

分类号: TP3  
学号: 20232109002

密级:  
单位代码: 10759

# 石河子大学

## 硕士学位论文



### 基于深度学习的小麦籽粒蛋白质含量定量分析方法研究

学位申请人	赵永航
指导教师	李栓明 副教授 张勃 高级工程师
申请学位门类级别	专业硕士
学科、专业名称	电子信息
研究方向	新一代信息技术
所在学院	机械电气工程学院

中国·新疆·石河子  
2026年5月

分类号: TP3  
学号: 20232109002

密级:  
单位代码: 10759

# 石河子大学

## 硕士学位论文

### 基于深度学习的小麦籽粒蛋白质含量定量分析方法研究

学位申请人	赵永航
指导教师	李栓明 副教授 张勃 高级工程师
申请学位门类级别	专业硕士
学科、专业名称	电子信息
研究方向	新一代电子信息技术
所在学院	机械电气工程学院

中国·新疆·石河子  
2026年5月

**Research on a Deep Learning Based Method for Quantitative  
Analysis of Wheat Grain Protein Content**

A Dissertation Submitted to

**Shihezi University**

In Partial Fulfillment of the Requirements

for the Degree of

**Master of Engineering**

By

**Zhao Yonghang**

**(Electronic Information)**

Dissertation Supervisor: Associate Prof. Li Shuanming

May, 2026

# 石河子大学学位论文独创性声明及使用授权声明

## 学位论文独创性声明

本人所提交的学位论文是在我导师的指导下进行的研究工作及取得的研究成果。据我所知，除文中已经注明引用的内容外，本论文不包含其他个人已经发表或撰写过的研究成果。对本文的研究做出重要贡献的个人和集体，均已在文中作了明确的说明并表示谢意。

研究生签名：赵永航

时间：2026年5月20日

## 使用授权声明

本人完全了解石河子大学有关保留、使用学位论文的规定，学校有权保留学位论文并向国家主管部门或指定机构送交论文的电子版和纸质版。有权将学位论文在学校图书馆保存并允许被查阅。有权自行或许可他人将学位论文编入有关数据库提供检索服务。有权将学位论文的标题和摘要汇编出版。保密的学位论文在解密后适用本规定。

研究生签名：赵永航

时间：2026年5月20日

导师签名：李柱明

时间：2026年5月20日

## 摘要

小麦是全球重要的粮食作物之一，其籽粒蛋白质含量是评价小麦营养品质和加工品质的关键指标。传统蛋白质检测方法主要依赖化学分析或精密仪器，存在检测周期长、操作复杂且对样品具有破坏性等问题，难以满足现代农业生产对快速、无损检测技术的迫切需求。近红外光谱技术因具有快速、无损、绿色环保等优点，在农产品品质检测领域展现出广阔的应用前景。然而，近红外光谱数据通常呈现高维度、变量间强共线性以及复杂非线性关系等特点，传统建模方法在深层特征挖掘方面存在一定局限。因此，探索融合深度学习的近红外光谱分析技术，对提高小麦蛋白质含量预测精度具有重要意义。本研究以小麦籽粒近红外光谱数据为研究对象，围绕样本划分、数据增强、光谱预处理、特征变量筛选以及深度学习建模等关键问题开展系统研究，构建小麦蛋白质含量预测模型，并开发基于深度学习的光谱指标预测系统。主要研究内容与成果如下：

(1) 针对样本量有限的问题，首先在训练集上，引入基于梯度惩罚的 Wasserstein 生成对抗网络 (WGAN-GP) 进行数据增强，通过学习原始样本的概率分布生成具有相似统计特性的合成数据。通过系统分析不同训练轮次下生成光谱的特征空间分布、目标变量分布以及 FID、MMD、SWD 等多指标变化趋势，确定模型在 3500-5000epochs 区间内达到最优性能。进一步通过对比试验，6 倍数据扩充方案在精度与稳定性之间取得最佳平衡，有效提升了后续建模的稳定性与泛化能力。

(2) 针对原始光谱中存在的噪声干扰、基线漂移及散射效应等问题，系统比较了多种预处理方法的效果。采用特征波段筛选相关算法提取关键特征波段。结果表明，筛选得到的特征波段主要集中在 1100-1400 nm、1500-1900 nm 和 1900-2500 nm 区域，与蛋白质中 N-H、C-H 等官能团的振动吸收特征密切相关，为后续模型构建提供了有效的光谱信息，并利用偏最小二乘回归 (PLSR) 进行了初步建模验证。

