

袁凯,李武阶,庞晶. 基于决策树算法的鄂东地区冰雹识别技术. 应用气象学报,2023,34(2):234-245.

DOI: 10.11898/1001-7313.20230209

基于决策树算法的鄂东地区冰雹识别技术

袁 凯 李武阶* 庞 晶

(湖北省武汉市气象台, 武汉 430040)

摘 要

冰雹是对流性天气常见的灾害之一,雷达是识别冰雹强有力的工具,为克服现有方法主观性强、特征量阈值不明确以及虚警率高的不足,探究机器学习算法用于冰雹识别的可行性,基于决策树算法利用 2015 年 1 月 1 日—2021 年 12 月 31 日鄂东地区冰雹灾情资料、武汉多普勒天气雷达以及探空资料,将湿球温度高度引入冰雹识别因子中,并根据命中率、虚警率和临界成功指数定量评估其识别能力。结果表明:仅包含回波强度的决策树(强度决策树)和包含回波强度和湿球温度高度的决策树(强度-高度决策树)均能有效识别冰雹,强度-高度决策树较强度决策树的命中率和临界成功指数均小幅提高,且虚警率明显降低;强度决策树识别冰雹的关键因子为组合反射率因子,底层多为 0.5° 和 1.5° 仰角反射率因子,强度-高度决策树的关键因子为 0.5° 仰角反射率因子,底层多为风暴的整体强度属性;个例分析显示强度-高度决策树减少了湿球 0°C 层高度较高时的虚警次数,展现出良好的应用前景。

关键词: 冰雹; 决策树算法; 湿球温度高度; 检验评估

引 言

冰雹是指直径不小于 5 mm 的固态降水,是对流性天气常见的附属灾害^[1],也是湖北省的主要灾害性天气之一^[2-3]。由于突发性强,来势猛烈,破坏性大,冰雹常给鄂东地区造成人员伤亡和财产损失:2009 年 11 月 9 日鄂州发生冰雹天气,造成 1 人死亡,直接经济损失 1152 万元;2015 年 4 月 3 日夜间黄冈蕲春遭受冰雹袭击,降雹时间持续 15 min 以上,雹粒小的如乒乓球,大的如拳头,共造成 15323 人受灾,直接损失 1833 万元;2020 年 3 月 21 日夜间咸宁 4 个乡镇出现冰雹天气,1323 人受灾。

冰雹识别一直是业务天气预报的难点,近年随着多普勒天气雷达网的不断加密、雷达探测技术的发展以及其他学科先进成果的不断引入,以多普勒天气雷达产品为基础的冰雹识别技术取得长足进步。俞小鼎等^[4]指出回波悬垂与弱回波区、有界弱

回波区、垂直累积液态水含量或者垂直累积液态水含量异常大值、三体散射以及强烈的风暴顶辐散等是多普勒天气雷达预警强冰雹的主要特征参量。刁秀广等^[5]指出垂直累积液态水含量及其密度对于冰雹特别是强冰雹的识别具有很好的指示意义;胡胜等^[6]基于广东省 12 个个例,总结归纳了大冰雹风暴单体的典型雷达回波特征。王萍等^[7]、王莎等^[8]和吴剑坤等^[9]分别建立了包括反射率因子、垂直累积液态水含量和回波顶高等多个雷达产品的冰雹识别与预警模型。上述研究详细讨论了强冰雹单体在雷达回波上的三维结构,归纳出强冰雹风暴典型的三维雷达回波特征,并予以定量描述,提高了临近时段内强冰雹的识别准确率。王洪等^[10]利用珠海与澳门共建的中国首部具有双偏振功能的 S 波段多普勒天气雷达对一次华南春季超级单体雹暴进行分析,发现大冰雹除了高反射率因子外,两个偏振方向反射率因子的相关系数较低,同时还具有较低的差分反射率因子。冯晋勤等^[11]、刁广秀等^[12]对比分析冰

2022-09-19 收到, 2022-12-17 收到再改稿。

资助项目: 武汉市气象局临近预报创新团队项目(2022 年度)

* 通信作者, 邮箱: 1669625159@qq.com

雹的双偏振量,发现冰雹在下降过程中差分反射率因子和比差分相移由负值转为正值,表明冰雹在下降过程中随着温度升高逐渐由干冰雹融化为大雨滴或外包水膜的冰雹。徐舒扬等^[13]基于双偏振量优化了水凝物识别算法中的参数。虽然增加双线偏振功能可加强雷达对大冰雹以及冰雹状态的识别能力^[14],但目前除珠江三角洲、长江三角洲和京津冀等地以外,我国大部地区的短时临近预报业务仍主要以多普勒天气雷达为基础,鄂东地区亦是如此。

实际发生的冰雹中绝大多数为雷达回波三维特征并不典型的小冰雹,主要依靠人工雷达图像判别,效率较低且依赖大量当地统计结果和较为丰富的预报经验^[15-16]。另外,目前业务中使用较为广泛的单站雷达 ROSE2.0 和短时临近预报系统(server weather automatic nowcasting, SWAN)在采用新的冰雹指数后冰雹识别命中率有所提高,但虚警率也较高,达到 90%以上^[17-19]。

鉴于此,国内众多专家学者将具有解决分类问题和非线性建模能力的决策树算法引入冰雹识别中,效果良好。王萍等^[20]基于系统结构疏密性特征、移出率、液态水含量以及累积液态水含量等特征,利用决策树算法实现了对冰雹和短时强降水的分类识别,结果表明:该方法对雷达站 50 km 范围内由雷暴系统产生冰雹的命中率为 79.8%,虚警率为 3.5%,平均临界成功指数达 0.8。方德贤等^[21]在风暴跟踪识别算法的基础上发展风暴分类识别技术,结果表明:该技术能有效提高冰雹识别的命中率,降低虚警率。郑建琴等^[22]利用决策树算法对天津地区的历史冰雹个例进行识别,得到天津地区冰雹云的雷达回波判别指标,检验结果表明,该指标明显优于常用判别函数构建的方法。

上述研究的冰雹识别因子仅包括回波强度和回波的垂直伸展,未考虑特殊温度层对冰雹生成和融化的影响,而濮文耀等^[23]指出降落地面雹粒的大小取决于开始融化时冰雹的尺寸和融化层的高度,俞小鼎^[24]认为这个融化层接近于湿球 0℃ 层高度,同时指出简单且有效判别有无冰雹,特别是大冰雹的依据为强回波相对于湿球 0℃ 和湿球 -20℃ 的高度^[4]。因此,本文基于决策树算法利用鄂东地区的历史冰雹灾情资料和雷达资料,将湿球温度高度引入冰雹识别因子,开展基于人工智能的冰雹识别技术研究,并详细检验评估其识别能力,为鄂东地区冰雹天气的监测和预警提供客观参考,为科学防雹决

策提供技术支持。

1 资料和方法

1.1 资料

本文使用以下 3 种资料:①2015 年 1 月 1 日—2021 年 12 月 31 日湖北武汉站 S 波段多普勒天气雷达的基数据,并对反射率因子进行杂波抑制^[25];②2015 年 1 月 1 日—2021 年 12 月 31 日鄂东地区冰雹灾情资料,鄂东指包含武汉、孝感、黄冈、黄石、咸宁和鄂州 6 个地(市)的湖北东部;③2015 年 1 月 1 日—2021 年 12 月 31 日武汉国家基本气象观测站每日 08:00 和 20:00(北京时,下同)的常规高空探测资料。

