

花生油中黄曲霉毒素 B1 含量的荧光光谱表征

刘翠玲^{1,2} 殷莺倩^{1,2} 张善哲^{1,2} 孙晓荣^{1,2} 李佳琮^{1,2} 吴静珠^{1,2}

(1. 北京工商大学人工智能学院, 北京 100048;
2. 北京工商大学北京市食品安全大数据技术重点实验室, 北京 100048)

摘要:花生油在生产过程中极易受到黄曲霉毒素 B1 (Aflatoxin B1, AFB1) 的污染。针对 AFB1 的传统检测方法操作繁琐、时效性差等问题,利用荧光光谱技术快速检测花生油中 AFB1 含量。首先通过三维荧光光谱确定最佳激发波长,使用 K-means 和自组织映射 (Self organizing map, SOM) 聚类算法对花生油中 AFB1 含量进行超标与否的定性鉴别,准确率均达 95% 以上;其次使用 2 种预处理算法和 2 种降维算法,选出竞争自适应重加权采样 (Competitive adaptive reweighted sampling, CARS) 为最佳的波长选择方法;随后将回声状态网络 (Echo state network, ESN) 用于 AFB1 的定量建模,同时与其他模型作比较,结果显示 CARS-ESN 模型获得了最佳 AFB1 含量预测效果;最后将麻雀搜索算法 (Sparrow search algorithm, SSA) 用于对 ESN 参数进行寻优,最终预测集决定系数达 0.984,均方根误差达 2.13 $\mu\text{g/kg}$ 。结果表明了荧光光谱技术结合 ESN 预测花生油中 AFB1 含量的可行性,为在线检测食用油中真菌毒素含量系统的开发提供理论依据。

关键词: AFB1 含量; 三维荧光光谱; 特征提取; 回声状态网络; 麻雀搜索算法

中图分类号: O657.3 文献标识码: A 文章编号: 1000-1298(2023)12-0389-08 OSID: 

Characterization of Aflatoxin B1 in Peanut Oil by Fluorescence Spectrum

LIU Cuiling^{1,2} YIN Yingqian^{1,2} ZHANG Shanzhe^{1,2} SUN Xiaorong^{1,2} LI Jiacong^{1,2} WU Jingzhu^{1,2}

(1. School of Artificial Intelligence, Beijing Technology and Business University, Beijing 100048, China
2. Beijing Key Laboratory of Big Data Technology for Food Safety, Beijing Technology and Business University, Beijing 100048, China)

Abstract: Peanut oil is susceptible to aflatoxin B1 (AFB1) contamination during the production process. In response to the problems of tedious operation and poor timeliness of the traditional detection method for AFB1, fluorescence spectroscopy was utilized for the rapid determination of AFB1 in peanut oil. Firstly, the optimal excitation wavelength was determined by three-dimensional fluorescence spectroscopy. K-means and self organizing map (SOM) clustering algorithm were used to qualitatively identify the AFB1 content in peanut oil with an accuracy of over 95%. Next, two preprocessing algorithms and two dimensionality reduction algorithms were used. Competitive adaptive reweighted sampling (CARS) was selected as the best wavelength selection method. The echo state network (ESN) was then used for quantitative modeling of AFB1. Compared with other models, the results showed that the CARS-ESN model obtained the best prediction of AFB1 content. Finally, the sparrow search algorithm (SSA) was used to find the optimal ESN parameters. The final test set coefficient of determination reached 0.984, with root mean square error of 2.13 $\mu\text{g/kg}$, demonstrating the feasibility of fluorescence spectroscopy technique combined with ESN to predict the AFB1 content in peanut oil. The research result can provide a theoretical basis for the development of an online system for the detection of fungal toxin content in edible oils.

Key words: content of AFB1; three-dimensional fluorescence spectroscopy; feature extraction; echo state network; sparrow search algorithm

0 引言

真菌毒素是真菌代谢产生的一种强致癌污染物,以黄曲霉毒素 B1 (Aflatoxin B1, AFB1) 毒性最强,而花生油特别容易受 AFB1 污染^[1]。因此,花生油中 AFB1 的定量检测非常重要,有必要开发一种快速、灵敏、简单和高效的检测技术,保障食品安全。目前,国内外常用的食用植物油中 AFB1 检测的方法包括酶联免疫吸附筛查法、高效液相色谱法等^[2-3]。这些传统检测方法操作过程繁琐,无法满足对食用植物油品质和安全进行实时监测的要求,因而迫切需要一种快速无损的检测方法。

随着科技的发展,无损检测技术因其检测简单快速等优点广泛应用于农产品安全检测领域。文献[4]利用近红外光谱技术结合支持向量机算法对稻谷中 AFB1 含量进行了预测,校正集决定系数达到 0.913,标准偏差和预测偏差分别为 1.186 和 1.267。文献[5]利用 BP-AdaBoost 方法建立表面增强拉曼散射光谱与 AFB1 浓度之间的定量关系模型,预测集相关系数达到 0.928 3。文献[6]使用拉曼光谱技术实现食用油样本中 AFB1 含量的检测分析,最终循环神经网络在预测集中的决定系数达到 0.95。然而使用近红外、拉曼等光谱技术检测黄曲霉毒素时,大多是通过检测农产品霉变后的挥发性气味、内部化学变化以及外部表现特性的变化,从而间接性地预测农产品毒素含量,并非是直接根据毒素的性质进行检测,无特异性。

