

研究报告

基于XGBoost方法的常见人体农兽药及化学污染物暴露与糖尿病的相关性研究

卢宇红¹,李孜孜²,刘芝霖¹,苏畅²,王惠君²,张兵²,侯艳³

(1. 北京大学公共卫生学院生物统计系,北京 100041;2. 中国疾病预防控制中心营养与健康所/国家卫生健康委微量元素与营养重点实验室,北京 100050;3. 北京大学公共卫生学院生物统计系/北京大学临床研究所,北京 100041)

摘要:目的 基于Lasso变量筛选和XGBoost模型探讨人体农兽药及化学污染物暴露水平与糖尿病患病风险相关性。方法 2018—2019年,在中国石家庄和杭州进行的“降低成年超重者营养相关慢性病风险的适宜身体活动量研究”调查,选取86例糖尿病患者与410例非糖尿病患者样本并提取问卷调查中个人基本情况数据、体格测量、血生化数据和血清农兽药及化学污染物暴露浓度数据。采用Lasso筛选变量,再分别引入逻辑回归模型与XGBoost模型进行拟合,比较AUC评估拟合效果,并进行变量重要性排序。结果 Lasso筛选出2-乙基己基二苯基磷酸酯、全氟辛酸与全氟十一烷酸3种农兽药及化学污染物暴露与人群糖尿病患病有关,重要性排序为全氟辛酸>2-乙基己基二苯基磷酸酯>全氟十一烷酸,XGBoost模型(AUC=0.83)显著优于逻辑回归模型(AUC=0.64)($P<0.05$)。结论 Lasso适用于糖尿病患病影响因素的筛选,且XGBoost模型具有较好的复杂数据拟合能力。2-乙基己基二苯基磷酸酯、全氟辛酸与全氟十一烷酸是人群糖尿病患病的重要影响因素。

关键词:机器学习;糖尿病;农兽药残留;化学污染物暴露;XGBoost

中图分类号:R155 文献标识码:A 文章编号:1004-8456(2023)05-0652-06

DOI:10.13590/j.cjfh.2023.05.003

Association analysis between common pesticide and veterinary drug exposure in humans and diabetes mellitus based on XGBoostLU Yuhong¹, LI Zizi², LIU Zhilin¹, SU Chang², WANG Huijun², ZHANG Bing², HOU Yan³

(1. Department of Biostatistics, Peking University, Beijing 100041, China; 2. National Institute for Nutrition and Health, Chinese Center for Disease Control and Prevention, Key Laboratory of Trace Element Nutrition of National Health Commission, Beijing 100050, China; 3. Department of Biostatistics, Peking University, Peking University Clinical Research Center, Beijing 100041, China)

Abstract: Objective To explore the association between pesticide and veterinary drug exposure and the incidence of diabetes mellitus based on Lasso feature selection and the XGBoost model. **Methods** A cross-sectional study was conducted in Shijiazhuang and Hangzhou, China, between 2018 and 2019, enrolling 80 participants with diabetes and 410 healthy controls. The basic personal information, physical measurements, blood biochemical data, and serum exposure concentration data concerning agricultural or veterinary drugs and chemical pollutants were extracted using a questionnaire. Lasso was used for screening variables. Logical regression and XGBoost models were introduced for data fitting. The area under the curve (AUC) was compared to evaluate the fitting effect, and the variables were ranked by importance. **Results** Three features from pesticide and veterinary drug exposure, namely PFOA, PFUDa and EHDPP, were selected as related to diabetes, ranking PFOA> EHDPP> PFUDa. The XGBoost model (AUC = 0.83) performed significantly better than the logistic regression model (AUC = 0.64) in this dataset ($P<0.05$). **Conclusion** Lasso is suitable for screening the factors influencing diabetes. The XGBoost model still had a strong ability to fit complex relationships between various influencing factors. PFOA, PFUDa and EHDPP are significant risk factors for diabetes.

