

# 基于极限梯度提升和探地雷达时频特征的水泥路面脱空识别

张 军<sup>1,2</sup>, 姜文涛<sup>1,2</sup>, 张 云<sup>3,4</sup>, 罗婷倚<sup>3</sup>, 余秋琴<sup>3</sup>, 杨 哲<sup>1,2</sup>

(1. 长安大学公路养护装备国家工程实验室, 陕西 西安 710064; 2. 长安大学道路施工技术与装备教育部重点实验室, 陕西 西安 710064;  
3. 广西北投公路建设投资集团有限公司, 广西 南宁 530028; 4. 广西交科集团有限公司, 广西 南宁 530007)

**摘要:** 针对探地雷达(GPR)数据解译依赖于人工经验, 存在费时费力和主观偏差的问题, 提出了基于极限梯度提升(XGBoost)和GPR时频特征的水泥路面脱空识别方法。采用正演模拟、室内试验和现场试验获得了脱空病害数据源, 建立含有标签的脱空GPR数据集; 通过重采样方法统一GPR数据采样频率, 并对预处理后的GPR数据进行时频域特征提取, 建立了包含18个时域和12个频域特征的数据集。以时频域特征为输入, 是否存在脱空病害为输出, 采用XGBoost算法构建脱空识别模型, 并与随机森林(RF)和人工神经网络(ANN)算法进行对比。结果表明, 模型的识别准确率排序为 XGBoost (98.10%) > ANN (95.10%) > RF (93.17%), XGBoost 模型识别精度最高, 并能在实际路面上准确定位脱空区域。

**关键词:** 道路养护; 探地雷达(GPR); 脱空病害; 极限梯度提升(XGBoost); 时频域特征

中图分类号: U418

文献标志码: A

## Cement Pavement Void Identification Based on XGBoost and GPR Time-frequency Features

ZHANG Jun<sup>1,2</sup>, JIANG Wentao<sup>1,2</sup>, ZHANG Yun<sup>3,4</sup>, LUO Tingyi<sup>3</sup>, YU Qiuqin<sup>3</sup>, YANG Zhe<sup>1,2</sup>

(1. National Engineering Laboratory of Highway Maintenance Equipment, Chang'an University, Xi'an 710064, China; 2. Key Laboratory of Road Construction Technology and Equipment of the Ministry of Education, Chang'an University, Xi'an 710064, China; 3. Guangxi Beitou Highway Construction Investment Group Co., Ltd., Nanning 530028, China; 4. Guangxi Transportation Science and Technology Group Co., Ltd., Nanning 530007, China)

**Abstract:** Ground penetrating radar (GPR) is an effective method of void detection, but GPR data

interpretation depends on human experience, being time consuming and laborious, or even existing subjective bias. To address above issues, a cement pavement void identification method based on XGBoost and GPR time-frequency features was proposed. To automatically identify cement pavement void area, the finite difference time domain method, lab and field tests were carried out, and GPR void dataset with label was created. Then, the resampling method was used to obtain the same sample frequency. Thirty time and frequency domain features, including 18 time-domain features and 12 frequency-domain features, were extracted from the post-processed GPR data. Taking the time-frequency domain feature as an input, and void label as an output, XGBoost was used to build a void identification model. The random forest (RF) and artificial neural network (ANN) were also trained to compare with XGBoost. The comparison results indicate that the accuracy ranking is XGBoost (98.10%) > ANN (95.10%) > RF (93.17%). The accuracy of the XGBoost method is the highest and verified by field tests.

**Keywords:** road maintenance; ground penetrating radar (GPR); void disease; extreme gradient boosting (XGBoost); time-frequency domain features

水泥路面具有强度高、耐久性好的特点, 但存在水泥路面层与基层刚度不匹配问题。在温度和车辆荷载的重复作用下, 面层底部出现脱空区域并逐步扩展, 最终导致路面出现断板。脱空区域的横向尺寸是影响路面结构承载力的重要参数<sup>[1]</sup>, 因此准确识别水泥路面脱空的横向尺寸对水泥路面的养护工作具有指导意义。

收稿日期: 2022-07-08

基金项目: 广西省交通运输行业重点科技项目(19-09); 陕西省自然科学基金基础研究计划(2022JM-249); 陕西省交通厅项目(20-30X)

第一作者: 张 军, 副教授, 工学博士, 主要研究方向为路面无损检测。E-mail: zhangjun@chd.edu.cn



论文  
拓展  
介绍

探地雷达(GPR)是利用高频电磁波在介质中的反射和散射来实现浅层成像和定位的高分辨率深层无损探测技术,通过路面电磁特性变化定性或定量地识别地下目标体,因此被广泛应用于地下空洞<sup>[2]</sup>、管线<sup>[3-4]</sup>和路面病害<sup>[5]</sup>的定位与识别。路面病害的识别集中于基于B-scan的图谱分析和基于A-scan的机器学习。图谱分析法是借助深度迁移学习对雷达B-scan图谱中的病害区域进行识别<sup>[6-8]</sup>,已用于传力杆<sup>[9-10]</sup>、管道<sup>[11]</sup>、裂缝<sup>[12]</sup>和其他目标的智能识别<sup>[13]</sup>。王辉等<sup>[14]</sup>设计了级联结构的卷积神经网络,用于双曲线目标分类。针对脱空区域的图谱特征识别,笔者采用浅层混合网络Resnet18-YOLOv2实现了机场跑道空洞区域的定位<sup>[6]</sup>。然而,深度迁移学习需要大量准确标记的样本图谱和大量取芯验证,加上路面病害区域无固定的宽高比,难以在图谱上准确标注病害区域,导致病害区域的边界定位存在误差,因此需要研究病害边界的高精度识别方法。

相比图谱识别,基于A-scan信号特征的病害识别在横向尺寸上获得更高的精度,目前常见的建模方法有支持向量机(SVM)<sup>[15-16]</sup>、人工神经网络(ANN)<sup>[7]</sup>、随机森林(RF)<sup>[17]</sup>和极限梯度提升(XGBoost)等方法<sup>[18-19]</sup>。覃晖等<sup>[20]</sup>将GPR信号分段并提取方差、标准绝对偏差和四阶矩等3个特征,对隧道衬砌空洞病害的识别准确率为93.56%,但时域特征抗强干扰能力较差。为提高模型的抗干扰能力,周辉林等<sup>[21]</sup>采用2个时域信号特征(幅值、平均绝对偏差)和4个小波近似系数作为SVM输入建模,公路路基病害检测识别的准确率为92.7%,但未明确病害类型。为获得路面病害的敏感特征参数,笔者借鉴旋转设备的故障识别方法,用28个时频特征表征病害,通过主成分分析(PCA)降维和ANN构建了沥青路面水损害识别模型,识别准确率达到92.4%<sup>[7]</sup>。杜豫川等<sup>[19]</sup>将每道A-scan数据划分为 $p$ 段,分别提取能量、方差、偏度和对数功率谱等特征值,并将特征值输入XGBoost用于模型训练,对路面脱空病害的识别准确率为96%,但段数 $p$ 需要根据经验设置。现有研究为脱空病害识别提供了参考,但仍存在以下问题:①未考虑不同的GPR设备天线频率和采样频率之间存在的差异;②模型的泛化能力不强,需要人工设置合理的模型参数;③模型中病害区域的数据样本量小。因此,还需要研究高效、高精度的脱空病害识别方法。

