

基于深度学习探究胰蛋白酶催化蛋白质的 特异性酶切预测

刘洋¹, 黄峻洪¹, 岑幸仪¹, 肖楚乔², 李淳淳¹, 陈选³, 李桂香¹, 徐巨才^{1*}, 刘磊^{1*}

(1. 五邑大学药学与食品工程学院, 广东江门 529020) (2. 广东工业大学生物医药学院, 广东广州 510006)

(3. 湖南农业大学园艺学院, 湖南长沙 410128)

摘要: 胰蛋白酶催化蛋白质的特异性酶切预测对于蛋白质的理论降解和结构分析具有重要意义。该研究采用卷积神经网络和长短期记忆网络构建了一种基于深度学习的胰蛋白酶催化蛋白质酶解预测模型, 并探讨了不同超参数对模型性能的影响。结果表明: 较低的学习率、较高的 Batch size 和较浅的卷积层数对模型训练的稳定性 and 收敛性有利, 可保障较好的预测效果, 模型的较优工作参数为: 学习率 0.001、Batch size 512、卷积层数 1。在此条件下, 模型在 PDX010627 数据集上的准确率为 0.950、特异性为 0.987、精确度为 0.986、召回率为 0.961、F1 分数为 0.973, 显示出良好的预测能力和稳定性。进一步应用于不同物种的公开数据集, 模型的准确率均保持在 0.920 以上, AUC 值、精确率、召回率和 F1 分数均在 0.900~0.989 之间, 说明模型在胰蛋白酶催化蛋白质特异性酶切预测体系中具备较强的泛化能力, 可有效提升酶切位点预测的准确性和可靠性。本研究将有望为蛋白质组学鉴定及空间结构分析提供了一种新思路, 推动蛋白质酶解研究与生物信息学的发展。

关键词: 胰蛋白酶; 酶切预测; 深度学习; 蛋白酶解; 蛋白质组学

文章编号: 1673-9078(2025)11-68-76

DOI: 10.13982/j.mfst.1673-9078.2025.11.1140

Prediction of Trypsin-catalyzed Protein Cleavage Specificity Based on Deep Learning

LIU Yang¹, HUANG Junhong¹, CEN Xingyi¹, XIAO Chuqiao², LI Zhunchun¹, CHEN Xuan³, LI Guixiang¹,
XU Jucai^{1*}, LIU Lei^{1*}

(1. School of Pharmacy and Food Engineering, Wuyi University, Jiangmen 529020, China)

(2. School of Biomedical and Pharmaceutical Sciences, Guangdong University of Technology, Guangzhou 510006, China)

(3. College of Horticulture, Hunan Agricultural University, Changsha 410128, China)

Abstract: The prediction of trypsin-catalyzed specific proteolysis is vital for guiding the theoretical degradation and structural analysis of proteins. In this study, a deep learning-based model was constructed to predict trypsin-catalyzed protein cleavage using convolutional neural networks (CNN) and long short-term memory (LSTM) networks. The impact of various hyperparameters on the model's performance was also explored. The results showed that a lower learning rate, higher batch size, and fewer convolutional layers were beneficial for the model's training stability, ensuring superior predictive outcomes. The optimal parameters were

引文格式:

刘洋, 黄峻洪, 岑幸仪, 等. 基于深度学习探究胰蛋白酶催化蛋白质的特异性酶切预测[J]. 现代食品科技, 2025, 41(11): 68-76.

LIU Yang, HUANG Junhong, CEN Xingyi, et al. Prediction of trypsin-catalyzed protein cleavage specificity based on deep learning [J]. Modern Food Science and Technology, 2025, 41(11): 68-76.

收稿日期: 2024-08-04; 修回日期: 2024-09-27; 接受日期: 2024-09-29

基金项目: 国家自然科学基金青年科学基金项目 (32402276); 广东省科技创新战略专项市县科技创新支撑 (大专项 + 任务清单) 项目 - 暨 2023 年江门市关键核心技术“揭榜挂帅”制项目 (2023780200060009632); 广东省基础与应用基础研究基金联合基金青年基金项目 (2022A1515110711); 广东省“百千万工程”项目 (BQW2024001)

作者简介: 刘洋 (1998-), 男, 硕士, 研究方向: 食品生物信息技术与多组学技术, E-mail: huhaoaaq@163.com

通讯作者: 徐巨才 (1991-), 男, 博士, 讲师, 研究方向: 食品生物信息技术与智造技术, E-mail: xujucai.happy@163.com; 共同通讯作者:

刘磊 (1982-), 男, 博士, 研究员, 研究方向: 农产品加工与品质调控, E-mail: liulei309@tom.com

identified as a learning rate of 0.001, batch size of 512, and one convolutional layer. In that case, the model achieved an accuracy of 0.950, specificity of 0.987, precision of 0.986, recall of 0.961, and an F1 score of 0.973 on dataset PXD010627, demonstrating excellent predictive capability and stability. Furthermore, when applied to publicly available datasets from different species, the model maintained an accuracy >0.920, with AUC values, precision, recall, and F1 scores ranging between 0.900 and 0.989. This indicates a strong generalization ability of the model in predicting specific cleavage sites of trypsin-catalyzed proteins which could considerably enhance the accuracy and reliability of cleavage site predictions. We hope this work could offer new ideas for protein identification and spatial structure analysis in proteomics, promoting advancements in proteolytic research and bioinformatics.