(3) 在深度学习方面，一维卷积神经网络 (1DCNN)、CNN-LSTM、CNN-LSTM-Attention 以及增强型 CNN-LSTM-Attention 等多种模型的预测性能。结果表明，增强型 CNN-LSTM-Attention 模型通过融合多尺度卷积、双向 LSTM、自适应融合注意力机制、残差连接及层归一化策略，实现了对光谱数据多层次特征的协同建模，性能在所有模型中表现最优。其中，在 SG-MS-C-UV 特征组合条件下，模型测试集  $R^2$  为 0.8939，RPD 值达到 4.67，显示出极高的预测精度与泛化能力。同时，引入 SHAP 方法对最优模型进行可解释性分析，揭示了关键波长变量及其交互作用对预测结果的贡献。

**关键词：**近红外光谱；小麦籽粒；蛋白质含量；深度学习；生成对抗网络



## Abstract

Wheat is one of the most important food crops in the world, and its grain protein content is a key indicator for evaluating wheat nutritional quality and processing quality. Traditional protein detection methods mainly rely on chemical analysis or precision instruments, which suffer from long detection cycles, complex operations, and sample destructiveness, making it difficult to meet the urgent demand for rapid and non-destructive testing in modern agricultural production. Near-infrared spectroscopy technology, with its advantages of speed, non-destructiveness, and environmental friendliness, has shown broad application prospects in the field of agricultural product quality detection. However, near-infrared spectral data typically exhibit characteristics such as high dimensionality, strong multicollinearity among variables, and complex nonlinear relationships, while traditional modeling methods have certain limitations in deep feature extraction. Therefore, exploring near-infrared spectroscopy analysis techniques integrated with deep learning is of great significance for improving the prediction accuracy of wheat protein content. This study takes wheat grain near-infrared spectral data as the research object, systematically investigates key issues including sample partitioning, data augmentation, spectral preprocessing, feature variable selection, and deep learning modeling, constructs a wheat protein content prediction model, and develops a deep learning-based spectral indicator prediction system. The main research contents and achievements are as follows:

(1) To address the problem of limited sample size, the Wasserstein Generative Adversarial Network with Gradient Penalty (WGAN-GP) was introduced for data augmentation on the training set, generating synthetic data with similar statistical characteristics by learning the probability distribution of original samples. Through systematic analysis of the feature space distribution, target variable distribution, and multiple indicators such as FID, MMD, and SWD of generated spectra under different training epochs, the model was determined to achieve optimal performance within the 3500-5000 epochs interval. Further comparative experiments showed that the 6-fold data augmentation scheme achieved the best balance between accuracy and stability, effectively enhancing the stability and generalization ability of subsequent modeling.

(2) To address issues such as noise interference, baseline drift, and scattering effects in raw spectra, the effects of various preprocessing methods were systematically compared. Feature band selection algorithms were employed to extract key characteristic bands. The results indicated that the selected characteristic bands were mainly concentrated in the regions of 1100-1400 nm, 1500-1900 nm, and 1900-2500 nm, which

are closely related to the vibrational absorption characteristics of functional groups such as N-H and C-H in proteins, providing effective spectral information for subsequent model construction. Preliminary modeling validation was conducted using Partial Least Squares Regression (PLSR).

(3) In terms of deep learning, the predictive performance of multiple models including One-Dimensional Convolutional Neural Network (1DCNN), CNN-LSTM, CNN-LSTM-Attention, and Enhanced CNN-LSTM-Attention was compared. The results showed that the Enhanced CNN-LSTM-Attention model, which integrates multi-scale convolution, bidirectional LSTM, adaptive fusion attention mechanism, residual connections, and layer normalization strategies, achieves collaborative modeling of multi-level features in spectral data, demonstrating the best performance among all models. Notably, under the SG-MSU-UVE feature combination condition, the model achieved a test set  $R^2$  of 0.8939 and an RPD value of 4.67, demonstrating extremely high prediction accuracy and generalization capability. Meanwhile, the SHAP method was introduced for interpretability analysis of the optimal model, revealing the contribution of key wavelength variables and their interactions to the prediction results.