以水泥路面脱空病害为研究对象,采用正演模拟、室内模型和现场试验相结合的方法建立了标准

脱空数据集,其中包含10 732条脱空数据和10 251条正常数据。针对GPR天线频率和采样参数不一致的问题,提出用重采样和标准化方法统一数据采样信息和幅值范围。在时频特征提取方法<sup>[22]</sup>的基础上,增加了平均能量和偏度<sup>[19]</sup>2个时域特征,将脱空数据集进行特征处理,构建了30个时频特征数据集。采用XGBoost算法建立脱空识别模型,并通过取芯验证了模型的准确性。

## 1 材料与方法

### 1.1 水泥路面脱空病害

在实际服役中水泥路面受到温度荷载和车辆荷载的耦合作用,出现由板底塑性变形导致的板底脱空,最终形成断板。板底脱空包括充气脱空和含水脱空,充气脱空是板底脱空的早期形态。图1展示了板底脱空的部位。随着时间的推移,脱空尺寸逐渐扩大,雨水渗透后形成含水脱空。

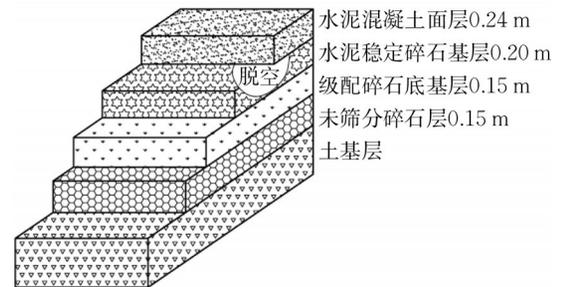


图1 水泥路面板底脱空

Fig.1 Void under cement pavement slab

### 1.2 脱空数据采集

以板底充气脱空为例,简化充气脱空的形状,使用16~300 mm不等的矩形脱空和圆形脱空进行室内试验和正演模拟,并综合现场试验数据构建脱空数据集。正演模拟和室内试验提供标准脱空特征,并为现场判断脱空提供依据,现场试验数据用于模型训练和验证模型的准确性。脱空数据集构成如图2所示。

#### 1.2.1 正演模拟

gprMax正演模拟是以时域有限差分为基础的雷达信号仿真算法,被广泛应用于GPR的病害模拟,以获得病害区域的信号特征<sup>[23-24]</sup>。构建如图3所示的脱空模型,C1~C4为圆形脱空,R1~R4为矩形脱空。根据实际路面的材料特性,设置脱空、水泥混凝土和半刚性基层的介电常数分别为1、7和9,水泥混凝土厚度为240 mm,仿真参数如表1所示。采用gprMax 3.1.5进行正演模拟,获得如图4所示的结

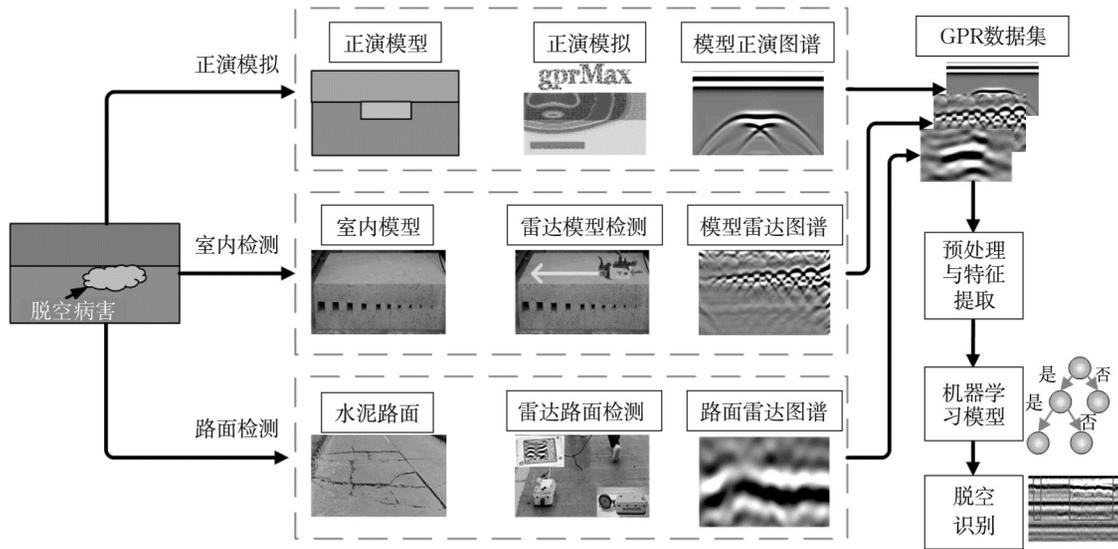


图2 脱空数据集构建

Fig.2 Void dataset construction

果,脱空病害区域均表现为双曲线特征,其宽度与病害横向尺寸相关。

表 1 gprMax 仿真模型参数

Tab.1 Parameters of gprMax simulation model

参数	数值
模型尺寸/(mm×mm)	1 000×5 100
线中心频率/MHz	800/900
网格尺寸/mm	$\Delta x = \Delta y = \Delta z = 2$
时窗/ns	16
采样频率/GHz	42.66
天线间距/m	0.14
激励源类型	Ricker

1.2.2 室内试验

从仿真结果可知,脱空形状并不影响脱空特征。为此,室内模型仅考虑不同尺寸的矩形脱空,构建了横向尺寸分别为 100、90、80、70、60、50、40、30、25、

16 mm 的 10 个矩形脱空区域 A1~A10,脱空区域的中心深度和中心间隔分别为 200 mm 和 150 mm。模型材料为型号 C30 的水泥混凝土,尺寸为 2 070 mm × 400 mm,如图 5 所示。意大利 IDS 公司的 RIS 型 900 MHz 和美国 US 公司的 1 000 MHz 天线沿着如图 5 所示的雷达测线方向对模型中的脱空区域进行多次数据采集,具体雷达参数设置如表 2 所示。获得的雷达图谱如图 6 所示,分别为 RIS 和 US 雷达实测剖面图谱。

探地雷达的纵向分辨率与雷达天线的频率有关,具体关系如下所示:

$$\Delta r = \frac{\lambda_c}{4} \tag{1}$$

式中: $\Delta r$ 为探地雷达的纵向分辨率; $\lambda_c$ 为探地雷达波在介质中传播的波长。

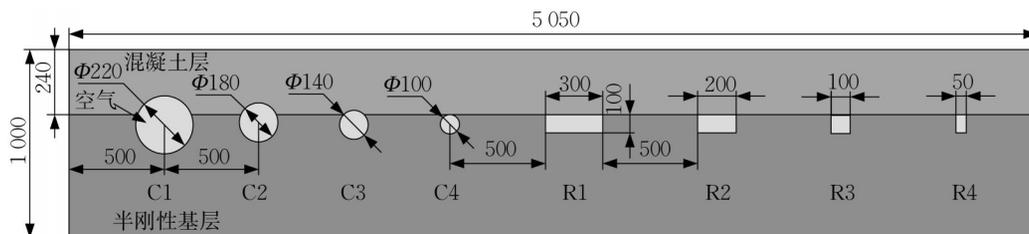


图3 正演脱空模型(单位:mm)

Fig.3 Simulated void model(unit:mm)

以波速  $0.1 \text{ m} \cdot \text{ns}^{-1}$  为例,900 MHz 和 1 000 MHz 的电磁波在混凝土的波长分别为 111 mm 和 100 mm,根据式(1)可得最小可探测脱空尺寸分别约为 28 mm 和 25 mm,因此 2 个天线难以分辨 25 mm 和 16 mm 两处脱空区域,导致这两处区域的 GPR 图谱

