

杨学冬,韩丽君,王榕,等.一种基于深度学习的苦味肽精准预测方法[J]. 轻工学报,2023,38(3):11-16. YANG X D,HAN L J,WANG R,et al. An accurate identification method of bitter peptides based on deep learning [J]. Journal of Light Industry,2023,38(3):11-16. DOI:10.12187/2023.03.002

# 一种基于深度学习的苦味肽精准预测方法

杨学冬1,韩丽君1,王榕1,王宏伟2,王晓1,3

- 1. 郑州轻工业大学 计算机与通信工程学院,河南 郑州 450001:
- 2. 郑州轻工业大学 食品与生物工程学院,河南 郑州 450001;
- 3. 河南省食品安全数据智能重点实验室,河南 郑州 450001

摘要:鉴于湿实验方法已无法满足快速鉴别苦味肽的需求,提出一种新颖的融合了传统手工特征和预训练深度特征的预测方法 Bitter-Fus。该方法首先使用预训练蛋白质序列语言模型自动地从多肽序列中提取深度学习特征,然后将该特征输入长短期记忆(LSTM)网络中进行降维处理以保留与多肽序列最相关的深度特征,最后将降维后的深度特征与传统氨基酸组成(AAC)方法提取的手工特征融合并输入前馈神经网络中构建预测模型。验证实验结果表明:预测方法 Bitter-Fus 在 10 折交叉验证测试中获得了 0.902 的准确性和 0.805 的马修斯相关系数,在独立数据集测试中准确性和马修斯相关系数分别达到 0.930 和 0.862,明显优于当前最先进的苦味肽预测方法 BERT4Bitter 和 iBitter-SCM。

关键词:苦味肽;深度学习;特征提取;特征融合

中图分类号:TS201.2;TP399 文献标识码:A 文章编号:2096-1553(2023)03-0011-06

## 0 引言

食物经蛋白酶水解后,会生成不同风味的寡聚肽,也称风味肽。苦味肽作为一种重要的风味肽,主要来源于酪蛋白、玉米蛋白、火腿、热米粒、黄酒等<sup>[1]</sup>,一般由带有疏水侧链的氨基酸残基(≤8个)组成,其中疏水侧链与苦味受体结合产生苦味,而苦味程度随着氨基酸残基数量的增加而增强,但当氨基酸残基超过8个时,苦味程度不再继续增强<sup>[2-3]</sup>。苦味肽中的疏水性氨基酸主要包括亮氨酸、苯丙氨

酸和甘氨酸 3 种。其中,亮氨酸主要存在于豆类中,从大豆水解蛋白中分离出的苦味肽其 C 端通常都含有一个亮氨酸残基,同时,在发酵不充分的纳豆中也发现有部分苦味肽的 N 端和 C 端含有亮氨酸残基<sup>[4]</sup>。苯丙氨酸存在于大多数苦味肽中,其自身带有微弱的苦味,且其氨基和羧基末端酰基化修饰能进一步增强相应苦味肽的苦味<sup>[5]</sup>。人工合成的甘氨酸肽中有 3 种短肽(Gly-Leu、Val-Gly 和 Gly-Phe-Gly)具有苦味,也有一些含甘氨酸的短肽(如 Gly-Gly、Gly-Ala 和 Gly-Glu)不具有苦味<sup>[6]</sup>。由于苦味

收稿日期:2022-12-15;修回日期:2023-02-15;出版日期:2023-06-15

基金项目:国家自然科学基金青年科学基金项目(32101976);河南省科技攻关项目(232102210020);河南省高等学校青年骨干教师培养项目(2019GGJS132);河南省高等学校重点科研项目(22A520013,23B520004)

作者简介:杨学冬(1980—),男,河北省唐山市人,郑州轻工业大学实验师,主要研究方向为生物信息学、食品信息学。E-mail: yangxd@zzuli.edu.cn

通信作者:王晓(1982—),男,河南省卫辉市人,郑州轻工业大学副教授,博士,主要研究方向为生物信息学、食品信息学。 E-mail:wangxiao@zzuli.edu.cn 肽会降低食物的味觉品质,因而人类和大部分动物会本能地厌恶苦味食物。为了改善人们对食物的味觉体验,苦味肽及苦味抑制技术得到了国内外学者的广泛关注。另外,苦味肽通常具有抗菌、抗血栓、抗氧化、血管紧张素转化酶抑制活性等生物活性<sup>[7]</sup>,特别是其天然、安全、高效的广谱抗菌活性<sup>[8-9]</sup>,可作为抗生素的替代药物。因此,对苦味肽的精准鉴别对于食品和药物研究开发具有重要意义。