1.2 冰雹个例及雷达特征量的选取

S 波段雷达扫描半径为 230 km,但对距离雷达中心小于 30 km 的区域无法扫描到风暴中上部,对距离雷达中心大于 150 km 的区域无法扫描到风暴底部^[19],因此本文所选用的冰雹个例和所有计算与提取的雷达特征量均位于距离雷达中心 30~150 km 范围内。此外,为保证冰雹发生时间和位置的准确性,本文只选取灾情记录详细、有图片和视频等相关材料佐证且冰雹发生位置与最近时刻雷达回波较为吻合的个例,共计 70 个。

结合文献^[26-29],同时考虑温度层结对冰雹生长与降落的影响,本文以组合反射率因子(R_C)、0.5° 仰角反射率因子($R_{0.5}$)、1.5° 仰角反射率因子($R_{1.5}$)和 2.4° 仰角反射率因子($R_{2.4}$)、垂直累积液态水含量(L)和 45 dBZ 回波的最大高度(H_{45})为冰雹的回波强度特征量,以湿球 0℃ 层高度(H_0)和湿球 -20℃ 层高度(H_{-20})为湿球温度高度特征量,将上述 2 种不同属性的 8 个特征量作为武汉地区冰雹识别的因子。雷达特征量全部采用降水模式(VCP21)下的极坐标格式(空间分辨率为 $1^\circ \times 1$ km),其中 45 dBZ 回波的最大高度指 45 dBZ 回波的最大伸展高度;湿球 0℃ 层高度和湿球 -20℃ 层高度则是根据冰雹发生前最邻近时刻的探空资料计算得到^[24,30]。

1.3 冰雹样本和非冰雹样本的提取

每个冰雹个例雷达特征量的提取过程:由于冰雹发生时刻与雷达观测时间大多数情况下并不同步,同时考虑到降雹点在灾情搜集以及经纬网格坐标和雷达极坐标转换过程中的偏差,需要进行时空

协同,因此本文选取距离冰雹发生时间最近的雷达观测时刻作为冰雹雷达特征量的提取时刻,并以离冰雹发生位置最近的像素点为中心,提取其周围 $3^{\circ}\times 3$ km范围内9个像素点上的特征量作为冰雹的特征值,即冰雹样本。共得到630个冰雹样本,每个冰雹样本为1组由同1个像素点上全部特征量数值所组成的数列。

由于强对流天气发生概率小^[31],且冰雹是对流风暴强烈发展的附属灾害,随机选取的非冰雹样本,绝大多数情况下是晴空回波或弱的降水回波,很难抽取到强对流风暴单体的回波,导致评分虚高。为尽可能抑制资料不平衡带来的不利影响,增强算法的业务实用性,本文根据冰雹样本和非冰雹样本1:1的原则,以组合反射率因子为标准,按照一定比例选取对应强度的非冰雹个例(选取标准和对应样本量见表1),并以此为基础提取对应特征量,共得到630个非冰雹样本。汇总上述冰雹和非冰雹样本,随机抽取其中的四分之三为训练样本,进行算法的训练和参数优化,剩余的四分之一为检验样本,用于检验评估算法的识别效果。

表1 非冰雹样本选取标准及对应的样本量
Table 1 Selection standard and corresponding quantity of non-hail samples

组合反射率因子/dBZ	样本量
40.0~44.9	63
45.0~49.9	158
50.0~54.9	188
55.0~59.9	158
≥ 60.0	63

1.4 检验方法

为了客观定量评估决策树算法对冰雹的识别能力,本文采用命中率、虚警率和临界成功指数3个指标量化评估冰雹识别结果,具体计算流程^[32]:对每个检验样本,将决策树算法对该样本的识别结果与其样本标签进行比较,若算法识别结果为有冰雹,样本标签为冰雹样本,记正确识别1次;若算法识别结果为有冰雹,而样本标签为非冰雹样本,记空报1次;若算法识别结果为非冰雹,样本标签为冰雹样本,记漏报1次;分别统计总的正确识别次数、总空报次数和总漏报次数,计算临界成功指数、命中率和虚警率。

2 决策树算法

决策树算法是一种运用概率与图论中的树对决

策中的不同方案进行比较,从而获得最优方案的风险型决策方法,是数据挖掘的一种重要方式^[33]。决策树是一种类似树结构的表示法,由节点、分支和叶子3部分组成,树中每个内部节点表示1个属性的测试,每个叶节点代表1个类,根节点是全部的开始,包含全部检验样本。决策树算法通俗易懂,分类效果良好,已广泛应用到农业气象以及对流天气识别中^[34]。本文所选用的分类回归树是一种通用的树生长算法,既可以作为分类树,也可以作为回归树,除了具备一般决策树算法的优点之外,还具有对目标变量和预测变量的概率分布无要求,对缺失值不敏感,计算量小,运算快等特点。

分类回归树算法通过计算基尼指数选定最佳特殊属性,并决定该属性的最佳二值切分点,具体流程:①根节点分裂,即对所有样本的全部特征量进行分割,并计算每组特征量及其分割值组合的信息增益,选取增益最大的特征量和分割值作为根节点及其阈值;②根节点分裂完成后,逐个迭代每个子节点,判别其成为叶子或需要继续分裂。若为叶子,则将该节点标注为叶子,并将该节点上样本最多的类别作为该叶子的类别(即冰雹或非冰雹),若需要继续分裂,则以该节点所包含的样本量作为本次分裂的总样本量,重复第①步,寻找最优特征量和分割值的组合,将该子节点一分为二,成为两个新的子节点;③不断迭代直到无节点需要分裂,最后形成一个树状判断结构,结构的终点均为叶子^[35]。

为防止算法过拟合,本文采取预剪枝策略以增强算法的鲁棒性,具体做法为控制决策树模型的深度,100组随机试验的结果表明:当特征量只包含回波强度时,深度为5的决策树算法效果最优;当特征量包含回波强度和湿球温度高度时,深度为7的决策树算法效果较佳。最后分别设定深度,并建立各自的决策树。

3 结果检验

3.1 整体评分

由决策树算法评分(表2)可见:仅有回波强度的决策树(强度决策树)的平均命中率为0.88,而包含回波强度与湿球温度高度的决策树算法(强度-高度决策树)的平均命中率为0.93,较强度决策树提高5.68%;强度决策树的虚警率为0.12,强度-高度决策树的虚警率为0.07,较强度决策树下降41.67%;