不清晰,但可以使用更高天线频率的 GPR 设备或者三维 GPR 检测小尺寸的脱空区域。

1.2.3 现场试验

对广西桂林永福县永鹿路部分路面进行数据采集,部分路面已经出现断板(见图 7b)。由专业检测

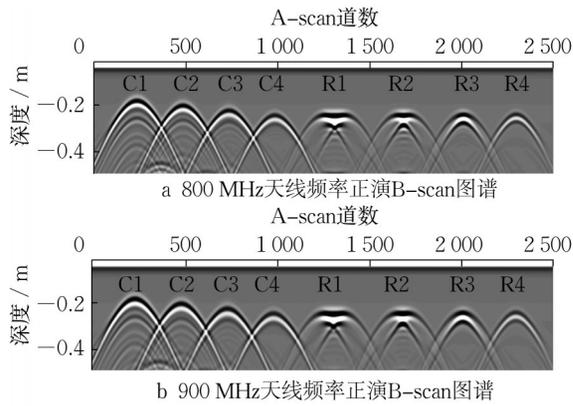


图 4 正演脱空模型的 B-scan 结果

Fig.4 B-scan results of simulated void model

表 2 探地雷达系统参数设置

Tab.2 GPR system parameter settings

雷达系统	天线中心频率/MHz	采样频率/GHz	A-scan 采样点数	采样间距/mm
RIS 雷达	900	16	471	2
US 雷达	1 000	40	467	10

人员和安全员协调下进行 GPR 检测,如图 7d 所示。试验设备包括 MALA ProEx 系统、800 MHz 天线和安装了 GroundVision 的个人电脑,采样频率为 14.5 GHz,每道采样点数为 308,时窗触发间隔为 0.01 s。

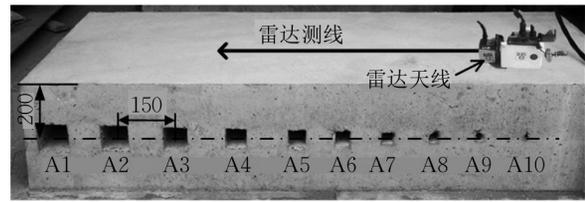


图 5 室内试验水泥板模型

Fig.5 Model of lab test

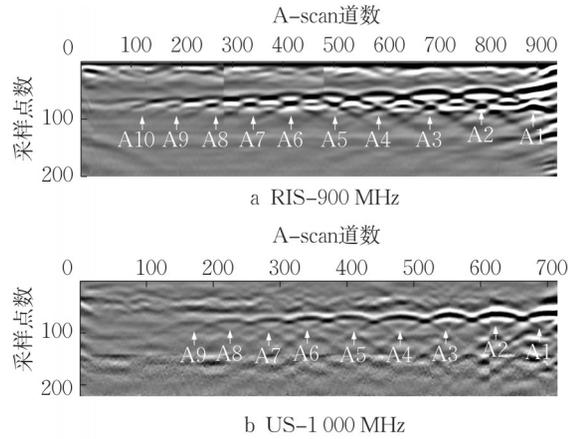


图 6 室内模型雷达实测剖面图谱

Fig.6 GPR results of lab model

典型路面图谱如图 8 所示。

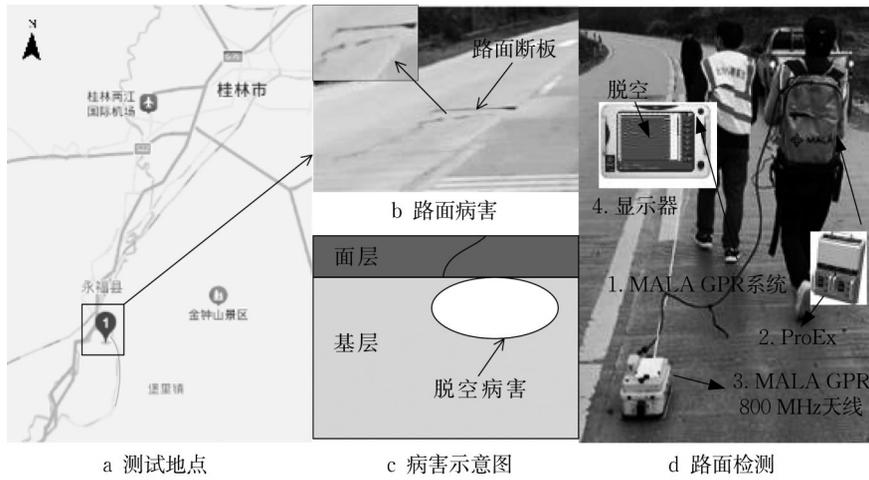


图 7 现场试验

Fig.7 Field test

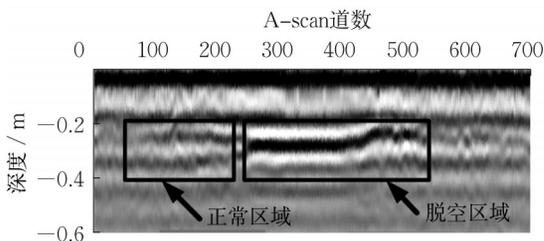


图 8 现场路面 B-scan 图谱

Fig.8 B-scan results of field test

采用以上 3 种试验方法,共获取 60 万道 A-scan 数据,正演模拟 5 万道,室内模型 5 万道,现场检测 50 万道。使用 ReflexW 软件对上述数据进行分析,并对数据集中的脱空区域数据给出“1”的真值标签,正常区域数据给出“0”的真值标签。经过人工筛选后,选择 20 983 道 A-scan 数据作为数据集,其中 10 623 道为室内和现场数据(脱空 6 745 道,正常 3 878 道),10 360 道为正演数据(脱空 3 987 道,正常 6 373 道),

合计脱空数据 10 732道,正常数据 10 251道。

### 1.3 GPR信号处理方法

对正反演 GPR 数据进行分析,如图 9 所示。与现场数据相比,正演数据中没有噪声和干扰,而室内试验与现场试验数据均存在干扰和噪声。因雷达设备的不同,室内模型数据的幅值明显低于现场路面数据的幅值,但在脱空病害部位具有相似的回波形状。因此,需要对 GPR 数据进行预处理,才能将正反演数据结果都用于后期建模。

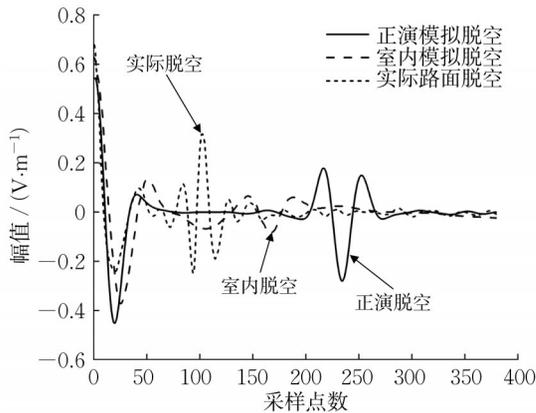


图9 正演模拟、室内和现场试验的A-scan数据对比

Fig.9 Comparison of GPR signal among simulation, lab and field tests

通过对比,确定了如图 10 所示的七步预处理方法以增强脱空特征,包括去直流漂移(去除直流漂移)、零点校正(去除天线与地面距离的误差)、能量增益(增强深层病害信号)、背景去除(减去平均道)、巴特沃斯带通滤波(滤除高频杂波干扰)、滑动平均和 F-K 偏移(纠正脱空区域偏移成分)。同时,为了解决采样频率不一致,采用重采样方法将原始数据转为具有相同采样频率的数据。

