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

基于深度学习和上采样的杉木单木枯损模型构建

胡浩然 孙玉军 谢运鸿 梁瑞婷 王宝莹

High-precision individual tree mortality prediction for Chinese fir using deep learning and oversampling

Hu Haoran, Sun Yujun, Xie Yunhong, Liang Ruiting, Wang Baoying

引用本文:
胡浩然, 孙玉军, 谢运鸿, 梁瑞婷, 王宝莹. 基于深度学习和上采样的杉木单木枯损模型构建[J]. 北京林业大学学报, 2025, 47(12):39–47. doi: 10.12171/j.1000–1522.20250103
Hu Haoran, Sun Yujun, Xie Yunhong, Liang Ruiting, Wang Baoying. High–precision individual tree mortality prediction for Chinese fir using deep learning and oversampling[J]. Journal of Beijing Forestry University, 2025, 47(12):39–47. doi: 10.12171/j.1000–1522.20250103

在线阅读 View online: https://doi.org/10.12171/j.1000–1522.20250103

您可能感兴趣的其他文章

Articles you may be interested in

面向钻蛀振动实时侦听的深度学习模型压缩

Deep learning model compression for real–time listening of boring vibration

北京林业大学学报, 2021, 43(6): 92–100. https://doi.org/10.12171/j.1000–1522.20200100

基于Mask R–CNN和迁移学习的无人机遥感影像杉木单木树冠提取

UAV remote sensing image extraction of single tree crown of Chinese fir based on Mask R–CNN and transfer learning

北京林业大学学报, 2024, 46(3): 153–166. https://doi.org/10.12171/j.1000–1522.20210343

基于非线性分位数混合效应构建杉木树高曲线模型

Construction of *Cunninghamia lanceolata* tree height curve model based on nonlinear quantile mixed effect

北京林业大学学报, 2023, 45(11): 33–41. https://doi.org/10.12171/j.1000–1522.20220496

基于级联循环网络的林木生长参数预测

Prediction of tree growth parameters based on cascaded recurrent network

北京林业大学学报, 2023, 45(8): 94–108. https://doi.org/10.12171/j.1000–1522.20230027

气候敏感的杉木树高胸径非线性混合效应模型研建

Development of climate–sensitive nonlinear mixed–effects tree height–DBH model for *Cunninghamia lanceolata*

北京林业大学学报, 2023, 45(9): 52–61. https://doi.org/10.12171/j.1000–1522.20230052

基于Sparse–DenseNet模型的森林火灾识别研究

Forest fire identification based on Sparse–DenseNet model

北京林业大学学报, 2020, 42(10): 36–44. https://doi.org/10.12171/j.1000–1522.20190371

DOI:10.12171/j.1000−1522.20250103

基于深度学习和上采样的杉木单木枯损模型构建

胡浩然 孙玉军 谢运鸿 梁瑞婷 王宝莹
（北京林业大学森林资源环境管理国家林业局重点实验室，北京 100083）

摘要:【目的】杉木是我国福建省重要用材树种，针对福建将乐国有林场杉木单木枯损预测精度不高的问题，构建基于深度学习的诊断模型，为杉木科学抚育提供参考。**【方法】**基于 2011—2024 年 79 块固定样地数据，采用 SMOTE、MAHAKIL 等上采样技术及其混合策略处理类间不平衡问题，建立深度学习单木枯损模型。采用 F1 分数、赤池信息量准则（AIC）和贝叶斯信息量准则（BIC）等多维度指标评价模型性能，并进一步与传统 logistic 回归模型进行对比，评估模型性能与稳定性。**【结果】**（1）混合采样策略显著提升模型性能，其中 MAHAKIL 混合上采样模型的 AIC 均值（874）和 BIC 均值（949）整体优于 SMOTE 混合上采样模型（AIC=1 038，BIC=1 117）。最优方案为 MAHAKIL-R300 混合采样，该模型 F1 分数达 0.959，临界阈值确定为 0.646。深度学习单木枯损模型性能显著优于传统 logistic 回归模型。（2）在 8 个影响因子中，对杉木单木枯损影响贡献最大的前 3 个因子依次为胸径、胸径基尼系数、大于对象木胸径的其他林木的胸高断面积之和（C1）。其中，胸径基尼系数、C1 与枯损概率呈正相关，而胸径与枯损概率呈负相关，表明林木竞争和结构不均衡显著增加了单木枯损风险。**【结论】**本研究构建的基于上采样技术与深度学习的杉木单木枯损模型，具有更高的预测精度和稳定性，能有效识别枯损关键影响因子。该模型可为杉木林精准抚育和科学经营提供可靠的决策支持工具。

关键词：杉木；枯损模型；深度学习；上采样；诊断模型；精准抚育
中图分类号：S791.27 **文献标志码：**A **文章编号：**1000−1522(2025)12−0039−09
引文格式：胡浩然，孙玉军，谢运鸿，等. 基于深度学习和上采样的杉木单木枯损模型构建［J］. 北京林业大学学报，2025，47(12)：39−47. Hu Haoran, Sun Yujun, Xie Yunhong, et al. High-precision individual tree mortality prediction for Chinese fir using deep learning and oversampling［J］. Journal of Beijing Forestry University, 2025, 47(12): 39−47.

High-precision individual tree mortality prediction for Chinese fir using deep learning and oversampling

Hu Haoran Sun Yujun Xie Yunhong Liang Ruiting Wang Baoying
(State Forestry & Grassland Administration Key Laboratory of Forest Resources & Environmental Management, Beijing Forestry University, Beijing 100083, China)

Abstract:［**Objective**］Chinese fir (*Cunninghamia lanceolata*) is a key timber species in Fujian Province, China. To address the issue of low prediction accuracy for individual tree mortality at the Jiangle State-owned Forest Farm, this study constructs a deep learning-based diagnostic model, providing a reference for the scientific tending of Chinese fir.［**Method**］Using data from 79 permanent sample plots from 2011 to 2024, SMOTE, MAHAKIL, and hybrid oversampling strategies were employed to address class imbalance issues. A deep learning-based individual tree mortality model was established. Model performance was evaluated using multidimensional metrics such as F1 score, AIC, and BIC, and further compared with traditional logistic regression models to assess performance and stability.［**Result**］The hybrid sampling strategies significantly improved model performance. The MAHAKIL hybrid oversampling model achieved