收稿日期:2022-05-16

基金项目:国家重点研发计划(2019YFC1605100);国家自然科学基金(81573155,82173615)

作者简介:卢宇红 女 硕士研究生 研究方向为生物统计 E-mail:2719836467@qq.com

通信作者:苏畅 男 研究员 研究方向为营养与食品卫生 E-mail:suchang@nih.chinacdc.cn

Key words: Machine learning; diabetes; pesticide residues; veterinary drug exposure; XGBoost

糖尿病是人群胰岛 B 细胞分泌功能失调引起的胰岛素分泌减少,进而导致血糖升高及一系列脂肪、蛋白质、碳水化合物的代谢紊乱症状,并最终造成慢性健康损伤^[1]。2017 年国际糖尿病联盟的统计数据显示,我国糖尿病患病率逐年升高,目前患病人数达到 1.14 亿人,高居世界第一,由此也带来了沉重的社会经济负担^[2]。糖尿病的影响因素除了遗传、肥胖、高血压、生活方式、年龄等常见因素以外,人体中的农兽药及化学污染物暴露可能也会影响血糖的代谢^[3]。我国作为农兽药生产与使用大国,虽然国家出台相关使用标准予以限制,但人体长期暴露于低毒混合农兽药污染环境仍可能对身体健康造成损害^[4]。目前农兽药与人群健康的相关研究多聚焦于农兽药的急性毒性方面,其与糖尿病等慢性病的相关研究较为缺乏,同时在农兽药的生产使用中伴随产生的其他化学污染物暴露也值得关注。

近年来,在卫生健康领域使用机器学习方法开展数据挖掘等工作中的应用已经越来越普及,相比于传统的统计方法,机器学习在如何更好地对非线性数据关系的变量进行拟合优化方面具有独特的优势,也已成为大数据时代的重要统计分析工具。本研究采用 Lasso (Least Absolute Shrinkage Selection Operator) 进行变量筛选,并使用逻辑回归模型与机器学习中 XGBoost (eXtreme Gradient Boosting) 模型进行建模与预测,分析农兽药及化学污染物暴露与糖尿病之间的关联,为流行病学中相关问题提供参考。

1 材料与方法

1.1 数据来源

本研究来自 2018—2019 年中国疾病预防控制中心营养与健康所、河北医科大学及浙江省疾病预防控制中心协作开展的“降低成年超重者营养相关慢性病风险的适宜身体活动量研究”的调查数据。该调查采取问卷调查与实验室检查相结合的方式,在中国的南北方各选取了一个城市(石家庄与杭州),并在每个城市选择两个社区进行调查,共确定了 496 名 40~66 岁的成年居民作为研究对象。该项目已获得中国疾病预防控制中心营养与健康研究所的伦理批准(编号 2019-023),研究前每位参与者都签署了知情同意书。

1.2 调查方式与内容

所有调查问卷都由工作人员现场进行询问、讲

解并完成填写。问卷调查表包括受试者性别、年龄、文化程度等基本情况信息以及吸烟、饮酒等生活方式信息。体格测量由工作人员为所有受试者进行测量,其中身高测量采用 SECA206 量高卷尺,体质量测量采用 SECA877 体重秤。身体质量指数 (Body mass index, BMI) = 体质量 / 身高² (kg/m²)。血糖采用己糖激酶法测定、血清甘油三酯采用酶测定,血清总胆固醇采用胆固醇氧化酶法测定,胰岛素采用电化学方法测定。血清中各种农兽药及化学污染物暴露浓度使用高效液相色谱-质谱联用方法测定。

1.3 判定标准

BMI 在 18.5~24 kg/m² 为正常,24 kg/m² 以上为超重或肥胖。糖尿病通过问卷结果和空腹血糖值综合评定,根据《中国 2 型糖尿病防治指南(2020 年版)》将血糖值 ≥ 6.1 mmol/L 或问卷回答为糖尿病的参与者纳入患者组^[5]。按照《中国成人血脂异常防治指南(2016 年修订版)》的标准,甘油三酯 ≥ 2.26 mmol/L 为偏高,总胆固醇 ≥ 5.2 mmol/L 为偏高^[6]。血清胰岛素 > 15 mU/L 判断为偏高^[7]。对农兽药及化学污染物暴露按检出率进行了整理,将检出率大于 30% 的农兽药及化学污染物暴露纳入分析^[8]。