不同天线频率以及数据源差异均导致电磁波幅值差异,因此采用标准化处理方法,将 GPR 数据处理成均值为 0 和标准差为 1 的数据,具体方法如下所示:

$$\bar{X} = \frac{X - \mu(X)}{\sigma} \quad (2)$$

式中: $X$ 为探地雷达采集的数据矩阵; $\bar{X}$ 为标准化后的数据矩阵; $\sigma$ 为整体数据方差; $\mu(X)$ 为整体数据均值。

### 1.4 特征提取

参照水损害的特征提取方法<sup>[7]</sup>,对每一条 A-scan 数据提取了 16 个时域特征和 12 个频域特征,另外增加平均能量和偏度( $P_{18}$ )2 个时域特征,累计 30

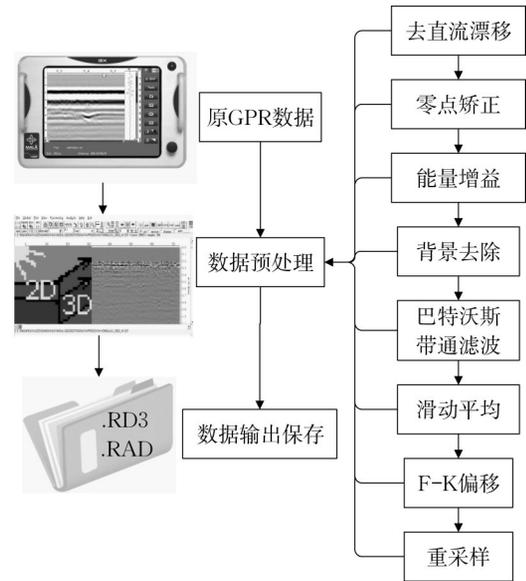


图10 雷达数据预处理流程

Fig.10 Flow chart of radar data preprocessing

个特征( $P_1 \sim P_{30}$ )。 $P_8, P_{17}, P_{18}$ 的计算式如下所示:

$$P_8 = \sqrt{\frac{1}{n-1} \sum_{i=1}^n (x_i - P_1)^2} \quad (3)$$

$$P_{17} = \frac{1}{n} \sum_{i=1}^n x_i^2 \quad (4)$$

$$P_{18} = \frac{1}{n} \sum_{i=1}^n \left( \frac{x_i - P_{17}}{P_8} \right)^3 \quad (5)$$

式中: $x_i$ 为该道 A-scan 第  $i$  个采样点数据,  $i=1, 2, \dots, n$ ,  $n$  是每条 A-scan 的采样点数。对提取的 30 个时频域特征采用式(2)进行标准化处理。图 11 为 3 种数据源中脱空区域和正常区域特征参数的对比结果。对比正常和脱空的信号特征,发现正演模型的特征中除  $P_6, P_9, P_{19}, P_{27}$  外,其余特征存在明显的差异;在室内模型的特征数据中,除  $P_{19}, P_{21}$  和  $P_{22}$  外,其余特征均存在明显差异;在实际路面特征中,除  $P_3$  和  $P_5$  外,其余特征数据均存在明显差异。结果表明,在这 3 种数据源中,正常与脱空数据的特征  $P_2, P_4, P_7, P_8, P_{10}, P_{11}, P_{13} \sim P_{17}, P_{20}, P_{24} \sim P_{26}, P_{28}$  和  $P_{30}$  均存在明显的差异,说明所选的时频域特征可用于脱空病害表征。

为了实现脱空识别,对数据集的 20 983 道数据进行时域和频域特征的提取,提取的时频域特征组成一个  $30 \times N$  ( $N=20\,983$ ) 的特征数据集矩阵。将 1.2 节中给出的对应每一条 A-scan 数据的真值标签移植到时频域特征数据集中,构成含有真值标签的  $31 \times N$  特征数据集,这个数据集用于后续模型训练。

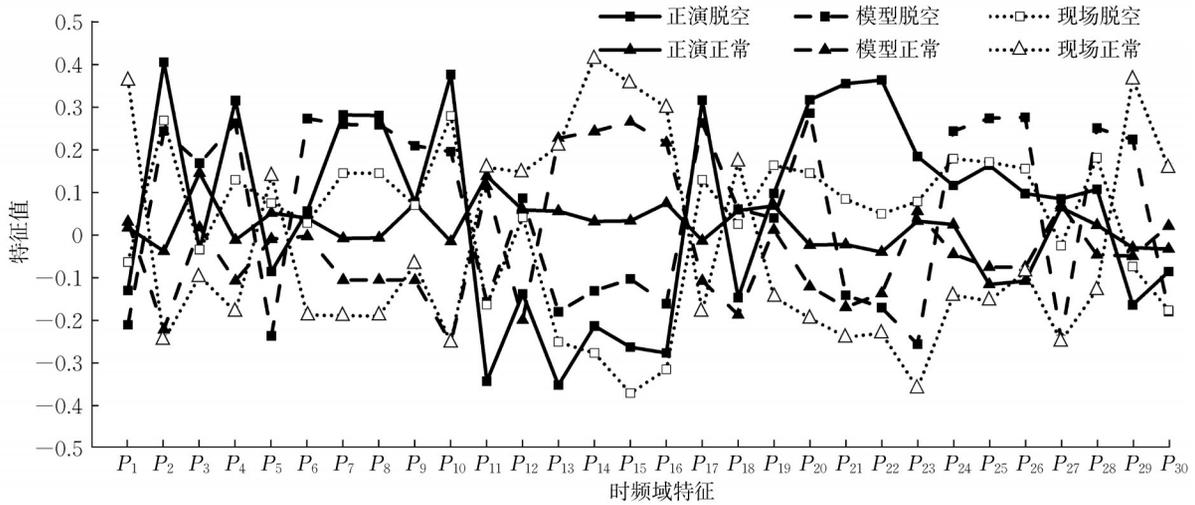


图 11 30 个时频域特征值对比

Fig.11 Comparison of 30 time-frequency domain feature values

### 1.5 特征重要性分析

对提取的 30 个时频域特征进行特征重要性分析,选择最小的特征子集,可加快模型训练速度和提升识别性能。特征重要性分析常用的方法有卡方检验、互信息、基于树模型和费希尔信息的方法。由于各种特征分析方法存在评价指标的差异,因此采用基于多种方法融合的分析方法来选取重要特征。特征重要性分析流程如图 12 所示。

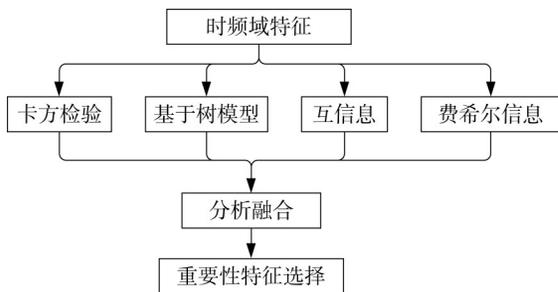


图 12 特征重要性分析流程

Fig.12 Flow chart of feature significance analysis

由于 4 种特征重要性分析方法的评价得分范围存在差异,单纯地将得分相加不能得到客观的最终得分,因此对 4 种分析方法的得分进行归一化处理,归一化结果相加得到最终的重要性评价得分。4 种分析方法归一化和相加的最终得分如图 13 所示,时频域特征重要性分析的最终得分排序如图 14 所示。