收稿日期: 2025−03−24 **修回日期:** 2025−11−27
基金项目: 国家自然科学基金项目（31870620），林业科学技术推广项目（[2019]06）。
第一作者: 胡浩然。主要研究方向：林业模型模拟。Email：905091277@qq.com 地址：100083 北京市海淀区清华东路 35 号北京林业大学林学院。
责任作者: 孙玉军，博士生导师，教授。主要研究方向：林业模型模拟。Email：sunyj@bjfu.edu.cn 地址：同上。
本刊网址: http://j.bjfu.edu.cn；http://journal.bjfu.edu.cn

mean AIC (874) and BIC (949) that were superior to those of the SMOTE hybrid oversampling model (AIC=1 038, BIC=1 117). The optimal strategy was the MAHAKIL-R300 hybrid sampling, with an F1-score of 0.959 and a critical threshold set at 0.646. The performance of the deep learning individual tree mortality model was significantly better than that of the traditional logistic regression model. (2) Among the eight influencing factors, the top three factors contributing the most to individual tree mortality of Chinese fir were, in order, diameter at breast height (DBH), Gini coefficient of DBH, and the total basal area of trees with DBH greater than that of the subject tree (C1). The Gini coefficient of DBH and C1 were positively correlated with mortality probability, while DBH was negatively correlated with mortality probability, indicating that tree competition and structural imbalance significantly increased the risk of individual tree mortality. [**Conclusion**] The individual tree mortality model for Chinese fir constructed in this study, based on oversampling techniques and deep learning, demonstrated higher prediction accuracy and stability, effectively identifying key influencing factors of mortality. This model can provide a reliable decision support tool for the precision nurturing and scientific management of Chinese fir forests.

Key words: *Cunninghamia lanceolata*; mortality model; deep learning; oversampling; diagnostic model; precision nurturing

陆地生态系统包含了森林、荒漠、草原等多种类型，森林作为其中覆盖面积最大且物种最丰富的生态系统[1]，在维持全球生态平衡、调节气候变化、保护生物多样性中扮演着不可替代的角色。杉木（*Cunninghamia lanceolata*）作为我国特有的速生用材树种，在我国南方地区广泛种植，具有生长快、材质优良、单位面积蓄积量高、病虫害相对较少等显著优势，但同时也存在喜光、不耐寒、不耐旱、浅根系等生物学特性。已有研究[2]表明，杉木人工林蓄积量最高可达 300 m^3/hm^2，在保障我国木材供给安全、维持区域生态平衡、促进天然林保护工程实施以及助力碳中和目标实现等方面发挥着至关重要的战略作用。

在森林生长发育与群落演替过程中，林木枯损与自然更新共同驱动着森林动态变化[3]。林木枯损作为森林生态系统演替的重要环节，不仅影响林分结构调整、养分循环、能量流动和物种共存格局，还深刻影响森林的生态功能与服务价值[4]。从生态学视角看，林木枯损是指在林木自身生理生态特性与外界环境胁迫因素共同作用下，林木个体生命力逐渐衰退直至完全消失的复杂生态过程。其发生机制涉及多层次、多尺度的因素交互，既包括树木自身的遗传特性、生理状态和竞争地位,也受到立地条件、气候变化、生物干扰等外源性因素的影响[5]。在以往的森林生态学研究中，影响林木枯损的因子通常可分为两大类：一是密度依赖性因素，主要包括树木个体大小（胸径、树高、冠幅等）、种内竞争与种间竞争强度、林分密度等内源性因子[6]；二是密度无关性因素，主要涵盖气候因子（温度、降水、极端气候事件）、立地条件（土壤类型、坡向坡位、海拔）、光照资源可获得性以及土壤养分含量等外源性环境因子[7−8]。这两类因素往往相互作用、协同影响，共同决定着单木的生存概率与林分动态。

准确预测单木枯损是深入理解森林动态、优化森林经营决策的关键科学问题。相较于林分枯损模型，单木枯损模型因其能够捕捉个体间的差异性和竞争关系的复杂性，在预测精度和生态解释力方面具有显著优势，因而被广泛应用于森林动态模拟、森林抚育规划和可持续经营管理中。传统的单木枯损建模方法主要基于传统统计学理论和经验公式，通过构建以胸径、树高、冠幅等基本生长变量为自变量的回归方程，利用 logistic 回归、Weibull 生存分析等参数化方法构建模型，并通过设定临界概率阈值判断单木存活状态[9−10]。这类传统方法虽然在统计理论上较为成熟，但在处理复杂非线性关系和高维交互作用时存在明显局限。近年来，随着计算能力的提升和算法理论的发展，机器学习与深度学习技术因其强大的非线性映射能力和特征自动提取能力，为枯损建模提供了全新的技术范式。大量研究[11]表明，深度学习模型在枯损预测精度上显著优于传统回归方法。例如，美国加利福尼亚地区运用随机森林算法预测干旱诱发的树木死亡率[12]，机器学习方法也有效评估了澳大利亚森林中干旱导致的树木死亡风险[13]。然而，针对福建将乐国有林场所处的亚热带季风气候区这一特定环境，夏季高温多雨且台风频发，杉木因其浅根系特性易发生风倒和枯损，但当前针对该地区杉木单木枯损模型的构建及其驱动因子的系统性研究仍相对缺乏，这在一定程度上制约了该地区杉木人工林的精准经营与可持续管理。

在实际森林样地调查过程中，枯损木数量相较于活立木在样本总量中的占比通常较低，这导致模

型训练过程中容易出现类间不平衡(class imbalance)问题。在二分类数据集中,若不平衡比(imbalance ratio, IR,定义为多数类样本数与少数类样本数的比值)超过 10∶1,传统分类算法往往会在预测过程中过度倾向于样本数量较多的类别[14],进而导致模型对少数类(即枯损木)的识别能力严重不足,出现高误判率和低召回率问题。这一现象不仅降低了模型的实际应用价值,也可能导致错误的经营决策。尽管类间不平衡问题在林业领域的关注度相对较低,但在医疗诊断[15]、金融风险评估[16]等领域已得到广泛研究和证实。为有效解决类间不平衡问题,研究者们提出了多种技术路径,其中上采样(oversampling)技术通过在数据预处理阶段扩充少数类样本平衡类别分布,被证明能够显著提升模型对少数类的识别精度[17]。然而,不同上采样技术在森林枯损预测中的适用性及其与深度学习算法的协同效应尚需进一步探索和验证。