传统的湿实验方法鉴别苦味肽需要耗费大量的 人力、物力、财力和时间成本。随着后基因组时代的 来临,新产生的多肽序列呈指数级增长,湿实验方法 已无法满足快速鉴别苦味肽的需求。因此,迫切需 要开发基于计算技术的预测方法,以便快速精确地 区别苦味肽和非苦味肽。手工设计的特征提取方法 主要分为基于氨基酸组成(AAC)[10]的方法、基于氨 基酸物理化学属性的方法和基于序列进化信息的方 法。这些特征提取方法大都基于单一视角提取特 征,并不能全面刻画多肽序列的整体特性,故影响了 预测模型的鉴别能力。与手工设计的特征提取方法 不同,深度学习方法可以直接从原始多肽序列数据 中自动学习具有强鉴别能力的特征,已广泛应用于 生物信息学和计算生物学领域[11-15]。目前,仅有 iBitter-SCM<sup>[16]</sup>和 BERT4Bitter<sup>[17]</sup>这两种预测方法可 直接从多肽的氨基酸序列中鉴别苦味肽。其中, iBitter-SCM 方法通过手工设计的特征提取方法从多 肽序列中提取有鉴别能力的特征,进而结合传统机 器学习方法构建预测模型; BERT4Bitter 方法使用 BERT<sup>[18]</sup>预训练模型自动地从多肽序列中提取特 征,预测性能显著优于 iBitter-SCM 方法。

基于此,本文拟提出一种新颖的融合传统手工特征和预训练深度特征的深度学习方法 Bitter-Fus。该方法首先使用预训练蛋白质序列语言模型 Prot-Bert<sup>[19]</sup>自动地从多肽序列中提取深度学习特征,然后将该深度学习特征输入长短期记忆网络(LSTM)<sup>[20]</sup>中进行降维处理,以保留与多肽序列最相关的深度特征,最后将降维处理后的深度特征与基于氨基酸组成的手工特征融合后,输入前馈神经网络中,进而构建苦味肽预测模型实现苦味肽鉴别,并结合相关验证方法验证该预测方法的准确性。

## 1 相关方法与模型构建

#### 1.1 数据集

为了与已有方法进行公平比较,本文使用已有 研究中通用的数据集 BTP640<sup>[16-17]</sup>进行预测模型的 构建和评测,其中,BTP640 数据集包含 320 个苦味 肽和320个非苦味肽,已剔除重复的及包含 X、B、 U、Z 等不确定氨基酸残基的多肽序列。为了训练预 测模型和验证模型的泛化能力,BTP640 数据集以 8:2 的比例被随机分割为训练集 BTP-CV 和独立 测试集 BTP-TS。训练集 BTP-CV 包含 256 个苦味 肽和256个非苦味肽,而独立测试集BTP-TS包含 64 个苦味肽和 64 个非苦味肽。BTP640 数据集中 的苦味肽均从文献中手工收集且经过实验验证, 而非苦味肽数据集由于科学意义较小,其中只有 很小一部分经过实验验证。因此,在生物活性肽 序列数据库 BIOPEP<sup>[21]</sup>中随机生成与苦味肽数目 相同的多肽序列,将其作为非苦味肽数据集(前提 是它们不存在于苦味肽数据集中)用于训练预测 模型。这种将随机序列作为阴性数据集的方法使 用了先前研究中构建阴性数据集的标准过 程[22-23],该过程基于发现随机序列为正的概率非 常低的假设。

#### 1.2 多肽序列特征提取

1.2.1 传统手工特征提取 前期实验发现,与其他 手工设计的特征提取方法相比,AAC 方法的预测精度较好。因此,本文使用 AAC 方法提取多肽序列的 手工特征。氨基酸是组成多肽序列的基本物质,自 然界中共有 20 种基本氨基酸。多肽序列可看作是由不同数量基本氨基酸通过不同排列组合构成的。氨基酸组成则是指这些基本氨基酸在多肽序列中的 出现频率。假设 P 是任意一个多肽序列, $f(a_i)$  表示 氨基酸  $a_i(i=1,2,\cdots,20)$  在该多肽序列中出现的次数,则多肽序列 P 可表示为如下向量形式:

$$P = (p_1, p_2, \dots, p_i, \dots, p_{20})^{\mathrm{T}}$$
$$p_i = f(a_i) / \sum_{i=1}^{20} f(a_i)$$