**Key words:** Near-infrared spectroscopy; Wheat grains; Protein content; Deep learning; Generative adversarial network

# 目 录

摘要 .....	I
Abstract .....	III
第 1 章 绪论 .....	1
1.1 研究背景 .....	1
1.2 国内外研究现状 .....	2
1.2.1 光谱预处理与特征筛选 .....	2
1.2.2 基于深度学习的近红外光谱定量分析方法 .....	3
1.2.3 基于小样本的深度学习策略研究 .....	5
1.3 研究内容与技术路线 .....	6
1.3.1 研究内容 .....	6
1.3.2 技术路线 .....	8
第 2 章 数据增强算法 .....	9
2.1 数据集详情 .....	9
2.2 数据集划分策略 .....	10
2.2.1 数据集划分方法介绍 .....	10
2.2.2 本文所采用数据集划分方法 .....	12
2.3 基于梯度惩罚的 Wasserstein 生成对抗网络 .....	13
2.3.1 数据归一化 .....	13
2.3.2 WGAN-GP .....	14
2.3.2.1 生成对抗网络基本原理 .....	14
2.3.2.2 Wasserstein 距离与 WGAN .....	14
2.3.2.3 梯度惩罚机制 .....	15
2.3.3 模型参数设置 .....	15
2.3.3.1 网络结构与参数初始化 .....	16
2.3.3.2 优化器与训练参数 .....	17
2.3.3.3 损失函数与约束条件 .....	17
2.3.4 结果分析 .....	18
2.3.4.1 光谱生成收敛过程分析 .....	18
2.3.4.2 特征空间分布演化分析 .....	19
2.3.4.3 目标变量分布收敛分析 .....	20

2.3.4.4	最优训练轮次综合评估	22
2.3.4.5	最优数据扩充方案选择	24
2.4	本章小结	25
第 3 章	光谱预处理与特征波段筛选	26
3.1	光谱预处理	26
3.1.1	SG 平滑算法	26
3.1.2	多元散射校正	27
3.1.3	标准正态变量变换	28
3.1.4	去趋势法	29
3.1.5	导数光谱法	30
3.1.6	生成光谱预处理	32
3.2	特征变量选择分析	33
3.2.1	UVE 变量选择	33
3.2.2	CARS 变量选择	34
3.2.3	SPA 变量选择	35
3.3	偏最小二乘回归	36
3.4	本章小结	40
第 4 章	卷积神经网络的小麦籽粒蛋白质含量预测模型构建	42
4.1	引言	42
4.2	基于一维卷积神经网络的小麦蛋白质含量预测模型构建	42
4.2.1	一维卷积神经网络模型原理	42
4.2.2	一维卷积神经网络模型结构	43
4.2.3	模型参数设置及训练	44
4.3	基于 CNN-LSTM 的小麦蛋白质含量预测模型构建	48
4.3.1	LSTM 原理	48
4.3.2	CNN-LSTM 模型结构	49
4.3.3	模型参数设置及训练	49
4.4	基于 CNN-LSTM-Attention 的小麦蛋白质含量预测模型构建	54
4.4.1	模型结构	54
4.4.2	基于加权平均的自注意力机制	55
4.4.3	模型参数设置及训练	55
4.5	增强型 CNN-LSTM-Attention 模型构建	60
4.5.1	模型原理	60
4.5.2	模型结构	60

4.5.2.1 输入层 .....	61
4.5.2.2 多尺度 CNN 特征提取模块 .....	61
4.5.2.3 深度可分离卷积模块 .....	61
4.5.2.4 双向 LSTM 序列建模层 .....	62
4.5.2.5 自适应融合注意力机制 .....	62
4.5.2.6 残差连接与层归一化 .....	62
4.5.2.7 输出层 .....	63
4.5.3 动态加载机制 .....	63
4.5.4 模型训练及分析 .....	63
4.6 SHAP 值分析与模型可解释性 .....	67
4.7 本章小结 .....	69
第 5 章 基于卷积神经网络的小麦籽粒蛋白质含量定量分析系统构建 .....	71
5.1 引言 .....	71
5.2 系统总体架构设计 .....	71
5.3 系统开发环境 .....	72
5.4 系统功能测试与应用 .....	73
5.4.1 系统功能测试 .....	73
5.4.2 系统应用 .....	73
5.4.2.1 系统登录 .....	74
5.4.2.2 光谱数据导入 .....	74
5.4.2.3 模型预测 .....	75
5.4.2.4 预测结果可视化 .....	75
5.4.2.5 预测结果导出 .....	76
5.5 本章小结 .....	77
第 6 章 总结与展望 .....	78
6.1 研究总结 .....	78
6.2 研究展望 .....	79
参考文献 .....	80
致谢 .....	85
作者简介 .....	86



