training set compounds (descriptor space of the training set), using statistically based method, and assuming the data distribution to be normal. The interpolation region was defined by a two-dimensional descriptor space demarcated a rectangle in one plane, which is the interval between the minimum and maximum values of the training dataset.

Moreover, Zhang et al. (2017) [42] developed a method based on ensemble support vector machine and random forest for the prediction of carcinogenicity that was implemented on an online web server (CarcinoPred-EL [43]) together with a user-friendly version called CarcinoPred-EL [43]. Authors claimed a broader AD of their model compared to the state of the art, though no explicit method for its calculation was provided.

Additionally, Guan et al. (2018) [44] developed a variety of ML algorithms for the prediction of carcinogenic properties, comparing the applicability domain of each chemical dataset against an independent external validation dataset comprised of pharmaceutical chemicals, visualizing the applicability domain by the PCA of seven physicochemical properties (descriptors), thus allowing for the visual representation and comparison of the number and types of molecules included in the datasets. However, weighting each prediction by an exact measure of the applicability domain remains the realm of future works.

On the other hand, Bloomingdale and Mager (2019) [45] developed ML models for the prediction of chemotherapy-induced peripheral neuropathy, determining the applicability domain by means of the software package ADMET Predictor™.

The cytotoxic effect of chemical substances is an important endpoint that has been traditionally obtained by means of the application of expensive and arduous *in vivo* models to predict drug pharmacology and toxicity. Consequently, ML methods are being developed by different researchers for their use in the web servers ProTox-II [46] and admetSAR [47–51]. Six physicochemical and topological properties (descriptors) were used to define the applicability domain, thus analyzing the distribution of these properties in all the training sets of the predictive models [50]. This was performed in such a manner that compounds with either the molecular weight or the AlogP higher than 99% or lower than 99% of the training set, which would be regarded as warning signs, as well as outside the domain if higher than the maximum of the training set. For the other four descriptors, only compounds out of upper bounds were tagged.

Moreover, Webel et al. (2020) [52] predicted cytotoxicity using a DL approach trained with a dataset of over 34,000 compounds, implementing a Deep Taylor Decomposition method to identify the substructures responsible of the cytotoxic effects, as well as making use of cytotoxicity maps for a visual structural interpretation of the relevance of these substructures, as a novel attempt to identify the applicability domain with the aim to avoid the black box issue and obtain a deeper mechanistic understanding of the model.

Considering that convulsive seizures or epilepsy is a troublesome area when predicting toxicological endpoints in the preclinical safety assessment of drugs developments, Antanasijević et al. (2017) [53] developed a groundbreaking modular NN approach to predict the anticonvulsant activity of succinimides, analyzing the AD of the model through the previously proposed simple statistical approach based on the theory of the standardization by Roy et al. (2015) [54].

Taking the importance of the Hemolytic toxicity into account as an endpoint for small molecules, Zheng et al. (2020a,b) [55,56] built ML-based models from a manually collected hemolytic toxicity dataset of more than 800 small molecules, providing a software package (e-Hemolytic-Regression) with automatic verification of AD to the input molecule, with an average-similarity > 0.15 , $0 < MW < 1500$, and $0 < cLogP < 10$.

Additionally, Plisson, Ramírez-Sánchez and Martínez-Hernández (2020) [57] developed ML models capable of predicting the hemolytic activity of antimicrobial peptides (AMPs), comparing 14 algorithms including decision tree (CART), random forest (RF), gradient boosting (GBC), adaptive boosting (AB), logistic regression (LOGREG), support-vector machine (SVM) and K-nearest neighbors (KNN) classifiers, and assessing nine outlier detection (OD) methods to define the applicability domains through the Mahalanobis distance (MD) that allowed for the reduction of multidimensional datasets (*i.e.*, 56 descriptors) into a single dimension.

Conversely, Feng et al. (2021) [58] predicted the reproductive toxicity of chemicals using ensemble learning methods and molecular fingerprints in a study that was conducted by a team of researchers from various institutions in China and Israel; they developed ensemble learning models to predict the reproductive toxicity of compounds using support vector machine, random forest, and extreme gradient boosting methods and 9 molecular fingerprints calculated for a dataset containing 1,823 chemicals. The best prediction performance was achieved by the Ensemble-Top12 model, with an accuracy rate of 86.3%, sensitivity at 82.0%, specificity at 90.2%, and an area under the receiver operating characteristic curve of 0.937 in 5-fold cross-validation. The AD of the standout model (*i.e.*, Ensemble-Top12) was defined by calculating the Tanimoto distance of the training set based on AD2D, EState, KR, MACCS, Pubchem, and FP4 fingerprints, thus identifying outliers and compounds residing outside the AD. Moreover, the effect of different Tanimoto distance values on the model performance was evaluated to identify a suitable distance threshold, thereby balancing the model's predictability with a Tanimoto distance of 0.3 recommended as the AD threshold. Additionally, several fingerprint features related to the chemical reproductive toxicity were identified. Using the Ensemble-Top12 model could offer an advantage in predicting reproductive toxicity in early drug development phases.

Furthermore, Zhao et al. (2021) [59] developed binary classification models to predict the mitochondrial toxicity of chemicals, collecting 3,407 chemicals associated with mitochondrial toxicity from literature and databases and using nine molecular fingerprints and five ML methods to construct 45 prediction models. Then, the models were assessed using a 10-fold cross validation and a test set, and their applicability domain was defined using the Euclidean distance method. Seven structural alerts related to mitochondrial toxicity were identified, providing valuable assistance to pharmaceutical chemists in the early stages of drug design, as well as contributing to the assessment of mitochondrial toxicity of environmental chemicals. The ML algorithms used to build the mitochondrial toxicity prediction models were KNN, LR, RF, SVM, and XGB. The applicability domain of the prediction models was defined using the Euclidean distance method, employing appropriate K and Z values ranging from 3 to 6 for K and ranging from 0.6 to 0.9 for Z, optimally 3

and 0.7, respectively.

Considering the influence of plasma protein binding (PPB) in drug efficacy and toxicity, Yuan et al. (2020) [60] devised several ML-based QSAR models to predict PPB, defining the AD of the models by comparing the Euclidean distance between the problem compound and its adjacent neighbor in the training set.

Likewise, Chou & Lin (2023) [61] reviewed the emerging paradigm for integrating physiologically based pharmacokinetic (PBPK) modeling with ML, including obtaining time-concentration PK data and/or ADME parameters from publicly available databases, developing ML-based approaches to predict ADME parameters, and incorporating the ML models into PBPK models to predict PK summary statistics. Additionally, the neural network architecture called "neural ordinary differential equation (Neural-ODE)" was discussed, with improved predictive capabilities compared to other ML methods, concluding the high potential of ML approaches to facilitate the efficient development of sturdy PBPK models for a sizable number of chemicals, even though the applicability domain of the models or the techniques used for its determination were not specifically stated.

Bearing in mind the significance of neurotoxicity as a major cause of drug withdrawal, Changsheng Jiang et al. (2020) [62] used eight ML algorithms to predict chemical neurotoxicity, and thus obtained the applicability domain of the models by calculating the standard deviation distance (SDD) and leverage distance of the training set. Similarly, they employed a Williams' plot to display the distribution in two dimensions, with the aim of visually describing the scope of the AD. Moreover, the effect of each compound on the model was evaluated using Cook's distance. All AD analyses were performed through Python scripts.

If alterations in the normal functioning of estrogen and androgen receptors (ER and AR) may cause endocrine disruption and lead to adverse effects on health, Cui et al. (2019) [63] developed a series of ML models to predict drug-induced rhabdomyolysis (DIR), thereby identifying structural alerts responsible for DIR and providing the best model [64]. The AD was appraised based on the concept of distance-to-model (DM).

Additionally, Lagares et al. (2019) [65] used a CPNN based on 2D molecular descriptors to develop a model capable of predicting the probability of a compound to interact with P-gp, which is important in a toxicological assessment during drug discovery. The AD was analyzed using the ED between molecules and the central neuron of the NN, which represents the interval between a central node (c_i) in the Kohonen layer and an input pattern (X).

Table 2. Applicability domain in AI methodologies in toxicology and medicinal chemistry.

Computational methodology	Aim	Algorithm applied for the AD evaluation.	Key considerations	Literature
k-NN	Predicting Drug-Induced liver injury (DILI).	Euclidean distances from KNIME with the Enalos nodes.	Concise data on applicability domain not reported due to confidentiality issues	[27]
k-NN, SVC, DT, DL, and BA.	Predicting Drug-Induced Liver Injury (DILI)	Bayesian probability-like and applicability score	Assay Central™ software package	[28]
k-NN, MLP, RF, SVM, LR, CT, FLDA, Bnet, NB, and RF.	Predicting Drug-Induced Liver Injury (DILI)	Consensus-based decision. Five different approaches. If the problem molecule lies outside of the bounds in at least three of the AD methods (consensus-3), it is considered unreliable.	Freely available software package used. AD methods taken from the Ambit Discovery software.	[30]
NN, SVM, RF, and k-NN.	Hepatotoxicity prediction	Not provided	Applicability Domain covering 90.6% of the TCMSP.	[32]
SVM.	Hepatotoxicity prediction	Not provided	Diverse set of xenobiotics used to ensure the wider applicability of the model.	[33]
SVM, NB, DT, GB, RF, TE, PNN, MLP, and FR.	Skin sensitization	Euclidean distances	Applicability Domains assessed with Enalos Domain - Similarity node, freely available software.	[34]
SVM.	Cardiotoxicity	Similarity-based approach	Publicly released model (http://drugdesign.riken.jp/hERGdb).	[38]
CP NN.	Carcinogenesis	Distance to the descriptor range in the dataset, and similarity score with the six most similar chemicals in the training set.	Java-based web application “CAESAR Application”.	[40]
PNN.	Carcinogenesis	Distance to the descriptor space of the training set.	Distance calculated with statistically based methods, assuming a normal data distribution.	[41]
SVM, and RF.	Carcinogenesis	Not provided	CarcinoPred-EL, http://112.126.70.33/toxicity/CarcinoPred-EL/about.html).	[42]
Adaboost, k-NN, DT, MLP, and RF.	Carcinogenic properties	Principal Components Analysis of seven descriptors	Exact measure of the applicability domain stays as future work.	[44]
NN.	Prediction of Chemotherapy-Induced Peripheral Neuropathy	Not provided	Software package ADMET Predictor to estimate the AD.	[45]
RF, SVM, NB and, k-NN.	Cytotoxic effect of chemical substances	Similarity-based approach	Implemented in web servers ProTox-II (http://tox.charite.de/prottox_II/) and admetSAR (http://lmmmd.ecust.edu.cn/admetSar1) and (http://lmmmd.ecust.edu.cn/admetSar2/)	[47–51]
DL.	Cytotoxic effect of chemical substances	Deep Taylor Decomposition method to identify cytotoxic substructures.	Cytotoxicity maps used for a visual structural interpretation	[52]

Continued on next page

Computational methodology	Aim	Algorithm applied for the AD evaluation.	Key considerations	Literature
NN.	Anticonvulsant activity of succinimides	Statistical approach based on the theory of standardization	Based on the approach previously proposed by Roy et al. (2015) [54]	[53]
k-NN, SVM, RF, GBM, and ECFP.	Hemolytic toxicity	Similarity-based approach	Software package (e-Hemolytic-Regression) with automatic checking of applicability domain (AD) to the input molecule	[55,56]
DT, RF, GBC, AB, LR, SVM, and k-NN.	Hemolytic activity of antimicrobial peptides (AMPs)	Outlier detection (OD) methods to define the applicability domains through the Mahalanobis distance (MD)	Scripts available at: https://github.com/plissonf/ML-guided-discovery-and-design-of-non-hemolytic-peptides	[57]
k-NN, SVR, RF, BT, and GB.	Plasma protein binding (PPB)	Euclidean distance between problem compound and its adjacent neighbor in the training set	AD threshold of DT with a constant value of Z within 0.5-3.0.	[60]
BGR, ETR, GPR, k- NN, MLP, Nu-SVR, RF, and SVR.	Neurotoxicity	Standard deviation distance (SDD) and leverage distance of the training set.	Williams' plot used to visually describe the AD.	[62]
ASNN, SVM, LR, PLS, RF, and DNN.	Drug-induced rhabdomyolysis (DIR)	Distance to model (DM)	Best model provided: (https://ochem.eu/model/32214665)	[63]
CP NN.	Predicting the probability of a compound to interact with P-gp	Euclidean Distance (ED) between molecules and the central neuron of the neural network.	Comparing the Training Set (TR) and the Test Set (TE) chemical coverage versus false predicted chemical space.	[65]
SVM, RF, and EGB	Reproductive toxicity	Tanimoto distance of the training set.	Tanimoto distance of 0.3 recommended as the AD threshold.	[58]
KNN, LR, RF, SVM, and XGB.	Mitochondrial toxicity of chemicals	Euclidean distance method.	Ranging from 3 to 6 for K and from 0.6 to 0.9 for Z, optimally 3 and 0.7 respectively.	[59]
Neural-ODE.	Physiologically Based Pharmacokinetic (PBPK) modeling.	Not provided	Publicly available databases used to predict PK summary statistics.	[61]

Abbreviations: Adaptive Boosting (AB), Neural Networks (NN), Associative Neural Networks (ASNN), Bayesian Algorithms (BA), Bagging Regressor (BGR), Bayes Network (Bnet), Extra-Trees Regressor (ETR), Boost Tree (BT), Classification Tree (CT), Counter Propagation (CP NN), Decision Tree (DT), Deep Learning (DL), Deep Neural Network (DNN), Extended-Connectivity Fingerprint (ECFP), Extreme Gradient Boosting (XGB), Fast Stagewise Multivariate Linear Regression (FSMLR), Fisher's Linear Discriminant Analysis (FLDA), Fuzzy Rules (FR), Gaussian Process Regression (GPR), Gradient Boosting (GB), k-Nearest Neighbor (k-NN), Logistic Regression (LR), Multilayer Perceptron (MLP), Naïve Bayes (NB), Neural Network (NN), Neural Ordinary Differential Equation (Neural-ODE) Nu-Support Vector Regression (Nu-SVR), Probabilistic Neural Network (PNN), Random Forest (RF), Support Vector Classification (SVC), Support Vector Machine (SVM), Support Vector Regression (SVR), Traditional Chinese Medicine Systems Pharmacology Database (TCMSP), Tree Ensemble (TE), Multiple Linear Regression Analysis (LRA), Partial Least Squares (PLS).