强度决策树的临界成功指数为 0.80,强度-高度决策树临界成功指数为 0.86,较强度决策树提高 7.5%。

由强度决策树(图 1)可知,根节点为组合反射率因子,表明强度决策树判断鄂东地区是否降雹的

表 2 不同决策树算法的评分

Table 2 Scores of different decision trees

决策树算法	命中率	虚警率	临界成功指数
强度决策树	0.88	0.12	0.80
强度-高度决策树	0.93	0.07	0.86

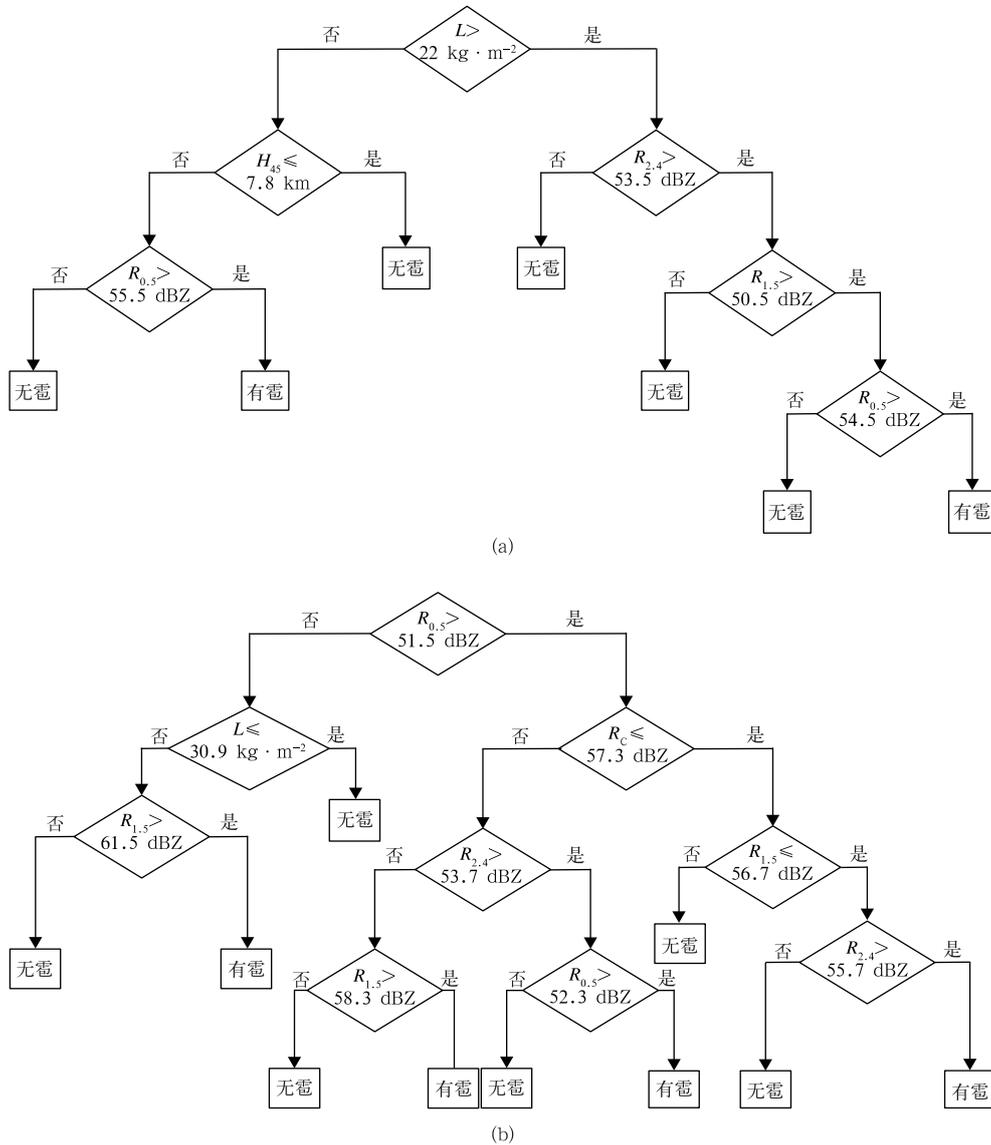


图 1 鄂东地区判别降雹的强度决策树

(a) $R_c \leq 56.7$ dBZ, (b) $R_c > 56.7$ dBZ

Fig. 1 Decision tree of intensity for hail in eastern Hubei

(a) $R_c \leq 56.7$ dBZ, (b) $R_c > 56.7$ dBZ

最重要因子是组合反射率因子,其前 3 位的分类规则如下:①当 $R_c \leq 56.7$ dBZ, $L \leq 22$ $\text{kg} \cdot \text{m}^{-2}$ 且 $H_{45} \leq 7.8$ km 时,该单体不会降雹;②当 $R_c > 57.3$ dBZ, $R_{0.5} > 51.5$ dBZ, $R_{2.4} \leq 53.7$ dBZ 且 $R_{1.5} > 58.3$ dBZ 时,该单体降雹;③当 $R_c > 57.3$ dBZ, $R_{0.5} \leq$

51.5 dBZ, $L > 30.9$ $\text{kg} \cdot \text{m}^{-2}$ 且 $R_{1.5} > 61.5$ dBZ 时,该单体降雹。