选取综合得分大于 1 的 21 个时频域特征作为识别模型的输入。进一步地去除  $31 \times N$  数据集中非重要的特征,得到含有真值标签的  $22 \times N$  数据集,用于识别模型的训练。

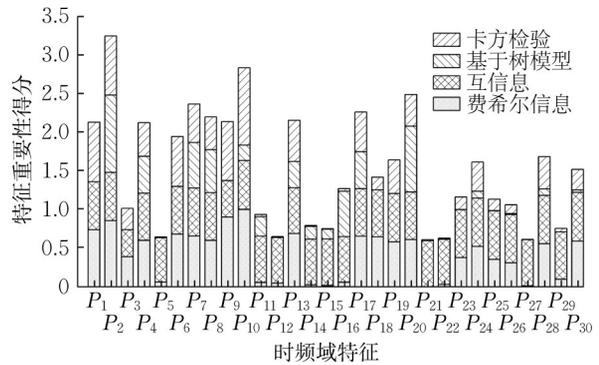


图 13 特征重要性得分归一化结果

Fig.13 Normalization results of feature significance score

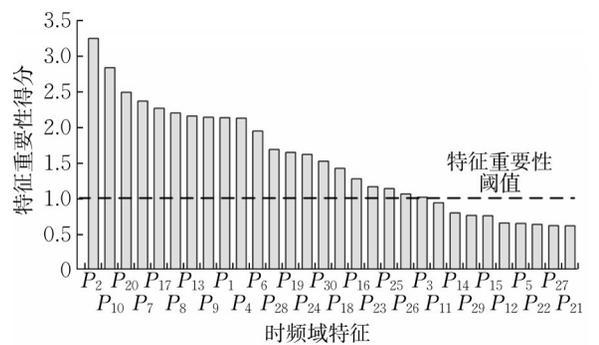


图 14 特征重要性得分排序

Fig.14 Ranking of feature significance

## 2 基于 XGBoost 的路面脱空识别算法

### 2.1 XGBoost 算法

对于 XGBoost 算法,为避免模型过拟合,加快训练速度,提升拟合及预测精度,损失函数应用了二阶泰勒展开,在目标函数中加入正则项。将含有  $N$  个

样本  $m$  个特征的数据集分为2类,数据集

$$D = \{(X_{11}, \dots, X_{1m}, y_1), \dots, (X_{i1}, \dots, X_{im}, y_i), \dots, (X_{N1}, \dots, X_{Nm}, y_N)\}, y_i \in \{0, 1\} \quad (6)$$

建立二分类模型,样本  $X_i$  在第  $k$  轮预测值  $\hat{y}_{i,k}$  的表达式如下所示:

$$\hat{y}_{i,k} = \sum_{k=1}^K f_k(X_i), f_k \in F \quad (7)$$

$$F = \{f(X) = \omega_{q(X)}, q \rightarrow T, \omega \in \mathbf{R}^T\}$$

式中:  $K$  表示模型中总的决策树数;  $F$  表示回归树空间;  $\omega$  表示回归树叶子节点权重;  $q$  表示回归树结构, 把每个样本节点映射到对应叶子节点的索引;  $q(X_i)$  表示样本  $X_i$  所在的叶子节点;  $T$  表示叶子节点数。

XGBoost 算法的目标函数包括损失函数和正则项两部分, 定义如下:

$$O_i = \sum_{i=1}^N L(y_i, \hat{y}_i(t)) + \sum_{j=1}^T W(f_j) \quad (8)$$

$$W(f_j) = \gamma T + \frac{1}{2} \lambda \sum_{j=1}^T \omega_j^2 \quad (9)$$

式中:  $\sum_{i=1}^N L(y_i, \hat{y}_i(t))$  为损失函数项;  $W(f_j)$  为目标函数的正则项, 由叶子节点数  $T$  和叶子节点权重  $\omega$  表示对模型复杂度的惩罚, 防止模型过拟合;  $\gamma$  和  $\lambda$  为模型超参数。在第  $j$  轮迭代时, 加入新的模型  $f_j$ , 对式(8)进行改写并按二阶泰勒级数展开, 得到如下目标函数:

$$O_i = \sum_{i=1}^n (L(y_i, \hat{y}_j(t-1)) + G_i f_j(X_i) + \frac{1}{2} H_i f_j^2(X_i)) + W(f_j) \quad (10)$$

式中:  $G_i = \sum_{i \in I_j} g_i = \sum_{i \in I_j} L'(y_i, \hat{y}_j(t-1))$  为损失函数的一阶梯度统计;  $H_i = \sum_{i \in I_j} h_i = \sum_{i \in I_j} L''(y_i, \hat{y}_j(t-1))$  为损失函数的二阶梯度统计。迭代获得模型的最小  $O$  值时表示模型训练完成。当模型预测结果大于 0.5 时, 将预测结果置为 1, 否则结果置为 0。

## 2.2 横向尺寸的权重设定

GPR 检测过程中存在噪声、干扰和天线跳动问题, 连续的 GPR 数据中存在单道或者连续几道的干扰信号, 致使识别模型对此类数据给出错误的识别结果。小尺寸的脱空病害也会对识别结果产生干扰, 降低识别结构的可读性。采用如图 15 所示的高斯权重系数, 当连续的 GPR 数据识别结果中存在孤立点时, 将雷达图谱的识别结果中孤立点两侧的识别结果乘以高斯权重系数进行综合考虑。应用时将

图 15 中的中心点(第  $i$  个识别结果)对应的数据作为该道识别结果的权重, 其左右 5 个权重系数分别赋给该道左右相邻 5 道, 通过与设定的阈值进行比较来过滤孤立点。

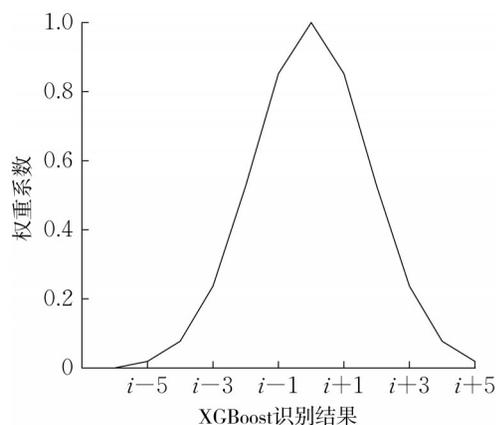


图 15 高斯权重系数曲线

Fig. 15 Gaussian weight coefficient curve

## 2.3 识别模型建立流程

将 GPR 信号数据集通过特征提取和归一化后构建特征数据集, 并分为训练集和测试集, 训练集用于 XGBoost 模型训练, 测试集用于模型的识别性能测试。预测模型建立流程如图 16 所示。

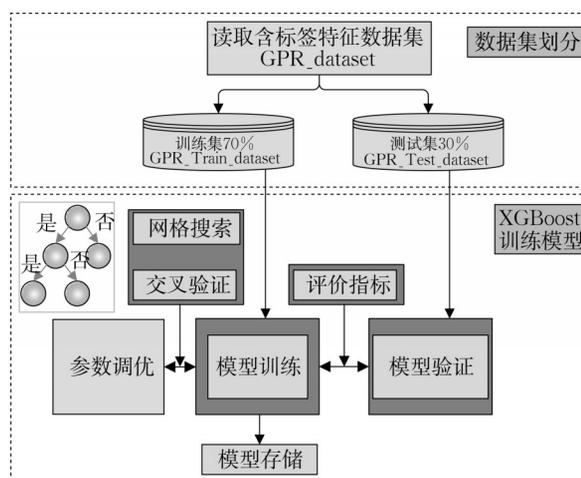


图 16 XGBoost 模型训练流程

Fig. 16 Flow chart of XGBoost model training

(1) 将含有真值标签的时频特征数据集 ( $22 \times N$ ) 随机抽取 70% 作为 XGBoost 模型训练集样本, 30% 作为模型测试集样本。