4. Applicability domain in food safety

Food security is when the entire population always has access to sufficient, safe and nutritious food that meets their dietary needs and taste preferences for an active and healthy life [66]. However, there are currently many known harmful substances of natural and anthropogenic origin that can cause harm once inserted into the agri-food chain, including pharmacological and phytosanitary residues, environmental contaminants, harmful and unhealthy substances derived from food processing or from materials in contact with food, additives and banned food substances, among others [67].

Classical analytical techniques based on high performance liquid chromatography (HPLC) and either gas chromatography (GC) or UV-Vis spectrophotometry have evolved over the last decade towards the use of new non-destructive methods such as spectroscopic technologies, including hyperspectral imaging [68], fluorescence spectroscopy [69], near-infrared spectroscopy, Fourier transform infrared and Raman spectroscopy, biosensors [70], electronic nose [71] and electronic tongue [72]. These techniques are used in combination with various ML algorithms (see Table 3).

Histamine analysis in tuna can be performed using thin-layer chromatography (TLC) with surface-enhanced ultrasensitive Raman scattering (SERS), which was supported by ML methods that improve the reliability and reproducibility of identifications [73]. Thus, TLC-SERS combined with ML analysis is postulated as a reliable, sensitive and accurate technique for *in situ* detection and quantification of seafood allergens and avoidance of histamine poisoning. The model performance was evaluated for the training and test datasets and compared based on four criteria: the correlation coefficient squared (R^2), the root mean square error of cross-validation (RMSECV), the root mean square error of prediction (RMSEP) and the ratio of prediction deviation (RPD). The TLC-SERS quantitative detection method of Tan et al., (2019) [73] was used to disclose histamine at a detection level up to 10 ppm, which was limited to a tuna matrix with rapid, cost-effective and quantitative *in situ* detection. In general, histamine detection could be extrapolated to seafood; however, this application domain has yet to be validated.

The formation of acrylamide during the baking of cookies can be prevented by monitoring quality parameters such as oven temperature, humidity and browning index by means of several analytical techniques such as IR spectroscopy and colorimetry using ML based nonlinear polynomial (PLN) and NNs models in the forward and inverse phases [74]. The cookie baking's modelling process has been performed on conventional domestic ovens, though the specially designed NNs models could be embedded to the automatized industrial ovens.

On the other hand, the acrylamide content in cookies can be performed using a PCA and least squares support vector machine (LS-SVM) combined with fractal color for classification and results predictions [75]. Therefore, the method has been validated to quantify acrylamide in cookies through their color for a single recipe. In principle, the same method could be extended to cookies made with different recipes or processing methods. Similarly, the detection of acrylamide in images of French fries can be performed by automatic image processing and support vector machine classification [76]. The results of the study indicate that it can be an effective, non-destructive and sensitivity technique for use in food quality control of foodstuffs. The identification of acrylamide was carried out on either fried or baked potatoes, with a methodology that achieved an accuracy of 98.33%. However, the type of potatoes used and the possible interference with the original color were not specified.

Pesticide residues are a significant factor in food safety. A machine-vision-based method was combined with hyperspectral imaging and supported with CNNs to detect a variety of pesticides in apples [77]. Post-harvest samples of Fushi apples were collected as the study material. Specifically, four pesticides (chlorpyrifos, carbendazim, and two mixed pesticides) and one inactive control of the

same concentration of chlorpyrifos (100 ppm) were used. Although the method features the advantages of low time cost and high robustness, it can lead to problems when two hyperspectral images are in the same band and have similar appearances. In these cases, the network is prone to detection errors, thus highlighting a limitation of the application domain.

Aflatoxin is a virulent and strong type of carcinogenic substance and is widely found in peanuts, maize and their agricultural products. Hyperspectral imaging supported with convolutional neural networks showed a high and improved accuracy compared to traditional models to detect aflatoxin in peanuts [68]. In this case, the overall recognition rate obtained scores more than 95% in aflatoxin contaminated peanuts. The methodology could be extended for use in sorting machines.

Furthermore, a visible-near infrared hyperspectral image of lettuce with DL algorithms allowed for improved lead and cadmium content prediction, thus achieving a reasonable performance [78]. On this occasion, Italian annual lettuces grown at the University of Jiangsu, China, were selected, in which the samples had been grown without soil. However, the study was limited to two heavy metals (Cd and Pb) and a single type of lettuce variety. The researchers propose that the methodology could be sensitive to other varieties of lettuce, locations, and other agronomic factors. Another methodology for predicting Pb, Cd, and Hg residues in fish (*Carassius carassius*) samples in a market in China (Zhenjiang) has been reported using a low-cost and simple optical electronic system coupled with ML approaches [72]. This work displays the ability to simultaneously and quantitatively predict heavy metal residues in fish and could potentially be applied to other fish products.

Recently, a portable nano-biosensor system integrating SVM algorithms provided good results with an on-site and sensitive detection of several antibiotic residues in cow milk [70]. The nano-biosensors were constructed with gold nanoparticles highly selective for four widely used antibiotics in the field of veterinary medicine: Kanamycin, Ampicillin, Oxytetracycline and Sulfadimethoxine. Overall, this technology offers a combination of portability and sensitivity, thus making it suitable for on-site analysis for daily farm screening. It could be selectively developed against other antibiotics by modifying the specific recognition points of nano-biosensor.

A rapid on-line detection method was demonstrated for beverage preservatives based on X-ray absorption spectroscopies (XAS), complemented with DNN and SVM classifiers for data classification and prediction [79]. Benzoic acid, potassium sorbate, sodium dehydrogenate and propyl p-hydroxybenzoate were analyzed at standard concentrations; samples with the preservative content of the beverage exceeded the standard, thereby achieving rapid on-line detection of preservative content in beverages. The proposed method could be extended to other preservatives in market beverages. Analogously, ML methods have been applied to terahertz (THz) spectroscopy techniques to measure benzoic acid in wheat flour, with a high correlation coefficient [80]. In this context, the authors limit the results to the accurate detection and quantification of benzoic acid in wheat flour, without providing data on other additives or the extrapolation of the method to other food matrices. Similarly, benzoic acid and chitosan as food additives in fruit juices can be detected by electronic nose using chemometrics [71]. For quantitative monitoring, SVM, RF, extreme learning machine (ELM) and partial least squares regression (PLSR) were applied to establish regression models between E-nose signals. The methodology has been validated on Satsuma mandarins. The fruit was pressed to extract the juice and filtered to remove solid particles.

Table 3. Applicability domain in AI methodologies in foods.

Foodstuff	Origin in Food	Identification technique	Algorithm applied	Applicability domain	Potential to	Literature
Tuna	Histamine	Thin layer chromatography (TLC)-ultra-sensitive surface enhanced Raman scattering (SERS)	Principal component analysis-support vector regression (PCA-SVR algorithm)	Tuna	Seafood	[73]
Cookies	Acrylamide, in fried, baked or roasted food	Infrared moisture analyzer at 130°C (moisture content); K-type thermocouple (temperature); CIE L*, a*, and b* color values by colorimeter (browning index)	neural network (NN)	Baking cookies in domestic conventional oven	Industrial and domestic food baking processes	[74]
Biscuits	Acrylamide, in fried, baked or roasted food	Image acquisition and measure of Fractal color and the traditional L*a*b*, RGB (red, green blue), CMYK (cyan, magenta, yellow, black) color models	Least squares- support vector machine (LS-SVM)	Baking cookies	Biscuit baking in other formulations	[75]
Potato chips	Acrylamide, in fried, baked or roasted food	Image processing	Continuous wavelets transform (CWT) with Morlet wavelet/leave one out cross validation-based Support Vector Machine classification (SVM)	Fried or baked potato. No variety of potato been specified	Any type of potato.	[76]
Apple	Pesticides,	Hyperspectral imaging (Otsu segmentation algorithm)	Convolutional neural network (CNN, AlexNet)	Chlorpyrifos, carbendazim and mixture of these pesticides	Other pesticides with characteristic hyperspectral imagen	[77]
Peanut	Aflatoxins	Hyperspectral imaging system (Reshape image by pixel-level)	Convolutional neural network (CNN)	Aflatoxin in peanuts.	Aflatoxin in peanut in industrial sorting machines.	[68]
Lettuce	Heavy metals	Hyperspectral imaging technology	Wavelet transform-stack convolution auto-encoder (WT-SCAE)/support vector machine regression (SVR)	Cd and Pb in a variety of lettuce	Heavy metals in vegetables and fruit	[78]

Continued on next page

Foodstuff	Origin in Food	Identification technique	Algorithm applied	Applicability domain	Potential to	Literature
Fish	Heavy metals	Electronic tongue system	Extreme learning machine (ELM)	Pb, Cd and Hg in fish	Other fish products	[72]
Raw milk	Antibiotics	Nano biosensors	Support vector machine (SVM)	Four antibiotics in raw milk; Kanamycin, Ampicillin, Oxytetracycline, sulfadimethoxine	Others selectively antibiotics	[70]
Beverages	Additives	X-ray absorption spectrum (XAS)	Deep neural network (DNN) and Support vector machine (SVM)	Benzoic acid, potassium sorbate, sodium dehydrogenate and propyl phydroxybenzoate.	Other additives	[79]
Wheat flour	Additives	Terahertz spectroscopy	Least squares-support vector machine (LS-SVM)	Benzoic acid in wheat flour	Other flours	[80]
Fruit juices	Additives	Electronic nose (e-nose)	Random Forest (RF) and extreme learning machine (ELM)	Benzoic acid and chitosan in citrus juices	Food additives in juices or other types of food productions	[71]

Abbreviations: CMYK, cyan, magenta, yellow, black; CWT, continuous wavelet transform; DNN, deep neural network; ELM, extreme learning machine; LSSVM, least squares-support vector machine; PCA-SVR, principal component analysis-support vector regression; RGB, red, green, blue; SERS, surface enhanced Raman scattering; SVM, support vector machine; SVR, support vector machine regression; TLC, thin layer chromatography; WT-SCAE, wavelet transform-stack convolution auto-encoder; XAS, X-ray absorption spectrum.

5. Applicability domain in field of toxicity in environmental sciences

As previously mentioned, computational toxicity models should only be used to make predictions within the domain by interpolation [81]. Regarding environmental sciences, sizable bibliographical documentation is comparable to the applications of ML, only recently highlighting the necessity to determine and specify the applicability domain, as to stress the limits in the models developed and provide added value to them.

Yu & Zeng (2022) [82] carried out a study on the classification of pesticide aquatic toxicity affecting fish using a random forest algorithm-based model by means of eight molecular descriptors to develop a QSAR model for 1,106 toxicity datapoints of organic pesticides to various fish species. The optimal RF model was found to have high prediction accuracies for both training and test sets, suggesting it could be useful in predicting the toxicity of pesticides to fish. However, they did not provide specific information about the applicability domain of the models used in this work.

Li et al. (2023) [83] presented a study on the ecotoxicological QSAR modelling of the acute toxicity of fused and non-fused polycyclic aromatic hydrocarbons (FNPAHs) against two aquatic organisms using a genetic algorithm (GA) plus multiple linear regression (MLR) approach to establish QSAR models of the two aquatic toxicity endpoints: *Daphnia magna* (48 h LC50) and *Oncorhynchus mykiss* (96 h LC50). The AD of the QSAR models used in this work was determined through leverage and standardization methods. The leverage method was applied to verify the presence of structural outliers, while the standardization method was used to identify response outliers for compounds having standardized residuals greater than three standard deviation units for cross-validation, showing that all test compounds fell within the AD, thus implying that the prediction for *Daphnia magna* acute toxicity can be reliably interpolated.