由图 2 可知,强度-高度决策树的根节点为 0.5° 仰角反射率因子,表明该算法将最接近地面的回波强度作为判别鄂东地区是否降雹的最主要因子,其

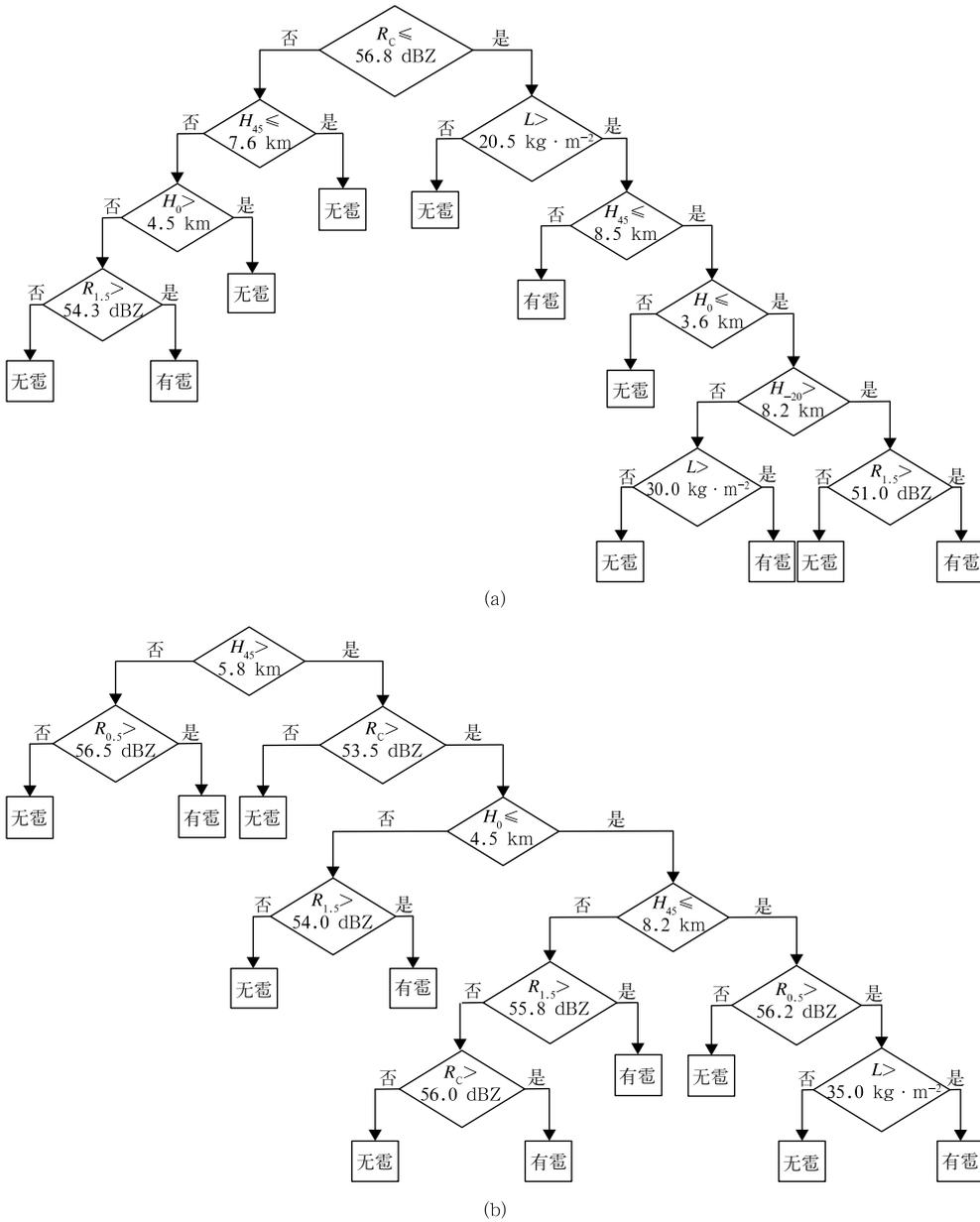


图 2 鄂东地区判别降雹的强度-高度决策树

(a) $R_{0.5} \leq 51.8 \text{ dBZ}$, (b) $R_{0.5} > 51.8 \text{ dBZ}$

Fig. 2 Decision tree of intensity-height for hail in eastern Hubei

(a) $R_{0.5} \leq 51.8 \text{ dBZ}$, (b) $R_{0.5} > 51.8 \text{ dBZ}$

前 3 位的分类规则如下: ①当 $R_{0.5} \leq 51.8 \text{ dBZ}$, $R_c \leq 56.8 \text{ dBZ}$, 且 $L \leq 20.5 \text{ kg} \cdot \text{m}^{-2}$ 时, 该单体不会降雹; ②当 $R_{0.5} > 56.2 \text{ dBZ}$, $8.2 \text{ km} > H_{45} > 5.8 \text{ km}$, $R_c > 53.5 \text{ dBZ}$, $H_0 \leq 4.5 \text{ km}$ 且 $L > 35 \text{ kg} \cdot \text{m}^{-2}$ 时, 该单体降雹; ③当 $R_{0.5} \leq 51.8 \text{ dBZ}$, $R_c > 56.8 \text{ dBZ}$, $H_{45} > 7.6 \text{ km}$, $H_0 \leq 4.5 \text{ km}$, 且 $R_{1.5} > 54.3 \text{ dBZ}$ 时, 则该单体降雹。

对比强度决策树和强度-高度决策树的结构可

知, 强度决策树的根节点为组合反射率因子, 最底层的叶节点多为 0.5° 和 1.5° 仰角反射率因子, 而强度-高度决策树的根节点为 0.5° 仰角反射率因子, 中间层多条路径均包含回波垂直伸展与湿球温度高度的对比, 尤其是与湿球 0°C 层高度的对比, 最底层多为风暴的整体属性(垂直累积液态水含量或组合反射率因子), 换言之, 当某个对流单体的组合反射率因子达到一定强度, 同时低层仰角的反射率因子也

超过一定阈值时,强度决策树大概率将其判别为降雹风暴,这种判断大多数情况下正确,但若此时环境大气的湿球 0℃层高度和湿球 -20℃层高度较高,考虑到强回波伸展高度不够或冰雹在下降过程中的融化作用,强度-高度决策树可能无法识别冰雹,从而使其虚警率明显低于强度决策树,这与俞小鼎等^[36-37]的研究成果一致,即强的低层反射率因子、高悬的强回波和适宜的融化层高度是判别地面降雹的关键因子。

3.2 个例分析

3.2.1 2020 年 3 月 26 日鄂州冰雹个例

2020 年 3 月 26 日午后至傍晚,受高空槽和中低层切变线的共同影响,鄂东地区出现春季典型强

对流天气,16:18 鄂州市鄂城区碧石渡镇黄咀村附近出现小冰雹(方位角为 115.1°,距离为 55.5 km),直径约为 0.5~1.5 cm,持续时间为 5~10 min。由武汉多普勒天气雷达的组合反射率因子和过降雹点的反射率因子剖面(图 3)可知,降雹点上空组合反射率因子为 62.0 dBZ,0.5°,1.5°和 2.4°仰角的反射率因子分别为 59.5,62.0 dBZ 和 57.0 dBZ,垂直累积液态水含量为 45 kg·m⁻²,45 dBZ 回波的最大伸展高度为 7.8 km。当日 08:00 武汉探空计算的湿球 0℃层和湿球 -20℃层高度分别为 3.3 km 和 6.4 km,基于上述回波强度和湿球温度高度的特征值,强度决策树和强度-高度决策树均成功识别出鄂州地区的冰雹(图 4),且两者识别冰雹的位置也近

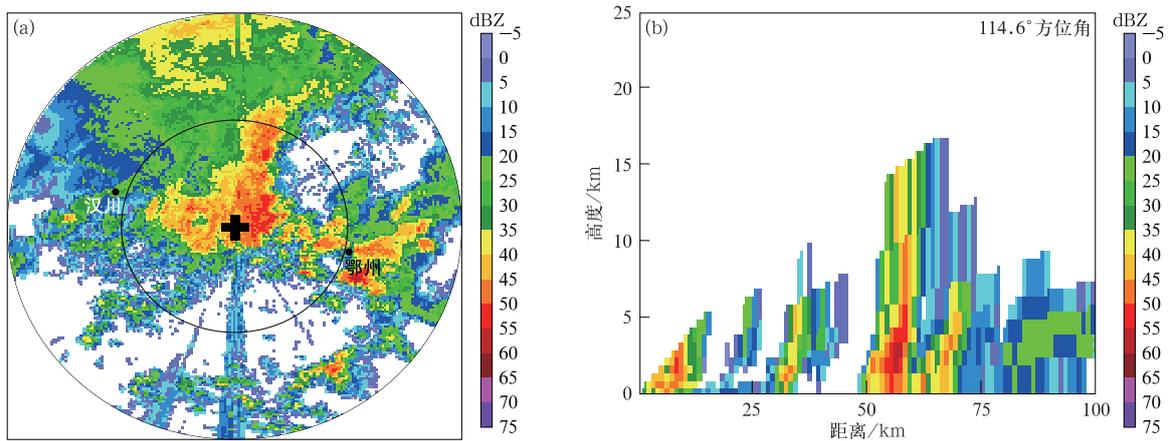


图 3 2020 年 3 月 26 日 16:18 武汉雷达组合反射率因子(相邻距离圈相距 50 km,下同)(a)和过降雹点的反射率因子剖面(b)

Fig. 3 Combined reflectivity(the distance between adjacent rings is 50 km,hereinafter)(a) and reflectivity factor profile(b) of hail point of Wuhan radar at 1618 BT 26 Mar 2020

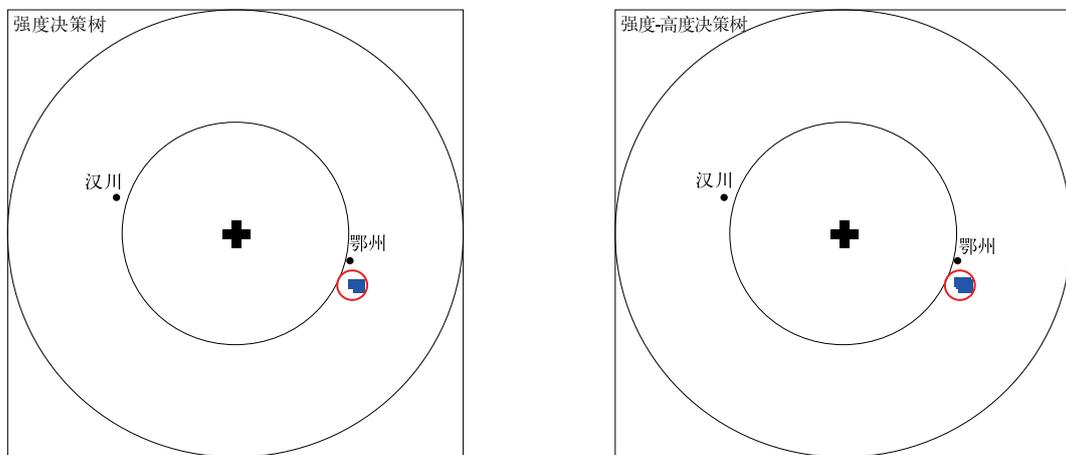


图 4 2020 年 3 月 26 日 16:18 决策树的识别结果(红色圆圈表示以降雹点为中心,半径为 5 km 的区域)

