

DOI:10.7522/j.issn.1000-0240.2022.0006

WEN Hong, WU Xiyong, ZHAO Siyuan, et al. Snow avalanche susceptibility evaluation in the central Shaluli Mountains of Tibetan Plateau based on machine learning method[J]. Journal of Glaciology and Geocryology, 2022, 44(6):1694–1706. [文洪, 巫锡勇, 赵思远, 等. 基于机器学习法的青藏高原沙鲁里山系中段雪崩易发性评价研究[J]. 冰川冻土, 2022, 44(6):1694–1706.]

基于机器学习法的青藏高原沙鲁里山系中段雪崩易发性评价研究

文 洪^{1,2}, 巫锡勇², 赵思远³, 边 瑞², 周桂宇¹, 孟少伟⁴, 孙春卫²

(1. 宜宾学院 智能制造学部, 四川 宜宾 644007; 2. 西南交通大学 地球科学与环境工程学院, 四川 成都 611756; 3. 四川大学 水利水电学院 水力学与山区河流开发保护国家重点实验室, 四川 成都 610065; 4. 中铁二院工程集团有限责任公司, 四川 成都 610031)

摘 要: 青藏高原广泛发育、暴发频繁的雪崩对既有交通廊道造成严重威胁。采用高预测精度的机器学习算法对该类区域雪崩易发性进行评价,可快速、有效地对雪崩风险进行区域性评估。以青藏高原沙鲁里山系中段山区雪崩为研究对象,通过室内解译与现场验证相结合的方式识别并建立雪崩编目数据库,同时采用GIS、遥感等定量化提取技术,通过方差膨胀因子(VIF)筛选出14个评价因子,在此基础上利用支持向量机(SVM)、决策树(DT)、多层感知器(MLP)、K最邻近法(KNN)共4种机器学习模型对雪崩易发性进行评价并完成指数图的绘制,并采用Kappa系数和ROC曲线进行准确性检验。评价结果显示,SVM、DT、MLP、KNN的易发性指数分别在[0,0.964]、[0,0.815]、[0,0.995]、[0,1]范围内。精度检验结果显示这4种模型均具有较好或很好的预测精度,其中SVM模型的Kappa系数和AUC值均为最高,其AUC值高达0.912。结果表明研究区内雪崩易发性高的区域主要分布在夷平面以上的格聂山、日拱山等地,极高易发区平均海拔约4 939 m,高易发区平均海拔约4 859 m。该区域雪崩对川藏公路和在建的川藏铁路影响较小。该研究可为横穿沙鲁里山系的川藏铁路等重大工程建设的雪崩防灾减灾工作提供科学依据和方法借鉴。

关键词: 雪崩; 易发性评价; 机器学习; 沙鲁里山系; 青藏高原

中图分类号: P954; X4 **文献标志码:** A **文章编号:** 1000-0240(2022)06-1694-13

0 引言

雪崩是指多雪山区积雪在重力驱动下快速向下崩落的现象^[1],由其引发的直接灾害或链生灾害对人类生命、建筑环境、交通、生态系统构成严重威胁^[2]。中生代中期以来,青藏高原经历了多次强烈的隆升和夷平作用,不仅为雪崩发育创造了有利的地形条件,也改变了周围的大气环流形势和气候纬向地带性分布^[3],为雪崩发育提供了有利的气象条件。自1960年以来,青藏高原以全球两倍的升温速率持续变暖,预计在本世纪末“亚洲水塔”气温将激增4℃^[4]。持续的气候变暖一方面使得极端降水事

件更加频繁、降水强度增大,另一方面使得冰雪融化速率加剧,导致雪崩事件显著增加^[5]。目前青藏高原周缘形成了喜马拉雅南坡雪崩区、藏东南雪崩区、川西滇西北雪崩区等^[6],愈发频繁的雪崩灾害引起了学界和公众越来越多的关注^[7-9]。快速、有效地获得雪崩发育位置与雪崩易发区域,是后续进行雪崩针对性防灾减灾的基础性工作。因此,探索构建一套科学合理、有效易行的雪崩易发性评价体系,是应对雪崩灾害风险增加的必然需求,对青藏高原的城乡规划以及川藏铁路等重大工程建设均具有十分重要的意义。

收稿日期: 2021-05-16; 修订日期: 2021-10-18

基金项目: 第二次青藏高原综合科学考察研究项目(2019QZKK0905);宜宾学院计算物理四川省高校重点实验室开放课题基金资助项目(412-2020JSWLYB001);宜宾学院科研培育项目(412-2020PY09)资助

作者简介: 文洪,博士研究生,主要从事雪崩时空演化规律及控灾机理研究. E-mail: geowenhong@qq.com

通信作者: 巫锡勇,教授,主要从事特殊岩土及地质灾害方面的科研与教学工作. E-mail: wuxiyong@126.com

瑞士等多雪国家根据雪崩频率和冲击力建立雪崩风险分区和制图标准^[10],并在此基础上根据雪崩的风险程度对山区进行土地利用规划与管理。该标准已广泛应用于俄罗斯、加拿大、美国等欧美雪崩多发的国家^[11-13],并将雪崩风险降低到可接受水平。这种雪崩风险区划和制图标准依赖于长期观测数据的积累。然而,青藏高原极端恶劣的工作条件为详细的野外雪崩监测造成了极大的制约和巨大的成本,无法有效支撑青藏高原雪崩灾害的大面积风险区划工作,因此需要一套能在区域上快速对青藏高原进行风险识别和区划的方法,使重大雪崩灾害点的监测布置以及灾害防治更具针对性。近年随着人工智能的快速发展,机器学习算法已被许多研究者应用于地震预测、地下水储量变化预测、降水数据订正、滑坡易发性制图等领域^[14-18]。机器学习算法在灾害易发性评价方面的引入,弥补了传统二元统计方法工作量大、主观性强、预测结果精度低等缺点,为灾害预测和预防提供了重要的理论方法^[19-20]。而对于雪崩观测记录档案短缺的地区,基于遥感解译和野外调查所获得的学习样本数据库,采用机器学习算法开展雪崩易发性评价和制图,可为区域性的灾害风险预估提供重要参考。已有部分研究尝试将机器学习算法应用于区域雪崩的遥感自动检测^[21]、雪崩搬运物质易发性评价^[22]、雪崩易发性制图^[23-25],但目前对青藏高原极端地形条件和气象条件控制下的雪崩易发性认识仍有不足,机器学习算法在青藏高原雪崩易发性评价的适用性,还有待深入研究。

本文通过遥感解译和野外调查验证,识别了青藏高原沙鲁里山系中段山区 536 处雪崩,构建了研究区的雪崩空间数据库,在此基础上根据雪崩发育对地形地貌、气候气象、积雪特性等各因素的响应特征,采用 GIS、遥感等定量化提取技术并通过方差膨胀因子(VIF)筛选出 14 个控制雪崩时空分异发育的评价因子,采用支持向量机(SVM)、决策树(DT)、多层感知器(MLP)、K 最邻近法(KNN)共 4 种机器学习方法获取雪崩易发性指数图,验证机器学习算法在青藏高原这类雪崩记录档案资料短缺的多雪山区的适用性,同时为当地雪崩减灾防灾指出重点设防区域。

1 研究区概况

1.1 自然地理环境概况

沙鲁里山系位于四川省西部,属青藏高原东部横断山区北端中部山脉,呈南北走向[图 1(a)],海拔多在 4 000 m 以上,为无数纵横交错的峡谷、河谷所组成的巨大山原,主要高峰有雀儿山(6 168 m)、格聂山(6 204 m)、海子山夏塞峰(5 833 m)等。本文选取的研究区沙鲁里山系中段西侧以金沙江为界,东至理塘县喇嘛垭乡附近,北达白玉县盖玉镇附近,南到巴塘县波密乡附近,总面积约 7 124.46 km²。G318 国道(川藏公路)、川藏铁路自东向西横穿研究区[图 1(b)]。