Key words: trypsin; cleavage prediction; deep learning; proteolysis; proteomics

由胰蛋白酶催化的蛋白质特异性水解体系^[1-3]在食品、生物医药研究领域应用广泛,明晰该过程胰蛋白酶作用规律及蛋白质降解特征,对于指导蛋白质的组学鉴定与空间结构分析具有重要意义^[4,5]。然而,由于酶解过程极为复杂,胰蛋白酶作用位点受蛋白序列残基分布、酶解条件及蛋白结构等多种因素影响,尚缺乏有效的酶解预测模型。刘辉等^[6]提出了一种基于马尔科夫链的预测方法,用于识别蛋白序列中的胰蛋白酶作用位点,但由于该方法忽视了肽段间各氨基酸残基的长程作用,以致模型泛化性不足,尚待进一步验证。

目前,胰蛋白酶的作用位点预测大多仍主要基于Keil规则^[7-9],即认为胰蛋白酶在酶解过程中主要切割序列中精氨酸或赖氨酸残基所在位置的羧基端(C端)^[6]。然而,实践过程中发现酶切位点漏切现象时有发生(据观察,漏切率可高达40%)^[10]。这种现象不断在提示,胰蛋白酶的酶解行为不仅受到底物氨基酸序列的影响,还与蛋白质的结构、酶解条件^[11-13](如pH值、温度、酶与底物的比例等)等密切相关。由于酶解是一种概率性事件,胰蛋白酶虽对特定位点有切割倾向,但这种倾向并不等同于绝对的切割确定性。因此,基于Keil规则的预测方法存在明显局限性,难以准确预测酶切位点与特征。当前,亟需建立一种更为有效的蛋白序列特异性酶切预测方法。

为此,本文主要研究并构建了一种基于深度学习的胰蛋白酶催化蛋白质特异性酶切预测模型,以模型灵敏度、特异性、假阳性率、精确度等为主要指标,探讨了模型各超参数对预测效果的影响并确定了其较优参数。在此基础上,将模型应用于不同物种数

据集进行了泛化应用评价和验证,以期蛋白质的特异性酶催化预测分析提供一种可行的方法,并为蛋白质的组学鉴定、空间结构以及多肽释放分析提供一种新思路。

1 方法

1.1 数据集的准备

1.1.1 数据收集与分割

本文选取三个不同物种(*H. sapiens*、*D. melanogaster*和*M. musculus*)的蛋白质组学分析结果构建了4个数据集(见表1)。每个数据集按照7:3的比例进行分割,分别组成训练集与验证集质谱数据^[14]。

1.1.2 数据预处理

为保障模型的预测可靠性,仅保留与胰蛋白酶切割作用直接相关的多肽数据,清洗、剔除其余无关多肽信息。设定多肽最大漏切数为2,即每个肽段最多包含3个K/R位点,采用滑动剪切窗口,设置窗口长度为15,即潜在酶切位点的N端和C端各包含7个氨基酸残基。由于识别出的肽段长度至少为7个氨基酸,因此仅保留那些在潜在酶切位点N端拥有7个氨基酸残基的肽段,而潜在酶切位点C端至少包含1个氨基酸的肽段均被保留。筛选后的肽段将逐一映射回原蛋白质序列,并提取酶切位点两侧各7个氨基酸的序列,作为模型的训练数据集。根据酶切位点在肽段中的位置,将其标记为0或1(分别表示未切割/已切割,MC/C),以作为训练数据的标签。此过程如图1a所示。

表 1 公开数据集来源
Table 1 Sources of public datasets

数据集	仪器	物种	鉴定肽段	筛选肽段	训练数据
PXD007158 ^[15]	Q Exactive	<i>H. sapiens</i> (<i>HeLa</i>)	86 940	81 019	10 000
PXD010627 ^[16]	Q Exactive	<i>D. melanogaster</i>	119 212	111 529	10 000
PXD014777 ^[17]	timsTOF Pro	<i>H. sapiens</i> (<i>HeLa</i>)	49 990	49 400	10 000
PXD013478 ^[18]	Q Exactive	<i>M. musculus</i>	59 211	56 100	9 726