综上所述,本研究以福建将乐国有林场杉木人工林为研究对象,综合考虑区域气候特征、林分结构特点和数据特征,拟采用多种上采样技术(SMOTE、MAHAKIL 等)对类间不平衡样本进行系统性预处理,并结合深度学习方法(基于 TensorFlow 框架的全连接神经网络)构建杉木单木枯损预测模型。研究旨在:(1)评估不同上采样策略对模型性能的提升效果,筛选最优采样方案;(2)构建高精度的深度学习枯损预测模型,并与传统 logistic 回归模型进行对比;(3)识别影响杉木单木枯损的关键驱动因子,揭示多因子交互作用机制。研究将为福建地区杉木人工林的精准抚育、科学经营和可持续管理提供理论依据和技术支撑,同时为其他区域类似树种的枯损预测研究提供方法学参考。

1　研究区概况

福建将乐国有林场位于福建省西北部将乐县内(26°25′ ~ 27°04′N, 117°05′ ~ 117°40′E)。属山地地貌,地形复杂,海拔 200 ~ 700 m[18]。该区域为典型的海洋性季风气候,年均温 18.9 ℃,年降水量 1 665 mm,无霜期 289 d[19]。土壤以红壤为主,腐殖质层厚,养分含量较高,主要乔木树种有杉木、马尾松(*Pinus massoniana*)、木荷(*Schima superba*)等。

2　研究数据与研究方法

2.1　数据来源

2011—2024 年,在福建将乐国有林场设置 79 块杉木林样地,包含幼龄林、中龄林、近熟林、成熟林及过熟林 5 个龄组。样地中杉木为优势树种,其他树种还包括马尾松、木荷等。每木检尺获取胸径、树高数据,标记枯损木,并计算株数密度、胸径基尼系数等林分结构指标。

2.2　研究方法

2.2.1　竞争因子计算

除单木尺度因子外,样地中林木竞争因子同样影响林木枯损。本文根据何潇等[20]开发的 R 语言 ForestStatTool 工具包,通过 CI.nondist 函数计算 8 个基于单木胸径和树高的距离无关竞争指数,进而评估其对枯损的贡献。样地基本信息见表 1。

$$C1 = \sum_{i=1}^{n} \pi \left(\frac{D_i}{2} \right)^2 \tag{1}$$

$$C2 = \frac{\sum_{i=1}^{m} D_i}{D_{\text{target}}} \tag{2}$$

表 1　样地基本信息统计

Tab. 1　Summary statistics of sample plots

变量	平均值	最大值	最小值	标准差
胸径(DBH)/cm	16.6	53.5	2.9	6.71
树高(H)/m	13.9	34.1	2.2	5.41
株数密度(N)/(株·hm^{-2})	1 692	4 050	525	755
胸径基尼系数(GINI)	0.158	0.304	0.070	0.05
大于对象木胸径的其他林木的胸高断面积之和(C1)	27.97	70.58	0.00	15.77
其他林木胸径总和与对象木胸径的比(C2)	106.91	604.96	13.31	69.13
对象木胸径与样地林木平均胸径(均方直径)的比(C3)	0.95	3.13	0.16	0.29
对象木胸径与样地中林木的最大胸径的比(C4)	0.57	1.00	0.08	0.20
对象木的胸高断面积与平均胸高断面积的比(C5)	1.00	9.85	0.02	0.60
对象木的胸高断面积与最大胸高断面积的比(C6)	0.370	1.000	0.006	0.23
其他林木胸高断面积总和与对象木胸高断面积的比(C7)	158.68	3 908.96	0.47	220.82
大于对象木胸径的林木胸高断面积和与样地单位面积胸高断面积以及样地树高空间指数的比(C8)	0.23	0.62	0.00	0.14

$$\mathrm{C3}=\frac{D_{\mathrm{target}}}{\sqrt{\frac{\sum_{i=1}^{z}D_i^2}{z}}} \tag{3}$$

$$\mathrm{C4}=\frac{D_{\mathrm{target}}}{D_{\max}} \tag{4}$$

$$\mathrm{C5}=\frac{z\pi\left(\frac{D_{\mathrm{target}}}{2}\right)^2}{\sum_{i=1}^{z}\pi\left(\frac{D_i}{2}\right)^2} \tag{5}$$

$$\mathrm{C6}=\frac{\pi\left(\frac{D_{\mathrm{target}}}{2}\right)^2}{\pi\left(\frac{D_{\max}}{2}\right)^2} \tag{6}$$

$$\mathrm{C7}=\frac{\pi\left(\frac{D_{\mathrm{target}}}{2}\right)^2}{\sum_{i=1}^{m}\pi\left(\frac{D_i}{2}\right)^2} \tag{7}$$

$$\mathrm{C8}=\frac{S\sum_{i=1}^{n}\pi\left(\frac{D_i}{2}\right)^2}{\sum_{i=1}^{z}\pi\left(\frac{D_i}{2}\right)^2\times\frac{zH_{\mathrm{target}}}{\sum_{i=1}^{z}H_i}} \tag{8}$$

式中：C1~C8 含义见表 1。D_{target}为对象木胸径，$D_{\max}$样地中胸径最大树木的胸径，H_{target}为对象木树高，D_i为第i株树木的胸径，H_i为第 i 株树木的树高，z 为样地中林木数量，m为其他胸径的林木数量，n为大于对象木胸径的林木数量，S 为样地面积。

2.2.2 模型构建方法

本研究采用 TensorFlow 深度学习框架构建深度学习模型，该框架由谷歌团队开发，是一个广泛应用于深度学习的开源框架。目前 TensorFlow 技术已广泛应用于病例识别、自然语言处理等领域[21]。PyTorch 深度学习框架作为目前最主流的两个深度学习框架之一，起源于 2002 年开发的 Torch 框架，基于 Python 对操作进行简化。而 TensorFlow 相较 PyTorch 的主要优势在于其更丰富的模型选择以及优异的深度学习可视化程度[22]。TensorFlow 包含的技术主要分为全连接神经网络、卷积神经网络和循环神经网络。其中全连接神经网络主要用于处理二分类问题，也适用本研究中对于单木枯损模型的建立。

在深度学习中，模型构建的隐藏层是神经网络的核心组成部分，负责对输入数据进行非线性变换和特征抽象。隐藏层主要作用包括：(1)非线性建模，隐藏层通过叠加非线性激活函数，使网络能够拟合复杂的非线性关系；(2)决策边界复杂化，通过多层非线性变换，隐藏层能将输入空间映射到高维空间，使原本线性不可分的数据变得可分。假设第 l 层为隐藏层，数据输入表达公式为

$$\boldsymbol{h}^{(l)}=\sigma\left(\boldsymbol{W}^{(l)}\boldsymbol{h}^{(l-1)}+\boldsymbol{b}^{(l)}\right) \tag{9}$$