(2) 将训练集样本输入算法模型中进行训练, 并采用交叉验证和网格搜索方法调整模型的  $\max\_depth$  (最大树深) 和  $n\_estimators$  (子学习器个数) 等超参数, 获得最优的预测模型。

(3) 将测试集样本输入预测模型,检验模型的预测准确率,并存储预测模型。训练后获得的最优模型将用于后续路面脱空病害识别。

### 2.4 模型评价标准

模型识别结果的评价指标可以从混淆矩阵中导出。具体的分类评价标准及其公式和含义如表 3 所示。表 3 中,  $N$  表示测试集样本总数,  $M_{ii}$  表示模型对类别  $c_i$  准确识别的样本个数,  $M_{si}$  表示类别  $c_i$  样本个数,  $M_{is}$  表示模型识别为类别  $c_i$  的样本个数。上述指标越接近 1, 代表模型的分类效果越好。

表 3 准确性评价指标

Tab.3 Accuracy evaluation index

评价指标	含义	公式
$P$	识别真阳性与所有识别阳性的比值	$P = \frac{M_{ii}}{M_{is}}$
$R$	识别真阳性与全部真阳性的比值	$R = \frac{M_{ii}}{M_{si}}$
$F_1$	$P$ 和 $R$ 的倒数和的倒数	$F_1 = \frac{1}{1/P + 1/R}$
$A$	总测试集中所有准确识别的比例	$A = \frac{M_{ii}}{N} \times 100\%$

## 3 试验结果分析

### 3.1 模型优化

为提高模型的准确率,采用五折交叉验证和网格搜索方法调优模型的关键参数<sup>[25]</sup>。XGBoost 模型中  $max\_depth$  和  $n\_estimators$  对预测模型的过拟合性敏感,并影响模型收敛的稳定性。根据 XGBoost 模型超参数调参时的经验,将  $max\_depth$  调优范围设置为  $[2, 20]$ ,  $n\_estimators$  调优范围设置为  $[10, 2000]$ , 比较在不同的最大树深和不同的子学习器个数的情况下模型的收敛速度与识别精度。根据图 17a 中模型  $max\_depth$  参数调优曲线,将  $max\_depth$  设置为 11, 模型的准确率为 94.12%。根据图 17b 中  $n\_estimators$  调优曲线,将  $n\_estimators$  设置为 1000, 模型的准确率为 94.14%。对预测模型其他参数采取同样的调优操作,最终确定模型的最优参数组合,如表 6 所示。

### 3.2 模型性能对比

使用表 4 中调优后的模型超参数数值和 70% 的数据集进行 XGBoost 模型训练。为比较模型性能,相同的数据集还用于 RF 和 ANN 模型训练。模型建模平台均使用 Python 语言,在 CPU-AMD R7 5800H(3.20 GHz)、16 GB RAM 和 GPU-NVIDIA GeForce RTX 3060 Laptop(6 GB) 计算机上进行。

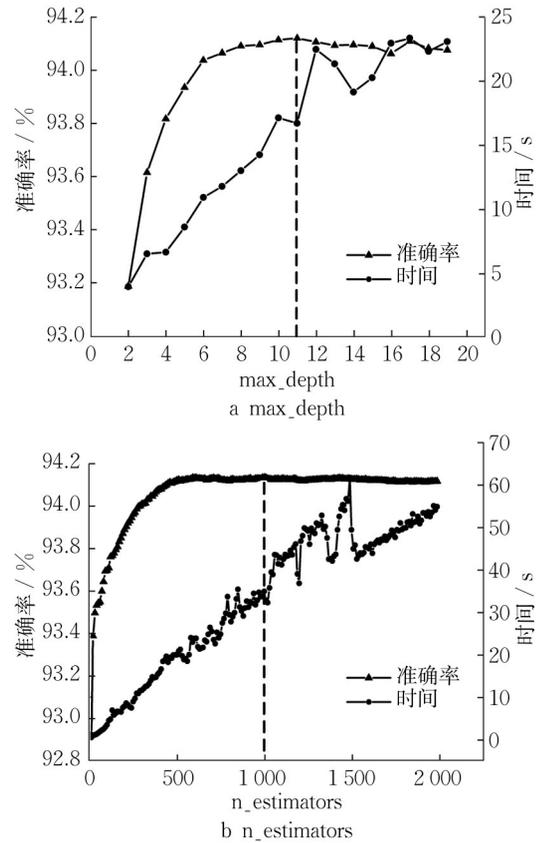


图 17 参数调优曲线

Fig.17 Parameter tuning curve

表 4 模型参数调优值

Tab.4 Parameter values of optimal model

参数名称	含义	设置值
$max\_depth$	最大树深	11
$learning\_rate$	学习率	0.01
$n\_estimators$	子学习器个数	1000
$subsample$	样本采样率	0.6
$reg\_lambda$	L2 正则化系数	1

将验证集(30%的数据集)输入建立的 3 种识别模型中,计算 3 种识别模型在验证集上的评价指标。由表 5 可见,3 种模型在测试集上的识别准确率排序为 XGBoost (99.63%) > ANN (99.30%) > RF (99.28%)。从识别准确率看, XGBoost 模型比另外 2 种模型更有效地区分雷达数据中的脱空病害和正常数据,对测试集中的脱空数据产生的误判率更低。

为了确定所提方法的有效性,在测试数据集

表 5 模型识别性能评价

Tab.5 Performance evaluation of recognition models

模型	$P$	$R$	$F_1$	$A/\%$
XGBoost	1.000 0	0.990 0	0.990 0	99.63
ANN	0.990 0	0.990 0	0.990 0	99.30
RF	0.990 0	0.980 0	0.980 0	99.28

(950道数据)上进行对比,结果如图18所示。可以看出,3种识别模型结果中都存在如图18中E标识所指的单道(孤立点)A-scan误判结果。XGBoost模型的识别结果中存在的单道A-scan误判情况明显优于RF和ANN模型的识别结果。统计3种模型的识别准确率和消耗时间,如表6所示。XGBoost模型对水泥路面雷达数据的识别准确率达98.10%,识别过程共耗时0.141 s;RF算法的识别准确率为93.17%,识别过程共耗时0.196 s;ANN模型的识别准确率为95.10%,识别过程共耗时0.203 s。与RF和ANN模型相比,XGBoost模型具有更高的预测准确率和更快的识别速度。