1.4 质量控制

所有调查人员都必须参与培训,通过考核的调查员才能参与调查和数据收集工作。研究设备由国家项目组负责采购和配送,采用统一的调查问卷进行询问。调查过程中由专项负责人在现场进行抽查,核对数据并检查问卷的错误及遗漏,协同调查人员完成工作。农兽药浓度测定时加入标准化血清样本作质量控制。

1.5 统计学分析

用均值法插补缺失值,有序分类变量按顺序赋值。对于基线变量,其中性别、BMI、吸烟、饮酒、甘油三酯等二分类选项赋予 0 与 1 的值,受教育程度和生活质量自评,分别赋予 0~2 和 0~3 的值,对于年龄直接纳入数字进行分析。对于农兽药及化学污染物暴露变量,将低于检出限的样本分为低于检出限组,检出限以上再按中位数分为低浓度组和高浓度组。将浓度进行对数转化,以进一步排除异常值的影响,并直接使用数值进行下一步分析。在 Lasso 变量筛选后,使用传统的逻辑回归模型分别对基线变量与所有变量进行拟合,强调农兽药及化学污染物暴露在预测中的作用,再使用 XGBoost 模型对所有变量进行拟合。为避免对数据的过拟合,

将所有样本按 7:3 划分为训练集和测试集,在训练数据上做五折交叉验证。测试集上的 ROC 曲线下面积 (Area under curve, AUC) 用于评定效果。对 XGBoost 模型中变量 (机器学习中称为特征) 按重要性进行排序,分析农兽药及化学污染物暴露与基线变量对糖尿病预测准确性的影响程度。本研究采用 python 3.9 作为分析软件,使用 pandas 模块进行预处理,使用 sklearn 模块进行模型的构建与拟合,使用 GridSearchCV 模块对模型参数自动进行调整,提高模型性能。检验水平 $\alpha=0.05$, 双侧 $P<0.05$ 表示差异具有统计学意义。

2 结果

2.1 基本情况

调查总共获得 496 份样本,其中 86 人患有糖尿病,定义为糖尿病组,余 410 人为非糖尿病组。调查对象基本情况见表 1。糖尿病组中男性占比为 60.5% (52/86),非糖尿病组中男性占比为 43.2% (177/410)。样本平均年龄为 50.6±5.7 岁,糖尿病组平均年龄为 51.6±5.6 岁,略高于非糖尿病组的 50.4±5.8 岁。本研究共对 106 种农兽药及化学污染物暴露进行了检测,其中有 11 种农兽药及化学污染物暴露检出率大于 30%,并纳入后续分析,其包括:咪鲜胺 (Prochloraz)、多西环素 (Doxycycline)、呋喃它酮 (Furaltadone)、4-氯苯氧乙酸 (4-Chlorophenoxyacetate, 4-CPA)、2-乙基己基二苯基磷酸酯 (2-Ethylhexyl diphenyl phosphate, EHDPP)、全氟辛酸 (Perfluorooctanoic acid, PFOA)、全氟壬酸 (Perfluorononanoic acid, PFNA)、全氟癸酸 (Perfluorodecanoic acid, PFDA)、全氟己烷磺酸 (Perfluorohexane sulfonate, PFHxS)、全氟辛烷磺酸 (Perfluorooctane sulfonate, PFOS)、全氟十一烷酸 (Perfluoroundecanoic acid, PFuDA),农兽药及化学污染物暴露的分布情况见表 2,其中 PFHxS 和 PFOS 在所有参与者体内均被检测到。

2.2 Lasso 变量筛选

Lasso 通过在模型损失函数中添加惩罚项将不够重要的变量系数收缩为零达到筛选变量的目标^[9],增大惩罚项系数 (以下简称惩罚值)可以减少选择的变量个数 (图 1),筛选后的变量能提供更高的分类准确性^[10]。GridSearchCV 模块可以根据交叉验证的平均 AUC 自动挑选最优惩罚值,结果显示惩罚值为 0.07 时 AUC 最高。Lasso 筛选出的非农兽药及化学污染物暴露变量为受教育程度、甘油三酯、胰岛素和生活质量自评;农兽药残留变量分别为 PFOA、PFuDA 和 EHDPP。

