# 基于人工智能和互联网时代的化合物毒性预测

徐 涛[1] 秦 越[1*] 单 伟[1] 贾辰阳[2] 朱加良[1] 邓志光[1] 李卓玥[1] 刘丹会[1]

([1]中国核动力研究设计院核反应堆系统设计技术重点实验室 成都 610213;[2]华中师范大学化学学院 农药与化学生物学教育部重点实验室 武汉 430079)

**摘 要** 随着商品中所含各种化合物的不断使用,人们日益关注其对人类及生态环境的安全危害。在过去的几年里,通过计算方法预测化合物毒性已经显示出极大的潜力。在此,总结了常用的机器学习和深度学习算法在建立毒性预测模型上的优缺点,并系统回顾了近三年发表的可免费访问的毒性预测网络服务器。此外,还讨论了基于人工智能和互联网时代下毒性预测所面临的机遇和挑战。希望指导人们合理选择算法和网络服务器进行建模及化合物毒性评估。

**关键词** 人工智能 深度学习 机器学习 毒性预测 网络服务器

## Toxicity Prediction Based on Artificial Intelligence and the Internet Era

Xu Tao[1], Qin Yue[1*], ShanWei[1], Jia Chenyang[2], Zhu Jialiang[1], Deng Zhiguang[1], Li Zhuoyue[1], Liu Danhui[1]

([1] Science and Technology on Reactor System Design Technology Laboratory, Nuclear Power Institute of China, Chengdu, 610213; [2] Key Laboratory of Pesticide & Chemical Biology, Ministry of Education, College of Chemistry, Central China Normal University, Wuhan, 430079)

**Abstract** With the continuous using compounds contained in commodities, there is growing concern about their harm to human and ecological environment safety. In the past few years, computational techniques have showed their potential to predict toxicity of compounds. Here, we summarize the advantages and drawbacks of machine learning and deep learning algorithms for establishing toxicity prediction models, and systematically review the freely accessible toxicity prediction web servers for *in silico* toxicity prediction in the past three years. Additionally, the opportunities and challenges of toxicity prediction based on artificial intelligence and internet are discussed. It is hoped that this paper can provide help in guiding people to rationally choose algorithms and web servers for modeling and toxicity evaluation.

**Keywords** Artificial intelligence, Deep learning, Machine learning, Toxicity prediction, Web server

"One Health"理念的提出,让人们越加意识到人类、动物和环境是相互关联的一个整体,在努力确保人类的健康和继续生存的同时,必须考虑到其他生态物种和环境间复杂的相互联系与依存[1~3]。随着医药、农药、食品添加剂、化妆品等商品的不断使用,人们越来越关注其中所含的化合物对人类和生态物种潜在的安全危害。2006 年,欧盟制定了化学品的注册、评估、授权和限制条例(REACH),以保护人类和生态环境健康[4]。实际上,我们生活的环境中充满了有毒化合物,其中一些毒性相对较低或者只有在长期摄入后才会产生毒性效应,而另外一些毒性相对较大,短时的接触或摄入就会对生物健康造成威胁甚至死亡。在巨大的化学空间内,绝大多数化合物的毒性尚未得到研究,即使是已被登记销售的化合物,仍然缺乏对其潜在毒性的全面评估[5]。例如,每年因药物不良反应会造成 7.5 万至 13.7 万人死亡,产生 1770 件严重的卫生事件,被认为是美国第四大致死原因;因膳食摄入造成的高胆固醇、高血糖等疾病占全球慢性疾病死亡的 23%[6~8]。因此,提前对化合物的潜在毒性进行评估,对保护人类和其他生态物种的健康有十分重要的意义。

* 联系人,**秦越** E-mail:qyqingyue@ formail. com

2021-06-08 收稿,2022-08-09 修回,2022-09-01 接受

传统的体内和体外毒理学试验可以研究化合物对生物体的毒性反应、毒性作用机制,以及帮助人们制定安全参考剂量等,但是,随着待测化合物数量的不断增长、高昂的测试费用、耗时的实验周期和出于动物保护等原因,仅依赖于传统的实验测试方法已不能完全满足人们对化合物毒性风险的评估需求。随着计算机科学的发展以及跨学科、跨专业的合作,通过计算机资源进行化合物的毒性预测成为一种有效的替代方法[9,10]。早在20世纪80年代,科学家便建立定量构效关系(QSAR)模型来预测化合物的毒性,即根据已知化合物的结构或理化性质与其生理活性之间建立定量关系,从而预测具有相似结构的化合物的毒性,因为这些化合物最有可能通过相同的作用机制表现出毒理学效应[11~14]。传统的QSAR模型基本上是基于统计学方法建立的,如线性回归分析、多元分析、早期浅层神经网络模型。实验数据的非线性可能会影响模型的准确性,此外,这些方法难以提取更抽象的特征,并一直被认为有高噪声、过拟合的特点,从而无法进行高精度的预测[15]。如今的大数据时代,人工智能算法以其优异的性能不仅广泛应用于自然语言处理、图像识别、语音识别、汽车自动驾驶,同时应用到了计算化学、生物信息学等多个学科领域,在化合物安全风险评估方面也发挥着自身的优势[16]。近年来,已发展了一系列基于人工智能算法的毒性预测模型来有效地评估化合物的毒性,即使用复杂的算法让计算机能够从数据中学习并做出预测,例如:支持向量机(Support vector machines,SVM)、贝叶斯分类器(Bayesian classifiers)、决策树(Decision trees,DTs)、k近邻(k-nearest neighbors,kNN)、随机森林(Random forest,RF)、人工神经网络(Artificial neural networks,ANN)等[17~20]。由于人工智能算法快速、经济、准确、能够处理大量数据及复杂问题等优点,越来越多的研究者使用机器学习或深度学习方法来优化传统的QSAR模型并进行化合物毒性的预测[21],常见的毒性预测端点包括:急性毒性、肝脏毒性、心脏毒性、细胞毒性、基因毒性、致癌性、诱变性等,另外,也包括对一些常见生态物种的急性毒性预测,如鼠毒、蜂毒、鸟毒、水生生物毒性等[13,22~26]。这些基于人工智能的毒性预测模型可以快速经济地帮助研究者预测化合物的毒性,以合理的方式来提前避免潜在的不利影响,此外,这些技术可以用于药物发现的早期阶段,有效筛选出低毒的化合物,并且为分子的优化提供指导。

不仅如此,通过利用便捷的互联网技术,在过去几年中这些模型不再仅以复杂的公式、代码等形式展现在文献中,极大地限制了它们的使用,研发者将其嵌入到在线网络平台中,以交互式界面的形式展示给用户,如MouseTox[27],ProTox II[28],admetSAR[29]等。据统计,截至2021年1月,admetSAR在2012年首次发表以来在web of science的被引频次已超过670次,在2019年更新发表后,一年时间内引用频次高达70余次,均为高被引论文[29,30]。由此可见,发展基于人工智能的化合物毒性预测网络服务器是科研工作者的需求,这些免费的服务器提供了便捷的输入形式、快速的结果反馈,不需要用户有专业的计算机或者毒理学研究背景便可以方便地使用这些模型,不仅为更多的科研工作者提供便利的使用条件、节省了实验的成本和时间,而且进一步促进了计算机预测毒性的发展。然而,目前仍然缺乏对基于人工智能预测化合物毒性的算法及相关网络服务器的系统归纳总结。