Lavado et al. (2021) [84] unveiled new QSAR models to predict acute toxicity affecting shrimp *T. platyurus*, using publicly available data, developed using two different techniques, partial least squares and gradient boosting machine, providing promising results in the identification of some descriptors with an important impact on aquatic toxicity affecting *T. platyurus*. Additionally, they provided a mechanistic interpretation of the results, which may be useful for experts and regulators. Two different ML approaches were used in this work: partial least squares and gradient boosting machine. Two approaches were used to determine the applicability domain of the QSAR models: a standardization approach and a leverage technique, suggesting that carboxin and chlorpropham were outside the AD of the models. In the standardization approach, after standardizing the descriptors in the training set, 99.7% of the samples remained in the range of $\mu \pm 3\sigma$, which was the space where most of the training set compounds belong. Any compound outside this region is dissimilar to the rest of the compounds and was considered outside the AD. The second approach was based on calculation of the leverage (h) for each chemical and defined a threshold that acted as an upper bound limit. Test compounds with leverage values $h > 3p/n$, where p is the number of descriptors and n is the number of molecules, were chemically different from training set compounds.

Sun et al. (2021) [85] developed QSAR models of acute oral toxicity of Polycyclic Aromatic Hydrocarbons (PAHs) to rats using simple 2D descriptors and interspecies toxicity modelling with mice by GA and MLR following the strict validation principles of QSAR modelling recommended by OECD. The most reliable QSAR model comprised eight simple 2D descriptors with a definite physicochemical meaning; moreover, the authors established, validated, and employed interspecies toxicity (iST) models between rat and mouse to fill gaps in the data. Their developed models should be applicable to new PAHs falling within the AD of the models for rapid acute oral toxicity prediction. The applicability domain of the models used in this work was determined using the leverage

approach combined with the standardized residuals of response variable and analyzed the AD using the PCA method.

Banjare, Singh & Roy (2021) [86] developed predictive classification based QSTR models for toxicity studies into diverse pesticides on multiple avian species, with coverage of a large dataset (516) of diverse pesticides found in three avian species. Models were developed using linear discriminant analysis method with genetic algorithm for feature selection from 2D descriptors. The mechanistic interpretation suggested that presence of phosphate, halogens (Cl, Br), ether linkage, and NCOO influence the avian toxicity. Model reliability was verified via the application of the standardization approach of the AD. The developed models provided a priori toxic and non-toxic classification for unknown pesticides (inside AD), with a particular emphasis on organophosphate pesticides. Plus, the interspecies toxicity correlation and predictions encouraged for their further applicability for the fulfilment of data gaps in vital missing species. ML methodologies used were specifically a linear discriminant analysis (LDA) and a genetic algorithm for feature selection from 2D descriptors. The models' applicability domain was determined through the standardization approach, revealing that a significant percentage of compounds were inside the AD for each model. This method calculates the standardized descriptor values for each compound and determines if they fall within a predefined range. If the standardized descriptor values fall outside this range, the compound is considered an outlier or outside the AD. Moreover, the structural analysis of outside AD compounds indicated that mostly outside AD compounds were cyan, thiophosphate, amide, and long chain of hydrocarbon for aquatic avian species, while amide, phosphoramidate, hydrophosphoric acid, and thiourea type of compounds were outside the AD for terrestrial avian species.

Samanipour et al. (2022) [87] connected molecular descriptors to toxicity by means of a QSAR regression model and a direct classification model to predict acute fish toxicity. A random forest QSAR regression model was developed, optimized, validated, and tested using the curated descriptors as independent variables and the experimentally defined LC₅₀ values as dependent variables. Authors compared the model-based AD and the training set AD to avoid extrapolation. The AD assessment was performed by calculating the leverage of each chemical compared to the training set.

Banjare et al. (2023) [88] published a study on the aquatic toxicity prediction of diverse pesticides on two algal species using a QSTR modeling approach, with the aim of identifying the toxic nature of diverse pesticides on the aquatic compartment. The QSTR models were developed by MLRs, and the GA was used for variable selection. Thus, the developed GA-MLR models were found to be statistically robust and reliable. The mechanistic interpretation showed that certain chemical fragments influenced pesticides' toxicity towards the algal species. Additionally, the developed models were applied to pesticides without an experimental value to assess the cumulative toxicity of pesticides on the aquatic environment using a PCA. The leverage approach and the prediction indicator were used to determine the AD.

Hao et al. (2022) [89] developed binary and multi-classification models using ML algorithms to predict the acute oral toxicity of nitroaromatic compounds (NACs) in rats, based on a comprehensive dataset containing 371 NACs with experimental median lethal dose values against rat (LD₅₀) determined through oral exposure. Thus, seven ML algorithms were used to develop the models, including LR, RF, KNN, Naïve Bayes (NB), SVM, NN, and DT, and the ADs of the models were determined using a distance-based similarity approach, defining the distance threshold AD as a function of the mean Euclidean distance of each compound to its nearest neighborhood within the modelling set; therefore, the predicted result was considered unreliable if the mean distance between a query molecule and its three nearest training neighbors was greater than the domain threshold. Developed models had a strong internal robustness and good external prediction power. Structural

alerts closely associated with oral acute toxicity were identified using information gain and substructure frequency analysis. Additionally, the study proved that the toxicity of NACs could be reduced via structural modification, which entails implications for the design of “green and safe” NACs in industrial production, thus facilitating an environmental risk assessment and the design of green and safe chemicals.

Xu et al. (2022) [90] presented a study on the prediction of chemical aquatic toxicity using ML and DL approaches, constructing predictive models for four fish species, including bluegill sunfish, rainbow trout, fathead minnow, and sheepshead minnow, as well as global models with all four fish data. Approximately 1,874 compounds and their labels were collected from the knowledgebase of chemical environmental toxicity data on aquatic and terrestrial species ECOTOX, and literature. Conventional ML methods, DL architecture, and a graph convolutional network (GCN) were used to build predictive models. The classification accuracy of the best local model for each fish species was higher than 0.83. For the global models, two strategies including consistency prediction and probability threshold were adopted to improve the predictive capability but limit the AD. Thus, for 63% of compounds in domain, the accuracy was in the region of 0.97. The single-task GCN method was found to show specific advantages in performance compared to the other methods, whilst multitask GCN showed no advantages over the conventional ML methods. The conventional ML methods used included SVM, RF, KNN, and NN. The applicability domain of the global models was determined by using a probability threshold strategy, thus defining predictions that had probabilities higher than the threshold to be toxic or nontoxic; others were labelled as inconclusive using the coverage rate (CR) to evaluate the applicability of the models.

Tinkov, Grigorev & Grigoreva (2021) [91] presented a QSAR analysis of the acute toxicity of avermectins, antiparasitic agents used in medicine, veterinary medicine, and agriculture, towards *Tetrahymena pyriformis*, with adherence to OECD principles. The models were developed using various molecular descriptors and ML methods, specifically least squares-SVM and transformer CNNs. The applicability domain was determined using the ‘distance to model’ concept, particularly the ‘CONSENSUS-STD’ approach. Additionally, a structural interpretation of the QSAR model was performed, thus revealing significant molecular transformations that increase and decrease the acute toxicity of organic compounds, providing valuable information on the ecotoxicological characteristics of individual avermectins.

Zhu, Chen & Tao (2023) [92] carried out a comprehensive assessment of 39 QSPR models developed to predict aqueous solubility using multiple ML algorithms and descriptor screening methods using the CRITIC-TOPSIS method for the first time in the environmental model field. The XGB model based on SRM was selected as the optimal pathway for predicting aqueous solubility. The applicability domain of the models was determined using Williams plots, employed to set the boundary values to define outliers, thereby indicating chemicals that displayed a vastly different structure in comparison with other chemicals if the leverage value (h) was greater than the warning value (h^*), regarding as an outlier when the absolute standardized residuals (δ) value of data was >3 .

Xu et al. (2021) [93] developed binary classification models for predicting acute contact toxicity on honeybees using ML methods, collecting data on honeybees’ acute contact toxicity from three publicly available databases and using six ML methods combined with nine molecular fingerprints to establish 54 binary classification models that were validated using a 10-fold cross-validation and external validation. ML methods used in this work were the NN, C4.5 DT, KNN, NB, RF and SVM. The best model, which combined the SVM algorithm with the CDK extended fingerprint, obtained an AUC value of 0.924 and a CA value of 0.904 for a test set containing 136 pesticides. The applicability domain of the models was analyzed, excluding certain extreme compounds. Additionally, nine

structural alerts were identified by information gain and substructure frequency analysis, thereby preventing the potential toxicity of test chemicals. A specific applicability domain based on similarity was employed to avoid the prediction for new compounds with substantially different structures from those in the training set. Thus, the distance between a new molecule and its KNNs in the training set were compared based on a threshold of AD in such a manner that if the distance from at least one of KNNs in the training set exceeded the calculated threshold, this would be considered as an outlier.

The most representative studies in which prediction models have been used are the following: SVC and SVR models were applied to the surface water quality data (1995–2010) to optimize the monitoring program. The results showed that the nonlinear models performed better than the corresponding linear methods of classification and regression modeling [94]. Another research developed a global modeling tool capable of categorizing structurally diverse chemicals in several toxicity classes to predict their acute toxicity in fathead minnow using a set of selected molecular descriptors; the results showed good predictive levels and can be used as tools for predicting toxicities [94]. A study of 24 linear and ML models for the prediction of bioconcentration in fish was presented and important factors influencing accumulation were identified [68]. A study was carried out to view the performance of different bentonite lining materials on the removal efficiency of Cu (II) and Zn (II) from industrial leachates and an NN was used to show the significant levels of the coating materials analyzed in the removal efficiency [73]. A study developed an intelligent expert system (based on a feedback neural network trained through an extreme ML algorithm) to predict the content of pharmaceuticals in lettuce tissues irrigated with wastewater treatment plants (WWTP) reclaimed water with the results showing that the intelligent expert system was reliable [74]. A combination of five ML methods along with seven types of fingerprints and a set of molecular descriptors were devised; the results of this study provided critical information and useful tools for chemical estimation in environmental risk assessment. [79]. Cellular automata coupled with a neural network (CA-NN) were established to calculate the atmospheric dispersion of methane (CH_4) [95]. In another study, Gaussian-MLA models were applied to predict the dispersion of polluting gases whose results showed that it is a good model in the issue of identifying emission source parameters [85].

Other studies in the field of toxicity in environmental sciences have been considered, such as those that can be seen in Table 4, which shows a review of the main ML-based methodologies applied to environmental sciences and their AD and detected limitations.

Table 4. Applicability domain in ML methodologies applied to the field of toxicity in environmental sciences.

Environmental field	Study focus	Environmental issues	Algorithm applied	Applicability (explicit/implicit)	Domain	Models' limitations	Literature
Water	Pesticide toxicity	Pesticide aquatic toxicity to fish	RF	Not provided		AD not provided	[82]
Water	Toxicity of fused and non-fused polycyclic aromatic hydrocarbons.	Acute toxicity of fused and non-fused polycyclic aromatic hydrocarbons (FNFPAHs) against aquatic organisms	GA plus MLR	Leverage and standardization methods.		Aquatic toxicity endpoints: <i>Daphnia magna</i> (48 h LC50) and <i>Oncorhynchus mykiss</i> (96 h LC50)	[83]
Water	Acute Aquatic toxicity	Acute Aquatic toxicity towards <i>T. platyurus</i> .	PLS and GB	Standardization approach and leverage technique.		Carboxin and chlorpropham were outside the AD of the models	[84]
Polycyclic Aromatic Hydrocarbons (PAHs)	Acute oral toxicity	Acute oral toxicity of Polycyclic Aromatic Hydrocarbons to rats.	GA and MLR.	Leverage approach combined with the standardized residuals of response variable, and PCA method.		Applicable to new PAHs falling within the applicability domain (AD)	[85]
Pesticides	pesticides on avian species	Presence of phosphate, halogens (Cl, Br), ether linkage, and NCOO influence the avian toxicity.	LDA with GA.	Standardization approach		Mostly outside AD were cyano, thiophosphate, amide, and long chain of hydrocarbon for aquatic avian species, while amide, phosphoramidate, hydrophosphoric acid, and thiourea type of compounds were outside the AD for terrestrial avian species	[86]
Water	Fish toxicity	Intrinsic acute Fish Toxicity of Chemicals	RF	Leverage of each chemical compared to the training set.		LC ₅₀ values	[87]
Water	Toxicity of pesticides on the aquatic compartment.	Aquatic toxicity prediction of diverse pesticides on two algal species	MLRs and GA	Leverage approach and prediction reliability indicator.		Certain chemical fragments that influence the toxicity of pesticides towards the algal species	[88]
Nitroaromatic compounds (NACs).	Acute oral toxicity	Acute oral toxicity of nitroaromatic compound (NACs) in rats	LR, RF, kNN, NB, SVM, NNs and DT.	Distance-based similarity approach. Distance threshold AD as a function of the mean Euclidean distance of each compound to its nearest neighborhood within the modeling set.		Predicted result unreliable if the mean distance between a query molecule and its three nearest training neighbors was greater than the domain threshold.	[89]

