

# A Recognition Method for Ballasted Railway Subgrade Defects

Zhezhe Hou (✉ [houzhezhe@stdu.edu.cn](mailto:houzhezhe@stdu.edu.cn))

Shijiazhuang Tiedao University <https://orcid.org/0000-0003-1976-0379>

Weigang Zhao

Shijiazhuang University

Yong Yang

Shijiazhuang University

---

## Research Article

**Keywords:** Ground penetrating radar(GPR), Ballasted railway, Subgrade defects, Feature extraction, Sparse representation

**Posted Date:** April 21st, 2022

**DOI:** <https://doi.org/10.21203/rs.3.rs-1516531/v1>

**License:** © ⓘ This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

---

1  
2  
3  
4  
5  
6

# **A Recognition Method for Ballasted Railway Subgrade Defects**

Zhezhe Hou<sup>a,b\*</sup>, Weigang Zhao<sup>c</sup>, Yong Yang<sup>c</sup>

*a. School of Material Science and Engineering, Shijiazhuang Tiedao University, China*

*b. School of Mechanical Electronic and Control Engineering, Beijing Jiaotong University, 100083, China*

*c. Structure Health Monitoring and Control Institute, Shijiazhuang Tiedao University, China*

7 **Abstract:** A recognition method based on sparse representation was proposed to solve the problems in subgrade  
8 detection with ground penetrating radar (GPR), such as massive data, time-frequency and difference in experience.  
9 Based on the propagation characteristics of electromagnetic waves and radar images in subgrade defects, the phase axis  
10 of radar images of railway subgrade defects was different from that of normal subgrades. The sparse representation of  
11 railway subgrade defects was studied from the aspects of the time domain, and time-frequency domain with  
12 compressive sensing theory. The high-frequency horizontal details of the radar images were obtained to reduce  
13 redundant data fetching. The demixing points, energy, and variance per block were obtained as time domain  
14 eigenvalues, and the high-frequency horizontal details of the radar images and the energy spectrum of the wavelet  
15 multiscale spatial were acquired; thus, the feature vectors of the radar signal were extracted by sparse representation.  
16 Based on FCM-GANN (fuzzy C-means and generalized regression neural network), rapid recognition of the ballasted  
17 railway was realized. Experimental results indicated that the redundancy of data was reduced, and the accuracy of  
18 identification was greatly increased.

19 **Keywords:** Ground penetrating radar(GPR); Ballasted railway; Subgrade defects; Feature extraction; Sparse  
20 representation

## 21 1. INTRODUCTION

22 Railway subgrades have long been influenced by the environment, climate conditions, and trainload. Ground  
23 penetrating radar(GPR) has become the main technical means for the detection of railway subgrades[1-3]. Structural  
24 characteristics and water content are directly reflected by radar echoes. The key problem is how to extract the effective  
25 features of radar signals, which plays an important role in identifying railway subgrade defects.

26 Because of rapid, nondestructive, continuous and efficient detection, GPR technology has attracted considerable  
27 attention from railway operation and maintenance departments and has been gradually adopted for the continuous  
28 detection of railway subgrades [4-7]. However, the massive data, time-frequency and difference in experience cannot  
29 guarantee the timeliness and effectiveness of the data analysis, which has affected the generalization of GPR. In view of  
30 the above problems, in recent years, many countries and scholars have carried out evaluations of railway subgrades by  
31 GPR. For example, the preliminary GPR test was executed from 1993 to 1994 in Germany, and a set of GPR signal  
32 processing and defect location systems were developed after more than ten years, but the defect types were not  
33 comprehensive, and the automatic identification of the defect could not be carried out. Carnegie Mellon University  
34 developed an automated classification method to assess the quality of railway subgrades, but it was limited to  
35 identifying the ballast bed, and the defect types could not be distinguished.

36 In summary, GPR is a powerful tool to evaluate the state of railway subgrades[8-10]. The GPR data were collected in

37 a real life/laboratory environment. By interpretation of the GPR data, the reasonable target feature vector was extracted  
38 and the subgrade condition was assessed [11-13]. However, the processing of GPR data was challenging. Currently,  
39 there are primary approaches for data interpretation: the time domain approach[14, 15], the time-frequency  
40 approach[16], and the wavelet approach [17].

41 The time domain approach is applied to a single structure or obvious layers [15, 18]. Considering the different  
42 railway defects, segmented energy, variance and interface were extracted as the eigenvalues from GPR data, the defect  
43 types can be distinguished, whereas the subgrade sinkhole and subgrade settlement cannot be distinguished obviously as  
44 the interface was not clear enough. The GPR signals were processed by short-time Fourier transform (STFT) in the time  
45 and frequency domains simultaneously, and effectively accessed ballast fouling and moisture content [9]. This method  
46 needed to have a clear interface between clean and fouled ballast, which affected the accuracy. Wavelet transform was  
47 applied to interpret GPR data to evaluate ballast fouling, which can interpret GPR data continuously, automatically and  
48 without interpreters' subjectivity [13]. A combination of both time-frequency and discrete wavelet techniques could give  
49 the levels of fouling ballast [1] .

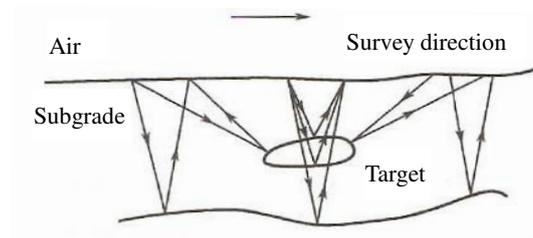
50 Traditional recognition methods need to seek eigenvalues from a large number of original data, and then reduce  
51 the data processing to achieve relatively accurate and rapid recognition. The sparse representation method  
52 overcame the limitation of the traditional sampling theorem. It selected the representative raw data as the sample  
53 according to the characteristics of the target, reduced the amount of data collected, and speeded up the data  
54 processing. A more concise and direct signal analysis method was provided. The sparse feature vector of ballasted  
55 railways was presented [19, 20]. The frequency and standard deviation were used as a combined feature. This approach  
56 transformed the scattering of landmines to geometrical features, which had a strong relationship with their physical  
57 characteristics. Compared with other feature extraction techniques, this method had a clear advantage in sparsity for the  
58 ballast layer or single structure, and obviously reduced the amount of data. According to the radar images of railway  
59 defects, vertical projection was used for the sparse characteristics, a support vector machine was used to identify

60 railway defects[21], whereas subgrade settlement and sinkholes were difficult to distinguish, which reduced the  
61 accuracy of identification.

