



基于双通道 GWO-KELM 的河北省主要蔬菜品种价格预测研究

贾新诚^{1,2}, 贾华^{1*}

1.大连海洋大学机械与动力工程学院, 辽宁大连 116023; 2.阜平县职业技术教育中心, 河北保定 073200

摘要: 针对河北省主要蔬菜价格序列的非线性、非平稳性与高噪声特征, 以及核极限学习机(KELM)参数依赖经验设定导致预测精度受限的问题, 构建一种融合灰狼优化算法(GWO)与双通道核极限学习机的多变量组合预测模型, 以提升蔬菜价格的短期预测精度。以2021年1月至2024年12月河北省10个主要蔬菜品种(大白菜、洋白菜、西红柿、黄瓜、土豆、青椒、大葱、大蒜、豆角、西葫芦)的周度批发价格数据为研究对象, 总样本量2 080条。同时采集气象、节日、农资、成品油、微博舆情、百度指数、宏观经济、食品消费等8类28个外部特征。采用均值插补和 3σ 原则进行数据清洗, Min-Max归一化消除量纲影响。首先通过格兰杰因果检验验证外部特征与价格的预测性因果关系, 并为保证模型简洁性和训练效率, 每个品种保留最显著的8个特征; 进而构建双通道并行KELM模型——通道1输入价格历史序列(4周), 通道2输入筛选后的外部特征, 两通道核矩阵通过可学习权重 α 进行加权融合。引入GWO算法对正则化系数 C 、核参数 σ_1 、 σ_2 及融合权重 α 进行自动寻优, 以验证集均方根误差(RMSE)为适应度函数, 种群规模15、迭代50次。同时与单通道价格历史模型(M1)、单通道特征拼接模型(M2)以及SVM、LSTM、GRU等3种主流模型进行对比。双通道GWO-KELM模型(M3)在10个品种上的平均绝对百分比误差(MAPE)为7.52%, 较M1(10.59%)降低3.07个百分点, 较M2(9.12%)降低1.60个百分点, 显著优于LSTM(9.87%)、GRU(9.56%)等深度学习方法。在高波动品种大葱上改善尤为显著(MAPE由13.76%降至9.85%)。融合权重 α 显示: 价格平稳品种(大蒜 $\alpha=0.81$, 土豆 $\alpha=0.76$)更依赖历史惯性, 高波动品种(大葱 $\alpha=0.48$, 黄瓜 $\alpha=0.54$)对外部特征依赖更强, α 与价格变异系数呈负相关($r \approx -0.73$)。外部特征贡献度分析表明, 气温、农资价格、节日舆情及成品油价格是影响蔬菜价格的重要先行指标。模型在未参与训练的茄子、芹菜、生菜上MAPE分别为7.98%、9.12%、12.06%, 各季节误差增幅控制在1个百分点以内。本研究构建的双通道GWO-KELM组合模型有效解决了KELM参数依赖经验设定的问题, 通过格兰杰因果验证与双通道核融合机制, 实现了时序特征与多源异构外部特征的自适应整合。在河北省主要蔬菜品种价格预测中具有高精度、强泛化能力和良好稳健性, 可为政府调控、企业运营和农户生产提供科学决策支持。

关键词: 灰狼优化算法; 核极限学习机; 格兰杰因果检验; 蔬菜价格预测; 河北省; 双通道模型

1 引言

蔬菜是城乡居民“菜篮子”的核心品类, 其价格稳定直接关系民生保障与农民增收^[1]。河北省作为京津冀“菜篮子”第一大省, 2024年蔬菜产量达5 717.3万吨, 同比增长5.91%, 但受自然条件、生产成本、流通环节及市场供需等多重因素影响, 蔬菜价格呈现

显著的周期性、季节性与突发性波动特征^[2]。小农户分散经营的生产决策往往滞后于市场变化, 易陷入“价高扩种、价低弃种”的恶性循环。因此, 构建高精度、可解释的蔬菜价格预测模型, 对于稳定市场、助农增收具有重要意义。