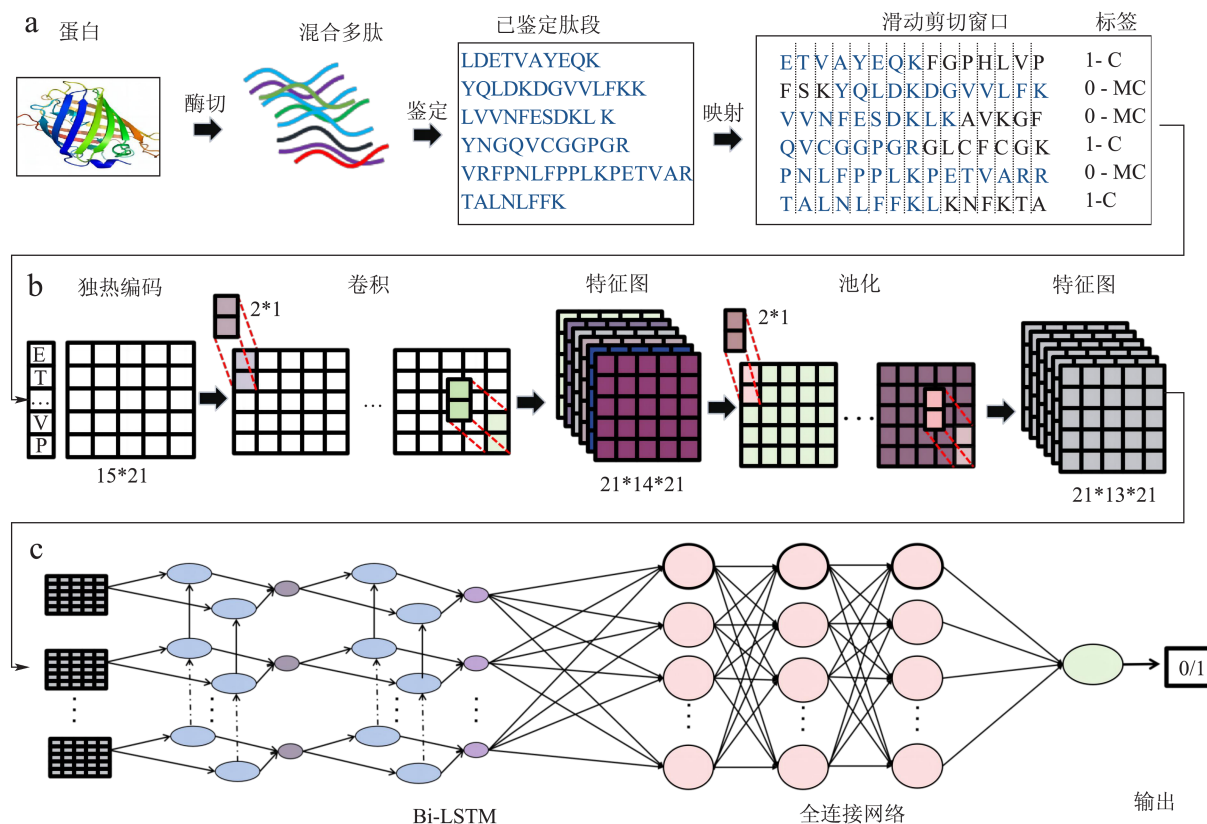


图 1 胰蛋白酶酶催化模型

Fig.1 Modeling the catalytic mechanism of pancreatic proteases

注: a 为数据预处理模块; b 为卷积网络模块; c 为序列特征提取模块。

1.2 模型的构建

模型构建技术路线图如图 1b、1c 所示, 主要由卷积神经网络、长短期记忆网络、全连接层等组成。

1.2.1 卷积神经网络

卷积神经网络^[19]旨在提取肽段序列间的特征(图 1b), 其具有一个大小为 2×1 的卷积核输出通道大小为 21 的卷积层, 将输入矩阵 $X^{15 \times 21}$ 转换为 $Y^{21 \times 14 \times 21}$, 该卷积算子在输入矩阵上总共滑动 294 次 (21×14) 并进行一系列的点积和加法如式子 1, 其中 $Y_{i,j,k}$ 表示卷积核的输出, $W_{m,0,k}$ 表示权重系数。网络的激活函数采用 Relu^[20], 其目的是在模型中添加非线性, 同时也使用了平均池化层, 用于降低数据维度并提取主要特征。增加 dropout 层, 以 0.3 的概率对输入进行随机置零, 以避免模型过拟合。

$$Y_{i,j,k} = \sum_{m=0}^1 W_{m,0,k} \cdot X_{i+m,j} \quad (1)$$

1.2.2 Bi-LSTM

在本文中, CNN 模块中提取的特征矩阵用于 Bi-LSTM 模块的输入, 从而提取序列之间的相关信息, 如图 1c 所示。本文使用双向循环神经网络, 目的在于可以有效捕获序列在前向和后向方向上的文本之间的

关系, 利用记忆细胞来记住长期的历史信息, 以便于从序列中获取全局信息^[21]。

1.3 模型的评估

为保证模型训练效果, 模型训练过程设定 epoch 数量 $10^{[22]}$, 同时为避免模型过拟合, 采用 BCELoss 作为损失函数^[23], 优化器选择 Adam^[24], 通过结合 Momentum 和 RMSprop 的优点, 使得模型可自适应调整学习率, 加速收敛并提高性能。模型训练利用前述不同物种数据集, 每个数据集中均包含等量的切割肽段与 1~2 个漏切的肽段, 每个数据集均采用训练-验证策略, 并通过 10 折交叉验证法以确保评估结果的准确性和可靠性。在模型评估方面, 使用灵敏度 (sensitivity)、特异性 (specificity)、假阳性率 (false positive rate)、精确度 (precision)、F1 分数以及受试者特征曲线下面积 (AUC) 等指标, 来评估模型在预测酶切与漏切数据上的性能。具体的计算方法如式 (2)~(6) 所述:

$$sensitivity = \frac{TP}{TP + FN} \quad (2)$$

$$specificity = \frac{TN}{FP + TN} \quad (3)$$

$$\text{false positive rate} = \frac{FP}{FP+TN} \quad (4)$$

$$\text{precision} = \frac{TP}{TP+FP} \quad (5)$$

$$F1 = 2 \times \frac{\text{precision} \times \text{sensitivity}}{\text{precision} + \text{sensitivity}} \quad (6)$$

式中:

TP、FP、TN、FN——分别代表真正例、假正例、真反例和假反例。

AUC (曲线下面积) 通过对由分类器的灵敏度 (真阳性率) 和假阳性率 (False Positive Rate, FPR) 之间关系构成的受试者工作特征曲线 (Receiver Operating Characteristic, ROC) 进行积分计算而得^[25]。

1.4 数据处理

数据处理主要采用五邑大学生物医药大数据中心 (Big Data Center for Biomedical Research) 进行, 该中心配备小型超算高性能集群 (计算能力理论峰值 ≥ 428 万亿次/秒, 存储能力 3.2 PB, 存储聚合带宽 ≥ 20 GB/s, 包含 120 台计算节点+3 台 GPU 节点+2 台八路胖节点)。本课题主要使用计算节点与 GPU 节点进行模型训练与测试, 其中计算节点具体设备参数为 2 颗 Xeon Gold 6148, 192 GB 内存, 1 块 240 GSSD, 2 个 100 GEDR 或 OPA 端口, GPU 节点具体设备参数为 2 颗 Xeon Gold 5118, 128 GB 内存, 2 块 1.2T 硬盘, 2 块 NVIDIA V100 GPU 卡, 1 个 100 GEDR 或 OPA 端口。使用 Prism 软件 (GraphPad, 美国) 进行统计分析与画图, 数据显示为平均值 $\pm$ 标准差。