荧光光谱技术具有分析灵敏度高、选择性强和使用简便的优点,在食品安全检测领域展现了巨大的发展潜力。其主要是针对毒素本身的结构,根据测量毒素荧光信号的强度从而判断毒素含量,具有特异性。而且通过激发光与发射光的结合,可以显著地减少背景信号的干扰,极大地提高检测系统的灵敏度。文献[7]采用荧光分光光度计对 AFB1 的标准溶液进行了检测分析,验证了荧光光谱技术分析 AFB1 含量的可行性。文献[8]结合荧光光谱技术,对白酒中的真菌毒素进行了检测研究,分析了各种滤波平滑处理的效果,决定系数均达 0.8 以上。然而,现有的荧光检测技术多集中于 AFB1 纯溶液以及其他水溶液,对食用油中真菌毒素含量的检测较少。将荧光技术与人工智能方法相结合进行建模分析,是当下荧光光谱分析的热点。回声状态网络(Echo state network, ESN)是一种递归神经网络,相比于传统的神经网络无论是建模方法还是学习过程都有很大的区别^[9]。文献[10]应用 ESN 对高光谱影像进行了分类。孙钰扬^[11]将 ESN 在公共建筑能

耗预测中进行应用。然而,现有研究中少有学者将 ESN 用于光谱检测领域。由于 ESN 具有非常出色的非线性映射能力,网络设计相对简单,在收敛速度和定量预测精度等方面有良好的表现,尤其适合解决各种预测问题^[12],故本文将 ESN 用于荧光光谱领域,对真菌毒素的含量进行预测。

本文基于 AFB1 在紫外光照射下可产生荧光信号特性,结合荧光光谱技术以花生油为研究对象检测其中的 AFB1 含量,实现花生油样品中 AFB1 超标与否的定性判别和 AFB1 污染程度的定量预测。结合不同的数据处理方法,将 ESN 用于 AFB1 含量的数据建模,并与反向传播神经网络(Back propagation neural network, BPNN)、径向基函数神经网络(Radial basis function neural network, RBFNN)和偏最小二乘回归(Partial least squares regression, PLSR)进行对比分析,最后用智能优化算法对最优组合模型的参数进行寻优。

1 材料与方法

1.1 实验样本

实验样本为含有 62、50、40、30、20、15、10、5、2、0 $\mu\text{g}/\text{kg}$ 的不同含量(质量比) AFB1 的花生油样本,由国家粮食和物资储备局科学研究院提供。该实验样本为市面上收集到的被污染 AFB1 的花生油,由国家粮食和物资储备局科学研究院根据国家标准 GB 5009.22—2016《食品安全国家标准 食品中黄曲霉毒素 B 族和 G 族的测定》中的同位素稀释液相色谱-串联质谱法对 AFB1 含量进行测定。

1.2 实验仪器及参数设置

实验仪器为爱丁堡 FS5 型荧光光谱仪,如图 1 所示。光源选用 150 W 无臭氧氙灯,发射检测器为单光子计数,选用 PMT-900 型光电倍增管,吸收检测器为紫外增强型硅光电二极管。采集样品三维荧光光谱,设置激发波长间隔为 10 nm,范围为 250 ~ 550 nm;发射波长间隔为 2 nm,范围为 260 ~ 750 nm。采集样品荧光发射光谱,选择 390 nm 为激发波长,设置步长间隔为 1 nm,采集范围为 400 ~ 800 nm,共采集 150 个花生油样品的荧光光谱。



图1 爱丁堡 FS5 型荧光光谱仪

Fig.1 Edinburgh FS5 fluorescence spectrometer

1.3 实验方法

1.3.1 样本划分

在建立花生油中 AFB1 的预测模型之前,需要将样本划分为校正集和预测集。本文采用基于联合 $X - Y$ 距离的样本划分法 (Sample set partitioning based on joint $X - Y$ distances, SPXY)。其中 X 代表多输入变量的光谱变量, Y 代表单输出变量的理化指标变量,通过综合考虑样本光谱数据和理化指标进而计算得到不同样本间距离^[13]。将样本按照 3:1 的比例划分校正集和预测集,用于构建 AFB1 含量预测模型的校正集和预测集样本信息,如表 1 所示。校正集 AFB1 含量涵盖了预测集的范围,两组数据集均含有 10 个不同 AFB1 含量的样本。

表 1 样本的数据集划分
Tab.1 Data set division of sample

数据集	数量	花生油中 AFB1 含量/($\mu\text{g}\cdot\text{kg}^{-1}$)		
		范围	平均值	标准差
校正集	112	0 ~ 62	25.63	20.60
预测集	38	0 ~ 62	18.40	18.40

1.3.2 光谱预处理

利用 FS5 型荧光光谱仪经过一系列调参后,发射光照射样品池中的花生油,油类物质内部的电子经过能级跃迁以及能量释放发射出荧光,从而获得所需要的荧光光谱数据信息^[14]。在荧光光谱数据中,除了含有被测样品待测成分信息外还包括各种仪器的噪声、杂散等无关信息,会干扰对物质的检测分析,应予以去除^[15]。本文对实验数据处理方法包括 S - G 卷积平滑 (Savitaky - Golay smoothing, S - G) 以及移动平均平滑 (Moving average smoothing, MAS)。S - G 卷积平滑主要原理是通过多项式对原始光谱数据进行平滑,进而达到能够最多地消除噪声信号,也能够保留光谱数据最有效的信息^[16]。MAS 是指选择一个合适的平滑窗口宽度 ($2x + 1$),用平滑窗口中心波长波点数的测量平均值来代替实际测量值,使整个波长得到平滑处理^[17]。