因此,本文对基于人工智能的化合物毒性预测方法进行了分类及总结,对近三年来搭建的毒性预测网络服务器进行了特点介绍、毒性预测端点统计、性能比较及实例应用展示,同时提出了基于人工智能和互联网时代的化合物毒性预测所面临的机遇和挑战。希望帮助人们了解基于人工智能和互联网时代的化合物毒性预测最新进展、仍存在的不足之处以及未来发展方向,为已具备毒理学知识的专业人员以及计算机与化学生物学等交叉学科的研究人员发展新的人工智能毒性预测模型和网络服务器提供思路和参考信息,同时指导专业的和非专业毒理学研究人员合理选择在线服务器进行化合物毒性评价。

## 1 用于建立毒性预测模型的人工智能算法

截止2020年,世界上的数据量预估将达到35万亿GB[31]。由于探索和分析大数据的需求,以及计算机CPU和GPU等硬件的完善,促进了机器学习、尤其是深度学习算法的发展。可以说人工智能正在改变着我们的日常生活,并被评选为2018年《麻省理工科技评论》全球十大突破性技术之一。近年来,关于人工智能在计算化学和

生命科学中的应用已发表了一些综述文章[31~35]，针对于化合物毒性预测方面的应用也在不断更新发展[21]。在此，我们将重点关注人工智能在毒性预测中的应用，举例介绍在化合物毒性预测中常见的传统的机器学习和深度学习方法，并对其优缺点进行比较。

## 1.1 传统的机器学习算法

支持向量机(SVM)是由 Vapnik 等[36]于 1995 年提出的，能处理小样本数据集的中高维问题，其基本模型即在特征空间上建构最佳的分割超平面，使得训练集上正类和负类样本能够最大的区分。SVM 是一种有效解决二分类问题的有监督学习算法，而不适用于多分类问题。对于线性问题，SVM 模型通过空间上的点映射来分离不同类别的点，使不同类别的点边界最大化。在引入核方法后，SVM 也可以用来解决非线性问题，使用核函数将非线性可分问题从原始的特征空间映射至更高维的特征空间，从而转化为线性可分问题[37]。例如，Cao 等[38]开发了一种基于核融合的 SVM 方法对 DSSTox 数据库中化合物的潜在毒性进行分类，独立验证的预测结果准确率最高可达 90.70%。

决策树(DTs)是一种可解释的机器学习方法，逻辑上以树的形式存在从而进行决策的多分类模型，包含一个根节点、若干个内部结点和若干个叶节点[36]。一般来说，DTs 的构建有两个基本步骤：选择属性和剪枝。选择分子属性作为对分子的“测试”，所选属性被视为非叶节点，每个分支代表一个“测试”结果的输出，每个叶节点代表一种分类结果。属性的选择决定了模型的预测准确度，使得每一个分支节点包含的数据尽可能属于划分结果的同一类别。但是，当分割过细时可能造成过度拟合，并且由于新数据和所训练的数据不同，在面对新的数据时错误率反而会上升。因此，为了降低过拟合的风险和树的复杂性，常使用剪枝算法对生成的树进行修剪。在近期的研究中，Karim 等[39]有效地使用 DTs 从成千上万的特征集合中获得最优数量的化合物 2D 特征，联合浅层神经网络开发了毒性预测模型，从而帮助化学家有效地进行有毒化合物的预筛选。实现了使用较少的计算机资源、训练时间以及化合物特征，获得高精度的预测模型，同时提高模型的可解释性。

随机森林(RF)算法运用了集成学习思想建模，以 DTs 作为基本单元，并引入 Bagging 思想，可以用来解决分类、回归等问题[40]。其用随机采样的子集而不是原始数据集进行训练，最终的结果是综合了所有 DTs 的输出结果：对于分类问题，根据每个 DTs 分类器的投票情况决定最终的分类结果；对于回归问题，由所有 DTs 预测值的均值决定最终的预测结果。因随机性的引入，使得 RF 不太可能过度拟合数据，且有较好的抗噪声能力，但是，当 DTs 数量较多时，将耗费大量的空间和时间。目前，RF 已被广泛应用于毒性预测模型的建立，例如，Shin 等[41]用 RF 回归方法建立亚慢性吸入毒性的 QSAR 模型，Tuulaikhuu 等[42]使用 RF 算法评估化合物、水温、LogP 等因素对淡水鱼产生的毒性影响。

k 邻近(kNN)算法是一种用于分类和回归的无监督学习算法，用于对特征空间中 k 个最邻近的样本进行计数分类，即如果一个在特征空间中的 k 个最邻近的样本中的大多数属于某一类别，则该样本也属于这个类别[43]。因此，kNN 算法是所有机器学习算法中最简单、最容易执行的算法之一，通常与其他特征选择算法相结合，但由于要对每一个待分类的样本计算其到全体已知样本的距离，因此该算法计算量较大，对内存的需求也较大。目前，该算法在预测化合物的基因毒性[44]、器官毒性[45]、急性毒性[46]等方面均有应用。

实际上，很多毒性预测模型在建立之初均用了多种机器学习方法，如 SVM、RF、朴素贝叶斯(naïve Bayes，NB)等，最终在多种模型中选择出表现最好的模型，目前，在许多毒性终点的预测上，如致癌性、致突变性和肝毒性，都取得了重大进展[47]。例如，Zhang 等[48]收集了 785 种药物诱导肝毒性化合物和 532 种无毒化合物，用 MACCS 和 FP4 两种分子指纹来描述化合物特征，使用 SVM、NB、kNN、分类树(classification tree，CT)、RF 五种机器学习方法建立了高精度的分类模型。其中，基于 FP4 指纹的 SVM 方法建立的模型最佳，训练集和测试集的总体预测准确率分别为 79.7% 和 64.5%，外部验证集的总体预测准确率为 75.0%。

## 1.2 深度学习算法

对于许多机器学习问题来说，在遇到复杂问题时需要通过人工的方式来提取有效的特征集合，耗时耗力，而基于深度学习的人工智能技术可以从数据中自动学习和提取更为复杂的特征表达，初始化架构的权重，并采用无监督学习，使用具有多层非线性处理单元的神经网络学习数据表