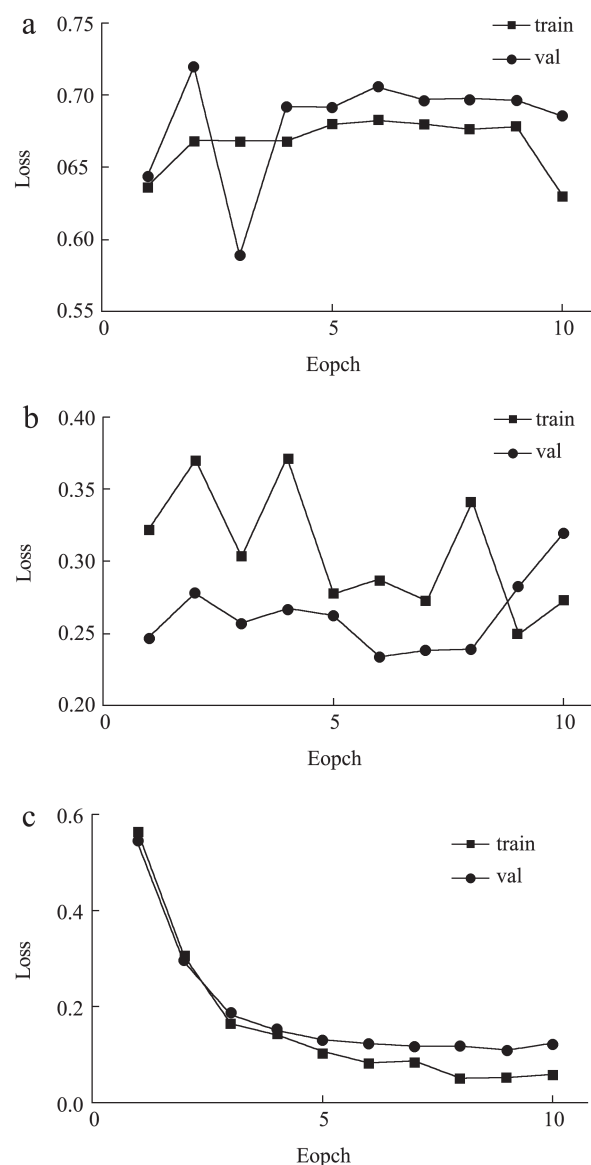
2 结果与讨论

2.1 不同超参数对模型预测效果的影响

2.1.1 学习率对模型的影响

学习率 (Learning rate)^[26-28] 是深度学习中的一个关键超参数, 它控制着优化过程中每次迭代时参数更新的步幅。若学习率设置过大, 可能导致优化参数在最小值附近波动, 难以稳定收敛; 而若设置过小, 则可能使优化参数收敛速度减慢, 需要更多迭代次数才能达到收敛。因此, 本实验在固定其他超参数的前提下, 选择五个不同的学习率 (0.1、0.05、0.01、0.005 和 0.001) 对模型进行训练, 以探讨学习率对模型预测性能的影响。结果如图 2 所示。由图可知, 当学习率为 0.1 和 0.05 时, 无论是训练集 (train) 还是测试集 (val), 模型损失函数 (Loss) 均出现大幅波动和震荡,

并且随着训练周期 (Epoch) 的增加, 损失值不降反增, 表明模型在训练过程中不稳定, 影响了收敛性, 这可能是由于学习率设置过高导致模型更新步幅过大, 特征提取学习不充分所致。当学习率为 0.01 和 0.005 时, 模型损失函数有所改善, 随着训练周期的增加, 损失值逐渐下降, 但仍存在一定波动和震荡。特别是从图 2e 可以看出, 当学习率为 0.005 时, 模型损失函数随着训练周期的增加迅速下降, 并在后期逐步稳定, 呈现出与大多数深度学习模型相似的梯度下降过程, 这反映在当前学习率下对于胰蛋白酶催化蛋白质的特异性酶切预测较为合适, 预测酶切具有实际参考价值。同时, 结合图 2f 可知, 此时模型测试集的准确率高达 0.95。综上所述, 选择学习率为 0.001 作为胰蛋白酶催化蛋白质酶解预测模型的较优参数。