式中 $,i=1,2,\cdots,20,p_i$  表示氨基酸  $a_i$  在多肽序列中的出现频率。

1.2.2 预训练深度特征提取 BERT 的模型架构

是一个多层双向变压器编码器,每层有 12 个或 24 个编码器块,分别用于 BERT 基带和 BERT 大型模型。其中,每个编码器块由一个多头自注意力子层和一个完全连接的前馈子层组成,在这两个子层周围部署剩余连接,进行层规范化。ProtBert 是 A. Elnaggar 等提出的预训练蛋白质序列语言模型,通过使用 UniRef100 和 BFD 这 2 个数据库的蛋白质序列微调原始 BERT 模型而获得,可直接从原始多肽序列数据中自动学习到具有强鉴别能力的特征。其中,UniRef100 数据库是一个广泛使用的参考蛋白质序列数据库,BFD 数据库合并了 UniProt 数据库中可使用的所有蛋白质序列及从多个元基因组测序项目翻译的蛋白质序列。本文使用预训练蛋白质序列语言模型 ProtBert 提取深度学习特征。

# 1.3 融合传统手工特征和预训练深度特征的 预测方法

融合传统手工特征和预训练深度特征的苦味肽 预测模型 Bitter-Fus 的整体架构如图 1 所示。首先, 从输入多肽序列中提取潜在特征。根据 BERT 模型 的要求,首先输入多肽序列标记化,即将字符 "[CLS]"和"[SEP]"添加到整个多肽序列的开头 和结尾,并添加位置编码。然后,将生成的向量输入 预训练蛋白质序列语言模型 ProtBert 中以提取深度 学习特征。经过 ProtBert 模型训练,每个输入的标 记(即每个氨基酸)创建了上下文感知嵌入,也就是 特征,输入的多肽序列被编码为1024维的高维特征 向量。接着,将获得的1024维的高维深度学习特征 向量输入长短期记忆(LSTM)网络中。LSTM 网络 基于多肽序列,对输入的特征信息进行提取,并对其 中重要的特征进行选择及降维。本文使用网格搜索 的方法确定 LSTM 网络的输出层参数,即对高维特 征降维后的维度。在实验中,将输出层维度分别设 置为 128、256、512、640 和 768 后进行训练,根据 10 折交叉验证的平均结果可知,输出层为512维时的 模型预测效果最佳。故通过 LSTM 网络降维,最终 获得 512 维的由 ProtBerT 模型提取的深度特征。随 后.将该深度特征与传统 AAC 方法提取的 20 维手 工特征进行特征融合,得到532维的融合特征。最 后,将该融合特征输入前馈神经网络(使用 Sigmoid

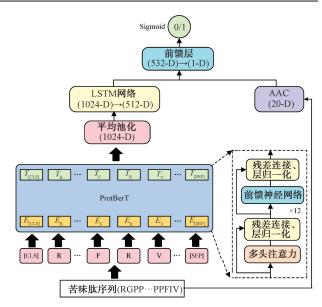


图 1 融合传统手工特征和预训练深度特征的 苦味肽预测模型 Bitter-Fus 的整体架构

Fig. 1 The overall architecture of a bitter peptide prediction model Bitter-Fus by fusing handcrafted features and pretrained deep features

分类),建立苦味肽分类预测模型。

#### 2 实验验证与分析

#### 2.1 评价指标

在苦味肽预测领域,通常采用以下 4 个评价指标定量评估预测模型的预测性能:准确性(Acc)、马修斯相关系数(MCC)、敏感性(Sn)和特异性(Sp)。其中,Acc 和 MCC 从总体分别反映预测模型的准确率和稳定性,而 Sn 和 Sp 则从局部分别反映预测模型对正类和负类样本的预测准确率。这 4 个评价指标的计算方法如下式所示:

$$Sn = 1 - \frac{N_{-}^{+}}{N^{+}} \qquad 0 \le Sn \le 1$$

$$Sp = 1 - \frac{N_{-}^{+}}{N^{-}} \qquad 0 \le Sp \le 1$$

$$Acc = 1 - \frac{N_{-}^{+} + N_{-}^{-}}{N^{+} + N^{-}} \qquad 0 \le Acc \le 1$$

$$MCC = \frac{1 - \left(\frac{N_{-}^{+}}{N^{+}} + \frac{N_{-}^{+}}{N^{-}}\right)}{\sqrt{\left(1 + \frac{N_{-}^{-} - N_{-}^{+}}{N^{+}}\right)\left(1 + \frac{N_{-}^{+} - N_{-}^{-}}{N^{-}}\right)}} \qquad -1 \le MCC \le 1$$

