许玉辰, 崔继馨, 王庆江,等. 基于不同机器学习方法的梨树需水量预测研究[J]. 智能计算机与应用,2026,16(5):94-101.
DOI:10.20169/j.issn.2095-2163.24090602

基于不同机器学习方法的梨树需水量预测研究

许玉辰[1], 崔继馨[1], 王庆江[2], 及歆荣[1]
(1 河北工程大学 信息与电气工程学院, 河北 邯郸 056038; 2 河北工程大学 园林与生态工程学院, 河北 邯郸 056038)

摘 要: 为探究不同机器学习算法在梨树需水量预测领域的应用效果,以位于河北南部 2009 年栽植的黄冠梨园为试验对象,获取梨园的生态因子,基于 FAO56 手册推荐的单作物系数法,计算 2018—2023 年间梨园内的黄冠梨树逐日需水量,并以此为基础组建黄冠梨树需水量数据集。选取温度、风速、土壤等影响黄冠梨树需水量的关键因子作为输入特征项,分别使用支持向量机回归(Support Vector Regression, SVR)、随机森林(Random Forest Regression, RF)、多项式回归(Polynomial Regression)、反向传播神经网络(Back-Propagation Neural Network, BPNN)4 种机器学习预测模型预测黄冠梨树的需水量。通过反复训练测试,比较 4 种模型的预测精度、泛化能力和稳定性。结果表明,4 种模型 R^2 值均在 0.90 以上,支持向量机回归模型整体表现最好, R^2、RMSE、MAE 和 max-error 分别为 0.92、0.73、0.52 和 3.21,且多次实验输出结果稳定。因此,支持向量机学习方法能够比较准确地预测黄冠梨树需水量,对黄冠梨园将来的智慧灌溉、精准灌溉具有重要的现实意义。

关键词: 需水量预测; 黄冠梨; 机器学习; SVR; BPNN; 随机森林; 多项式回归
中图分类号: S274.2 **文献标志码**: A **文章编号**: 2095-2163(2026)05-0094-08

Prediction of pear water demand based on different machine learning methods

XU Yuchen[1], CUI Jixin[1], WANG Qingjiang[2], JI Xinrong[1]
(1 School of Information and Electrical Engineering, Hebei University of Engineering, Handan 056038, Hebei, China;
2 School of Landscape and Ecological Engineering, Hebei University of Engineering, Handan 056038, Hebei, China)

Abstract: To explore the application effects of different machine learning algorithms in the field of pear tree water requirement prediction, an experiment was conducted in a Huangguan pear orchard planted in 2009, located in southern Hebei. Ecological factors of the orchard were collected, and the daily water requirements of Huangguan pear trees from 2018 to 2023 were calculated using the single crop coefficient method recommended by the FAO56 manual. Based on these calculations, a dataset of Huangguan pear tree water requirements was established. Key factors influencing the water needs of the pear trees, including temperature, wind speed, and soil conditions, were selected as input features. Four machine learning models—Support Vector Regression (SVR), Random Forest Regression (RF), Polynomial Regression, and Back-Propagation Neural Network (BPNN)—were used to predict the water requirements of the Huangguan pear trees. Through repeated training and testing, the predictive accuracy, generalization ability, and stability of the four models were compared. The results indicate that all four models achieved R^2 values above 0.90, with the Support Vector Regression model demonstrating the best overall performance. The R^2, RMSE, MAE, and max-error values for the SVR model were 0.92, 0.73, 0.52, and 3.21, respectively, with stable results across multiple experiments. Therefore, the Support Vector Machine learning method can accurately predict the water requirements of Huangguan pear trees, providing significant practical value for future smart and precision irrigation in pear orchards.

Key words: water requirement prediction; Huangguan pear; machine learning; SVR; BPNN; Random Forest; Polynomial Regression

0 引 言

中国水资源供需矛盾日益尖锐,农业用水浪费严重、水污染问题突出已成为严重制约国民经济可持续发展的瓶颈[1]。目前的果园多按照“丰水高产”的理论采用大水漫灌、多次灌溉、超量灌溉等方式,致使树冠营养生长旺盛、土壤板结以及养分流失,严重降低了水资源的利用效率[2]。河北省梨产量位于全国首

基金项目: 河北省现代农业产业技术体系梨产业创新团队建设项目(HBCT2024170204)。
作者简介: 许玉辰(2000—),女,硕士研究生。主要研究方向:机器学习,智能信息处理。
通信作者: 及歆荣(1978—),女,博士,副教授,硕士生导师。主要研究方向:分布式机器学习,无线传感器网络。Email:jixinrong2008@163.com。
收稿日期: 2024-09-06