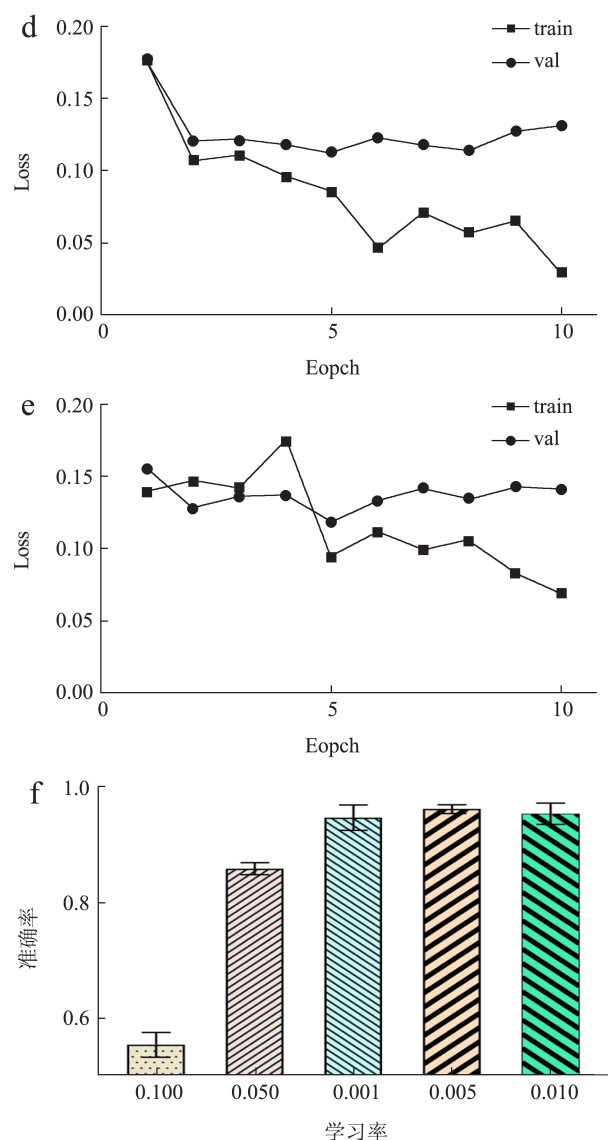


图2 学习率对模型的影响

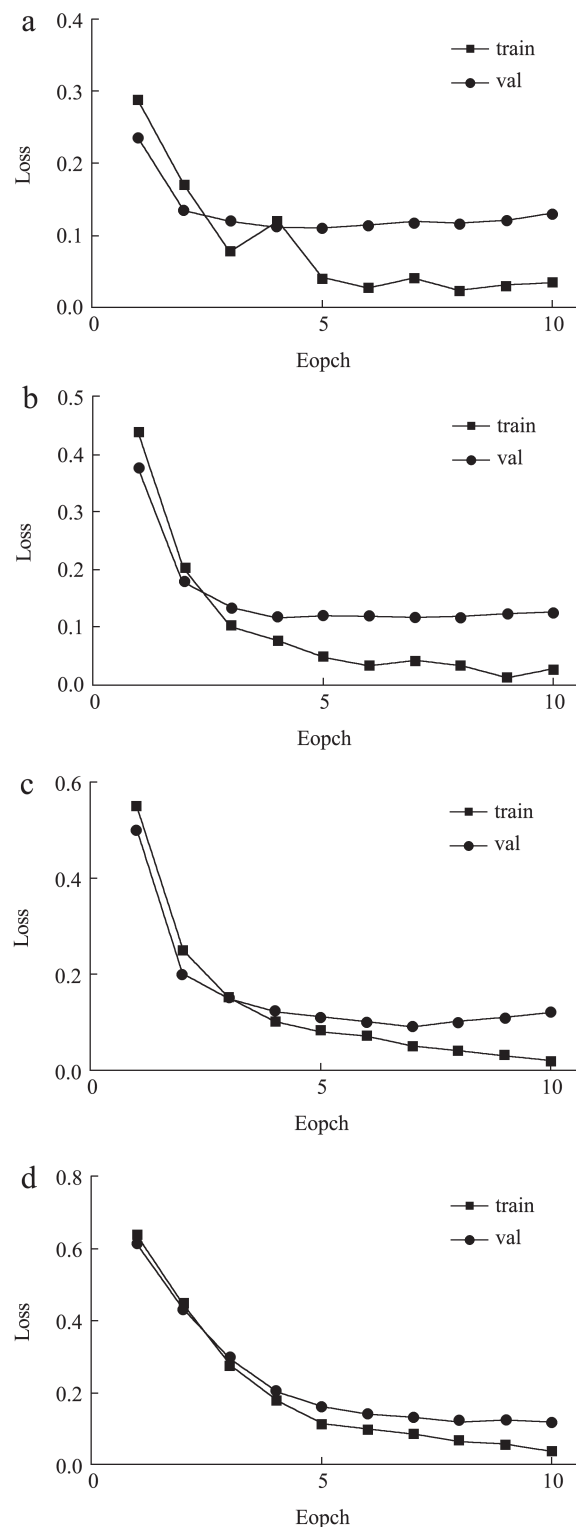
Fig.2 The impact of learning rate on model performance

注:a、b、c、d、e分别代表模型在学习率0.1、0.05、0.001、0.005、0.01时的损失函数;f为各学习率下模型的测试集准确率。

2.1.2 Batch size对模型的影响

Batch size 是指在每次迭代中用于计算梯度和更新模型权重的样本数量。选择适当的 Batch size 对模型训练的效率和性能至关重要^[29]。较小的 Batch size 可能增加训练过程的不稳定性,但有助于更快收敛;较大的 Batch size 则使训练更加稳定,但需更多计算资源,且可能无法达到最优解。本实验在前述实验基础上,选取五个不同的 Batch size (32、64、128、256、512) 对模型进行训练,以探讨 Batch size 对模型预测性能的影响。结果如图3所示。由图可知,当 Batch size 较小(32、64、128)时,模型在测试集上的损失函数过早稳定且不再显著下降,可能原因是较小的

Batch size 使得模型无法在小样本上有效提取序列与酶切之间的特征关系;当 Batch size 为 256 和 512 时(图 3d、3e),随着训练迭代周期增加,模型损失函数逐渐下降并趋于平稳,表明损失函数有效收敛,这表明在大样本的情况下模型可以有效学习到酶切的特征。结合图 3f,相比 Batch size 为 256 时,Batch size 为 512 时模型在测试集上的准确率更高,达 0.947。综合考虑,选择 512 作为 Batch size 的较优值。