示,包括深度神经网络(Deep neural networks,DNNs)、递归神经网络(Recurrent neural networks,RNNs)、卷积神经网络(Convolutional neural networks,CNNs)等[49,50],如图 1 所示。它们能够自动捕获相关描述符或属性之间的复杂非线性关系及响应,避免了人工进行特征选择和做出决策时的时间及主观限制。相对于传统的人工神经网络(ANNs),深度学习使用了大量的隐藏层,而 ANNs 由于早期计算机硬件的限制,通常只能提供一到两个隐藏层。由于更强大的 CPU 和 GPU 硬件的出现,深度学习可以在每一层使用更多的节点[31]。另外,深度学习在算法上也有很多改进,例如使用 Dropout[51] 和 DropConnect[52,53] 解决过拟合问题,应用修正线性单元(ReLU)[54] 避免梯度的消失和引入卷积层和池化层作为新型网络体系结构,从而能够使用大量的输入变量。基于深度学习算法的优势,目前已经将该技术逐步应用到化学和生物信息学领域[55,56]。

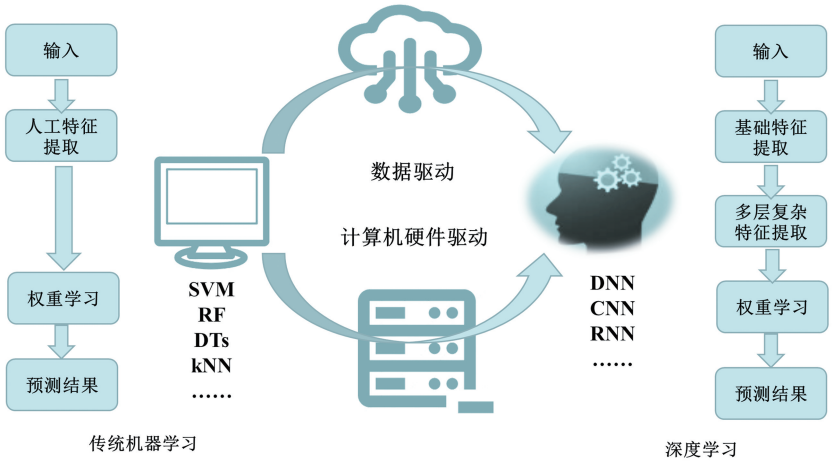


**图 1 传统机器学习算法和深度学习算法对比**

**Fig. 1 Comparison between traditional machine learning algorithm and deep learning algorithm**

首先是 DNNs,包含输入层、多个隐藏层和输出层,层与层之间神经元全连接,可以获取大量的输入特征,不同层次的神经元可以自动提取不同层次的特征[50,57]。例如,通过基因转录组数据直接提取基于机制的肝损伤信息,有望提高目前毒性预测模型的准确性。Wang 等[58] 利用转录组中包含的成千上万的输入变量作为机器学习的特征,建立了从转录组数据预测化学性肝损伤的深度神经网络模型,并且在外部验证中,DNNs 模型的性能表现优于 RF 和 SVM 模型。自动从输入特征中发现与解决分类问题,几乎不需要人工干预是 DNNs 的一个关键优势。原则上,只要有足够的训练集,DNNs 便能够处理大量的输入特征,然而,实际往往因训练数据的缺乏削弱了这种优势。

RNNs 允许同一隐藏层神经元之间的连接形成一个定向循环,即传递当前时刻处理的信息给下一时刻,利用上一时刻学习到的信息进行当前时刻的学习[49]。RNNs 可以将顺序数据作为输入特征,非常适合于语言建模等依赖时间的任务。利用长短期记忆技术(Long short term mermory network,LSTM),RNNs 可以减少梯度消失问题,能够学习长期依赖关系[59]。2020 年,Peng 等[60] 基于双向门控递归单元(BiGRU)的 RNNs 和全连接神经网络建立了分子毒性预测模型,是第一次利用化合物结构特征 SMILES 和理化性质同时进行毒性预测的工作,这种深度混合学习方法有效提高了预测准确性,表现优于 DeepTox、RF、SVM 和 KNN 等方法。

到目前为止,已有以下几种常用类型的输入数据被用于编码化学结构和训练神经网络:理化性质描述符、分子指纹、SMILES。采用 CNNs 不需要对分子描述符进行初步计算和特征选择,避免了显式的特征提取,输入符合 CNNs 图片大小要求的图片,就可以将化合物的化学结构和性质之间的直接关联作为人工神经网络的输入数据导入模型。典型的 CNNs 由卷积层、池化层、全连接层三个部分构成。卷积层用来进行特征提取;池化层用于数据降维,不仅可以减少运算量,还可以有效的避免过拟合;全连接层将前面各层的特征链接起来,组成输入的样本特征,从而输出结果。CNNs 因其能够有效的将含有大数据量的图片降维成较小的数据量处理,并且有效地保留图片特

征,另外,由于每个滤波器共享相同的参数,CNNs 大大减少了学习自由参数的数量,从而降低了内存消耗,提高了学习速度。它在图像识别方面胜过其他类型的机器学习算法。因此,CNNs 被广泛用于图像识别,并且在医学图像处理上得到了较为成功的应用。例如,皮肤癌是人类最常见的恶性肿瘤,主要通过肉眼诊断,Esteva 等[61]开发了一种新的深度 CNNs 图像分类模型来识别皮肤癌,仅使用图片和疾病标签作为输入,测试结果显示其在分类皮肤癌方面的能力可以与专业的皮肤科医生相媲美。又例如,我们用来治疗疾病的药物可能会对肝脏等造成损害,需要在药物发现的早期阶段识别和消除潜在的对肝脏有毒性的候选药物,Asilar 等[62]使用化合物的三维构象的图像,基于 CNNs 开发出了肝毒性预测模型,并且证明了利用深度神经网络进行基于图像的毒性预测的适用性。2014 年,Igarashi 等[63]建立了一个毒物基因组学数据库 Open TG-GATEs,存储了 170 种包括对肝和肾有损伤的药物等化合物对大鼠和人的基因表达和传统毒理学影响数据,以及大量肝、肾切片的数字病理图像。清晰准确的图片是应用 CNNs 建立模型的基础,而相关器官病理图像数据库的建立,将促进基于 CNNs 算法进行毒性预测的发展。

传统机器算法与深度学习算法的对比如表 1 所示。

**表 1 不同学习算法之间优劣对比**

**Tab. 1 Comparison of advantages and disadvantages between different algorithms**

| 算法 | | 优点 | 缺点 |
|---|---|---|---|
| 传统算法 | SVM | 能有效解决二分类问题 | 不适用于多分类问题 |
| | DTs | 具有可解释性 | 易过度拟合 |
| | RF | 具有较好的抗噪声能力 | 耗费大量的空间和时间 |
| | kNN | 方法简单容易执行 | 计算量大 |
| 深度学习 | DNNs | 不需要人工干预 | 训练集不够将影响输入特征识别 |
| | RNNs | 预测准确性优于传统机器算法 | 样本数据要求高 |
| | CNNs | 学习自由参数数量较少 | 样本数据要求高 |