研究区地貌大体上是缓和起伏的高原夷平面作为基础,这与构造隆升强烈、河流不断下切的喜马拉雅山和念青唐古拉山的地貌格局显著不

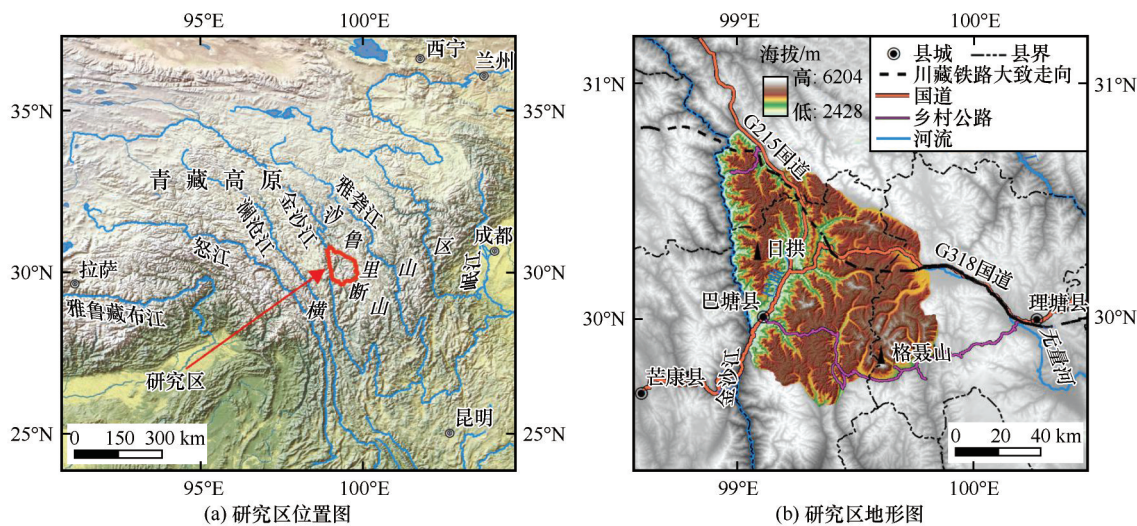


图1 研究区地理位置图
Fig. 1 Location of study area

同。夷平面海拔约4 500~4 700 m。夷平面以下河谷发育,多宽谷,山麓及斜坡上是茂密的高原森林,宽谷底部是弯曲的河道和密集的沼泽草甸。夷平面以上的古蚀残余山海拔多在6 000 m左右,峰顶终年积雪,是古代和现代冰川发育的中心,山脊呈刃状,坡壁地形复杂,存在大量海拔较高的、基岩裸露的常年积雪或季节性积雪区域。这些夷平面上的山岭相对高差多在500~1 500 m之间,也是雪崩赖以发育的地形基础。由于海拔高度、南北走向的山脉和大气环流的影响,研究区属高山高原气候,太阳辐射强,日温差大,降水季节分布不均,具有垂直分布明显和区域性差异大的特点。据毗邻的理塘县气象站(海拔3 948.9 m)监测数据,平均气温3.0℃,极端最高气温25.6℃,最低

气温-30.6℃,年平均地面温度5.9℃,年降雨量为722.2 mm。研究区降水主要集中于6、7、8三个月,季节性雪崩区固态降水量约300 mm左右,积雪层中深霜较为发育^[26],为雪崩的发生提供了丰富的物质条件。

1.2 雪崩编目数据库

雪崩运动过程不仅使积雪受到扰动、污染,在光谱上呈现明显的变化,还影响植物群落的发育和分布,形成了区别于其他高海拔地区自然灾害的遥感解译特征^[27-29]。主要有雪崩沟槽、雪崩碎屑尾、雪崩巨砾舌等地貌标志,植被缺失、植被群落改变等植被标志,以及雪崩雪堆光谱变化等。通过遥感解译,辅以野外调查验证[图2(a)、2(b)],可建立较为完整的雪崩编目数据库。

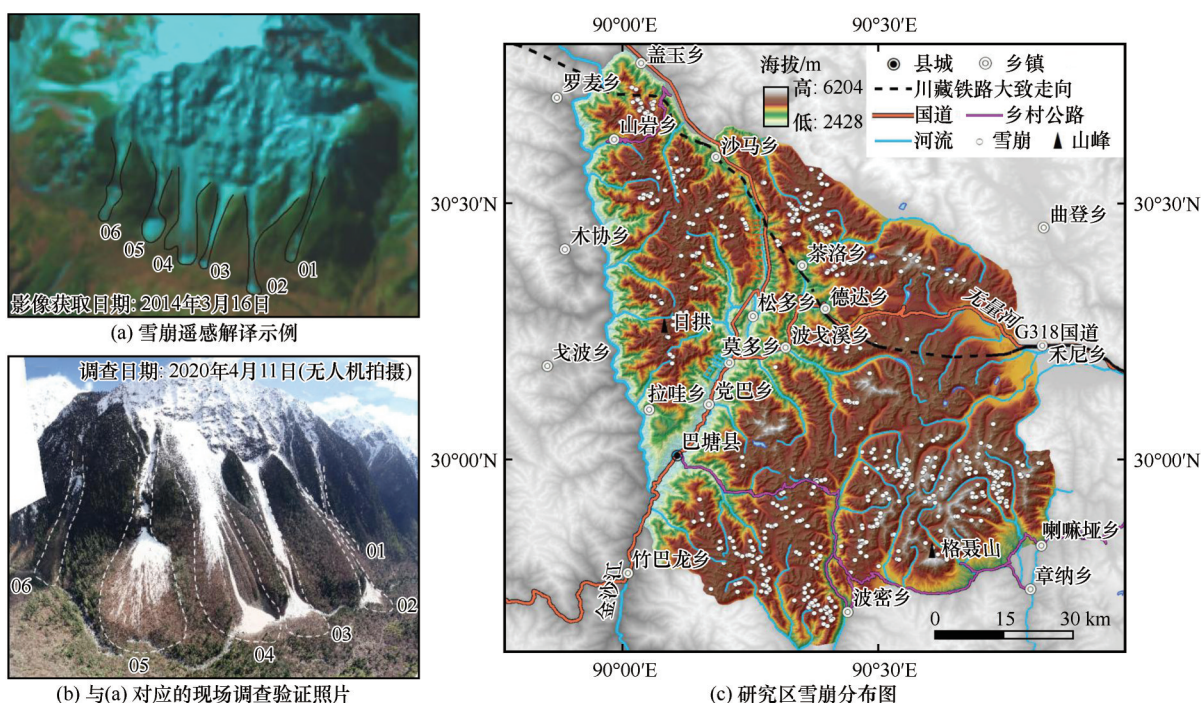


图2 雪崩遥感解译与分布图

Fig. 2 Snow avalanche distribution based on remote sensing interpretation

本文采用的遥感数据源主要有 Landsat 5、7、8 等。ALOS DSM (12.5 m 分辨率)用于雪崩地貌校验,冰川编目数据用于剔除高海拔区域的冰川这类在遥感影像上容易混淆的地物。笔者自2018年11月以来,先后5次到研究区开展现场调查。其中,前2次主要开展雪崩发育特征调查,结合文献资料,梳理形成雪崩堆积体光谱变化标志、雪崩地貌标志、雪崩活动区域植被分布特征等雪崩遥感解译标志;后3次现场调查主要是开展遥感解译结果的

现场验证。在遥感解译中获得了562个解译结果,通过野外调查验证,剔除了26个错误样本。这26个错误样本主要为岩崩,其地貌单元与雪崩有一定的相似之处,在积雪覆盖时容易被错误识别为雪崩。最终,获得536处雪崩样本数据[图2(c)]。通过GIS随机生成与雪崩样本同等数量的随机点作为非雪崩样本(536个),组成共计1 072个样本的数据集。其中,随机抽取846个(80%)样本用于机器学习建模,余下226个(20%)样本用于模型检验,训练