式中：$\boldsymbol{h}^{(l)}$为第l层神经元的输出向量；$\boldsymbol{W}^{(l)}$为第l层的权重矩阵；$\boldsymbol{h}^{(l-1)}$为第$l-1$层的输入向量；$\boldsymbol{b}^{(l)}$为第l层的偏置向量；σ为隐藏层中的激活函数。

激活函数作为隐藏层中传递神经元信息的桥梁，将输入信息在神经网络中从线性拟合转化为非线性拟合。本研究采用 ReLU 激活函数。

$$f(x)=\max(0,x)=\begin{cases}x, & \text{if } x\geqslant 0\\ 0, & \text{if } x<0\end{cases} \tag{10}$$

式中：$f(x)$为 ReLU 激活函数。

ReLU 函数虽为分段函数，但整体呈现非线性。在正向传播的过程中，ReLU 激活函数响应正值输入，负值输出均为 0 值，部分网络中的神经元被“抑制”，形成稀疏激活，同时抑制负值输入可减少参数依赖性，提升模型泛化能力，增强神经网络模型的鲁棒性。

本研究采用二分类 logistic 回归作为传统基准模型[23]。该模型能够有效地分析多个影响因素与单木枯损之间的关系，是构建单木尺度枯损概率预测模型的常用方法之一。

$$P=\frac{1}{1+\mathrm{e}^{(\beta_0+\beta_1X_1+\cdots+\beta_pX_p)}} \tag{11}$$

式中：P为单木枯损概率，$X_1,\cdots,X_p$为预测变量，β_0为截距，$\beta_1,\cdots,\beta_p$为各自变量对应的回归系数。

通过对数函数将非线性的概率模型转化为线性回归形式。

$$\ln(P/(1-P))=\beta_0+\beta_1X_1+\cdots+\beta_pX_p \tag{12}$$

2.2.3 混合采样方法

以二分类模型为例，当数据集中多数类和少数类样本数量比例不超过 4∶1 时，被认为是合理的建模比例[24]。针对样本集中枯损木与活立木严重不平衡的问题，本研究评估 4 种过采样方法。

(1)随机过采样(ROS)：随机从少数类中反复地抽取样本，将抽取到的样本放入原样本空间以组成新样本空间，从而扩充少类样本个数[25]。该抽样方法虽简单便捷，但会使少数类中存在大量相似数据，容易导致过拟合而被排除。

(2)SMOTE 算法：通过线性插值合成新样本，可缓解随机上采样的过拟合风险[26]。

$$X_{\mathrm{new}}=a+r(a_i-a) \tag{13}$$

式中：X_{new}表示新样本，r 是介于 0~1 的随机数，a_i为 k 个最近邻中随机选择的一个样本，a 为从少数类样本中随机选择的一个子样本。

（3）MAHAKIL 算法：基于马氏距离将少数类样本分为两个分区，选择来自同一分区的两个父样本求取平均值生成新的合成样本。不断持续子样本与亲代样本交叉结合，直至满足需要合成的样本数[27]。

$$\boldsymbol{\mu} = \frac{1}{N}\sum_{i=1}^{N} X_i \tag{14}$$

$$\boldsymbol{S} = \frac{1}{N-1}\sum_{i=1}^{N} (X_i - \boldsymbol{\mu})(X_i - \boldsymbol{\mu})^{\mathrm{T}} \tag{15}$$

$$X'_{\mathrm{new}} = X_i + \lambda \boldsymbol{d}_i \tag{16}$$

$$\boldsymbol{d}_i = \boldsymbol{S}^{-\frac{1}{2}} (X_i - \boldsymbol{\mu}) \tag{17}$$

式中：X_i 为少数类样本中第 i 个子样本，$\boldsymbol{\mu}$ 表示样本集的均值向量，$\boldsymbol{S}$ 表示样本集的协方差矩阵，λ表示随机数，$\boldsymbol{d}_i$为基于马氏距离的方向向量。X'_{new}表示基于 MAHAKIL 算法的新样本，N 为少数类样本总数。

（4）混合采样：整合 SMOTE 与 MAHAKIL 的优势，按不同比例混合以优化数据分布。研究[28]表明混合策略普遍优于单一方法。本研究将比较混合采样与单类上采样方式的预测结果，探究不同比例混合采样方式对模型拟合结果的影响。

2.2.4　模型评价方法

鉴于样本不平衡特性，本研究以 F1 分数为核心评价指标，其综合了精确率与召回率，能有效衡量模型对枯损木（少数类）的识别性能[29]。同时，引入赤池信息量准则（Akaike information criterion，AIC）和贝叶斯信息量准则（Bayesian information criterion，BIC）评估模型复杂度与拟合优度，其值越小，表明模型越优且过拟合风险越低。

3　结果与分析

3.1　模型拟合结果

由于原始数据集中存在类间不平衡问题，预测结果偏向数据集中占比较大的活立木。因此，采用单类上采样模型的精确率较高，但 F1 分数较低（表 2）。相较于原始数据，采用 SMOTE 上采样和 MAHAKIL 上采样虽可显著提高模型的召回率与 F1 分数，但 F1 分数均不到 0.5，不具有解释意义。

对比表 2，混合采样策略（表 3 ~ 6）使模型性能显著提升。在深度学习的 32 个模型中，基于 MAHAKIL 混合上采样模型的 AIC 均值为 874，BIC 均值为 949，整体优于 SMOTE 混合上采样模型（AIC = 1 038，BIC = 1 117），表明 MAHAKIL 更能抑制噪声样本生成。随着 ROS 采样比例的逐步上升，模型的

表 2　单类上采样模型拟合结果

Tab. 2　Fitting results of individual class oversampling model

采样方式	精确率	召回率	F1分数
原始数据	0.50	0.03	0.05
SMOTE上采样	0.18	0.31	0.23
MAHAKIL上采样	0.19	0.43	0.27

表 3　深度学习结合 ROS-SMOTE 混合上采样模型拟合结果

Tab. 3　Fitting results of deep learning combined with ROS-SMOTE hybrid oversampling

采样方式	精确率	召回率	F1分数	AIC	BIC
R200	0.976	0.920	0.947	921	1 000
R300	0.974	0.922	0.947	783	861
R400	0.958	0.930	0.944	1 021	1 099
R500	0.964	0.926	0.944	1 096	1 174
R600	0.968	0.926	0.946	1 092	1 170
R700	0.972	0.922	0.946	973	1 051
R800	0.957	0.928	0.942	1 135	1 213
R900	0.966	0.926	0.945	1 062	1 140
R1000	0.978	0.926	0.951	1 061	1 139
R1100	0.976	0.924	0.949	1 067	1 145
R1200	0.966	0.926	0.945	1 170	1 249
R1300	0.976	0.920	0.947	1 000	1 078
R1400	0.976	0.922	0.948	1 042	1 121
R1500	0.974	0.924	0.948	1 087	1 166
R1600	0.966	0.926	0.945	1 057	1 135
R1700	0.970	0.926	0.947	1 055	1 133