综上,在已报道的毒性预测模型中大多使用了机器学习方法,且均有良好的表现,近年来,由于大量实验数据的积累和 GPU 强大的并行计算能力推动了深度学习算法在毒性预测方面的发展。不同于机器学习,深度学习方法可以在不需要人工输入的情况下处理基于大量、异构或多维数据的复杂任务(如图 1 所示)。值得一提的是,尽管深度学习算法有革命性的突破,在预测毒性方面比传统的机器学习算法具有更大的潜力,但深度学习并不是总表现出更好的性能或提供更好的解决方案,也有相反的情况出现[64~67]。这些研究表明,在缺乏大量训练样本的情况下,机器学习在性能上可能优于深度学习,而且即使有了可用的大数据,深度学习算法也需要进行优化才能表现出优异的性能[68,69]。此外,传统的机器学习通常计算成本更低,对计算机硬件要求较低、时间消耗较少。因此,在建模中需要根据实际情况合理地选择算法,不能盲目使用某一种算法。

## 2 基于人工智能算法的在线毒性预测服务器

模型的建立为化合物的毒性评估带来了新方法,然而,如果这些模型仅在论文中被介绍和引用,这很可能会限制这些模型的使用、与已有方法的比较等。此外,即使提供了算法的源代码或二进制代码,专用计算机平台、特定操作系统、系统参数、编译器、库等也可能需要进行相应的配置,仍然阻碍了模型的使用[70]。与此相反,在网络平台上直接发布这些开发好的模型,一方面可以直接在开发程序的环境中执行程序,易于维护和更新系统,另一方面每个互联网端的用户都可以在他们的工作场所随时访问和使用这些模型,而不需要任何的特殊要求或者编程技能,由于强大的搜索引擎,在线网络资源还能更好地传播有关模型的信息。因此,将这些模型不再以程序代码的形式封装起来,而是通过互联网平台直接供用户使用,不论是对于开发者还是用户都具有十分重要意义。

近年来,在线的化合物毒性预测服务器不断被开发和使用。鉴于预测模型和网站后台技术的不断更新,我们重点对近三年发表的、免费供用户使用的化合物毒性预测服务器进行了调研和总结。表 2 中列出了 9 个在线网络平台,并对相关的模型算法、性能、毒性预测端点等内容进行了分析比较。下面,将具体介绍每个服务器的特点及其应用。

### 2.1 对人体毒性预测的网络服务器

对化合物潜在的基因毒性、细胞毒性、器官毒性等进行评估是至关重要的,尤其是在药物研发的早期阶段,以减少研发后期因候选化合物的毒副作用造成研发的失败。因此,提前对化合物的

表2 基于人工智能算法的在线毒性预测服务器

Tab. 2 Online toxicity prediction server based on artificial intelligence algorithm

| 服务器名称 | 毒性预测端点 | | 毒性预测模型 | | | | 网址[c] |
|---|---|---|---|---|---|---|---|
| | 数目 | 预测范围 | 数据集[a] | 算法[b] | 类型 | 性能评价 | |
| MouseTox | 1 | NIH/3T3细胞毒性 | 5416 | RF | 分类模型 | 0.915;五倍交叉验证精度 | http://enalos.insilicotox.com/MouseTox/ |
| CLC-Pred | 305 | (非)肿瘤细胞系毒性 | 59882 | Naïve Bayes | Pa/Pi值 | 0.93;留一法交叉验证精度 | http://way2drug.com/Cell-line/ |
| Vienna LiverTox Workspace | 15 | 肝损伤 | 84~1872 | RF, BayesNet, SVM, kNN, logistic regression algorithm, NB, rotation forest algorithm | 回归模型,分类模型 | 0.59~0.87;十倍交叉验证精度 | https://livertox.univie.ac.at/ |
| ADVERPred | 5 | 器官毒性(致畸性) | 684~911 | Bayesian | Pa/Pi值 | 0.67~0.77;留一法交叉验证精度 | http://www.way2drug.com/adverpred/ |
| ADMETlab | 7 | 急性毒性、器官毒性(致畸性)、基因毒性(致癌性) | 404~7619 | RF | 回归模型,分类模型 | 0.986;二倍交叉验证精度;0.689~0.844;五倍交叉验证精度 | http://admet.scbdd.com/ |
| eMolTox | 15 | 急性毒性、器官毒性(致畸性)、细胞毒性、基因毒性(致癌性) | 708~37462 | RF | 分类模型 | 0.98~0.83;显著性水平0.1时的有效性值 | http://xundrug.cn/moltox |
| ToxiM | 1 | 分子有无毒性 | 4112 | RF, kNN, CART, SVM | 分类模型 | 0.93;十倍交叉验证精度 | http://metagenomics.iiserb.ac.in/ToxiM/ |
| BeeTox | 1 | 蜂毒 | 891 | GACNN | 分类模型 | 0.837;十倍交叉验证精度 | http://beetox.cn/ |
| admetSAR | 21 | 急性毒性、器官毒性(致畸性)、基因毒性(致癌性)、生态毒性 | 195~10207 | SVM, RF, kNN, CNN | 回归模型,分类模型 | 0.477~0.627;$R^2$ 0.766~0.963;五倍交叉验证精度 | http://lmmd.ecust.edu.cn/admetsar2/ |

[a]预测模型中训练集和测试集的分子数量; [b]用于训练模型的算法; [c]最近访问时间2020年1月17日。

不良毒性效应进行评估可以帮助研发者合理筛选优先考虑进行实验测试的化合物,同时过滤掉那些看起来前景不太好的化合物。

Varsou等[27]用5416个化合物对NIH/3T3细胞毒性作用的实验数据,运用RF算法,开发了一个有效预测化合物对细胞毒性的模型,在5倍交叉验证中,模型精度可达0.915,并搭建了MouseTox云平台为用户提供直接访问的途径。用户通过输入分子的结构信息即可快速预测该分子的细胞毒性,反馈“有活性(化合物对NIH/3T3具有细胞毒性)”或“无活性(化合物对NIH/3T3不具有细胞毒性)”两种分类结果,并同时提供预测可信度“可靠”或“不可靠”以供用户合理参考。另外,该平台支持一次预测多个分子的细胞毒性。总之,MouseTox在计算机辅助药物发现过程中可用于不同化合物的细胞毒性风险评估,促进新化合物的虚拟筛选,有助于减少体外实验和动物试验。

CLC-Pred[71]也是专门用来预测化合物细胞毒性的网络服务器,与MouseTox不同,CLC-Pred能够预测的类型更为广泛,包括化合物对肿瘤和非肿瘤细胞系的细胞毒性。Lagunin等[72~74]根据ChEMBL中59882个化合物的实验数据,使用先前基于朴素贝叶斯(Naïve Bayes)算法开发的PASS算法创建了分类模型,从而预测化合物对不同类型的人类细胞系的细胞毒性。具体包括278个与不同器官或组织有关的肿瘤细胞系的细胞毒性,这对新药的开发或已知药物的重定向有非常重要的指导意义;27个属于不同器官或组织的正常细胞系的细胞毒性,可以有助于估计候选药物在其他治疗领域的开发安全性。用户上传一个分子结构,网页将自动反馈该化合物对肿瘤和非肿瘤细胞系的细胞毒性预测结果,包括:化合物具有“活性”(Pa值)和“非活性”(Pi值)的概率估计