1.3.3 特征波长提取

竞争性自适应重加权算法 (Competitive adaptive reweighted sampling, CARS) 是一种结合蒙特卡洛采样与偏最小二乘 (Partial least squares, PLS) 模型回归系数的特征变量选择方法,模仿达尔文理论中的“适者生存”的原则^[18]。每次通过自适应加权采样保留 PLS 模型中回归系数绝对值权重较大的点作为新的子集,然后基于新的子集建立 PLS 模型,选择 PLS 模型交互验证均方根误差最小的子集中的波长作为特征波长^[19]。随机森林 (Random forest,

RF) 是一种基于多棵决策树,适用于分析复杂交互特征的算法^[20]。它的原理是将来自学习样本的几个 bootstrap 样本构建的许多二叉决策树结合起来,并在每个节点上随机选择解释变量的子集。先对随机森林中的特征变量按照重要性进行降序排列并确定删除比例,然后删除相应比例不重要的指标,即得到一个新的特征集。再用新的特征集建立新的随机森林并计算每个特征的重要性并排序,根据得到的每个特征集和建立的随机森林计算对应的袋外误差率,将袋外误差率最低的特征集作为最后选定的特征集^[21]。

1.3.4 聚类分析方法

K-means 聚类算法是一种典型的无监督学习算法。对于给定样本集,其根据样本间距离将样本集划分为 K 个簇,在簇内各样本的距离尽量小且簇间样本的距离尽可能大的约束条件下,迭代求解 K 个簇的簇心并把每个样本归于距离其最近的簇心^[22]。

自组织映射 (Self organizing map, SOM) 神经网络具有良好的自组织性和可视化等特性,输入层接收数据,竞争层对输入进行比较分类。该网络主要利用迭代方法计算各输入向量与竞争层处理单元间的连接权值向量,通过竞争学习算法来不断调整连接权重使其越来越接近原输入向量,直到输入向量与连接权值之间总距离为最小或最大学习循环时,停止训练。通过将高维数据映射到结构和相互关系简单的低维空间中进行展示,从而实现数据可视化、聚类、分类等功能^[23]。

1.3.5 定量分析方法

回声状态网络 (ESN) 是一种新型的递归神经网络^[24],是由输入层、隐含层 (储备池) 和输出层构成的 3 层网络结构,具体结构如图 2 所示。

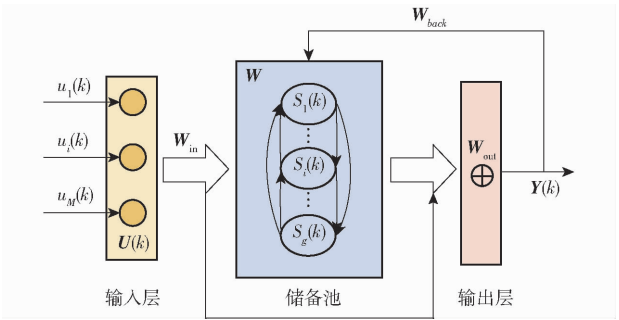


图 2 ESN 经典结构

Fig.2 Classical structure of ESN

设定网络具有 M 个输入节点, H 个隐含节点 (储备池), 且具有 L 个输出节点。在 k 时刻的输入 $U(k) = [u_1(k) \quad u_2(k) \quad \cdots \quad u_M(k)]^T$ 作用下,网络的状态方程为

$$\begin{cases} S(k) = f(WS(k-1) + W_{in}U(k) + W_{back}Y(k-1)) \\ Y(k) = f_{out}(W_{out}[S(k)^T U(k)^T] + B) \end{cases}$$

式中 $S(k)$ ——隐含节点的状态
 $Y(k)$ ——网络的输出
 W_{in} ——输入节点和隐含节点(储备池)的权连接矩阵
 W ——隐含节点自身的反馈连接矩阵
 W_{out} ——输入节点、隐含节点与输出节点的权连接矩阵
 W_{back} ——输出节点与隐含节点的权连接矩阵,体现输出对系统的影响
 f ——隐含节点的激活函数,常取双曲正切函数,以反映系统状态的非线性演化过程
 f_{out} ——输出函数,在时间序列的预测中,通常取恒等函数
 B ——输出激活函数的偏置

ESN 相比于传统的神经网络,无论是建模方法还是学习过程都有很大的区别。ESN 采用大量的稀疏、层内互连的隐含节点(储备池)来存储系统的状态,且系统的状态演化由输入、系统的历史状态和输出共同作用,以实现历史输入的记忆功能。其次,整个网络以储备池作为信息处理媒介,将输入信号从低维空间映射到高维空间,并且在高维空间中采用线性回归的形式对连接权重进行训练,这种方式在一定程度上避免了出现局部最小值的情况,有效解决了由于计算负荷过大导致网络输出失真的问题^[25]。

