

基于深度信念网络的 CYP450 2C9 抑制性分类

李自臣^{1,2} 史新宇^{2*} 禹龙³ 田生伟² 王梅⁴ 李莉⁴

¹(乌鲁木齐职业大学信息工程学院 新疆 乌鲁木齐 830002)

²(新疆大学软件学院 新疆 乌鲁木齐 830008)

³(新疆大学网络中心 新疆 乌鲁木齐 830046)

⁴(新疆医科大学药学院 新疆 乌鲁木齐 830011)

摘要 细胞色素 P450 2C9 (Cytochrome P450 2C9) 是人体肝脏中重要的代谢酶, 参与多种药物代谢, 约占 CYP450 蛋白总量的 15% ~ 20%。利用深度学习思想, 提出基于深度信念网络 DBN (Deep Belief Network) 的 CYP450 2C9 抑制性分类模型。实验选用 13 000 个化合物作为数据集, 采用 PubChem 和 MACCS 分子指纹进行分子结构表征。利用 DBN 的半监督学习方式从预处理后的特征中学习更本质的特征表示, 避免人工提取特征的过程, 实现 CYP450 2C9 的抑制性分类。实验结果表明: 在同等条件下, DBN 相比于 SVM 和 ANN 具有明显优势, 平均分类准确率为 80.6%, 灵敏度 (SE) 为 86.9%, 特异性 (SP) 为 66.2%, 对药物筛选和新药研发具有积极意义。

关键词 深度学习 CYP450 2C9 分子指纹 深度信念网络 支持向量机 人工神经网络

中图分类号 TP391

文献标识码 A

DOI:10.3969/j.issn.1000-386x.2019.02.034

CLASSIFICATION OF CYP450 2C9 INHIBITORS BASED ON DEEP BELIEF NETWORK

Li Zichen^{1,2} Shi Xinyu^{2*} Yu Long³ Tian Shengwei² Wang Mei⁴ Li Li⁴

¹(School of Information Engineering, Urumqi Vocational University, Urumqi 830002, Xinjiang, China)

²(School of Software, Xinjiang University, Urumqi 830008, Xinjiang, China)

³(Network Center, Xinjiang University, Urumqi 830046, Xinjiang, China)

⁴(College of Pharmacy, Xinjiang Medical University, Urumqi 830011, Xinjiang, China)

Abstract The cytochrome P450 2C9 is one of the important metabolic enzymes in the human liver, which involves in many kinds of drug metabolism and accounts for about 15% ~ 20% of the total CYP450 protein. Deep learning was adopted to propose a CYP450 2C9 inhibitory classification model based on deep belief network (DBN). In the experiment, 13 000 compounds were selected as data sets. PubChem and MACCS molecular fingerprints were used to characterize the molecular structure. We used DBN semi-supervised learning method to learn more essential feature representation from the features obtained by pre-processing, thus avoiding the process of extracting features manually and realizing classification of CYP450 2C9 inhibitors. Experimental results show that DBN has outstanding advantages compared with SVM and ANN at the same conditions. The average accuracy is 80.6%, the sensitivity (SE) is 86.9% and the specificity (SP) is 66.2%. It has a positive significance in drug screening and development.

Keywords Deep learning CYP450 2C9 Molecular fingerprint Deep belief network Support vector machine Artificial neural network

0 引言

吸收、分布、代谢、排泄和毒性问题是造成临床医药化合物失败的主要因素,约有 70% 的药物在临床实验中失败或退出市场^[1]。其中,新陈代谢决定一个化合物进入人体的命运,并最终控制该化合物是否具有毒副作用。细胞色素 P450 酶是药物代谢酶中的一个家族,它是肝脏中药物代谢的主要位点,负责人体 90% 以上的临床药物代谢^[2]。在代谢第一阶段,细胞色素 P450 酶的亚型主要通过氧化反应,修改各种各样的基质,使它们有更强的水溶性且易于被消除^[3]。

CYP2C9 作为细胞色素 P450 第二亚家族中的一个重要成员,在人体肝脏中,约占全部的 CYP450 蛋白总量的 20%,并且已存在于市场上的 16% 的临床药物由它负责代谢。例如抗惊厥类药物、抗凝血类药物、非甾体抗炎药及其他种类的药物等。它的抑制作用可能会导致药物较低的代谢速率及等离子体浓度的增加,并进一步导致药物产生毒性作用。因此,在早期的药物筛选和安全评价中,区分 CYP2C9 的抑制性和非抑制性成为重要的研究课题。