注：R200为第1步采用ROS上采样至200棵枯立木，第2步采用SMOTE上采样至2 000棵枯立木（满足活立木∶枯立木至少达到4∶1的建模要求），其他类推。下同。

表 4　深度学习结合 ROS-MAHAKIL 混合上采样模型拟合结果

Tab. 4　Fitting results of deep learning combined with ROS-MAHAKIL hybrid oversampling

采样方式	精确率	召回率	F1分数	AIC	BIC
R200	0.968	0.936	0.952	459	533
R300	1.000	0.922	0.959	428	502
R400	0.966	0.936	0.951	664	739
R500	0.978	0.930	0.953	769	844
R600	0.980	0.928	0.953	809	884
R700	0.982	0.926	0.951	886	962
R800	0.978	0.926	0.952	995	1 071
R900	0.982	0.924	0.951	1 041	1 117
R1000	0.980	0.924	0.950	1 126	1 203
R1100	0.978	0.924	0.948	1 191	1 268
R1200	0.974	0.924	0.957	1 274	1 351
R1300	0.993	0.924	0.958	791	864
R1400	0.996	0.924	0.956	821	895
R1500	0.991	0.924	0.956	861	935
R1600	0.991	0.924	0.954	918	992
R1700	0.987	0.924	0.954	958	1 033

表 5 logistic 模型结合 ROS-SMOTE 混合上采样模型拟合结果

Tab. 5 Fitting results of the logistic model combined with ROS-SMOTE hybrid oversampling

采样方式	精确率	召回率	F1分数	AIC	BIC
R200	0.621	0.743	0.676	4 423	4 463
R300	0.640	0.778	0.702	4 432	4 471
R400	0.693	0.795	0.740	4 322	4 362
R500	0.666	0.810	0.731	4 546	4 586
R600	0.655	0.775	0.710	4 330	4 376
R700	0.636	0.703	0.668	4 533	4 573
R800	0.655	0.805	0.722	4 381	4 421
R900	0.643	0.770	0.701	4 586	4 625
R1000	0.651	0.786	0.713	4 397	4 436
R1100	0.651	0.793	0.715	4 308	4 355
R1200	0.637	0.755	0.691	4 402	4 441
R1300	0.658	0.740	0.697	4 505	4 545
R1400	0.650	0.745	0.694	4 423	4 463
R1500	0.653	0.768	0.706	4 440	4 480
R1600	0.643	0.798	0.712	4 464	4 504
R1700	0.642	0.738	0.686	4 487	4 527

表 6 logistic 模型结合 ROS-MAHAKIL 混合上采样模型拟合结果

Tab. 6 Fitting results of the logistic model combined with ROS-MAHAKIL hybrid oversampling

采样方式	精确率	召回率	F1分数	AIC	BIC
R200	0.628	0.770	0.692	4 229	4 275
R300	0.630	0.766	0.691	4 222	4 262
R400	0.603	0.785	0.682	4 361	4 401
R500	0.631	0.798	0.705	4 235	4 275
R600	0.643	0.786	0.708	4 239	4 278
R700	0.632	0.788	0.701	4 252	4 298
R800	0.634	0.811	0.712	4 200	4 240
R900	0.632	0.808	0.709	4 229	4 268
R1000	0.628	0.788	0.699	4 138	4 178
R1100	0.635	0.798	0.707	4 189	4 229
R1200	0.620	0.753	0.680	4 256	4 303
R1300	0.632	0.810	0.710	4 230	4 270
R1400	0.627	0.773	0.693	4 225	4 265
R1500	0.639	0.75	0.690	4 204	4 250
R1600	0.654	0.783	0.713	4 228	4 274
R1700	0.629	0.743	0.682	4 177	4 223

AIC 值和 BIC 值会有一定的增加，说明 ROS 上采样增加了模型的过拟合性。

而对比传统 logistic 单木枯损模型拟合结果，深度学习模型在精确率、召回率、F1 分数、AIC 值与 BIC 值 5 项指标方面均有较高的提升。这一研究结果也进一步显示，在通过混合采样解决原始样本集类间不平衡问题后，深度学习模型相较传统 logistic 模型在处理大样本数据集时展现出更高的精度优势。

综合来看，最优方案为 MAHAKIL-R300 混合采样，其 AIC = 428，BIC = 502，F1 = 0.959，确定为最终预测模型。在 64 个训练模型中，临界阈值多大于 0.5（图 1），符合单木枯损预测中阈值 > 0.5 的典型特征[30]。其中，最优预测模型的临界阈值为 0.646。

3.2 模型因子特征重要性结果

以 MAHAKIL-R300 模型训练后各因子特征重要性排序（基于第一层权重），结合回归系数的正负性，探究各因子对单木枯损概率的影响[31]。在建模前，通过多重共线性检验剔除方差膨胀因子（variance inflation factor，VIF）值大于 10 的因子。建模后模型各参数见表 7。在 8 个影响因子中，胸径、胸径基尼系数与竞争因子 C1 为前 3 关键因子（图 2），其中胸径基尼系数和 C1 贡献值为正，胸径为负。