Continued on next page

Environmental field	Study focus	Environmental issues	Algorithm applied	Applicability (explicit/implicit)	Domain	Models' limitations	Literature
Water	Aquatic toxicity	Chemical aquatic toxicity to fish species	SVM, RF, kNN, NN and, GCN.	Probability threshold strategy		Four fish species, including bluegill sunfish, rainbow trout, fathead minnow, and sheepshead minnow.	[90]
Water	Acute toxicity of avermectins	Ecotoxicological characteristics of individual avermectins towards <i>Tetrahymena pyriformis</i> .	SVM and CNN.	Distance to model. Consensus approach.		Avermectins	[91]
Water	Ecological risk and toxicity of organic pollutants by means of hydrophobicity of compounds	Transportation of Contaminant molecules in the environment and the absorption capacity of organisms	XGB based on SRM	Williams plots		Aqueous solubility	[92]
Toxicity of pesticides	Acute contact toxicity on honeybees	Ecological risk assessment of pesticides	NN, DT, KNN, NB, RF and SVM.	Similarity distance.		Structurally diverse pesticides, excluding some extreme molecules.	[93]
Water	Water quality	Predicting the biochemical oxygen demand (BOD)	SVC	Dataset of 1500 water samples from 10 different locations monitored for 15 years			[94]
Water	Fathead minnow	Categorizing chemicals into toxicity classes	MLPN, RBFN, PNNs, GRNNs	In this study, a total of 573 compounds were selected.		Linear modeling-based structure-activity relationships of the chemicals may have complex non-linear dependence.	[96]
Water	Gammarus pulex	Prediction of chemical bioconcentration in a freshwater invertebrate	GRNNs	The data used here was a sub-selection with only one <i>C. carpio</i> fish species (n = 352) for modeling purposes.		Limitations of predictive performance may stem from the raw data.	[97]
Water	Fathead minnow	Prediction of the acute toxicity of the compounds to fathead minnow.	SVM, NN.	Most of the selected descriptors take the atomic properties as the weighting.		The model should only be used to make predictions by interpolation not extrapolation.	[98]
Soil	Lead and its compounds	Evaluating the heavy metal biosorption process	NN.	Lead working standard solutions were prepared for use in the experiments by dilution.		The main limitation of RSM assumes only nonlinear quadratic correlation.	[99]

Continued on next page

Environmental field	Study focus	Environmental issues	Algorithm applied	Applicability (explicit/implicit)	Domain	Models' limitations	Literature
Soil	Zn (II) ions	Predicting the percentage of adsorption efficiency for the removal of Zn (II) ions from the leachate by the hazelnut shell	NN	The performance of the proposed NNs modeling technique was compared to full factor experimental design.		The input variables are calculated as the initial pH of 8 and temperature of 40 °C.	[100]
Soil	Heavy metals (Cu, Mn and Ni)	Predicting the accumulation and transformation of heavy metals in the subarctic soil.	GRNNs, MLPN.	The samples were randomly split into training and test datasets.		Deterministic methods are impossible in principle.	[101]
Soil	Cu (II) and Zn (II)	Predicting the removal efficiency of Cu (II) and Zn (II) from bentonite, bentonite mixtures, natural zeolite, expanded vermiculite and pumice.	NN.	The proposed NN uses only the removal of Cu(II) and Zn(II).		Training procedures for NNs require long computer runs.	[102]
Soil	Lettuce crops	Predicting the concentration of carbamazepine and diclofenac in lettuces irrigated with reclaimed water.	ELM.	For this experiment the ELM technique implements a linear Kernel.		Further studies are required to investigate uptake in other types of crops.	[103]
Soil	Pesticides	Predicting the soil adsorption coefficient in pesticides.	GBDT.	Applicability domain was characterized by Euclidean distance.		The prediction accuracy was higher than those obtained in previous studies except for the model by Huuskonen.	[104]
Soil	Polycyclic aromatic hydrocarbons (PAHs)	Predicting the formation of PAHs in sediments of the Caspian Sea	NN.	The leverage points, which are located between the standardized residual values ± 3 , are in the applicability domain.		The GRNN neural network has an error greater than the MLP neural network in the different permutations.	[105]
Soil	Polycyclic aromatic hydrocarbons (PAHs)	Predicting the potential toxicity of PAHs in soils located in south Nigeria.	NN.	The dataset was divided into three for the purposes of training, validation and testing.		This study did not entirely consider all influencing factors responsible for soil toxicity.	[106]
Soil	Polycyclic aromatic hydrocarbons (PAHs)	Predicting temporal bioavailability changes of PAHs in contaminated, compost-modified soils	MLPN, RBF, SVR, M5P, M5R.	Empirical data from an experiment was used to predict temporal changes of PAH bioavailability.		The influence of compost on the bioavailability depended on the soil and compost type.	[107]
Wildlife	Apis mellifera	The presence of chemical contaminants in bees was investigated.	SVM, kNN, NB, DT, RF.	251 organic compounds from previous works were extracted.		nHBD and LogS indicated a negative correlation coefficients with HBT.	[108]

Continued on next page

Environmental field	Study focus	Environmental issues	Algorithm applied	Applicability (explicit/implicit)	Domain	Models' limitations	Literature
Wildlife	Bumblebees	The presence of pesticides associated with the health of bumblebees in the northern United States was evaluated.	MLMM.	A large USA bumblebee dataset was used.		Caution is needed against overinterpreting patterns of correlation between pathogens and bee declines.	[109]
Atmosphere	CH ₄	An atmospheric dispersion model was developed to calculate the atmospheric dispersion of methane in atmosphere.	NN.	A total of 95 simulations were used in the application of the cellular automaton.		Parameters defined by experiments and the Gaussian model are needed.	[95]
Atmosphere	Particulate matter (PM)	Forecasting future air pollution for several days in different areas of Seoul	DL.	Some weather data collected from January 1, 2015 to December 31, 2018.		RNN-based models require a lot of training time and computational resources.	[110]
Atmosphere	Air pollutants	Studying the transport and dispersion of polluting particles into the atmosphere.	AQ20.	A classification algorithm was used to find common patterns in weather data from the National Centers for Environmental Prediction.		A large number of simulations are needed.	[111]
Atmosphere	Atmospheric emissions	This study collected geolocated information from residential wood burning to study the spatial distribution of atmospheric emissions.	CNN.	A keyword dataset consisting of 94 keywords covering a wide range of heating systems was established.		Some limitations were identified for the input data.	[112]
Atmosphere	Dispersion of gases	Predicting the dispersion of pollutant gases	GMLNM.	In this model, 2832 training datasets and 3147 test sets were considered.		This built model may not be valid for all situations.	[113]
Atmosphere	Fine particles (Diameter <2.5 μm, PM _{2.5})	Predicting the spatial-temporal distributions of continuous daily PM _{2.5} concentrations in China	GW-GBM.	PM _{2.5} monitoring data was obtained for 1015 sites from 267 cities.		There were 479 missing values, meaning it was necessary to interpolate the estimated PM _{2.5} values.	[114]

Abbreviations: AQ20 algorithm (AQ20), neural network (NN), C4.5 decision tree (DT), Convolutional neural network (CNN), Deep Learning (DL), Extreme learning machine neural network (ELM), Gaussian-ML network model (GMLNM), Genetic Algorithm (GA), Generalized regression neural networks (GRNNs), Geographically-Weighted Gradient Boosting Machine (GW-GBM), Gradient boosting decision tree (GBDT), K-nearest neighbor (kNN), Linear Discriminant Analysis (LDA), M5 model tree (M5P), M5 rule (M5R), ML multi-variable model (MLMM), Multiple Linear Regressions (MLRs), Multilayer perceptron network (MLPN), Naive Bayes (NB), Partial Least Squares (PLS), Probabilistic neural networks (PNNs), Radial basis function (RBF), Radial basis function network (RBFN), Random forest (RF), Regression neural networks (GRNN), Support vector machines (SVM), Support vector regression (SVR).

6. Applicability domain in industrial Hygiene

Regarding toxicity in the context of industrial hygiene, an intense bibliographic review shows that, as far as current knowledge allows, no study into ML-based methodologies in this field where ADs were specifically and clearly defined has been discovered. While it is true that the use of ML in this sector is relatively novel, this definition and its limits into their approaches will be of interest in future studies, which will provide a certain value with regard to the usability of these models. Coelho et al. [115] used NNs to perform an improved risk assessment of occupational exposure to pesticides, integrating variables as individual susceptibility, physicochemical properties or type of exposure, achieving more than 90% precision in the evaluation of 142 pesticides in this study. Nevertheless, there is no clear definition in place on a certain AD since the model does not consider the biotransformation pathways, routes of exposure, contact time and individual susceptibility of the compounds.

Mansouri, et al. (2021)[116] developed the Collaborative Acute Toxicity Modeling Suite (CATMoS), which is a consensus model developed by an international collaboration of 35 research groups to predict acute oral toxicity, with the aim of predicting the latter based on five different endpoints: Lethal Dose 50 (LD50), U.S. Environmental Protection Agency hazard categories, Globally Harmonized System for Classification and Labeling hazard categories, highly toxic chemicals, and nontoxic chemicals. With this aim, an acute oral toxicity data inventory for 11,992 chemicals was compiled and made available to the participating groups, who submitted a total of 139 predictive models. A variety of modeling approaches were used, including ML techniques such as NNs, SVM, RF, KNN, DT, and NB. The applicability domain of the models used in CATMoS was assessed and combined into consensus predictions based on a weight-of-evidence approach, therefore predictions were only considered for chemicals that fell within the AD of each individual model.

Acosta-Jiménez et al. (2022) [117] devised a QSTR model for predicting the toxicity of carbamate compounds, developed using a set of 178 carbamate derivatives whose toxicities in rats through oral administration had been evaluated, and thoroughly validated the model using either tested or untested compounds falling within the applicability domain of the model. A genetic algorithm was applied over selected reliable molecular descriptors to obtain a model with 10 descriptors. Additionally, several regression approaches such as Ridge, Lasso, Backward-Forward selection, XGBoost and SVR were tested with the score R^2 in a range of 0.67 to 0.88. A Williams plot was utilized to view the model's AD and prove the relationship between standardized residuals and leverage values, revealing the robustness of the AD for the current QSAR models.

Kotzabasaki et al. (2021) [118] develop predictive nanoinformatics models for accurately predicting the genotoxicity of different types of multi-walled carbon nanotubes (MWCNTs), using a combination of unsupervised and supervised ML techniques, including PCA, SVM, RF, LR, and NB, as well as Bayesian optimization. The recursive feature elimination (RFE) method was applied to select the standout variables, showing that an RF model using only three features - "Length", "Zeta average", and "Purity" - was the most efficient for predicting the genotoxicity of MWCNTs, which exhibited an 80% accuracy on an external validation and high classification probabilities. The AD of the models was determined using the Leverage method. The thresholds were estimated for both the LR and RF models, and all test environment MWCNTs were found to be within the DOA of the model.

Wehr et al. (2022) [119] presented the development of a QSAR model called RespiraTox, aiming to predict human respiratory irritants and reduce reliance on animal testing established using molecular physicochemical and structural information following the OECD QSAR principles. The curated project database was comprised of 1,997 organic substances, with 1,553 being classified as irritating and 444 as non-irritating. Several ML methodologies were used in this work, including LR, RFs and

Gradient Boosted Decision Trees (GBTs), to determine the best classification method for predicting human respiratory irritants using the developed QSAR model, showing that the best classification was obtained by GBTs. The applicability domain for the QSAR model used in this work was determined using the Euclidean distance of a test compound to compounds within the training set, calculated using molecular descriptors. The default cutoff value of 0.5 for Z was taken. Alongside the classification and information on the AD, the web-based tool provides a list of structurally similar analogues together with their experimental data to facilitate expert review for read-across purposes. This study displays certain limitations, as the one related to the classification of compounds in this dataset which might depend on data annotation richness. A further limitation is that for non-irritating compounds, the concordance between the different sources is 100% due to the applied worst-case approach.

In turn, Zendehdel et al. [120] used NNs (MATLAB software) to determine the oxidative stress of factory workers exposed to hexavalent chromium with interesting results because their multivariate modeling can be used as an effective prediction of biochemical toxicity in the group of workers included in the study. Remarkably, it is limited to the group of workers studied that could be considered the AD. There is no data regarding female workers or other members of humanity. Therefore, the potential extrapolation of this model would be an approach worth contemplating in the future.

Vis-à-vis the level of exposure, Black et al. [121] predicted general or intermittent benzene exposures in tanker truck drivers using a NN based on either job-specific modules or questionnaires. Nevertheless, the AD was not well defined. Thus, although the study showed an optimistic supplement of the human assessor's opinions, the fact that there was no definition in terms of the range of age, gender and races of these drivers, tends to limit any realistic applicability for this model.

Regarding chemical exposure, Johnston et al. [122] used ML-based methodologies under the Estimation and Assessment of Substance Exposure (EASE) program to predict the exposure to chloroprene and toluene of a group of workers in a polychloroprene manufacturing plant. This approach does not show a clear applicability domain. However, the authors have indicated that the detection limit of the technique was 0.003 ppm in both cases, thereby avoiding the estimations to lower exposure concentrations. Aside from this, EASE features a highly generalized nature and needs to be adapted to the complexity of real exposure situations limits, thus defining clear aspects concerning to actual tasks, time, and process streams composition.