式中,*N*<sup>+</sup>表示苦味肽样本总数;*N*<sup>+</sup>表示被错误预测 为非苦味肽的苦味肽样本数量:*N*<sup>-</sup>表示非苦味肽样 本总数; $N_{+}$ 表示被错误预测为苦味肽的非苦味肽样本数量。

当  $N_{-}^{+}=0$ ,即没有苦味肽样本被错误预测为非苦味肽时,预测模型的 Sn=1。当  $N_{-}^{+}=N_{-}^{+}$ ,即所有苦味肽样本被错误预测为非苦味肽时,预测模型的 Sn=0。类似地,当  $N_{+}=0$ ,即没有非苦味肽样本被错误预测为苦味肽时,预测模型的 Sp=1。当  $N_{+}^{-}=N_{-}^{-}$ ,即所有非苦味肽样本被错误预测为苦味肽时,预测模型的 Sp=0。当  $N_{-}^{+}=N_{-}^{-}=0$ ,即所有苦味肽和非苦味肽样本都未被错误预测时,预测模型的 Acc=1,MCC=1。当  $N_{-}^{+}=N_{-}^{+}=1$ ,即所有苦味肽和非苦味肽样本都被错误预测时,预测模型的 Acc=0,MCC=1。当  $N_{-}^{+}=N_{-}^{+}/2$  且  $N_{+}^{-}=N_{-}/2$  时,预测模型的 Acc=0,MCC=0.5,MCC=0,意味着模型的预测结果与随机猜测相当。

#### 2.2 交叉验证

交叉验证是一种广泛用于评估机器模型并测试 其性能的方法。在预测分析中,10 折交叉验证测 试、独立数据集测试和 jackknife 测试是 3 种常用的 检验模型预测性能的交叉验证方法。本文采用 10 折交叉验证测试和独立数据集测试对模型进行评 估。在 10 折交叉验证测试中,首先将数据集拆分为 10 个相等的部分,依次选择其中 9 折作为训练集以 训练模型,剩余 1 折作为测试集以验证模型,随后保 存其验证结果。最后将获得的 10 次验证结果平均 值作为对模型性能的估计值。泛化能力决定了预测 模型对新输入的适应能力,而检验深度学习模型的 泛化能力非常重要,故本文使用独立数据集 BTP-TS 对模型进行独立测试,以评估模型的泛化能力和稳 定性。

#### 2.3 本文预测方法与已有预测方法的比较

为了公平比较,笔者在 BTP-CV 数据集上采用 10 折交叉验证方法对本文所提预测方法 Bitter-Fus 与 iBitter-SCM 预测方法和 BERT4Bitter 预测方法的 预测性能进行比较分析,3 种预测方法的 10 折交叉验证结果见表 1。由表 1 可知,与 BERT4Bitter 预测方法相比,Bitter-Fus 预测方法的 Acc 和 MCC 分别提高了 4.8%和 10.9%,Sn 和 Sp 分别提高了 6.2%和 3.4%。与 iBitter-SCM 预测方法相比,Bitter-Fus 预测方法的 Acc 和 MCC 分别提高了 3.6%和 7.2%,Sn 和 Sp 分别提高了 1.0%和 6.6%。10 折交叉验证结

表 1 3 种预测方法的 10 折交叉验证结果 Table 1 10-fold cross-validation results of 3 prediction methods

预测方法	Acc	MCC	Sn	Sp
iBitter-SCM	0.871	0.751	0. 913	0. 828
BERT4Bitter	0.861	0.726	0.868	0.854
Bitter-Fus	0.902	0.805	0.922	0.883

果表明,本文提出的预测方法 Bitter-Fus 在 4 个性能指标上都表现优异。

为了进一步评估本文所提预测方法 Bitter-Fus 的泛化性能,在独立数据集 BTP-TS 上将其与 iBitter-SCM 预测方法和 BERT4Bitter 预测方法进行比较分析。3 种预测方法的独立数据集测试结果见表2。由表2可知,Bitter-Fus 预测方法在 Acc、MCC 和 Sp 这3 项指标上均优于 BERT4Bitter 预测方法,其中,Sp 提高最多,为7.0%,Acc 和 MCC 分别提高了0.9%和2.1%。尽管 Bitter-Fus 预测方法的 Sn 略低于 BERT4Bitter 预测方法,但明显高于 iBitter-SCM 预测方法。考虑到独立数据集测试是最严格的交叉验证方法,较高的 Acc 和 MCC 更可取,由此表明,本文所提预测方法更有效也更稳定。综上所述,本文所提预测方法 Bitter-Fus 的整体性能优于目前最先进的苦味肽预测方法 BERT4Bitter 和 iBitter-SCM,且更精准和稳定。