Fig. 4 Identification of decision tree at 1618 BT 26 Mar 2020 (red circle area denotes the hail location with a radius of 5 km)

乎一致,与观测的误差均在5 km 以内。

3.2.2 2021年5月14日黄冈冰雹个例

2021年5月14日午后至傍晚,受高空槽、中低层切变线和低层西南急流的共同影响,鄂东地区出现混合型强对流天气,16:12 黄冈市罗田县骆驼坳镇涂家湾村附近出现冰雹(方位角为76.7°,距离为102 km),直径约为1.5~3 cm,持续时间为10~15 min。由武汉多普勒天气雷达的组合反射率因子和过降雹点的反射率因子剖面(图5)可知,降雹点上空组合反射率因子为63.0 dBZ,0.5°,1.5°和2.4°仰角反射率因子分别为49.5,63.0 dBZ和58.0 dBZ,垂直累积液态水含量为70 kg·m⁻²,45 dBZ 回波的最大伸展高度为10.8 km。当日08:00 武汉探空计算

的湿球0℃层和湿球-20℃层高度分别为4.5 km和8.0 km。根据上述特征值,强度决策树和强度-高度决策树均成功识别出罗田附近的冰雹(图6),但在降雹风暴南侧浠水境内以及东南侧黄梅境内的回波均被错误识别为冰雹单体,强度-高度决策树比强度决策树少2个虚警点。

3.2.3 2021年9月28日武汉和孝感冰雹个例

2021年9月28日午后至傍晚,受高空槽和中低层切变线的共同影响,鄂东地区出现强对流天气,19:48 孝感汉川市马口镇枣树村(方位角为275.5°,距离为47 km)和武汉市新沟镇前港村同时出现小冰雹,直径约为0.5~1.0 cm,持续时间为5~10 min。从武汉多普勒天气雷达的组合反射率因子

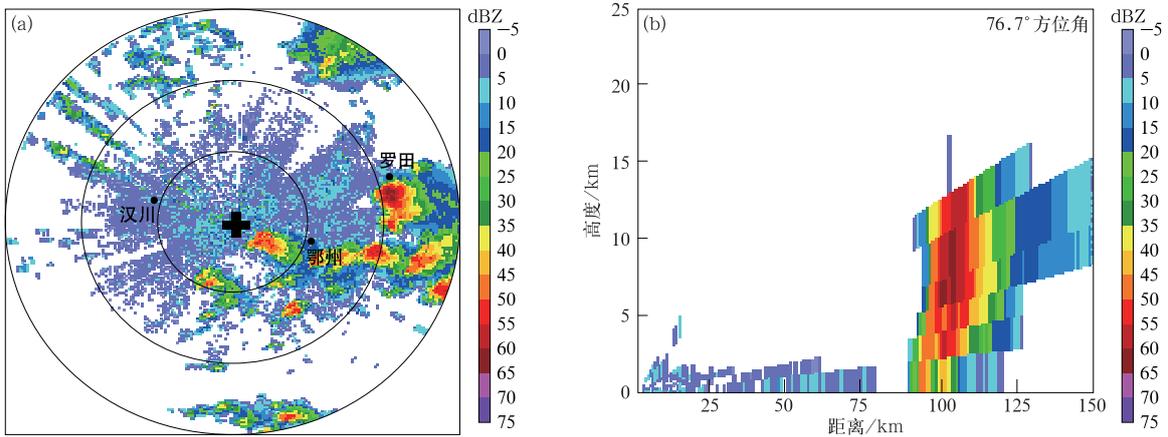


图5 2021年5月14日16:12 武汉雷达组合反射率因子(a)和过降雹点的反射率因子剖面(b)
Fig. 5 Combined reflectivity(a) and reflectivity factor profile(b) of hail point of Wuhan radar at 1612 BT 14 May 2021

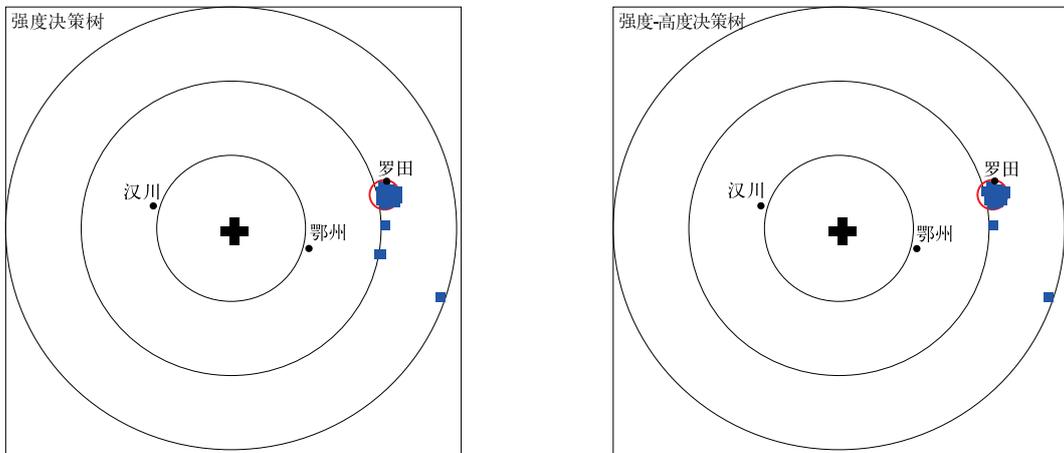


图6 2021年5月14日16:12 决策树识别结果(红色圆圈表示以降雹点为中心半径为5 km的区域)
Fig. 6 Identification of decision tree at 1612 BT 14 May 2021 (red circle area denotes the hail location with a radius of 5 km)

和过降雹点的反射率因子剖面(图 7)可知,汉川降雹点上空组合反射率因子为 62.0 dBZ,0.5°,1.5°和 2.4°仰角反射率因子分别为 55.0,54.0 dBZ 和 62.0 dBZ,垂直累积液态水含量为 47.5 kg·m⁻²,45 dBZ 回波的最大伸展高度为 11.8 km。当日 08:00 武汉探空计算的湿球 0℃层和湿球-20℃层高度分别为 4.3 km

和 7.8 km。根据上述特征值强度,决策树和强度-高度决策树均成功识别出汉川和武汉附近两个冰雹落区(图 8),但强度决策树算法将汉川西南侧的回波错误识别为冰雹单体,较强度-高度决策树多 1 个虚警点。