农产品价格预测方法经历了从传统统计模型(ARIMA、灰色模型^[3-5])到机器学习(BP、ELM、

收稿日期: 2026-05-13; 录用日期: 2026-07-15

基金项目: 河北省职业教育科学研究“十四五”规划2025年度课题(JZY25181)。

联系方式: 贾新诚, E-mail: djpc_jia@126.com。通信作者贾华, E-mail: jiahua110023@126.com。

SVM^[6-8])再到深度学习(LSTM、GRU^[9-10])的演进。核极限学习机(KELM)因其学习速度快、非线性拟合能力强而被广泛应用于时序预测^[11-12],但其性能高度依赖核参数 σ 和正则化系数C的经验选择^[13]。智能优化算法(如粒子群算法PSO、灰狼算法GWO)可有效解决参数寻优问题^[14],其中GWO因参数少、搜索机制平衡而表现更优^[15-16]。在农产品价格预测领域,已有研究将GWO与KELM结合^[17-19],但多数仅使用价格历史数据,未充分融入气象、舆情、政策等多源外部特征,且对时序特征与截面特征的异质性处理不足。

针对上述问题,本文提出一种双通道GWO-KELM组合预测模型。主要创新点如下:(1)首次将格兰杰因果检验用于蔬菜价格预测的外部特征筛选,从28个潜在变量中自适应选择各品种的显著影响因素;(2)构建双通道并行KELM架构,分别

处理时序依赖特征(价格历史)和横截面外部特征,并通过可学习权重 α 在核矩阵层面实现异构特征的最优融合;(3)以河北省10个主要蔬菜品种为对象,开展系统的多变量预测实验,验证模型的有效性和普适性。

2 数据与方法

2.1 数据来源与预处理

2.1.1 数据来源

蔬菜价格数据来源于河北省农业农村厅官网“河北省主要蔬菜品种批发市场价格周报”^[2],选取2021年第1周至2024年第52周(共208周)10个主要品种(大白菜、洋白菜、西红柿、黄瓜、土豆、青椒、大葱、大蒜、豆角、西葫芦)的周度批发价格,总计2080条记录。

外部特征数据共28个变量,详见表1。

表1 外部特征数据来源与变量列表

Table 1 Data sources and variable list of external features

类别	变量数量	具体变量	数据来源
气象	3	周平均气温、周总降水量、周日照时数	国家气象信息中心
节日	3	春节前4周、国庆前2周、中秋前2周	国务院办公厅节假日安排通知
农资	10	复合肥、尿素、氯化钾、氯化铵、液氨、硫酸钾、硫酸铵、碳酸氢铵、磷酸一铵、磷酸二铵价格(元/吨)	农业农村部全国农资价格行情表
成品油	2	汽油周度调价幅度、柴油周度调价幅度(元/吨)	国家发展改革委成品油调价公告
微博舆情	2	“菜价”“蔬菜”关键词的周度互动量(转发+评论+点赞)	微博公开数据
百度指数	5	“大蒜”“番茄”“白菜”“蔬菜”“黄瓜”的周均搜索热度	百度指数平台
宏观经济	2	城镇化率(%)、城镇居民人均可支配收入(元)	河北省统计局(年度数据线性插值到周度)
食品消费	1	全省居民人均蔬菜消费量(公斤)	河北省统计局(年度数据线性插值到周度)

2.1.2 数据预处理

所有时间序列按ISO周标准对齐为2021W1–2024W52(208周)。缺失值采用前后各2周有效值的均值插补;异常值基于 3σ 原则检测并修正。为消除量纲影响,对连续型变量进行Min-Max归一化,映射至[0,1]区间。归一化参数仅基于训练集(2021W1–2023W52,156周)计算,验证集和测试集采用同一映射。数据集划分:训练集(156周)、验证集(2024W1–W26,26周)、测试集(2024W27–W52,26周)。

2.2 外部特征筛选: 格兰杰因果检验

为验证外部特征与蔬菜价格之间的预测性因果关

系,并为多变量输入提供统计依据,本文首先对每个蔬菜品种分别进行格兰杰因果检验^[20]。检验步骤为:

(1)对每个价格序列和外部特征序列进行ADF单位根检验,非平稳序列取一阶差分直至平稳;(2)以AIC准则确定最优滞后阶数(最大滞后4周);(3)检验原假设“外部特征X不是价格Y的格兰杰原因”,若 $p<0.05$ 则判定存在显著因果关系。