1.3.6 群体智能优化算法

麻雀搜索算法(Sparrow search algorithm, SSA)是2020年提出的一种新兴的元启发式算法,它同属于基于群体的社会化特征优化的群智能算法^[26]。该算法通过不断更新个体位置,模拟麻雀觅食和反捕食行为。相比传统算法,SSA的结构简单易于实现,且控制参数较少,局部搜索能力较强。该算法在单峰多峰等基准函数上的表现优于粒子群算法、蚁

群算法等传统算法^[27]。该算法通过优化 ESN 模型中的储备池规模、学习率和正则化系数,寻找其最优组合,以提高模型预测精度。

1.3.7 模型评价指标

本研究采用准确率作为聚类分析的评价指标,采用校正集决定系数 R_c^2 、预测集决定系数 R_p^2 、校正均方根误差(RMSEC)以及预测均方根误差(RMSEP)作为定量分析模型的评价指标。决定系数越接近1,表明荧光光谱信息与花生油中 AFB1 的相关性越好,均方根误差越小,表明预测中产生的误差越小,即光谱建模效果越好^[28]。

2 结果与分析

2.1 荧光光谱分析

为了获取 AFB1 污染的花生油最佳激发波长,提取 2 mL 含有 62 $\mu\text{g/kg}$ AFB1 的花生油样本,将其置于石英比色皿中并扫描其三维荧光光谱。图 3a 为含 62 $\mu\text{g/kg}$ AFB1 花生油的三维荧光光谱图和对应的等高线图。可看到花生油共有两个荧光特征峰,峰 1 的最佳激发波长在 390 nm 附近,发射波长在 470 nm 附近,这部分荧光响应代表不饱和脂肪酸物质及黄曲霉毒素;峰 2 的最佳激发波长在 410 nm,发射波长在 670 nm 附近,这部分代表色素类物质,如花生素等^[29]。所以,AFB1 污染的花生油的最佳激发波长在 390 nm 附近,最佳发射波长在 470 nm 附近。

配制含不同含量 0 ~ 62 $\mu\text{g/kg}$ 的 AFB1 的花生油样本,观察其荧光响应曲线与黄曲霉素的变化规律。以最佳激发波长 390 nm 作为激发波长,扫描其发射光谱。发射波长范围设置为 400 ~ 800 nm,共得到 401 维的光谱数据。将各含量下的数据取平均,绘制出 AFB1 含量变化对比曲线图,如图 3b 所示。可发现其荧光响应的整体变化趋势是随着花生油中 AFB1 含量的增加,对应的荧光强度增强。并且在 450 ~ 500 nm 与 650 ~ 700 nm 处有明显差异。

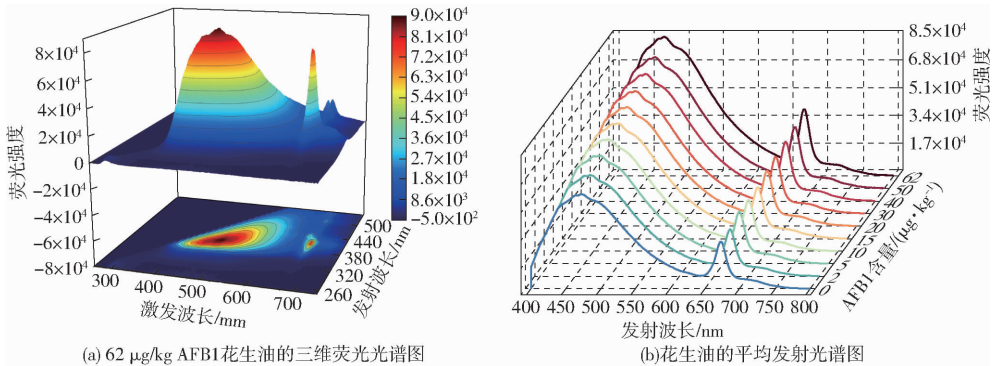


图 3 样品的荧光光谱图

Fig. 3 Fluorescence spectra of samples

通过数据分析发现样品的荧光光谱信息与 AFB1 含量存在规律性的变化,因此理论上可通过荧光光谱表征花生油中 AFB1 的含量。

2.2 AFB1 聚类结果

采用 K-means 聚类算法以及 SOM 聚类算法对 AFB1 污染的花生油进行无监督学习的聚类分析,识别结果分别如图 4 所示。设置目标类别数为 2,即根据国家标准 GB 2761—2017《食品安全国家标准 食品中真菌毒素限量》将花生油样本分为 AFB1 超标 ($>20\text{ }\mu\text{g/kg}$) 与不超标 ($\leq 20\text{ }\mu\text{g/kg}$) 两类。分别各选择一个样本作为每一类的初始聚类中心。对

比图 4a 和图 4b 可知,K-means 聚类后存在小样本的数据簇聚集,不同样本间存在交叉混合,部分数据可分性不明显。相比之下,SOM 的聚类可视化效果更好,光谱数据有其分别的数据簇,AFB1 超标与不超标的食用油的可分性比较明显。且从最终的判别结果来看,K-means 的准确率达 95.33%,SOM 的准确率达 96%。两种聚类算法的准确率均高于 95%,能够较为精准地判别 AFB1 污染花生油的程度,表明荧光强度与 AFB1 的相关性较强,该光谱对 AFB1 含量的检测敏感性较为敏感,且为后续定量分析奠定基础。

