北京林业大学学报
JOURNAL OF BEIJING FORESTRY UNIVERSITY

基于微钻阻力与LightGBM的活立木木材密度区间分类预测模型

唐政权 何夏薇 黄元 宁华珙 张方友 陈煜佳 孙嘉骏 刘天颐

An interval classification prediction model for standing tree wood density based on micro-drilling resistance and LightGBM

Tang Zhengquan, He Xiawei, Huang Yuan, Ning Huagong, Zhang Fangyou, Chen Yujia, Sun Jiajun, Liu Tianyi

引用本文:
唐政权, 何夏薇, 黄元, 宁华珙, 张方友, 陈煜佳, 孙嘉骏, 刘天颐. 基于微钻阻力与LightGBM的活立木木材密度区间分类预测模型[J]. 北京林业大学学报, 2026, 48(10):26–36. doi: 10.12171/j.1000–1522.20250449
Tang Zhengquan, He Xiawei, Huang Yuan, Ning Huagong, Zhang Fangyou, Chen Yujia, Sun Jiajun, Liu Tianyi. An interval classification prediction model for standing tree wood density based on micro–drilling resistance and LightGBM[J]. Journal of Beijing Forestry University, 2026, 48(10):26–36. doi: 10.12171/j.1000–1522.20250449

在线阅读 View online: https://doi.org/10.12171/j.1000–1522.20250449

您可能感兴趣的其他文章

Articles you may be interested in

无损检测指标与木材横纹局部抗压强度和密度的相关性

Correlations between nondestructive testing indicators and local transverse compressive strength as well as wood density

北京林业大学学报, 2024, 46(5): 136–142. https://doi.org/10.12171/j.1000–1522.20230367

基于广义加性模型的红松树干纵向木材密度模型研建

Developing a longitudinal wood density model for *Pinus koraiensis* stems based on generalized additive models

北京林业大学学报, 2026, 48(7): 48–57. https://doi.org/10.12171/j.1000–1522.20250465

基于beta回归的迎春5号杨树树干密度混合效应模型

Mixed effect model of stem density of *Populus nigra* × *P. simonii* based on beta regression

北京林业大学学报, 2023, 45(5): 67–78. https://doi.org/10.12171/j.1000–1522.20220450

广东省含参数分级的阔叶林分生物量生长模型

Stand biomass growth model of broadleaved forest with parameter classification in Guangdong Province of southern China

北京林业大学学报, 2022, 44(5): 19–33. https://doi.org/10.12171/j.1000–1522.20210403

耦合时序特征的林分类型遥感识别

Remote sensing classification of stand type coupled with time series features

北京林业大学学报, 2024, 46(1): 68–81. https://doi.org/10.12171/j.1000–1522.20230093

基于支持向量回归（SVR）的马尾松木材脱脂率预测

Deresination rate prediction of Masson pine wood based on support vector regression (SVR)

北京林业大学学报, 2025, 47(3): 151–161. https://doi.org/10.12171/j.1000–1522.20240359

DOI:10.12171/j.1000−1522.20250449

基于微钻阻力与 LightGBM 的活立木木材密度区间分类预测模型

唐政权[1]　何夏薇[1]　黄　元[1]　宁华珙[2]　张方友[2]　陈煜佳[1]　孙嘉骏[1]　刘天颐[1]
（1. 华南农业大学林学与风景园林学院，广东 广州 510642；2. 英德市林业科学研究所，广东 清远 513055）

摘要:【目的】微钻阻力法是活立木木材密度无损检测的主要技术之一，然而，现有基于该技术的回归预测模型易受仪器、操作及环境等多源干扰，存在稳定性差、通用性弱的问题，制约了其在工程实践（如木材分级、野外快速检测）中的应用，亦限制了科研中的可靠分析。为提升微钻阻力法在木材密度预测中的稳健性与实用性，本研究旨在探索一种能够包容测量不确定性的建模方法。**【方法】**通过转换建模思路，将连续密度预测问题重构为离散区间分类任务。以木荷为例，基于经数据预处理与异常值剔除后保留的 481 组微钻阻力－密度匹配数据，采用基于机器学习法（LightGBM）的区间分类方法构建预测模型，并与多种传统回归模型进行系统性对比。该方法通过密度区间输出包容测量波动，从而提升模型的适应能力。**【结果】**所构建的区间分类模型整体准确率达 0.876，宏平均 F_1 值为 0.878，在木材密度分级任务中表现稳定；传统回归模型的最优 R^2 为 0.771，更适用于连续密度预测。两类模型任务定位不同，分类指标与回归指标不宜直接比较。该区间分类模型在各密度区间上均表现稳健，精确率范围为 0.769 ~ 1.000；对于木材分级实践中常需明确区分的高、低密度两端区间，模型分类正确率均达到或超过 84%，显示出良好的判别能力。结果表明，区间分类策略能够提供比传统点估计更稳健的分级结果。**【结论】**所提出并验证的“区间分类”建模策略，通过将连续密度预测重构为区间分类任务，有效提升了微钻阻力法预测木材密度的实用性。该策略输出的密度等级直观适用于工程快速判断，其分类概率输出也为科学研究提供了结构化数据。研究为解决微钻阻力法应用中模型不稳定的共性问题，提供了一种侧重于稳健性与可用性的机器学习建模新视角。

关键词：微钻阻力法；木材密度；活立木；LightGBM；区间分类；木荷；有序分类；木材分级
中图分类号：S781.31　　**文献标志码：**A　　**文章编号：**1000−1522(2026)10−0026−11
引文格式：唐政权，何夏薇，黄元，等. 基于微钻阻力与 LightGBM 的活立木木材密度区间分类预测模型［J］. 北京林业大学学报，2026，48(10)：26−36. Tang Zhengquan, He Xiawei, Huang Yuan, et al. An interval classification prediction model for standing tree wood density based on micro-drilling resistance and LightGBM［J］. Journal of Beijing Forestry University, 2026, 48(10): 26−36.

An interval classification prediction model for standing tree wood density based on micro-drilling resistance and LightGBM

Tang Zhengquan[1]　He Xiawei[1]　Huang Yuan[1]　Ning Huagong[2]　Zhang Fangyou[2]
Chen Yujia[1]　Sun Jiajun[1]　Liu Tianyi[1]
(1. College of Forestry and Landscape Architecture, South China Agricultural University, Guangzhou 510642, China;
2. Yingde Institute of Forestry Science, Qingyuan 513055, China)

Abstract:［**Objective**］The micro-drilling resistance method is one of the principal techniques for non-destructive evaluation of wood density in standing trees. However, existing regression-based prediction models built upon this technique are susceptible to multi-source disturbances arising from instrumentation,

收稿日期：2025−11−11　**修回日期：**2026−10−07
基金项目：广东省林业局自然资源事务管理专项资金−生态林业建设 (省级组织实施) 项目（GDLZXY20220102，GDLZXY20230102），广东省林业局 2021 年省级生态公益林效益补偿资金省统筹经费项目。

第一作者：唐政权。主要研究方向：木荷材性变异。Email：1454828604@qq.com　地址：510642 广东省广州市天河区五山路 483 号华南农业大学林学与风景园林学院。
责任作者：刘天颐，副教授。主要研究方向：林木遗传育种。Email：tianyiliu@scau.edu.cn　地址：同上。