近年来,基于机器学习的 QSAR 建模方法已广泛应用于 CYP450 酶的抑制性分类。文献[4]以 Three-Point Pharmacophoric (3PP) 分子指纹特征作为 SVM 模型的输入,对 1 100 个化合物训练,在包含 238 个化合物的测试集上实现 CYP2C9 的抑制性分类。Cheng 等^[5]分别使用偏最小二乘判别分析法和 SVM 方法,结合两种基于分子指纹的描述符实现对 CYP2C9 的抑制性分类。李兰婷等^[6]采用逐步判别分析法和 K-均值聚类分析法(K-Means cluster analysis method)建立模型,对 81 个化合物进行训练和测试,取得了较好的效果。然而,这些方法大多基于有监督的浅层学习模型,其性能依赖分子特征的选择,容易出现维数灾难和局部最优等问题。

本文利用深度学习思想,提出了基于 DBN 的 CYP2C9 抑制性分类模型。通过堆叠多层受限玻尔兹曼机(RBM),利用其组合低层数据特征并充分挖掘分布式特征的能力,对分子特征进行更本质的学习。采用反向传播神经网络(BP)对多层 RBM 进行有监督的微调,完成对 CYP2C9 的抑制性分类。通过与浅层学习模型 SVM 和 ANN 进行对比,本文所提模型克服了浅层模型容易出现维数灾难和局部最优等问题,验证了深度信念网络模型对 CYP2C9 抑制性分类的有效性。

1 相关工作

1.1 数据源

本文所获取的数据集源于文献[7],它从 PubChem BioAssay 数据库中收集了 13 908 个结构多样的化合物(PubMed ID:AID410),为避免样本的重复和错误,所有化合物都经过处理和检验。原始数据集中包含一些信息,例如分子 ID、记录 ID、化合物的 SIMLES 结构,以及 CYP2C9 的抑制性和非抑制性标识。本文选取 13 000 条数据,并以 4:1 的比例将全部数据集分为训练集和测试集。数据集的详细类别分布见表 1。全部实验采用五折交叉验证评估模型的性能,避免实验的随机性和偶然性。数据集可从 Online Chemical Modeling Environment (OCMEM)中免费下载。

表 1 训练集和测试集的详细类别分布

类别	训练集	测试集
抑制性	6 857	1 715
非抑制性	3 543	885
全部	10 400	2 600

1.2 分子指纹

分子指纹技术是描述化合物结构属性的一种方法,通过检测分子结构中一些特定子结构(即分子结构片段)是否存在,从而把分子结构转化为一组二进制指纹序列^[8]。目前,有多种不同形式分子指纹,如 FP2、FP3、FP4 指纹、MACCS 指纹、Estate 指纹、Pubchem 指纹,以及 Daylight 指纹等。

1.3 分子指纹生成

本文采用新加坡国立大学 Yap 等^[9]开发的 PaDEL-Descriptor 描述符计算软件。该软件当前可计算 797 个描述符(1 维和 2 维 663 个,3 维 134 个)及 10 种不同类型的分子指纹。该软件免费且开源,有便于用户使用的图形用户界面,可运行在多个平台,接受多种类型的文件格式。利用该软件生成了常用的 PubChem 和 MACCS 分子指纹,其中 MACCS 根据 166 位结构片段词典进行编码产生二进制字符串。这些结构特征包括原子类型、化学键类型、原子环境类型与结构性质等。用 0 和 1 来表示分子中相关结构特征信息的存在与缺失,即当分子中存在某一结构特征时,就在预定义结构特征位点构成的位串(词典)中相应位置标记为 1,否则标记为 0^[10]。

1.4 分子指纹预处理

分子指纹维度过高,会增加模型的计算量和运行

时间。为避免分子特征冗余,本文按以下规则对分子指纹特征进行筛选:(1) 去除重复特征;(2) 去除全为零的特征;(3) 去除标准差等于零的特征。具体实现过程如下:

算法:分子指纹预处理

i 为单个样本, num 为样本数量, c 为特征个数

1. for i in $datas.num$:
2. if $data[i].std() == 0$: $remove.append(i)$
3. $c = datas.feature.num$
4. for i in $range(c)$
5. $v = datas[i]$
6. for j in $range(i+1, c)$:
7. if $equal(v, datas[j])$: $remove.append(c[j])$

2 模型介绍

2.1 深度信念网络 (DBN)

深度神经网络 DNN 利用多层神经网络训练模型, 不仅能克服一些浅层机器学习模型的局限性, 尽可能使用较少参数实现复杂函数逼近, 而且有很强的自学习能力。同时, 它能从原始高维特征中抽取多层分布表示。DBN 作为 DNN 的一种, 由 Hinton 于 2006 年提出, 已被广泛应用于多个领域^[11]。DBN 由多层 RBM 和 BP 网络组成。DBN 训练过程分两步: 预训练和微调。首先, 采用无监督学习方式对 RBM 每一层训练, 保证当特征向量被映射到不同特征空间时, 特征信息能尽可能多的保留; 然后利用最后一层 BP 网络以有监督训练方式微调整个 DBN 网络权重值。它以 RBM 输出向量作为自身输入向量训练一个实体关系分类器。DBN 结构如图 1 所示。

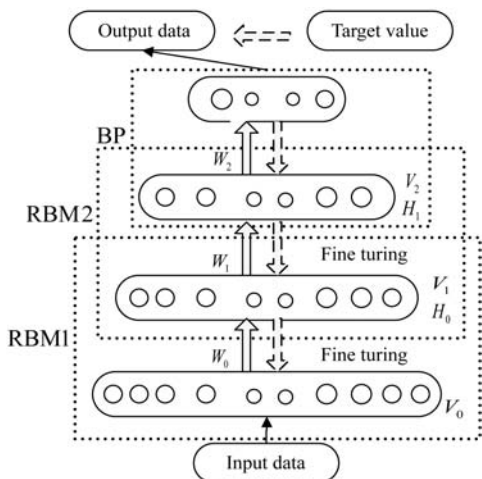


图 1 DBN 模型结构

2.2 受限玻尔兹曼机 (RBM)

受限玻尔兹曼机是一个生成式随机神经网络, 由

Hinton 和 Sejnowski 于 1986 年提出^[12]。它由可见单元和隐单元组成, 这些单元是二值变量, 状态为 0 或 1。全部神经网络是一个二部图, 可见层和隐藏层之间全连接, 层内之间无连接。RBM 结构如图 2 所示。

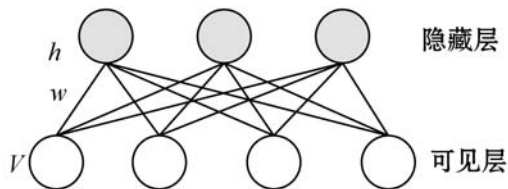


图 2 RBM 的结构

图 2 中, 可以看到 RBM 包含 4 个可见单元 (由 v 表示) 和 3 个隐藏单元 (由 h 表示), w 是一个 4×3 矩阵, 它表示可见层和隐藏层之间边的权重。受统计学能量函数概念的启发, RBM 引入能量函数的概念: “联合配置 (v, h)”, 被定义为:

$$E(v, h; \theta) = - \sum_{ij} W_{ij} v_i h_j - \sum_i b_i v_i - \sum_j a_j h_j \quad (1)$$

式中: θ 是 RBM 的一个参数, 表示为 $\{W, a, b\}$; w 表示可见层的基向量; b 表示隐藏层的基向量。根据玻尔兹曼分布, 可见单元和隐藏单元的联合概率如下:

$$P_{\theta}(v, h) = \frac{1}{Z(\theta)} \exp\left(\sum_{i=1}^D \sum_{j=1}^F W_{ij} v_i h_j + \sum_{i=1}^D v_i b_i + \sum_{j=1}^F h_j a_j \right) \quad (2)$$