样本和检验样本中的雪崩样本和非雪崩样本数量相同。

2 评价因子筛选与易发性评价模型

2.1 评价单元选择

评价单元是雪崩易发性评价因子定量提取、易发性评价计算最基础的单元。确定评价单元是区域雪崩易发性评价的一个重要步骤^[30-31]。目前,易发性评价单元有栅格单元、斜坡单元、行政单元等。栅格单元划分简单易行、客观且准确性高。因此,本文选取栅格单元作为评价分析单元。此外,栅格尺寸大小直接影响易发性评价结果的精度。考虑到研究区实际情况,选取 100 m×100 m 的栅格单元,共计 713 033 个栅格单元。

2.2 评价因子提取与筛选

雪崩形成的影响因素很多,包括积雪厚度、雪晶大小与形状、含水率、密度、雪层结构、硬度、雪温与温度梯度、海拔、相对高差、坡度、坡向、植被类型与覆盖率、风速、风向、降雪等^[1]。从总体上来说,这些影响因素可以归纳为地形地貌、气候气象、积雪特性及其他因素。评价因子的提取应考虑到因子的代表性,及其能否较全面的反映雪崩形成条件、能否量化表达等^[30,32]。因此,选取了以下 17 个可量化提取因素,包含:海拔、坡度、坡向、地面曲率、地形起伏度、地面粗糙度、地表切割深度、高程变异系数、地形湿度指数、植被覆盖指数、水系(距河流距离)、断层(距断层距离)、平均年降雪量、平均年降雪日数、1 月平均气温、年最大积雪深度、地表覆盖类型。其数据源如表 1 所示。

表 1 评价因子及数据源
Table 1 Evaluate factors and data sources

分类	评价因子	数据源
地形地貌	海拔、坡度、坡向、地面曲率、地形起伏度、地面粗糙度、地表切割深度、高程变异系数、地形湿度指数	ALOS DSM(12.5 m 分辨率)
气候气象	平均年降雪量、平均年降雪日数、1 月平均温度	研究区及周边各站点近 30 年的气象数据统计插值
积雪特性	年最大积雪厚度	研究区及周边各站点近 30 年的气象数据统计插值
	水系	全国地理信息资源目录服务系统的 1:25 万公开版基础地理数据
其他	断层	1:20 万区域地质图
	植被覆盖指数	Landsat 8 卫星影像数据
	地表覆盖类型	全国地理信息资源目录服务系统的 30 m 全球地表覆盖数据

易发性因子的筛选需要考虑到评价因子之间的多重共线性。多重共线性是指模型中的解释变量之间由于存在精确相关关系或高度相关关系而使结果不够客观准确^[33]。本文采用方差膨胀因子(VIF)检验评价因子之间的多重共线性,筛选出更准确的评价因子。VIF 的取值大于 1,VIF 值越接近于 1,多重共线性越轻,反之越重。通常以 10 作为判断边界。当 VIF<10,不存在多重共线性;当 10≤VIF<100,存在较强的多重共线性;当 VIF≥100,存在严重多重共线性^[34]。通过提取训练数据集及各样本的所有评价因子的值进行共线性诊断,结果如图 3 所示。依次剔除 VIF 值最大的因子(平均年降雪量、海拔、地形起伏度),最终筛选出因子为:坡度、坡向、地表曲率、地面粗糙度、地表切割深度、高程变异系数、地形湿度指数、植被覆盖指数、水系、断层、

平均年降雪日数、1 月平均气温、最大积雪厚度、地表覆盖类型,共计 14 个评价因子。在被剔除的因子中,平均年降雪量与海拔相关性较强,同时,又与最大积雪厚度存在较强的正相关关系,因而存在严重多重共线性;坡度、坡向、地表曲率、地形起伏度等地形地貌因子均为 DSM 基础数据通过 GIS 空间分析获得,因而检验出海拔、地形起伏度两个存在多重共线性的因子。最终选定的各评价因子 VIF 值均小于 10,其中最大值为 7.205,最小值为 1.014,满足多重共线性分析的要求。

2.3 评价模型

2.3.1 支持向量机(SVM)

SVM(Support Vector Machine)是一种基于结构风险最小化原则的重要监督学习二值分类器模型,已被广泛用于解决线性和非线性问题^[35]。SVM



图3 雪崩评价因子选择过程

Fig. 3 The Selection process of snow avalanche conditioning factors

的基本原理是通过预先选择的非线性关系将输入向量映射到高维特征空间,并在该空间中寻找最优分类超平面,使两类之间的分类区间最大化^[36],这个映射关系如下。

$$f(x) = [\omega \times \varphi(x)] + b \tag{1}$$

式中:ω是高维空间超平面的特征向量;φ是低维空间到高维空间变换的映射函数;b是阈值。

SVM的关键是核函数的构造。核函数主要用于度量相似性,包括Sigmoid核函数、径向基核函数(RBF)、多项式核函数、线性核函数等。通过RBF可以将样本映射到高维空间,对非线性样本的预测和分析有很好的效果^[37]。雪崩易发性评价是一个典型的受多种因素影响的非线性问题。因此,本文选择RBF作为核函数。RBF的表现主要受惩罚因子C和径向基函数参数γ的影响。

2.3.2 决策树(DT)

DT(Decision Tree)是一种用于分类和回归的

非参数有监督学习方法,其目标是创建一个模型,通过学习从数据特性中推断出的简单决策规则来预测目标变量的值^[38]。本文采用Python调用“Scikit-learn”库^[39]中的决策树算法。Scikit-learn中的决策树算法使用的是CART算法的优化版本,其特征选择是基于信息熵或者基尼系数实现的。

信息熵反应的是信息杂乱程度,信息越杂乱(越不纯),则信息熵越大;反之,信息熵越小。基尼系数在简化模型的同时还保留了熵模型的优点。基尼系数代表了模型的不纯度。基尼系数越小,不纯度越低,特征越好。这和信息增益(率)正好相反。此外,在不加限制的情况下,决策树会生长到衡量不纯度的指标最优,或者直到没有更多的特征可用为止。这样的决策树往往会产生过拟合问题。为了让决策树有更好的泛化性,需对决策树进行剪枝,主要涉及到限制树的最大深度、内部节点再划分所需最小样本数、叶子节点最少样本数等参数。

2.3.3 多层感知器(MLP)

MLP(Multilayer Perceptron)是一种前馈的人工神经网络模型,它将多个输入数据集映射到单个输出数据集(图4)。单层感知器只能学习线性函数,而MLP也可以学习非线性函数^[40],适用于雪崩易发性评价这类非线性问题。MLP的参数主要有隐藏层中的神经元数量、激活函数类型等。激活函数的作用是将非线性引入神经元的输出。MLP可使用任何形式的激活函数,但是为了使用反向传播算法进行有效学习,激活函数必须限制为可微函数。常用的激活函数有Sigmoid、Tanh和ReLU等函数。

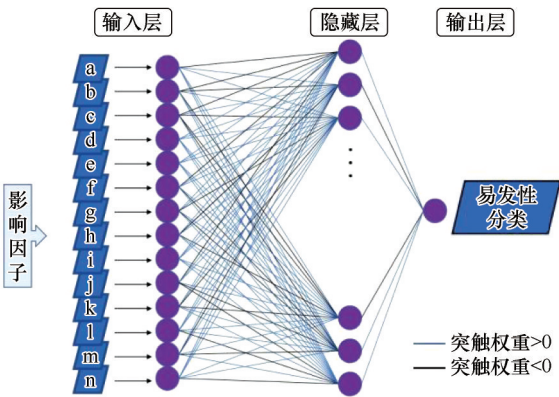


图4 MLP模型结构示意图

Fig. 4 Schematic diagram of MLP model structure

2.3.4 K最邻近法(KNN)

KNN(K-nearest neighbor)是一种通过找到在距离上离待分类样本最近的一些训练样本,并从这些

样本中预测待分类样本标签的方法^[41]。在分类决策中,KNN只根据最近的一个或多个训练样本的类别来预测待分类样本的类别。KNN方法思路简单,易于实现,不足之处是计算量较大,因为需要对每一个待分类的样本都要计算它到全体训练样本的距离,才能求得它的 k 个最邻近点。KNN主要依赖于周围有限的相邻样本, k 值的最佳选择是高度依赖于数据的。较大的 k 会抑制噪声的影响,但使分类边界不那么清晰。一般来说,KNN分类算法包括以下四个步骤^[41]:

- ①准备数据并对数据进行预处理;
- ②计算待分类点与其他训练样本点之间的距离;
- ③对每个距离排序,然后选择距离最小的 k 个点;
- ④根据少数服从多数的原则,将待分类点划分为 k 个点中占比最高的类别。

2.4 模型精度检验

2.4.1 Kappa系数

Kappa系数检验是一种用混淆矩阵检验模型预测结果与实际值一致性的方法。Kappa系数检验是用混淆矩阵来计算Kappa系数,将验证数据集中的雪崩样本标记为1,非雪崩样本标记为0,模型的预测值与实际值的关系如表2所示。

表2 实际值与预测值关系的二进制表
Table 2 Binary table of the relationship between actual and predicted values

实际值	模型预测值	样本数量
1	1	N_1
1	0	N_2
0	1	N_3
0	0	N_4

Kappa系数的计算公式如下:

$$Ka = \frac{P_a - P_e}{1 - P_e} \tag{2}$$

其中 P_a 为模型预测中正确划分的样本数与总样本数之比,计算公式如下:

$$P_a = \frac{N_1 + N_4}{N_1 + N_2 + N_3 + N_4} \tag{3}$$

P_e 为所有分类中预测样本数与实际样本数的乘积之和与总样本数的平方之比,计算公式为:

$$P_e = \frac{(N_1 + N_2)(N_1 + N_3) + (N_4 + N_2)(N_4 + N_3)}{(N_1 + N_2 + N_3 + N_4)^2} \tag{4}$$

Kappa系数值域在-1到1之间,该值通常大于

0。数值越大,说明评价模型的准确性越高^[42-43]。Kappa系数各数值区间以及意义详见表3。

表3 Kappa系数和AUC值域区间及意义
Table 3 The ranges of Kappa coefficient and AUC and their significance

Kappa系数		AUC	
区间	意义	区间	意义
[0.8, 1]	几乎完全一致	(0.9, 1]	模型精度高
[0.6, 0.8)	高度的一致性	(0.7, 0.9]	模型精度较高
[0.4, 0.6)	中等的一致性	(0.5, 0.7]	模型精度较低
[0.2, 0.4)	一般的一致性	[0, 0.5]	无预测价值
[0, 0.2)	极低的一致性		

2.4.2 ROC曲线

ROC曲线(Receiver Operating Characteristic curve)是根据X轴上的真阳性率(敏感度)和Y轴上的假阳性(1-特异性)在不同阈值处生成的图形来直观表示模型评价精度。敏感度和特异性实质上表示模型正确判断雪崩和非雪崩的概率,但这两个指标并不能显示模型性能的整体准确性,所以一般采用AUC(Area Under Curve)值来检验模型精度。AUC值是指ROC曲线与坐标轴围成的面积,是计算二值分类器性能的评价指标。AUC的值域为[0, 1],当AUC值越接近1表明模型预测准确性越高^[44-45],其各数值区间及意义详见表3。整个评价过程如图5所示。

3 雪崩易发性评价结果与讨论

3.1 雪崩易发性指数

通过GIS平台将训练样本和验证样本的各评价因子的值提取出来,借助于Python语言调用“Scikit-learn”库^[39]中的SVM、DT、MLP和KNN算法进行运算,同时调用网格搜索(Grid search)算法进行参数寻优。运算结束后,将结果导入GIS中输出栅格。在SVM模型中,通过网格搜索获得了最优超参数:C为1, γ 为0.1。将研究区各栅格单元的评价因子引入评价模型,计算易发性指数。最后得到雪崩易发性指数图如图6(a)所示,其值域为[0,0.964]。DT模型对输入样本进行训练之后自动构建分类规则,确定了不同分类特征及其阈值。经过调参寻优,形成了基于信息熵的分枝方法,最大深度为3层、分割内部节点所需的最小样本数为25、叶子节点上的最小样本数为17的决策树。该模型雪崩易发性指数图如图6(b)所示,其值域为