本刊网址：https://j.bjfu.edu.cn；https://journal.bjfu.edu.cn

operators, and environmental conditions, leading to poor stability and limited generalizability. These shortcomings constrain its application in engineering practice (e.g., wood grading and rapid field testing) and hinder reliable analysis in scientific research. To enhance the robustness and practical applicability of the micro-drilling resistance method for wood density prediction, this study aims to explore a modeling approach that accommodates measurement uncertainty. [**Method**] By transforming the modeling concept, the continuous density prediction problem was reconstructed as a discrete interval classification task. Taking *Schima superba* as an example, based on 481 sets of micro-drilling resistance–density paired data retained after data preprocessing and outlier removal, a predictive model was developed using an interval classification approach based on machine learning (LightGBM), and systematically compared with several conventional regression models. By outputting density intervals, this method accommodates measurement fluctuations, thereby enhancing the model's adaptability. [**Result**] The interval classification model achieved an overall accuracy of 0.876 and a macro-average F_1-score of 0.878, showing stable performance in wood density grading. Its precision across density intervals ranged from 0.769 to 1.000. For the high- and low-density intervals that are often critical for timber grading, the classification accuracy exceeded 84%, indicating good discriminative ability. These results suggest that the interval classification strategy can provide more robust grading results than conventional point estimation. [**Conclusion**] The proposed and validated "interval classification" modeling strategy, by reformulating continuous density prediction as an interval classification task, effectively enhances the practical applicability of the micro-drilling resistance method for predicting wood density. The density grades output by this strategy are intuitively suited to rapid engineering judgment, while its classification probability outputs provide structured data for scientific research. This study offers a new machine learning modeling perspective focused on robustness and usability, for addressing the common problem of model instability in the application of the micro-drilling resistance method.

Key words: micro-drilling resistance method; wood density; standing trees; LightGBM; interval classification; *Schima superba*; ordinal classification; wood grading

木材密度作为衡量木材材性的核心指标，显著影响木材的力学性能[1]、加工性能[2]及使用价值[3]，是木材加工、家具制造、建筑工程等领域选材的关键依据[4]。精准测定木材密度对提升木材资源利用效率[5]、推动林业产业高质量发展具有重要意义[6]。目前，木材密度测量方法主要包括体积法[7]、X 射线法[8]、Pilodyn 法[9]及微钻阻力法[10]等。其中，微钻阻力法通过电机控制钻针匀速钻入木材，依据钻针所受阻力间接推定木材密度，具备操作简便、对活立木损伤小（接近无损）、便携性强及数据处理简单等显著优势，已成为活立木木材密度无损测量的主要技术之一[11]。

近年来，众多学者致力于建立微钻阻力（R）与木材密度（ρ）间的预测模型，其研究路径主要沿两个方向推进：一是优化模型的数学形式，早期研究普遍采用简单的线性回归，但其决定系数（R^2）因干扰因素影响而在 0.21 ~ 0.87 之间剧烈波动[12−14]，模型稳定性与通用性差。为提升拟合优度，后续研究引入了更复杂的函数形式，例如线性−对数混合模型在跨树种建模中取得了 $R^2 = 0.961$ 的结果[10]，显著优于单一线性或对数模型。二是尝试将主要环境干扰因素显式纳入模型以校正偏差：研究表明，含水率显著影响马尾松（*Pinus massoniana*）、落叶松（*Larix* spp.）等树种的密度预测模型[15−16]，且理论上在模型中引入含水率因子可提升解释率。此外，近期研究进一步指出，仪器本身、仪器型号与操作者差异亦是影响模型稳定性的关键因素：不同微钻阻力仪之间因其本身性能，如进给速度、转速以及针与木材的摩擦特性[17]、仪器型号等不同而存在系统性偏差，而操作习惯则引入随机波动[18]，导致即使基于大量样本建立的模型也难以跨设备通用，需频繁重新校准。然而，这些改进均未突破传统"点估计"回归范式的固有局限。首先，其对数据质量与分布敏感，易受仪器本身性能、操作差异及环境瞬时波动等多源随机干扰的显著影响，导致模型在实际应用中的预测稳定性不足，性能指标波动大。更为关键的是，模型的泛化能力存在根本挑战：即便为单个树种专门建立的分模型，其性能也可能出现极大分化，如麻栎（*Quercus acutissima*）分模型的 R^2 可低至 0.397，而旨在降低建模工作量的多树种通用模型（总模型），其预测精度又显著低于特定树种的分模型[10]。这揭示了在"点

估计”框架下，模型精度与通用性难以兼顾。其次，旨在提升精度的改进往往以牺牲实用性与便捷性为代价，例如：依赖含水率等动态、难测变量的模型，虽可能提升实验室条件下的拟合度，却大幅增加了野外现场检测的复杂性与操作成本，因为含水率随树种、季节、环境而异，实时准确测定本身即为一项挑战；同样，针对不同性能状态或型号的仪器及操作者建立专属模型虽能提升局部精度，却以牺牲方法的标准化与可推广性为代价，弱化了其在实际应用中的便捷性与广泛适用潜力。因此，当前研究面临的关键挑战在于如何构建一个既能包容固有测量不确定性（包括仪器、操作、环境等多源干扰），又能为工程实践提供直观、稳定、无需复杂前置测量的密度预测模型。

为应对上述挑战，从“点估计”回归转向能够包容不确定性的“区间化输出”，成为一种潜在思路。区间概率理论关注不确定性条件下参数的可能取值范围及其概率分布，强调以区间而非单点刻画估计结果[19]，这为处理微钻阻力与密度测定中的离散性和测量波动提供了理论启示。借鉴这一思想，但并不在回归框架下构造连续密度真值的置信区间，而是将“区间化表达”与“概率化度量”落实到密度区间分类任务中：以微钻阻力为输入，以密度区间为输出类别，由机器学习模型给出样本属于各区间的分类概率。由此，模型不追求对连续密度真值的单点逼近，而是直接输出稳定的密度等级及分类信心，从而在任务层面包容仪器、操作和环境等因素引起的测量波动，提升微钻阻力法用于木材密度分级的稳健性与实用性。

实现这一区间分类预测，需要选择一种匹配林业研究现实条件（样本量有限、数据获取成本高）并能直接输出概率的算法。与需要海量训练数据的深度学习模型（如 ResNet、DenseNet）不同，梯度提升决策树框架下的 LightGBM 算法，通过其独特的优化策略，在中小规模数据集上通常能获得更优的泛化性能与效率[20]。它对于类别不平衡数据具有良好的适应性，并能直接提供区间概率输出，这使其成为构建稳健密度区间分类模型的更适宜选择。

木荷（*Schima superba*）作为我国南方重要的多功能树种，兼具生态价值与经济价值。其生长速度快、树干通直、材质均匀致密，力学性能优良，是地板、家具、工艺品制造的优质原料[21]；同时，木荷具有突出的抗火性能，是构建森林防火林带的核心树种[22]。发展木荷人工林不仅能增加国家优质木材储备，还能提升区域生态防护能力，实现生态保护与经济效益的协同提升[23]。然而，目前针对木荷木材密度的快速、无损预测研究仍较缺乏，现有微钻阻力预测模型在该树种上的适用性与稳定性尚未得到充分验证。这一现状使木荷成为合适的验证案例，可用以检验新型预测方法在应对树种特异性挑战方面是否有效。

基于此，本研究以广东省英德市木荷无性系种子园的木荷活立木为研究对象，采用基于机器学习的区间分类新范式，构建一种具有跨树种潜力的木材密度预测方法，以包容测量不确定性，应对微钻阻力法因多源干扰导致的预测不稳定问题。通过对比区间分类模型与传统回归模型在分级任务与连续预测任务中的表现，验证该范式在提升预测稳健性与结果可用性方面的表现。以木荷为实证案例，本研究既为木荷木材密度的快速分级提供实用预测工具，也为微钻阻力法在其他树种中的应用提供方法学参考。