$Z(\theta)$ 是一个归一化因子 (也称为配分函数), 采用 sigmoid 激活函数, 公式如下:

$$Z(\theta) = \frac{1}{1 + \exp(-\theta)} \quad (3)$$

学习 RBM 的任务是求出参数 θ 的值, 以拟合给定的训练数据, 参数 θ 可以通过最大化 RBM 在训练集上的对数似然函数 ($P(v)$) 学习得到, $P(v)$ 可由式 (4) 得到。最大化 $P(v)$ 等同于最大化 $\log(P(v)) = L(\theta)$, 如式 (5) 所示。

$$P_{\theta}(v) = \frac{1}{Z(\theta)} \sum_h \exp[v^T W h + a^T h + b^T v] \quad (4)$$

$$L(\theta) = \frac{1}{N} \sum_{n=1}^N \log P_{\theta}(v^{(n)}) \quad (5)$$

其中: 可见层节点集合的边缘分布为 $P_{\theta}(v)$, 然而计算 $P_{\theta}(v)$ 非常困难。因此, Hinton 等提出了对比散度算法解决这一问题。

经过这一步, RBM 提取出的特征向量作为 BP 模型的输入。BP 网络可以微调整个 DBN 网络。它的训练过程主要分为两步: 一是前向传播, 将输入特征向量沿输入端传播至输出端; 二是反向传播, 将 BP 网络的输出结果与正确结果比较得到误差, 然后将误差从输出端反向传播至输入端。

2.3 DBN模型对CYP2C9的分类流程

基于DBN模型的CYP2C9抑制性和非抑制性分类包含三个部分:数据预处理、DBN训练过程以及CYP2C9的分类过程。DBN的整个分类流程如图3所示。首先,用分子计算软件生成分子指纹特征并由SPSS19.0软件进行特征预处理。然后,根据2.2节介绍的算法训练DBN模型。它包括两个阶段:一是基于无标签数据训练多层RBM;二是采用BP网络微调整个DBN模型的参数。最后使用测试集评估模型的性能。

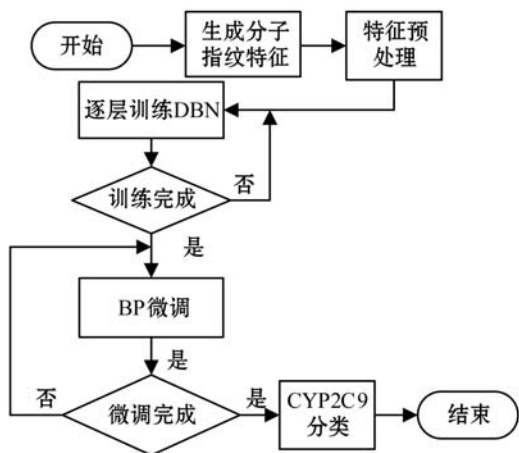


图3 DBN模型对CYP2C9的分类流程

3 实验与分析

3.1 模型评估标准

对二分类模型,有很多公认指标判定模型的性能。本文采用特异性(式(6))、敏感度(式(7)),总的分类准确率(式(8))以及马修斯相关系数(式(9))作为模型的评估标准。

$$SP = TN / (TN + FP) \times 100\% \quad (6)$$

$$SE = TP / (TP + FN) \times 100\% \quad (7)$$

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \times 100\% \quad (8)$$

$$MCC = \frac{TP \times TN - FP \times FN}{\sqrt{(TP + FP)(TP + FN)(TN + FP)(TN + FN)}} \quad (9)$$

其中,MCC常用于二分类检测,为验证测试实验结果是否平衡,其值在-1~1之间,越接近1模型评价效果越好。这些评估标准通过统计TP、FN、FP和TN的个数计算得出。TP表示真实值为抑制性,预测结果也是抑制性;FN表示真实值为抑制性,预测结果是非抑制性;FP表示真实值为非抑制性,预测为抑制性;TN表示真实值为非抑制性,预测也为非抑制性。

3.2 DBN、SVM和ANN参数信息

本文模型运行在Windows7系统上,使用MATLAB完成仿真实验。计算机的配置为: Intel i3处理器,4GB内存,主频率为2.4GHz。为得到模型最优分类性能,采用不同参数组合做了大量实验。表2列出了DBN和BP模型的详细初始化参数信息。

表2 DBN和BP的初始化参数信息

DBN初始化参数	
hiddensize	[655 300 50]
numepochs	200
momentum	0.9
alpha	0.1
batchsize	80
BP初始化参数	
numepochs	300
batchsize	80