位，黄冠梨以其品质好、价格较高且稳定的优点，已经成为河北省主栽品种之一，但黄冠梨在灌溉过程中灌溉需水量的多少仍然依赖于人工经验。智慧农业作为新时代新征程加快农业现代化进程的主攻方向，其中的智慧灌溉是未来农业的新趋势。智慧灌溉利用物联网、云计算、人工智能等新一代信息技术，对灌溉系统进行智能化控制和优化，实现水资源高效利用与作物生长需求相匹配的目标。针对黄冠梨树目前传统灌溉方式引发的水资源浪费问题，利用现代信息采集和处理技术构建其需水量预测模型，对实现黄冠梨智慧灌溉具有重要的现实意义。

近年来，随着机器学习算法的不断发展和广泛应用，叠加农业现代化发展的迫切要求，众多高校和科研院所的研究者们纷纷开展农业灌溉与物联网、机器学习相结合的研究[3]。李莉等提出随机森林（Random Forest）结合门控循环单元（Gated Recurrent Unit）神经网络的方法预测温室番茄结果前期蒸腾量[4]。谢家兴等对比了长短期记忆（Long short-term memory，LSTM）、极限学习机（Extreme learning machine，ELM）和广义回归神经网络（General regression neural network，GRNN）方法对柑橘园蒸散量的预测效果，指出 LSTM 模型的表现是 3 种模型中最优的[5]。何淑林等提出在 LSTM 神经网络模型的基础上加入注意力（Attention）机制来预测果树需水量[6]。目前结合机器学习的农作物需水量预测研究工作大多针对番茄、葡萄、小麦、玉米等作物，鲜见梨树需水量预测的相关研究。

本文研究使用欧洲中期天气预报中心 ERA5-Land 气象数据集，依据《FAO Irrigation and Drainage Paper NO. 56 Crop Evapo-transpiration》[7]中提出的单作物系数法，并结合 Penman-Monteith[8]公式计算梨树蒸散量作为梨树实际需水量，以此构建梨树需水量模型训练数据集。利用支持向量机、随机森林、多项式回归、神经网络模型 4 种机器学习方法进行模型训练，选取果园实测数据进行模型测试，并对 4 种模型的预测精度、泛化能力和稳定性 3 个方面进行对比分析。最终，挑选出适合黄冠梨树需水量预测的模型训练方法，以实现黄冠梨智慧灌溉系统中需水量预测模型的精准性与普适性。

1　理论框架

1.1　作物需水量的计算方法

能够准确计算出作物需水量是实现智能算法控制灌溉的关键技术要点。作物需水量指作物在正常健康的成长过程中维持生长所消耗的水量，主要取决于作物的种类、品种、生长阶段以及所处的气候和土壤条件，消耗于土壤和叶片表面的蒸腾作用、地表渗漏、地表径流等。对于梨树而言，由于其冠层茂盛叶片总面积较大，吸收的水分 95%消耗于叶片的蒸腾作用[9]。作物蒸散量（Evapotranspiration，ET）是指作物植株的蒸腾量与土壤的蒸发量之和，是梨树作物需水量的主要消耗因素，也是农业灌溉用水的主要消耗项[10]。对 ET 影响最大的因素包括作物系数、温度、相对湿度、风速、日辐射量等。常用的蒸散观测方法有蒸渗仪法、涡度相关法、基于能量平衡的波文比法等[11]。本文计算蒸散量使用的是单作物系数法，并将蒸散量作为作物需水量研究的基础数据。作物系数法计算作物实际蒸散量的公式来自《FAO Irrigation and Drainage Paper NO. 56 Crop Evapo-transpiration》[7]，即 FAO56 手册。通过参考作物蒸散量（ET_0）及作物系数（K_C）计算实际蒸散量，其中将各种天气状况的影响纳入 ET_0，作物特征纳入 K_C 系数。计算公式如下：

$$ET = K_C ET_0 \tag{1}$$

其中：ET、ET_0 分别为参考作物、目标作物蒸散量，单位 mm/d（毫米每天）；K_C 为目标作物的作物系数。

本文选用彭曼公式计算 ET_0，彭曼公式（又称彭曼蒙特斯公式，Penman-Monteith Equation）是由英国气象学家 H. L. 彭曼提出的，用于蒸散量 ET_0 的半经验半理论公式。根据宋妮等在关于 Penman 修正式和 Penman-Monteith 公式的差异分析文献[12]中所述，Penman-Monteith 公式受季节、地域等情况影响较小，稳定性较高。具体公式如下：

$$ET_0 = \frac{0.408\Delta(R_n - G) + \gamma \dfrac{900}{T + 273} u_2 (e_s - e_a)}{\Delta + \gamma(1 + 0.34u_2)} \tag{2}$$