On the contrary, Li et al. [123] defined an applicability domain for their study. In this way, an accurate prediction for respiratory occupational exposure to manganese dioxide using linear regression and NNs through a backpropagation algorithm was made. The authors clearly described six exposure influence factors to define the study's limits: intensity (none-high), frequency (0.1-0.9%), exposure circumstance (indoor-outdoor), distance between exposure source and measure site (0-50 m), time (1978-2007) and sites (different places on the plant). This well-defined applicability domain provided a greater prediction accuracy and reproducibility for the NN predictions, with the unique limitation that individual respiratory exposure was not contemplated. Nevertheless, personal sampling would not be admissible in routine surveillance for evident practical reasons. Additionally, Sottas et al. [124] used NNs to advance the level of occupational exposure to a variety of air pollutants. The authors correctly defined its AD for the following pollutants: isopropanol, dimethyl, ethyl amine, acetonitrile, chromium, tetrachloroethylene, ethyl alcohol, respirable dust, inhalable dust, total dust, benzene, lead, formaldehyde, oil mist, liquid petrolatum and n-heptane. However, not enough information was provided about used emulsions made with detergents and mixed pollutants. To have a complete AD, it would be interesting to define which detergents, pollutants and concentrations of each were used, as well as the fact to consider the biotransformation pathways, routes of exposure and individual susceptibility of the workers in the model.

Moayed and Shell [125,126] facilitated the interpretation of the relationships between the level of exposure and the effect provoked through a comparative study between the results offered by linear regression and NNs using questionnaires as input sources. This study established a wide range of parameters for its AD. They used data from males, with an average height of 179.8 ± 6.49 cm and weight of 87.42 ± 15.97 kg, smokers and non-smokers, any age group, any work experience, and any profession. Nevertheless, there was insufficient data for females and other ethnics groups different to Caucasian, meaning it is not possible to ensure the accurate of these models in these cases.

Insomuch as nanotubes toxicity, Gernand and Casman [127] used ML algorithms such as RF to study the toxicity of carbon nanotubes (CNT) and nanoparticles of various metal oxides (TiO₂, SiO₂, ZnO, MgO), thereby predicting toxicological properties of these new nanomaterials in a reliable way. Through this approach, an AD was implicitly defined, as the study included doses for CNT from 2 to 8,890 $\mu\text{g}/\text{kg}$, TiO₂ from 35 to 90,000 $\mu\text{g}/\text{kg}$, and SiO₂, ZnO and MgO from 1,000 to 5,000 $\mu\text{g}/\text{kg}$. However, lengths and diameters of these nanoparticles and the effect of impurities were not contemplated in the model. These parameters are especially relevant for CNT toxicity and therefore, their inclusion in future studies as well as their applicability domain would be beneficial.

Other authors, namely Concu et al. [128], Luan et. Al. [129], and Kleandrova et. Al. [130,131], achieved a simultaneous prediction of general toxicity profiles of nanoparticles (NP) under diverse experimental conditions using a unified QSAR/QSTR-perturbation model based on NNs, with an accuracy higher than 97% in both the training and validation sets. In this scenario, the AD might cover a wide range of parameters contemplated in the study (different chemical compositions, size, or shape of NP) with highly promising results, though there was no clear characterization on these approaches regarding the exact range of size or shape of NP that could be used to obtain a good prediction of toxicity. In a similar manner, Ramchandran and Gernand [132] used a wide AD of metal oxide NP with different shapes and sizes as an entry of a genetic algorithms to classify the toxicity of various metal oxide NP into different groups, based on their dose-response, thereby predicting patterns that explain their toxic behavior. Thus, they concluded that a simple relationship between the potency and physical-chemical characteristics of NP did not exist. Therefore, more physical and chemical properties of these NP must be studied for more grounded conclusions.

Table 5. Applicability domain in ML methodologies applied to the field of toxicity in industrial hygiene.

Toxic compounds	Origin in Industrial Hygiene	Identification technique	Algorithm applied	Applicability (explicit/implicit)	Domain	Limitations of the models	Literature
Pesticides	Occupational exposure to pesticides	Acute and chronic toxicity by means of oral, dermal, ocular and inhalation through dietary, recreational and/or occupational exposure.	NNs	Not provided, so limited to the pesticides under study.		The model does consider biotransformation pathways, contact time and individual susceptibility.	[115]
11,992 chemicals were taken into account.	Acute oral toxicity, with five different endpoints including Lethal Dose 50 (LD50).	Lethal Dose 50 (LD50), U.S. Environmental Protection Agency hazard categories, Globally Harmonized System for Classification and Labeling hazard categories, very toxic chemicals, and nontoxic chemicals	NN, SVM, RF, kNN, DT, NB.	Consensus predictions based on a weight-of-evidence approach,		Predictions were only considered for chemicals that fell within the AD of each individual model	[116]
Carbamate compounds	Oral toxicity.	Oral toxicity in rats.	Regression approaches such as Ridge, Lasso, Backward-Forward selection, XGBoost and Support Vector regression (SVR)	Williams plot		Limited to one specific class of organic compounds	[117]
Multi-walled carbon nanotubes (MWCNTs)	Genotoxicity	Genotoxicity was selected as the hazard endpoint.	PCA, SVM, RF, LR, NB.	Leverage method		Thresholds were estimated for both the LR and RF models.	[118]
RespiraTox - Development of a QSAR m Human respiratory irritants	Sensory irritation mediated by the interaction with sensory neurons, and direct tissue irritation.	Inhalation studies with acute and repeated exposure	LR, RFs, and GBTs.	Euclidean distance		Classification of compounds in this dataset might depend on data annotation richness and for non-irritating compounds, the concordance between the different sources is 100%, due to the applied worst-case approach.	[119]

Continued on next page

Toxic compounds	Origin in Industrial Hygiene	Identification technique	Algorithm applied	Applicability (explicit/implicit)	Domain	Limitations of the models	Literature
Pesticides	Agriculture		Back propagation algorithm	142 different compounds recorded on the National Pesticide Information Center		The model is generic and does not consider the biotransformation pathways, routes of exposure, contact time and individual susceptibility	[115]
Hexavalent chromium	Chromium compound manufacturing, chrome electroplating, and leather tanning, and welding	Ferricreducing ability of plasma (FRAP), thiol (SH) content and lipid peroxidation of plasma	Feed-forward back propagation algorithm	Males from Tehran, age 35±9.6 years, work history between 1 to 10 years.		There are no studies about females and another human races	[120]
Benzene	Fuel transport drivers	Job-specific modules or questionnaires	Back-propagation algorithm	Tanker truck drivers		There is no definition of age, gender, and races of these drivers	[121]
Toluene and chloroprene	Polychloroprene manufacturing plant	Gas chromatography	EASE (Estimation and Assessment of Substance Exposure)	Air samples with more than 0.003 ppm of Toluene or chloroprene		EASE must be adapted to the complexity of real exposure situations limits defining in more detail parameters concerning to actual tasks, time, and process streams composition.	[122]
Manganese dioxide	Surface mine site, Mn concentrator plant, Mn powder plant, metallurgical plant, and electrolytic manganese dioxide plant	Records of airborne manganese dioxide	Back-propagation algorithm	Six exposure influence factors: intensity (none-high), frequency (0.1-0.9%), exposure circumstance (indoor-outdoor), distance between exposure source and measure site (0-50m), time (1978-2007) and sites (different places on the plant)		Reflects the conditions of the workers' inhalation zone, but not individual respiratory exposure.	[123]

Continued on next page

Toxic compounds	Origin in Industrial Hygiene	Identification technique	Algorithm applied	Applicability (explicit/implicit)	Domain	Limitations of the models	Literature
18 airborne pollutants	Different companies.		Bayesian model	Isopropanol, dimethyl ethyl amine, acetonitrile, chromium, tetrachloroethylene, ethyl alcohol, respirable dust, inhalable dust, total dust, benzene, lead, formaldehyde, oil mist, liquid petrolatum, n-heptane and three emulsions and detergents with mixed pollutants		There is not definition about used emulsions and detergents with mixed pollutants. Bayesian model does not consider the biotransformation pathways, routes of exposure and individual susceptibility of the workers.	[124]
Non-described	Construction workers	Work Compatibility questionnaire	Feedforward backpropagation algorithm	Males, average height 179.8 ± 6.49 cm and weight 87.42 ± 15.97 kg, smokers and non-smokers, any age, any work experience, and any job.		There is not enough data for females and other ethnics groups different to Caucasian	[125,126]
Carbon nanotube (CNT) and metal oxide nanoparticles	Design new nanoparticles	Data from reported quantitative toxicity measures	Random forests	CNT from 2 to 8890 $\mu\text{g}/\text{kg}$, TiO ₂ from 35 to 90000 $\mu\text{g}/\text{kg}$, and SiO ₂ , ZnO and MgO from 1000 to 5000 $\mu\text{g}/\text{kg}$		Lengths, diameters, and impurities of CNT are not contemplated	[127]
Nanoparticles (NP)	A variety of experimental conditions	Compilation from literature of 260 differing NPs and 31 chemical compositions	QSTR/QSAR-perturbation model based on NNs	Solely metal-based to metallic oxide NPs, including silica-based NPs		Size and shapes of NP were not specified	[128–131]
Metal oxide nanoparticles	Design new nanoparticles	Data from different peer-reviewed journal articles	Genetic algorithm	Metal oxide nanoparticles (Titanium, Silica, Iron, Zinc, Cerium, Nickel, Cooper, Aluminium, Magnesium, Cobalt) from 4 to 63000nm size and 11 to 105 m^2/g specific surface		More physical and chemical properties must be studied	[132]

Abbreviations: Back propagation algorithm (BPA), Bayesian model (BM), Estimation and Assessment of Substance Exposure (EASE), Feed-forward back propagation algorithm (FFBPA), Gradient Boosted Decision Trees (GBTs), Genetic algorithm (GA), Logistic Regression (LR), Naïve Bayes (NB), Neural Networks (NNs), Principal Component Analysis (PCA), QSTR/QSAR-perturbation model based on NNs (QSTR-NN), Support Vectors Machine (SVM), Random forest (RF).

7. Conclusions and future work

This paper provides a new sight for investigating the use of QSAR models and modern NNs, along with a multidisciplinary review of the applicability domains of ML for toxicity prediction of chemical compounds. This approach can be used in various sectors, including medicinal chemistry, food science, environmental science, and industrial hygiene, to predict the toxicity of potential new drugs, to determine the limits of detection for harmful substances in food, to predict the toxicity limits of chemicals in the environment, and to establish exposure limits to harmful substances in the workplace, respectively. ML techniques, particularly NNs, are important in the development of predictive models of toxicity, which are capable of learning the relationship between input data (such as chemical descriptors) and output data (such as toxicity values) to predict the toxicity of new chemicals based on their descriptors. The most promising ML algorithms used to predict chemical toxicity properties include NNs, SVM, RF, KNN, and DL methods such as DCNNs and TBN. These techniques have shown great potential in predicting toxicity endpoints for regulatory purposes. Specifically, NNs are one of the most promising ML algorithms used to predict chemical toxicity properties. The most frequently used strategies to extrapolate untested chemical compounds are the extent of extrapolation, the effective prediction domain, the error estimation and residual standard deviation, and the similarity distance [24]

The development of computational models to predict toxicity indicating their range of applicability is of great interest, according to OECD, which has developed five principles for the use of computational techniques in a regulatory context, and these are internationally accepted standards for safety testing and assessment of chemicals. Thus, one of these principles is the need to define the AD, which refers to the range of chemical compounds for which a QSAR model can accurately predict toxicity. Therefore, this is a crucial concept in the development and practical use of computational models for toxicity prediction. By defining the AD, the reliability of the model can be ensured and spurious extrapolations can be prevented. The OECD guidelines are at the forefront of the publication of good practice guidelines for the replacement of animal experimentation by computational methodologies for the assessment of chemical compounds. Taking this into account, these techniques have the potential to provide more reliable results than traditional *in vivo* tests and can help minimize potential damage by predicting the toxicological properties of new chemicals before they are approved for production and commercialization. Here, we highlight the importance of defining the AD of a model through the different methods available. Some of the unsupervised methods include distance-based approaches, consensus-based decision methods, and the statistical analysis of the training set. For example, distance-based approaches calculate the distance between the problem compound and the training set compounds to determine whether the prediction is reliable, whilst consensus-based decision methods use multiple approaches to determine the applicability domain and to reach a consensus. Our statistical analysis of the results concludes that the interpolated prediction results are more reliable than extrapolated ones and can be used to define the requirements for determining an AD. Thus, the importance of model validation in a regulatory context according to the OECD principles is emphasized. Some limitations of this review include its limited scope. Thus, the study does not encompass some areas of potential interest such as the petrochemical industry, with distinct challenges and applications for toxicity prediction. Additionally, there are some extrapolation challenges; similar to the various AI techniques reviewed, the study identifies difficulties in accurately extrapolating toxicity predictions to untested chemical compounds. Moreover, the emphasis on medicinal chemistry and environmental science leaves other scientific domains like food science, and industrial hygiene with comparatively limited exploration of toxicity models. Additionally, it has been

discovered that there is an underutilization of Applicability Domains. Concretely, while stressing the importance of AD, the paper highlights that practical adoption remains restricted in medicinal chemistry and other scientific areas, limiting the predictive power of models. Finally, there are some validation and regulatory issues due to limited model validation and adherence to OECD principles, as the study does not explore in depth how these principles might differ across industries, potentially affecting model application and generalization. These limitations underscore the need for further research to address extrapolation challenges, assess applicability domains across various industries, and enhance the practical implementation of toxicity models while considering industry-specific nuances. In summary, this paper provides valuable insights into the theory and practice of AD in the context of toxicity problems and highlights the need for further research and development in this area.