表2中,hiddensize表示隐藏层神经元个数;numepochs表示训练迭代次数;momentum表示RBM初始化动量;alpha表示模型训练过程中初始化学学习率;batchsize表示每一次训练批量处理样本个数。对于SVM模型,采用LIBSVM(3.2版本,网址:<http://www.csie.ntu.edu.tw/~cjlin/libsvm>)实现仿真。为获取SVM最优分类性能,内核函数和代价因子的选择非常重要。可选内核参数有: linear、polynomial、RBF和sigmoid function。因RBF参数具有高效性和较低复杂性,选择RBF作为SVM内核参数,代价因子为5。对ANN模型,采用和BP相同参数。

3.3 DBN层数对分类结果的影响

在深度学习模型中,选择合适的DBN网络深度对CYP2C9的抑制性分类精度有一定影响。我们尝试了不同DBN网络结构(RBM的层数从1层到5层)。实验结果如表3所示。

表3 不同DBN层数在测试集上的准确率

DBN层数	抑制性			非抑制性			Q/%
	TP	FN	SE/%	TN	FP	SP/%	
RBM ¹	1 438	277	83.8	550	335	62.1	76.5
RBM ²	1 415	300	82.5	596	289	67.3	77.3
RBM ³	1 509	206	86.9	586	299	66.2	80.6
RBM ⁴	1 480	235	86.3	564	321	63.7	78.6
RBM ⁵	1 433	282	83.6	570	315	64.4	77.8

从表3可以看出,随着DBN模型层数增加,当模

型的层数从 1 层到 3 层时,模型总的分类准确率有所提升(准确率从 76.5% 增加到 80.6%)。当模型的层数从 3 层到 5 层时,模型总的分类准确率均有不同程度的下降。而且,模型层数的增加会使得训练过程更加复杂、计算时间也随之增加。因此,在后续的实验过程中,经过多方面的考虑,我们设置 DBN 的隐藏层层数为 3。

3.4 PubChem 和 MACCS 描述符对实验的影响

选择不同的分子指纹作为模型输入对 CYP2C9 的抑制性分类精度有一定影响。本文实验采用仅使用 PubChem 特征,仅使用 MACCS 特征,以及两者的特征组合分别作为模型的输入验证不同特征组合下模型的性能。实验结果如图 4 所示。

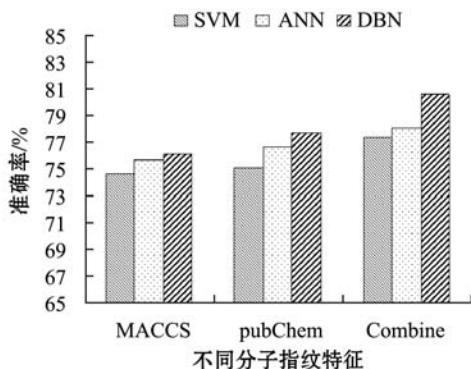


图 4 不同特征下的分类准确率

从图 4 中可以看出, PubChem 和 MACCS 分子指纹特征组合作为模型的输入(模型分别为 SVM、ANN 和 DBN),三个模型都获得了最好的分类性能。它们各自的分类精度分别为 78.3%、78.0%、80.6%,高于仅使用 MACCS 特征所得到的准确率:74.6%、75.7%、76.1%,以及仅使用 PubChem 特征所得到的分类准确率:75.1%、76.6%、77.7%。实验结果表明: PubChem 和 MACCS 特征组合给模型带来了新的信息,并且增加了模型的性能。同时将 MACCS 特征加入模型中提升了模型的性能,这一结果说明 MACCS 特征信息对 CYP2C9 的抑制性分类有积极影响。Michielan 等^[3]也得到了类似的结论:分子指纹特征对构建 CYP2C9 的抑制性分类模型具有重要贡献。因此实验证明了将 PubChem 和 MACCS 分子特征组合作为模型的输入可以进一步提升其分类准确率。

3.5 模型对比试验

为验证 DBN 模型对 CYP1A2 抑制性和非抑制性分类的有效性,基于相同的数据集和特征,将它与 ANN 和 SVM 模型进行了比较。以准确率和马修斯相关系数(MCC)作为衡量标准,结果分别如图 5 和图 6 所示。