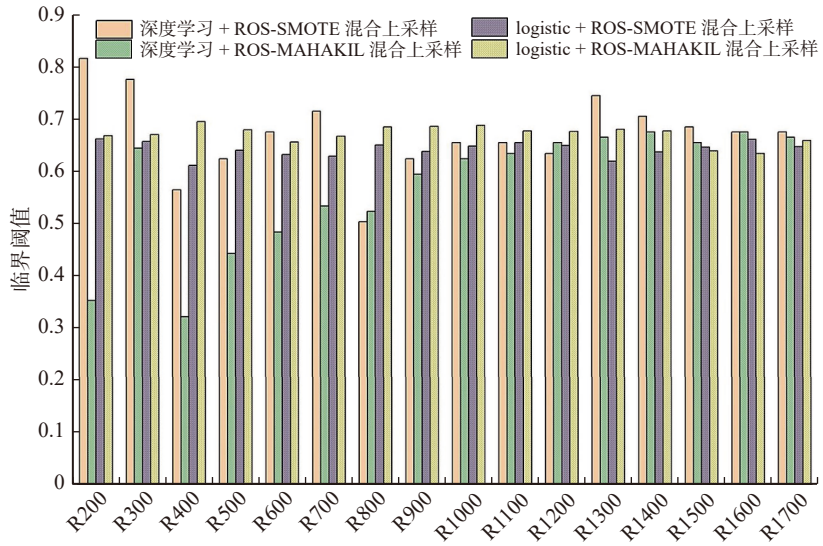


图 1 单木枯损模型临界阈值

Fig. 1 Critical threshold of individual tree mortality model

表 7　单木枯损模型拟合结果

Tab. 7　Fitting results of individual tree mortality model

变量	回归系数	标准误	*P*	置信区间	
				2.50%	97.50%
DBH	−1.241 5	0.200	0	−1.633	−0.850
H	−0.262 9	0.077	0.001	−0.414	−0.112
GINI（DBH）	1.485 6	0.096	0	1.297	1.674
N	−0.788 7	0.069	0	−0.924	−0.653
C1	1.456 4	0.100	0	1.261	1.652
C3	0.322 2	0.228	0.157	−0.124	0.769
C4	0.629 7	0.191	0.001	0.255	1.004
C8	−0.477 2	0.069	0	−0.613	−0.342

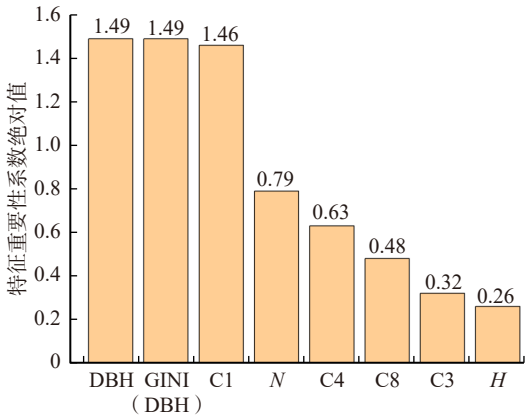


图 2　单木枯损模型特征重要性系数

Fig. 2　Characteristic importance coefficient of individual tree mortality model

小径级林木在林分中抵抗外界干扰能力较弱，且在光照、水分、营养物质的竞争中处于劣势，容易发生枯损现象，这可能是胸径与枯损概率负相关的主要原因。而胸径基尼系数与枯损概率正相关，这是因为胸径基尼系数反映林木胸径异质性，优势木虽占据主林层但面临更高风倒与病原体侵染风险，劣势木则因长期处于光限制与营养胁迫状态而易致枯损，整体上增加了群落的枯损概率。而作为邻体竞争强度的量化指标，C1 值越高，表明对象木受到来自更大径级林木的水平资源拦截越严重，生存压力增大，进而导致枯损概率升高。

4　讨　　论

单木枯损模型相较于林分尺度枯损模型，因其能够在个体水平上捕捉林木间的异质性、竞争关系和微环境差异，在预测精度和生态解释力方面具有内在优势[32]。本研究基于福建将乐国有林场 13 年动态监测数据，构建了杉木单木枯损预测模型，系统评估了不同建模方法的性能表现。已有研究[33]表明，传统统计学方法构建的单木枯损模型 F1 分数通常在 0.7~0.9，本研究的 logistic 回归模型结果进一步印证了这一结论。在 32 个混合采样方案下，logistic 模型的 F1 分数集中在 0.668 ~ 0.740，表现出一定的稳定性。而深度学习模型的拟合效果显著优于传统模型，最优模型的 F1 分数高达 0.959，较传统模型提升了约 30%。这一结果充分证实了深度学习模型在捕捉单木枯损这一复杂生态过程中的非线性响应、高阶交互作用和潜在特征方面具有强大的学习能力[34]。此外，本研究还发现 MAHAKIL 上采样技术在处理类间不平衡问题时优于传统 SMOTE 方法，这可能归因于 MAHAKIL 通过考虑样本间的马氏距离和多样性约束，能够更有效地抑制噪声样本生成，从而提高模型的泛化能力。

在林木枯损的二分类预测模型中，临界阈值的确定是影响模型实际应用效果的关键环节。传统二分类模型通常将 0.5 作为默认临界阈值，这隐含地假设两类事件的发生概率相等。然而，林木枯损作为一种小概率事件，其发生频率远低于林木存活，若简单地将临界阈值设定为 0.5，则等同于将枯损与存活事件的先验概率视为相等，这显然与森林生态系统的实际情况相悖，也不符合贝叶斯决策理论的基本原则。因此，针对类间不平衡数据集，需要根据具体的样本分布和业务需求设定大于 0.5 的临界阈值，以在精确率与召回率之间取得最优平衡。本研究基于 TensorFlow 框架中的全连接神经网络技术，通过优化分类错误率最小化目标函数，系统地确定了不同模型的最优临界概率。结果表明，64 个训练模型中绝大部分模型的临界阈值大于 0.5，这与以往针对小概率事件预测的结论[30]高度一致。最优模型（MAHAKIL-R300 混合采样方案）的临界阈值为 0.646，这一数值既考虑了样本的类别分布特征，也综合权衡了误判成本与漏判风险。从森林经营实践角度看，过低的阈值会导致大量假阳性（将健康木误判为枯损木），造成不必要的抚育成本和资源浪费；而过高的阈值则可能导致假阴性（将枯损风险木漏判为健康木），延误抚育时机，降低木材利用价值。因此，本研究确定的临界阈值为福建杉木人工林经营提供了科学的决策依据。

本研究通过深度学习模型的特征重要性评估，发现胸径、胸径基尼系数与竞争指数 C1 是影响福建将乐林场杉木单木枯损的 3 个关键因子，这一结果具有明确的生态学意义和管理启示。首先，胸径负向效应符合经典的竞争胁迫理论。在资源有限的森林环境中，小径级林木在光、水、养分竞争中处于劣势，长期生理胁迫导致抗性下降，易发生枯损，此效应在高密度人工林中尤为突出[35−36]。其次，胸径基尼系数的正向贡献揭示了林分结构异质性的“双重