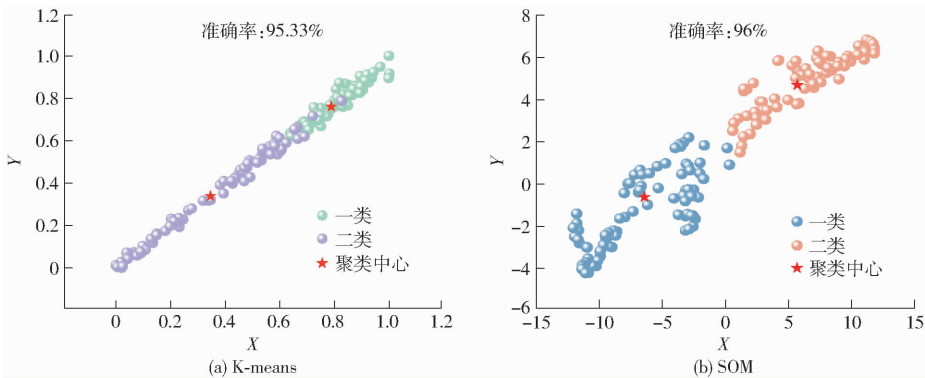


图 4 聚类结果
Fig. 4 Clustering results

2.3 AFB1 含量预测结果

2.3.1 特征降维结果

AFB1 污染的花生油样品光谱信息量大,通过特征降维算法可以减小数据维度来降低计算复杂度,同时保留数据的关键信息,以提高模型的准确性和效率。图 5a ~ 5c 为 CARS 算法的降维过程图,CARS 是一种基于回归系数的波长选择方法,以交叉验证均方差 (Root mean square error of cross

validation, RMSECV) 最小来确定波长变量的数目。图 5a 显示了随着采样次数增加,衰减指数 (EDF) 的作用使选择的波长数量快速下降。图 5b 反映了变量选择过程中 RMSECV 的变化,在开始迭代时,由于花生油数据中大量不相关的变量被消除,导致 RMSECV 从 1 次到 23 次时在逐渐减小。当采样次数到第 23 次后,由于部分重要信息被剔除,所以 RMSECV 又整体呈上升趋势。图 5c 显示了变量选

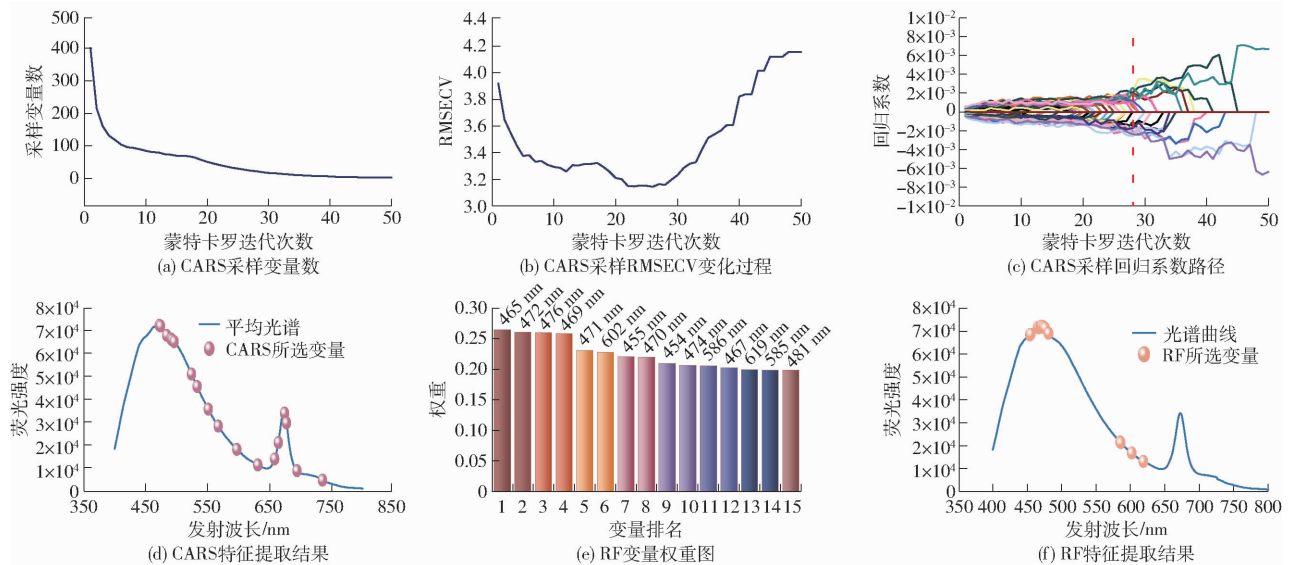


图 5 光谱特征降维结果

Fig. 5 Spectral feature dimensionality reduction results

择过程中各波长变量的回归系数的路径,当迭代次数为 23 时所选择的波长集合为最优集合,共提取了 19 个特征波长变量,特征波长提取结果如图 5d 所示。采用随机森林算法对变量进行重要性程度排序,设置决策树的数目为 200,被选择的特征阈值为 0.037。最终选取前 15 个特征变量,变量重要性降序排序结果如图 5e 所示,提取结果分布如图 5f 所示。可以看出,所提取的特征波长大多位于 450 ~ 500 nm 之间,说明该波段内特征变量的重要性较高,能更好地反映花生油中 AFB1 含量的变化特性。