62 In this paper, aiming at the problem of machine recognition of radar images of typical railway subgrade defects, and  
63 the characteristics of nonuniform medium and nonspecific shapes, the sparse distribution characteristics of the spatial  
64 structure of railway subgrade diseases were analyzed, and the normal subgrade and its defects were distinguished. The  
65 feature extraction of typical radar images in subgrade defects was studied from the aspects of the time domain,  
66 time-frequency domain and physical geometry. The optimal radar signal feature was established. Thus, sparse  
67 expression of the subgrade defects was obtained, and the rapid identification of ballast railway defects was achieved.

## 68 **2 STATE OF THE ART**

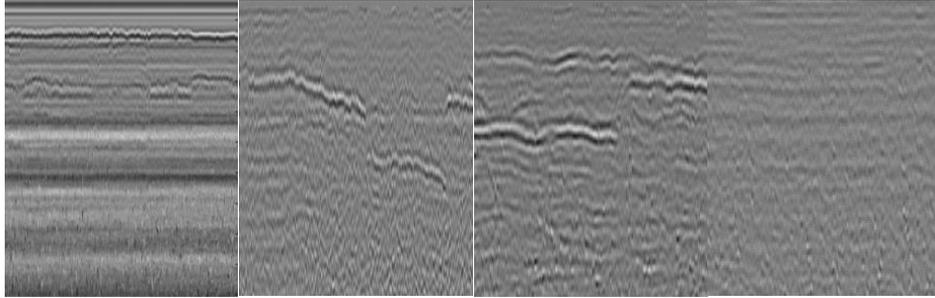
69 When GPR was used to detect railway subgrades, the subgrade defects were the detection targets. According to the  
70 waveform of GPR reflected waves, the location and growth degree of defects could be determined. Fig.1 shows the  
71 imaging principle of the target in the railway subgrade.



72  
73 Fig.1 The principle of the target in railway subgrade

74 Typical defects of the existing ballasted railway subgrade were mainly detected by GPR, including mud pumping,  
75 settlement and sinkholes. The phase axes of settlement were continuous and the parabola curves were generated. The  
76 radar images of the subgrade sinkhole showed uneven phase axes. The change in the phase axes could be identified with  
77 the settlement and sinkhole. The radar images of mud pumping had no variation of the same phase axis at all, and were  
78 similar to the snowflake images to distinguish the other subgrade states. The echo signals included target signals and  
79 noise signals such as antenna coupled waves, radio frequency signals, ground reflection, and random noise. However,  
80 the intensity of the target signals was weakened by noise, even masked. The echo signals were pre-processed to obtain  
81 high-resolution radar images. Based on the radar images, railway subgrade defects could be classified into two groups:

82 subgrade settlement and mud pumping. Subgrade settlement included settlement and sinkhole. From the formation  
 83 mechanism, due to the insufficient ground bearing capacity, ballast invasions into the subgrade bed, both types of  
 84 defects were caused, the settlement and the sinkhole had the obvious phase axis, and the axis locations were different  
 85 from the normal subgrade, which could be judged from the normal subgrade.



86  
 87 Fig. 2 The GPR images of the railway subgrade  
 88

89 Some sections of the Daqin and Shichang railways were detected by GPR, and a large quantity of railway subgrade  
 90 defects were sparse in radar images. Therefore, railway subgrade defects met the requirements of sparse theory.

### 91 3 METHODOLOGY

#### 92 3.1 Target sparsity and sparse imaging

93 For radar signals of fixed frequency, the mixer output is a linear frequency modulation signal, and its frequency  
 94 signal is as follows:

$$95 S_r(t) = A e^{-j2\pi\left(f_0 t + \frac{1}{2} k t^2\right)} \quad (1)$$

96 where  $A$  is the amplitude of the signals,  $f_0$  is the initial frequency,  $A$  is the transmit amplitude, and  $k$  is the frequency  
 97 modulation slope. The echo of a point with a distance of  $H$  is as follows:

$$98 S_H(t) = A \sigma e^{-j2\pi\left(f_0 t + \frac{1}{2} k t^2\right) + j2\pi H(\rho, i)/c} / S(H(\rho, i)) \quad (2)$$

99 where  $\rho$  is the target location,  $S_{(H)}$  is the attenuation factor,  $c$  is the electromagnetic wave velocity in a vacuum,  $i$  is the  
 100 echo channel number, and  $\sigma$  is the target reflection coefficient.

101 To speed up signal reading and processing speed, every 25 channels of radar signals formed a  $256 \times 256$  pixel image  
 102  $\varphi[\mu_x, \mu_y]$ . The coincidence rate of radar images is 50%. Combined with the sparsity of railway subgrade defects, the

103 relationship between the measurement target and space images is as follows:

$$104 \quad d(\mu_x, \mu_y, f) = \psi \pi_T(x, y, z) \quad (3)$$

105 where  $\pi_T(x, y, z)$  is the spatial position of the measurement target,  $d(\mu_x, \mu_y, f)$  is the frequency-space image, and  $\psi$   
106 is the space transformation basis matrix, that is, the dictionary.

### 107 3.2 Sparse matrix of radar signals

108 To establish the sparse matrix, the measurement target must be discrete in spatial position,  $i$  is the echo channel  
109 number,  $(x_i, y_i, z_i)$  is the spatial position, image space  $B$  is formed correspondingly by  $N$  pixels  $\{\pi_1, \pi_2, \dots, \pi_n\}$ , and  
110 each pixel  $\pi_i$  corresponds to the three-dimensional vector  $(x_i, y_i, z_i)$ . The  $i$ th vector radix of pixel  $\pi_i$  is as  
111 follows:

$$112 \quad [\psi_i]_j = \exp[-j\omega(t - \Gamma_i(\pi_j))] \quad (4)$$

113 where  $\omega$  is the frequency vector  $B$ , and the dictionary matrix corresponding to the echo channel is obtained through  
114 Formula (4). The  $P$  target echo is received by the  $i$ th echo channel

$$115 \quad \zeta_i(\omega) = \sum_{k=1}^P b(k) \exp[-j\omega(t - \Gamma_i(\pi_k))] \quad (5)$$

116 Formula (6) was converted to a vector

$$117 \quad \zeta_i(\omega) = \psi_i b \quad (6)$$

118 where  $b$  is the weighted steering vector of the target space,  $\pi_j$  is the partial position of the measurement target,  
119  $b_j = A\sigma_j/\Gamma(\pi_j)$ ; otherwise,  $b_j = 0$ .