风险”机制：在结构高度异质的林分中，优势木因冠层暴露度高、受风面积大，更易成为台风、病虫害等外部干扰的首要目标[37]；劣势木则在长期的资源竞争抑制下持续承受光照不足、根系竞争激烈等多重生理胁迫，导致抗逆性减弱。这种“两端受压”的格局使得林分整体枯损风险上升。第三，竞争指数 C1 的正向作用进一步证实了邻体竞争在枯损过程中的关键驱动作用。当对象木周围存在大量大径级优势木时，对象木可获取的光照、水分和养分资源被大幅压缩，限制其生长与碳储备，在遭遇环境胁迫时更易触发枯损。这 3 个因子的协同识别为精准抚育提供了明确的靶点：在森林经营实践中，应重点关注小径级受压木、高胸径基尼系数林分中的两端木（优势木和劣势木），以及竞争压力大的林木，通过适时间伐、择伐等抚育措施，调整林分结构，降低枯损风险，提高森林质量和木材产量。

本研究在模型精度与方法上取得进展，但仍存在三重局限。其一，环境因子覆盖不全。气候动态（温湿度、极端事件）、立地质量（土壤养分、地形）及病虫害历史等关键变量未能系统纳入，可能限制机制的完整解析。未来需融合遥感、气象与土壤调查数据，构建多维度环境矩阵以深化因果识别。其二，空间代表性不足。模型仅基于单一林场数据，跨区域适用性未经验证。不同气候区、立地类型及经营历史的杉木林枯损模式差异显著，未来应拓展多气候带、多海拔梯度、多经营强度的样地网络，发展区域或国家尺度普适性模型。其三，机制可解释性受限。深度学习“黑箱”特性削弱了结果对经营决策的理论支撑。未来应耦合可解释性机器学习与生态过程模型，构建“精度−机理”双优的混合预测系统，实现从“预测”到“阐释”的范式跃升，以更好地服务于森林经营决策和生态学理论发展。

5 结　　论

本研究构建了福建将乐国有林场杉木单木枯损混合采样深度学习模型。其中，深度学习结合 ROS-MAHAKIL 混合上采样模型精度最佳，F1 分数可达 0.959。对单木枯损识别概率贡献排名前 3 的因子依次为胸径、胸径基尼系数与大于对象木胸径的其他林木的胸高断面积之和（C1）。针对生存概率偏低的小径级受压木，应优先实施生长伐以释放资源；对高胸径基尼系数林分中的优势木与劣势木，需通过目标树经营与结构调控降低两极分化风险；对竞争指数 C1 高的林木，则应适时疏伐邻体以缓解胁迫。未来需拓展多区域验证并耦合可解释性算法，以更好地服务于森林经营决策。

参 考 文 献

[1] 贺鹏, 和敬渊, 张海军, 等. 华北落叶松人工林保留木生长与林分生产力对间伐的响应 [J]. 北京林业大学学报, 2025, 47(11): 48−57.
He P, He J Y, Zhang H J, et al. Responses of residual tree growth and stand productivity to thinning in *Larix gmelinii* var. *principis-rupprechtii* plantations[J]. Journal of Beijing Forestry University, 2025, 47(11): 48−57.

[2] 陈庆辉. 福建含笑与杉木人工林的生产力与水源涵养能力比较 [J]. 亚热带农业研究, 2025, 21(1): 20−26.
Chen Q H. Comparison of productivity and water conservation capacity between *Michelia fujianensis* and *Cunninghamia lanceolata* plantations[J]. Subtropical Agriculture Research, 2025, 21(1): 20−26.

[3] Runkle J R. Canopy tree turnover in old-growth mesic forests of eastern North America[J]. Ecology, 2000, 81(2): 554−567.

[4] Zhou X, Chen Q, Sharma R P, et al. A climate sensitive mixed effects diameter class mortality model for Prince Rupprecht larch (*Larix gmelinii* var. *principis-rupprechtii*) in northern China[J]. Forest Ecology and Management, 2021, 491: 119091.

[5] 张雄清, 王翰琛, 鲁乐乐, 等. 杉木单木枯损率与初植密度、竞争和气候因子的关系 [J]. 林业科学, 2019, 55(3): 72−78.
Zhang X Q, Wang H C, Lu L L, et al. Tree mortality in relation to planting density, competition and climate factors for Chinese fir plantation in Southern China[J]. Scientia Silvae Sinicae, 2019, 55(3): 72−78.

[6] 窦啸文, 汤孟平. 基于引力模型的林木竞争分析 [J]. 应用生态学报, 2022, 33(10): 2695−2704.
Dou X W, Tang M P. Gravitational model-based competitive analysis of trees[J]. Chinese Journal of Applied Ecology, 2022, 33(10): 2695−2704.

[7] 常晓敏, 余新晓, 贾国栋, 等. 基于随机森林的不同径级树木死亡影响因子研究 [J]. 生态学报, 2019, 39(21): 8003−8009.
Chang X M, Yu X X, Jia G D, et al. Influence factors of tree mortality in different size classes based on random forest[J]. Acta Ecologica Sinica, 2019, 39(21): 8003−8009.

[8] Lu M, Du H, Song T, et al. Drivers of tree survival in an evergreen-deciduous broadleaf karst forest in southwest China[J]. Forest Ecology and Management, 2021, 499: 119598.

[9] 闫明, 陈艳梅, 闫静, 等. 基于广义线性混合效应模型的森林树木死亡研究 [J]. 生态学报, 2024, 44(6): 2420−2436.
Yan M, Chen Y M, Yan J, et al. Studying forest tree mortality based on a generalized linear mixed-effects model[J]. Acta Ecologica Sinica, 2024, 44(6): 2420−2436.

[10] Lutz J A, Halpern C B. Tree mortality during early forest development: a long-term study of rates, causes, and consequences[J]. Ecological Monographs, 2006, 76(2): 257−275.

[11] Cho N, Agossou C, Kim E, et al. Machine-learning modeling on tree mortality and growth reduction of temperate forests with climatic and ecophysiological parameters[J]. Ecological Modelling, 2023, 483: 110456.

[12] Madakumbura G D, Goulden M L, Hall A, et al. Recent California tree mortality portends future increase in drought-driven forest die-off[J]. Environmental Research Letters, 2020, 15(12): 124040.

[13] Carnegie A J, Lawson S, Mason J, et al. Forest health and biosecurity in a changing world[J]. Australian Forestry, 2022, 85(4): 141−145.

[14] Jain S, Kotsampasakou E, Ecker G F. Comparing the performance of meta-classifiers: a case study on selected imbalanced data sets relevant for prediction of liver toxicity[J]. Journal of Computer-aided Molecular Design, 2018, 32(5): 583−590.

[15] Huang C, Huang X, Fang Y, et al. Sample imbalance disease classification model based on association rule feature selection[J]. Pattern Recognition Letters, 2020, 133: 280−286.

[16] Marinakos G, Daskalaki S. Imbalanced customer classification for bank direct marketing[J]. Journal of Marketing Analytics, 2017, 5(1): 14−30.

