



Deposited via The University of Leeds.

White Rose Research Online URL for this paper:

<https://eprints.whiterose.ac.uk/id/eprint/183177/>

Version: Accepted Version

Article:

Chen, J, Huang, H, Cohn, AG et al. (2022) A hierarchical DCNN-based approach for classifying imbalanced water inflow in rock tunnel faces. *Tunnelling and Underground Space Technology*, 122. 104399. ISSN: 0886-7798

<https://doi.org/10.1016/j.tust.2022.104399>

© 2022, Elsevier. This manuscript version is made available under the CC-BY-NC-ND 4.0 license <http://creativecommons.org/licenses/by-nc-nd/4.0/>.

Reuse

This article is distributed under the terms of the Creative Commons Attribution-NonCommercial-NoDerivs (CC BY-NC-ND) licence. This licence only allows you to download this work and share it with others as long as you credit the authors, but you can't change the article in any way or use it commercially. More information and the full terms of the licence here: <https://creativecommons.org/licenses/>

Takedown

If you consider content in White Rose Research Online to be in breach of UK law, please notify us by emailing eprints@whiterose.ac.uk including the URL of the record and the reason for the withdrawal request.

A hierarchical DCNN-based approach for classifying imbalanced water inflow in rock tunnel face

Jiayao Chen^{a,b}, Hongwei Huang^a, Anthony G. Cohn^{b,c,d,e,f}, Mingliang Zhou^{a,1}, Dongming Zhang^a, Jianhong Man^a

^a Key Laboratory of Geotechnical and Underground Engineering of Ministry of Education and Department of Geotechnical Engineering, Tongji University, Shanghai 200092, China

^b School of Computing, University of Leeds, LS2 9JT Leeds, United Kingdom

^c Department of Computer Science and Technology, Tongji University, Shanghai 211985, China

^d School of Civil Engineering, Shandong University, Jinan 250061, China

^e Luzhong Institute of Safety, Environmental Protection Engineering and Materials, Qingdao University of Science and Technology, Zibo 255000, China

^f School of Mechanical and Electrical Engineering, Qingdao University of Science and Technology, Qingdao 260061, China

Abstract: Accurate water inflow assessment in the under-construction rock tunnel sites is critical for the next optimized construction and rehabilitation strategy. In this paper, a deep convolutional neural networks (DCNN)-based method, named H-ResNet-34, is implemented to classify water inflow category from rock tunnel faces in under-construction highway tunnels in Yunnan, China. An image database is compiled, which contains 8,000 images in five different water inflow categories of rock tunnel faces, namely complete dry (CD), wet state (WS), dripping state (DS), flowing state (FS) and gushing state (GS). Herein, a crucial issue is the imbalanced images between damage and non-damage owing to the vast sample of datasets and between various damages due to varying damage occurrence rates, which bring enormous challenges for conventional DCNN models. Thus, a hierarchical classification structure is applied to overcome the issue of imbalanced images at two different levels: coarse-level and fine-level. The coarse-level distinguishes the dataset with non-damage (i.e. complete dry) images. The fine-level computes the occurrence probability of the image dataset with water inflow damage. The constructed framework is then trained, validated, and tested using tunnel face images with various water inflow categories. The testing results suggest that the proposed hierarchical classifier is well competent for water inflow classification for rock tunnel face images and can effectively alleviate the imbalanced data issue.

Keywords: Water inflow, Rock tunnel, Image classification, Imbalanced images, Deep convolutional neural network

1. Introduction

The assessment of water inflow is critical for the final classification of surrounding rock in rock tunnels under construction, owing to its significant impact on constructors and managers in case of tunnel collapse and water gushing accidents. It also provides a significant basis for continuing project strategies under limited construction schedules and engineering budgets. In general, the manual water inflow inspection approaches (e.g., tipping buckets, discharge vessels, weirs, etc) (Rálek and Hokr, 2013), which are widely employed under current practice (Hwang and Lu, 2007). Although correct flow rate values can be obtained by these contact manners, they are labour- and time-consuming, and even threatening the safety of engineers (Fernandez and Moon, 2010). Thus, there is an urgent to explore a vision-based inspection method that can identify the water inflow categories accurately in rock tunnel faces (Chen et al., 2021d).

From the perspective of engineering, visual inspection consists of human vision and machine vision.

Received 22 November 2019




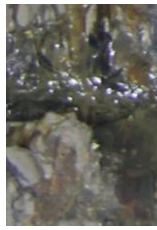

Received in revised form 19 January 2022

¹ Corresponding author.