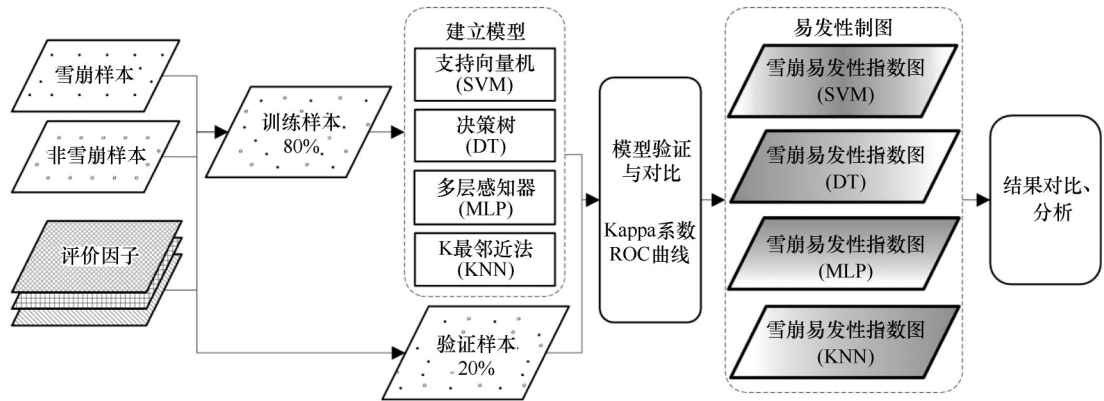


图5 雪崩易发性评价流程图

Fig. 5 Flow chart of snow avalanche susceptibility evaluation

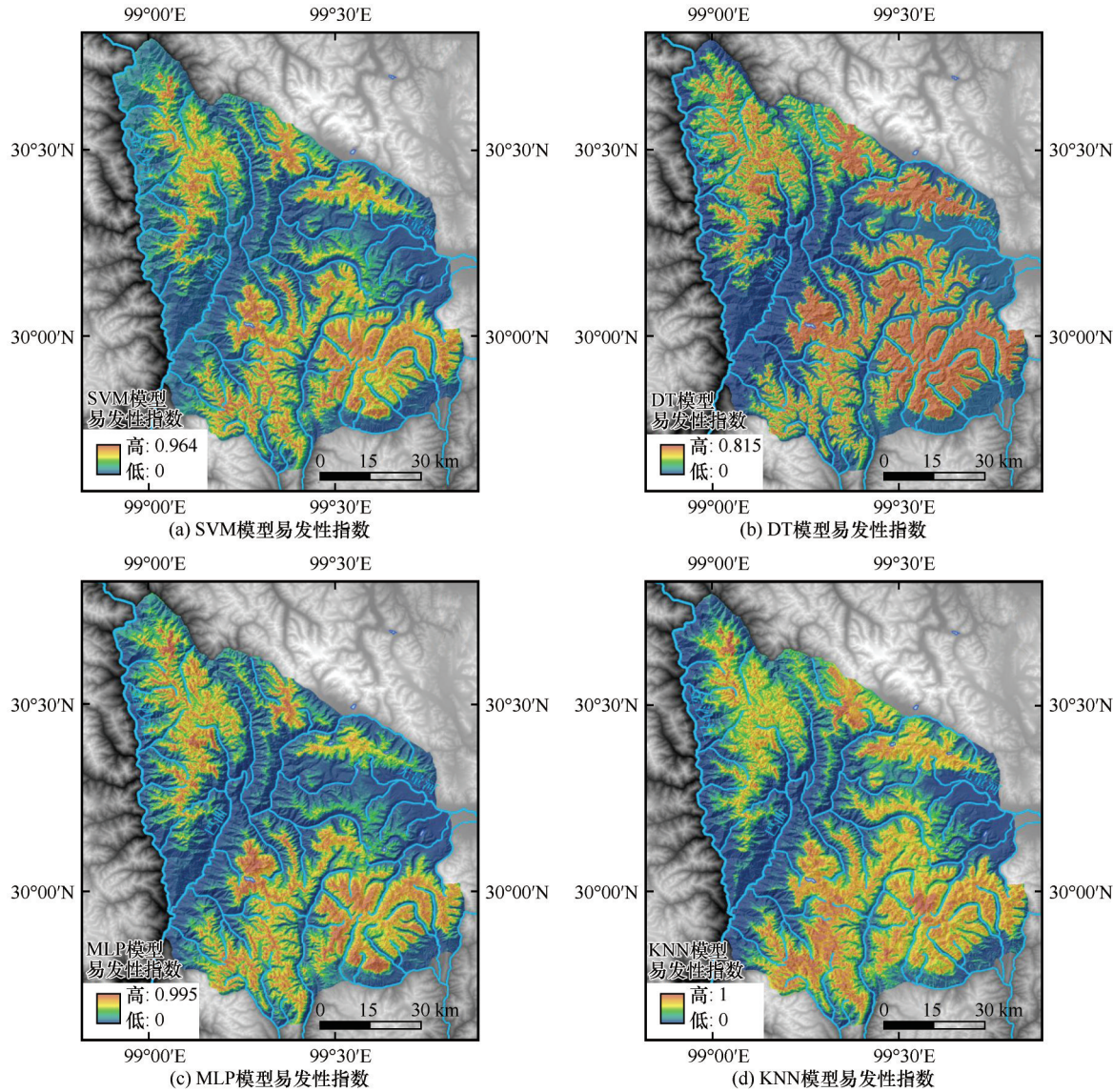


图6 雪崩易发性指数图

Fig. 6 Snow avalanche susceptibility index maps

[0,815]。在MLP模型中,通过调参寻优,构造了1个包含1个输入层、1个隐含层和1个输出层的3层神经网络,形成30个神经元,激活函数为Tanh。该模型雪崩易发性指数如图6(c)所示,其值域为[0,995]。

在KNN模型中,通过网格搜索调参,获得了最优超参数:N_neighbors(KNN中的“K”)为10,P为1(曼哈顿距离),Weights(权重)为Distance(权重和距离成反比,距离预测目标越近具有越高的权重)。该模型的雪崩易发性指数如图6(d)所示,其值域为[0,1]。

3.2 雪崩评价因子重要性分析

通过机器学习算法对训练样本的训练建模过程中,可以获得变量的重要性,即评价因子的权重,如图7所示。由于不同算法的原理不同,其计算得出的评价因子权重也并不完全一致。除去KNN模型中各个评价因子权重均相同外,其他3个模型得到的各评价因子权重在总体上存在着不同程度的类似和差异之处。其中,这3个模型中最重要

的因子都是1月平均气温。1月平均气温表征了雪崩发育区域的冷储条件,也是区别于雪崩不易发区域的重要特征。DT、MLP和SVM模型中重要性排序第2的因子分别为植被覆盖指数、植被覆盖类型和高程变异系数,排序第3的因子分别是最大积雪厚度、高程变异系数和坡度。此外,超过平均数的因子还有平均降雪日数、水系。植被覆盖指数和植被覆盖类型涉及到雪崩形成的下垫面状况,其中裸地、草地等有利于雪崩的形成,茂密的森林是雪崩形成和运动的主要阻碍。高程变异系数和坡度体现了局地地形变化,而高差悬殊和适宜的坡度区间正是积雪等斜坡物质运动形成的必要的地形条件。最大积雪厚度表征了雪崩形成的物质(积雪)条件。平均降雪日数是一年中降雪日数的总和,从侧面表征了当地降雪强度及频率等,体现了雪崩的物质(积雪)来源。

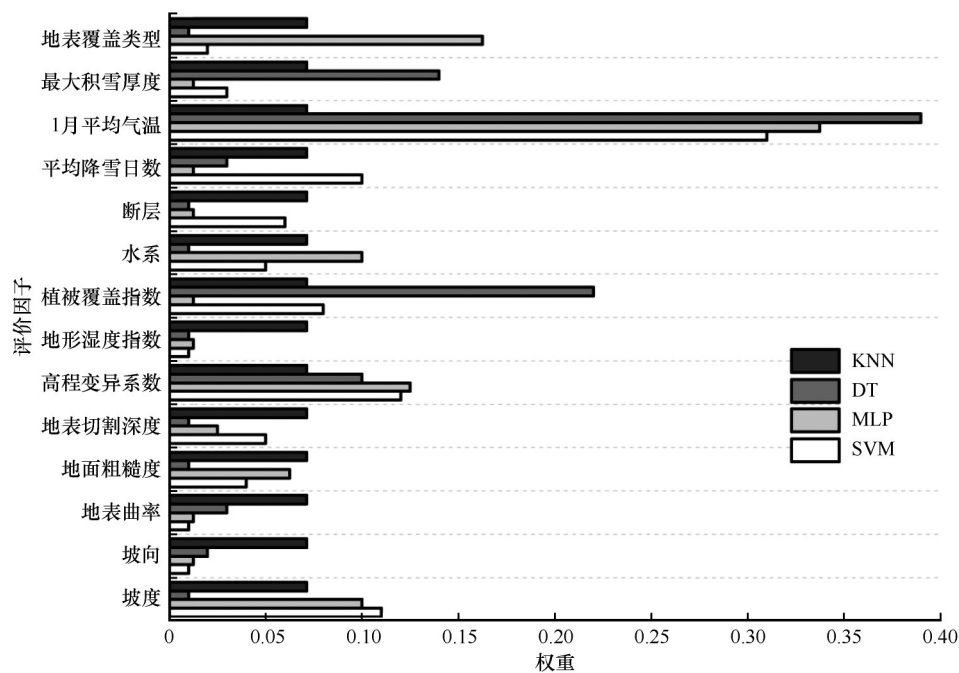


图7 评价因子权重条形图
Fig. 7 Bar chart of evaluation factor weight

3.3 模型精度检验结果与对比分析

验证数据集通过SVM、DT、MLP和KNN四种模型预测结果的相关值及Kappa系数如表4所示。SVM、DT、MLP和KNN模型的Kappa系数分别为0.720、0.570、0.711和0.672,除了DT外,其余的Kappa系数均大于0.6,表明SVM、MLP和KNN模型对验证数据集的预测结果与实际值存在高度的一致性,DT对验证数据集的预测结果与实际值存在中等的一致性。ROC曲线如图8所示。SVM、

DT、MLP和KNN模型的AUC值分别为0.912、0.801、0.891和0.903,均大于0.8,表明SVM和KNN模型的预测精度高,DT和MLP模型的预测精度较高。综合Kappa系数和ROC曲线检验的结果,这4中机器学习算法在雪崩易发性评价上都具有较好或很好的预测能力,其中SVM模型的Kappa系数和AUC值均为最高,为该项雪崩易发性评价精度最佳的模型。

此外,结合现场调查的认识,这4种机器学习算

表4 各模型结果相关值及Kappa系数
Table 4 Kappa coefficients and relevant values of the prediction results

	SVM	DT	MLP	KNN
N_1	102	98	98	105
N_2	8	12	12	5
N_3	22	34	19	30
N_4	83	71	86	75
Pe	0.502	0.503	0.501	0.503
P_a	0.860	0.786	0.856	0.837
K_a	0.720	0.570	0.711	0.672