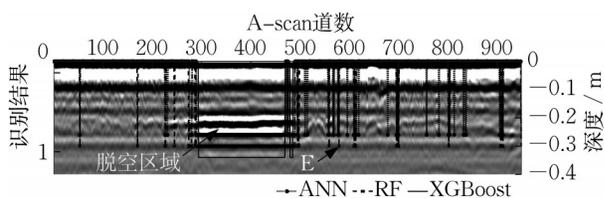


图18 XGBoost、RF和ANN模型识别结果

Fig.18 Performance of XGBoost, RF and ANN models

表6 模型识别性能

Tab.6 Performance of recognition models

模型	脱空识别准确率/%	预测时间/s
XGBoost	98.10	0.141
ANN	95.10	0.203
RF	93.17	0.196

## 4 模型适用性

### 4.1 正演模型数据

将图4中的正演模型数据采取与训练集相同的预处理流程,提取雷达数据的时频域特征,输入XGBoost识别模型中,获得如图19所示的预测结果。可以看出,XGBoost模型对正演图谱中大多数脱空区域都给出了“1”的识别结果,准确给出圆形脱空的双曲线特征范围,对矩形脱空的横向尺寸给出了较高准确率的识别。图20为XGBoost正演模型识别脱空尺寸和实际脱空尺寸的对比。从图20可见,XGBoost正演模型对矩形脱空和圆形脱空的横向尺寸的误差在8 mm以内。造成此现象的原因是:F-K偏移对仿真数据中的圆形脱空双曲线特征产生了误差。

### 4.2 现场验证

任取2段存在脱空区域的数据,考虑单个脱空

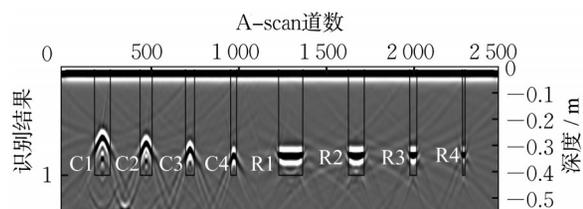


图19 XGBoost正演模型雷达数据预测结果

Fig.19 Prediction results of XGBoost on forward model

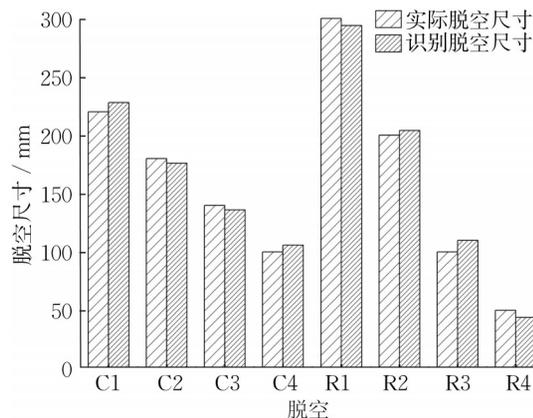


图20 脱空横向尺寸对比

Fig.20 Comparison of lateral dimension

和多处脱空的状态,现场检测GPR结果如图21a、b所示。按照图10预处理流程进行了特征提取后,将特征数据输入XGBoost模型,识别结果如图21所示。图21c、e为模型直接输出的识别结果。可知,脱空区域被准确识别,但识别结果中存在多道孤立点,具体原因如下:

(1) XGBoost模型给出连续的“1”的识别结果比孤立的“1”的识别结果更能说明此处路面脱空病害的存在,在对路面脱空病害尺度的评估中占有更大权重。

(2) XGBoost模型识别结果中存在稀疏的孤立“1”的结果,可能不是由路面脱空病害造成的,如数据集不全面;在进行雷达数据采集时存在干扰造成的雷达信号波动;存在早期的脱空病害(横向尺寸小于2 cm),但尺寸过小难以验证。

实际项目中主要关注大范围脱空区域,为减少孤立点对识别结果的干扰,引入识别结果的后处理方法。当一条A-scan的XGBoost模型识别结果为“1”,左右相邻的连续10条雷达数据的识别结果与高斯判断系数相乘并相加的和等于或大于5时,此条A-scan识别结果才为“1”。此外,如果一个测试段(10条A-scan)雷达数据中有90%的A-scan都被XGBoost模

型识别为“1”,整个测试段就被视为一个连续的脱空区域,XGBoost模型的结果都修改为“1”。

对XGBoost模型识别结果进行后处理,得到如图21d、f所示的后处理结果,后处理可以有效地去除孤立点的影响。在图21d中,已经形成了长尺寸的脱空区域,连续脱空区域会在车辆交变荷载等作用下发展为更大的脱空,需要及时采取补救

措施。GPR专家可以识别图21f中的脱空区域。然而,在对比度较低的雷达图谱中脱空区域很难评价。在0~100条和300~400条A-scan之间存在两处小尺寸的脱空区域,说明此脱空区域处于早期阶段;在200~300条A-scan存在相邻很近的两处脱空区域,说明这两处脱空会发展为连续的脱空区域。