1 材料与方法

1.1 材 料

试验地点位于广东省英德市林业科学研究所内的木荷无性系种子园（24°13′N，113°45′E），为省级木荷良种繁育基地。园区属亚热带季风气候，年平均气温 21.1 ℃，≥ 10 ℃ 年积温 6 200 ℃·d，年降水量 1 906.2 mm，年日照时数 1 631.7 h；土壤为赤红壤，pH 值 5.2 ~ 6.7，适宜木荷生长。选取种子园内保存较为完好的 164 个无性系，每无性系 3 株，共 492 株，为试验材料，经现场测定与初步有效性检查，共获得 488 组微钻阻力与基本密度初始匹配数据，用于后续预处理。

1.2 微钻阻力与基本密度测定

采用“微钻阻力值测定–钻取木芯–实验室密度测定”的流程[24]进行测定，具体操作如下：使用 IML-RESI PD 400 型木质检测仪在样本树胸径处沿南北方向进行微钻阻力测定。参考同类研究[17]并结合木荷的材性特征，测定过程中保持钻针倾斜角为 0°，并将进给速度和钻针转速分别设定为 50 cm/min 和 2 000 r/min。进给速度和钻针转速为仪器控制系统设定值，以保证测量条件的一致性。仪器自动记录以振幅百分比表示的微钻阻力值。测定过程中实时检查数据质量，如出现钻针未通过髓心或记录中断等异常情况，立即进行复测；复测数据正常则继续下一株样本的测定，复测后仍异常（如树干中空导致无法形成完整阻力—密度匹配数据）的样本不纳入后续分析；完成阻力测定后，标记阻力孔位置，并在同一胸径高度、同一南北方位，以原阻力孔为进锥导向，沿原针刺方向使用内径 1.2 cm 的生长锥钻取完整木芯。由于生长锥内径明显大于微钻阻力针直径，取

芯区域可覆盖原阻力测线及其邻近木质部，从而保证微钻阻力值与木芯基本密度在空间上的对应性。标注方向与编号后装入密封样品袋，供后续密度测定使用。

木材基本密度测定采用排水法[7]。木芯去除树皮与腐朽部分后，浸泡 72 h 至饱和，擦净表面水分后悬挂浸入盛有蒸馏水的量筒中，通过万分之一电子天平读取排开水的体积 V；随后将木芯置于 70 ℃烘箱中烘干至恒质量，称取其绝干质量 m；按公式(1)计算基本密度。按上述步骤重复测定 3 次，以 3 次测定结果的平均值作为该木芯的基本密度。

$$\rho = \frac{m}{V} \tag{1}$$

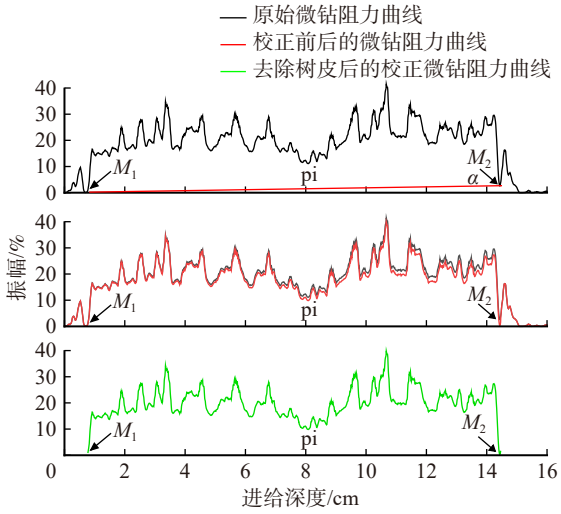
式中：ρ 为木材基本密度；m 为木芯绝干质量；V 为木芯饱和吸水后排开水的体积。

1.3　数据预处理

数据预处理包括树皮去除、去趋势处理(图 1)与平均微钻阻力值计算[25−26]，具体步骤如下：首先进行树皮去除：使用仪器配套软件 PD - Tools Pro 将原始阻力曲线数据(rgp 格式)导出为 txt 格式，导入 Origin 软件绘制原始阻力曲线，根据曲线特征识别树皮区域(曲线起始段急剧上升部分)与木质部区域，确定去皮直径的微钻起点(M_1)与终点(M_2)，剔除树皮区域数据，保留木质部区域数据；随后进行去趋势处理：从 M_1 到 M_2，因钻针摩擦累积，微钻阻力呈现递增趋势，导致木质部区域阻力值比实际值偏大，需进行去趋势处理。根据式(2)计算阻力递增斜率。

$$\tan\alpha = \frac{R_2 - R_1}{L} \tag{2}$$

式中：α 为钻针摩擦导致的阻力递增角度；R_1 为 M_1 的微钻阻力值；R_2 为 M_2 的微钻阻力值；L 为去皮直径总长度；$\tan\alpha$ 为阻力递增斜率。



α 为钻针摩擦导致的阻力递增角度，pi 为髓芯。M_1 为去皮直径微钻起点，M_2 为去皮直径微钻终点。

图 1　微钻阻力曲线预处理过程

Fig. 1　Preprocessing of micro - drilling resistance profiles

再根据式(3)对木质部区域每个密度点的微钻阻力值进行校正：

$$R_i' = R_i - (L_i \times \tan\alpha) \tag{3}$$

式中：R_i'为校正后的第 i 个密度点微钻阻力值；R_i 为校正前的第 i 个密度点微钻阻力值；L_i 为第 i 个密度点到 M_1 的距离。

最后，计算平均阻力值。对校正后的木质部区域所有密度点的微钻阻力值取平均值，作为该样本树的最终平均微钻阻力值，用于后续建模分析。

1.4　异常值处理与数据集划分

为提升建模数据质量，采用学生化残差法[27]与马氏距离法[28]两种统计方法识别并剔除异常样本(异常样本指微钻阻力与基本密度之间存在显著偏差，可能由测量误差、样本异常等导致)。首先，以微钻阻力为自变量(x)、基本密度为因变量(y)，构建线性回归模型式(4)。

$$y = \beta_0 + \beta_1 x + \varepsilon \tag{4}$$

式中：β_0 为截距，β_1 为斜率，ε 为随机误差。

基于此模型，按照式(5)和式(6)计算每个样本的残差及其学生化残差，将$|r_i| > 1.96$(约 95% 置信水平)的样本视为异常值。

$$e_i = y_i - \bar{y}_i \tag{5}$$

式中：e_i 为第 i 个样本的残差；y_i 为实际密度值；$\bar{y}_i$ 为回归预测值。

$$r_i = \frac{e_i}{\delta\sqrt{1 - h_{ii}}} \tag{6}$$

式中：r_i 为学生化残差；e_i 为残差；δ为残差标准差；h_{ii} 为杠杆值。

同时，将每个样本视为二维向量 $x = (x, y)$，按式(7)计算其马氏距离 D^2。当 $D^2 > \chi^2_{0.95}(2) = 5.991$ 时，判定该样本为异常值；其中，$\chi^2_{0.95}(2)$为自由度 2、显著性水平 $\alpha = 0.05$ 下的卡方分布临界值。

$$D^2 = (x-\mu)(x-\mu)^{\mathrm{T}}\boldsymbol{\Sigma}^{-1}(x-\mu) \tag{7}$$

式中：D^2 为样本的马氏距离，x 为样本特征向量，μ 为均值向量，$\boldsymbol{\Sigma}$ 为协方差矩阵，上标 T 表示转置。

异常值检测与剔除过程在 Python 3.11 环境中实现，主要依赖 scikit-learn、SciPy 与 pandas 库。为确定最终用于数据清洗的方法，将比较两种方法剔除异常值后数据的分布情况，选择能使“微钻阻力—密度”关系更集中、合理的方案。