### 120 3.3 Random sampling frequency

121 Assume that the collected radar signals are made up of huge amounts of one-dimensional signals of length  $L$  with a  
122 sparsity of  $k$  (that is, it contains  $k$  nonzero values), which form a large quantity of sample data. Due to the sparsity of  
123 railway subgrade defects in space, this paper proposes the compressive sensing method for data sampling, that is, a  
124 small number of signals represent all signals, to construct the target images. The measurement matrix should be a  
125 random matrix that is not related to the dictionary. In this paper, the Bernhard matrix composed of 0 and 1 elements is  
126 selected as the observation matrix. Therefore,  $M$  random rows are extracted from the  $L \times L$  identity matrix to obtain the  
127 measurement matrix  $\phi_i$  corresponding to the  $i$ th echo channel. The measured value is

128 
$$\beta_i = \phi_i \xi_i = \phi_i \psi_i b \tag{7}$$

129 The measurement matrix  $\phi_i$  of each echo channel is different.

130 **3.4 Sparse matrix algorithm**

131 To obtain the space guidance vector  $b$ , echo channels  $K$  were selected. The dictionary matrix is  $\psi = [\psi_1^T, \psi_2^T \dots, \psi_K^T]$ ,  
 132 the measurement matrix is  $\phi = \text{diag}[\phi_1, \phi_2 \dots, \phi_K]$ , the measurement value is  $\beta = [\beta_1^T, \beta_2^T \dots, \beta_K^T]$ , and reconstruction  $b$   
 133 becomes a constrained problem to solve the convex optimization problem.

134 
$$\hat{b} = \text{argmin} \|b\|_1 \quad \text{s.t.} \quad \beta = \phi \psi b \tag{8}$$

135 where Formula (8) is only in the absence of noise. The measured value corresponding to the  $i$ th echo channel with  
 136 noise:

137 
$$\beta_i = \phi_i \xi_i = \phi_i \psi_i b + \mu_i \tag{9}$$

138 where  $\mu_i = \phi_i n_i \sim N(0, \sigma^2)$  and  $n_i \sim N(0, \sigma^2)$  is aliasing noise. Convex optimization of the improved  $l_1$  norm under  
 139 constraint conditions:

140 
$$\hat{b} = \text{arg min} \|b\|_1 \tag{10}$$

141 where  $A = \phi \psi$ ,  $\varepsilon = \sigma \sqrt{2 \lg N}$ . Formula (10) can be used to create the target images.

142 **3.5 Feature representation of railway subgrade defects**

143 **1. Feature extraction**

144 **1) Feature extraction of GPR signals based on a two-dimensional wavelet**

145 The railway subgrade is a kind of continuous medium, and the phase axes of radar images are continuous except for  
 146 localized mutations because of the correlation of the radar signals, as shown in **Fig.3**.The first image is the low  
 147 frequency detail (see above left), while the other three images are the outline of the image, that is, the horizontal,  
 148 vertical and diagonal details are the high frequency details.

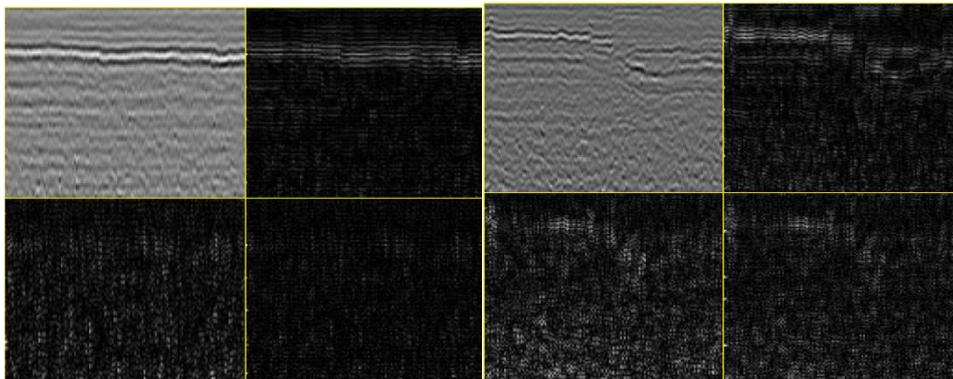


Fig. 3 GPR images of railway subgrade defects for normal subgrade and settlement

150 The radar images of the subgrade have obvious phase axes, and the axes are almost continuous. In most cases, the  
 151 axes slowly vary. A two-dimensional Harr wavelet transform is applied to analyze the radar images, as shown in Fig.3.  
 152 The horizontal detail highlights the radar signals with a large correlation in the horizontal direction, and it shows the  
 153 phase axes of the radar images (see right-hand chart, above).  
 154

2) Feature extraction of GPR signals based on the time domain

155 Based on the continuity and disorder of the phase axes of the subgrade, the time domain characteristics of the  
 156 subgrade defects are established. A signal of length N is divided into M blocks, and each block image is divided into K  
 157 segments by length, as shown in Fig.4. The coincidence rate between the images is 50%. The features of subgrade are as  
 158 followed:  
 159

$$E_i = \sum_{j=M-i/2+1}^{M-i/2+M} A_j^2 \quad (11)$$

$$\sigma_i^2 = \frac{1}{M-1} \sum_{j=M-i/2+M}^{M-i/2+M} (A_j - \bar{A}_i)^2 \quad (12)$$

160 where  $i = 0, 1, 2, \dots, K-1$ ;  $E_i$  is the energy of the  $i$ th segment;  $A_j$  is the amplitude of the  $j$ th sampling point;  $\sigma^2$  is the  
 161 sample variance of the  $i$ th segment;  $\bar{A}_i$  is the mean amplitude of the  $i$ th segment.  
 162  
 163