值、细胞系名称、组织名称、肿瘤类型等，如果一个化合物的 Pa 值和 Pa-Pi 值越大，则预测在实验或临床试验中其有活性的可能性越大。肝脏中表达的转运体是维持胆汁流动的关键，抑制这些转运蛋白可能导致肝毒性和药物-药物相互作用。Vienna LiverTox Workspace[75] 能够预测化合物成为肝脏转运体的底物或抑制剂的可能性，以及它引起高胆红素血症、胆汁淤积和药物性肝损伤的风险，共计 15 个端点。建立模型的算法涉及 SVM、RF、NB、BayesNet、kNN、逻辑回归算法等。在十倍交叉验证中，模型的最高精度可达 0.87。用户输入分子结构后，会得到每个端点的预测分类情况（“positive”或“negative”）及对应的打分，得分越高表明该项预测的可信度越大。这可以帮助药物化学家在药物开发的早期阶段对化合物进行筛选，并指导毒理学家对候选化合物进行安全评估。但是，该服务器需要用户注册才能正常使用。

除了以上单独预测化合物细胞毒性、器官毒性的服务器外，也有一些服务器可以同时预测对人体产生不同毒性的端点。如 ADVERPred[76]，与 CLC-Pred[71] 中运用的算法类似，利用 PASS 算法开发出预测 5 种最常见和最严重的药物不良反应模型：心肌梗死、心律失常、心力衰竭、肝毒性和肾毒性，每种效应的训练集平均有 850 种药物。用户输入某个化合物的结构后，将反馈相应不良反应的预测 Pa 值和 Pi 值，从而指导用户合理地选择候选化合物进行下一步的实验。例如，对于罗非昔布（rofecoxib）的预测结果，心肌梗死和心衰得到较高的 Pa 值（>0.5）和较低的 Pi 值（~0.03），推测罗非昔布存在较为严重的心脏毒性，该预测结果与实际数据相符，2004 年其因可增加心血管事件（心梗和卒中）而被宣布退出市场[77,78]。

ADMETlab[79] 提供了 7 个毒性端点的预测，包括心脏 hERG 毒性、肝脏毒性、药物性肝损伤、诱变性、皮肤过敏性、急性经口毒性和每日最大推荐剂量。研究者系统地比较了不同的方法（SVM，RF，NB，RP，PLS，DTs）及不同的分子特征（11 个描述符组和 5 种指纹图谱）对每个毒性端点的预测准确性，从而构建了一系列性能良好的预测模型。其中回归模型的二倍交叉验证精度最高可达 0.986，分类模型的五倍交叉验证精度最高可达 0.844。用户通过输入一个分子，网页将展示不同预测端点的“预测值”、“可能性”、“建议”、“含义”和“参考文献”。对于回归模型，“预测值”显示为数值和单位；在分类模型中，用“+”或“-”的数目清晰和直观地表示“预测值”。根据可能性大小、反馈的建议和每个端点含义中总结的推荐阈值，用户可以更加合理地对预测结果进行判断，并有针对性地对化合物进行进一步的优化。同时，ADMETlab 提供了多样的搜索方式（名称或结构准确搜索、理化性质值域搜索、结构相似性搜索）供用户直接查询该平台已存储化合物的 28 万多条信息。以 ADMETlab 工具中药物毒性预测为例，其预测算法流程如图 2 所示。

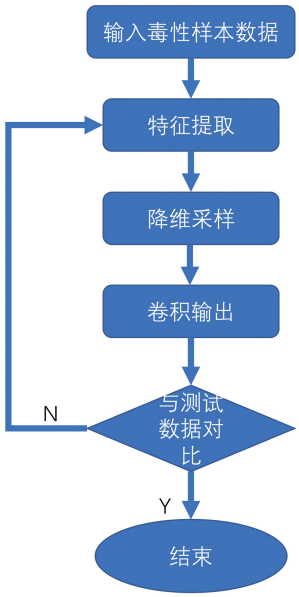


**图 2 ADMETlab 工具预测算法流程**

**Fig. 2 Predictive algorithm of ADMETlab**

eMolTox[80] 中运用 RF 算法、通过 174 组与毒理学相关的体外/体内实验数据集构建模型，用于预测化合物肝毒性、心脏毒性、肾毒性、细胞毒性、基因毒性、线粒体毒性以及对呼吸系统和皮肤的敏感性在内的 15 个毒性端点。然而，基于机器学习方法构建的 QSAR 模型经常会遇到这样的问题，即不能保证该模型能够很好地预测化学空间中的所有分子，因此，该服务器为用户提供了“期望误差率”这一选项，用于提前估计预测结果的可信度[81~83]。用户可以输入分子结构、名称或 CAS 号进行毒性预测，网页将反馈一个表包含该化合物所有可能的潜在毒性端点，以及每个端点的预测可信度打分值，而且还提供了数据库中有相似结构的化合物的毒性信息，可以通过这些信息进一步辅助用户对预测结果进行判断。特别是，eMolTox 还提供了该化合物的结构分析，对可

能造成某一潜在的毒性结构片段进行了标注,为用户进行化合物结构的优化提供了指导。

此外,ToxiM[84]是一个基于机器学习开发的二元分类器,用于综合的预测小分子有无毒性。模型根据已知的 1263 个无毒化合物和 2849 个有毒化合物,以及 RF、kNN、CART 和 SVM 算法建立,并从中选择最佳的机器学习方法(RF)用于分类。在 10 倍交叉验证中,基于描述符、指纹和混合的分类模型准确率均约为 93%。用户可以提交一个或多个化合物进行毒性预测,服务器将自动反馈化合物打分值帮助用户判断分子有无毒性。另外,该服务器还提供了“Caco-2 渗透率”和“水溶解度”预测模块,以及“相似性搜索”模块帮助用户识别与毒素数据库中存在的最相似的分子。

## 2.2 对生态物种毒性预测的网络服务器

对于专门预测生态物种的在线网络服务器目前还比较少见,大多数的研究仍集中于构建预测模型,例如:用 SVM、kNN、RF、DTs 等机器学习算法构建模型来预测化合物对蜜蜂的急性触杀毒性 $LD_{50}$ 值,对鸟类和鱼类的急性毒性分类等[23~26]。特别是,2020 年 Wang 等[85]构建了首个基于深度学习算法预测蜜蜂毒性的云平台 BeeTox。作者将无向图(Undirected graph,UG)和注意卷积神经网络(Attention convolutional neural networks,ACNN)相结合,建立了一种新的深度图注意卷积神经网络(Graph attention convolutional neural networks,GACNN)模型用于预测化合物对蜜蜂的毒性。用户输入化合物结构后点击预测,结果将划分为“有毒”和“无毒”两种类型。通过与其他常用的机器学习模型(包括 SVM、逻辑回归、DNN)进行比较,发现 GACNN 模型有较好的预测性能,最高提高了约 6%,且有较好的稳定性。另外,该服务器提供了所有预测集和测试集的化合物结构及对蜜蜂毒性的毒性信息,以及化合物对蜜蜂的毒性机制并可以通过分子相似性进行映射。总之,该平台免费为用户提供了 891 种化合物的蜂毒信息,并可对任意化合物进行蜂毒预测,促进了生态物种毒性预测的相关网络服务器的发展。