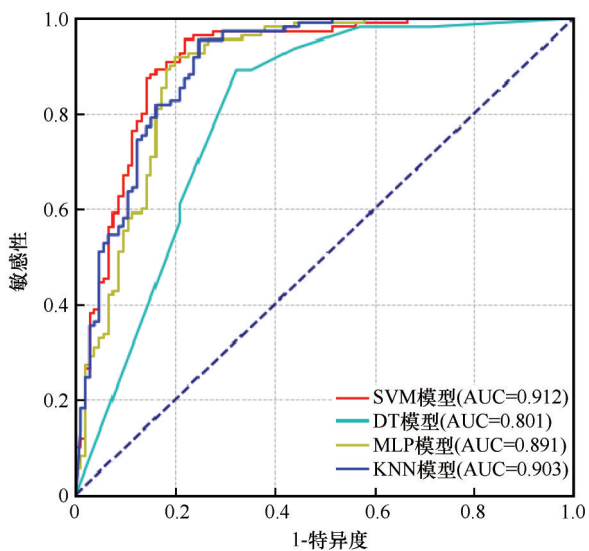


图8 验证数据集ROC曲线
Fig. 8 ROC curve of prediction rate for the four models with validation dataset

法得到的易发性指数图是合理的。易发性指数图中易发性指数较高的区域主要分布在海拔较高的区域,在金沙江、巴曲等干热河谷两岸区域的易发性指数最低,这与野外调查中雪崩发育的区域基本一致。同时,易发性指数图还指明了此前遥感解译中未曾解译到的部分区域,有助于研究区内更多雪崩范围的遥感解译。机器学习算法在雪崩易发性建模过程中,对各评价因子的值域进行解析,建立了基于学习样本的雪崩分类函数或分类规则,然后对研究区内各评价单元进行分类计算,得到各评价单元雪崩易发分类的原始倾向评分。理论上,这些结果也与各评价因子的雪崩易发值域的空间分布叠加结果基本一致。模型精度结果也表明了这4种模型均具有较好的预测精度。因此,这4种机器学习算法均适用于沙鲁里山系雪崩易发性制图。

3.4 雪崩易发性分区

采用自然间断法将AUC值最高的SVM模型易发性指数图划分为极高、高、中、低和极低易发性区,得到雪崩易发性区划图(图9)。其中,极高、高、中、低和极低易发性区分别占总面积的13.1%、12.9%、11.1%、17.6%和45.3%。易发性高的区域主要分布在格聂山、日拱山等地,多位于夷平面(海拔约4 500~4 700 m)以上,海拔较高。其中,极高易发区平均海拔约4 939 m,高易发区平均海拔约4 859 m。这些区域基岩裸露,在雪季多有积雪。研究区虽然远离大洋,但是研究区南侧的三江并流区的怒江、澜沧江、金沙江等深切河谷构成了南来湿润气流北上的通道。该区域年均降雪量(300 mm)虽说没有帕隆藏布流域(约1 000 mm)等地那么大,但是降雪量分布较集中,往往集中在每年降雪最大的2~3次降雪过程中。特别是在每年春季,孟加拉海水汽顺深切河谷北上进入研究区,在高海拔地区降落大量湿雪,为雪崩的发生奠定了物质基础。夷平面上的蚀余山经过长期的冰川作用、流水作用后,古冰斗和雪蚀洼地地貌发育,成为良好的储雪场地。蚀余山的坡度条件为雪崩的运动提供了足够的动力条件。而在夷平面(海拔约4 500~4 700 m)以下,一方面年平均气温较高,年降雪相对较少,特别是在靠近金沙江一侧,为典型的干热河谷,据巴塘气象站监测数据,年极端最大积雪深度仅4 cm,出现时间为2006年12月13日;另一方面,山麓及斜坡上多是茂密的高原森林,茂密的森林在强降雪天气时能够在很大程度上阻碍新雪的沉降和再次分配,还能形成锚点,提供抗滑力锁固积雪层。因而在夷平面高程以下区域,雪崩并不易发。

从雪崩易发性指数图和区划图可以看出,雪崩对现G318国道(川藏公路)、G215国道影响较小。王彦龙在其专著《川藏公路雪害研究》中记录的川藏公路海子山垭口段的雪害主要为风吹雪,并无雪崩记录^[26]。此外,川藏铁路海子山越岭段线路行走标高略低于现G318国道,且为隧道形式从毛垭草原西侧穿越折多山至德达附近,其受雪崩影响相较于川藏公路更小。经德达后北上,过茶洛乡、沙马乡等地附近,直至金沙江边,主要以隧道形式穿行于沙鲁里山系中。这一段隧道进出口选址高程均较低、周边植被茂密,受雪崩影响较小。但是,研究区北部盖玉乡前往山岩乡的越岭路段、南部波密乡至巴塘县的越岭路段,主要位于雪崩高易发区。由

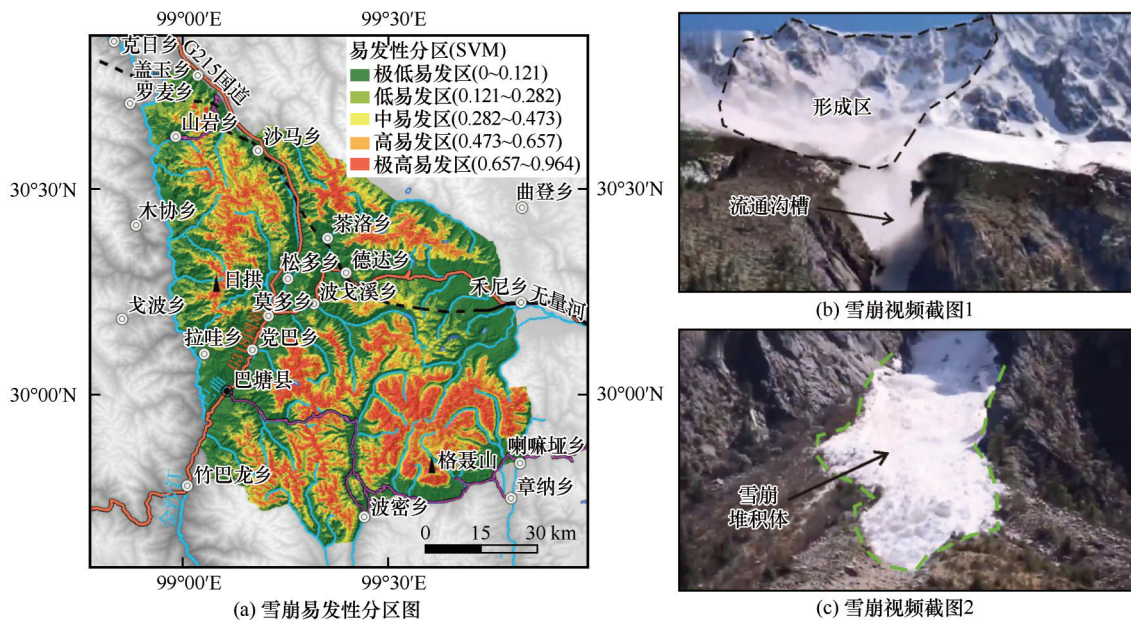


图9 基于SVM的雪崩易发性分区图

Fig. 9 Zoning map of snow avalanche susceptibility based on SVM

于这两段公路均属于乡村公路,行车量较少,且较偏僻,现代通讯不一定覆盖到,其雪崩活动鲜见报道。不过,在研究区南部的格聂山主峰附近有2次关于雪崩的报道。其中一次是在2006年12月,美国著名登山家查理·福勒(Charlie Fowler)和克里斯汀·博斯科夫(Christine Boskoff)尝试从格聂东壁攀登的时候,在5300m的冰川附近遭遇雪崩遇难^[46]。另外一次是在2020年5月,一只徒步探险队在理塘县章纳乡老冷古寺附近拍到格聂主峰东坡(99.6409°E, 29.8108°N)正在流动的沟槽型湿雪崩视频[图9(b)、9(c)为该次雪崩视频截图]^[47]。这两次雪崩事件均发生在研究区内夷平面以上海拔较高的人迹罕至的区域。