完成异常值剔除后，对保留的数据按 8∶2 的比例进行分层随机划分，其中测试集样本数按比例向

上取整确定，训练集为剩余样本；训练集用于模型构建与参数训练，测试集用于独立评估模型的泛化性能[29]。

1.5 预测模型构建

模型的构建与评估均在 Python 3.11 平台上完成，依托 scikit-learn（v1.2）与 lightgbm（v3.3.5）库实现。其中，区间概率模型以 LightGBM 为分类器，将密度区间作为分类目标进行训练；作为对比的 7 种传统回归模型（包括线性回归、二次多项式回归、三次多项式回归、决策树回归、神经网络回归、随机森林回归及支持向量回归）则统一在 scikit-learn 框架下构建。为保证结果可复现，对决策树、随机森林、神经网络及 LightGBM 等含随机过程的模型，随机种子统一设为 42；线性回归、多项式回归与支持向量回归本身不含随机过程，无需设置随机种子。回归模型采用 5 折交叉验证评估泛化能力，交叉验证按样本原始顺序划分，不额外引入随机性；区间分类模型采用以密度区间为分层变量的分层随机划分，随机种子设为 42。

1.5.1 区间概率模型构建

区间分类模型以微钻阻力为输入特征、以密度区间为输出类别，输出样本属于各密度区间的分类概率；该概率为分类条件概率，表征模型对密度等级归属的置信程度。建模时，以微钻阻力为输入变量（x），密度区间为输出变量（y），构建区间分类预测模型。构建高性能分类模型的关键前提之一是确立合理的密度区间划分方案。本研究中微钻阻力对应的木材基本密度分布并不均匀（图 2）。图 2 分别给出了基本密度与微钻阻力的单变量分布，可见二者均存在明显的集中区和稀疏区，但分布形态并不完全重合。这种分布不均与测量波动意味着等宽划分易导致类别不平衡，因此本研究采用分位数法确定区间边界。该方法使区间边界随数据密度自适应调整：在样本密集区域，相邻分位点间距较小，区间边界更细，有助于捕捉局部特征；在样本稀疏区域，分位点间距较大，区间相应加宽，从而保障各区间具有基本的样本基础[30]。具体而言，对于 k 个区间，取 0，1/k，2/k，…，1 共 k+1 个分位点作为初始边界，使各区间样本数大致均衡；若出现样本过少的区间，则与相邻区间合并，以保证分类模型有足够的训练样本。为确定最优划分，系统设计并比较了 3 ~ 6 个区间的划分方案。评价方案时，采用了多准则分析框架，从分类性能与工程适用性两个维度综合考量：在分类性能维度，以测试集准确率、宏平均 F_1 值为核心指标，衡量模型的判别精度与稳健性；在工程适用性维度，以平均区间宽度（反映分级的精细程度）和区间宽度变异系数（反映各区间划分的一致性）为核心指标，评估分级结果在实际应用中的分辨能力与可操作性。通过系统比较不同方案在上述指标上的表现，选取在分类性能与工程适用性上达到最佳平衡的划分方案，作为最终建模依据。

考虑到样本量有限，未采用 ResNet、DenseNet 等深度学习分类算法，而是选用 LightGBM（梯度提升决策树）算法构建区间概率模型。该算法具有训练效率高、能有效处理类别不平衡数据等优势，适合小样本分类任务[20]。

1.5.2 回归模型构建

以微钻阻力为自变量（x），基本密度为因变量（y），构建 7 种回归模型，包括线性回归、二次多项式回归、三次多项式回归、决策树回归、神经网络回归、随机森林回归及支持向量回归。

2 结果与分析

2.1 异常值的筛选与剔除

对学生化残差法与马氏距离法的异常值剔除效果进行对比（图 3）。学生化残差法能够更精准识别

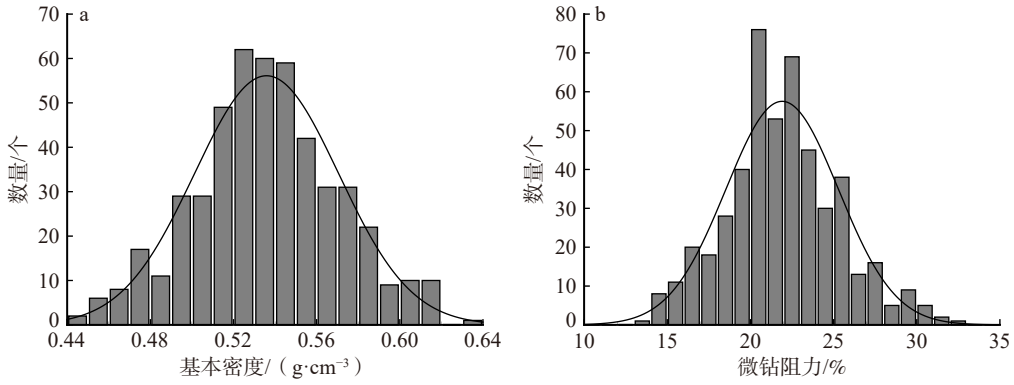


图 2 基本密度（a）与微钻阻力（b）数据分布

Fig. 2 Distribution of density and micro-drilling resistance data

注：两图分别为基于异常值剔除前的初始匹配数据的单变量分布，用于说明区间划分依据，不用于展示二者一一对应关系，n = 488。

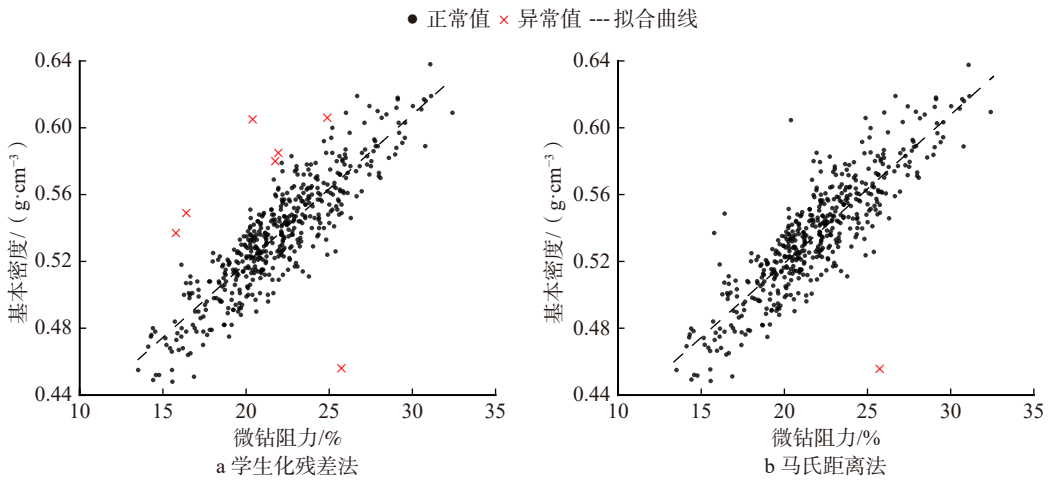


图 3　学生化残差和马氏距离剔除异常值结果

Fig. 3　Comparison of outlier removal results using Studentized Residuals and Mahalanobis Distance