表 2 3 种预测方法的独立数据集测试结果
Table 2 The results of 3 prediction methods tested
in independent data sets

预测方法	Acc	MCC	Sn	Sp
iBitter-SCM	0. 844	0. 688	0.844	0.844
BERT4Bitter	0. 922	0.844	0. 938	0.906
Bitter-Fus	0. 930	0.862	0.891	0. 969

# 3 结论

本文针对苦味肽的精准鉴别问题,提出了一种新颖的融合传统手工特征和预训练深度特征的预测方法。该方法首先使用预训练蛋白质语言模型ProtBert 提取多肽序列的深度学习特征,并与传统手工特征相融合,然后使用 LSTM 网络进行特征降维处理,保留与多肽序列最相关的特征以使模型整体预测性能大幅度提高。10 折交叉验证测试和独立数据集测试结果表明,该方法的 Acc 分别达到

0.902 和 0.930, MCC 分别达到 0.805 和 0.862, 且 与当前最先进的苦味肽鉴别方法 BERT4Bitter 和 iBitter-SCM 相比, Acc、MCC、Sn 和 Sp 这 4 项性能指 标均有明显提高。在苦味肽的鉴别预测领域,本文 所提预测方法有望成为一种有潜力的预测工具。此 外,将传统手工特征与 ProtBert 深度学习特征融合 的特征工程模块也可应用于其他基于蛋白质序列的 预测任务中,并可使下游预测模型获得更全面的序 列信息,进而使预测模型得以更充分的训练从而获 得更好的预测效果。

#### 参考文献:

- [1] 刘桃妹. 风味增强肽及其应用研究[J]. 江苏调味副食品,2014(3):5-8.
- [2] 王知非,林璐,孙伟峰,等. 苦味肽和苦味受体研究进展[J]. 中国调味品,2016,41(9):152-156.
- [3] 毕继才,崔震昆,张令文,等. 苦味传递机制与苦味肽研究进展[J].食品工业科技,2018,39(11):333-338.
- [4] 司阔林,李志国,李玲玉,等.干酪苦味肽的形成及脱苦方法研究进展[J].食品工业,2021,42(3):267-271
- [5] 郭兴峰,魏芳,周祥山,等. 苦味肽的形成机理及脱苦技术研究进展[J]. 食品研究与开发,2017,38(21): 207-211.
- [6] 应欣,张连慧,陈卫华.蛋白水解物苦味形成、评价及功能活性的研究进展[J].中国粮油学报,2017,32(12):141-146.
- [7] 邓尚贵,余妙灵,甄兴华,等. 苦味肽抗氧化活性延长食品保鲜[J]. 食品安全质量检测学报,2020,11(2):375-380.
- [8] 杨保军,梁琪,宋雪梅.基于计算机虚拟技术研究牦牛乳硬质干酪苦味肽的抑菌活性差异[J].食品与生物技术学报,2021,40(12):75-87.
- [9] 杨保军,梁琪,宋雪梅. 牦牛乳干酪苦味肽 ACE 抑制活性表征的分子机制[J]. 中国食品学报,2022,22(5):8-17.
- [10] CHOU K C. Using amphiphilic pseudo amino acid composition to predict enzyme subfamily classes [J]. Bioinformatics, 2005, 21(1):10-19.
- [11] BO W C, QIN D Y, ZHENG X, et al. Prediction of bitterant and sweetener using structure-taste relationship models based on an artificial neural network [J]. Food Research International, 2022, 153:110974.
- [12] HUANG W K, SHEN Q C, SU X B, et al. BitterX: A tool for understanding bitter taste in humans [J]. Scientific