2.3.2 建模结果

对含 AFB1 的花生油样本的光谱数据采用 S - G 卷积平滑和 MAS 法进行预处理,然后分别用 CARS 和 RF 算法进行特征提取,最后将提取的 19 维和 15 维数据用于建立 ESN 定量分析模型。为了验证 ESN 模型的整体性能优于传统的前馈神经网络和经典的线性机器学习模型,本研究同时引入 BPNN 和 RBFNN 以及 PLSR 进行训练,将计算结果与 ESN 进行对比分析。ESN 的参数包括储备池规模(hidden)、更新速率(Ir)以及正则化系数(reg)。首先确定储备池的规模即神经元的个数,数目越多其拟合能力越强。由于 ESN 仅通过调整输出权值来线性拟合输出结果,所以一般 ESN 需要远大于常规神经网络的节点规模,故本实验参数设置 hidden 为 300,Ir 为 0.05,reg 为 1。BPNN 采用 tansig 作为激活函数,通过设置隐含层神经元的个数 epochs 以及学习率 Ir 来调节模型的精度^[30];RBFNN 通过设置模型中径向基的扩展速度 spread 来调节模型的精度^[31];PLSR 通

过调节主成分数来改变建模分析的变量数,其最优取值通过二十折交叉验证而获得^[32]。

表 2 为不同模型的建模结果对比。从模型来看,非线性模型的预测精度均优于 PLSR,表明该荧光光谱数据较复杂,用一般的线性关系表征结果不够理想。而 ESN 模型计算的 R_p^2 最高,且 RMSEP 最低,表明其预测能力最好。证实了 ESN 模型的结构与其他神经网络相比空间复杂度更低,计算量相对较小而且非线性映射能力更强,更适用于分析较为复杂的光谱数据。这归因于 ESN 网络的建立和训练过程是完全分开的,输入权重是随机初始化产生的,不需要训练,输出权重是唯一需要调整的部分,从而显著地提高了收敛速度。所以综合来看,ESN 模型性能更优越,预测能力和稳定性更好。从特征波长提取算法来看,CARS 算法提取的特征变量数目为 19,占全光谱变量的 4.7%,RF 算法提取的特征变量数为 15,占全光谱变量的 3.7%,均简化了模型。且从建模结果来看,两种降维方法均能有效提升模型精度,说明全波段光谱存在大量冗余信息,通过特征波长的筛选能够提取出与 AFB1 含量相关性较强的光谱信息。其中,CARS - ESN 为花生油中 AFB1 含量预测的最优组合模型,校正集与预测集决定系数分别为 0.978 和 0.973。该组合模型相比全波段建模的预测集决定系数增加 0.026,均方根误差降低 1.35 $\mu\text{g}/\text{kg}$ 。综上所述,荧光光谱结合 ESN 模型预测花生油中 AFB1 含量是可行的,CARS 所提取的特征波段最能表征花生油中 AFB1 含量,CARS - ESN 为花生油中 AFB1 含量预测的最优组合模型。

表 2 不同算法建模结果对比
Tab.2 Comparison of modeling results of different algorithms

模型	特征提取算法	特征数	参数	校正集		预测集	
				R_c^2	RMSEC/($\mu\text{g}\cdot\text{kg}^{-1}$)	R_p^2	RMSEP/($\mu\text{g}\cdot\text{kg}^{-1}$)
ESN	无	401	(230,0.8,0.01)	0.962	1.940	0.947	3.750
	CARS	19	(230,0.8,0.01)	0.978	1.157	0.973	2.400
	RF	15	(230,0.8,0.01)	0.970	1.531	0.961	2.527
BPNN	无	401	(8,0.01)	0.951	2.888	0.922	4.869
	CARS	19	(8,0.01)	0.962	1.910	0.953	4.168
	RF	15	(8,0.01)	0.967	1.727	0.951	5.326
RBFNN	无	401	0.5	0.960	3.210	0.938	4.149
	CARS	19	0.5	0.965	2.740	0.952	3.270
	RF	15	0.5	0.966	3.120	0.945	3.830
PLSR	无	401	11	0.948	3.305	0.932	4.385
	CARS	19	8	0.951	2.802	0.946	3.041
	RF	15	6	0.959	4.750	0.944	5.099

2.3.3 模型优化结果

用 SSA 优化算法对 CARS - ESN 组合模型的 3

个参数储备池规模(hidden),更新速率(Ir)和正则化系数(reg)进行优化,以使得 ESN 训练迭代更快、

性能更好。由于 SSA 算法最大迭代次数对预测精度和运行时间影响较大,考虑到计算时间,同时为避免陷入局部最优,设置最大迭代次数为 30。SSA 算法迭代曲线如图 6a 所示,可看到 CARS-SSA-ESN 模型经过少量迭代迅速搜索到最优值,优化后的参数 hidden 为 144,Ir 为 0.01,reg 为 0.001。图 6b 显示了 CARS-SSA-ESN 模型真实值与预测值的线性拟合结果,可以看出 AFB1 预测值均分布在真实值附近,能够较准确地判别花生油中 AFB1 含量。