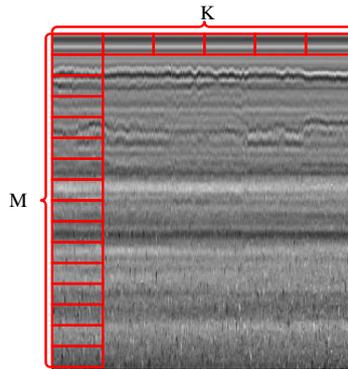
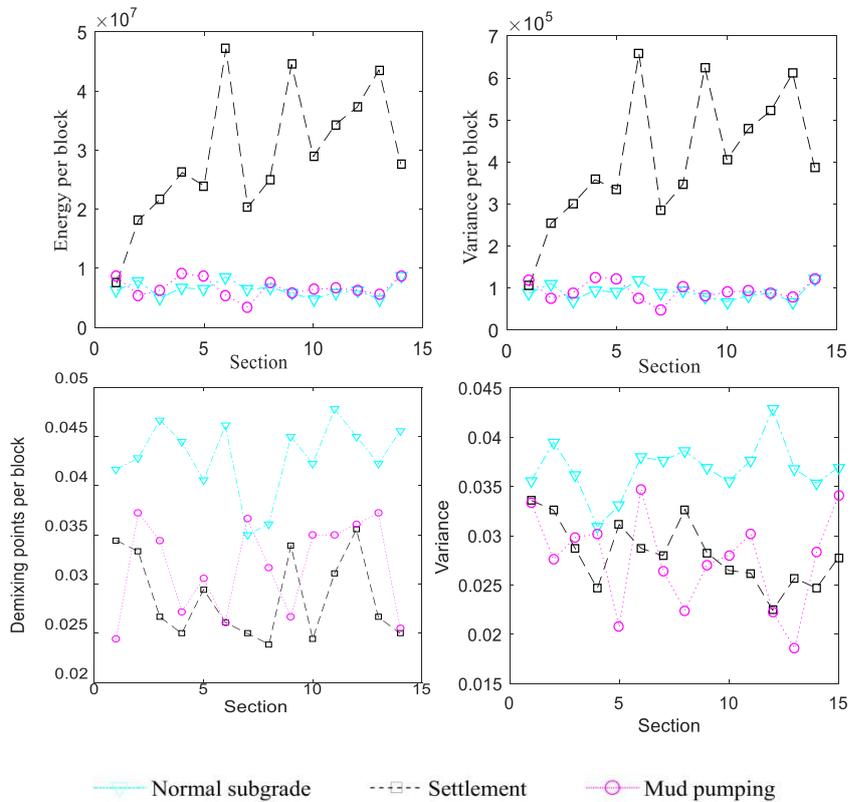


Fig.4 The segments of radar image

164 It is very important to choose appropriate M and K for the identification of subgrade defects. On the one hand, the  
 165 larger the M value is, the faster the image processing and the faster the recognition speed, but the recognition rate is also  
 166 reduced; In contrast, the smaller the M value, the smaller the image block, and the higher the recognition rate of radar  
 167  
 168

169 images, but the recognition speed will decrease. On the other hand, the speed of processing data is affected directly by  
 170 the K value.



171

172

173

Fig.5 Eigenvalue curve

174

175

176

177

178

179

180

181

182

183

Due to the serious changes of the subgrade defects, the energy and the phase axis change obviously. The space location, energy, and variation of defects are different from those of the normal subgrade, as shown in Fig.5. The energy per block and the variation per block can distinguish the normal subgrade from the defects. The phase axes of settlement apparently decline, the energy of the fault increases obviously, and the variation and demixing points per block are distinguished from the subgrade settlement. Due to mud pumping, the interfaces of mud pumping are not obvious. In addition, the high conductivity of the mud pumping makes the energy of the radar image low. Judging from the energy and variance of the radar images, the subgrade defects can be identified.

Thus, the energy per block, the variation per block, the variation and the demixing points per block are used as feature vectors to identify the position and the range of subgrade defects.

### 3) Horizontal energy spectrum

184 The characteristics of the subgrade radar signal are different at each scale. Each scale energy has different  
 185 contributions to the total energy. The main part of the signal is identified according to the characteristics of the energy  
 186 spectrum. **The component energy of wavelet decomposition** at the  $J$ th scale is shown as follows:

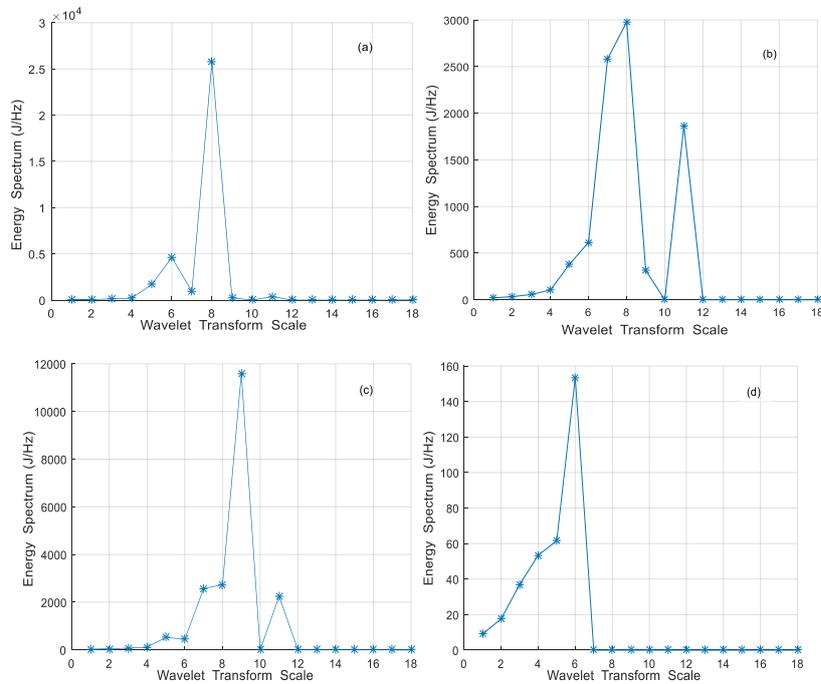
$$187 \quad E_j^A f(n) = \sum_{n=1}^N (A_j f(n))^2 \quad (13)$$

$$188 \quad E_j^D f(n) = \sum_{n=1}^N (D_j f(n))^2 \quad j = 1, 2, \dots, J \quad (14)$$

189 where,  $A_j f(n)$  is the low-frequency reconstructed signal at the  $J$ th wavelet decomposition, and  $D_j f(n)$  is the  
 190 high-frequency reconstructed signal at the  $J$ th wavelet decomposition.  $E_j^A f(n)$  and  $E_j^D f(n)$  are the low- and  
 191 high-frequency signal energies, respectively, at the  $J$ th wavelet decomposition.