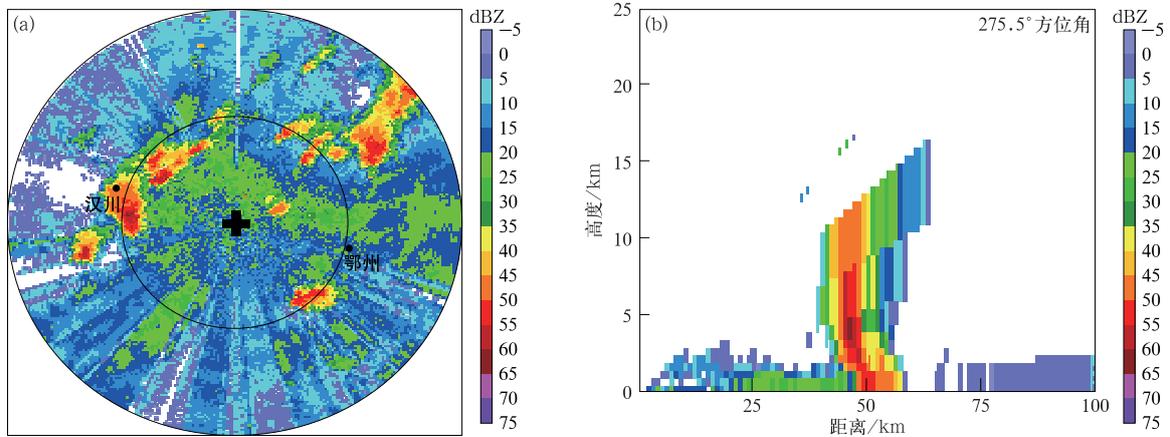


图 7 2021 年 9 月 28 日 19:48 武汉雷达组合反射率因子(a)和过汉川降雹点反射率因子剖面(b)

Fig. 7 Combined reflectivity(a) and reflectivity factor profile(b) of Hanchuan hail point of Wuhan radar at 1948 BT 28 Sep 2021

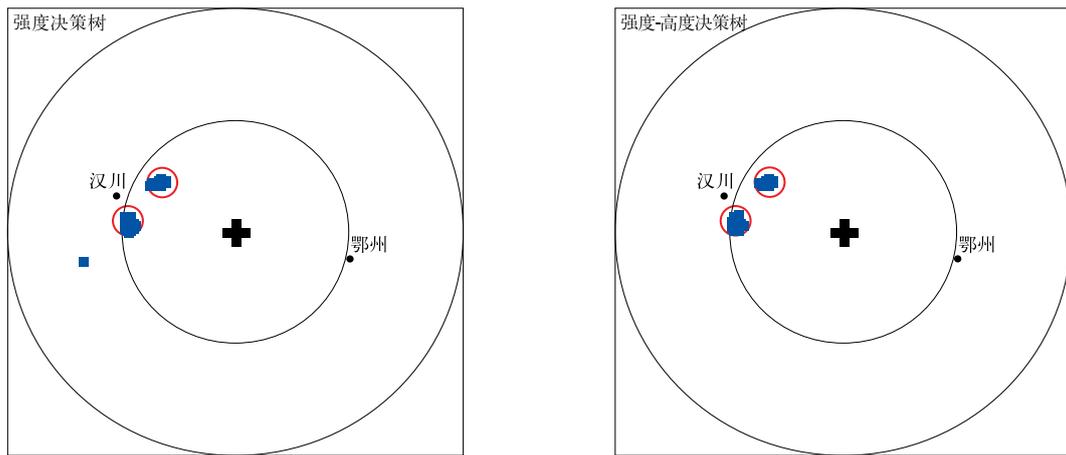


图 8 2021 年 9 月 28 日 19:48 决策树识别结果(红色圆圈表示以降雹点为中心半径为 5 km 的区域)

Fig. 8 Identification of decision tree at 1948 BT 28 Sep 2021 (red circle area denotes the hail location with a radius of 5 km)

4 结论与讨论

本文利用 2015 年 1 月 1 日—2021 年 12 月 31 日鄂东地区的冰雹灾情资料、武汉多普勒天气雷达基数据以及高空探空资料,将湿球温度高度引入冰

雹识别的因子中,并采用较为严格的非冰雹样本选取标准,基于决策树算法开展人工智能技术在冰雹识别中的应用研究,并根据命中率、虚警率和临界成功指数 3 个指标检验评估其识别能力,得到以下主要结论:

- 1) 强度决策树的命中率为 0.88,虚警率为 0.12,

临界成功指数为 0.8;强度-高度决策树的命中率、虚警率和临界成功指数分别为 0.93,0.07 和 0.86,其中命中率和临界成功指数较强度决策树分别提高 5.68%和 7.5%,虚警率降低 41.67%。

2) 由强度决策树和强度-高度决策树的结构可知,强度决策树的根节点为组合反射率因子,最底层的叶节点多为 0.5° 和 1.5° 仰角反射率因子,而强度-高度决策树的根节点为 0.5° 仰角反射率因子,中间层中多条路径均包含回波强度垂直伸展与特殊湿球温度高度的对比,尤其是与湿球 0°C 层高度的对比,最底层多为风暴的整体属性。

3) 由个例分析可知,当湿球 0°C 层高度较低时,强度决策树和强度-高度决策树的识别结果近乎一致,均能有效识别降雹落区;当湿球温度高度较高,尤其湿球 0°C 层高度较高时,强度-高度决策树的虚警次数减少,虚警率明显降低。

决策树算法硬件要求低、移植方便,在已部署双偏振雷达的台站可结合回波强度和差分反射率、相关系数等偏振信息获取多维度的冰雹特征量阈值,进一步提高冰雹的识别能力,相关工作已在华南地区开展,并取得良好效果^[38]。冰雹个例的数量和质量是影响算法性能的关键因素,下一步将丰富冰雹样本量、优化样本扩充技术,不断提升资料的科学性和应用价值,同时尝试与多种机器学习算法进行融合,建立多算法融合识别模型,提升业务系统算法的多样性^[39-40]。由于雷达探测的局限性影响远距离(大于 150 km)和近距离(小于 30 km)冰雹识别的准确性,后期将按距离建立不同的决策树算法,提高算法的业务适用性。冰雹是对流性天气强烈发展的附属灾害,机器学习算法对复杂的物理机制缺乏解释能力,而物理机制的深入研究是统计分析的重要基础,未来随着冰雹,特别是直径为 5~20 mm 的小冰雹发生机理的不断研究以及机器学习算法的不断迭代升级,利用机器学习算法识别冰雹将会取得更有意义的成果^[35]。