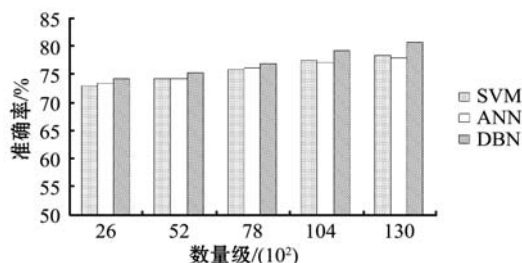


图 5 DBN、ANN 和 SVM 的分类准确率

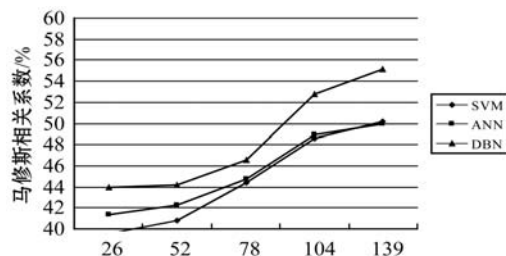


图 6 DBN、ANN 和 SVM 的 MCC 系数

从图 5 和图 6 能够得出,随着数据量的增加,DBN、SVM 以及 ANN 模型的性能均有所提升。实验结果表明,模型在丰富和大量的样本条件下能够学习更加多样的特征,进而提升了分类准确率。当数据量增加到 13 000 条时,相较于 SVM 和 ANN,DBN 模型获得了最好的分类准确率。原因是:不同于 ANN,DBN 模型避免了权重值随机分配,采用无监督预训练学习过程,可以提供一个更合适的初始值,从特征中抽取出多级的分布式表示,可以更好地挖掘分子结构的规律性。同时,DBN 模型是一个深层网络结构,当问题规模变得更加复杂时,它能克服一些浅层神经网络相对较弱的泛化能力及容易陷入过拟合的问题。

4 结 语

本文基于一个相对较大且结构多样的数据集,采用深度信念网络探讨了分子结构与区分 CYP2C9 的抑制性判别关系,验证了不同分子指纹特征对模型分类结果的影响。同时与 ANN 和 SVM 进行比较,验证了 DBN 模型对 CYP2C9 抑制性分类的有效性。因此,本文的研究有助于在药物研发阶段对 CYP1A2 的抑制性进行快速评估,对药物筛选具有一定的指导作用。

参 考 文 献

- [1] Klon A E. Machine learning algorithms for the prediction of hERG and CYP450 binding in drug development [J]. Expert Opinion on Drug Metabolism & Toxicology, 2010, 6(7): 821-833.
- [2] Graaf C D, Vermeulen N P E, Feenstra K A. Cytochrome P450 in silico: an integrative modeling approach [J]. Journal of Medicinal Chemistry, 2005, 48(8): 2725-2755.

线性插值计算。实验结果表明:AKAZE-LATCH 算法比原 AKAZE 算法匹配正确率提升 10%,运行速度提升 7%;KAZE-LATCH 算法比原 KAZE 算法匹配正确率提高 15%,匹配速度仅为 KAZE 算法的 75%,可广泛应用于对匹配速度和精度要求较高的场景。