192 Multiscale wavelet decomposition is used to extract feature vectors of the energy spectrum, and the wavelet energy  
 193 spectrum of railway subgrade and typical defects is established. According to the spectrum features and classification of  
 194 the wavelet decomposition, subgrade images require more than 13 orders of wavelet decomposition. The  
 195 high-frequency part is extracted by multiscale decomposition; thus the number of points is reduced.

196



197 **Fig.6** Multiscale wavelet energy spectrum of subgrade (a) the normal subgrade; (b) subgrade settlement; (c) subgrade sinkhole;  
 198 (d) mud pumping  
 199

200 The high-frequency energy spectrum of typical subgrade defects in the multiscale space of wavelet decomposition  
 201 is shown in **Fig.6**. Through analysis of the energy spectrum, it can be seen that the characteristic peaks of the normal

202 subgrade, sinkhole, and settlement are all 8, and the characteristic peak of mud pumping is 6. The energy spectrum of  
 203 the normal subgrade is as high as  $2.5 \times 10^{-4}$  J/Hz, and the energy spectrum of mud pumping is as low as 160 J/Hz. The  
 204 energy spectra of the two other subgrade defects are so similar that it is difficult to distinguish between them.

205 **4. RESULT ANALYSIS AND DISCUSSION**

206 **4.1 Target detection and identification process**

207 According to the different starting points of the feature extraction, target features can be divided into two categories:  
 208 the characteristics in the time domain are determined by the principal component analysis, the frequency and the energy  
 209 value of the sub block energy, the spatial characteristics of the lower dimensional space are obtained; the low frequency  
 210 information of the radar image and the energy spectrum of the wavelet multiscale spatial are acquired, and thus provide  
 211 a theoretical basis for the image feature extraction. Then according to the feature classification, the identification  
 212 process is as follows: the classification of railway subgrade defects is established, sparse dictionary and measurement  
 213 matrix are constructed according to the radar images, then sparse imaging of the detection area are carried out by CS  
 214 Algorithm, the pre-identification is carried out according to two-dimension wavelet, the remaining data are reduced 1/4.  
 215 According to the extracted features in the time domain and multiscale wavelet energy spectrum, the subgrade defects  
 216 are identified by. The flow is shown in Fig. 7.

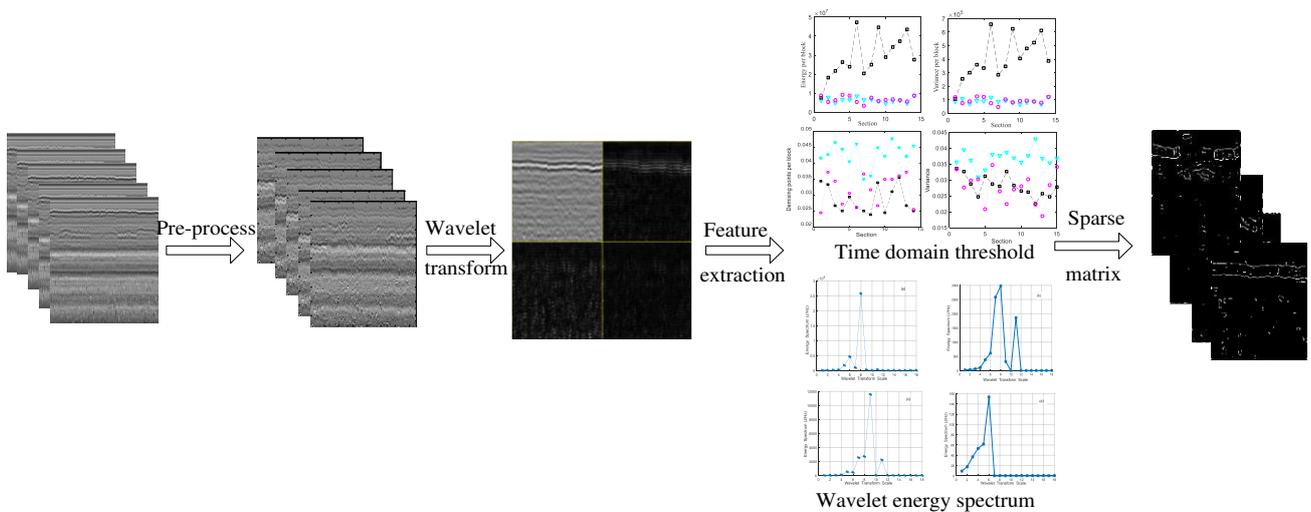


Fig.7 The flow of subgrade defects

219 The subgrade defects are sparse from the graphical distribution based on the CS algorithm. The identification

220 parameters of typical defects are as follows:

221 (1) Peaks of the multiscale wavelet energy spectrum of the subgrade.

222 (2) The time-domain features, such as energy per block, the variation per block, the variation and the demixing points  
223 per block.

#### 224 4.2 Identification of subgrade defects based on FCM

225 The 25 channels of the radar signals were selected as the image size. The coincidence rate between the images was  
226 50%. Accordingly, 100 MHz and 400 MHz radar antennas were adopted to detect the Daqin Railway subgrade. The  
227 working parameters of the GPR were set as follows: the interval of the sample was 5 cm, the depth resolution was up to  
228 0.2 m, and the rate was set to 100 scan/s. A total of 1084 subgrade sinkholes, 970 mud pumping defects, and 1534  
229 subgrade settlements were selected as the test samples.