参考文献

- [1] 郑永光,周康辉,盛杰,等.强对流天气监测预报预警技术进展.应用气象学报,2015,26(6):641-657.
Zheng Y G,Zhou K H,Sheng J,et al. Advance in techniques of monitoring, forecasting and warning of severe convective weather. *J Appl Meteor Sci*, 2015, 26(6): 641-657.
- [2] 徐双柱,韦慧红.关于强对流天气的几点思考.暴雨灾害,2016,35(3):197-202.
- Xu S Z,Wei H H. Some thoughts on the weather forecast of severe convective storms. *Torrential Rain Disaste*, 2016, 35(3):197-202.
- [3] 施望芝,金琪,郭施,等.湖北一次冰雹天气过程的落区诊断分析和预报.热带气象学报,2004,20(2):212-217.
Shi W Z,Jin Q,Guo S,et al. An analysis and forecast for the area of an hail weather in Hubei Province. *J Trop Meteor*, 2004, 20(2): 212-217.
- [4] 俞小鼎,王迎春,陈明轩,等.新一代天气雷达与强对流天气预警.高原气象,2005,24(3):456-464.
Yu X D,Wang Y C,Chen M X,et al. Severe convective weather warnings and its improvement with the introduction of the NEXRAD. *Plateau Meteor*, 2005, 24(3): 456-464.
- [5] 刁秀广,朱君鉴,黄秀韶,等.VIL和VIL密度在冰雹云判据中的应用.高原气象,2008,27(5):1131-1139.
Diao X G,Zhu J J,Huang X S,et al. Application of VIL and VIL density in warning criteria for hailstorm. *Plateau Meteor*, 2008, 27(5): 1131-1139.
- [6] 胡胜,罗聪,张羽,等.广东大冰雹风暴单体的多普勒天气雷达特征.应用气象学报,2015,26(1):57-65.
Hu S,Luo C,Zhang Y,et al. Doppler radar features of severe hailstorm in Guangdong Province. *J Appl Meteor Sci*, 2015, 26(1): 57-65.
- [7] 王萍,潘跃.基于显著性特征的大冰雹识别模型.物理学报,2013,62(6):515-524.
Wang P,Pan Y. Severe hail identification model based on saliency characteristics. *Acta Physica Sinica*, 2013, 62(6): 515-524.
- [8] 王莎,沙勇,宋金妹,等.冀东地区冰雹云多普勒雷达参数特征分析.气象,2019,45(5):713-722.
Wang S,Sha Y,Song J M,et al. Characteristic analysis of hail cloud Doppler radar parameters in the eastern Hebei Province. *Meteor Mon*, 2019, 45(5): 713-722.
- [9] 吴剑坤,陈明轩,秦睿,等.变分回波跟踪算法及其在对流临近预报中的应用研究.气象学报,2019,77(6):999-1014.
Wu J K,Chen M X,Qin R,et al. The variational echo tracking method and its application in convective storm nowcasting. *Acta Meteor Sinica*, 2019, 77(6): 999-1014.
- [10] 王洪,吴乃庚,万齐林,等.变分回波跟踪算法及其在对流临近预报中的应用研究.气象学报,2018,76(1):92-103.
Wang H,Wu N G,Wan Q L,et al. Analysis of S-band polarimetric radar observations of a hail-producing supercell. *Acta Meteor Sinica*, 2018, 76(1): 92-103.
- [11] 冯晋勤,张深寿,吴陈锋,等.双偏振雷达产品在福建强对流天气过程中的应用分析.气象,2018,44(12):1565-1574.
Feng J Q,Zhang S S,Wu C F,et al. Application of dual polarization weather radar products to severe convective weather in Fujian. *Meteor Mon*, 2018, 44(12): 1565-1574.
- [12] 刁秀广,李芳,万夫敬.两次强冰雹超级单体风暴双偏振特征对比.应用气象学报,2022,33(4):414-428.
Diao X G,Li F,Wan F J. Comparative analysis on dual polarimetric radar observations of two severe hail-producing supercells. *J Appl Meteor Sci*, 2022, 33(4): 414-428.

- zation features of two severe hail supercells. *J Appl Meteor Sci*, 2022, 33(4): 414-428.
- [13] 徐舒扬, 吴翀, 刘黎平. 双偏振雷达水凝物相态识别算法的参数改进. *应用气象学报*, 2020, 31(3): 350-360.
Xu S Y, Wu C, Liu L P. Parameter improvements of hydrometeor classification algorithm for the dual-polarimetric radar. *J Appl Meteor Sci*, 2020, 31(3): 350-360.
- [14] 俞小鼎, 郑永光. 中国当代强对流天气研究与业务进展. *气象学报*, 2020, 78(3): 391-418.
Yu X D, Zheng Y G. Advances in severe convective weather research and operational service in China. *Acta Meteor Sinica*, 2020, 78(3): 391-418.
- [15] 王婷波, 周康辉, 郑永光. 我国中东部雷暴活动特征分析. *气象*, 2020, 46(2): 189-199.
Wang T B, Zhou K H, Zheng Y G. Statistic analysis of thunderstorm characteristics in central and eastern China. *Meteor Mon*, 2020, 46(2): 189-199.
- [16] 周康辉, 郑永光, 韩雷, 等. 机器学习在强对流监测预报中的应用进展. *气象*, 2021, 47(3): 274-289.
Zhou K H, Zheng Y G, Han L, et al. Advances in application of machine learning to severe weather monitoring and forecasting. *Meteor Mon*, 2021, 47(3): 274-289.
- [17] 王瑾, 刘黎平. WSR-88D 冰雹探测算法在贵州地区的评估检验. *应用气象学报*, 2011, 22(1): 96-106.
Wang J, Liu L P. The evaluation of WRS-88D hail detection algorithm over Guizhou Region. *J Appl Meteor Sci*, 2011, 22(1): 96-106.
- [18] 周康辉. 基于卷积神经网络的强对流天气预报方法研究. 北京: 中国科学院大学, 2021.
Zhou K H. Convective Weather Forecasting with Convolutional Neural Networks. Beijing: University of Chinese Academy of Sciences, 2021.
- [19] 刘新伟, 蒋盈沙, 黄武斌, 等. 基于雷达产品和随机森林算法的冰雹天气分类识别及预报. *高原气象*, 2021, 40(4): 898-908.
Liu X W, Jiang Y S, Huang W B, et al. Classified identification and nowcast of hail weather based on radar products and random forest algorithm. *Plateau Meteor*, 2021, 40(4): 898-908.
- [20] 王萍, 高毅, 李聪. 50 km 以内雷暴系统的分类识别方法研究. *气象*, 2016, 42(2): 230-237.
Wang P, Gao Y, Li C. Method study of classification and recognition of thunderstorm system less than 50 km. *Meteor Mon*, 2016, 42(2): 230-237.
- [21] 方德贤, 李红斌, 董新宁, 等. 风暴分类识别技术在人工防雹中的应用. *气象*, 2016, 42(9): 1124-1134.
Fang D X, Li H B, Dong X N, et al. Application of storm auto-classification technology in artificial hail prevention. *Meteor Mon*, 2016, 42(9): 1124-1134.
- [22] 郑建琴, 路明月, 王曙东, 等. 基于决策树的天津地区冰雹天气雷达因子分析. *气象科技*, 2017, 45(2): 349-354.
Zheng J Q, Lu M Y, Wang S D, et al. Analysis of hail weather based on decision-making tree using radar data in Tianjin. *Meteor Sci Technol*, 2017, 45(2): 349-354.
- [23] 濮文耀, 李红斌, 宋煜, 等. 0℃层高度的变化对冰雹融化影响的分析与应用. *气象*, 2015, 41(8): 980-985.
Pu W Y, Li H B, Song Y, et al. Analysis and application of the effect of 0℃ layer height on melting hail. *Meteor Mon*, 2015, 41(8): 980-985.
- [24] 俞小鼎. 关于冰雹的融化层高度. *气象*, 2014, 40(6): 649-654.
Yu X D. A note on the melting level of hail. *Meteor Mon*, 2014, 40(6): 649-654.
- [25] Steiner M, Smith J. Use of three-dimensional reflectivity structure for automated detection and removal of non-precipitating echoes in radar data. *J Atmos Ocean Technol*, 2014, 40(6): 649-654.
- [26] 刘伯骏, 张亚萍, 黎中菊, 等. 一种基于地面实况的降雹风暴体客观标识方法. *应用气象学报*, 2021, 32(1): 78-90.
Liu B J, Zhang Y P, Li Z J, et al. An objective hailstorm labeling algorithm based on ground observation. *J Appl Meteor Sci*, 2021, 32(1): 78-90.
- [27] 王研峰, 黄武斌, 王聚杰, 等. 一次甘肃天水强冰雹的雷达回波特征及成因分析. *高原气象*, 2019, 38(2): 368-376.
Wang Y F, Huang W B, Wang J J, et al. Analysis on the characteristic of radar echo and the causes of a strong hail in Tianshui City of Gansu Province. *Plateau Meteor*, 2019, 38(2): 368-376.
- [28] 郭欣, 郭学良, 陈宝君, 等. 一次大冰雹形成机制的数值模拟. *应用气象学报*, 2019, 30(6): 651-664.
Guo X, Guo X L, Chen B J, et al. Numerical simulation on the formation of large-size hailstone. *J Appl Meteor Sci*, 2019, 30(6): 651-664.
- [29] 张曦, 黄兴友, 刘新安, 等. 北京大兴国际机场相控阵雷达强对流天气监测. *应用气象学报*, 2022, 33(2): 192-204.
Zhang X, Huang X Y, Liu X A, et al. The hazardous convective storm monitoring of phased-array antenna radar at Daxing International Airport of Beijing. *J Appl Meteor Sci*, 2022, 33(2): 192-204.
- [30] Knox J A, Nevius D S, Knox P N. Two simple and accurate approximations for wet-bulb temperature in moist conditions with forecasting applications. *Bull Amer Meteor Soc*, 2017, 98(9): 1897-1906.
- [31] 修媛媛, 韩雷, 冯海磊. 基于机器学习方法的强对流天气识别研究. *电子设计工程*, 2016, 24(9): 4-7.
Xiu Y Y, Han L, Feng H L. The identification of strong convective weather based on machine learning methods. *Electronic Design Engineering*, 2016, 24(9): 4-7.
- [32] 陈训来, 刘军, 郑群峰, 等. 基于卷积门控循环单元神经网络的临近预报方法研究. *高原气象*, 2020, 40(2): 411-423.
Chen X L, Liu J, Zheng Q F, et al. A study on radar echo nowcasting based on convolutional gated recurrent unit neural network. *Plateau Meteor*, 2020, 40(2): 411-423.
- [33] 章国材. 强对流天气分析与预报. 北京: 气象出版社, 2011.