其中，R_n 为净辐射，单位 MJ/（m^2 · d）；G 为土壤热通量，单位 MJ/（m^2 · d），当计算时段小于一个月时可忽略不计；γ 为干湿表常数，单位 kPa/℃；T 为平均温度，单位℃；e_s、e_a 分别为饱和水汽压、实际水汽压，单位 kPa；Δ 为饱和水汽压与温度曲线的斜率，单位 kPa/℃；u_2 为两米高度处风速，单位 m/s。部分主要参数的计算公式如下：

$$R_n = R_{ns} - R_{nl} = 0.77 R_A \left(a + b \frac{n}{N}\right) -$$

$$\sigma\left(\frac{T_{\max,K}^4 + T_{\min,K}^4}{2}(0.34 - 0.14\sqrt{e_a})\right)\left(1.35\frac{R_s}{R_{so}} - 0.35\right) \quad (3)$$

$$\gamma = 0.665 \times 10^{-3} \times 101.3 \times \left(\frac{293 - 0.0065H}{293}\right)^{5.26} \quad (4)$$

$$\Delta = \frac{4098\left(0.6108\exp\left(\frac{17.27T}{T + 273.3}\right)\right)}{(T + 237.3)^2} \quad (5)$$

式中，R_{ns}、R_{nl}、R_A、R_S、R_{SO} 分别为净短波辐射、净长波辐射、碧空太阳总辐射、短波辐射、晴空太阳辐射，单位 MJ/(m² · d)；$T_{\max,K}$、$T_{\min,K}$ 分别为最大、最小绝对温度，1K = 1 ℃ +273.16；σ 为 Stefan-Boltzmann 常数，为 4.903×10⁻⁹ MJ/(K⁴ · m² · d)；H 为海拔高度，单位 m。更具体的参数及计算公式可见参考文献[13-15]。

根据 FAO56 手册以及相关文献研究确定，由于不同物候期蒸散量的差异，同一作物的 K_C 会随着物候期的推移而变化。为了便于研究，本文将黄冠梨生长季的生长过程大致分为 4 个物候期，从新叶开始萌芽至叶片地面垂直投影覆盖率达 10% 为萌芽期(Germination stage)，从叶片地面垂直投影覆盖 10% 到全覆盖为新梢速生长和幼果期(Rapid-growing stage，Young fruit stage)，从叶片地面垂直投影全覆盖到果实成熟为果实迅速膨大和成熟期(Fruit-expansion stage，maturation stage)，从果实成熟采收开始到叶片自然落完称为果实采后落叶期(Leaf-fall stage)。新叶萌芽后一段时间内，K_C 值较小。从萌芽期的初始值(K_{ini})开始，K_C 值开始增加，在黄冠梨新梢速长期，枝条叶片快速发育达到最茂盛阶段，K_C 达到最大值($K_{C_{mid}}$)。在果实采后落叶期，由于黄冠梨叶片逐渐衰老，Kc 开始下降，直到整个生长季结束时达到较低值，即 $K_{C_{end}}$[7]。单作物系数法的广义作物系数曲线如图 1 所示。

黄冠梨树为多年生木本作物，根据邯郸地区观察可知，从萌芽到落叶整个生长季为 3 月 1 日—10 月 31 日，具体物候期划分及其对应作物系数[16]见表 1，无叶期为 11 月—次年 2 月，采用默认作物系数 0.4。

经过黄冠梨各物候期及其相应作物系数的确定，可构建黄冠梨作物系数曲线。该曲线反映了黄冠梨作物系数随生长阶段的变化规律。其形态体现了黄冠梨树整个生长季中，树体与叶片地面垂直投影变化对 ET_C/ET_0 比值的影响。从曲线上可以得到生长季内任意时期的 K_C 因子和 ET_C。

黄冠梨树的作物系数曲线如图 2 所示。

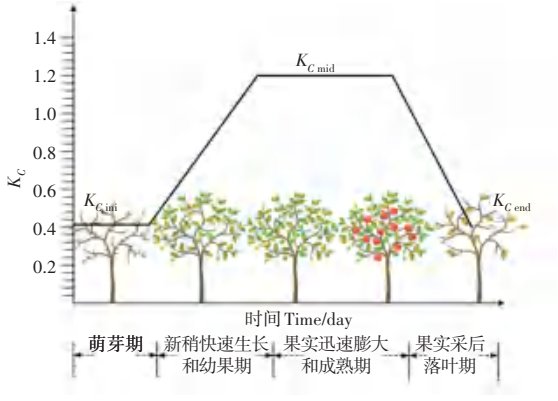


图 1 单作物系数法的广义作物系数曲线

Fig. 1 Generalized crop coefficient curve for the single crop coefficient method

表 1 黄冠梨树物候期划分及作物系数(K_C)取值

Table 1 Phenological stages and corresponding crop coefficient (K_C) values for Huangguan pear tree