检验结果(表2)显示,气温、农资价格、节日、舆情等变量与蔬菜价格存在显著的格兰杰因果关系($p<0.05$),证实了将这些特征纳入预测模型的合理性。在此基础上,为保证模型简洁性和训练

效率,每品种保留显著性最高的8个外部特征。以西红柿为例,最终保留的变量包括周日照时数($p=0.012$)、尿素价格($p=0.008$)、汽油调价幅度($p=$

0.021)、蔬菜百度指数($p=0.035$)、人均蔬菜消费量($p=0.043$)等8个。表2以5个代表性品种展示检验结果。

表2 代表性品种格兰杰因果检验结果($p<0.05$ 的特征)

Table 2 Granger causality test results for representative varieties (features with $p<0.05$)

品种	显著外部特征	滞后阶数
大白菜	周平均气温、复合肥价格、蔬菜百度指数	2
西红柿	周日照时数、尿素价格、汽油调价、蔬菜消费量	1
黄瓜	周降水量、氯化钾价格、中秋节虚拟变量、菜价微博互动量	3
大葱	周平均气温、柴油调价、春节虚拟变量、大葱百度指数	2
土豆	磷酸二铵价格、城镇化率、人均可支配收入	1

综合10个品种,被选中频率最高的外部特征依次为:周平均气温(8/10)、尿素价格(7/10)、复合肥价格(7/10)、蔬菜百度指数(6/10)、春节虚拟变量(6/10)、汽油调价幅度(5/10)、周日照时数(5/10)。

2.3 双通道核极限学习机(Dual-Channel KELM)

2.3.1 标准KELM回顾

给定训练集 $\{(x_i, y_i)\}_{i=1}^N$, KELM引入核矩阵 $\Omega_{ij} =$

$K(x_i, x_j)$, 输出权重为: $\beta = \left(\frac{1}{C} + \Omega \right)^{-1} y$

预测函数为 $f(x) = [K(x, x_1), \dots, K(x, x_N)]\beta$ 。本文采用径向基核函数 $K(x_i, x_j) = \exp\left(-\frac{\|x_i - x_j\|^2}{2\sigma^2}\right)$ 。

2.3.2 双通道并行融合架构

设每个样本的特征分为两部分: $x(1)_i$ (时序特征, 前4周价格) 和 $x(2)_i$ (外部特征, 经格兰杰筛选后的8个变量)。分别计算两个核矩阵:

$$\Omega_{ij}^{(1)} = \exp\left(-\frac{\|x_i^{(1)} - x_j^{(1)}\|^2}{\sigma_1^2}\right), \quad \Omega_{ij}^{(2)} = \exp\left(-\frac{\|x_i^{(2)} - x_j^{(2)}\|^2}{\sigma_2^2}\right)$$

融合核矩阵为: $\Omega_{\text{fused}} = \alpha\Omega^{(1)} + (1-\alpha)\Omega^{(2)}$

其中 $\alpha \in [0, 1]$ 为可学习的融合权重。最终输出权重 $\beta = (1/C + \Omega_{\text{fused}})^{-1} y$ 。该架构的优点在于: 两个通道在核函数层面独立学习各自特征空间的相似性, 避免了直接特征拼接带来的信息干扰, 且 α 可自适应平衡时序惯性 (α 大) 与外部冲击 (α 小) 的贡献。

2.4 灰狼优化算法(GWO)与超参数寻优

GWO算法模拟灰狼群体的社会等级和捕食行为^[14]。本研究使用GWO对双通道KELM的4个关键超参数

进行自动寻优:

正则化系数 $C \in [10^{-2}, 10^4]$

时序通道核参数 $\sigma_1 \in [10^{-3}, 10^3]$

特征通道核参数 $\sigma_2 \in [10^{-3}, 10^3]$

融合权重 $\alpha \in [0, 1]$

优化目标为最小化验证集均方根误差(RMSE)。

GWO参数设置: 种群规模15, 最大迭代次数50, 收敛因子 a 从2线性递减至0。每个品种独立运行优化。

2.5 对比模型与评价指标

设置6种对比模型以全面评估M3的性能:

SVM: RBF核, 网格搜索优化 C 和 γ 。

LSTM: 两层LSTM(每层64单元), Dropout=0.2, Adam优化器。

GRU: 两层GRU(每层64单元), Dropout=0.2, Adam优化器。

M1(单通道价格历史): 仅使用前4周价格的标准KELM。

M2(单通道特征拼接): 将价格历史与外部特征直接拼接的标准KELM。

M3(双通道并行融合, 本文提出): 双通道KELM+核矩阵加权融合。

评价指标: 平均绝对误差(MAE)、平均绝对百分比误差(MAPE)、均方根误差(RMSE)。

3 结果与分析

3.1 GWO参数优化结果

以西红柿为例, GWO寻得的最优超参数为: $C=1236.8$, $\sigma_1=2.47$, $\sigma_2=0.83$, $\alpha=0.62$ 。所有品种的优化参数及 α 值见表3。

表 3 各品种 GWO 优化后的双通道 KELM 超参数

Table 3 Optimized dual-channel KELM hyperparameters for each variety by GWO

品种	C	σ_1	σ_2	α
大白菜	892.4	1.85	0.72	0.51
洋白菜	1045.2	2.13	0.68	0.53
西红柿	1236.8	2.47	0.83	0.62
黄瓜	1567.3	1.26	0.45	0.54
土豆	678.5	4.12	1.54	0.76
青椒	1123.6	2.58	0.91	0.58
大葱	2034.7	0.92	0.38	0.48
大蒜	453.2	5.67	2.31	0.81
豆角	1345.9	3.24	1.12	0.60
西葫芦	987.3	1.94	0.59	0.55

α 值反映了品种对外部信息的依赖程度：大蒜（0.81）、土豆（0.76）等耐储藏、价格平稳的品种主要依赖价格历史惯性；大葱（0.48）、黄瓜（0.54）等高波动品种对外部特征（气温、油价等）更为敏感。相关性分析显示， α 与价格

变异系数呈显著负相关（ $r=-0.73$ ， $p<0.05$ ），验证了直觉认知。

3.2 预测性能对比

三种模型在测试集（2024W27–W52）上的预测误差如表 4 所示。

表 4 多种模型在 10 个品种上的平均预测误差对比（测试集）

Table 4 Comparison of average prediction errors of multiple models on 10 varieties (test set)

模型	平均 MAE (元/公斤)	平均 MAPE (%)	平均 RMSE (元/公斤)
SVM	0.41	11.23	0.53
LSTM	0.35	9.87	0.46
GRU	0.34	9.56	0.44
M1 (单通道价格历史)	0.33	10.59	0.42
M2 (单通道特征拼接)	0.29	9.12	0.36
M3 (双通道融合)	0.27	7.52	0.33

从表 4 可以看出，本文提出的 M3 取得了最低的预测误差。与主流深度学习方法相比，M3 的平均 MAPE 较 LSTM（9.87%）降低 2.35 个百分点，较 GRU（9.56%）降低 2.04 个百分点。与传统的 SVM 相比，优势更为明显（降低 2 个百分点）。值得注意的是，M2（单通道特征拼接）的 MAPE（9.12%）已优于 LSTM 和 GRU，表明在 KELM 框架下引入外部特征的提升效果，甚至超过了从传统网络切换到复杂深度网络带来的增益。而 M3 在

M2 基础上进一步通过双通道核融合实现了 1.60 个百分点的额外提升，验证了核矩阵加权融合机制的有效性。

此外，为验证特征筛选的有效性，本文补充了对照实验：使用全部 28 个外部特征训练双通道 GWO-KELM 模型（记为 M3-All）。结果显示（表 5），M3-All 的平均 MAPE 为 7.68%，略高于筛选后的 M3（7.52%），表明适度剔除冗余特征不仅未损失有效信息，反而通过降低噪声提升了预测精度。

表 5 特征筛选对照实验（全特征 vs 筛选特征）

Table 5 Feature selection comparison experiment (all features vs. selected features)