微钻阻力与基本密度关联显著偏离整体趋势的样本，剔除的 7 组异常样本均分布在拟合曲线 95% 置信区间外，且剔除后剩余数据的离散程度明显降低；马氏距离法虽能剔除部分异常样本，但根据可视化散点图可以发现，剔除异常值效果较差。综合两种方法的处理效果，学生化残差法在识别与密度－阻力线性趋势显著偏离的异常点时更具针对性，因此本研究选用该方法进行最终的数据清洗，共保留 481 组有效数据（训练集 384 组，测试集 97 组）用于建模。

2.2　回归模型性能分析

基于构建的 7 种回归模型（图 4），其性能评估结果如表 1 所示，从预测精度、误差水平与泛化能力三个维度分析。

2.2.1　预测精度（测试集 R^2）

各回归模型的预测精度间差异明显。线性回归与二次多项式回归的 R^2 均为 0.771，并列最高，二次多项式并未带来 R^2 的提升，说明在本研究数据范围内，微钻阻力与木荷木材基本密度间以线性关联为主，非线性特征不明显或可忽略；随机森林回归（R^2 = 0.736）与决策树回归（R^2 = 0.708）具备一定预测能力，但精度低于前两者，说明树模型在捕捉钻阻－密度线性关联上无明显优势；三次多项式回归（R^2 = 0.762）精度略低于线性与二次多项式回归，推测因过高阶次导致模型过度拟合训练数据，丢失部分泛化能力；神经网络回归（R^2 = 0.182）与支持向量机回归（R^2 = −0.037）精度极低，其中支持向量机回归 R^2 为负，表明这两种模型在本研究小样本（384 组）且未深度调参的情况下，无法有效捕捉钻阻－密度的关联规律。

2.2.2　误差水平（MAE 与 RMSE）

性能较优的模型（R^2 > 0.7）误差水平集中在较低范围：线性回归与二次多项式回归的误差相近，MAE 分别为 0.014 5 和 0.014 4 g/cm³、RMSE 均为 0.018 0 g/cm³，证实两种模型具有较好的预测稳定性；随机森林回归与决策树回归 MAE（0.015 2 ~ 0.015 9 g/cm³）、RMSE（0.019 3 ~ 0.020 3 g/cm³）略高于前者，误差控制能力稍弱；三次多项式回归（MAE = 0.014 7 g/cm³、RMSE = 0.018 3 g/cm³）虽误差水平接近最优模型，但结合其 R^2 下降趋势，进一步说明过拟合对模型误差控制的潜在影响；神经网络回归（MAE = 0.028 1 g/cm³、RMSE = 0.034 0 g/cm³）与支持向量机回归（MAE = 0.030 5 g/cm³、RMSE = 0.038 3 g/cm³）误差显著偏高，其 MAE 值已明显高于本研究中性能较优的线性与多项式回归模型（MAE 约 0.014 g/cm³），亦难以满足木材分级实践对精度的要求。

2.2.3　泛化能力（交叉验证 R^2 均值）

线性回归（交叉验证 R^2 = 0.790）与二次多项式回归（交叉验证 R^2 = 0.784）的泛化能力最强，且与测试集 R^2 差值较小（分别约为 0.019、0.013），表明两种模型在不同数据子集上的性能稳定，受数据分布波动影响小；随机森林回归（交叉验证 R^2 = 0.756）与决策树回归（交叉验证 R^2 = 0.732）交叉验证 R^2 均值低于前两者，且与测试集 R^2 差值较大（分别约为 0.020、0.024），泛化能力较弱；三次多项式回归交叉验证 R^2 = −43.789（数据异常，推测因过拟合导致交叉验证时误差剧增），进一步证实其过拟合风险；神经网络回归（交叉验证 R^2 = −0.290）与支持向量机回归（交叉验证 R^2 = −0.059）交叉验证 R^2 为负，表明两种模型无法泛化到新数据，不具备实用价值。

2.3　区间概率模型性能分析

2.3.1　不同区间划分方案比较

对不同区间划分方案（3 ~ 6 个区间）综合分析（表 2）。区间数量 3 方案具有最高的分类准确率（0.907）和宏平均 F_1 值（0.908），但其平均区间宽度较大（0.063 g/cm³），分级粒度较粗，各等级覆盖范围