Use of AI tools declaration

The authors declare they have not used Artificial Intelligence (AI) tools in the creation of this article.

Acknowledgments

This research is funded by Universidad Internacional de La Rioja under IMPESTIA project (2021-2023) and PLeNTaS (PID2019-111430RB-I00).

Conflict of interest

All authors declare no conflicts of interest in this paper.

References

1. National Research Council, Toxicity testing in the 21st century: A vision and a strategy, in *National Academies Press*, 2007, 1–196. Available from: <https://doi.org/10.17226/11970>
2. H. Sun, M. Xia, C. P. Austin, R Huang, Paradigm shift in toxicity testing and modeling, *AAPS J.*, **14** (2012), 473–480. <https://doi.org/10.1208/s12248-012-9358-1>
3. I. Fischer, C. Milton, H. Wallace, Toxicity testing is evolving! *Toxicol. Res. (Camb)*, **9** (2020), 67–80. <https://doi.org/10.1093/toxres/tfaa011>
4. S. Gibb, Toxicity testing in the 21st century: A vision and a strategy, *Reprod. Toxicol.*, **25** (2008), 136–138. <https://doi.org/10.1016/j.reprotox.2007.10.013>
5. K. A. Ford, Refinement, reduction, and replacement of animal toxicity tests by computational methods, *ILAR J.*, **57** (2016), 226–233. <https://doi.org/10.1093/ilar/ilw031>
6. C. Jean-Quartier, F. Jeanquartier, I. Jurisica, A. Holzinger, In silico cancer research towards 3R, *BMC Cancer*, **18** (2018), 408. <https://doi.org/10.1186/s12885-018-4302-0>
7. E. Pérez Santín, R. Rodríguez Solana, M. González García, M. Del Mar García Suárez, G. David Blanco Díaz, M. Dolores Cima Cabal, et al., Toxicity prediction based on artificial intelligence: A multidisciplinary overview, *Wiley Interdiscip. Rev. Comput. Mol. Sci.*, **11** (2021), e1516. <https://doi.org/10.1002/wcms.1516>
8. G. J. Myatt, L. D. Beilke, K. P. Cross, In Silico Tools and their Application, In: *Comprehensive Medicinal Chemistry III*, Oxford, Elsevier, 2017, 156–176. Available from: <https://doi.org/10.1016/B978-0-12-409547-2.12379-0>
9. R. Todeschini, V. Consonni, P. Gramatica, 4.05-Chemometrics in QSAR, In: *Comprehensive Chemometrics*, Oxford, Elsevier, (2009), 129–172. <https://doi.org/10.1016/B978-044452701-1.00007-7>

10. Committee 37th Joint Meeting of the Chemicals, OECD principles for the validation, for regulatory purposes, of (quantitative) structure–activity relationship models, 2019. Available from: <https://www.oecd.org/chemicalsafety/risk-assessment/37849783.pdf>
11. OECD (Organisation for Economic Co-operation and Development, Quantitative Structure-Activity Relationships Project [(Q)SARs], 2023. Available from: <https://www.oecd.org/chemicalsafety/risk-assessment/oecdquantitativestructure-activityrelationshipsprojectqsars.htm>.
12. ECHA (European Chemicals Agency), REACH: Regulation (EC) No 1907/2006. Available from: <https://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=OJ:L:2007:136:0003:0280:en:PDF>
13. European Commission, JRC QSAR Model Database, *Joint Research Centre (JRC)*, 2020. Available from: <https://data.jrc.ec.europa.eu/dataset/e4ef8d13-d743-4524-a6eb-80e18b58cba4>
14. S. C. Peter, J. K. Dhanjal, V. Malik, N. Radhakrishnan, M. Jayakanthan, D. Sundar, Quantitative Structure-Activity Relationship (QSAR): Modeling Approaches to Biological Applications, In *Encyclopedia of Bioinformatics and Computational Biology*, Oxford, Academic Press, 2019, 661–676. <https://doi.org/10.1016/B978-0-12-809633-8.20197-0>
15. K. Roy, S. Kar, R. N. Das RN, QSAR/QSPR Modeling: Introduction, In: *A Primer on QSAR/QSPR Modeling: Fundamental Concepts*, Cham, Springer International Publishing, 2015, 1–36. <https://doi.org/10.1007/978-3-319-17281-1>
16. G. J. Hwang, H. Xie, B. W. Wah, D. Gašević, Vision, challenges, roles and research issues of Artificial Intelligence in Education, *Comput. Education: Artif. Intell.*, **1** (2020), 100001. <https://doi.org/10.1016/j.caeai.2020.100001>
17. S. Agatonovic-Kustrin, R. Beresford, Basic concepts of artificial neural network (ANN) modeling and its application in pharmaceutical research, *J. Pharm. Biomed. Anal.*, **22** (2000), 717–727. [https://doi.org/10.1016/S0731-7085\(99\)00272-1](https://doi.org/10.1016/S0731-7085(99)00272-1)
18. R. Jabbar, R. Jabbar, S. Kamoun, Recent progress in generative adversarial networks applied to inversely designing inorganic materials: A brief review, *Comput. Mater. Sci.*, **213** (2022), 111612. <https://doi.org/10.1016/j.commatsci.2022.111612>
19. G. Gómez-Jiménez, K. Gonzalez-Ponce, D. J. Castillo-Pazos, A. Madariaga-Mazon, J. Barroso-Flores, J. Barroso-Flores, et al., Chapter Four-The OECD Principles for (Q)SAR Models in the Context of Knowledge Discovery in Databases (KDD), In: *Advances in Protein Chemistry and Structural Biology*, Academic Press, 2018, 85–117. <http://dx.doi.org/10.1016/bs.apcsb.2018.04.001>
20. A. Morger, F. Svensson, S. Arvidsson McShane, N. Gauraha, U. Norinder, O. Spjuth, Assessing the calibration in toxicological in vitro models with conformal prediction, *J. Cheminform.*, **13** (2021), 1–14. <https://doi.org/10.1186/s13321-021-00511-5>
21. U. Norinder, Traditional machine and deep learning for predicting toxicity endpoints, *Molecules*, **28** (2023), 217. <https://doi.org/10.3390/molecules28010217>
22. M. Nascimben, L. Rimondini, Molecular toxicity virtual screening applying a quantized computational SNN-Based framework, *Molecules*, **28** (2023), 1342. <https://doi.org/10.3390/molecules28031342>
23. J. Li, D. Luo, T. Wen, Q. Liu, Z. Mo, Representative feature selection of molecular descriptors in QSAR modeling, *J. Mol. Struct.*, **1244** (2021), 131249. <https://doi.org/10.1016/j.molstruc.2021.131249>
24. A. Tropsha, 4.07-Predictive Quantitative Structure–Activity Relationship Modeling, In: *Comprehensive Medicinal Chemistry II*, Oxford, Elsevier, 2007 149–165. <http://dx.doi.org/10.1016/B0-08-045044-X/00248-0>
25. A. M. Davis, 3.15-Quantitative Structure-Activity Relationships, In: *Comprehensive Medicinal Chemistry III*, Oxford, Elsevier, 2017, 379–392.

26. E. Benfenati, J. R. Chrétien, G. Gini, Chapter 6-Validation of the models, In: *Quantitative Structure-Activity Relationships (QSAR) for Pesticide Regulatory Purposes*, Amsterdam, Elsevier, 2007, 185–199. <https://doi.org/10.1016/B978-044452710-3/50008-2>
27. E. Kotsampasakou, G. F. Ecker, Predicting Drug-Induced Cholestasis with the Help of Hepatic Transporters-An in Silico Modeling Approach, *J. Chem. Inf. Model.*, **57** (2017), 608–615. <https://doi.org/10.1021/acs.jcim.6b00518>
28. E. Minerali, D. H. Foil, K. M. Zorn, T. T. Lane, S. Ekins, Comparing machine learning algorithms for predicting Drug-Induced liver injury (DILI), *Mol. Pharm.*, **17** (2020), 2628–2637. <https://doi.org/10.1021/acs.molpharmaceut.0c00326>
29. Collaborations Pharmaceuticals, Inc. <http://tomocomd.com/apps/ptoxra>, Assay Central, 2023. Available from: <https://www.collaborationspharma.com/assay-central>
30. J. R. Mora, Y. Marrero-Ponce, C. R. García-Jacas, A. S. Causado, Ensemble Models Based on QuBiLS-MAS Features and Shallow Learning for the Prediction of Drug-Induced Liver Toxicity: Improving Deep Learning and Traditional Approaches, *Chem. Res. Toxicol.*, **33** (2020), 1855–1873. <https://doi.org/10.1021/acs.chemrestox.0c00030>
31. ToMoCoMD framework, SiliS-PTOXRA 2023. Available from: <http://tomocomd.com/apps/ptoxra>
32. Q. Wu, C. Cai, P. Guo, M. Chen, X. Wu, J. Zhou, et al., In silico Identification and mechanism exploration of hepatotoxic ingredients in traditional Chinese medicine, *Front Pharmacol*, **10** (2019), 1–15. <https://doi.org/10.3389/fphar.2019.00458>
33. F. Hussain, S. Basu, J. J. H. Heng, L. H. Loo, D. Zink, Predicting direct hepatocyte toxicity in humans by combining high-throughput imaging of HepaRG cells and machine learning-based phenotypic profiling, *Arch. Toxicol.*, **94** (2020), 2749–2767. <https://doi.org/10.1007/s00204-020-02778-3>
34. P. Di, Y. Yin, C. Jiang, Y. Cai, W. Li, Y. Tang, et al., Prediction of the skin sensitising potential and potency of compounds via mechanism-based binary and ternary classification models, *Toxicol. Vitro*, **59** (2019), 204–214. <https://doi.org/10.1016/j.tiv.2019.01.004>
35. KNIME Open for innovation, End to End Data Science, 2023. Available from: <https://www.knime.com/>
36. KNIME Open for innovation, Community Extensions, 2023. Available from: <https://www.knime.com/community>
37. NovaMeechanics Ltd, Cheminformatics & Nanoinformatics Excellence, 2023. Available from: <https://novamechanics.com/>
38. K. Ogura, T. Sato, H. Yuki, Y. Cai, W. Li, Y. Tang, Support Vector Machine model for hERG inhibitory activities based on the integrated hERG database using descriptor selection by NSGA-II, *Sci. Rep.*, **9** (2019), 1–7. <https://doi.org/10.1038/s41598-019-47536-3>
39. Construction of Drug Discovery Informatics System by Japan Agency for Medical Research and Development, AMED Cardiotoxicity Database, 2023. Available from: <https://drugdesign.riken.jp/hERGDdb/>.
40. N. Fjodorova, M. Vračko, M. Novič, A. Roncaglioni, E. Benfenati, New public QSAR model for carcinogenicity, *Chem. Cent. J.*, **4** (2010), 1–15. <https://doi.org/10.1186%2F1752-153X-4-S1-S3>
41. K. P. Singh, S. Gupta, P. Rai, Predicting carcinogenicity of diverse chemicals using probabilistic neural network modeling approaches, *Toxicol. Appl. Pharmacol.*, **27** (2013), 465–475. <https://doi.org/10.1016/j.taap.2013.06.029>
42. L. Zhang, H. Ai, W. Chen, Z. Yin, H. Hu, J. Zhu, et al., CarcinoPred-EL: Novel models for predicting the carcinogenicity of chemicals using molecular fingerprints and ensemble learning methods, *Sci. Rep.*, **7** (2017), 1–14. <https://doi.org/10.1038/s41598-017-02365-0>