## 2.3 综合的毒性预测网络服务器

目前,同时能够预测基因毒性、细胞毒性、器官毒性以及对不同物种个体毒性的服务器还比较少见。admetSAR[29,30]是一个能够综合预测多种毒性端点的网络服务器,包括器官毒性、基因毒性和生态毒性等 15 个毒性端点,如:肝脏毒性、对眼睛的损伤、致癌作用、诱变作用等。其中对于生态毒性预测,涵盖的物种有鱼、蜜蜂、四膜虫、甲壳类水生生物和哺乳动物。这些模型包括二分类、三分类模型和回归模型,且大多运用 SVM、kNN、RF 等机器学习算法建立,特别是,急性经口毒性和四膜虫的急性毒性模型运用了图卷积神经网络(Graph convolutional neural network,GCNN)建立。admetSAR 不仅是一个服务器,同时是一个数据库。用户可以对特定的化合物进行预测,并且支持一次同时预测多个化合物(不超过 10 个),网页将快速反馈预测结果及可能性,同时,提供用户通过化合物的 CAS 号、通用名、SMILES 和理化性质值域进行特定的搜索,并能够快速与库中已有的 9.6 万个化合物进行匹配。总之,admetSAR 将为生物毒理学、环境毒理学的研究提供丰富的资源和便利的条件。

综上所述,目前基于人工智能算法的在线毒性预测服务器总共分为针对人体毒性预测、针对生态物种毒性预测以及综合的毒性预测网络服务器三种类型。由于受关注程度有限,有关专门预测生态物种的网络服务器还比较少见。由于不同软件工具针对的毒性分析类型各有侧重,现有的在线工具还无法针对同一类型的试样结果进行横向比较。对比几类服务器,其中 eMolTox 的使用受众较广泛,admetSAR 服务器普适性和通用性更强。

值得一提的是,在线网络服务器为研究者提供便利条件的同时,用户必须考虑到预测结果的不确定性。由于建立模型时选用的训练集、使用的算法参数、选择的分子描述符和采样方法的不同,同一化合物的毒性端点预测结果可能会不一致,开发者也不可能对所有模型进行全面的测试比较。因此,用户可以根据预测结果中提供的可能性打分,或者通过不同平台的多次预测,综合考虑预测结果从而提高判断的准确性。总之,与任何计算机辅助预测一样,所有预测结果仅是提供参考作用,供研究者合理地进行分子设计和实验,最终的结论还是应与实际的实验相结合。

# 3 人工智能和互联网时代下毒性预测面临的机遇与挑战

通过体内或体外实验进行化合物的毒性评估

既消耗资金又消耗时间,运用计算机预测结果来合理的指导实验是一种有效的替代方法。在过去的几十年里,机器学习算法,如定量结构活性关系(QSAR)模型被开发出来,可以快速且廉价地预测数百万候选化合物的生物毒性。当进入大数据时代,除了机器学习,深度学习算法逐步发展起来,是一种更强大、更有效的方式来处理大量的数据集及更为复杂的问题。而在线预测平台的不断开发,更是极大地方便了科研工作者将这些模型算法运用到实际的预测中,为化合物毒性风险评估提供了指导,也促进了预测模型的进一步发展与改进。总之,随着人工智能和互联网时代的发展,为化合物的毒性预测提供了新的机遇,同时也面临着一些挑战:

(1)实验数据的问题。①大量且高质量的数据会直接影响到模型的性能和可靠性。尽管已有开源的化学信息学和生物信息学数据库,如 PubChem[86,87]、ChEMBL[88~90]、ExCAPE-DB[91] 辅助该项研究,但实际可用的数据仍然有限。此外,毒性测试的实验数据大部分是发表在论文中,或者由制药公司测试产生且不对外开放使用。因此,需要利用计算机技术,一方面提供数据挖掘等有效方法获得可用数据,另一方面,要尽量用少的数据开发出可靠模型[92,93]。②需要建立基准库。与注释计算机语音和视觉数据不同,准确地提取化学分子的有用信息并将其特征化,从而转化成适用于机器学习算法的形式,往往需要该领域的专家进行监督[94~96]。而且,由于大多数新算法都是以不同的数据集为基准的,这使得衡量所提出方法的性能具有挑战性,也使算法的发展受到限制。因此,需要建立一个用于分子机器学习的基准库,同时管理多个公共数据集,提供高质量且开源的分子特征和相关算法,如 MoleculeNet[97]。

(2)模型的可解释性。与以知识为基础的毒性预测方法不同,深度学习作为一种基于神经网络的人工智能算法,通过模仿人脑的复杂结构来认知和处理问题,因此,科学家尚不能解释它们是如何使用这些数据得出结论的。目前,尽管模型具有一定的预测准确性,但是,计算机领域的专家仍不能通过合理的解释,如通过化合物结构与毒性之间的关系来说服实验科学家相信这些结果。此外,用于机器学习的一些描述分子特征的描述符如理化性质、分子指纹、拓扑性质等也不适用于深度学习[98]。因此,深度学习模型仍然扮演着“黑盒子”的角色,很难揭示其中的毒理学机制[36]。

(3)预测结果的可靠性。在所有的预测模型中,即便是(外部)测试集有较高的预测准确度,但在实际应用中仍会有假阳性或者假阴性结果产生,这可能与训练集的覆盖范围有关[99]。这就要求模型的开发者不断更新模型,也要求用户不能完全依赖于某次的预测结果,可以结合多种模型的结果综合参考,合理进行判断,最终仍是需要与实验相结合。

(4)网络安全问题。用户在使用在线网络服务器进行毒性评估时,会将自己预测的化合物结构上传,由于这是一个公开的平台,因此,可能将化合物的信息泄露,而对于研发者,尤其是药物研发者来说,这些结构往往涉及到保密[100]。所以,平台开发者需要建立一个安全的环境确保用户的信息不被泄密,也可以提供能够下载的软件包,供用户自行使用,而此时同时需要关注软件或代码可能会无意中传播病毒或间谍软件。

(5)缺乏一个预测化合物全毒性的平台。“One health”理念强调包括人、动物和环境乃至整个生态系统在内的整体健康观,因此,在化合物毒性评价方面也应遵循该理念,不仅关注化合物对人类的健康,同时也要重视其对其他不同生态物种的安全问题。这点在对农药分子进行环境风险评估及登记注册时显得尤为重要[101,102]。然而,目前的在线毒性预测平台重点关注于细胞、器官、人体毒性,尽管已出现了 admetSAR 2.0[30] 可以综合预测包括基因、器官、生态毒性的服务器,但其涉及到的毒性端点和生态物种仍有限。因此,需要建立一个全面预测化合物对不同毒性端点和物种的毒性的平台,以方便人们直接使用并进行化合物的毒性风险评价。