4 结论

本文通过遥感解译辅以野外调查验证,构建了沙鲁里山系中段雪崩编目数据库,借助SVM等4种机器学习算法开展雪崩易发性评价,得出以下结论。

(1)通过遥感解译识别雪崩562处,结合野外调查验证,剔除了26个错误样本,共计获得536处雪崩样本数据,建立了较为完整的雪崩编目数据库。选取17个可量化提取的评价因子,通过方差膨胀因子(VIF)检验评价因子之间的多重共线性,筛选出坡度、坡向、地表曲率等共计14个评价因子。

(2)采用SVM、DT、MLP、KNN机器学习算法

训练模型,获得的易发性指数分别在 $[0, 0.964]$ 、 $[0, 0.815]$ 、 $[0, 0.995]$ 、 $[0, 1]$ 范围内,其Kappa系数分别为0.720、0.570、0.711和0.672, AUC值分别为0.912、0.801、0.891和0.903。结果表明这4种模型均具有较好或很好的预测精度,适用于沙鲁里山系中段雪崩易发性评价,其中SVM模型的Kappa系数和AUC值均为最高,为该项雪崩易发性评价精度最佳的模型。机器学习算法建模过程中获得的主要影响因子有1月平均气温、植被覆盖指数、植被覆盖类型、高程变异系数、最大积雪厚度、坡度等。

(3)该区域雪崩极高、高、中、低和极低易发性区分别占总面积的13.1%、12.9%、11.1%、17.6%和45.3%。雪崩极高易发区和高易发区主要分布在格聂山、日拱山等地,多位于夷平面以上,极高易发区平均海拔约4939m,高易发区平均海拔约4859m。雪崩对现G318国道(川藏公路)、G215国道影响较小,对横穿研究区内的川藏铁路的影响相较于川藏公路更小。但是,研究区北部盖玉乡前往山岩乡的越岭路段、南部波密乡至巴塘县的越岭路段主要位于雪崩高易发区。该结果可为横穿沙鲁里山系的川藏铁路等重大工程建设的雪崩防灾减灾工作提供科学依据和方法借鉴。

参考文献(References):

- [1] Schweizer J, Bruce Jamieson J, Schneebeli M. Snow avalanche formation[J]. Reviews of Geophysics, 2003, 41(4):

- 1-25.
- [2] Podolskiy E A, Izumi K, Suchkov V E, et al. Physical and societal statistics for a century of snow-avalanche hazards on Sakhalin and the Kuril Islands (1910 - 2010) [J]. *Journal of Glaciology*, 2014, 60(221): 409-430.
- [3] Zhang Dianfa, Fengquan L, Jianmin B. Eco-environmental effects of the qinghai-Tibet plateau uplift during the Quaternary in China [J]. *Environmental Geology*, 2000, 39 (12) : 1352-1358.
- [4] Wang Tao, Zhao Yutong, Xu Chaoyi, et al. Atmospheric dynamic constraints on tibetan plateau freshwater under paris climate targets [J]. *Nature Climate Change*, 2021, 11 (3) : 219-225.
- [5] Keiler M, Knight J, Harrison S. Climate change and geomorphological hazards in the eastern European Alps[J]. *Philosophical Transactions Series A, Mathematical, Physical, and Engineering Sciences*, 2010, 368(1919): 2461-2479.
- [6] McClung D M. Avalanche character and fatalities in the high mountains of Asia[J]. *Annals of Glaciology*, 2016, 57(71): 114-118.
- [7] Zhang Gen, Sun Chunwei, Yang Chengye, et al. Study on the developmental characteristics and distribution pattern of grooved avalanche in palongzangbu river basin[J]. *Plateau Science Research*, 2021, 5(1): 35-43. [张根, 孙春卫, 杨成业, 等. 帕隆藏布流域沟槽型雪崩发育特征及分布规律研究[J]. *高原科学研究*, 2021, 5(1): 35-43.]
- [8] Hu Wentao Hu, Yao Tandong, Yu Wusheng, et al. Advances in the study of glacier avalanches in high Asia[J]. *Journal of Glaciology and Geocryology*, 2018, 40(6): 1141-1152. [胡文涛, 姚檀栋, 余武生, 等. 高亚洲地区冰崩灾害的研究进展[J]. *冰川冻土*, 2018, 40(6): 1141-1152.]
- [9] Wang Shijin, Wen Jiahong. Characteristics, influence of cryosphere disaster and prospect of discipline development[J]. *Bulletin of Chinese Academy of Sciences*, 2020, 35(4): 523-530. [王世金, 温家洪. 冰冻圈灾害特征、影响及其学科发展展望[J]. *中国科学院院刊*, 2020, 35(4): 523-530.]
- [10] Schweizer J, Bartelt P, van Herwijnen A. Snow avalanches [M]//*Snow and Ice-related Hazards, Risks and Disasters*. Amsterdam: Elsevier, 2015: 395-436.
- [11] Haeberli W, Whiteman C. Snow and ice-related hazards, risks, and disasters [M]. Waltham, MA: Academic Press, 2021: 259-296.
- [12] Jamieson B, Stethem C. Snow avalanche hazards and management in Canada: challenges and progress[J]. *Natural Hazards*, 2002, 26(1): 35-53.
- [13] Seliverstov Y, Glazovskaya T, Shnyparkov A, et al. Assessment and mapping of snow avalanche risk in Russia[J]. *Annals of Glaciology*, 2008, 49: 205-209.
- [14] Chen Hao, Ning Chen, Nan Zhuotong, et al. Correction of the daily precipitation data over the tibetan plateau with machine learning models [J]. *Journal of Glaciology and Geocryology*, 2017, 39(3): 583-592. [陈浩, 宁忱, 南卓铜, 等. 基于机器学习模型的青藏高原日降水数据的订正研究[J]. *冰川冻土*, 2017, 39(3): 583-592.]
- [15] Choubin B, Mosavi A, Alamdarloo E H, et al. Earth fissure hazard prediction using machine learning models[J]. *Environmental Research*, 2019, 179: 108770.
- [16] Xiong Pan, Tong Lei, Zhang Kun, et al. Towards advancing the earthquake forecasting by machine learning of satellite data [J]. *Science of the Total Environment*, 2021, 771: 145256.
- [17] Youssef A M, Pourghasemi H R, Pourtaghi Z S, et al. landslide susceptibility mapping using random forest, boosted regression tree, classification and regression tree, and general linear models and comparison of their performance at Wadi Tayyah Basin, Asir Region, Saudi Arabia [J]. *Landslides*, 2016, 13(5): 839-856.
- [18] Fan Yubin, Guo Weina, Ke Changqing. Texture-assisted glacier recognition based on SAR image[J]. *Journal of Glaciology and Geocryology*, 2019, 41(6): 1326-1334. [范宇宾, 郭唯娜, 柯长青. 纹理特征辅助的 SAR 影像冰川识别[J]. *冰川冻土*, 2019, 41(6): 1326-1334.]
- [19] Kavzoglu T, Colkesen I, Sahin EK. *Landslides: theory, practice and modelling* [M]: Springer International Publishing, 2019: 283-301.
- [20] Huang Faming, Cao Zhongshan, Guo Jianfei, et al. Comparisons of heuristic, general statistical and machine learning models for landslide susceptibility prediction and mapping[J]. *Catena*, 2020, 191: 104580.
- [21] Yang Jinming, Li Chengzhi, Li Lanhai, et al. Automatic detection of regional snow avalanches with scattering and interference of C-band SAR data[J]. *Remote Sensing*, 2020, 12(17): 2781.
- [22] Choubin B, Borji M, Hosseini F S, et al. Mass wasting susceptibility assessment of snow avalanches using machine learning models[J]. *Scientific Reports*, 2020, 10(1): 18363.
- [23] Choubin B, Borji M, Mosavi A, et al. Snow avalanche hazard prediction using machine learning methods[J]. *Journal of Hydrology*, 2019, 577: 123929.
- [24] Mosavi A, Shirzadi A, Choubin B, et al. Towards an ensemble machine learning model of random subspace based functional tree classifier for snow avalanche susceptibility mapping[J]. *IEEE Access*, 2020, 8: 145968-145983.
- [25] Rahmati O, Ghorbanzadeh O, Teimurian T, et al. Spatial modeling of snow avalanche using machine learning models and geo-environmental factors: comparison of effectiveness in two mountain regions[J]. *Remote Sensing*, 2019, 11(24): 2995.
- [26] Wang Yanlong. *Snow disaster and prevention along Sichuan-Tibet Highway* [M]. Beijing: Ocean Press, 1993. [王彦龙. *川藏公路沿线雪害与防治* [M]. 北京: 海洋出版社, 1993.]
- [27] Singh K K, Singh D K, Thakur N K, et al. Detection and mapping of snow avalanche debris from Western Himalaya, India using remote sensing satellite images[J]. *Geocarto International*, 2020: 1-19.
- [28] Bühler Y, Hüni A, Christen M, et al. Automated detection and mapping of avalanche deposits using airborne optical remote sensing data[J]. *Cold Regions Science and Technology*, 2009, 57(2/3): 99-106.
- [29] Lato M J, Frauenfelder R, Bühler Y. Automated detection of snow avalanche deposits: segmentation and classification of optical remote sensing imagery [J]. *Natural Hazards and Earth System Sciences*, 2012, 12(9): 2893-2906.
- [30] Lee S. Current and future status of GIS-based landslide susceptibility mapping: a literature review [J]. *Korean Journal of Remote Sensing*, 2019, 35(1): 179-193.
- [31] Hearn G J, Hart A B. Landslide susceptibility mapping: a practitioner's view[J]. *Bulletin of Engineering Geology and the Environment*, 2019, 78(8): 5811-5826.
- [32] Pourghasemi H R, Teimoori Yansari Z, Panagos P, et al. Analysis and evaluation of landslide susceptibility: a review on articles published during 2005-2016 (periods of 2005-2012 and 2013-2016) [J]. *Arabian Journal of Geosciences*, 2018, 11(9): 1-12.