品种	M3-All (全 28 特征) MAPE(%)	M3 (筛选后 8 特征) MAPE(%)	差值 (百分点)
大白菜	10.18	9.73	-0.45
洋白菜	9.31	8.94	-0.37
西红柿	6.92	6.65	-0.27
黄瓜	8.34	8.02	-0.32
土豆	6.45	6.18	-0.27
青椒	7.81	7.49	-0.32
大葱	10.12	9.85	-0.27
大蒜	3.96	3.84	-0.12
豆角	6.48	6.21	-0.27
西葫芦	8.66	8.33	-0.33
平均	7.68	7.52	-0.16

3.3 特征重要性分析

基于双通道架构中特征通道的贡献及格兰杰检验 p 值，以西红柿为例进行 SHAP 值分析^[21]，各外部特征的相对贡献为：周日照时数贡献度最高（24%），其次是尿素价格（22%）、汽油调价（18%）、蔬菜百度指数（15%）、人均蔬菜消费量（12%）。这与西红柿喜光、氮肥敏感、运输成本影响显著等农业经

济规律吻合。

3.4 普适性与稳健性检验

3.4.1 跨品种泛化

将 M3 模型直接应用于未参与训练的茄子、芹菜、生菜（2024 年测试集），结果见表 6。M3 对三种未知品种均取得显著优于 M1 的效果，MAPE 降幅在 2.5 - 3.2 个百分点之间，表明模型具有良好的泛化能力。

表 6 模型对未参与训练品种的预测误差

Table 6 Prediction errors of the model on untrained varieties

品种	M1 MAPE(%)	M3 MAPE(%)	提升
茄子	10.43	7.98	-2.45pp
芹菜	12.00	9.12	-2.88pp
生菜	15.27	12.06	-3.21pp

3.4.2 季节稳健性

将 2024 年测试集按季节划分为春季（W14-17）、夏季（W18-26）、秋季（W27-39）、冬季（W40-52），

计算 M3 的平均 MAPE（表 7）。各季节误差保持稳定，夏季提升幅度略大，可能与夏季极端天气、油价波动等外部因素影响增强有关。

表 7 模型分季节预测性能

Table 7 Seasonal prediction performance of the model

季节	M3 平均 MAPE(%)	M1 平均 MAPE(%)	提升
春季	6.89	10.02	+3.13pp
夏季	7.94	11.21	+3.27pp
秋季	7.58	10.68	+3.10pp
冬季	7.48	10.45	+2.97pp

4 讨论

4.1 外部特征的局限性

本文构建的外部特征体系虽未直接包含周边省份（北京、天津、山东）的蔬菜价格变量，但通过百度指数（“蔬菜价格”等关键词的搜索热度）和微博舆情（“菜价”互动量），部分捕捉了消费者跨区域关注和网络信息传播对河北省内价格预期的影响。此外，成品油价格变量间接反映了跨区域物流成本，而农资价格（如复合肥、尿素）的跨省趋同性也在一定程度上代理了周边产区的生产成本变化。然而，便捷的交通条件和日益普及的网购渠道确实为省外蔬菜进入河北市场提供了便利，可能对本地价格产生直接的替代效应。受数据可得性限制，本文未能直接量化这种跨区域竞争和渠道替代效应，未来研究将纳入北京新发地市场批发价格、山东寿光蔬菜价格指数以及主流电商平台的蔬菜销售数据，构建更完整的跨区域价格传导模型。

4.2 模型优势与适用性

与 LSTM、GRU 等深度学习方法相比，双通道 GWO-KELM 在参数量更少（仅需优化 4 个超参数）、训练速度更快的同时，取得了更高的预测精度。这主要得益于：（1）核方法对非线性关系的高效建模；（2）双通道架构对异构特征的自适应融合；（3）GWO 对超参数的全局优化。本模型尤其适用于中小样本、多变量、非线性强的农业时序预测场景。

5 结论与展望

5.1 结论

本文以河北省 10 个主要蔬菜品种为研究对象，提出一种融合格兰杰因果验证与双通道 GWO-KELM 的多变量价格预测模型，主要结论如下：

（1）格兰杰因果检验有效验证了气温、农资价格、节日舆情、成品油价格等外部特征与蔬菜价格之间的显著预测性因果关系，为多变量输入提供了统计依据。

（2）双通道并行 KELM 通过独立计算时序核矩阵和特征核矩阵并以可学习权重融合，显著优于单通道价格历史模型（M1）、特征拼接模型（M2）以及 LSTM、GRU 等深度学习方法。M3 的平均 MAPE 为 7.52%，较 M1 降低 3.07 个百分点，较 M2 降低 1.60 个百分点，较 LSTM 降低 2.35 个百分点。