参 考 文 献

- [1] Lowe D G. Distinctive image features from scale-invariant keypoints [J]. *International Journal of Computer Vision*, 2004, 60(2): 91-110.
- [2] Bay H, Ess A, Tuytelaars T, et al. SURF: speeded up robust features [J]. *Computer Vision and Image Understanding*, 2008, 110(3): 346-359.
- [3] Morel J M, Yu G. ASIFT: A New Framework for Fully Affine Invariant Image Comparison [J]. *Siam Journal on Imaging Sciences*, 2010, 2(2): 438-469.
- [4] Codruta O A, Cosmin A, Philippe B. Decolorizing images for robust matching [J]. *International Conference on Image Processing*, 2010, 9(2): 26-29.
- [5] Leutenegger S, Chli M, Siegwart R. BRISK: binary robust invariant scalable keypoints [C]//*Proceedings of the IEEE International Conference on Computer Vision (ICCV)*. Barcelona, Spain: IEEE, 2011: 2548-2555.
- [6] Rublr E, Rabaud V, Konolige K, et al. ORB: an efficient alternative to SIFT or SURF [C]//*Proceeding of International Conference on Computer Vision (ICCV)*, Barcelona, Spain: IEEE 2011: 2564-2571.
- [7] Alahi A, Ortiz R, Vandergheynst P. FREAK: Fast retina keypoint [C]//*Computer Vision and Pattern Recognition*. Providence USA: IEEE, 2011: 510-517.
- [8] Alcantarilla P F, Nuevo J, Bartoli A. Fast explicit diffusion for accelerated features in nonlinear scale spaces [C]//*Proceedings of British Machine Vision Conference*. Berlin, Germany: Springer, 2013: 131-142.
- [9] Grewenig S, Weickert J, Bruhn A. From box filtering to fast explicit diffusion [C]//*Proceedings of the 32nd DAGM conference on Pattern recognition*. Springer-Verlag, 2010: 533-542.
- [10] Levi G, Hassner T. LATCH: Learned arrangements of three patch codes [C]//*IEEE Winter Conference on Applications of Computer Vision*. IEEE Computer Society, 2016: 1-9.
- [11] Ouyang P, Yin S, Liu L, et al. A fast extreme illumination robust feature in affine space [C]//*Proceedings of the 2014 22nd International Conference on Pattern Recognition*. IEEE, 2014: 2365-2370.
- [12] Weickert J, Romeny B M T H, Viergever M A. Efficient and reliable schemes for nonlinear diffusion filtering [J]. *IEEE Trans Image Process*, 1998, 7(3): 398-410.
- [13] Lenc K, Gulshan V, Vedaldi A. VLBenchmarks [OL]. 2018. <http://www.vlfeat.org/benchmarks/>.
- [14] Yang X, Cheng K T. LDB: An ultra-fast feature for scalable Augmented Reality on mobile devices [C]//*2012 IEEE International Symposium on Mixed and Augmented Reality (ISMAR)*. IEEE Computer Society, 2012.
- [15] Yang X, Cheng K T. Local difference binary for ultrafast and distinctive feature description [J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2014, 36(1): 188-194.
- [16] Calonder M, Lepetit V, Strecha C, et al. BRIEF: computing a local binary descriptor very fast [J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2012, 34(7): 1281-1298.
- [17] Revaud J, Weinzaepfel P, Harchaoui Z, et al. EpicFlow: edge-preserving interpolation of correspondences for optical flow [C]//*2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. IEEE, 2015: 1164-1172.

(上接第 193 页)

- [3] Michielan L, Terfloth L, Gasteiger J, et al. Comparison of multilabel and single-label classification applied to the prediction of the isoform specificity of cytochrome P450 substrates [J]. *Journal of Chemical Information and Modeling*, 2009, 49(11): 2588-2605.
- [4] Byvatov E, Baringhaus K, Schneider G, et al. A virtual screening filter for identification of cytochrome P450 2C9 (CYP2C9) inhibitors [J]. *QSAR & Combinatorial Science*, 2007, 26(5): 618-628.
- [5] Cheng F, Yu Y, Shen J, et al. Classification of cytochrome P450 inhibitors and noninhibitors using combined classifiers [J]. *Journal of Chemical Information and Modeling*, 2011, 51(5): 996-1011.
- [6] 李兰婷, 李燕, 王永华, 等. 基于分子参数的 CYP2C9 抑制剂的预测模型 [J]. *计算机与应用化学*, 2007, 24(7): 939-943.
- [7] Novotarskyi S, Sushko I, Körner R, et al. A comparison of different QSAR approaches to modeling CYP450 1A2 inhibition [J]. *Journal of Chemical Information and Modeling*, 2011, 51(6): 1271-1280.
- [8] 闵建亮. 基于 2D 分子指纹和非平衡数据集的药物与受体交互作用预测研究 [D]. 景德镇: 景德镇陶瓷大学, 2014.
- [9] Yap C W. PaDEL-descriptor: an open source software to calculate molecular descriptors and fingerprints [J]. *Journal of Computational Chemistry*, 2011, 32(7): 1466-1474.
- [10] 晁丽. 细胞色素 P450 抑制剂虚拟筛选与分子对接 [D]. 重庆: 重庆大学, 2014.
- [11] Hinton G E, Salakhutdinov R R. Reducing the dimensionality of data with neural networks [J]. *Science*, 2006, 313(5786): 504-507.
- [12] Hinton G E, Sejnowski T J. Learning and relearning in Boltzmann machines [C]//*Parallel Distributed Processing: Explorations in the microstructure of Cognition*. MIT Press, 1986: 282-317.