表 1 调查对象基本情况

Table 1 Basic information of participants

变量	糖尿病组 (n=86)	非糖尿病组 (n=410)	合计(n=496)
性别			
女	34(39.5)	233(56.8)	267(53.8)
男	52(60.5)	177(43.2)	229(46.2)
年龄/岁	51.6±5.6	50.4±5.8	50.6±5.7
BMI/(kg/m ²)			
正常	10(11.6)	46(11.2)	56(11.3)
超重/肥胖	76(88.4)	364(88.8)	440(88.7)
吸烟			
是	36(41.9)	108(26.3)	144(29.0)
否	50(58.1)	302(73.7)	352(71.0)
饮酒			
是	62(72.1)	253(61.7)	315(63.5)
否	24(27.9)	157(38.3)	181(36.5)
甘油三酯/(mmol/L)			
正常	47(54.7)	302(73.7)	349(70.4)
偏高	39(45.3)	108(26.3)	147(29.6)
胰岛素/(mU/L)			
正常	56(65.1)	323(78.8)	379(76.4)
偏高	30(34.9)	87(21.2)	117(23.6)
总胆固醇/(mmol/L)			
正常	48(55.8)	263(64.1)	311(62.7)
偏高	38(44.2)	147(35.9)	185(37.3)
受教育程度			
初中及以下	20(23.3)	79(19.3)	99(20.0)
高中及中专	39(45.3)	127(31.0)	166(33.5)
大学及以上	27(31.4)	204(49.8)	231(46.6)
生活质量自评			
较差	17(19.8)	109(26.6)	126(25.4)
一般	48(55.8)	187(45.6)	235(47.4)
良好	19(22.1)	112(27.3)	131(26.4)
优越	2(2.3)	2(0.5)	4(0.8)

注:年龄使用平均数±标准差表示,其余变量使用频数(频率)表示

2.3 拟合结果

将 Lasso 筛选后的 7 个变量分类为基线变量与农兽药及化学污染物暴露变量,首先使用基线变量建立逻辑回归模型,测试集上的 AUC 为 0.64。接着使用所有变量建立逻辑回归模型,测试集上 AUC 为 0.78,预测效果上升,提示 3 种农兽药及化学污染物暴露在糖尿病的预测中起到明显作用。最后将所有变量导入 XGBoost 建模,在测试集上得到了 0.83 的 AUC,进一步提高了预测效果,表明 XGBoost 更加适合拟合农兽药及化学污染物暴露与糖尿病的关系。3 个模型 ROC 曲线的比较见图 2。XGBoost 的特征重要性排序见图 3,结果显示农兽药及化学污染物暴露中,全氟辛酸、2-乙基己基二苯基磷酸酯在糖尿病预测模型中有重要影响,对预测准确性的贡献甚至超过了某些基线变量。

3 讨论

Lasso 结果显示,PFOA、PFuDA 和 EHDPP 三种农兽药及化学污染物暴露与糖尿病相关性较高。PFOA、PFuDA 同属全氟烷基酸化合物(Per- and

表 2 11 种检出率高于 30% 的农兽药及化学污染物暴露分布情况[n(%)]

Table 2 Exposure distribution of 11 agricultural and veterinary drugs and chemical pollutants with detection rates higher than 30% [n(%)]