## 4 结束语

综上所述,随着计算机科学的发展,计算机辅助化合物毒性预测为传统的实验测试化合物毒性提供了新的思路和方法。近年来,人工智能也逐渐应用于毒性预测领域,研究者通过传统的机器学习、深度学习算法建立了预测不同毒性端点的模型,并通过互联网技术将这些可靠的模型嵌入到在线的网络服务器中,为更多的科研工作者提供了帮助。基于此,本文系统介绍了常用于建立毒性预测模型的人工智能算法,总结了近三年

来搭建的毒性预测网络服务器及其特点,另外,提出了基于人工智能和互联网时代的化合物毒性预测所面临的机遇和挑战。希望为发展新的毒性预测模型及相关网络资源提供参考,同时也指导用户合理选择在线服务器进行化合物的毒性预测研究。

## 参 考 文 献


[1] Godlee F, Waters A. Brit. Med. J., 2018, 362.
[2] Amuasi J H, Lucas T, Horton R, et al. Lancet, 2020, 395 (10235): 1469~1471.
[3] Kahn L H. Nature, 2017, 543 (7647): S47~S47.
[4] REACH: The New EU Chemicals Policy. 2006.
[5] Tweedale A C. J. Appl. Toxicol., 2017, 37 (1): 92~104.
[6] Ernst F R, Grizzle A J. J. Am. Pharm. Assoc., 2001, 41 (2): 192~199.
[7] Wester K, Jonsson A K, Spigset O, et al. Brit. J. Clin. Pharm., 2008, 65 (4): 573~579.
[8] Lim E, Pon A, Djoumbou Y, et al. Nucl. Acids Res., 2010, 38, D781~D786.
[9] Zhang L, McHale C M, Greene N, et al. Environ. Mol. Mutagen., 2014, 55 (9): 679~688.
[10] Thomas R S, Bahadori T, Buckley T J, et al. Toxicol. Sci., 2019, 169 (2): 317~332.
[11] Quinn F R, Neiman Z, Beisler J A. J. Med. Chem., 1981, 24 (5): 636~639.
[12] Veith G D. SAR QSAR Environ. Res., 2004, 15 (5-6): 323~330.
[13] Muster W G, Breidenbach A, Fischer H, et al. Drug Discov. Today, 2008, 13 (7-8): 303~310.
[14] Khan K, Benfenati E, Roy K. Ecotoxicol. Environ. Safety, 2019, 168, 287~297.
[15] Liu R, Madore M, Glover K P, et al. Toxicolog. Sci., 2018, 164 (2): 512~526.
[16] Mullard A. Nature, 2017, 549 (7673): 445~447.
[17] Raies A B, Bajic V B. WIREs-Comput. Mol. Sci., 2016, 6 (2): 147~172.
[18] Roncaglioni A, Toropov A A, Toropova A P, et al. Curr. Opin. Pharm., 2013, 13 (5): 802~806.
[19] Yang H, Sun L, Li W, et al. Front. Chem., 2018, 6: 30.
[20] Svetnik V, Liaw A, Tong C, et al. J. Chem. Inf. Comput. Sci., 2003, 43 (6): 1947~1958.
[21] Wu Y, Wang G. Int. J. Mol. Sci., 2018, 19 (8): 2358.
[22] Li X, Chen L, Cheng F, et al. J. Chem. Inf. Model., 2014, 54(4): 1061~1069.
[23] Li X, Zhang Y, Chen H, et al. J. Chem. Inf. Model., 2017, 57 (12): 2948~2957.
[24] Zhang C, Cheng F, Sun L, et al. Chemosphere, 2015, 122: 280~287.
[25] Sun L, Zhang C, Chen Y, et al. Toxicol. Res., 2015, 4 (2): 452~463.
[26] Li F, Fan D, Wang H, et al. Toxicol. Res., 2017, 6 (6): 831~842.
[27] Varsou D-D, Melagraki G, Sarimveis H, et al. Food Chem. Toxicol., 2017, 110: 83~93.
[28] Banerjee P, Eckert A O, Schrey A K, et al. Nucl. Acids Res., 2018, 46 (W1): W257~W263.
[29] Cheng F, Li W, Zhou Y, et al. J. Chem. Inf. Model., 2012, 52 (11): 3099~3105.
[30] Yang H, Lou C, Sun L, et al. Bioinformaticsm, 2018, 35 (6): 1067~1069.
[31] Chen H, Engkvist O, Wang Y, et al. Drug Discov. Today, 2018, 23 (6): 1241~1250.
[32] Rifaioglu A S, Atas H, Martin M J, et al. Brief. Bioinform., 2019, 20 (5): 1878~1912.
[33] Lipinski C F, Maltarollo V G, Oliveira P R, et al. Front. Robot. AI, 2019, 6: 108.
[34] Ekins S. Pharm. Res., 2016, 33 (11): 2594~2603.
[35] Goh G B, Hodas N O, Vishnu A. J. Comput. Chem., 2017, 38 (16): 1291~1307.
[36] Zhang L, Tan J, Han D, et al. Drug Discov. Today, 2017, 22 (11): 1680~1685.
[37] Cao D-S, Dong J, Wang N-N, et al. Chemometr. Intell. Lab. Syst, 2015, 146: 494~502.
[38] Cao D S, Zhao J C, Yang Y N. SAR QSAR Environ. Res., 2012, 23 (1-2): 141~153.
[39] Karim A, Mishra A, Newton M A H, et al. ACS Omega, 2019, 4 (1): 1874~1888.
[40] Mistry P, Neagu D, Trundle P R, et al. Soft Comput., 2016, 20 (8): 2967~2979.
[41] Shin J H, Lee B H, Lee S K. Bull. Korean Chem. Soc., 2019, 40 (8): 819~825.
[42] Tuulaikhuu B-A, Guasch H, Garcia-Berthou E. Environ. Sci. Pollut. Res., 2017, 24 (11): 10172~10181.
[43] Ertugrul O F, Tagluk M E. Appl. Soft Comput., 2017, 55: 480~490.
[44] Sinha M, Pandit S, Singh P. J. Ind. Chem. Soc., 2019, 96 (7): 957~966.
[45] An Y R, Kim J Y, Kim Y S. Mol. Cell. Toxicol., 2016, 12 (1): 1~6.
[46] Chavan S, Friedman R, Nicholls I A. Int. J. Mol. Sci., 2015, 16 (5): 11659~11677.
[47] Zhang L, Zhang H, Ai H, et al. Curr. Topics Med. Chem., 2018, 18 (12): 987~997.
[48] Zhang C, Cheng F, Li W, et al. Mol. Inform., 2016, 35 (3-4): 136~144.
[49] Jing Y, Bian Y, Hu Z, et al. AAPS J., 2018, 20 (3): 58.
[50] LeCun Y, Bengio Y, Hinton G. Nature, 2015, 521 (7553): 436~444.
[51] Srivastava N, Hinton G, Krizhevsky A, et al. J. Mach. Learn. Res., 2014, 15: 1929~1958.
[52] Sakai Y, Pedroni B U, Joshi S, et al. IEEE J. Em. Sel. Top. C., 2019, 9 (4): 658~667.
[53] Liu M, Song J, Wang Z, et al. Optimization of Deep Convolution Neural Network Based on Sparse DropConnect//