230 Cluster analysis is multivariate analysis method in mathematical statistics that uses mathematical methods to  
231 quantitatively analyze and determine the relationship of samples to objectively categorize the types of things. However,  
232 the boundaries between things were sometimes blurred and sometimes exact. The boundary between the subgrade  
233 sinkhole and settlement was fuzzy, which led to the low accuracy of neural network technology.

234 FCM (fuzzy C-means) was used to determine the category by the degree of membership. The vector  $x_i (i=1, 2, 3, \dots, n)$   
235 was divided into  $c$  vague categories, and the cluster center of each category was found to minimize the objective  
236 function of the fuzziness index. The membership matrix elements  $u_{ij}$  of FCM ranged from (0, 1), and the comprehensive  
237 membership was normalized to 1:

$$238 \sum_{i=1}^c u_{ij} = 1, \quad \forall_j = 1, 2, \dots, n \quad (15)$$

239 The objective function of FCM was as follows:

$$240 J = \sum_{i=1}^n \sum_{j=1}^c (u_{ij})^m \|x_i - x_j\| \quad (16)$$

241 where  $m$  is the fuzzy weighted index, and  $v_j$  is the clustering center of the  $J_{th}$  class.

242 The FCM algorithm was as follows:

243 1) The subgrade defects were divided into three categories: sinkhole, mud pumping, and settlement, and the fuzzy  
244 weight index was determined;

245 2) The clustering center ( $v$ ) and individual fuzzy membership matrix ( $u$ ) of each kind of subgrade defect were set;  
246 thus, fuzzy clustering was analyzed;

247 3) According to the clustering, the subgrade defects were classified, the corresponding mean value center ( $v$ ) was  
248 obtained, and the distance between the individual (subgrade defect type) in the class and the mean value center was  
249 obtained; thus, the corresponding fuzzy clustering statistical results were obtained.

250 The FCM algorithm was adopted to analyze the test sample, and the results are shown in **Table 1**. As shown in **Table**  
251 **1**, the classification accuracy of mud pumping and settlement was 82.3% and 77.1%, respectively, and the classification  
252 accuracy of sinkholes was only 45.5%. The difference in settlement and sinkhole was small, and the distance between  
253 the individual (subgrade defect types) in the class and the mean value center was small; thus, the two defects cannot be  
254 classified accurately.

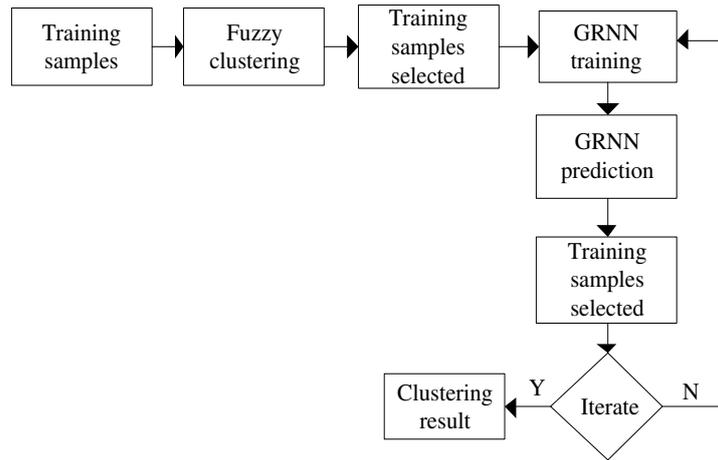
#### 255 **4.3 Identification of subgrade defects based on FCM-GANN**

256 The generalized regression neural network (GRNN) is a radial basis neural network and is a method of regression  
257 analysis of the nonindependent variables relative to the independent variables. The GRNN has a strong nonlinear  
258 mapping ability, fault tolerance and robustness, especially in learning speed and approximation ability. According to the  
259 fuzzy boundaries and considerable data of railway subgrade defects, FCM and GRNN algorithms were combined to  
260 identify the subgrade defects. The specific algorithm was as follows:

261 1) Based on the FCM, the GRNN was used to predict the type of training samples;

262 2) The corresponding mean value center ( $v$ ) and the distance between the individual (subgrade defect type) in the  
263 class and the mean value center were recalculated, and the data closest to the center were selected as the training  
264 samples of the network;

265 3) After repeated calculations, the final network cluster was obtained.



**Fig.8** The FCM-GRNN algorithm

266  
267

268 The detection samples were processed by combining them with FCM and GRNN according to the fuzzy division and  
269 large data of the railway subgrade defects. **Table 1** shows that railway subgrade defects were effectively identified by  
270 the FCM and FCM-GRNN algorithms. The settlement and sinkhole characteristics were similar. The classification  
271 boundary was not obvious. The recognition rate of the two kinds of defects was relatively low.

272 The effective classification and identification of settlement and mud pumping were performed by the FCM-GRNN  
273 algorithm, and the classification accuracy reached 100%. The classification accuracy of subgrade sinkholes was 59.1%,  
274 and compared with the 45.5% classification accuracy of the FCM, the classification recognition was improved to some  
275 extent.