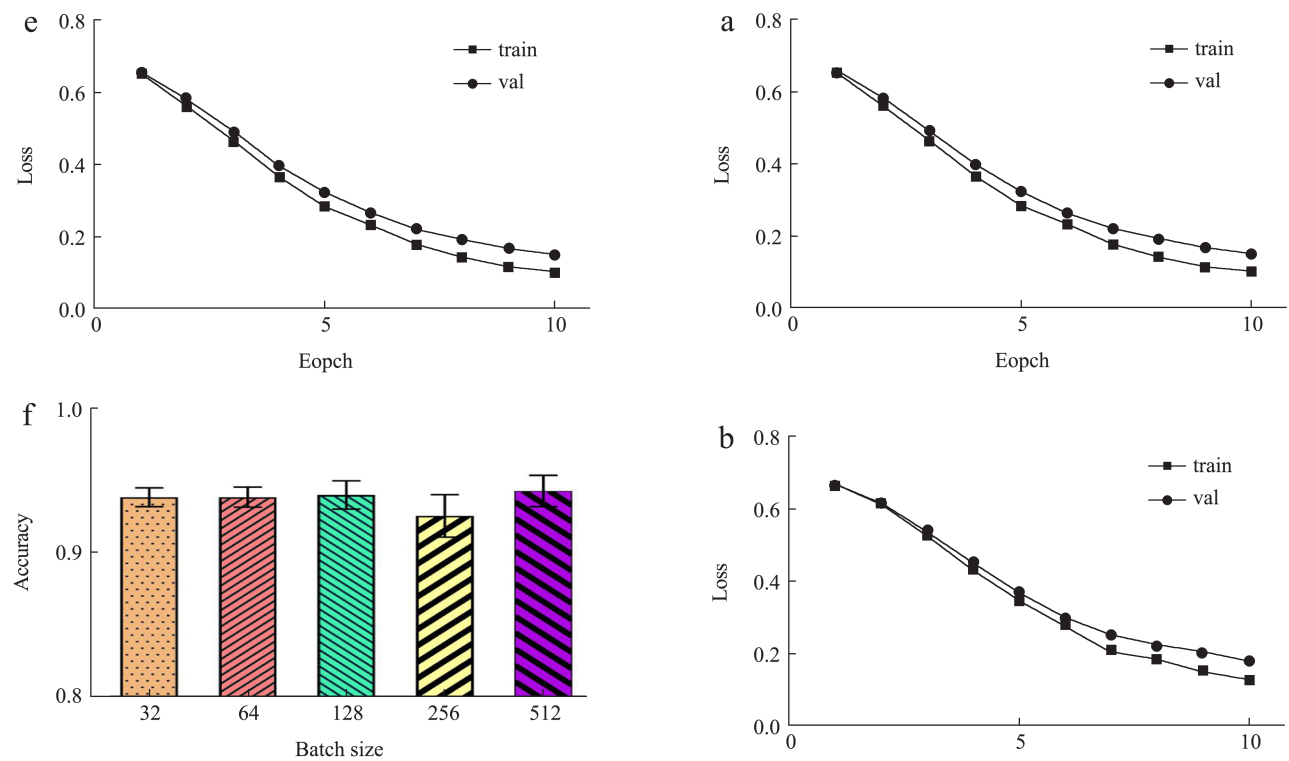


图 3 Batch size 对模型的影响

Fig.3 Effect of batch size on model performance

注：a~e 分别代表模型在 Batch size 为 32、64、128、256、512 时的损失函数；f 为各 Batch size 下模型的测试集准确率。

2.1.3 卷积层数的影响

为进一步改善模型的预测效果，本实验研究了卷积神经网络（CNN）1、2、3 和 4 层卷积层对模型预测效果的影响。在不同层数的实验中，均配有相应数量的池化层和正则化层，以维持网络结构的平衡。结果如图 4 所示。结果显示，当卷积层数为 1 时，模型损失函数输出值随着训练周期增加而缓慢下降，最终收敛值在四个实验中最小。这表明，在胰蛋白酶作用位点预测的公开数据集上，较浅的卷积网络结构更有效地提取和学习特征，实现更高的分类准确率，可能原因是胰蛋白酶酶切蛋白具有较强的特异性，容易学习到酶切特征进而不需要太复杂的网络结构。随着卷积层数增加，尽管特征提取能力增强，但也带来了过拟合风险和计算复杂度的增加。实验中，增加卷积层数并未显著提升性能，反而可能导致测试集上性能下降。综合考虑模型性能、训练效率及过拟合风险，选择单层卷积层的 CNN 作为特征提取器，以确保模型能够高效地学习和提取多肽序列中的关键特征，并保持较低的计算复杂度和良好的泛化能力。

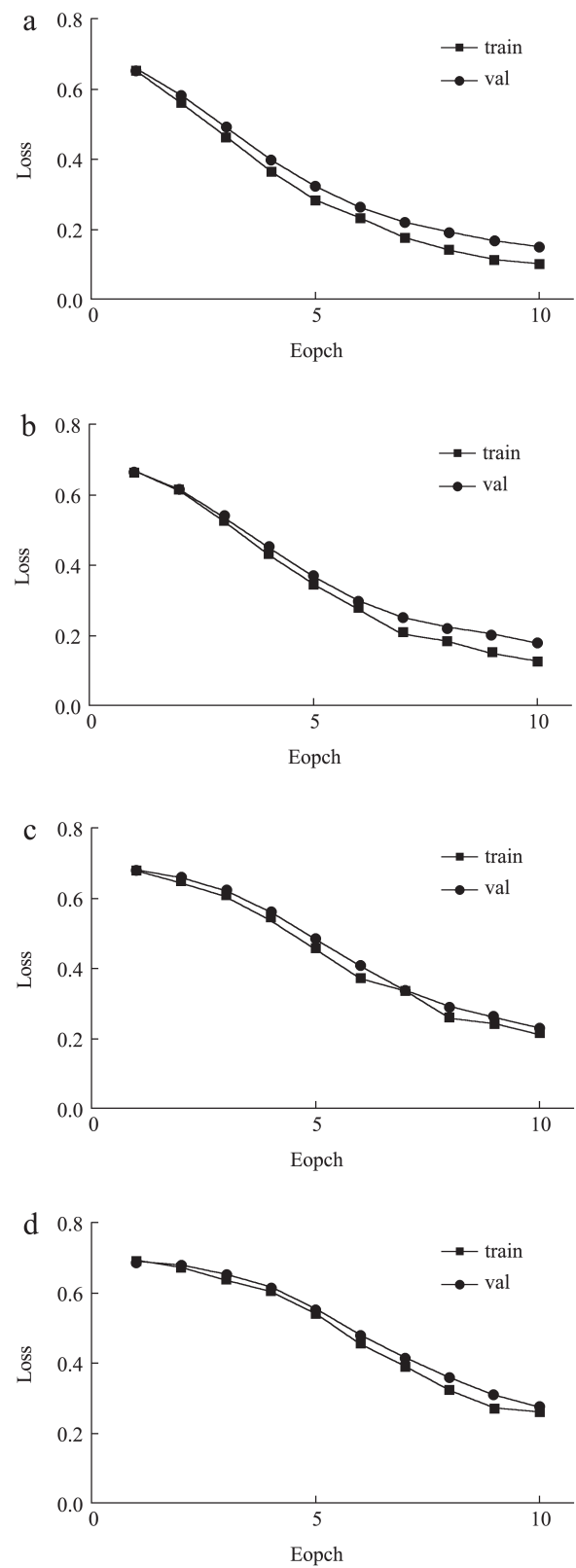


图 4 卷积层数对模型的影响

Fig.4 Effect of convolutional layers on model performance

注：a~d 分别代表模型 CNN 卷积层数为 1、2、3、4 的损失函数图像。

2.1.4 模型的准确性评价

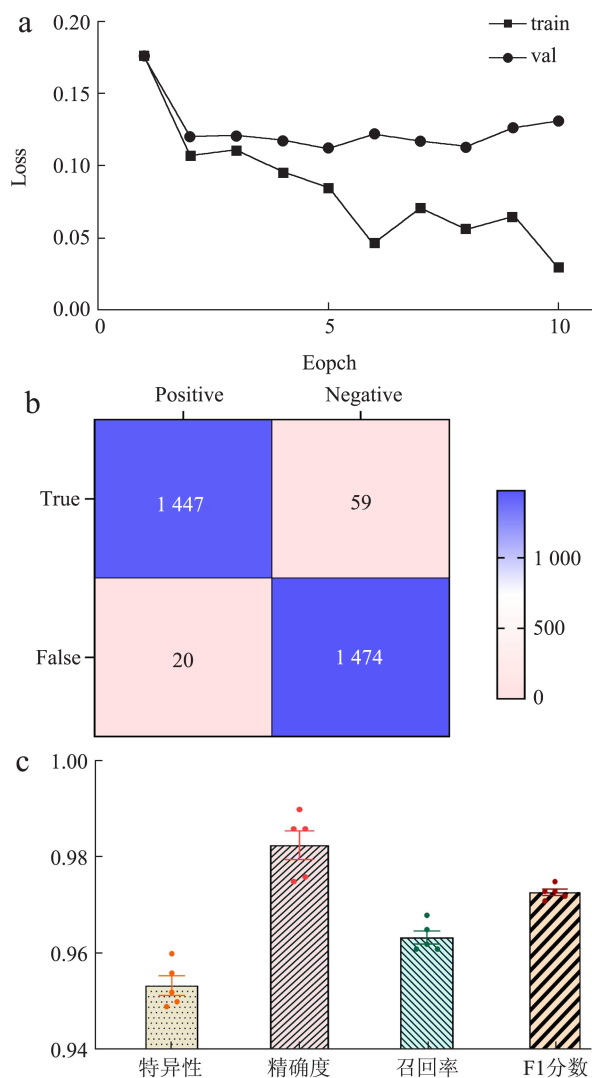


图5 模型在 PDX010627 数据集上的应用评价

Fig.5 Evaluation of model performance on the PDX010627 dataset

注: a、b 分别代表模型在 PDX010627 数据集上的准确率、混淆矩阵; c 代表模型在 PDX010627 数据集上的特异性、精确度、召回率及 F1 分数。