物候期	时间(月日)	作物系数
萌芽期	03.01—03.31	0.80
新梢快速生长和幼果期	04.01—06.30	0.80—1.20
果实迅速膨大和成熟期	07.01—08.31	1.20
果实采后落叶期	09.01—10.31	1.20—0.85



图 2 黄冠梨树作物系数(K_C)曲线图

Fig. 2 Crop coefficient (K_C) curve chart for Huangguan pear trees

1.2 基于机器学习的预测方法

回归预测是统计学和机器学习领域的一个重要分支，旨在建立自变量与因变量之间的映射关系，实现对未知因变量值的预测。机器学习方法在处理回归预测问题时，展现出了强大的建模能力和灵活性。传统的线性回归模型虽然简单易懂，但往往难以捕捉数据中的复杂关系。而机器学习方法，能够自动学习并适应数据的内在模式，从而构建出更加具有精确性和鲁棒性的预测模型。以下介绍 4 种经典的

机器学习回归模型：支持向量回归、多项式回归、随机森林回归以及神经网络。

（1）支持向量回归（Support Vector Regression，SVR）基于支持向量机（SVM）的原理进行扩展，将数据映射到高维特征空间中，使得在这个高维空间中自变量与因变量具有良好的线性回归特性。在高维空间中拟合线性模型后，再将拟合结果映射回原始空间，从而得到非线性回归模型。SVR 模型的鲁棒性强、泛化性能好、灵活性高，但性能受参数选择影响较大，且在大规模数据集上训练耗时较长[17]。

（2）随机森林回归（Random Forest Regression，RF）通过构建多棵决策树并对所有决策树的预测结果取平均值来提升预测精度和模型稳定性。在随机森林中，每棵决策树都是基于原始数据集的随机子集和特征子集进行训练，这种方式能有效减少过拟合，提高模型的泛化能力。但由于需要训练多棵决策树，因此随机森林的训练时间通常比单棵决策树更长[18]。

（3）多项式回归（Polynomial Regression）使用一个 n 次多项式函数来近似描述目标变量和输入变量之间的关系。在线性回归的基础上，通过引入输入特征的非线性项（例如平方项、立方项等），来拟合更复杂的非线性的数据关系，增加了模型的灵活性。虽然多项式回归可以适应非线性关系，但其可能无法捕捉到某些复杂的非线性模式。

（4）BP 神经网络（Back－Propagation Neural Network，BPNN），即反向传播神经网络，是 Rumelhart 与 McClelland 等科学家在 1986 年提出的一种神经网络架构。其核心机制在于，当网络输出与预设目标不符时，会触发误差反向传播过程。此过程中，误差信号从输出层逆向传递至隐藏层，并最终抵达输入层，每一层都通过计算偏导数的方式将误差分配到其前一层的神经元上。基于这些误差分配，系统自动调整神经元之间的连接权重以及各自的激活阈值。通过迭代优化这一过程，BP 神经网络不断优化其内部结构，调整输入层、隐藏层与输出层之间连接的强度与阈值，使误差沿着梯度下降的方向逐步减小。这一过程持续至误差降低到规定范围之内或达到预设的迭代次数上限。BP 神经网络具有卓越的复杂模式分类能力和强大的多维函数映射特性，特别适用于处理难以用简单线性模型描述的复杂非线性问题，展现出了极高的灵活性和适应性[19]。此外，其还具有自学习和自适应能力，能够通过训练自动提取输入和输出数据之间的内在规律。

在以上 4 种机器学习方法中，支持向量回归通过寻找一个超平面来最大化预测值与实际值之间的间隔，对噪声和异常值具有较好的鲁棒性；随机森林回归通过集成多个决策树的结果，提高了预测的稳定性和准确性；多项式回归通过引入高次项来构建一个多项式函数来适应非线性关系具有灵活性和广泛适用性等优点，但也存在过拟合、对异常值敏感等缺点。神经网络能够通过学习大量的数据样本，自动提取出数据的深层特征，从而构建出高度复杂的预测模型。

2　实验设计

2.1　实验目的

本实验将黄冠梨树最近 5 年内的生态因子数据设置为最近 3 年期和最近 5 年期两种不同规模的训练数据集，选用了 4 种经典的机器学习方法：支持向量机回归（SVR）、随机森林（RF）、多项式回归（Poly）和反馈型神经网络（BPNN）分别进行模型训练，选取果园实测数据进行模型测试。从模型预测正确率、泛化能力和稳定性三方面进行模型预测结果的对比分析，以期得到整体效果较好的需水量预测模型，进而挑选出适合黄冠梨树需水量预测的模型训练方法。

2.2　数据集准备

本研究的试验区域位于河北省邯郸市永年区河北工程大学现代农业示范基地（东经 114.5°，北纬 36.77°），海拔 65 m。气象数据来源于 ERA5－Land（欧洲中期天气预报中心再分析数据集）。该数据集使用 H－TESSEL 作为基础陆面模型，空间分辨率为 9 km（约为 0.08°），采用（八面体）简化高斯网格（表示为 TCo1279），以小时为步长提供数据，其中累积量数据计算方法为 UTC 时间当日 00:00 时到当前时间的累积量，累计时间最多为 24 小时。