（3）融合权重 α 揭示了品种异质性：价格平稳品

种（大蒜 $\alpha=0.81$ ，土豆 $\alpha=0.76$ ）更依赖历史惯性，高波动品种（大葱 $\alpha=0.48$ ，黄瓜 $\alpha=0.54$ ）对外部特征依赖更强， α 与价格变异系数呈负相关（ $r \approx -0.73$ ）。这一发现为不同品种的价格预测策略差异化提供了量化依据。

（4）模型在未参与训练的品种及不同季节上均表现稳健，验证了其良好的泛化能力和实用性。

5.2 展望

未来工作可从以下方面深化：（1）数据维度拓展：纳入周边省份（北京、天津、山东）的蔬菜批发价格、电商平台销售数据以及极端天气预警信息，构建跨区域价格传导模型；（2）模型轻量化：探索知识蒸馏、模型剪枝等技术，降低双通道 KELM 的计算成本，便于部署到移动端或边缘设备；（3）实时预测系统：开发可视化决策支持平台，整合实时数据流，实现周度价格的滚动预测和动态预警，服务政府调控、企业经营和农户生产。

6 数据可用性

本数据集已在《农业大数据学报》认证数据仓储——国家农业科学数据中心（<https://www.agridata.cn>）正式发布，可通过以下唯一标识符访问。

CSTR: <https://cstr.cn/17058.11.sciencedb.agriculture.00315>

DOI: <https://doi.org/10.57760/sciencedb.agriculture.00315>

数据服务系统网址: <https://www.agridata.cn>

数据引用格式:

贾新诚,贾华. 2021—2024 年河北省主要蔬菜品种批发价格及多源影响因素数据集[DS/OL]. V1. Science Data Bank, 2026[2026-06-14]. <https://cstr.cn/17058.11.sciencedb.agriculture.00315>. CSTR:17058.11.sciencedb.agriculture.00315.

参考文献

- [1] 许世卫,李乾川,栾汝朋,等. 农产品市场监测预警深度学习智能预测方法. 智慧农业(中英文), 2025, 7(1): 57-69.
Xu S W, Li Q C, Luan R P, et al. Agricultural market monitoring and early warning: An integrated forecasting approach based on deep learning. Smart Agriculture, 2025, 7(1): 57-69.
- [2] 宗义湘,宋淼鑫,王逸群. 2024—2025 年河北省蔬菜产业形势分析与预测[M]//河北农业农村经济发展报告(2025). 北京: 社会科学文献出版社, 2025: 42-59.
Zong Y X, Song M X, Wang Y Q. Analysis and forecast of vegetable industry situation in Hebei Province, 2024-2025[M]//Agricultural and