- Reports, 2016, 6(1):23450.
- [13] ZHANG Y Q, ZHU G C, LI K W, et al. HLAB; Learning the BiLSTM features from the ProtBert-encoded proteins for the class I HLA-peptide binding prediction [J]. Briefings in Bioinformatics, 2022, 23(5); bbac173.
- [14] WANG Z F, LEI X J. Prediction of RBP binding sites on circRNAs using an LSTM-based deep sequence learning architecture [J]. Briefings in Bioinformatics, 2021, 22(6): bbab342.
- [15] GUO Y C, YAN K, LYU H W, et al. PreTP-EL; Prediction of therapeutic peptides based on ensemble learning [J]. Briefings in Bioinformatics, 2021, 22(6); bbab358.
- [16] CHAROENKWAN P, YANA J, SCHADUANGRAT N, et al. iBitter-SCM: Identification and characterization of bitter peptides using a scoring card method with propensity scores of dipeptides [J]. Genomics, 2020, 112(4):2813– 2822.
- [17] CHAROENKWAN P, NANTASENAMAT C, HASAN M M, et al. BERT4Bitter: A bidirectional encoder representations from transformers (BERT)-based model for improving the prediction of bitter peptides [J]. Bioinformatics, 2021,37(17):2556-2562.
- [18] DEVLIN J, CHANG M W, LEE K, et al. BERT: Pre-training of deep bidirectional transformers for language understanding [C]//Association for Computational Linguistics. Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies. Minneapolis, Minnesota: Association for Computational Linguistics, 2019: 4171–4186.
- [19] ELNAGGAR A, HEINZINGER M, DALLAGO C, et al. ProtTrans: Towards cracking the language of lifes code through self-supervised deep learning and high performance computing [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2022, 44(10):7112-7127.
- [20] HOCHREITER S, SCHMIDHUBER J. Long short-term memory [J]. Neural Computation, 1997, 9 (8): 1735 1780.
- [21] MINKIEWICZ P, DZIUBA J, IWANIAK A, et al. BIOPEP database and other programs for processing bioactive peptide sequences [J]. Journal of AOAC International, 2008, 91(4):965-980.
- [22] GAUTAM A, CHAUDHARY K, KUMAR R, et al. In silico approaches for designing highly effective cell penetrating peptides [J]. Journal of Translational Medicine, 2013, 11 (1):74.
- [23] KUMAR R, CHAUDHARY K, CHAUHAN J S, et al. An in silico platform for predicting, screening and designing of antihypertensive peptides [J]. Scientific Reports, 2015, 5(1):12512.

# An accurate identification method of bitter peptides based on deep learning

YANG Xuedong<sup>1</sup>, HAN Lijun<sup>1</sup>, WANG Rong<sup>1</sup>, WANG Hongwei<sup>2</sup>, WANG Xiao<sup>1,3</sup>

- 1. College of Computer and Communication Engineering , Zhengzhou University of Light Industry , Zhengzhou 450001 , China;
  - 2. College of Food and Bioengineering, Zhengzhou University of Light Industry, Zhengzhou 450001, China;
    - 3. Henan Key Laboratory of Food Safety Data Intelligence, Zhengzhou 450001, China

Abstract: Given that wet experimental methods were no longer adequate for the rapid identification of bitter peptides, this paper presented Bitter-Fus, a novel predictive deep learning method incorporating traditional manual features and pre-trained deep features. Firstly, the method automatically extracted deep learning features from peptide sequences using a pre-trained protein sequence language model, then fed the deep learning features into a long short-term memory (LSTM) network for dimensionality reduction to retain the most relevant features. Finally, the reduced-dimensional deep features were fused with the manual features composed of traditional amino acids composition (AAC) method and passed into the feedforward neural network to construct a prediction model. The validation experimental results showed that the prediction method Bitter-Fus obtained an accuracy precision value of 0.902 and a Mathews correlation coefficient value of 0.805 in a 10-fold cross-validation, and an accuracy precision value of 0.930 and a Mathews correlation coefficient value of 0.862 in the independent dataset test, which significantly outperformed the current state-of-the-art bitter peptide prediction methods BERT4Bitter and iBitter-SCM.

**Key words:** bitter peptide; deep learning; feature extraction; feature fusion

(责任编辑:杨晓娟)

# 本刊数字网络传播声明

本刊已许可中国知网,万方数据资源系统,维普网,国家科技学术期刊开放平台,博看网,超星,中国科技论文在线,中教数据库,CA,EBSCOhost,Ulrichs,FSTA等在其系列数据库产品中以数字化方式复制、汇编、发行、信息网络传播本刊全文。其相关著作权使用费与本刊稿酬一并支付。作者向本刊提交文章发表的行为即视为同意我刊上述声明。