- Zhang G C. Analysis and Forecast for Severe Convective Weather. Beijing: China Meteorological Press, 2011.
- [34] 李颖, 陈怀亮. 机器学习技术在现代农业气象中的应用. 应用气象学报, 2020, 31(3): 257-266.
- Li Y, Chen H L. Review of machine learning approaches for modern agrometeorology. *J Appl Meteor Sci*, 2020, 31(3): 257-266.
- [35] 史达伟, 张静, 曹庆, 等. 基于决策树算法的海州湾地区海雾预测. 气象科学, 2022, 42(1): 136-142.
- Shi D W, Zhang J, Cao Q, et al. Research on sea fog diagnosis in Haizhou Bay based on decision tree algorithm. *J Meteor Sci*, 2022, 42(1): 136-142.
- [36] 俞小鼎, 周小刚, 王秀明, 等. 雷暴与强对流临近天气预报技术进展. 气象学报, 2012, 70(3): 311-337.
- Yu X D, Zhou X G, Wang X M, et al. The advances in the nowcasting techniques on thunderstorms and severe convection. *Acta Meteor Sinica*, 2012, 70(3): 311-337.
- [37] 俞小鼎, 王秀明, 李万莉, 等. 雷暴与强对流临近预报. 北京: 气象出版社, 2020.
- Yu X D, Wang X M, Li W L, et al. Thunderstorm and Strong Convection Nowcasting. Beijing: China Meteorological Press, 2020.
- [38] 李博勇, 胡志群, 郑佳锋, 等. 利用贝叶斯方法改进华南地区冰雹识别效果. 热带气象学报, 2021, 37(1): 112-125.
- Li B Y, Hu Z Q, Zheng J F, et al. Using Bayesian method to improve hail identification in South China. *J Trop Meteor*, 2021, 37(1): 112-125.
- [39] 刘海知, 徐辉, 包红军, 等. 机器学习分类算法在降雨型滑坡预报中的应用. 应用气象学报, 2022, 33(3): 282-292.
- Liu H Z, Xu H, Bao H J, et al. Application of machine learning classification algorithm to precipitation landslides forecasting. *J Appl Meteor Sci*, 2022, 33(3): 282-292.
- [40] 刘娜, 熊安元, 张强, 等. 强对流天气人工智能应用训练基础数据集构造. 应用气象学报, 2021, 32(5): 530-541.
- Liu N, Xiong A Y, Zhang Q, et al. Development of basic dataset of severe convective weather for artificial intelligence training. *J Appl Meteor Sci*, 2021, 32(5): 530-541.

Hail Identification Technology in Eastern Hubei Based on Decision Tree Algorithm

Yuan Kai Li Wujie Pang Jing

(Wuhan Meteorological Observatory of Hubei Province, Wuhan 430040)

Abstract

Hail refers to the solid precipitation with a diameter greater than or equal to 5 mm caused by convection. Hail is also one main disastrous weather phenomenon in eastern Hubei, while Doppler weather radar is the most favorable tool for hail identification. At present, there are two hail identification methods used in the actual operation in eastern Hubei, one is artificial conceptual model, the other is the self-contained identification technology in short-time and proximity prediction system. The conceptual model needs to be judged by human, which is too subjective and threshold values of radar echo characteristic are not clear, while the false alarm rate(FAR) of existing automatic technologies in prediction system are too high. To overcome the shortcoming of the above methods, feasibilities of machine learning algorithms for hail identification are explored and a decision tree algorithm is established. Based on hail disaster data of Wuhan, Huanggang, Huangshi, Ezhou, Xianning and Xiaogan, Doppler weather radar data and convention high altitude sounding data of Wuhan from 2015 to 2021, the height of wet bulb 0°C (HWB0) and the height of wet bulb -20°C (HWB-20) are introduced into the hail identification factors, and artificial intelligence technology is applied in hail recognition. The performance is evaluated according to probability of detection (POD), FAR and critical success index(CSI). The result shows that both the decision tree algorithm with radar echo intensity (intensity decision tree) and the decision tree algorithm with radar echo intensity and wet bulb temperature height (intensity-height decision tree) can identify hail effectively. The POD results of the two decision tree algorithms are higher than 0.88, while the FAR are lower than 0.12, and the CSI are higher than 0.8, but the intensity-height decision tree performs better, with the POD and CSI increased by 5.68% and 7.5% than intensity decision tree respectively, while the FAR decrease 41.67%. The key factor of hail recognition by intensity decision tree is the combined reflectivity factor, and the bottom layer is the reflectivity factor of 0.5° and 1.5° elevation. The key factor of intensity-height decision tree is the reflectivity factor of 0.5° elevation and the judgment modules of radar echo extension height with the height of wet bulb temperature, especially with HWB0 included in the middle, and the bottom layer is the strength attributes of storm (vertically integrated liquid water and combined reflectivity). The analysis results of three cases with different occurrence time, location and hail size show that, due to the introduction of height of wet bulb temperature, the intensity-height decision tree reduces the number of empty alarm when the height of HWB0 and HWB-20 are high, especially when the HWB0 is high, thus reduces its FAR and improves its CSI, which indicate its potential wide prospect for operational application.

Key words: hail; decision trees; height of wet bulb temperature; examination and evaluation