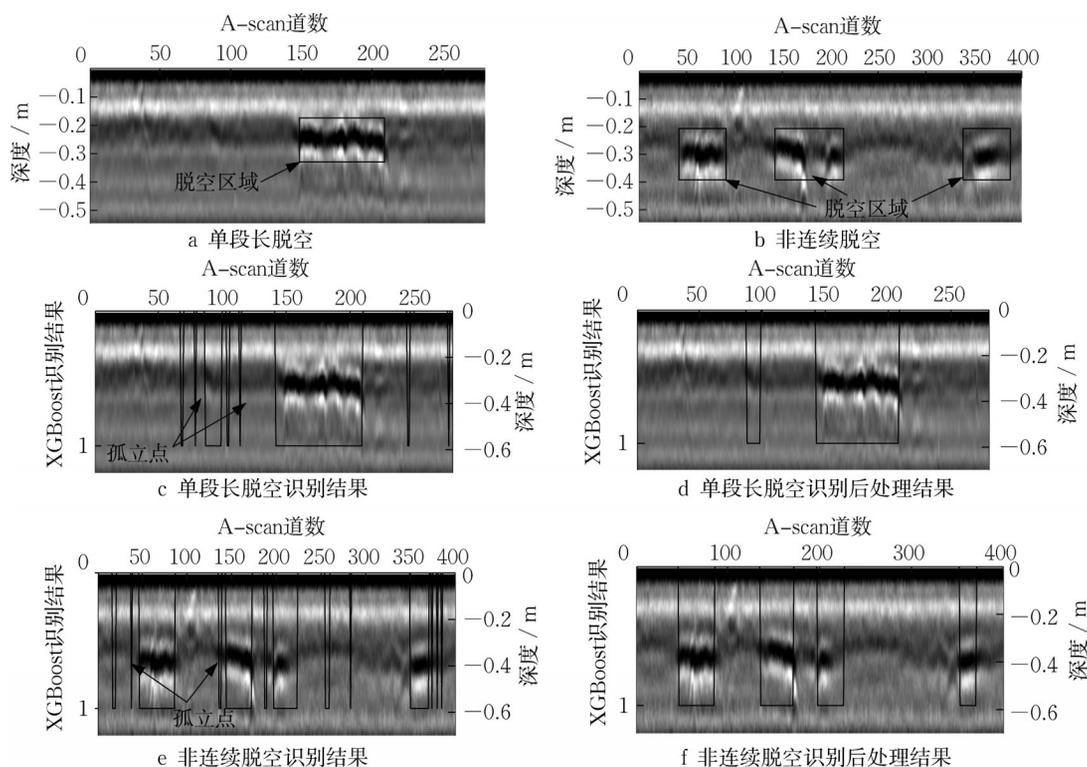


图21 实际路面脱空病害识别

Fig.21 Identification of actual pavement void disease

上述验证结果表明,所构建的模型可有效地确定路面脱空区域。为进一步验证模型的准确性,在封闭交通的条件下,对模型识别出脱空的位置(见图22a)采

用破碎镐进行破碎,并通过针孔摄像头对破碎孔进行观察,结果如图22b所示。针孔成像结果验证了脱空病害的存在,进一步验证了算法的准确性。

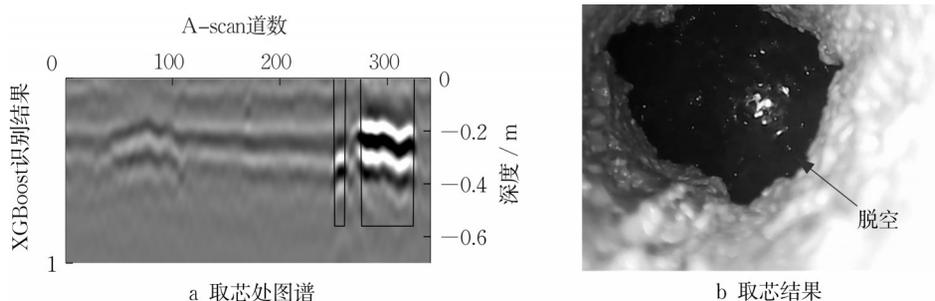


图22 取芯验证

Fig.22 Verification by coring

## 5 结论

(1)分析了GPR在脱空和正常区域的时域和频

域响应差异,提出了18个时域和12个频域的时频域特征,用以表征脱空和正常区域的GPR数据。通过正演模拟、室内试验和现场试验构建脱空数据集来

训练 XGBoost 模型,实现了脱空病害的自动识别。

(2) 提出了重采样和标准化的 GPR 数据处理方法,统一不同 GPR 数据的采样频率,解决了不同采样频率和不同天线频率对识别模型的影响,可以解决不同源的数据使用和识别问题。

(3) 对比了 XGBoost、RF 和 ANN 的模型, XGBoost 模型最优,脱空识别准确率达到 98.10%,高于 ANN(95.10%)和 RF(93.17%)。

#### 作者贡献声明:

张 军:论文构思,论文撰写和修改。

姜文涛:模型构建,程序设计,论文撰写及修改。

张 云:开展室内和现场试验。

罗婷倚:开展现场试验和数据分析。

余秋琴:试验数据分析。

杨 哲:孤立点的权重系数设计。

#### 参考文献:

- [1] 王春和,胡通海,崔海涛,等. 探地雷达技术用于地下空洞塌陷灾害探测的创新与实践[J]. 测绘通报, 2013(S1): 13.  
WANG Chunhe, HU Tonghai, CUI Haitao, *et al.* The innovation and practice of ground penetrating radar technology used for disasters investigation especially for underground cavity or collapse detection [J]. *Bulletin of Surveying and Mapping*, 2013(S1): 13.
- [2] 刘迎春,覃晖. 隧道壁后空洞检测的探地雷达试验[J]. 实验室研究与探索, 2022, 41(1): 53.  
LIU Yingchun, QIN Hui. Model test of tunnel lining void detection using ground penetrating radar [J]. *Research and Exploration in Laboratory*, 2022, 41(1): 53.
- [3] 栗宝鹃,张美多,王志豪,等. 水工混凝土结构钢筋分布及锈蚀检测[J]. 工程地球物理学报, 2021, 18(2): 257.  
LI Baojuan, ZHANG Meiduo, WANG Zhihao, *et al.* Steel bar distribution and corrosion detection of hydraulic concrete structure [J]. *Chinese Journal of Engineering Geophysics*, 2021, 18(2): 257.
- [4] 彭建,杨泽帆,白洁,等. 基于探地雷达的地下管线埋深估计方法[J]. 雷达科学与技术, 2022, 20(1): 79.  
PENG Jian, YANG Zefan, BAI Jie, *et al.* Depth estimation of underground pipeline using ground penetrating radar [J]. *Radar Science and Technology*, 2022, 20(1): 79.
- [5] 朱兆荣,赵守全,秦欣,等. 探地雷达在铁路隧道衬砌质量无损检测中的应用研究[J]. 工程地球物理学报, 2021, 18(5): 703.  
ZHU Zhaorong, ZHAO Shouquan, QIN Xin, *et al.* Application of ground penetrating radar in nondestructive testing of railway tunnel lining quality [J]. *Chinese Journal of Engineering Geophysics*, 2021, 18(5): 703.
- [6] ZHANG J, LU Y M, YANG Z, *et al.* Recognition of void defects in airport runways using ground-penetrating radar and shallow CNN [J]. *Automation in Construction*, 2022, 138: 104260.
- [7] ZHANG J, ZHANG C, LU Y M, *et al.* In-situ recognition of moisture damage in bridge deck asphalt pavement with time-frequency features of GPR signal [J]. *Construction and Building Materials*, 2020, 244: 118295.
- [8] 王辉,欧阳缙,廖可非,等. 基于深度学习的 GPR B-SCAN 图像双曲线检测方法[J]. 电子学报, 2021, 49(5): 953.  
WANG Hui, OUYANG Shan, LIAO Kefei, *et al.* GPR B-SCAN image hyperbola detection method based on deep learning [J]. *Acta Electronica Sinica*, 2021, 49(5): 953.
- [9] XIANG Z, RASHIDI A, OU G. An improved convolutional neural network system for automatically detecting rebar in GPR data [C]//*Computing in Civil Engineering 2019: Data, Sensing, and Analytics*. Reston: American Society of Civil Engineers, 2019: 422-429.
- [10] 张东昊,覃晖. 基于探地雷达和深度学习的隧道初期支护检测方法[J]. 现代隧道技术, 2020, 57(S1): 174.  
ZHANG Donghao, QIN Hui. Tunnel primary support detection using ground penetrating radar and deep learning [J]. *Modern Tunnelling Technology*, 2020, 57(S1): 174.
- [11] KO H-Y, KIM N-G. Performance analysis of detecting buried pipelines in GPR images using faster R-CNN [J]. *Journal of Convergence for Information Technology*, 2019, 9(5): 21.
- [12] DINH K, GUCUNSKI N, ZAYED T. Automated visualization of concrete bridge deck condition from GPR data [J]. *NDT & E International*, 2019, 102: 120.
- [13] LI S, GU X, XU X, *et al.* Detection of concealed cracks from ground penetrating radar images based on deep learning algorithm [J]. *Construction and Building Materials*, 2021, 273: 121949.
- [14] 王辉,欧阳缙,刘庆华,等. 基于深度学习的探地雷达二维剖面图像结构特征检测方法[J]. 电子与信息学报, 2022, 44(4): 1284.  
WANG Hui, OUYANG Shan, LIU Qinghua, *et al.* Structure feature detection method for ground penetrating radar two-dimensional profile image based on deep learning [J]. *Journal of Electronics & Information Technology*, 2022, 44(4): 1284.
- [15] XIE X Y, QIN H, YU C, *et al.* An automatic recognition algorithm for GPR images of RC structure voids [J]. *Journal of Applied Geophysics*, 2013, 99: 125.
- [16] 郑艾辰,赵浩然,谭冰心,等. 基于 SVM 的隧道衬砌空洞填充物雷达图像识别研究[J]. 现代隧道技术, 2022, 59(2): 45.  
ZHENG Aichen, ZHAO Haoran, TAN Bingxin, *et al.* Radar image recognition of tunnel lining cavity fillings based on SVM [J]. *Modern Tunnelling Technology*, 2022, 59(2): 45.
- [17] BREIMAN L. Random forests [J]. *Machine Learning*, 2001, 45(1): 5.
- [18] CHEN T Q, GUESTRIN C. XGBoost: a scalable tree boosting system [C]//*Proceedings of the 22nd ACM SIGKDD*

(下转第 121 页)