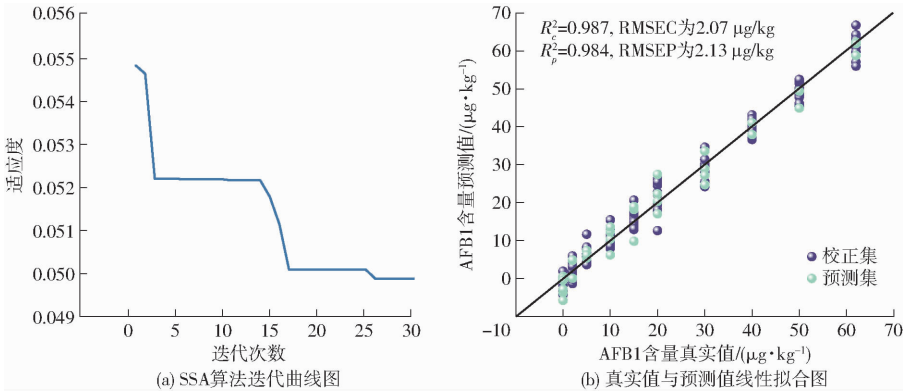


图 6 CARS-SSA-ESN 建模结果图

Fig. 6 Diagrams of CARS-SSA-ESN modeling results

3 结论

- (1)测定含 62 μg/kg AFB1 的花生油的三维荧光光谱图,发现其在 390 nm 激发波长、470 nm 发射波长以及 410 nm 激发波长、670 nm 发射波长处均呈现荧光特征峰。
- (2)使用 K-means 和 SOM 聚类算法对花生油中 AFB1 含量进行超标与否的定性鉴别,准确率均达 95% 以上。
- (3)基于 CARS 和 RF 分别得到 19、15 个特征

- 变量,两种降维方法均有效剔除了与建模相关度低的光谱波段,提高了信噪比与后期建模的精度。其中,CARS 算法消除冗余变量的性能最佳。
- (4)为了评估 ESN 模型的性能,对比了两种经典的神经网络 BPNN、RBFNN 以及经典线性机器学习模型 PLSR 的建模结果。结果表明 ESN 模型性能更优越,预测能力和稳定性更好,具有较强的拟合能力。
- (5)引入 SSA 算法对 ESN 模型的参数进行寻优,结果表明 SSA 算法局部搜索能力较强,对 ESN 参数的优化能够提升模型的稳健性。

参 考 文 献

[1] KARUNARATHNA N B, FERNANDO C J, MUNASINGHE D M S, et al. Occurrence of aflatoxins in edible vegetable oils in Sri Lanka[J]. Food Control, 2019, 101: 97-103.

[2] 陆斌. 酶联免疫吸附技术在食品安全检测中的应用技术[J]. 现代食品, 2022, 28(17): 132-134.

LU Bin. Application of enzyme-linked immunosorbent assay in food safety testing[J]. Modern Food, 2022, 28(17): 132-134. (in Chinese)

[3] 徐洪文, 朱瑜, 徐华, 等. 高效液相色谱法测定食用植物油中 6 种真菌毒素[J]. 中国油脂, 2020, 45(11): 77-83.

XU Hongwen, ZHU Yu, XU Hua, et al. Detection of six mycotoxins in edible vegetable oil by high performance liquid chromatography[J]. China Oils and Fats, 2020, 45(11): 77-83. (in Chinese)

[4] 张强, 刘成海, 孙井坤, 等. 基于支持向量机的稻谷黄曲霉毒素 B1 近红外无损检测[J]. 东北农业大学学报, 2015, 46(5): 84-88.

ZHANG Qiang, LIU Chenghai, SUN Jingkun, et al. Near-infrared spectroscopy nondestructive determination of aflatoxin B1 in paddy rice based on support vector machine regression[J]. Journal of Northeast Agricultural University, 2015, 46(5): 84-88. (in Chinese)

[5] 杨明秀. 基于表面增强拉曼光谱技术的花生油中 AFB1 检测方法研究[D]. 镇江: 江苏大学, 2018.

YANG Mingxiu. Study on detection of AFB1 in peanut oil based on surface-enhance Raman spectroscopy technology [D]. Zhenjiang: Jiangsu University, 2018. (in Chinese)