[17] Bennin K E, Keung J, Phannachitta P, et al. MAHAKIL: diversity based oversampling approach to alleviate the class imbalance issue in software defect prediction[J]. IEEE Transactions on Software Engineering, 2018, 44(6): 534−550.

[18] 孙雅楠, 刘亚静, 孙钊, 等. 将乐国有林场杉木径向生长及其对气候的响应 [J]. 北京林业大学学报, 2024, 46(2): 18−27.
Sun Y N, Liu Y J, Sun Z, et al. Radial growth of *Cunninghamia lanceolata* and its response to climate in Jiangle National Forest Farm, Fujian Province of Eastern China[J]. Journal of Beijing Forestry University, 2024, 46(2): 18−27.

[19] 吴明钦. 福建杉木人工林树冠结构研究 [D]. 北京: 北京林业大学, 2014.
Wu M Q. A study on crown structure of *Cunninghamia lanceolata* plantation in Fujian[D]. Beijing: Beijing Forestry University, 2014.

[20] 何潇, 雷相东, 周超凡, 等. R 程序包 ForestStatTool 在林学单木和林分因子统计中的应用 [J]. 北京林业大学学报, 2024, 46(10): 125−135.
He X, Lei X D, Zhou C F, et al. Application of R program package ForestStatTool in statistics of individual tree and stand factors in forestry[J]. Journal of Beijing Forestry University, 2024, 46(10): 125−135.

[21] Rimal K, Shah K B, Jha A K. Advanced multi-class deep learning convolution neural network approach for insect pest classification using TensorFlow[J]. International Journal of Environmental Science and Technology, 2023, 20(4): 4003−4016.

[22] 黄玉萍, 梁炜萱, 肖祖环. 基于 TensorFlow 和 PyTorch 的深度学习框架对比分析 [J]. 现代信息科技, 2020, 4(4): 80−82,87.
Huang Y P, Liang W X, Xiao Z H. Comparative analysis of deep learning frameworks based on TensorFlow and PyTorch[J]. Modern Information Technology, 2020, 4(4): 80−82,87.

[23] 马武. 蒙古栎林单木生长模型系研究 [D]. 北京: 中国林业科学研究院, 2012.
Ma W. IStudy on single tree growth model system of *Quercus mongolica* forest[D]. Beijing: Chinese Academy of Forestry, 2012.

[24] Maalouf M, Trafalis T B. Rare events and imbalanced datasets: an overview[J]. International Journal of Data Mining, Modelling and Management, 2011, 3(4): 375−388.

[25] Estabrooks A, Jo T, Japkowicz N. A multiple resampling method for learning from imbalanced data sets[J]. Computational Intelligence, 2004, 20(1): 18−36.

[26] Chawla N V, Bowyer K W, Hall L O, et al. SMOTE: synthetic minority over-sampling technique[J]. Journal of Artificial Intelligence Research, 2002, 16: 321−357.

[27] Yao L, Lin T B. Evolutionary mahalanobis distance-based oversampling for multi-class imbalanced data classification[J]. Sensors, 2021, 21(19): 6616.

[28] Yang Y, Khorshidi H A, Aickelin U. A review on over-sampling techniques in classification of multi-class imbalanced datasets: Insights for medical problems[J]. Frontiers in Digital Health, 2024, 6: 1430245.

[29] Diallo R, Edalo C, Awe O O. Machine learning evaluation of imbalanced health data: a comparative analysis of balanced accuracy, MCC, and F1 score[M]// Awe O O, Vance E A. Practical statistical learning and data science methods. Cham: Springer Nature Switzerland, 2024: 283−312.

[30] 王涛. 大兴安岭东部地区天然落叶松和白桦基础单木模型研究 [D]. 哈尔滨: 东北林业大学, 2022.
Wang T. Basic individual tree model research for natural larch and birch in the eastern of Daxing'an Mountains[D]. Harbin: Northeast Forestry University, 2022.

[31] 徐宝涵. 基于驾驶人视角的车辆运行侧向风险量化方法研究 [D]. 南京: 南京林业大学, 2025.
Xu B H. Research on the quantification method of vehicle lateral operation risk based on the driver's perspective[D]. Nanjing: Nanjing Forestry University, 2025.

[32] 李春明, 赵丽芳, 李利学. 基于混合效应模型和零膨胀模型方法的蒙古栎林分水平枯损模型 [J]. 林业科学, 2019, 55(11): 27−36.
Li C M, Zhao L F, Li L X. Modeling stand-level mortality of Mongolian oak (*Quercus mongolica*) based on mixed effect model and zero-inflated model methods[J]. Scientia Silvae Sinicae, 2019, 55(11): 27−36.

[33] 姚丹丹, 徐奇刚, 闫晓旺, 等. 基于贝叶斯方法的蒙古栎林单木枯死模型 [J]. 北京林业大学学报, 2019, 41(9): 1−8.
Yao D D, Xu Q G, Yan X W, et al. Individual-tree mortality model of Mongolian oak forests based on Bayesian method[J]. Journal of Beijing Forestry University, 2019, 41(9): 1−8.

[34] Yang Z, Duan G, Sharma R P, et al. Predicting individual tree mortality of *Larix gmelinii* var. *principis-rupprechtii* in temperate forests using machine learning methods[J]. Forests, 2024, 15(2): 374.

[35] 李春明. 基于广义线性混合效应模型的蒙古栎林单木枯损建模及影响因子分析 [J]. 林业科学研究, 2020, 33(6): 105−113.
Li C M. Modeling of individual tree mortality and analysis of influence factor in Mongolian oak stand based on generalized linear mixed effect model[J]. Forest Research, 2020, 33(6): 105−113.

[36] 郭嘉. 秦岭天然松栎混交林树高胸径关系及竞争因子分析 [D]. 杨凌: 西北农林科技大学, 2019.
Guo J. Analysis on the relationship between height and diameter at breast height and competitive factors of natural pine-oak mixed forest in Qinling Mountains[D]. Yangling: Northwest A&F University, 2019.

[37] 周泽宇, 冯林艳, 闫星蓉, 等. 华北落叶松人工林林分枯损株数随机效应预测模型 [J]. 林业科学, 2022, 58(10): 67−78.
Zhou Z Y, Feng L Y, Yan X R, et al. Prediction model of stand mortality of *Larix principis-rupprechtii* plantation in the core area of Winter Olympic Games[J]. Scientia Silvae Sinicae, 2022, 58(10): 67−78.

（责任编辑　孟　瑶
责任编委　张会儒）