变量	糖尿病组 (n=86)	非糖尿病组 (n=410)	合计 (n=496)
咪鲜胺			
<LOD	50(58.1)	275(67.1)	325(65.5)
低浓度	17(19.8)	68(16.6)	85(17.1)
高浓度	19(22.1)	67(16.3)	86(17.3)
多西环素			
<LOD	42(48.8)	176(42.9)	218(44.0)
低浓度	22(25.6)	117(28.5)	139(28.0)
高浓度	22(25.6)	117(28.5)	139(28.0)
呋喃它酮			
<LOD	56(65.1)	209(51.0)	265(53.4)
低浓度	7(8.1)	108(26.3)	115(23.2)
高浓度	23(26.7)	93(22.7)	116(23.4)
4-氯苯氧乙酸			
<LOD	66(76.7)	231(56.3)	297(59.9)
低浓度	11(12.8)	88(21.5)	99(20.0)
高浓度	9(10.5)	91(22.2)	100(20.2)
2-乙基己基二苯基磷酸酯			
<LOD	43(50.0)	296(72.2)	339(68.3)
低浓度	18(20.9)	60(14.6)	78(15.7)
高浓度	25(29.1)	54(13.2)	79(15.9)
全氟辛酸			
<LOD	0(0.0)	2(0.5)	2(0.4)
低浓度	77(89.5)	170(41.5)	247(49.8)
高浓度	9(10.5)	238(58.0)	247(49.8)
全氟壬酸			
<LOD	41(47.7)	109(26.6)	150(30.2)
低浓度	38(44.2)	135(32.9)	173(34.9)
高浓度	7(8.1)	166(40.5)	173(34.9)
全氟癸酸			
<LOD	42(48.8)	103(25.1)	145(29.2)
低浓度	39(45.3)	136(33.2)	175(35.3)
高浓度	5(5.8)	171(41.7)	176(35.5)
全氟己烷磺酸			
<LOD	0(0.0)	0(0.0)	0(0.0)
低浓度	51(59.3)	197(48.0)	248(50.0)
高浓度	35(40.7)	213(52.0)	248(50.0)
全氟辛烷磺酸			
<LOD	0(0.0)	0(0.0)	0(0.0)
低浓度	52(60.5)	196(47.8)	248(50.0)
高浓度	34(39.5)	214(52.2)	248(50.0)
全氟十一烷酸			
<LOD	79(91.9)	217(52.9)	296(59.7)
低浓度	7(8.1)	93(22.7)	100(20.2)
高浓度	0(0.0)	100(24.4)	100(20.2)

注:<LOD 表示低于检测限;检出限以上按中位数分为低浓度组和高浓度组

polyfluoroalkyl substances, PFAS), 常用于烹饪设备、食品包装、防水防污织物处理以及农药加工过程中的添加剂^[11]。PFAS 自二战后就在全球大量使用, 不易降解的特性使其在环境中大量蓄积, 在饮用水^[12]及大部分普通人群的血清中^[13]都能检测到 PFAS。PFAS 也是本研究检出率和检出种类最多的一类物质, 反映了人类处于 PFAS 广泛暴露下的现状。流行病学研究显示 PFOA 等 PFAS 与糖尿病及相关代谢物存在关联, 如美国全国健康和营养检查调查的结果发现 PFOA 与糖尿病患病率之间存在很强的正相关^[14]。动物研究表明, PFAS 通过影响肝脏中的 PI3K-AKT 信号通路减少肝糖原合成, 抑制

糖异生并诱导胰岛素超敏反应, 破坏葡萄糖稳态^[15]。此外, 由于 PFAS 影响人体多种代谢调控机制, 其与高血压^[16]、高尿酸^[17]、高脂血症^[18]等慢性病同样存在相关性。用作阻燃剂和包装材料的 EHDPP, 虽然没有其与糖尿病相关的直接证据, 但在鸡胚胎肝细胞的研究中发现, EHDPP 会影响细胞葡萄糖代谢相关基因与通路, 使肝细胞的糖异生和糖原分解出现异常, 导致葡萄糖代谢失衡^[19]。农兽药本身及其生产使用中伴随产生的其他化学污染物暴露与糖尿病的关联及作用机制还需要更多的研究。