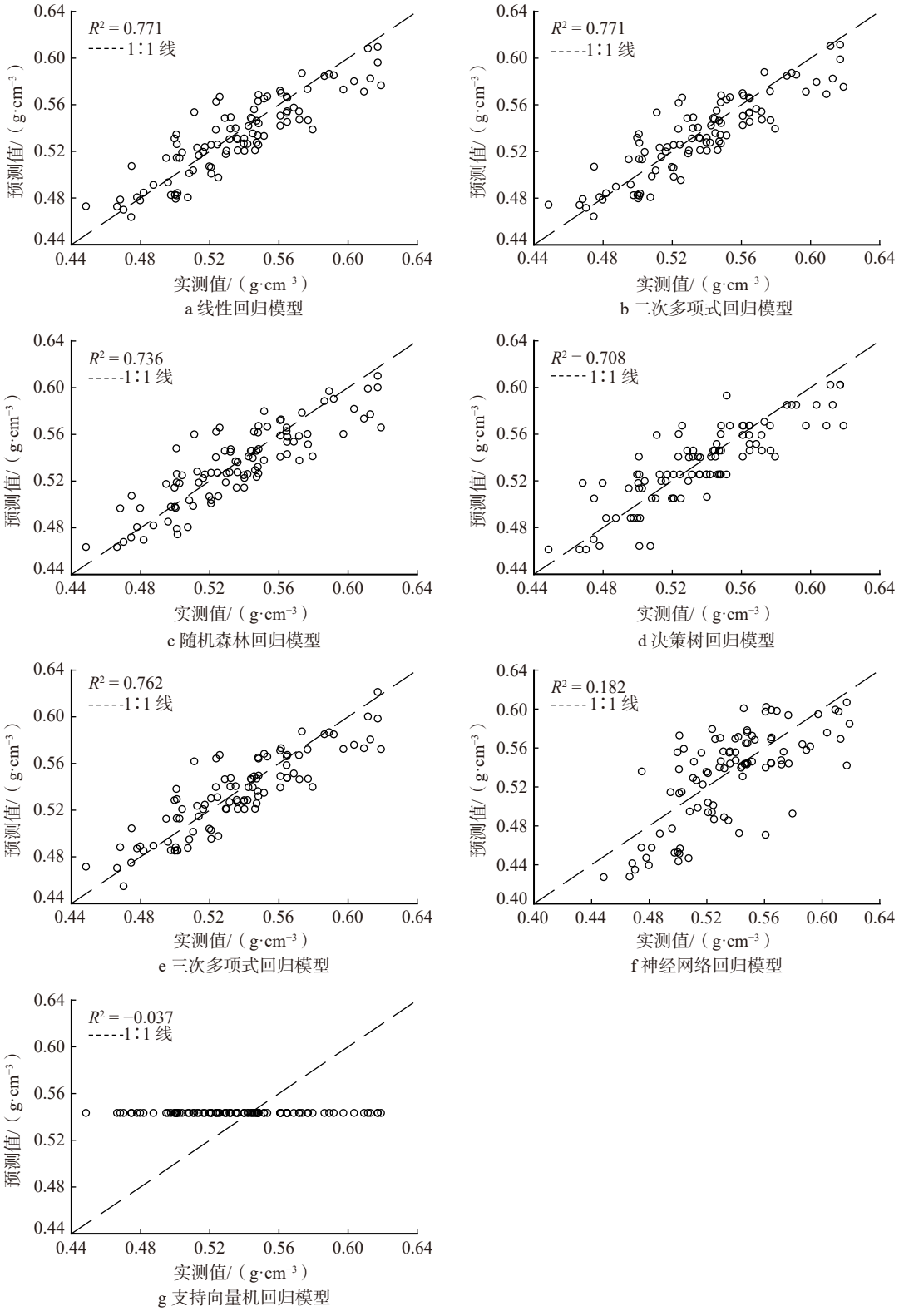


图 4 不同回归模型拟合结果

Fig. 4 Fitting results of different regression models

宽，在需要一定区分精度的工程场景中实用性受限。区间数量 4 方案的分类性能较 3 区间略有下降，准确率和宏平均 F_1 值分别为 0.876 和 0.878，但仍维持在较高水平；同时，其平均区间宽度缩减约 24%（至 0.048 g/cm^3），分级精细度明显提升。需要指出的是，区间数量 4 方案的区间宽度变异系数由区间数量 3 的 0.394 升至 0.502，表明各区间宽度的一致性有所下降，这是提高分级精细度所付出的代价。但该代价处于可接受范围：区间数量 4 方案在分类精度和区间一致性上仍明显优于区间数量 5 和 6。区间数量 5 和 6 虽进一步缩小平均区间宽度（分别至 0.038 和 0.032 g/cm^3），但准确率分别降至 0.825 和 0.701，

表 1　不同回归模型性能比较

Tab. 1　Performance comparison of different regression models

模型	测试集			交叉验证R^2均值
	R^2	MAE/(g·cm^{-3})	RMSE/(g·cm^{-3})	
线性回归	0.771	0.014 5	0.018 0	0.790
二次多项式回归	0.771	0.014 4	0.018 0	0.784
随机森林回归	0.736	0.015 2	0.019 3	0.756
决策树回归	0.708	0.015 9	0.020 3	0.732
三次多项式回归	0.762	0.014 7	0.018 3	−43.789
神经网络回归	0.182	0.028 1	0.034 0	−0.290
支持向量机回归	−0.037	0.030 5	0.038 3	−0.059

注: MAE为平均绝对误差, RMSE为均方根误差, 交叉验证R^2均值代表泛化能力。

表 2　不同密度区间划分方案在测试集上的性能比较

Tab. 2　Performance comparison of different density interval partition schemes

区间数量	分类性能		工程适用性	
	准确率	宏平均F_1值	平均区间宽度/(g·cm^{-3})	区间宽度变异系数
3	0.907	0.908	0.063	0.394
4	0.876	0.878	0.048	0.502
5	0.825	0.824	0.038	0.714
6	0.701	0.686	0.032	0.668

注: 准确率、宏平均F_1值为核心指标, 衡量模型的判别精度与稳健性; 平均区间宽度(反映分级的精细程度)和区间宽度变异系数(反映各区间划分的一致性)为核心指标, 评估分级结果在实际应用中的分辨能力与可操作性。区间宽度变异系数按“标准差/均值”计算, 标准差采用总体标准差口径。

宏平均 F_1 值分别降至 0.824 和 0.686, 且区间宽度变异系数分别为 0.714 和 0.668, 均高于区间数量 4, 说明划分过细并未形成更合理的区间结构, 反而可能因类别边界重叠和样本支撑不足而损害模型判别能力与结果可靠性。综上, 区间数量 3 方案侧重分类精度但分级较粗, 区间数量 5 及以上方案则因过细划分导致性能劣化; 区间数量 4 方案虽在区间宽度一致性上较区间数量 3 有所下降, 但在分类精度、分级精细度与工程可解释性之间取得了相对最优的折中, 故本研究最终采用区间数量 4 划分方案构建预测模型。为便于表述, 将区间 Ⅰ ~ Ⅳ依次简称为低、中低、中高、高密度区间, 密度区间(g/cm^3)划分为 Ⅰ [0.45, 0.51]、Ⅱ (0.51, 0.54]、Ⅲ (0.54, 0.56]、Ⅳ (0.56, 0.64]。

2.3.2　区间概率模型性能

为评估区间概率模型性能, 本研究从整体分类性能和各密度区间分类性能两方面, 综合分析了准确率、精确率、召回率与 F_1 值。

2.3.2.1　整体分类性能

如表 3 和图 5 所示, 区间概率模型在 97 个测试样本中正确分类 85 个, 整体准确率达到 0.876, 宏平均 F_1 值为 0.878, 表明模型具有优异的整体分类性能。各密度区间的 F_1 值均达到或高于 0.800, 其中高密度区间表现最佳(F_1 = 0.957), 中高密度区间次

表 3　区间概率模型性能比较

Tab. 3　Performance comparison of the interval probability model across density intervals

密度区间/(g·cm^{-3})	精确率	召回率	F_1值	测试集样本量
[0.45, 0.51]	0.875	0.840	0.857	25
(0.51, 0.54]	0.769	0.833	0.800	24
(0.54, 0.56]	0.880	0.917	0.898	24
(0.56, 0.64]	1.000	0.917	0.957	24
平均值	0.881	0.877	0.878	

注: 准确率反映模型整体分类正确的比例; 精确率衡量被预测为某类的样本中实际属于该类的比例; 召回率评估实际属于某类的样本中被正确预测的比例; F_1值为精确率与召回率的调和平均数, 综合反映模型在各类别上的判别能力。

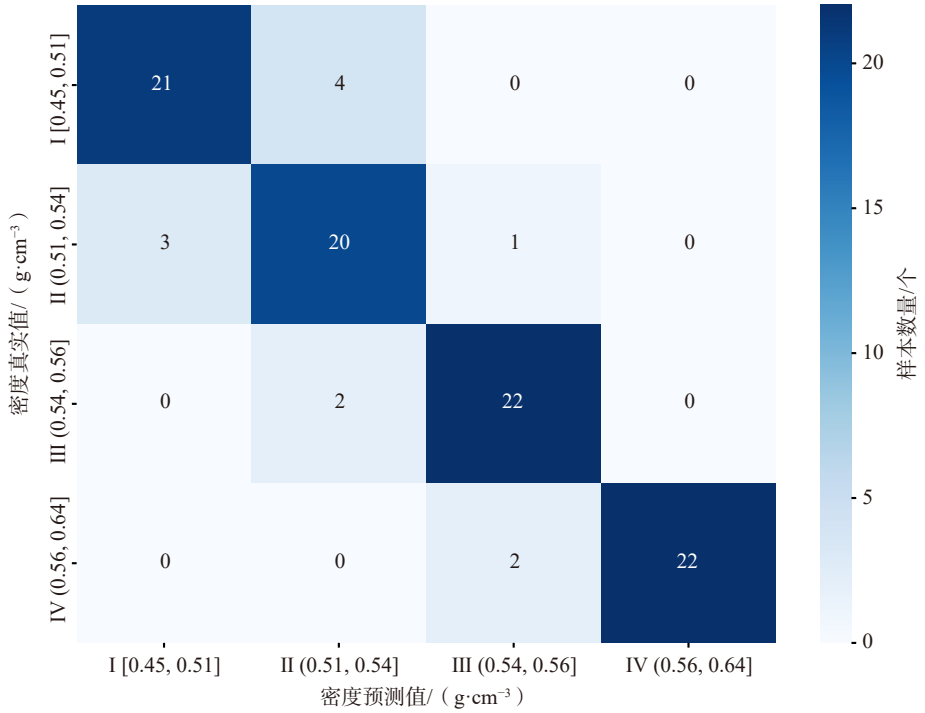


图 5 密度区间概率模型测试集混淆矩阵

Fig. 5 Confusion matrix for the density interval probability model on the test set