Journal of Physics: Conference Series. IOP Publishing, 2018, 1061(1): 012014.
[54] Qiu S, Xu X, Cai B. FReLU: flexible rectified linear units for improving convolutional neural networks//2018 24th international conference on pattern recognition (icpr). IEEE, 2018: 1223~1228.
[55] Ferreira L L G, Andricopulo A D. Future Med. Chem., 2019, 11 (5): 371~374.
[56] Angermueller C, Parnamaa T, Parts L, et al. Mol. Systems Biol., 2016, 12 (7): 878.
[57] Lee H, Grosse R, Ranganath R, et al. Commun. ACM, 2011, 54(10): 95~103.
[58] Wang H, Liu R, Schyman P, et al. Front. Pharm., 2019, 10: 42.
[59] Hochreiter S, Schmidhuber J. Neural Comput., 1997, 9 (8): 1735~1780.
[60] Peng Y, Zhang Z, Jiang Q, et al. Methods, 2020, 179: 55~64.
[61] Esteva A, Kuprel B, Novoa R A, et al. Nature, 2017, 542 (7639): 115~118.
[62] Asilar E, Hemmerich J, Ecker G F. J. Chem. Inf. Model., 2020, 60 (3): 1111~1121.
[63] Igarashi Y, Nakatsu N, Yamashita T, et al. Nucl. Acids Res., 2015, 43 (D1): D921~D927.
[64] Idakwo G, Thangapandian S, Luttrell J, et al. Front. Physiol., 2019, 10: 1044.
[65] Mayr A, Klambauer G, Unterthiner T, et al. Chem. Sci., 2018, 9 (24): 5441~5451.
[66] Huang R, Xia M, Sakamuru S, et al. Nat. Commun., 2016, 7: 10425.
[67] Huang R, Xia M, Nguyen D-T, et al. Front. Environ. Sci., 2016, 3: 85.
[68] Fuentes A F, Yoon S, Lee J, et al. Front. Plant Sci., 2018, 9: 1162.
[69] Matsuzaka Y, Uesawa Y. Front. Bioeng. Biotech., 2019, 65.
[70] Tetko I V. Drug Discov. Today, 2005, 10 (22): 1497~1500.
[71] Lagunin A A, Dubovskaja V I, Rudik A V, et al. PloS One 2019, 13 (1): e0191838.
[72] Poroikov V V, Filimonov D A, Borodina Y V, et al. J. Chem. Inf. Comp. Sci., 2000, 40 (6): 1349~1355.
[73] Lagunin A, Filimonov D, Poroikov V. Curr. Pharm. Design., 2010, 16 (15): 1703~1717.
[74] Filimonov D A, Lagunin A A, Gloriozova T A, et al. Chem. Heterocycl. Compd., 2014, 50 (3): 444~457.
[75] Montanari F, Knasmüller B, Kohlbacher S, et al. Front. Chem., 2020, 7: 899.
[76] Ivanov S M, Lagunin A A, Rudik A V, et al. J. Chem. Inf. Model., 2018, 58 (1): 8~11.
[77] Sibbald B. Can. Med. Assoc. J., 2004, 171 (9): 1027~1028.
[78] Maxwell C B, Jenkins A T. Am. J. Health-Syst Pharm., 2011, 68 (19): 1791~1804.
[79] Dong J, Wang N N, Yao Z J, et al. J. Cheminform., 2018, 10: 9.
[80] Ji C, Svensson F, Zoufir A, et al. Bioinformatics, 2018, 34 (14): 2508~2509.
[81] Norinder U, Carlsson L, Boyer S, et al. J. Chem. Inf. Model., 2014, 54 (6): 1596~1603.
[82] Norinder U, Carlsson L, Boyer S, et al. Regul. Toxicol. Pharm., 2015, 71 (2): 279~284.
[83] Shafer G, Vovk V. J. Mach. Learn. Res., 2008, 9: 371~421.
[84] Sharma A K, Srivastava G N, Roy A, et al. Front. Pharm., 2017, 8: 880.
[85] Wang F, Yang J-F, Wang M-Y, et al. Sci. Bull., 2020, 65 (14): 1184~1191.
[86] Kim S, Chen J, Cheng T J, et al. Nucl. Acids Res., 2019, 47 (D1): D1102~D1109.
[87] Wang Y L, Bryant S H, Cheng T J, et al. Nucl. Acids Res., 2017, 45 (D1): D955~D963.
[88] Wassermann A M, Bajorath J. Binding D B, et al. Expert Opin. Drug Discov., 2011, 6 (7): 683~687.
[89] Gaulton A, Hersey A, Nowotka M, et al. Nucl. Acids Res., 2017, 45 (D1): D945~D954.
[90] Mendez D, Gaulton A, Bento A P, et al. Nucl. Acids Res., 2019, 47 (D1): D930~D940.
[91] Sun J, Jeliazkova N, Chupakin V, et al. J. Cheminform., 2017, 9: 17.
[92] Hernandez Medrano I, Tello Guijarro J, Belda C, et al. Int. J. Interact. Multi. Artif. Intell., 2018, 4 (7): 8~12.
[93] Gavrilov D, Melerzanov A, Schelkunov N, et al. Artificial intelligence image recognition inhealthcare//2018 International Conference on Artificial Intelligence Applications and Innovations (IC-AIAI). IEEE, 2018: 24~26.
[94] Rogers D, Hahn M. J. Chem. Inf. Model., 2010, 50 (5): 742~754.
[95] Miller G A. Commun. ACM, 1995, 38 (11): 39~41.
[96] Deng J, Dong W, Socher R, et al. Imagenet: A large-scale hierarchical image database//2009 IEEE conference on computer vision and pattern recognition. IEEE, 2009: 248~255.
[97] Wu Z, Ramsundar B, Feinberg E N, et al. Chem. Sci., 2018, 9 (2): 513~530.
[98] Kearnes S, McCloskey K, Berndl M, et al. J. Comput. Aid. Mol. Des., 2016, 30 (8): 595~608.
[99] Segall M D, Barber C. Drug Discov. Today, 2014, 19 (5): 688~693.
[100] Jia C-Y, Li J-Y, Hao G-F, et al. Drug Discov. Today, 2020, 25 (1): 248~258.
[101] Mostafalou S, Abdollahi M. Arch. Toxicol., 2017, 91 (2): 549~599.
[102] Madariaga-Mazon A, Osnaya-Hernandez A, Chavez-Gomez A, et al. Toxicol. Res., 2019, 8 (2): 146~156.