实验使用 2018－2023 年间共 52 584 条数据，具体包含每小时的温度、露点温度、大气压强、10 m 处 V 风分量、10 m 处 U 风分量，土壤温度、土壤湿度、地表净短波（太阳）辐射累积量，其中土壤部分的数据分别取 3 个层级中的温度、湿度（Layer1 为 0－7 cm；Layer2 为 7－28 cm；Layer3 为 28－100 cm）。

将原始小时级的数据按北京时间日期分组，每 24 条分为一组提取日值数据。根据每小时温度取日平均温度、日最高温度、日最低温度，根据 10 m 处 V 风、U 风分量计算平均总风速，根据累积量测量规

则取每日地表净短波(太阳)辐射量,其余的数据如露点温度、大气压强、三层土壤温湿度等取每日平均值。根据上述 1.1 作物需水量的计算方法中的计算公式,对每条数据包含的信息进行计算得到日需水量,作为数据的真实值。由此得到共 2 183 条可用数据。使用其中的月份、日期、平均温度、最高温度、最低温度、风速、气压、3 层土壤温度、3 层土壤湿度共 13 项特征及其对应真实值组建模型训练集和测试集。

为了缩小不同属性的数据量纲的影响,对数据的特征进行缩放处理,此处选用极差标准化缩放做数据的归一化处理,所有数据均归一化到(0,1)区间。

2.3 模型训练及预测

本实验基于 Python 3.9.6 环境实现,其编译信息为: tags/v3.9.6:db3ff76, Jun 28 2021, 15:26:21, 运行环境为 MSC v.1929 64 bit (AMD64),运行机器为 Intel Core™ i5-8250U CPU @1.60 GHz。在实验中为了减小初始化时的随机性对各机器学习模型性能的影响,同时为了模型训练结果的可复现性,对随机参数 random_state 进行了统一设置。利用决定系数(R^2)、平均绝对误差(Median Absolute Error,MAE)、均方根误差(Root Mean Squared Error,RMSE)等评价参数对不同模型的预测效果进行比较,以寻找更加适合梨园灌溉系统实际使用的预测模型训练方法。

(1)SVR 模型的训练使用 Scikit-learn 库中的支持向量机回归对象实现。创建 SVR 模型,设置使用径向基函数(Radial Basis Function, RBF)作为核函数。正则化参数 C;设置为 10 以控制正则化的强度。指定不敏感损失函数的参数 ε 的带宽为 0.1,若预测值与实际值的差的绝对值小于 ε,则不会计入损失函数。γ 设置为'scale'以适应输入数据的特征尺度。

(2)RF 模型的训练使用 sklearn.ensemble 模块中的 RandomForestRegressor 类实现。以 CART 决策树作为弱学习机,且随机森林与 CART 决策树一致,采用基尼系数作为特征选择的依据。设置随机森林中决策树的数量为 200,叶节点所需最小样本数量为 1,内部节点(非叶子节点)所需的最小样本数为 5,寻找最佳分割点时,考虑的特征数为 $\sqrt{13}$ 个,对每棵树不设置最大深度即默认其可无限增长。

(3)多项式回归模型的训练使用 Scikit-learn 库中的 PolynomialFeatures 和 LinearRegression 实现。指定多项式的度数(即最高次幂)degree = 2。通过 PolynomialFeatures 将数据转换为多项式特征,其中包含了原始特征的所有多项式组合。将所有的这些特征组合传递给 LinearRegression 进行拟合。在 LinearRegression 中,模型通过遍历系数组合以最小化预测值与实际值之间的平方误差,构建出对应多项式公式。

(4)BPNN 模型的训练使用 Keras 库(集成在 TensorFlow 中,常用于深度学习模型构建的高级 API)构建序贯模型。本文研究的是回归问题,故输出层使用单个输出单元,隐藏层使用 ReLU(Rectified Linear Unit)激活函数,能够加速训练过程并缓解梯度消失问题。输出层使用线性激活函数以保证输出值的连续性。

以上算法的具体参数在实验中均使用网格搜索方法,以决定系数 R^2 为评价指标进行选优,决定系数 R^2 用于衡量模型预测值与真实值的拟合程度,其计算公式为:

$$R^2 = 1 - \frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)}{\sum_{i=1}^{n}(y_i - \hat{y}_i)} \tag{6}$$