- Rural Economy Development Report of Hebei (2025). Beijing: Social Sciences Academic Press, 2025: 42-59.
- [3] 彭红星,郑楷航,黄国彬,等. 基于 BP、LSTM 和 ARIMA 模型的蔬菜价格预测. 中国农机化学报, 2020, 41(4): 193-199.
- Peng H X, Zheng K H, Huang G B, et al. Vegetable price prediction based on BP, LSTM and ARIMA models. Journal of Chinese Agricultural Mechanization, 2020, 41(4): 193-199.
- [4] 丁慧娟,张金磊,陈建中,等. ARIMA 模型和灰色模型在农产品价格预测中的应用比较. 安徽农业科学, 2018, 46(24): 191-194.
- Ding H J, Zhang J L, Chen J Z, et al. Comparison of ARIMA Model and Grey Model in agricultural product price forecasting. Journal of Anhui Agricultural Sciences, 2018, 46(24): 191-194.
- [5] Manogna R L, Dharmaji V, Sarang S. Enhancing agricultural commodity price forecasting with deep learning. Scientific Reports, 2025, 15: 20903.
- [6] 郭婷婷,陈新国. 基于 PCA-ELM 的我国粮食价格预测研究. 价格月刊, 2015(12): 21-26.
- Guo T T, Chen X G. The prediction study on China's grain price based on PCA-ELM. Prices Monthly, 2015(12): 21-26.
- [7] 贾宁,郑纯军. 基于 LSTM-DA 神经网络的农产品价格指数短期预测模型. 计算机科学, 2019, 46(S11): 62-65,71.
- Jia N, Zheng C J. Short-term Forecasting Model of agricultural product price index based on LSTM-DA Neural Network. Computer Science, 2019, 46(S11): 62-65,71.
- [8] Paul R K, Yeasin M, Kumar P, et al. Machine learning techniques for forecasting agricultural prices: A case of brinjal in Odisha, India. PLoS One, 2022, 17(7): e0270553.
- [9] 方雪清,吴春胤,俞守华,等. 基于 EEMD-LSTM 的农产品价格短期预测模型研究. 中国管理科学, 2021, 29(11): 68-77.
- Fang X Q, Wu C Y, Yu S H, et al. Research on Short-Term Prediction Model of agricultural product prices based on EEMD-LSTM. Chinese Journal of Management Science, 2021, 29(11): 68-77.
- [10] 李洁琼,刘振. 基于 CNN 和 GRU 的农产品价格预测模型研究. 湖北农业科学, 2023, 62(1): 177-181.
- Li J Q, Liu Z. Research on the price forecast model of agricultural products based on CNN and GRU. Hubei Agricultural Sciences, 2023, 62(1): 177-181.
- [11] Huang G B, Zhu Q Y, Siew C K. Extreme learning machine: Theory and applications. Neurocomputing, 2006, 70(1/2/3): 489-501.
- [12] Huang G B, Zhou H M, Ding X J, et al. Extreme learning machine for regression and multiclass classification. IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics), 2012, 42(2): 513-529.
- [13] 刘合兵,韩晶晶,马新明,等. 基于 EMD-ELM 模型的农产品价格预测研究. 农业大数据学报, 2020, 2(3): 68-74.
- Liu H B, Han J J, Ma X M, et al. Research on agricultural product price prediction based on the EMD-ELM model. Journal of Agricultural Big Data, 2020, 2(3): 68-74.
- [14] Mirjalili S, Mirjalili S M, Lewis A. Grey wolf optimizer. Advances in Engineering Software, 2014, 69: 46-61.
- [15] 轩彤,修子涵,初洪龙. Ensemble 融合模型的建立及其在面粉价格预测中的应用. 农业大数据学报, 2026, 8(1): 36-47.
- Xuan T, Xiu Z H, Chu H L. Ensemble integration model and its application in flour price forecasting. Journal of Agricultural Big Data, 2026, 8(1): 36-47.
- [16] Feng M D, Duan Y H, Wang X, et al. Carbon price prediction based on decomposition technique and extreme gradient boosting optimized by the grey wolf optimizer algorithm. Scientific Reports, 2023, 13: 18447.
- [17] Wang G, Xu S, Chen Z, et al. A hybrid model integrating variational mode decomposition and intelligent optimization for vegetable price prediction. Agriculture, 2025, 15(9): 919.
- [18] Li F Y, Meng X Y, Zhu K, et al. Research on ginger price prediction model based on deep learning. Agriculture, 2025, 15(6): 596.
- [19] 侯颖,孙坦,崔运鹏,等. 基于优化神经网络时间序列模型的蔬菜价格预测方法. 智慧农业(中英文), 2025, 7(5): 78-87.
- Hou Y, Sun T, Cui Y P, et al. Vegetable price prediction based on Optimized Neural Network Time Series Models. Smart Agriculture, 2025, 7(5): 78-87.
- [20] Granger C W J. Investigating causal relations by econometric models and cross-spectral methods. Econometrica, 1969, 37(3): 424-438.
- [21] Lundberg S M, Lee S I. A unified approach to interpreting model predictions[C]//NIPS'17: Proceedings of the 31st International Conference on Neural Information Processing Systems, 2017: 4768-4777. <https://arxiv.org/pdf/1705.07874>.

引用本文: 贾新诚,贾华. 基于双通道 GWO-KELM 的河北省主要蔬菜品种价格预测研究[J]. 农业大数据学报,2026,8(3):305-312. DOI: 10.19788/j.issn.2096-6369.000166.