基于上述较优参数,在数据集 PDX010627 上训练模型,所得结果如图 5 所示。由图 5a 可知,随着训练周期的增加,模型在训练集和测试集上的准确率稳步上升,最终稳定在约 0.95。这表明模型在充分训练后,能够准确分类胰蛋白酶的酶切与漏切情况。为全面评估模型预测效果,计算混淆矩阵^[30-32](包括真正例(TP)、假正例(FP)、真反例(TN)和假反例(FN))如图 5b 所示,基于该矩阵,进一步计算模型的特异性、精确度、召回率及 F1 分数等重要评价指标如图 5c。结果表明,模型的特异性高达 0.960,表明在实际为负例的样本中,模型能正确识别出 98.7% 的

负例,表现出优异的负例识别能力。同时,模型的精确度为 0.986,说明在模型预测为正例的样本中,有 98.6% 是真正例,显示了高度的正例预测准确性。此外,模型的召回率为 0.961,表明在实际为正例的样本中,模型能成功预测出 96.1% 的正例,进一步证明了模型在捕获正例方面的良好表现。综上所述,基于 *D.melanogaster* (PDX010627) 数据集训练的模型在准确分类胰蛋白酶的酶切与漏切情况方面表现出色,其高特异性、精确度和召回率使其在指导实际酶解实验中的应用具有重要的潜力和价值。

2.2 模型在不同物种数据集上的泛化应用评价

为进一步评价模型泛化能力,将模型应用于不同物种数据集(包括 PDX010627、PDX010158、PDX013478 和 PDX014777)的训练和预测中,评估并对比其性能,如图 6 所示。由图可知,模型在这些数据集上均表现出了较为良好的预测性能,在四个数据集上的准确率分别达 0.95、0.92、0.92 和 0.96,显示出其对多肽序列的高识别准确性。进一步计算其 AUC 值分别为 0.94、0.92、0.92 和 0.96,表明模型具备较高的分类能力。同时,模型的精确率、召回率和 F1 分数均不低于 0.92,表明模型在负例识别方面性能优异,且在不同物种数据集上表现出良好的稳定性和可靠性。值得注意的是,模型在 *H. sapiens* 数据集(PDX014777)和 *D. melanogaster* 数据集(PDX010627)上的表现更为优异,可能的原因是这两个物种包含了广泛的多肽序列信息和多样的酶切位点,同时其蛋白具有较一致的特征,这有助于模型更容易捕捉和学习这些特征,从而提高预测的准确性。整体而言,该模型在不同物种数据集上均展现出了较高的准确率、AUC 值、精确率、召回率、F1 分数和特异性,具备较强的泛化能力,可适用于不同物种下由胰蛋白酶催化的蛋白质酶切预测任务。

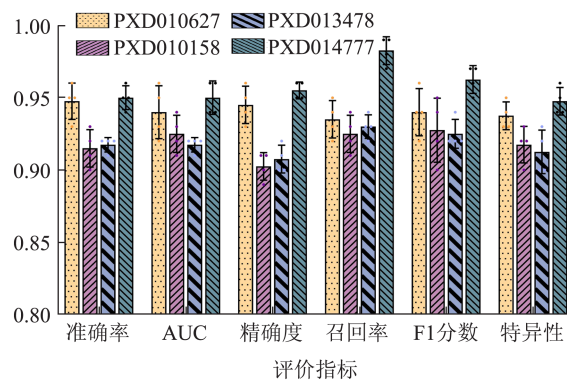


图6 模型在不同数据集上的表现与评价

Fig.6 Performance and evaluation of the model across different datasets

3 结论

本文结合卷积神经网络与长短期记忆网络构建了一种胰蛋白酶催化蛋白质的特异性酶切预测模型,通过探讨不同超参数对模型预测效果的影响,确定了模型的较优参数为:学习率 0.001、批处理大小 512、卷积层和池化层各 1 层等关键参数。在该条件下,模型在 *D. melanogaster* 胰蛋白酶酶解数据集(PXD010627)上的准确率高达 95%,且在其他物种数据集上同样表现出了优异的预测性能,灵敏度、特异性、精确度、召回率和 F1 分数均大于 0.920,显示模型具有良好的泛化能力和稳定性,证实了模型在胰蛋白酶催化蛋白质特异性酶切预测中的可靠性。本文所构建的深度学习模型提高了对胰蛋白酶酶切位点的预测准确性,将有望为蛋白质的组学鉴定、空间结构和多肽释放分析^[33,34]提供新思路。未来,研究将进一步拓展前述模型在非特异性蛋白酶催化蛋白质水解预测中的泛化应用,以推动蛋白质组学与多肽组学研究的快速发展。