- [6] DENG J H, ZHANG Z J, LI M C, et al. Feasibility study on Raman spectra-based deep learning models for monitoring the contamination degree and level of aflatoxin B1 in edible oil[J]. *Microchemical Journal*, 2022, 180: 107613.
- [7] 殷金坚. 基于诱导荧光的黄曲霉素光谱检测系统的研制[D]. 长春:吉林大学, 2006.
YIN Jinjian. Design of system in detecting aflatoxin spectrum based on induced fluorescence[D]. Changchun: Jilin University, 2006. (in Chinese)
- [8] 杜阳锋. 基于三维荧光光谱的食品中真菌毒素检测方法[D]. 杭州:浙江大学, 2011.
DU Yangfeng. The detection method of mycotoxin in food based on three-dimensional fluorescence spectrometry[D]. Hangzhou: Zhejiang University, 2011. (in Chinese)
- [9] GOUHEI T, TOSHIYUKI Y, JEAN B H, et al. Recent advances in physical reservoir computing; a review[J]. *Neural Networks*, 2019, 115:100–123.
- [10] 员安然. 基于轻量网络与增量学习的高光谱影像分类[D]. 西安:西安电子科技大学, 2021.
YUAN Anran. Hyperspectral image classification based on lightweight network and incremental learning[D]. Xi'an: Xidian University, 2021. (in Chinese)
- [11] 孙钰扬. 基于回声状态网络的改进模型在公共建筑能耗预测中的应用[D]. 重庆:重庆大学, 2021.
SUN Yuyang. An improved model based on echo state network for energy consumption prediction in public buildings[D]. Chongqing: Chongqing University, 2021. (in Chinese)
- [12] NAOYA T, YUKI M, SHUN-ICHI S, et al. Quantitative analysis of nonlinear optical input/output of a quantum-dot network based on the echo state property[J]. *Optics Express*, 2022, 30(9): 14669–14676.
- [13] HAN T, LINNA Z, MING L, et al. Weighted SPXY method for calibration set selection for composition analysis based on near-infrared spectroscopy[J]. *Infrared Physics & Technology*, 2018, 95: 88–92.
- [14] WU X J, TIAN R L, SUN M F, et al. Application of fluorescence spectra and matrix analysis in the classification of edible vegetable oils[J]. *Spectroscopy and Spectral Analysis*, 2016, 36(7): 2155–2161.
- [15] LI Y, WANG G Z, GUO G S, et al. Spectral pre-processing and multivariate calibration methods for the prediction of wood density in Chinese white poplar by visible and near infrared spectroscopy[J]. *Forests*, 2022, 13(1):62.
- [16] ZHAO A, TANG X, ZHANG Z, et al. Optimizing Savitzky–Golay parameters and its smoothing pretreatment for FTIR gas spectra[J]. *Journal of Instrumental Analysis*, 2016, 36(5): 1340–1344.
- [17] 李霞. 基于近红外光谱技术的赤霞珠干红葡萄酒品质指标检测方法研究[D]. 杨凌:西北农林科技大学, 2018.
LI Xia. Rapid detection of quality index in cabernet sauvignon red wine based on near-infrared spectroscopy[D]. Yangling: Northwest A&F University, 2018. (in Chinese)
- [18] 郭文川, 朱德宽, 张乾, 等. 基于近红外光谱的掺伪油茶籽油检测[J]. *农业机械学报*, 2020, 51(9): 350–357.
GUO Wenchuan, ZHU Dekuan, ZHANG Qian, et al. Detection on adulterated oil-tea camellia seed oil based on near-infrared spectroscopy[J]. *Transactions of the Chinese Society for Agricultural Machinery*, 2020, 51(9): 350–357. (in Chinese)
- [19] SONG X Z, DU G R, LI Q Q, et al. Rapid spectral analysis of agro-products using an optimal strategy: dynamic backward interval PLS-competitive adaptive reweighted sampling[J]. *Analytical and Bioanalytical Chemistry*, 2020, 412: 2795–2804.
- [20] 王斌, 何丙辉, 林娜, 等. 基于随机森林特征选择的茶园遥感提取[J]. *吉林大学学报*, 2022, 52(7): 1719–1732.
WANG Bin, HE Binghui, LIN Na, et al. Remote sensing extraction of tea plantation based on random forest feature selection [J]. *Journal of Jilin University*, 2022, 52(7): 1719–1732. (in Chinese)
- [21] WANG Z. Research on feature selection methods based on random forest[J]. *Tehnicki Vjesnik*, 2023, 30(2): 623–633.
- [22] 李翠玲, 李余康, 谭昊然, 等. 基于 K-means 聚类 and RF 算法的葡萄霜霉病检测分级方法[J]. *农业机械学报*, 2022, 53(5): 225–236, 324.
LI Cuiling, LI Yukang, TAN Haoran, et al. Grading detection method of grape downy mildew based on K-means clustering and random forest algorithm[J]. *Transactions of the Chinese Society for Agricultural Machinery*, 2022, 53(5): 225–236, 324. (in Chinese)
- [23] DOMINIK O. Clustering-based adaptive self-organizing map[J]. *Lecture Notes in Computer Science*, 2021, 12854: 182–192.
- [24] HU R H, TANG Z R, SONG X Y, et al. Ensemble echo network with deep architecture for time-series modeling[J]. *Neural Computing and Applications*, 2021, 33: 4997–5010.
- [25] 王林, 王燕丽, 安泽远. 改进粒子群算法优化回声状态网络的电力需求预测研究[J]. *计算机工程与科学*, 2022, 44(8): 1457–1466.
WANG Lin, WANG Yanli, AN Zeyuan. Echo state networks with improved particle swarm optimization algorithm for electricity demand forecasting[J]. *Computer Engineering & Science*, 2022, 44(8): 1457–1466. (in Chinese)
- [26] 尹航, 吕佳威, 陈耀聪, 等. 基于 LightGBM–SSA–ELM 的新疆羊舍 CO₂ 浓度预测[J]. *农业机械学报*, 2022, 53(1): 261–270.
YIN Hang, LÜ Jiawei, CHEN Yaocong, et al. Prediction of CO₂ concentration in Xinjiang breeding environment of mutton sheep based on LightGBM–SSA–ELM[J]. *Transactions of the Chinese Society for Agricultural Machinery*, 2022, 53(1): 261–270. (in Chinese)