CITE: JIA Xincheng, JIA Hua. Prediction of Main Vegetable Prices in Hebei Province Based on Dual-Channel GWO-KELM[J]. Journal of Agricultural Big Data, 2026,8(3): 305-312. DOI: 10.19788/j.issn.2096-6369.000166

Prediction of Main Vegetable Prices in Hebei Province Based on Dual-Channel GWO-KELM

JIA Xincheng^{1,2}, JIA Hua^{1*}

1. College of Mechanical and Power Engineering, Dalian Ocean University, Dalian 116023, Liaoning, China; 2. Fuping County Vocational and Technical Education Center, Baoding 073200, Hebei, China

Abstract: To address the nonlinear, non-stationary, and high-noise characteristics of main vegetable price series in Hebei Province, as well as the problem that the kernel extreme learning machine (KELM) relies on empirical parameter settings, which limits its prediction accuracy, this paper constructs a multivariate combined forecasting model integrating the grey wolf optimizer (GWO) and a dual-channel KELM to improve short-term vegetable price prediction accuracy. The weekly wholesale price data of 10 major vegetable varieties (Chinese cabbage, cabbage, tomato, cucumber, potato, green pepper, scallion, garlic, bean, and zucchini) in Hebei Province from January 2021 to December 2024 are collected, with a total sample size of 2,080 records. Eight categories of external features (meteorological, holiday, agricultural material prices, refined oil prices, Weibo public opinion, Baidu index, macro-economic indicators, and food consumption) comprising 28 variables are also collected. Missing values are imputed by mean interpolation, outliers are corrected based on the 3σ principle, and Min-Max normalization is applied to eliminate dimensional effects. First, the Granger causality test is used to verify the predictive causal relationship between external features and vegetable prices, and for the sake of model parsimony and training efficiency, the eight most significant features per variety are retained. Then, a dual-channel parallel KELM model is constructed: Channel 1 takes the price history (4 weeks) and Channel 2 takes the selected external features. The kernel matrices of the two channels are fused by a learnable weight α . The GWO algorithm is introduced to automatically optimize the regularization coefficient C , kernel parameters σ_1 and σ_2 , and the fusion weight α , using the root mean square error (RMSE) on the validation set as the fitness function with a population size of 15 and 50 iterations. The proposed model (M3) is compared with a single-channel model using only price history (M1), a single-channel feature-concatenation model (M2), as well as three mainstream models (SVM, LSTM, GRU). The dual-channel GWO-KELM model (M3) achieves an average mean absolute percentage error (MAPE) of 7.52% across the 10 varieties, which is 3.07 percentage points lower than that of M1 (10.59%) and 1.60 percentage points lower than that of M2 (9.12%), and significantly outperforms deep learning methods such as LSTM (9.87%) and GRU (9.56%). The improvement is particularly significant for highly volatile varieties, such as scallion (MAPE reduced from 13.76% to 9.85%). The fusion weight α shows that price-stable varieties (garlic $\alpha=0.81$, potato $\alpha=0.76$) rely more on historical inertia, while highly volatile varieties (scallion $\alpha=0.48$, cucumber $\alpha=0.54$) depend more on external features, with a negative correlation between α and the coefficient of price variation ($r \approx -0.73$). Contribution analysis of external features indicates that temperature, agricultural material prices, holiday/opinion indicators, and refined oil prices are important leading indicators for vegetable prices. The model also exhibits good generalization on untrained varieties (eggplant, celery, lettuce), with MAPEs of 7.98%, 9.12%, and 12.06%, respectively, and its seasonal error increase is controlled within 1 percentage point. The dual-channel GWO-KELM model constructed in this paper effectively solves the problem of empirical parameter setting of KELM. By integrating Granger causality verification and dual-channel kernel fusion, it achieves adaptive integration of temporal features and heterogeneous external features. The model demonstrates high accuracy, strong generalization ability and good robustness in forecasting main vegetable prices in Hebei Province, and can provide scientific decision support for government regulation, enterprise operations and farmer production.

Keywords: Grey Wolf Optimizer; kernel extreme learning machine; Granger causality test; vegetable price prediction; Hebei Province; dual-channel model