除了农兽药及化学污染物暴露的影响, 甘油三

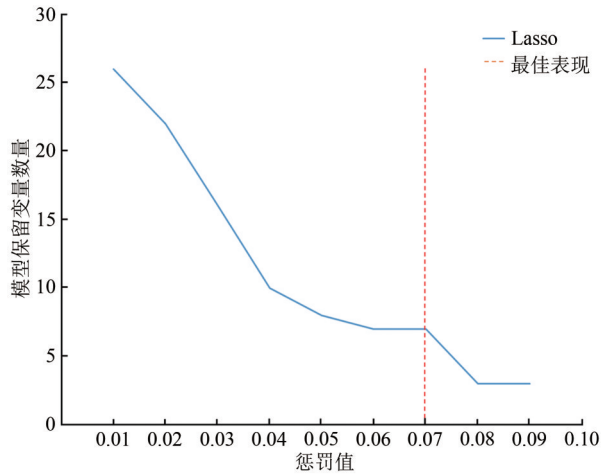


图1 模型保留变量数随惩罚值变化

Figure 1 The change of number of variables retained in model as the penalty

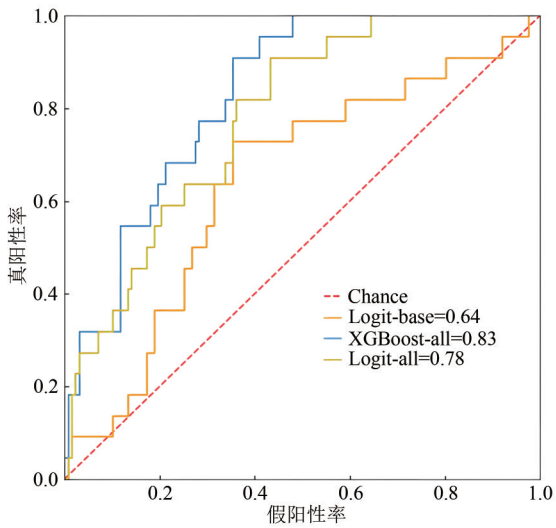


图2 基于基线变量逻辑回归模型与基于所有变量的逻辑回归模型及XGBoost模型ROC曲线比较

Figure 2 Comparison of ROC curve among logistic regression based on baseline variables, XGBoost and logistic regression based on all variables

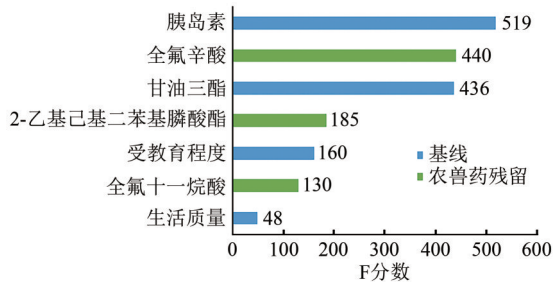


图3 XGBoost特征重要性排序

Figure 3 Ranking of feature importance in XGBoost model

酯、胰岛素水平、受教育程度和生活质量自评也与糖尿病正相关。研究表明,糖异生与脂质代谢相关联,糖尿病会引起肝的糖脂代谢改变,使血脂异常成为糖尿病的常见并发症^[20]。脂肪组织中甘油三酯异常积累,而异位脂肪存储也会促进糖尿病的发

展,两者相互影响^[21]。胰岛素分泌减少是2型糖尿病的直接原因,因此变量重要性排序显示胰岛素水平是最重要的变量。但胰岛素及其类似物也是糖尿病的治疗药物^[22],所以糖尿病患者体内可能存在外源胰岛素,使得胰岛素水平并不能完全预测糖尿病的发生。

受教育程度越高,了解糖尿病相关知识越多,有助于人们有意识地采取各种措施,预防糖尿病的发生。社会经济的发展提高了人们的生活质量,同时也使不少人养成了不健康的生活习惯,如过多的营养素摄入、较少的体力活动、更多的视屏时间,这些行为都是糖尿病的危险因素。此外,一些常见的影响因素如年龄、性别等可能是研究的潜在混杂因素,在纳入Lasso进行变量筛选后,发现其对于预测的重要性不如现存变量,去除这些混杂因素能够更清晰地分析农兽药及化学污染物暴露等变量与糖尿病的关联。