之（F_1 = 0.898），整体呈现“极端密度区间分类性能更优”的特征。各区间精确率平均值为 0.881，召回率平均值为 0.877，进一步说明模型在不同类别间均保持较高的识别稳定性。

2.3.2.2 各区间分类性能

结合混淆矩阵（图 5）对各密度区间的分类表现进行细部分析。低密度区间：精确率为 0.875，召回率为 0.840，25 个测试样本中有 21 个被正确分类，误判样本多集中于区间上边界（0.50 ~ 0.51 g/cm^3），说明该区域与相邻区间存在一定特征重叠。中低密度区间：作为过渡区间，因密度特征与相邻类别重叠，精确率略低（0.769），但召回率仍达 0.833，具备较好的实际样本覆盖能力。中高密度区间：分类性能优异，精确率为 0.880，召回率为 0.917，24 个样本中有 22 个被正确判别。高密度区间：分类效果最优，精确率达 1.000，召回率为 0.917，仅 2 个样本因微钻阻力轻微波动被误判，展现出优秀的判别能力。

综上，区间概率模型在各密度区间上均表现出良好的分类稳健性，尤其在高、低密度区间判别准确率高，能满足工程实践中对木材密度等级的快速、可靠判定需求。

2.4 两类模型的对比分析

从实用场景角度对比回归模型与区间概率模型，回归模型（以二次多项式回归为例）虽能提供精确的密度点估计值（如 0.523 g/cm^3），但需结合误差范围（MAE = 0.014 4 g/cm^3）才能判断可靠性，且受数据离散性影响显著，部分样本预测偏差较大，需结合误差范围解读。区间概率模型通过输入实测木材的微钻阻力值便能直接输出密度所属区间（例如，输入一个具有代表性的微钻阻力值时，模型可直接输出密度区间 0.51 ~ 0.54 g/cm^3），无需额外计算误差，其结果更贴合工程中“木材等级判断”的实际需求，且抗干扰能力强（即使钻阻数据轻微波动，仍能正确分类）。从性能稳定性看，回归模型中仅线性与二次多项式回归表现可靠，其他模型实用性差；区间概率模型各区间性能均衡，宏平均 F_1 值达 0.878，在分级稳定性上表现更均衡。这种稳定性使得该模型既能为工程决策提供可靠依据，也能为科学研究提供高质量、可解释的分类数据。

3 讨 论

本文的区间分类模型定位于木材分级决策，而非连续密度的精确估计。若研究目标需要连续密度值，或需解析微钻阻力–密度关系，传统点估计回归仍不可替代。分类指标与回归 R^2 含义不同，不能直接等同。

3.1 区间分类更契合木材分级决策

区间分类模型直接输出密度等级及分类概率，比连续点估计更契合木材分级决策的实际需求。工程实践中，木材分级往往只需要判断样本属于哪一密度等级，而不必先获得精确密度值再换算等级[31]；区间模型省去了“点估计 + 误差范围”的二次判断，

使现场快速分级更直接。同时，分类概率本身可作为预测置信度，便于使用者判断结果的可靠程度。这种概率化输出也可为遗传育种等研究提供等级化、概率化的补充信息，有助于描述木材密度等生物性状的自然变异[11−12]。

3.2　评价目标从拟合优度转向分级稳健性

相对于既有微钻阻力－密度回归研究，本研究的推进不在于提高 R^2，而在于将评价目标从拟合优度转向分级稳健性。既有研究多通过优化函数形式[10]或引入含水率[15−16]、仪器型号和操作者[17−18]等校正变量来提升拟合精度，但前者受数据分布影响大，R^2 可在 0.21 ~ 0.87 间波动[12−14]，后者又依赖野外难以实时获取的变量，并削弱跨设备通用性。本研究通过任务重构，不再以连续值拟合精度为核心目标，而是以分类准确率、宏平均 F_1 值等分类指标评价模型，使其更贴近木材分级决策。本研究中，表现最佳的回归模型 R^2 为 0.771，而区间模型整体准确率为 0.876，宏平均 F_1 值为 0.878。因此，上述分类指标的优势应理解为等级判别稳健性的提升，而非连续密度拟合精度的提高。

3.3　区间划分可兼顾分级精度与粒度

区间数量 4 方案并非天然最优，而是在木荷数据分布与工程分级需求共同约束下，分类精度与分级粒度之间的最佳折中。区间划分的核心是在类别可区分性与分级可操作性之间寻求平衡[30]。区间数量 3 方案准确率虽高，但分级过粗；区间数量 5 和 6 方案边界更细，却因特征重叠导致性能下降。区间数量 4 方案以损失约 3 个百分点分类精度的代价，将平均区间宽度缩减约 24%，同时维持较高的准确率和宏平均 F_1 值，并保证了高、低密度两端区间的判别能力。这表明，将连续密度预测转化为适度粒度的分类问题，是包容测量不确定性的有效途径。就应用而言，最优区间数依赖数据分布、样本量和具体分级标准；换树种或换应用场景时，应重新评估区间边界，不能直接套用木荷方案。

3.4　区间分类范式具备跨树种应用潜力

区间分类范式不仅在木荷上取得了稳定的分级性能，也为其他树种的密度预测提供了一种值得探索的方法路径。与依赖含水率[15−16]、仪器型号和操作者[17−18]等校正变量的传统模型不同，区间分类通过任务重构在输出端包容测量不确定性，降低了对含水率、仪器型号和操作者等校正变量的依赖，而后者正是传统回归模型跨树种推广困难的主要原因之一。因此，该范式具备跨树种应用的潜力。当然，其普适性仍需在更多树种、仪器和样本条件下进一步检验；未来可扩展多树种数据，在不同仪器与操作条件下检验该范式的稳健性，并进一步利用区间概率输出所包含的信心信息，开展密度等级的可靠性评估与群体比较分析。本文结论基于木荷单一树种数据，跨树种应用前仍需针对目标树种重新划分密度区间并开展独立验证。

4　结　　论

本研究针对微钻阻力法预测木材密度易受多源干扰的问题，将连续密度预测重构为密度区间分类任务，并以木荷为例构建了基于机器学习法（LightGBM）的区间分类模型。主要结论如下。

（1）将微钻阻力法木材密度预测由连续点估计重构为区间分类，可有效包容测量不确定性，提升模型在木材分级中的稳健性与实用性。以木荷为案例，基于 LightGBM 的区间分类模型在测试集上取得了 0.876 的整体准确率和 0.878 的宏平均 F_1 值，表明从“预测精确密度值”转向“输出稳定密度等级”具有可行性。

（2）模型在木荷木材的不同密度区间均表现出稳定的分类能力。其中，在木材分级实践中常需区分的高密度区间（0.56, 0.64] g/cm^3 与低密度区间 [0.45, 0.51] g/cm^3，分类正确率均达到或超过 84%，表明该模型能满足快速分级的基本需求。

（3）区间分类范式为解决微钻阻力法模型稳定性差的问题提供了新思路，其输出的密度等级直观、分类概率可表征置信程度，兼顾工程快速判断与科研数据需求。

本研究基于木荷单一树种数据，区间分类范式在其他树种中的适用性仍需进一步检验；未来可扩展多树种数据，并针对目标树种重新划分密度区间。

参　考　文　献

[1] Ramírez Correa J A, Marín Vélez A M, Torres-Andrade P, et al. Influence of tree diameter and height on the physical and mechanical properties of *Retrophyllum rospigliosii* wood [J]. Forests, 2025, 16(4): 593.