## 第 1 章 绪论

### 1.1 研究背景

蛋白质是构成人体细胞与组织的基础，是生命活动的重要物质载体。小麦作为全球约 60%人口的主粮，为人类提供了约 19%的膳食热量和 21%的蛋白质<sup>[1]</sup>，在全球粮食安全和优质蛋白供给体系中具有重要的战略地位<sup>[2]</sup>。研究表明，由于消费基数庞大，谷物（尤其是小麦）已成为人类膳食中最重要的蛋白质来源之一，在资源利用效率和环境可持续性方面具有显著优势<sup>[3]</sup>。

小麦是人类植物性蛋白的重要来源，其籽粒蛋白质主要由清蛋白、球蛋白、麦胶蛋白和麦谷蛋白四类组成<sup>[4]</sup>。其中，麦胶蛋白与麦谷蛋白共同构成面筋网络，赋予面团良好的粘弹性，从而决定面团的流变学特性及最终的加工品质。因此，小麦蛋白质的含量及其组成结构直接影响产品用途（如面包、面条和糕点等），不同加工产品对蛋白质含量的需求存在显著差异<sup>[5]</sup>。

从营养学角度来看，小麦籽粒蛋白质含量通常占干重的 10%–15%<sup>[6]</sup>。此外，小麦还富含 B 族维生素及铁、锌等微量元素，是兼具较高蛋白含量与营养密度的重要粮食资源<sup>[2]</sup>。由于蛋白质含量不仅是衡量营养品质的核心指标，也是划分小麦等级（强筋、中筋、弱筋）的关键依据<sup>[4]</sup>，并直接影响面粉加工性能与市场价值，因此，实现其快速、准确检测对于保障粮食安全及推动优质小麦产业发展具有重要意义<sup>[7]</sup>。

传统小麦蛋白质含量检测方法通常依赖昂贵的精密仪器，并需要对样品进行破坏性处理，不仅检测流程复杂、耗时较长，而且难以满足现代农业对大规模、高效率品质检测的需求。为在保证样品完整性的同时实现高效检测，融合光谱技术与信息化技术的无损检测方法逐渐成为研究热点。当前常见的无损检测技术包括介电特性检测<sup>[8]</sup>、声学检测<sup>[9]</sup>、电子鼻与电子舌<sup>[10]</sup>、核磁共振<sup>[11]</sup>、机器视觉<sup>[12]</sup>及光学检测技术<sup>[13]</sup>等。其中，光学检测技术因具有非接触、高效率、无损伤及易于自动化等优点，在农产品品质检测领域展现出重要应用价值<sup>[14]</sup>。

在光学检测技术中，高光谱技术能够在可见光至近红外范围内获取连续光谱信息，这些信息能够反映小麦籽粒的化学成分及物理特性。通过对光谱数据进行分析，可以实现对小麦蛋白质含量的定量预测。然而，传统光谱分析方法在处理高维数据时往往依赖人工特征提取和线性建模，难以充分挖掘光谱数据中的复杂非线性关系，这在一定程度上限制了模型的预测精度与泛化能力。

## 1.2 国内外研究现状

### 1.2.1 光谱预处理与特征筛选

光谱分析方法主要基于光与物质的相互作用原理，利用物质的光谱对物质进行定性、定量和结构分析<sup>[15]</sup>。通过光谱曲线，可以反映出农作物的相关特征，光谱分析对农产品品质检测方面意义重大<sup>[13]</sup>。

光谱分析技术提供了关于各种扫描样本的光谱成分的完整信息，主要基于波长范围750至2500nm内电磁辐射的吸收<sup>[16]</sup>。光在待测样品中传输时，会发生吸收、反射和折射等现象，从而使光束发生变化产生光谱曲线。光在组织中吸收的主要成分与组织的化学组成成分有关，例如，在可见/近红外波段，光的吸收主要是由分子中C-H, N-H, O-H等含氢基团振动的倍频及合频产生，而含氢基团可以反应待测样品的化学性质，因此光谱曲线可以反映出农产品的含水量、蛋白质含量或者可溶性固形物含量等特征<sup>[13]</sup>。得益于化学计量学理论和方法的不断发展，近红外光谱（near-infrared, NIR）技术得到了显著进步<sup>[17]</sup>。

王旭等<sup>[18]</sup>用傅里叶变换近红外漫反射法测定单粒小麦种子的蛋白质含量，结果表明，相关系数 $>0.98$ ，可满足单粒小麦蛋白质的快速检测。