43. CarcinoPred-EL, Prediction of chemical carcinogenicity using ensemble learning methods, Available from: <http://112.126.70.33/toxicity/CarcinoPred-EL/about.html>
44. D. Guan, K. Fan, I. Spence, S. Matthews, Combining machine learning models of in vitro and in vivo bioassays improves rat carcinogenicity prediction, *Regul. Toxicol. Pharm.*, **94** (2018), 8–15. <https://doi.org/10.1016/j.yrtph.2018.01.008>
45. P. Bloomingdale, D. E. Mager, Machine learning models for the prediction of chemotherapy-induced peripheral neuropathy, *Pharm. Res.*, **36** (2019), 35. <https://doi.org/10.1007/s11095-018-2562-7>
46. Team ProTox-II, ProTox-II-Prediction Of Toxicity Of Chemicals, 2023. Available from: https://tox-new.charite.de/protox_II/
47. D. R. Tonholo, V. G. Maltarollo, T. Kronenberger, I. R. Silva, P. O. Azevedo, R. B. Oliveira, et al., Preclinical toxicity of innovative molecules: In vitro, in vivo and metabolism prediction, *Chem. Biol. Interact.*, **315** (2020), 108896. <https://doi.org/10.1016/j.cbi.2019.108896>
48. P. Banerjee, A. O. Eckert, A. K. Schrey, R. Preissner, ProTox-II: A webserver for the prediction of toxicity of chemicals, *Nucleic. Acids. Res.*, **46** (2018), 257–263. <https://doi.org/10.1093/nar/gky318>
49. F. Cheng, W. Li, Y. Zhou, J. Shen, Z. Wu, G. Liu, et al., AdmetSAR: A comprehensive source and free tool for assessment of chemical ADMET properties, *J. Chem. Inf. Model.*, **52** (2012), 3099–3105. <https://doi.org/10.1021/ci300367a>
50. H. Yang, C. Lou, L. Sun, J. Li, Y. Cai, Z. Wang, et al., AdmetSAR 2.0: Web-service for prediction and optimization of chemical ADMET properties, *Bioinformatics*, **35** (2019), 1067–1069. <https://doi.org/10.1093/bioinformatics/bty707>
51. Y. Gu, C. Lou, Y. Tang, AdmetSAR-A valuable tool for assisting safety evaluation. In *QSAR in Safety Evaluation and Risk Assessment*, Academic Press, (2023), 187–201. <https://doi.org/10.1016/B978-0-443-15339-6.00004-7>
52. H. E. Webel, T. B. Kimber, S. Radetzki, M. Neuenschwander, M. Nazaré, A. Volkamer, Revealing cytotoxic substructures in molecules using deep learning, *J. Comput. Aided. Mol. Des.*, **34** (2020), 731–746. <https://doi.org/10.1007/s10822-020-00310-4>
53. D. Antanasijević, J. Antanasijević, N. Trišović, G. Ušćumlić, V. Pocajt, From classification to regression multitasking QSAR modeling using a novel modular neural network: Simultaneous prediction of anticonvulsant activity and neurotoxicity of succinimides, *Mol. Pharm.*, **14** (2017), 4476–4484. <https://doi.org/10.1021/acs.molpharmaceut.7b00582>
54. K. Roy, S. Kar, P. Ambure, On a simple approach for determining applicability domain of QSAR models, *Chemometr. Intell. Lab. Syst.*, **145** (2015), 22–29. <https://doi.org/10.1016/j.chemolab.2015.04.013>
55. S. Zheng, J. Xiong, Y. Wang, G. Liang, Y. Xu, F. Lin, Quantitative prediction of hemolytic toxicity for small molecules and their potential hemolytic fragments by machine learning and recursive fragmentation methods, *J. Chem. Inf. Model.*, **60** (2020), 3231–3245. <https://doi.org/10.1021/acs.jcim.0c00102>
56. S. Zheng, Y. Wang, W. Liu, W. Chang, G. Liang, Y. Xu, et al., In Silico prediction of hemolytic toxicity on the human erythrocytes for small molecules by machine-learning and genetic algorithm, *J. Med. Chem.*, **63** (2020), 6499–6512. <https://doi.org/10.1021/acs.jmedchem.9b00853>
57. F. Plisson, O. Ramírez-Sánchez, C. Martínez-Hernández, Machine learning-guided discovery and design of non-hemolytic peptides, *Sci. Rep.*, **10** (2020), 1–19. <https://doi.org/10.1038/s41598-020-73644-6>
58. H. Feng, L. Zhang, S. Li, L. Liu, T. Yang, P. Yang, et al., Predicting the reproductive toxicity of chemicals using ensemble learning methods and molecular fingerprints, *Toxicol. Lett.*, **340** (2021), 4–14. <https://doi.org/10.1016/j.toxlet.2021.01.002>

59. P. Zhao, Y. Peng, X. Xu, Z. Wang, Z. Wu, W. Li, et al., In silico prediction of mitochondrial toxicity of chemicals using machine learning methods, *J. Appl. Toxicol.*, **41** (2021), 1518–1526. <https://doi.org/10.1002/jat.4141>
60. Y. Yuan, S. Chang, Z. Zhang, Z. Li, S. Li, P. Xie, et al., A novel strategy for prediction of human plasma protein binding using machine learning techniques, *Chemometr. Intell. Lab.*, **199** (2020), 103962. <https://doi.org/10.1016/j.chemolab.2020.103962>
61. W. C. Chou, Z. Lin, Machine learning and artificial intelligence in physiologically based pharmacokinetic modeling, *Toxicol. Sci.*, **191** (2023), 1–14. <https://doi.org/10.1093/toxsci/kfac101>
62. C. Jiang, P. Zhao, W. Li, Y. Tang, G. Liu, In silico prediction of chemical neurotoxicity using machine learning, *Toxicol. Res. (Camb)*, **9** (2020), 164–172. <https://doi.org/10.1093%2Ftoxres%2Ftaa016>
63. X. Cui, J. Liu, J. Zhang, Q. Wu, X. Li, In silico prediction of drug-induced rhabdomyolysis with machine-learning models and structural alerts, *J. Appl. Toxicol.*, **39** (2019), 1224–1232. <https://doi.org/10.1002/jat.3808>
64. I. Sushko, S. Novotarskyi, R. Körner, A. Pandey, M. Rupp, W. Teetz, et al., Online chemical modeling environment (OCHEM): web platform for data storage, model development and publishing of chemical information, *J. Comput. Aided. Mol. Des.*, **25** (2011), 533–554. <https://doi.org/10.1007/s10822-011-9440-2>
65. L. M. Lagares, N. Minovski, M. Novič, Multiclass classifier for P-glycoprotein substrates, inhibitors, and non-active compounds, *Molecules*, **24** (2019), 24102006. <https://doi.org/10.3390%2Fmolecules24102006>
66. FAO The State of Food Insecurity in the World 2001, Rome, 2002. Available from: <http://www.fao.org/3/y1500e/y1500e.pdf>
67. P. A. Luning, F. Devlieghere, Safety in the agri-food chain, Wageningen Academic Pub, 2006. Available from: <https://doi.org/10.3920/978-90-76998-77-0>
68. Z. Han, J. Gao, Pixel-level aflatoxin detecting based on deep learning and hyperspectral imaging, *Comput. Electron. Agric.*, **164** (2019), 104888. <https://doi.org/10.1016/j.compag.2019.104888>
69. F. R. Bertani, L. Businaro, L. Gambacorta, A. Mencattini, D. Brenda, D. Di Giuseppe, et al., Optical detection of aflatoxins B in grained almonds using fluorescence spectroscopy and machine learning algorithms, *Food Control*, **112** (2020), 107073. <https://doi.org/10.1016/j.foodcont.2019.107073>
70. P. Gutiérrez, S. E. Godoy, S. Torres, P. Oyarzún, I. Sanhueza, V. Díaz-García, et al., Improved antibiotic detection in raw milk using machine learning tools over the absorption spectra of a problem-specific nanobiosensor, *Sensors*, **16** (2020), 4552. <https://doi.org/10.3390/s20164552>
71. S. Qiu, J. Wang, The prediction of food additives in the fruit juice based on electronic nose with chemometrics, *Food Chem.*, **230** (2017), 208–214. <https://doi.org/10.1016/j.foodchem.2017.03.011>
72. F. Han, X. Huang, E. Teye, Novel prediction of heavy metal residues in fish using a low-cost optical electronic tongue system based on colorimetric sensors array, *J. Food Process. Eng.*, **42** (2019), 12983. <https://doi.org/10.1111/jfpe.12983>
73. A. Tan, Y. Zhao, K. Sivashanmugan, K. Squire, A. X. Wang, Quantitative TLC-SERS detection of histamine in seafood with support vector machine analysis, *Food Control*, **103** (2019), 111–118. <https://doi.org/10.1016%2Fj.foodcont.2019.03.032>
74. H. Isleroglu, S. Beyhan, Prediction of baking quality using machine learning based intelligent models, *Heat Mass Transfer*, **56** (2020), 2045–2055. <https://doi.org/10.1007/s00231-020-02837-6>
75. H. Lu, H. Zheng, Fractal colour: A new approach for evaluation of acrylamide contents in biscuits, *Food Chem.*, **134** (2012), 2521–2525. <https://doi.org/10.1016/j.foodchem.2012.04.085>

76. A. Yadav, N. Sengar, A. Issac, M. K. Dutta, Image processing based acrylamide detection from fried potato chip images using continuous wavelet transform, *Comput. Electron. Agric.*, **145** (2018), 349–362. <https://doi.org/10.1016/j.compag.2018.01.012>
77. B. Jiang, J. He, S. Yang, H. Fu, T. Li, H. Song, et al., Fusion of machine vision technology and AlexNet-CNNs deep learning network for the detection of postharvest apple pesticide residues, *Artif. Intell. Agricul.*, **1** (2019), 1–18. <https://doi.org/10.1016/j.aiaa.2019.02.001>
78. X. Zhou, J. Sun, Y. Tian, B. Lu, Y. Hang, Q. Chen, et al., Hyperspectral technique combined with deep learning algorithm for detection of compound heavy metals in lettuce, *Food Chem.*, **321** (2020), 126503. <https://doi.org/10.1016/j.foodchem.2020.126503>
79. W. Hu, S. Chen, Y. Li, Q. Wang, Z. Fang, X-ray absorption spectrum combined with deep neural network for on-line detection of beverage preservatives, *Rev. Sci. Instrum.*, **89** (2018), 103108. <https://doi.org/10.1063/1.5048281>
80. X. Sun, K. Zhu, J. Liu, J. Hu, X. Jiang, Y. Liu, Terahertz spectroscopy determination of benzoic acid additive in wheat flour by machine learning, *J. Infrared Millim. Terahertz Waves*, **40** (2019), 466–475. <https://doi.org/10.1007/s10762-019-00579-z>
81. N. Nikolova-Jeliazkova, J. Jaworska, An approach to determining applicability domains for QSAR group contribution models: An Analysis of SRC KOWWIN, *Alt-Altern. Lab. Anim.*, **33** (2005), 461–470. <https://doi.org/10.1177/026119290503300510>
82. X. Yu, Q. Zeng, Random forest algorithm-based classification model of pesticide aquatic toxicity to fishes, *Aquatic Toxicol.*, **251** (2022), 106265. <https://doi.org/10.1016/j.aquatox.2022.106265>
83. F. Li, G. Sun, T. Fan, N. Zhang, L. Zhao, R. Zhong, et al., Ecotoxicological QSAR modelling of the acute toxicity of fused and non-fused polycyclic aromatic hydrocarbons (FNFPAHs) against two aquatic organisms: Consensus modelling and comparison with ECOSAR, *Aquatic Toxicol.*, **255** (2023), 106393. <https://doi.org/10.1016/j.aquatox.2022.106393>
84. G. J. Lavado, D. Baderna, D. Gadaleta, M. Ulte, K. Roy, E. Benfenati, Ecotoxicological QSAR modeling of the acute toxicity of organic compounds to the freshwater crustacean *Thamnocephalus platyurus*, *Chemosphere*, **280** (2021), 130652. <https://doi.org/10.1016/j.chemosphere.2021.130652>
85. G. Sun, Y. Zhang, L. Pei, Y. Lou, Y. Mu, J. Yun, et al., Chemometric QSAR modeling of acute oral toxicity of Polycyclic Aromatic Hydrocarbons (PAHs) to rat using simple 2D descriptors and interspecies toxicity modeling with mouse, *Ecotoxicol. Environ. Safe*, **222** (2021), 112525. <https://doi.org/10.1016/j.ecoenv.2021.112525>
86. P. Banjare, J. Singh, P. P. Roy, Predictive classification-based QSTR models for toxicity study of diverse pesticides on multiple avian species, *Environ. Sci. Pollut. Res.*, **28** (2021), 17992–18003. <https://doi.org/10.1007/s11356-020-11713-z>
87. S. Samanipour, J. W. O'Brien, M. J. Reid, K. V. Thomas, A. Praetorius, From Molecular Descriptors to Intrinsic Fish Toxicity of Chemicals: An Alternative Approach to Chemical Prioritization, *Environ. Sci. Technol.*, (2022), 1–9. <https://doi.org/10.1021/acs.est.2c07353>
88. P. Banjare, J. Singh, E. Papa, P. P. Roy, Aquatic toxicity prediction of diverse pesticides on two algal species using QSTR modeling approach, *Environ. Sci. Pollut. Res.*, **30** (2023), 10599–10612. <https://doi.org/10.1007/s11356-022-22635-3>
89. Y. Hao, T. Fan, G. Sun, F. Li, N. Zhang, L. Zhao, et al., Environmental toxicity risk evaluation of nitroaromatic compounds: Machine learning driven binary/multiple classification and design of safe alternatives, *Food Chem. Toxicol.*, **170** (2022), 113461. <https://doi.org/10.1016/j.fct.2022.113461>
90. M. Xu, H. Yang, G. Liu, W. Li, In silico prediction of chemical aquatic toxicity by multiple machine learning and deep learning approaches, *J. Appl. Toxicol.*, **42** (2022), 1766–1776. <https://doi.org/10.1002/jat.4354>