式中,y_i 为第 i 个真实值,$\hat{y}_i$ 为第 i 个预测值。R^2 值越接近 1,表示模型预测越准确。

为了防止模型过拟合,参数寻优阶段对于各模型均使用交叉验证的方法获得评价参数[20]。将原始数据集共分成 K 个子集,每次剩下一个子集作为测试集,其余 $K-1$ 个子集作为训练集来评估模型。整个过程中每个子集都有机会作为测试集,"划分-训练-评价"流程会重复 K 次。这样可以得到 K 组评估结果,取这 K 组评估结果的平均值作为对模型性能的最终估计。本实验中取 $K=5$。

构建模型完成之后,分别使用规模不同的两个数据集——"数据集 3"和"数据集 5"对 4 个模型进行训练。其中"数据集 3"包含 2020-2022 共 3 年间的 1 091 条数据,"数据集 5"包含 2018-2022 共 5 年间的 1 819 条数据。模型训练完成后使用 2023 年间 364 条数据作为测试集,通过预测值与真实值间的计算得到决定系数(R^2)、平均绝对误差(MAE)、均方根误差(RMSE)、最大误差(Max-Error)几项评价参数。具体计算公式如下:

$$\mathrm{MAE} = \frac{1}{n}\sum_{i=1}^{n}|y_i - \hat{y}_i| \tag{7}$$

$$\mathrm{RMSE}=\sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i-\hat{y}_i)^2} \tag{8}$$

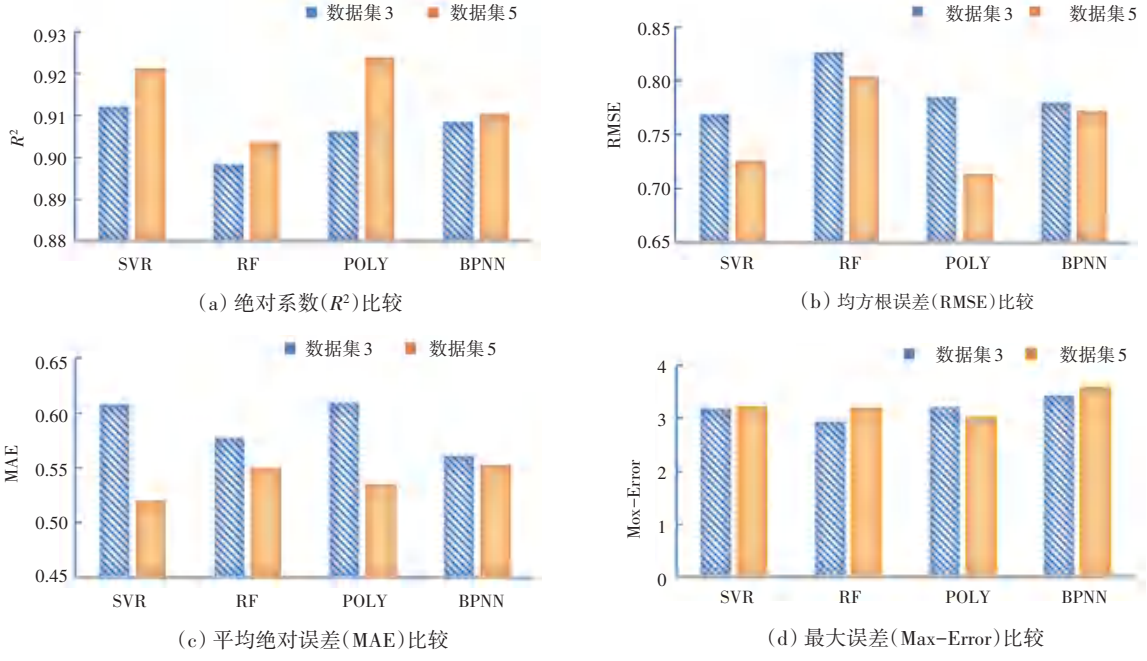
$$\mathrm{Max-Error}=\max_{1\leqslant i\leqslant n}|y_i-\hat{y}_i| \tag{9}$$

其中，y_i 为第 i 个真实值，$\hat{y}_i$ 为第 i 个预测值。每次训练时保证对 4 种模型使用相同训练集和相同数据次序进行训练，避免训练集因素对不同模型的性能比较结果产生影响。

3　实验结果分析

分别使用规模较小的数据集 3 和规模较大的数据集 5 对各模型重复训练 100 次，每次训练后在测试集上测试可得到决定系数（R^2）、均方根误差（RMSE）、平均绝对误差（MAE）、最大误差（Max-Error）4 项评价结果。分别对评价结果取 100 次均值，得到各模型在两个数据集上的各项测试结果，如图 3 所示。