[2] Pinkowski G, Piernik M, Wołpiuk M, et al. Effect of chip thickness and tool wear on surface roughness and cutting power during up-milling wood of different density [J]. BioResources, 2024, 19(4): 9234−9248.

[3] Apungu J, Antwi K, Appiah-Kubi E, et al. Investigating physical and mechanical properties of mahogany root wood (*Khaya ivorensis*) for its utilization [J]. Drvna Industrija, 2025, 76(2): 189−199.

[4] Ali A C. Physical-mechanical properties and natural durability of lesser used wood species from Mozambique[D]. Uppsala: Swedish University of Agricultural Sciences, 2011.

[5] De Souza Cardoso F, José Vinha Zanuncio A, Deroco Martins G, et al. Innovative methodology for determining basic wood density

using multispectral images and MAPIR RGNIR camera [J]. IEEE Access, 2025, 13: 166581−166592.

[6] 孙燕良, 张厚江, 朱磊, 等. 木材密度检测方法研究现状与发展 [J]. 森林工程, 2011, 27(1): 23−26.
Sun Y L, Zhang H J, Zhu L, et al. Research and development of the method for testing wood density [J]. Forest Engineering, 2011, 27(1): 23−26.

[7] GB/T 1927.5—2021 无疵小试样木材物理力学性质试验方法 第 5 部分: 密度测定 [S].
GB/T 1927.5—2021 Test methods for physical and mechanical properties of small clear wood specimens: part 5: determination of density [S].

[8] Gaitan-Alvarez J, Moya R, Berrocal A. The use of X-ray densitometry to evaluate the wood density profile of *Tectona grandis* trees growing in fast-growth plantations [J]. Dendrochronologia, 2019, 55: 71−79.

[9] Charria D L G, Rodríguez V N, Oviedo M B S, et al. Estimación indirecta de densidad por el uso de “pilodyn” en *Eucalyptus pellita* f. Muell [J]. Colombia Forestal, 2014, 17(2): 181.

[10] 姚建峰, 郭旭展, 符利勇, 等. 微钻阻力法间接测量木材密度 [J]. 林业科学, 2022, 58(9): 138−147.
Yao J F, Guo X Z, Fu L Y, et al. Indirect measurement of wood density by micro drill resistance method [J]. Scientia Silvae Sinicae, 2022, 58(9): 138−147.

[11] Nickolas H, Williams D, Downes G, et al. Application of resistance drilling to genetic studies of growth, wood basic density and bark thickness in *Eucalyptus globulus* [J]. Australian Forestry, 2020, 83(3): 172−179.

[12] Isik F, Li B L. Rapid assessment of wood density of live trees using the Resistograph for selection in tree improvement programs [J]. Canadian Journal of Forest Research, 2003, 33(12): 2426−2435.

[13] Downes G M, Lausberg M, Potts B M, et al. Application of the IML Resistograph to the infield assessment of basic density in plantation eucalypts [J]. Australian Forestry, 2018, 81(3): 177−185.

[14] Cheng L T, Dai J, Yang Z G, et al. Variation of larch wood property indexes based on nondestructive testing data [J]. BioResources, 2020, 15(2): 2906−2923.

[15] Cheng L T, Dai J, Yang Z G, et al. Theoretical and experimental research on moisture content and wood property indexes based on nondestructive testing [J]. BioResources, 2020, 15(1): 1600−1616.

[16] Yao J F, Zhao Y B, Lu J, et al. Research on the wood density measurement in standing trees through the micro drilling resistance method [J]. Forests, 2024, 15(1): 175.

[17] Gendvilas V, Downes G M, Lausberg M, et al. Predicting wood density using resistance drilling: the effect of varying feed speed and RPM [J]. Forests, 2024, 15(4): 579.

[18] Gendvilas V, Lee D J, Kain D P, et al. Predicting wood density using resistance drilling: the effect of instrument and operator [J]. Forests, 2024, 15(1): 157.

[19] Kabir S, Wagner C, Ellerby Z. Toward handling uncertainty-at-source in AI: a review and next steps for interval regression [J]. IEEE Transactions on Artificial Intelligence, 2024, 5(1): 3−22.

[20] Ke G, Meng Q, Finley T, et al. LightGBM: A Highly Efficient Gradient Boosting Decision Tree[C]//Advances in Neural Information Processing Systems 30 (NIPS 2017). Curran Associates Inc., 2017: 3146−3154.

[21] 王云鹏, 张蕊, 周志春, 等. 10 年生木荷生长和材性性状家系变异及选择 [J]. 南京林业大学学报 (自然科学版), 2020, 44(5): 85−92.
Wang Y P, Zhang R, Zhou Z C, et al. A variation and selection of growth and wood traits for 10-year-old *Schima superba* [J]. Journal of Nanjing Forestry University (Natural Sciences Edition), 2020, 44(5): 85−92.

[22] 彭徐剑, 何晨, 詹庆斌. 3 种南方防火树种理化性质的试验研究 [J]. 林产工业, 2021, 58(3): 27−31.
Peng X J, He C, Zhan Q B. Experimental research on physical and chemical properties of three fire-prevention tree species in South China [J]. China Forest Products Industry, 2021, 58(3): 27−31.

[23] 韦昌幸. 48 年生木荷人工林生产力及经济效益评价研究 [D]. 长沙: 中南林业科技大学, 2014.
Wei C X. Productivity analysis and economic benefit evaluation of48a *Schima superba* plantation [D]. Changsha: Central South University of Forestry & Technology, 2014.

[24] 彭建, 鲁鑫钥, 余小晏, 等. 基于 Pilodyn 的楠木木材密度预测模型的建立 [J]. 西北林学院学报, 2023, 38(6): 269−275.
Peng J, Lu X Y, Yu X Y, et al. Establishment of wood density prediction model of *Phoebe zhennan* based on pilodyn data [J]. Journal of Northwest Forestry University, 2023, 38(6): 269−275.

[25] Fundova I, Funda T, Wu H X. Non-destructive wood density assessment of Scots pine (*Pinus sylvestris* L.) using Resistograph and Pilodyn [J]. PLoS One, 2018, 13(9): e0204518.

[26] Trachta M A, Acosta J J, Walker T D. Removing friction and bark improves drill resistance data for selection on wood density in *Pinus taeda* plantations [J]. Forest Products Journal, 2024, 75(1): 26−33.

[27] Kobayashi Y. New precise model of studentized principal components [J]. Communications in Statistics-Theory and Methods, 2024, 53(2): 487−504.

[28] Garthwaite P H, Elfadaly F G, Crawford J R. Modified confidence intervals for the Mahalanobis distance [J]. Statistics & Probability Letters, 2017, 127: 131−137.

[29] 周飞燕, 金林鹏, 董军. 卷积神经网络研究综述 [J]. 计算机学报, 2017, 40(6): 1229−1251.
Zhou F Y, Jin L P, Dong J. Review of convolutional neural network [J]. Chinese Journal of Computers, 2017, 40(6): 1229−1251.

[30] 谈询滔, 顾依依, 阮彤, 等. 数据集分类可用性评估的置信区间方法 [J]. 计算机科学, 2019, 46(1): 78−85.
Tan X T, Gu Y Y, Ruan T, et al. Confidence interval method for classification usability evaluation of data sets [J]. Computer Science, 2019, 46(1): 78−85.

[31] DS/EN 338: 2016 Structural timber-strength classes [S].

（责任编辑 吴 娟
责任编委 李文彬）