赖立群等<sup>[19]</sup>以一阶导数预处理方法分析处理近红外光谱数据，结果表明，该方法可以有效解决样品在检测时受到其他组分和背景干扰的问题。

Antonov等<sup>[20]</sup>将适应于紫外-可见分光光度法的逐步滤波方法应用到近红外光谱分析领域，改进了Savitzky-Golay (SG)方法避免了光谱在波长尺度上的导数衰减的问题，一定程度上解决了信噪比退化的问题。

光谱预处理的核心目标在于削弱或消除由非化学因素引入的系统性变异，从而增强与目标组分相关的有效信号。根据其作用机制，主流预处理方法可分为三大类。第一类为散射校正，旨在消除由样品颗粒度、密度不均等因素引起的乘性散射效应。其中，多元散射校正 (MSC) 和标准正态变量变换 (SNV) 是最为经典和广泛使用的技术。研究表明，在对模糊棉籽油含量进行快速测定时，采用SNV预处理能显著提升模型性能<sup>[21]</sup>；而在烟草叶片水分含量预测中，MSC被证明是四种对比方法中最有效的预处理手段<sup>[22]</sup>。第二类为基线/趋势校正，主要用于抑制加性基线漂移并增强峰形分辨能力。一阶导数 (1st-Derivative) 和二阶导数 (2nd-Derivative) 通过数学微分操作，能够有效消除恒定或线性基线偏移，并锐化重叠峰。Savitzky-Golay (SG) 平滑微分法则在提供导数信息的同时兼具降噪功能，被证实是高地大麦水分、 $\beta$ -葡聚糖和支链淀粉含量PLS预测模型的最优预处理方法<sup>[23]</sup>。第三类为降噪，其目的是在保留光谱整体轮廓的前提下抑制高频随机噪声，移动平均 (Moving Average) 和小波去噪 (Wavelet Denoising) 是常用策略。

在荔枝糖分的无损检测研究中，系统比较了包括移动平均在内的九种预处理方法，以寻找最优组合<sup>[24]</sup>。

在完成预处理后，特征选择（或称波长选择）成为进一步优化模型的关键步骤。其核心思想是从全谱中识别出对目标组分响应最敏感、鲁棒性最强的特征子集，从而减少冗余变量、降低模型复杂度、提高计算效率并增强模型的泛化能力。特征选择方法主要分为三类：过滤法（Filter）、包装法（Wrapper）和嵌入法（Embedded）。过滤法依据变量自身的统计特性进行筛选，计算高效但忽略了变量间的交互作用。例如，基于皮尔逊相关系数和互信息的集成评价函数 CONMI 被提出，以同时考量线性和非线性关系，用于分类任务中的特征选择<sup>[25]</sup>。包装法将特征子集的选择过程与特定预测模型的性能直接挂钩，通过迭代搜索（如遗传算法 GA、模拟退火 SA）来寻找最优子集，虽计算成本高但通常能获得更优的预测性能。Fan 等人利用遗传算法进行特征波长选择，成功应用于 NIR 光谱建模<sup>[26]</sup>；而针对鱼粉质量评估，研究者则采用了频率自适应二进制差分进化算法来选择特征波段<sup>[27]</sup>。嵌入法在模型训练过程中自动执行特征选择，如 LASSO 回归通过 L1 正则化实现变量收缩与选择。在大豆种子品种鉴别研究中，一种集成方法结合了 PLS-DA、随机森林（Random Forest, RF）和 LASSO 的重要性排序，实现了高精度的品种识别<sup>[28]</sup>。此外，Max-Relevance Min-Redundancy（mRMR）作为一种先进的过滤法，通过最大化所选特征与目标的相关性，同时最小化特征间的冗余性，在 NIR 波长选择中展现出优越性能<sup>[29]</sup>。

现阶段，多种线性和非线性化学计量学模型已被应用于光谱数据分析中，包括偏最小二乘回归（Partial least squares regression, PLSR）<sup>[30]</sup>模型、人工神经网络（Artificial neural network, ANN）<sup>[31]</sup>模型、支持向量机（Support vector machine, SVM）<sup>[32]</sup>模型和随机森林（RF）模型<sup>[33]</sup>等。这些算法能够快速分析近红外光谱，准确地确定样品的类型以及其中重要组成成分的含量，使得分析更加高效。