由图 3 可以看出，随着增大训练数据集的规模，各模型的 R^2 值普遍增大，MAE 和 RMSE 都有所减小，max-error 变化不甚明显。这表明了训练集的规模大小一定程度上影响了模型的拟合效果，更多的训练样本训练出的模型拟合效果更好。其中多项式回归模型受数据集规模影响最大，其 R^2 值从 0.906 提升至 0.924，MAE 和 RMSE 分别减少了 12%（降幅为 0.074）和 9%（降幅为 0.071）。SVR 模型的 MAE 降幅最多，表明增大训练集规模有效提升了 SVR 模型的准确性。相比之下，BPNN 模型的效果虽然也有小幅提升，但整体来说受训练集规模影响最小。从评价指标横向比较来看 SVR 和多项式回归模型的预测效果更优，R^2 值能够达到 0.92 左右，MAE 和 RMSE 值也相对更低。RF 和 BPNN 模型的 R^2 值都在 0.90 左右，Max-Error 值随训练集规模增大略有上升，受到了模型随机性和整体预测次数增大的影响。



（a）绝对系数（R^2）比较　（b）均方根误差（RMSE）比较

（c）平均绝对误差（MAE）比较　（d）最大误差（Max-Error）比较

图 3　各模型在不同规模训练集上的表现

Fig. 3　Performance of various models on different sizes of training datasets

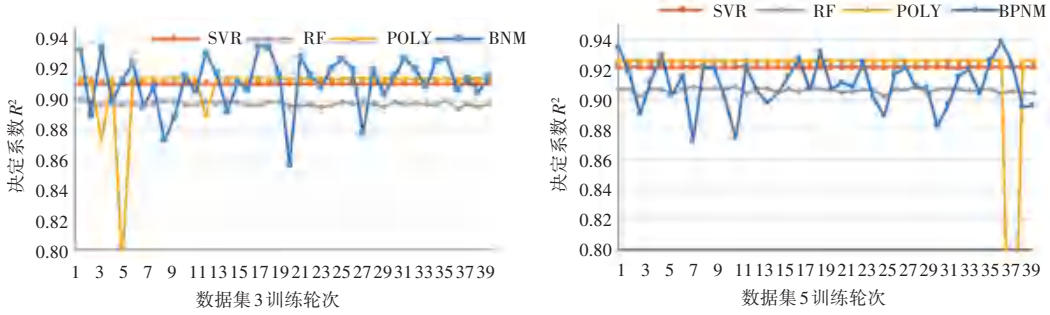
将日需水量按每 10 天取和，得到每旬作物需水量，将 4 种模型旬需水量预测值与旬需水量真实值输出为折线图可以直观看到各模型拟合情况，如图 4 所示。可以看到各模型的预测结果基本符合梨树需水量的真实走势，但在峰值部分出现预测结果偏低的情况。

使用两个训练集分别对各模型进行重复训练测试以验证各模型的稳定性。取 100 次实验的前 40 次结果进行可视化输出，将测试结果（以 R^2 值为综合标准）绘制折线图，如图 5 所示。

可以看出 4 种模型在两个训练集上的稳定性表现存在差异。其中 BPNN 模型 R^2 值上下波动明显，幅度较大，表明其预测效果不够稳定。这是因为 BPNN 模型本身就具有一定的不确定性，即使每次

使用相同训练集也难以得到稳定的实验结果。

RF 模型预测效果有一定浮动但整体波动范围不大,尚在可接受范围之内。两种模型的预测结果波动情况均因为训练集规模的增大而呈现略微增大的趋势。多项式回归模型平均准确率最高但在多次测试中小概率会产生跳水式巨大误差。SVR 模型预测效果较佳且具有良好的稳定性,在 4 种模型中综合表现最好。



(a) 数据集 3 训练拟合图 (b) 数据集 5 训练拟合图

图 4 黄冠梨旬需水量拟合折线图

Fig. 4 Fitted line chart of ten-dayl water requirement for huangguan pears

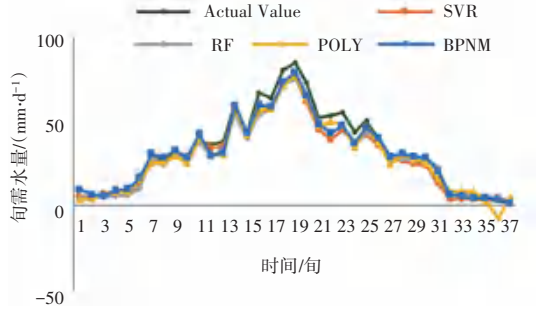


图 5 四种模型在重复实验下的 R^2 值折线图

Fig. 5 Line chart of R^2 values for four models under repeated experiments

4 结束语

为构建适合梨树的需水量预测模型,本研究基于 ERA5-Land 气象数据集,采用单作物系数法和 Penman-Monteith 公式计算得到黄冠梨树蒸散量,以此作为其实际需水量的方式构建了样本数据集,利用支持向量机、随机森林、多项式回归、神经网络 4 种经典机器学习模型训练方法进行梨树需水量模型训练和模型预测,并对 4 种预测模型的正确率、泛化能力以及稳定性进行了比较。实验结果表明,增加模型的训练样本数据量,4 种模型的预测正确率均有提高,其中多项式回归和 SVR 提升效果非常明显,表明增加训练样本数可明显提升模型预测正确率;从 MAE 和 RMSE 来看,SVR 和多项式回归的泛化能力较好;从多次实验来看预测模型的稳定性,SVR 稳定性最好。综上,在本研究测试的 SVR、RF、多项式回归和 BPNN 四种学习方法中,SVR 整体表现最佳。因此,在黄冠梨树智慧灌溉系统中,可选择 SVR 作为其需水量预测模型训练方法。