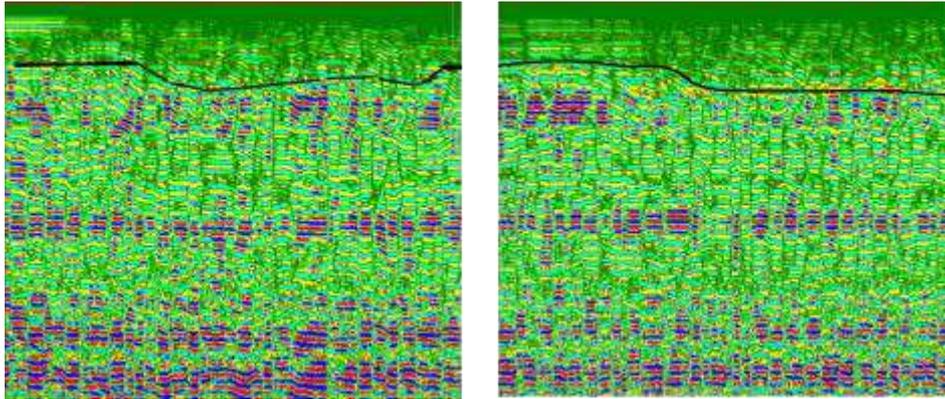
276

**Table 1** The accuracy of defects recognition in railway subgrade

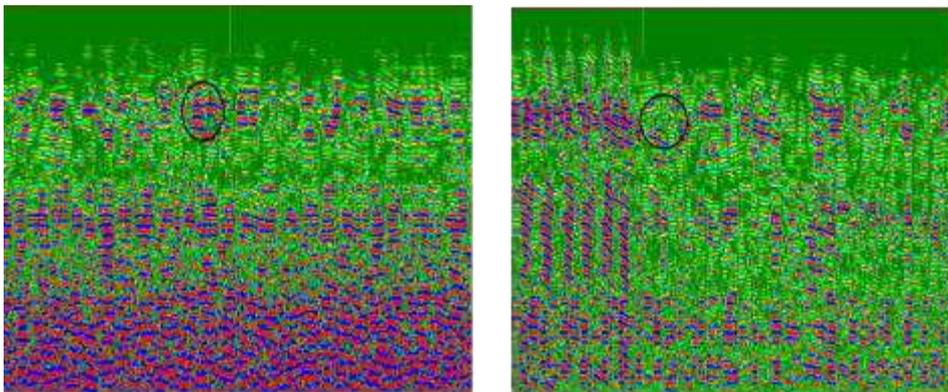
Input	Number	Output	Number		Classification Accuracy	
			FCM	FCM-GRNN	FCM	FCM-GRNN
Sinkhole	1084	Sinkhole	493	641		
		Settlement	443	345	45.5%	59.1%
		Mud Pumping	148	98		
Mud Pumping	970	Sinkhole	128	0		
		Settlement	44	0	82.3%	100%
		Mud Pumping	798	970		
Settlement	1534	Sinkhole	307	0		
		Settlement	1183	1534	77.1%	100%
		Mud Pumping	44	0		

277 **Table 1** showed that some results from Daqin railway by the recognition method. Practice proved that predict  
278 results were in good coincidence with drilling and underground workings materials.

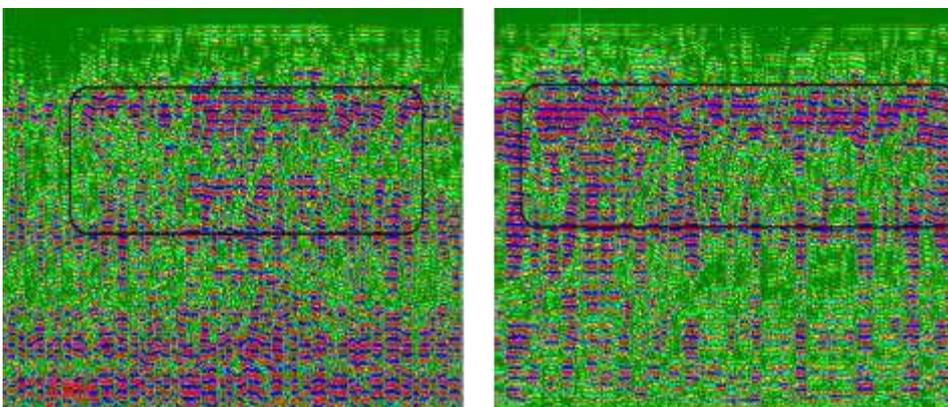
279 The images of railway defects are shown in **Fig. 9-11**. There were some defects, such as settlement, sinkhole and mud  
280 pumping, in the railway sungrades. The defects were inferred to be the remains of the artificial mining cave. Under  
281 long-term traffic loading and water erosion conditions, the structure of rock and soil was gradually destroyed, and its  
282 load-carrying capacity gradually decreased, eventually leading to collapse. The collapse was not filled completely to  
283 form a loose area, which resulted in subgrade settlement, and water-enriched regions formed mud pumping.



284  
285 Fig.9 GPR images of subgrade settlement



286  
287 Fig.10 GPR images of subgrade sinkhole



288  
289 Fig.11 GPR images of mud pumping

290 **5 Conclusion**

291 To reduce the redundant data and solve the difficulties in identifying subgrade defects, we combined the  
292 distribution of subgrade defects and GPR radar images; and constructed a method of sparse representation to identify  
293 subgrade defects. The following conclusions were as followed:

294 The feature extraction of typical radar images in subgrade defects was studied from the aspects of time domain and  
295 physical geometry. The optimal radar feature was established. Two-dimensional time-frequency features of the radar  
296 images in typical subgrade defects were analyzed, and the energy spectrum of the wavelet multiscale spatial was  
297 extracted. Sparse representations of subgrade defects were obtained.

298 Based on the features of subgrade defects, fuzzy C-means (FCM) and generalized regression neural networks  
299 (GRNNs) were used as the recognition algorithms for subgrade defects. The recognition rates of the subgrade defects all  
300 improved. Rapid identification of railway subgrade defects was realized.

301 The study combined theoretical analysis with field experiments and constructed a sparse representation method to  
302 identify the subgrade defects. The identification method overcame the influence of redundant data and promoted GPR  
303 application for the detection of subgrade defects. However, the sinkhole and settlement of railway subgrades were  
304 distinguished completely in the study. Hence, identification methods for subgrade defects should be further discussed.

#### 305 **Declarations:**

#### 306 **Computer code availability**

307 Source code implementing the algorithm described in this work can be obtained from the following git repository:  
308 <http://www.ilovematlab.com/>. This software is written in Matlab (version R2016a) . It should run on any modern Linux  
309 x86 computer supporting the aforementioned packages.

#### 310 **Competing interests**

311 The authors declare that they have no conflict of interest.

#### 312 **Funding**

313 This study was supported by Key R & D projects of Hebei Province: Research on cooperative perception and digital

314 operation technology of transportation infrastructure operation situation (19210804D).

### 315 **Authors' contributions**

316 Zhezhe Hou designed experiments and wrote the manuscript; Yanliang Du carried out experiments; Weigang Zhao and  
317 Yong Yang analyzed experimental results.

### 318 **Acknowledgements**

319 The authors express their thanks to Mr. Guolin Li, from Railway 19th Bureau Group Co., LTD, and graduate students  
320 Meng Zhao for help conducting of the tests.