91. O. V. Tinkov, V. Y. Grigorev, L. D. Grigoreva, QSAR analysis of the acute toxicity of avermectins towards *Tetrahymena pyriformis*, *SAR QSAR Environ. Res.*, **32** (2021), 541–571. <https://doi.org/10.1080/1062936x.2021.1932583>
92. T. Zhu, Y. Chen, C. Tao, Multiple machine learning algorithms assisted QSPR models for aqueous solubility: Comprehensive assessment with CRITIC-TOPSIS, *Sci. Total Environ.*, **857** (2023), 159448. <https://doi.org/10.1016/j.scitotenv.2022.159448>
93. X. Xu, P. Zhao, Z. Wang, X. Zhang, Z. Wu, W. Li, et al., In silico prediction of chemical acute contact toxicity on honey bees via machine learning methods, *Toxicol. Vitro.*, **72** (2021), 105089. <https://doi.org/10.1016/j.tiv.2021.105089>
94. K. P. Singh, N. Basant, S. Gupta, Support vector machines in water quality management, *Anal. Chim. Acta.*, **703** (2011), 152–162. <https://doi.org/10.1016/j.aca.2011.07.027>
95. P. Lauret, F. Heymes, L. Aprin, A. Johannet, Atmospheric dispersion modeling using Artificial Neural Network based cellular automata, *Environ. Modell. Software*, **85** (2016), 56–69. <https://doi.org/10.1016/j.envsoft.2016.08.001>
96. K. P. Singh, S. Gupta, P. Rai, Predicting acute aquatic toxicity of structurally diverse chemicals in fish using artificial intelligence approaches, *Ecotoxicol. Environ. Safe.*, **95** (2013), 221–233. <https://doi.org/10.1016/j.ecoenv.2013.05.017>
97. T. H. Miller, M. D. Gallidabino, J. I. MacRae, S. F. Owen, N. R. Bury, L. P. Barron, Prediction of bioconcentration factors in fish and invertebrates using machine learning, *Sci. Total. Environ.*, **648** (2019), 80–89. <https://doi.org/10.1016/j.scitotenv.2018.08.122>
98. N. X. Tan, P. Li, H. B. Rao, Z. R. Li, X. Y. Li, Prediction of the acute toxicity of chemical compounds to the fathead minnow by machine learning approaches, *Chemom. Intell. Lab. Syst.*, **100** (2010), 66–73. <https://doi.org/10.1016/j.chemolab.2009.11.002>
99. D. Bingöl, M. Hecan, S. Eleveli, E. Kılıç, Comparison of the results of response surface methodology and artificial neural network for the biosorption of lead using black cumin, *Bioresour. Technol.*, **112** (2012), 111–115. <https://doi.org/10.1016/j.biortech.2012.02.084>
100. N. G. Turan, B. Mesci, O. Ozgonenel, Artificial neural network (ANN) approach for modeling Zn(II) adsorption from leachate using a new biosorbent, *Chem. Eng. J.*, **173** (2011), 98–105. <https://doi.org/10.1016/j.cej.2011.07.042>
101. A. P. Sergeev, A. G. Buevich, E. M. Baglaeva, A. V. Shichkin, Combining spatial autocorrelation with machine learning increases prediction accuracy of soil heavy metals, *Catena*, **174** (2019), 425–435. <https://doi.org/10.1016/j.catena.2018.11.037>
102. N. G. Turan, E. B. Gümüşel, O. Ozgonenel, Prediction of heavy metal removal by different liner materials from landfill leachate: Modeling of experimental results using artificial intelligence technique, *Sci. World J.*, **2013** (2013), 240158. <https://doi.org/10.1155/2013/240158>
103. M. González García, C. Fernández-López, A. Bueno-Crespo, R. Martínez-España, Extreme learning machine-based prediction of uptake of pharmaceuticals in reclaimed water-irrigated lettuces in the Region of Murcia, Spain, *Biosyst. Eng.*, **177** (2019), 78–89. <https://doi.org/10.1016/j.biosystemseng.2018.09.006>
104. Y. Kobayashi, T. Uchida, K. Yoshida, Prediction of Soil Adsorption Coefficient in Pesticides Using Physicochemical Properties and Molecular Descriptors by Machine Learning Models, *Environ. Toxicol. Chem.*, **39** (2020), 1451–1459. <https://doi.org/10.1002/etc.4724>
105. J. Sayyad Amin, H. Rajabi Kuyakhi, A. Bahadori, Prediction of formation of polycyclic aromatic hydrocarbon (PAHs) on sediment of Caspian Sea using artificial neural networks, *Petrol. Sci. Technol.*, **37** (2019), 1987–2000. <https://doi.org/10.1080/10916466.2018.1496111>

106. R. Olawoyin, Application of backpropagation artificial neural network prediction model for the PAH bioremediation of polluted soil, *Chemosphere*, **161** (2016), 145–150. <https://doi.org/10.1016/j.chemosphere.2016.07.003>
107. G. Wu, C. Kechavarzi, C. Li, S. Wu, S. J. Pollard, H. Sui, et al., Machine learning models for predicting PAHs bioavailability in compost amended soils, *Chem. Eng. J.*, **223** (2013), 747–754. <https://doi.org/10.1016/j.cej.2013.02.122>
108. X. Li, Y. Zhang, H. Chen, H. Li, Y. Zhao, Insights into the Molecular Basis of the Acute Contact Toxicity of Diverse Organic Chemicals in the Honey Bee, *J. Chem. Inf. Model.*, **57** (2017), 2948–2957. <https://doi.org/10.1021/acs.jcim.7b00476>
109. S. H. McArt, C. Urbanowicz, S. McCoshum, R. E. Irwin, L. S. Adler, Landscape predictors of pathogen prevalence and range contractions in US bumblebees, *Proc. R. Soc. B: Biol. Sci.*, **284** (2017), 2017181. <https://doi.org/10.1098/rspb.2017.2181>
110. G. Yang, H. M. Lee, G. Lee, A hybrid deep learning model to forecast particulate matter concentration levels in Seoul, South Korea, *Atmosphere*, **11** (2020), 348. <https://doi.org/10.3390/atmos11040348>
111. G. Cervone, P. Franzese, Y. Ezber, Z. Boybeyi, Risk assessment of atmospheric emissions using machine learning, *Nat. Hazard. Earth. Syst. Sci.*, **8** (2008), 991–1000. <https://doi.org/10.5194/nhess-8-991-2008>
112. S. Lopez-Aparicio, H. Grythe, M. Vogt, M. Pierce, I. Vallejo, Webcrawling and machine learning as a new approach for the spatial distribution of atmospheric emissions, *PLoS One*, **13** (2018), 0200650. <https://doi.org/10.1371/journal.pone.0200650>
113. D. Ma, Z. Zhang, Contaminant dispersion prediction and source estimation with integrated Gaussian-machine learning network model for point source emission in atmosphere, *J. Hazard. Mater.*, **311** (2016), 237–245. <https://doi.org/10.1016/j.jhazmat.2016.03.022>
114. Y. Zhan, Y. Luo, X. Deng, H. Chen, M. L. Grieneisen, X. Shen, et al., Spatiotemporal prediction of continuous daily PM_{2.5} concentrations across China using a spatially explicit machine learning algorithm, *Atmos. Environ.*, **155** (2017), 129–139. <https://doi.org/10.1016/j.atmosenv.2017.02.023>
115. C. Coelho, M. R. Martins, N. Lima, H. Vicente, J. Neves, An assessment to toxicological risk of pesticide exposure, In: *Communications in Computer and Information Science*, 2016, 139–150. https://doi.org/10.1007/978-3-319-44672-1_12
116. K. Mansouri, A. L. Karmaus, J. Fitzpatrick, G. Patlewicz, P. Pradeep, D. Alberga, et al., CATMoS: Collaborative acute toxicity modeling suite, *Environ Health Perspect*, **129** (2021), 47013. <https://doi.org/10.1289/EHP8495>
117. E. H. Acosta-Jiménez, L. A. Zárate-Hernández, R. L. Camacho-Mendoza, S. González-Montiel, J. Alvarado-Rodríguez, C. Z. Gómez-Castro, et al. QSTR Modeling to Find Relevant DFT Descriptors Related to the Toxicity of Carbamates, *Molecules*, **27** (2022), 5530. <https://doi.org/10.3390/molecules27175530>
118. M. Kotzabasaki, I. Sotiropoulos, C. Charitidis, H. Sarimveis, Machine learning methods for multi-walled carbon nanotubes (MWCNT) genotoxicity prediction, *Nanoscale. Adv.*, **3** (2021), 3167–3176. <http://dx.doi.org/10.1039/D0NA00600A>
119. M. M. Wehr, S. S. Sarang, M. Rooseboom, P. J. Boogaard, A. Karwath, S. E. Escher, RespiraTox—Development of a QSAR model to predict human respiratory irritants, *Regul. Toxicol. Pharm.*, **128** (2022), 105089. <https://doi.org/10.1016/j.yrtph.2021.105089>
120. R. Zendehdel, S. V. Shetab-Boushehri, M. R. Azari, V. Hosseini, H. Mohammadi, Chemometrics models for assessment of oxidative stress risk in chrome-electroplating workers, *Drug Chem. Toxicol.*, **38** (2015), 174–179. <https://doi.org/10.3109/01480545.2014.922096>

121. J. Black, G. Benke, K. Smith, L. Fritschi, Artificial neural networks and job-specific modules to assess occupational exposure, *Ann. Occup. Hyg.*, **48** (2004), 595–600. <https://doi.org/10.1093/annhyg/meh064>
122. K. L. Johnston, M. L. Phillips, N. A. Esmen, T. A. Hall, Evaluation of an artificial intelligence program for estimating occupational exposures, *Ann. Occup. Hyg.*, **49** (2005), 147–153. <https://doi.org/10.1093/annhyg/meh072>
123. Y. N. Li, F. T. Luo, Y. M. Jiang, Y. R. Lu, J. L. Huang, Z. B. Zhang, A prediction model of occupational manganese exposure based on artificial neural network, *Toxicol. Mech. Method.*, **19** (2009), 337–345. <https://doi.org/10.1080/15376510902918392>
124. P. E. Sottas, J. Lavoué, R. Bruzzi, D. Vernez, N. Charrière, P. O. Droz, An empirical hierarchical Bayesian unification of occupational exposure assessment methods, *Stat. Med.*, **28** (2009), 75–93. <https://doi.org/10.1002/sim.3411>
125. F. A. Moayed, R. L. Shell, Developing the function of ‘magnitude-of-effect’ (MoE) for artificial neural networks to demonstrate the causal effect of exposure variables on outcome variable, *Ann. Occup. Hyg.*, **55** (2011), 143–151. <https://doi.org/10.1093/annhyg/meq080>
126. F. A. Moayed, R. L. Shell, Application of artificial neural network models in occupational safety and health utilizing ordinal variables, *Ann. Occup. Hyg.*, **55** (2011), 132–142. <https://doi.org/10.1093/annhyg/meq079>
127. J. M. Gernand, E. A. Casman, Nanoparticle characteristic interaction effects on pulmonary toxicity: A random forest modeling framework to compare risks of nanomaterial variants, *ASCE-ASME J. Risk Uncertain Eng. Syst. B: Mech. Eng.*, **2** (2016), 021002. <https://doi.org/10.1115/1.4031216>
128. R. Concu, V. V. Kleandrova, A. Speck-Planche, M. N. D. S. Cordeiro, Probing the toxicity of nanoparticles: A unified in silico machine learning model based on perturbation theory, *Nanotoxicology*, **11** (2017), 891–906. <https://doi.org/10.1080/17435390.2017.1379567>
129. F. Luan, V. V. Kleandrova, H. González-Díaz, J. M. Ruso, A. Melo, A. Sperck-Planceh, et al., Computer-aided nanotoxicology: Assessing cytotoxicity of nanoparticles under diverse experimental conditions by using a novel QSTR-perturbation approach, *Nanoscale*, **6** (2014), 10623–10630. <http://dx.doi.org/10.1039/c4nr01285b>
130. V. V. Kleandrova, F. Luan, H. González-Díaz, J. M. Ruso, A. Speck-Planche, M. N. D. Cordeiro, Computational tool for risk assessment of nanomaterials: Novel QSTR-perturbation model for simultaneous prediction of ecotoxicity and cytotoxicity of uncoated and coated nanoparticles under multiple experimental conditions, *Environ. Sci. Technol.*, **48** (2014), 14686–14694. <https://doi.org/10.1021/es503861x>
131. V. V. Kleandrova, F. Luan, H. González-Díaz, J. M. Ruso, A. Speck-Planche, M. N. D. Cordeiro, Computational ecotoxicology: Simultaneous prediction of ecotoxic effects of nanoparticles under different experimental conditions, *Environ. Int.*, **73** (2014), 288–294. <https://doi.org/10.1016/j.envint.2014.08.009>
132. V. Ramchandran, J. M. Gernand, Examining the in vivo pulmonary toxicity of engineered metal oxide nanomaterials using a genetic algorithm-based dose-response-recovery clustering model, *Comput. Toxicol.*, **13** (2020), 100113. <https://doi.org/10.1016/j.comtox.2019.100113>



AIMS Press

© 2023 the Author(s), licensee AIMS Press. This is an open access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/4.0>)