参考文献

[1] 王秀鹃. 农业节水的路径组合与绩效研究[D]. 泰安:山东农业大学, 2019.

[2] 单长河, 汪艳萍. 超量灌溉对农田耕作区水土肥流失的影响[J]. 中国农业资源与区划, 2015, 36(6): 11-16.

[3] 陆红飞, 王涛, 乔冬梅, 等. 物联网在农业灌溉中的应用:从灌溉自动化到智慧灌溉[J]. 灌溉排水学报, 2023, 42(1): 87-99.

[4] 李莉, 李伟, 耿磊, 等. 基于 RF-GRU 的温室番茄结果前期蒸腾量预测方法[J]. 农业机械学报, 2022, 53(3): 368-376.

[5] 谢家兴, 高鹏, 孙道宗, 等. 基于长短期记忆的柑橘园蒸散量预测模型[J]. 农业机械学报, 2020, 51(S2): 351-356.

[6] 何淑林, 刘慧敏, 金立强, 等. 基于神经网络算法的果树需水预测研究[J]. 灌溉排水学报, 2022, 41(1): 19-24.

[7] ALLEN R G, PRUITT W O, WRIGHT J L, et al. A recommendation on standardized surface resistance for hourly calculation of reference ET_0 by the FAO56 Penman-Monteith method [J]. Agricultural Water Management, 2005, 81 (1): 1-22. DOI:10.1016/j.agwat.2005.03.007.

[8] 马钊, 李鹏程, 刘洪伟, 等. 基于辐射改进 Penman-Monteith 模型估算粮食主产区参考作物蒸散量[J]. 节水灌溉, 2024 (3): 24-33.

[9] 唐书玥, 任思琪, 李子怡, 等. 我国不同参考作物蒸散经验模型适用性分析及参数修正[J]. 中国农业大学学报, 2023, 28(10): 1-14.

[10] 尹起. 机器学习预测参考作物蒸散量在智能灌溉系统中的应用研究[D]. 乌鲁木齐:新疆大学, 2021.

[11] 武佳乐, 李银坤, 张钟莉, 等. 基于单作物系数法的温室秋季生菜蒸散量估算及验证[J]. 灌溉排水学报, 2024, 43(4): 9-14.

[12] 宋妮, 孙景生, 王景雷, 等. 基于 Penman 修正式和 Penman-Monteith 公式的作物系数差异分析[J]. 农业工程学报, 2013, 29(19): 88-97.

[13]黄松，戴佳琦，赵宝山，等. 气候变化下广东省参考作物蒸散量变化趋势及其影响因素分析[J]. 人民珠江，2023，44(12)：35-43.

[14]付晓刚，周亚红，毕攀，等. 基于彭曼公式的作物灌溉需水量确定方法研究[J]. 人民黄河，2009，31(5)：85-87.

[15] IPPOLITO M，CARO D D，CANNAROZZO M，et al. Evaluation of daily crop reference evapotranspiration and sensitivity analysis of FAO Penman-Monteith equation using ERA5-Land reanalysis database in Sicily，Italy [J]. Agricultural Water Management，2024，295：108732. DOI：10.1016/J.AGWAT.2024.108732.

[16]杨天一，王军，张红梅，等. 基于单作物系数法的华北平原典型农业生态系统蒸散规律研究[J]. 中国生态农业学报(中英文)，2022，30(3)：356-366.

[17]胡乐怡，付晓雷，蒋晓蕾，等. 基于 LSTM、RF、SVR 三种机器学习方法的径流预测研究[J]. 水文，2024，44(5)：17-24. DOI：10.19797/j.cnki.1000-0852.20230347.

[18]于浩. 基于随机森林和 LSTM 网络的股指预测模型优化研究[D]. 哈尔滨：哈尔滨商业大学，2023.

[19]杨昊. 基于 BP 神经网络和模糊控制的智能灌溉系统设计[D]. 淮南：安徽理工大学，2019.

[20] TIAN Z. A combined prediction approach based on wavelet transform for crop water requirement [J]. Water Supply，2020，20(3)：1016-1034. DOI：10.2166/ws.2020.024.