### 321 **References**

- 322 [1] L. B. Ciampoli, A. Calvi, and F. D'Amico, "Railway Ballast Monitoring by GPR: A Test-Site Investigation," *Remote*  
323 *Sensing*, vol. 11, no. 20, Oct 2019, Art. no. 2381.
- 324 [2] L. Ciampoli, V. Gagliardi, C. Clementini, D. Latini, F. Del Frate, and A. Benedetto, "Transport Infrastructure  
325 Monitoring by InSAR and GPR Data Fusion," *Surveys in Geophysics*, vol. 41, no. 3, pp. 371-394, May 2020.
- 326 [3] J. Xiao and L. Liu, "Suppression of Clutters Caused by Periodic Scatterers in GPR Profiles With Multibandpass  
327 Filtering for NDT&E Imaging Enhancement," *Ieee Journal of Selected Topics in Applied Earth Observations and*  
328 *Remote Sensing*, vol. 10, no. 10, pp. 4273-4279, Oct 2017.
- 329 [4] A. Benedetto, F. Tosti, L. B. Ciampoli, A. Calvi, M. G. Brancadoro, and A. M. Alani, "Railway ballast condition  
330 assessment using ground-penetrating radar - An experimental, numerical simulation and modelling  
331 development," *Construction and Building Materials*, vol. 140, pp. 508-520, Jun 1 2017.
- 332 [5] A. K. Khamzin, A. V. Varnavina, E. V. Torgashov, N. L. Anderson, and L. H. Sneed, "Utilization of air-launched  
333 ground penetrating radar (GPR) for pavement condition assessment," *Construction and Building Materials*, vol.  
334 141, pp. 130-139, Jun 15 2017.
- 335 [6] Z. Tong, J. Gao, and H. Zhang, "Innovative method for recognizing subgrade defects based on a convolutional  
336 neural network," *Construction and Building Materials*, vol. 169, pp. 69-82, Apr 30 2018.
- 337 [7] F. Tosti, L. B. Ciampoli, A. Calvi, A. M. Alani, and A. Benedetto, "An investigation into the railway ballast  
338 dielectric properties using different GPR antennas and frequency systems," *Ndt & E International*, vol. 93, pp.  
339 131-140, Jan 2018.
- 340 [8] L. B. Ciampoli, F. Tosti, M. G. Brancadoro, F. D'Amico, A. M. Alani, and A. Benedetto, "A spectral analysis of  
341 ground-penetrating radar data for the assessment of the railway ballast geometric properties," *Ndt & E*  
342 *International*, vol. 90, pp. 39-47, Sep 2017.
- 343 [9] S. Liu, Q. Lu, H. Li, and Y. Wang, "Estimation of Moisture Content in Railway Subgrade by Ground Penetrating  
344 Radar," *Remote Sensing*, vol. 12, no. 18, Sep 2020, Art. no. 2912.
- 345 [10] S. S. Artagan, L. B. Ciampoli, F. D'Amico, A. Calvi, and F. Tosti, "Non-destructive Assessment and Health  
346 Monitoring of Railway Infrastructures," *Surveys in Geophysics*, vol. 41, no. 3, pp. 447-483, May 2020.
- 347 [11] S. S. Artagan and V. Borecky, "Advances in the nondestructive condition assessment of railway ballast: A focus  
348 on GPR," *Ndt & E International*, vol. 115, Oct 2020, Art. no. 102290.
- 349 [12] B. E. Barrett, H. Day, J. Gascoyne, and A. Eriksen, "Understanding the capabilities of GPR for the measurement  
350 of ballast fouling conditions," *Journal of Applied Geophysics*, vol. 169, pp. 183-198, Oct 2019.

- 351 [13] J. Sadeghi, M. E. Motieyan-Najar, J. A. Zakeri, B. Yousefi, and M. Mollazadeh, "Improvement of railway ballast  
352 maintenance approach, incorporating ballast geometry and fouling conditions," *Journal of Applied Geophysics*,  
353 vol. 151, pp. 263-273, Apr 2018.
- 354 [14] A. Lbc, A. Ac, and A. J. T. R. P. Eo, ""Test-site operations for the health monitoring of railway ballast using  
355 Ground-Penetrating Radar" - ScienceDirect," vol. 45, pp. 763-770, 2020.
- 356 [15] Y. Li, Z. Zhao, W. Xu, Z. Liu, and X. Wang, "An effective FDTD model for GPR to detect the material of hard  
357 objects buried in tillage soil layer," *Soil & Tillage Research*, vol. 195, Dec 2019, Art. no. 104353.
- 358 [16] J. Zhang *et al.*, "In-situ recognition of moisture damage in bridge deck asphalt pavement with time-frequency  
359 features of GPR signal," *Construction and Building Materials*, vol. 244, May 30 2020, Art. no. 118295.
- 360 [17] Y. Yang and W. Zhao, "Curvelet transform-based identification of void diseases in ballastless track by  
361 ground-penetrating radar," *Structural Control & Health Monitoring*, vol. 26, no. 4, Apr 2019, Art. no. e2322.
- 362 [18] A. Alsabhan, D. Fratta, B. J. Warren, J. M. Tinjum, and T. B. Edil, "Using Time Domain Reflectometry to  
363 Determine Depth of Fouling and Fouling Type in Railway Track Substructure," *Geotechnical Testing Journal*, vol.  
364 42, no. 1, pp. 156-179, Jan 2019.
- 365 [19] W. Shao, A. Bouzerdoun, S. L. J. I. T. o. G. Phung, and R. Sensing, "Sparse Representation of GPR Traces With  
366 Application to Signal Classification," vol. 51, no. 7, pp. 3922-3930, 2013.
- 367 [20] W. Shao, A. Bouzerdoun, S. L. Phung, L. Su, C. J. I. T. o. G. Rujikiatkamjorn, and R. Sensing, "Automatic  
368 Classification of Ground-Penetrating-Radar Signals for Railway-Ballast Assessment," vol. 49, no. 10, pp.  
369 3961-3972, 2011.
- 370 [21] H. Zhezhe, Z. Weigang, and D. J. B. T. T. B. Yanliang, "Feature extraction for ballasted railway subgrade defects  
371 based on sparse representation," vol. 55, no. 9, pp. 766-778, 2017.
- 372
- 373