参考文献

- [1] TAVANO O L, BERENGUER-MURCIA A, SECUNDO F, et al. Biotechnological applications of proteases in food technology [J]. Comprehensive Reviews in Food Science and Food Safety, 2018, 17(2): 412-436.
- [2] KAUR J, SINGH P K. Trypsin detection strategies: a review [J]. Critical Reviews in Analytical Chemistry, 2022, 52(5): 949-967.
- [3] HABINSHUTI I, NSENGUMUREMYI D, MUHOZA B, et al. Recent and novel processing technologies coupled with enzymatic hydrolysis to enhance the production of antioxidant peptides from food proteins: A review [J]. Food Chemistry, 2023: 136313.
- [4] SUN B, LIU Z, LIU J, et al. The utility of proteases in proteomics, from sequence profiling to structure and function analysis [J]. Proteomics, 2023, 23(6): 2200132.
- [5] 范丽琪,郁晓艺,刘磊,等.肽组学分析技术及其在食品研究中的应用进展[J].现代食品科技,2023,39(8):352-359.
- [6] 刘辉,张纪阳,孙汉昌,等.基于马尔科夫链的胰蛋白酶肽段蛋白酶切位点概率预测[J].生物物理学报,2011,27(6):528-536.
- [7] YEN C Y, RUSSELL S, MENDOZA A M, et al. Improving sensitivity in shotgun proteomics using a peptide-centric database with reduced complexity: protease cleavage and SCX elution rules from data mining of MS/MS spectra [J]. Analytical Chemistry, 2006, 78(4): 1071-1084.
- [8] KEIL B, KEIL B. Essential substrate residues for action of endopeptidases [J]. Specificity of Proteolysis, 1992, 43: 228.
- [9] RODRIGUEZ J, GUPTA N, SMITH R D, et al. Does trypsin cut before proline? [J]. Journal of Proteome Research, 2008, 7(1): 300-305.
- [10] WALMSLEY S J, RUDNICK P A, LIANG Y, et al. Comprehensive analysis of protein digestion using six trypsins reveals the origin of trypsin as a significant source of variability in proteomics [J]. Journal of Proteome Research, 2013, 12(12): 5666-5680.
- [11] DEL RIO A R, KEPPLER J K, BOOM R M, et al. Protein acidification and hydrolysis by pepsin ensure efficient trypsin-catalyzed hydrolysis [J]. Food & Function, 2021, 12(10): 4570-4581.
- [12] 关致颖,潘建章,方群.蛋白质组学分析中蛋白质快速酶解方法的研究进展[J].分析科学学报,2022,38(6):771-778.
- [13] 洪佳伟,肖柳柳,王爱琳,等.核桃抗氧化肽的酶法制备、鉴定及其对脂氧合酶的抑制机理[J].食品与生物技术学报,2024, 43(6):112-118.
- [14] LIU Y, MI Y, MUELLER T, et al. Multi-omic measurements of heterogeneity in HeLa cells across laboratories [J]. Nature Biotechnology, 2019, 37(3): 314-322.
- [15] HAMMERSCHMIDT P, OSTKOTTE D, NOLTE H, et al. CerS6-derived sphingolipids interact with Mff and promote mitochondrial fragmentation in obesity [J]. Cell, 2019, 177(6): 1536-1552.
- [16] GÄRTNER S M, HUNDERTMARK T, NOLTE H, et al. Stage-specific testes proteomics of Drosophila melanogaster identifies essential proteins for male fertility [J]. European Journal of Cell Biology, 2019, 98(2-4): 103-115.
- [17] PRIANICHNIKOV N, KOCH H, KOCH S, et al. MaxQuant software for ion mobility enhanced shotgun proteomics [J]. Molecular & Cellular Proteomics, 2020, 19(6): 1058-1069.
- [18] JUST P A, CHARAWI S, DENIS R G, et al. Lkb1 suppresses amino acid-driven gluconeogenesis in the liver [J]. Nature Communications, 2020, 11(1): 6127.
- [19] ALZUBAIDI L, ZHANG J, HUMAIDI A J, et al. Review of deep learning: concepts, CNN architectures, challenges, applications, future directions [J]. Journal of Big Data, 2021, 8: 1-74.
- [20] AGARAP A F. Deep learning using rectified linear units (relu) [J]. arXiv Preprint arXiv, 2018, 1803: 08375.
- [21] ASLAN M F, UNLERSEN M F, SABANCI K, et al. CNN-based transfer learning-BiLSTM network: A novel approach for COVID-19 infection detection [J]. Applied Soft Computing, 2021, 98: 106912.
- [22] CHEN Y, YUILLE A, ZHOU Z. Which layer is learning faster? a systematic exploration of layer-wise convergence rate for deep neural networks [C]// The Eleventh International Conference on Learning Representations, 2023.
- [23] LIN T-Y, GOYAL P, GIRSHICK R, et al. Focal loss for dense object detection [C]// Proceedings of the IEEE international conference on computer vision, 2017: 2980-2988.
- [24] KINGMA D P, BA J. Adam: A method for stochastic optimization [J]. arXiv Preprint arXiv, 2014, 1412: 6980.

- [25] GÖREKE V, SARı V, KOCKANAT S. A novel classifier architecture based on deep neural network for COVID-19 detection using laboratory findings [J]. *Applied Soft Computing*, 2021, 106: 107329.
- [26] ZEILER M D. Adadelta: an adaptive learning rate method [J]. *arXiv Preprint arXiv*, 2012, 1212: 5701.
- [27] TANG S, ZHU Y, YUAN S. An improved convolutional neural network with an adaptable learning rate towards multi-signal fault diagnosis of hydraulic piston pump [J]. *Advanced Engineering Informatics*, 2021, 50: 101406.
- [28] SHAZEER N, STERN M. Adafactor: Adaptive learning rates with sublinear memory cost [C]// *International Conference on Machine Learning*, 2018: 4596-4604.
- [29] RADIUK P M. Impact of training set batch size on the performance of convolutional neural networks for diverse datasets [J]. *Information Technology and Management Science*, 2017, 20(1): 20-24.
- [30] XU J, ZHANG Y, MIAO D. Three-way confusion matrix for classification: A measure driven view [J]. *Information Sciences*, 2020, 507: 772-794.
- [31] LUQUE A, CARRASCO A, MARTÍN A, et al. The impact of class imbalance in classification performance metrics based on the binary confusion matrix [J]. *Pattern Recognition*, 2019, 91: 216-231.
- [32] KRSTINIĆ D, BRAOVIĆ M, ŠERIĆ L, et al. Multi-label classifier performance evaluation with confusion matrix [J]. *Computer Science & Information Technology*, 2020, 1: 1-14.
- [33] 范丽琪,刘洋,杨宝,等.酸溶性大豆蛋白酶解产物的制备及特性研究[J].*食品工业科技*,2025,46(3):53-59.
- [34] 方磊,夏羽菡,任杰,等.小麦水解蛋白肽段组成及结构序列分析[J].*食品与生物技术学报*,2024,43(6):104-111.