逻辑回归模型是解决二分类问题的经典广义线性模型,但农兽药及化学污染物暴露与糖尿病的关系并不是简单的线性关系,在变量数较多的情况下,要正确找出所有交互项加入模型也十分困难,导致逻辑回归模型难以达到较好的拟合效果。因此本研究使用了拟合能力更强的XGBoost模型,XGBoost作为梯度提升决策树(Gradient boosting decision tree,GBDT)的一种,通过优化可学习的弱分类器来提高模型的准确性,代价函数中加入的正则项一定程度上解决了GBDT的过拟合问题,使模型具有较好的泛化能力^[23],同时拥有更快的运行速度与更高的灵活性。对于拟合能力更好的神经网络,其往往需要大量的样本进行训练,而在中小型数据中XGBoost具有更好的效果,基于决策树的算法也比神经网络拥有更好的解释性,可以在拟合非线性复杂关系的同时对变量的重要性进行排序,这对于实施有效的干预与决策非常重要。此外,在对线性回归影响较大的异常值和噪声方面,XGBoost也具有较好的鲁棒性。本研究保留了传统的逻辑回归模型进行对比,AUC从0.64增加到0.78,证明农兽药及化学污染物暴露变量可以有效提升预测效果,而在拟合所有变量时AUC从0.78增加到0.83展现了XGBoost模型具有更强的区分能力,更适合应用于非线性复杂数据的建模问题。

本研究存在一定的局限性,研究样本仅来自于两个城市,而农兽药及化学污染物暴露受地域影响较大,仅采用两个城市的样本数据并不能很好地展示农兽药及化学污染物暴露的真实分布,增加取样地点有利于提高结果的准确度和泛化性。由于没

有足够证据表明 EHDPP 与糖尿病直接相关,我们应谨慎对待该结论,最终筛选出的农兽药及化学污染物暴露依然需要相关的研究进一步验证。鉴于 PFAS 的广泛暴露与健康危害,相关部门应考虑给予其足够的重视。本研究获得的模型 AUC 值较高,具备较好的区分能力,因此本研究能够为糖尿病的预防工作提供重要科学依据,为机器学习方法在流行病学研究中的应用提供参考,同时也为探索农兽药及化学污染物暴露对身体健康的影响提供了有效的科研方法与思路。