### 1.2.2 基于深度学习的近红外光谱定量分析方法

为了从高维光谱数据中提取待测样品的品质信息，需要建立光谱数据与目标变量之间的数学映射关系，从而实现对样品中目标成分含量的预测或对样品类别的判别，并将模型应用于新样品，实现快速无损检测<sup>[34]</sup>。然而，传统化学计量学方法在建模过程中存在流程繁琐、拟合性能有限以及鲁棒性较差等问题，难以满足当前小麦蛋白质含量无损检测对高精度与智能化的需求。

在基于光谱技术的食品质量定量分析中，核心任务是建立光谱特征与目标属性（如蛋白质含量）之间的回归模型。传统方法往往依赖人工特征工程，难以充分挖掘高维光谱数据中的复杂信息。随着信息技术的发展，深度学习作为一种数据驱动方法，能够通过自动学习提取数据中的潜在特征，并拟合复杂的非线性关系，实现端到端建模，从而

有效替代人工特征提取过程<sup>[35]</sup>。这一方法已被证明是一种有前景且具有优势的方法，能够从原始光谱数据中自动提取与目标属性高度相关的深层特征<sup>[36]</sup>。

凭借其强大的特征自动提取能力，深度学习已被广泛应用于农产品品质无损检测领域<sup>[37]</sup>。但需要指出的是，深度学习模型本身无法直接获取农产品内部品质信息，其有效性依赖于与光谱技术的结合。光谱技术（如近红外光谱 NIR 和高光谱成像 HSI）提供反映内部化学成分的基础数据，而深度学习则对这些复杂数据进行建模分析，二者形成“光谱采集—智能解析”的协同模式<sup>[38]</sup>。在这一框架下，通过融合数据预处理、特征提取和模型构建等步骤，可以将原始光谱数据转化为更高层次的抽象特征，从而显著提升模型的表达能力与自适应能力<sup>[39]</sup>。

在实际应用中，这种端到端建模框架已在小麦蛋白质含量检测中得到验证。例如，研究者利用高光谱成像（HSI）获取小麦籽粒光谱信息，并结合一维卷积神经网络（1D-CNN），实现了蛋白质含量的实时、无损和高精度预测<sup>[40]</sup>。此外，还有研究基于近红外光谱（NIRS）数据，通过 Savitzky-Golay 平滑等方法进行预处理后，构建并行卷积神经网络（Parallel CNN）模型，显著提升了预测准确性与模型稳定性<sup>[41]</sup>。

作为深度学习的重要代表，卷积神经网络（CNN）在图像识别领域取得了显著成果，并发展出 AlexNet<sup>[42]</sup>、VGG<sup>[43]</sup>、GoogLeNet<sup>[44]</sup>和 ResNet<sup>[45]</sup>等经典结构。近年来，CNN 逐渐被引入光谱分析领域，并在农业信息化研究中得到广泛应用，如农产品形状分类、质量分级及品种识别等<sup>[46]</sup>。研究表明，将深度学习与光谱技术相结合的方法在作物品质评估中具有良好的可靠性<sup>[47]</sup>。

进一步研究表明，卷积神经网络能够直接从高维原始光谱数据中提取有效信息，无需依赖复杂的预处理和人工特征设计<sup>[48]</sup>。其多层非线性结构能够逐层提取特征，将底层数据逐步映射为高层抽象表达，从而学习复杂的光谱特征表示。随着研究深入，CNN 在光谱分析中的应用不断增多，尤其是在一维光谱分析任务中发挥着越来越重要的作用<sup>[49]</sup>，并能够显著提升模型对数据的拟合能力和整体表现<sup>[50]</sup>。

在具体研究中，Acquarelli 等<sup>[51]</sup>提出了一种仅包含单卷积层的 CNN 模型，并在多个数据集上验证其性能，结果表明模型平均准确率达到 96%，且对数据预处理依赖较低。杨友等<sup>[47]</sup>将 CNN 与特征选择回归方法相结合，通过多种预处理提升光谱信噪比，并构建集成模型进行小麦蛋白质含量预测，结果表明 CNN 的引入显著增强了模型的非线性建模能力，提高了预测性能。此外，Chen 等<sup>[49]</sup>提出端到端 CNN 集成模型用于玉米品质评价，在多个指标预测中优于传统方法；Prasad 等<sup>[52]</sup>则设计轻量级 HResNeXt 模型，在降低计算复杂度的同时实现了高效特征提取。

在模型结构方面，一维卷积神经网络（1D-CNN）因其在序列数据局部特征提取中的优势，被广泛应用于光谱分析。其通过卷积核沿波长维度滑动，能够有效捕捉与化学成分相关的局部特征（如吸收峰），从而替代传统人工特征提取方法。例如，在糙米蛋