- [33] Thompson C G, Kim R S, Aloe A M, et al. Extracting the variance inflation factor and other multicollinearity diagnostics from typical regression results [J]. *Basic and Applied Social Psychology*, 2017, 39(2): 81-90.
- [34] Tamura R, Kobayashi K, Takano Y, et al. Mixed integer quadratic optimization formulations for eliminating multicollinearity based on variance inflation factor[J]. *Journal of Global Optimization*, 2019, 73(2): 431-446.
- [35] Wiederschain G Y. Data mining techniques for the life sciences [J]. *Biochemistry (Moscow)*, 2011, 76(4): 494.
- [36] Suthaharan S. Support vector machine [M]//*Machine Learning Models and Algorithms for Big Data Classification*. Boston, MA: Springer US, 2016: 207-235.
- [37] Liu Quanzhong, Chen C, Zhang Yang, et al. Feature selection for support vector machines with RBF kernel[J]. *Artificial Intelligence Review*, 2011, 36(2): 99-115.
- [38] Wu Xueling, Ren Fu, Niu Ruiqing. Landslide susceptibility assessment using object mapping units, decision tree, and support vector machine models in the Three Gorges of China[J]. *Environmental Earth Sciences*, 2014, 71(11): 4725-4738.
- [39] Pedregosa F, Varoquaux G, Gramfort A, et al. Scikit-learn: Machine learning in Python[J]. *The Journal of Machine Learning Research*, 2011, 12: 2825-2830.
- [40] Ramchoun H, Amine M, Idrissi J, et al. Multilayer perceptron: architecture optimization and training [J]. *International Journal of Interactive Multimedia and Artificial Intelligence*, 2016, 4(1): 26.
- [41] Kramer O. Dimensionality reduction with unsupervised nearest neighbors[M]: Springer, 2013: 13-23.
- [42] Kavzoglu T, Kutlug Sahin E, Colkesen I. An assessment of multivariate and bivariate approaches in landslide susceptibility mapping: a case study of Duzkoy district[J]. *Natural Hazards*, 2015, 76(1): 471-496.
- [43] Kraemer HC. Kappa Coefficient[J]. *Wiley Statsref: Statistics Reference Online*, 2014: 1-4.
- [44] Shahabi H, Jarihani B, Tavakkoli Piralilou S, et al. A semi-automated object-based gully networks detection using different machine learning models: A case study of Bowen catchment, Queensland, Australia [J]. *Sensors (Basel, Switzerland)*, 2019, 19(22): 4893.
- [45] Mandrekar J N. Receiver operating characteristic curve in diagnostic test assessment [J]. *Journal of Thoracic Oncology*, 2010, 5(9): 1315-1316.
- [46] Association Sichuan-mountaineering. Circulate a notice on the confirmation of a U. S. citizen died in an accident in the climbing of Genie Mountain [EB/OL]. 2006. <https://www.8264.com/viewnews-19879-page-1.html>. [四川省登山协会. 关于在格聂山登山遇难的一名美国公民身份被证实的情况通报 [EB/OL]. 2006. <https://www.8264.com/viewnews-19879-page-1.html>.]
- [47] Province Construction-management-preparatory-group-of-Ganzigennie-Shenshan-scenic-area-in-Sichuan. An snow avalanche released in Ganzi, Sichuan Province: a hiking team witnessed the whole process[EB/OL]. 2020. <http://culture.ifeng.com/c/7wB6ieIaGv6>. [四川甘孜格聂神山景区建设管理筹备组. 四川甘孜格聂神山发生雪崩, 一支徒步探险队目睹全过程直呼震撼 [EB/OL]. 2020. <http://culture.ifeng.com/c/7wB6ieIaGv6>.]

Snow avalanche susceptibility evaluation in the central Shaluli Mountains of Tibetan Plateau based on machine learning method

WEN Hong^{1,2}, WU Xiyong², ZHAO Siyuan³, BIAN Rui², ZHOU Guiyu¹,
MENG Shaowei⁴, SUN Chunwei²

(1. Faculty of Intelligence Manufacturing, Yibin University, Yibin 644000, Sichuan, China; 2. Faculty of Geosciences and Environmental Engineering, Southwest Jiaotong University, Chengdu 611756, China; 3. State Key Laboratory of Hydraulics and Mountain River Engineering, College of Water Resource & Hydropower, Sichuan University, Chengdu 610065, China; 4. China Railway Eryuan Engineering Group Co. Ltd, Chengdu 610031, China)

Abstract: Snow avalanches, which are widely and frequently developed at high elevations, seriously threatens the built traffic corridors in the Tibetan Plateau. Susceptibility evaluation of snow avalanche via machine learning model with a high forecast accuracy can be applied to quickly and effectively assess the regional avalanche risk. This paper took the central Shaluli Mountain region as the study area, in which the snow avalanche inventory was established through remote sensing interpretation and field investigation verification. We quantitatively extracted 17 evaluation factors via GIS-based analysis, and these factors were selected through the variance expansion factor (VIF). Four machine learning models containing SVM, DT, MLP and KNN were used to compile the susceptibility index map of snow avalanches, and kappa coefficient and ROC curve were used to verify the accuracy. The results suggested that the susceptibility indexes obtained from SVM, DT, MLP and KNN were in the range of $[0, 0.964]$, $[0, 0.815]$, $[0, 0.995]$ and $[0, 1]$, respectively. The accuracy test results show that these four models all have good prediction accuracy. Among them, the SVM model is the best. The results also indicated that the areas with the high snow avalanche susceptibility mainly distributed in Genie Mountain and Rigong Mountain, most of which were above the planation surface of the Tibetan Plateau. The average altitude of the extremely high snow-avalanche-prone areas is 4 939 m, while the average altitude of the high snow avalanche-prone areas is 4 859 m. The snow avalanche has low perniciousness on the Sichuan-Tibet Highway and the Sichuan-Tibet Railway in the study area. This study can provide theoretical basis and method reference for disaster prevention and mitigation of snow avalanche along Sichuan-Tibet Railway and other major projects across Shaluli Mountains region.

Key words: snow avalanche; susceptibility evaluation; machine learning; Shaluli Mountains; Tibetan Plateau

(责任编辑: 牛富俊)