参考文献

- [1] MAYFIELD J. Diagnosis and classification of diabetes mellitus: New criteria [J]. *American Family Physician*, 1998, 58(6): 1355-1362, 1369-1370.
- [2] LUO Z, FABRE G, RODWIN V G. Meeting the challenge of diabetes in China[J]. *International Journal of Health Policy and Management*, 2020, 9(2): 47-52.
- [3] MA R C W. Epidemiology of diabetes and diabetic complications in China[J]. *Diabetologia*, 2018, 61(6): 1249-1260.
- [4] KIM K H, KABIR E, JAHAN S A. Exposure to pesticides and the associated human health effects [J]. *Science of the Total Environment*, 2017, 575: 525-535.
- [5] 中华医学会糖尿病学分会. 中国 2 型糖尿病防治指南(2020 年版)[J]. *国际内分泌代谢杂志*, 2021(5): 482-548. Chinese Diabetes Society. Guideline for the prevention and treatment of type 2 diabetes mellitus in China (2020 edition) [J]. *International Journal of Endocrinology and Metabolism*, 2021(5): 482-548.
- [6] 诸骏仁, 高润霖, 赵水平, 等. 中国成人血脂异常防治指南(2016 年修订版)[J]. *中国循环杂志*, 2016, 31(10): 937-953. ZHU J R, GAO R L, ZHAO S P, et al. 2016 Chinese guideline for the management of dyslipidemia in adults [J]. *Chinese Circulation Journal*, 2016, 31(10): 937-953.
- [7] 康怡, 张宝和, 徐洪涛, 等. 北京市部分中老年男性高胰岛素血症患者代谢紊乱及胰岛功能评估[J]. *首都医科大学学报*, 2016, 37(3): 385-390. KANG Y, ZHANG B H, XU H T, et al. Evaluation of metabolic disorders and pancreatic islet function in aged male patients with hyperinsulinemia in Beijing [J]. *Journal of Capital Medical University*, 2016, 37(3): 385-390.
- [8] YOU L, ZHENG F J, SU C, et al. Metabolome-wide association study of serum exogenous chemical residues in a cohort with 5 major chronic diseases [J]. *Environment International*, 2022, 158: 106919.
- [9] KANG C Z, HUO Y H, XIN L H, et al. Feature selection and tumor classification for microarray data using relaxed Lasso and generalized multi-class support vector machine [J]. *Journal of Theoretical Biology*, 2019, 463: 77-91.
- [10] KHAIRE U M, DHANALAKSHMI R. Stability of feature selection algorithm: A review [J]. *Journal of King Saud University- Computer and Information Sciences*, 2022, 34(4): 1060-1073.
- [11] MITRO S D, LIU J X, JAACKS L M, et al. Per- and polyfluoroalkyl substance plasma concentrations and metabolomic markers of type 2 diabetes in the Diabetes Prevention Program trial [J]. *International Journal of Hygiene and Environmental Health*, 2021, 232: 113680.
- [12] DOMINGO J L, NADAL M. Human exposure to per- and polyfluoroalkyl substances (PFAS) through drinking water: A review of the recent scientific literature [J]. *Environmental Research*, 2019, 177: 108648.
- [13] CRINNION W J. The CDC fourth national report on human exposure to environmental chemicals: what it tells us about our toxic burden and how it assist environmental medicine physicians [J]. *Alternative Medicine Review: a Journal of Clinical Therapeutic*, 2010, 15(2): 101-109.
- [14] HE X W, LIU Y X, XU B, et al. PFOA is associated with diabetes and metabolic alteration in US men: National health and nutrition examination survey 2003-2012[J]. *Science of the Total Environment*, 2018, 625: 566-574.
- [15] YAN S M, ZHANG H X, ZHENG F, et al. Perfluorooctanoic acid exposure for 28 days affects glucose homeostasis and induces insulin hypersensitivity in mice[J]. *Scientific Reports*, 2015, 5: 11029.
- [16] MIN J Y, LEE K J, PARK J B, et al. Perfluorooctanoic acid exposure is associated with elevated homocysteine and hypertension in US adults[J]. *Occupational and Environmental Medicine*, 2012, 69(9): 658-662.
- [17] KATARIA A, TRACHTMAN H, MALAGA-DIEGUEZ L, et al. Association between perfluoroalkyl acids and kidney function in a cross-sectional study of adolescents [J]. *Environmental Health: A Global Access Science Source*, 2015, 14: 89.
- [18] SPRATLEN M J, PERERA F P, LEDERMAN S A, et al. The association between perfluoroalkyl substances and lipids in cord blood [J]. *The Journal of Clinical Endocrinology & Metabolism*, 2019, 105(1): 43-54.
- [19] SHEN J Y, ZHANG Y Y, YU N Y, et al. Organophosphate ester, 2-ethylhexyl diphenyl phosphate (EHDPP), elicits cytotoxic and transcriptomic effects in chicken embryonic hepatocytes and its biotransformation profile compared to humans [J]. *Environmental Science & Technology*, 2019, 53(4): 2151-2160.
- [20] JIANG S Z, YOUNG J L, WANG K, et al. Diabetic-induced alterations in hepatic glucose and lipid metabolism: The role of type 1 and type 2 diabetes mellitus (Review) [J]. *Molecular Medicine Reports*, 2020, 22(2): 603-611.
- [21] SØNDERGAARD E, NIELSEN S. VLDL triglyceride accumulation in skeletal muscle and adipose tissue in type 2 diabetes [J]. *Current Opinion in Lipidology*, 2018, 29(1): 42-47.
- [22] GRECO D S, BROUSSARD J D, PETERSON M E. Insulin therapy [J]. *Veterinary Clinics of North America: Small Animal Practice*, 1995, 25(3): 677-689.
- [23] CHEN T Q, GUESTRIN C. XGBoost: A scalable tree boosting system [C]. *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. New York, NY, USA: ACM, 2016: 785-794.