E-mail address: zhoum@tongji.edu.cn

41 Herein, machine vision-based inspection is designed for seeking to understand and automate tasks that the
 42 human visual system can do. The rock mass rating (RMR) is one of the most internationally recognized
 43 discrimination methods (Bieniawski, 1988), which uses the measured water inflow rate to classify the
 44 following five water inflow categories, namely completely dry, damp, wet, dripping, and flowing (listed in
 45 Table 1) (Warren et al., 2016). It is hard to distinguish the damp state and wet state by naked eyes since the
 46 rate range of damp state is less than 10 liters/min, and that of wet is 10-25 liters/min, which can only be
 47 differentiated by the flowmeter measurement. Hence, for the vision-based task, damp and wet are unified into
 48 the wet category to perform the manual image tagging. Additionally, a water gushing state is added as one of
 49 the water inflow categories because of its sudden and disastrous nature in the rock tunnelling project. Typical
 50 tunnel face images of the five water inflow categories are shown in the corresponding columns in Table 1. To
 51 provide a comprehensive assessment of water inflow status, a full surface inspection of consecutive multiple
 52 tunnel faces is essential (Cai et al., 2022; Man et al., 2022; Zarei et al., 2013). Nevertheless, the remaining
 53 challenge for the field engineers is to establish an efficient identification method of water inflow under
 54 frequently changing construction processes (Zarei et al., 2012).

55
 56 **Table 1.** Water flow statistics and example images based on rock mass rating (RMR)

Water inflow	None	< 10 liters/min	10-25 liters/min	25-125 liters/min	>125 liters/min	Tunnel disaster
General conditions in RMR	Completely dry	Damp	Wet	Dripping	Flowing	Gushing
Classification in this study	CD	Wet		Dripping	Flowing	Gushing
Example images						

57
 58 A representative traditional visual inspection method is the geological sketch method, which is a labour-
 59 consuming task with strong subjectivity, and heavily relies on the experience and meticulousness of the
 60 inspectors (Santos et al., 2018; Sou-Sen Leu, 2011). Inspectors have to tolerate the safety risks caused by close
 61 contact with the tunnel working face to conduct the inspection and then detect potential regions of damage
 62 through stop-and-check (Cai et al., 2021). However, geological sketch methods are still widely used in many
 63 countries due to the limitations of project budgets and computational technology (Li et al., 2017; Marjoribanks,
 64 2010).

65 Geological radar detection is another typical method used to identify damage by analysing medium
 66 reflection signals, which can quantify damage with a high level of automation by professional personnel, and
 67 is excellent at detecting abnormal geological environments (Annan et al., 1991; Guo et al., 2019; Koopialipoor
 68 et al., 2019b; Liu et al., 2010). However, this method urgently needs to be strengthened in terms of its
 69 sensitivity to geological hazards with little change and to simultaneously and efficiently acquire multiple
 70 comprehensive types of information, such as joints and fractures, and ground water (Chen et al., 2021a).

71 The latest trend is to employ computer vision methods for automated task classification (Chen et al.,
 72 2021b; Kumar et al., 2018; Nhat-Duc et al., 2018). The traditional image processing method is initially used

73 to extract thousands of feature parameters to obtain the target characteristics with relatively backward
74 computing equipment, to pre-design feature extractors and pre-process images before training (Chen et al.,
75 2021a; Chen et al., 2021e). In this regard, the operational efficiency and the friendliness of the developed
76 frameworks are undoubtedly reduced, and thus an efficient and high-precision image detection method is
77 urgently needed.

78 Deep convolutional neural networks (DCNNs) have shown admirable end-to-end performance for visual
79 detection tasks and can learn abstract features and reveal the rules of input and outputs by self-deep learning
80 (Chen et al., 2020; Zhao et al., 2021). This method has been widely used with images for classification and
81 detection tasks. DCNNs are made up of neurons for different functions that learn a large number of significant
82 parameters such as weights and biases, while input data is transformed into output data (Huang et al., 2020).
83 At present, civil engineering fields using DCNNs are primarily concentrated on the foundation pit (Fang et al.,
84 2018), buildings (Martinez-Murcia et al., 2018; Nhat-Duc et al., 2018), shield tunnels (Huang et al., 2020;
85 Zhou et al., 2021), and municipal utility engineering (Kumar et al., 2018), providing significant practical
86 experience for the intellectualization and informatization of engineering construction. However, the existing
87 research mainly focused on the operation and maintenance stages to facilitate management and repair. Few
88 studies applied DCNNs in the under-construction site, especially for the underground excavation rock face (Lü
89 et al., 2017).

90 For addressing the deficiencies as mentioned above, a digital photography method is proposed and
91 adopted to obtain raw water inflow images (3968×2240 pixels) of the rock tunnel face in batches from
92 highway tunnels under construction in Yunnan, China. Five water inflow categories (complete dry (CD), wet
93 state (WS), dripping state (DS), flowing state (FS), and gushing state (GS)) were classified manually to
94 establish the target image dataset. In general, the proportion of damage-free images is the largest, while the
95 distribution of other damage images is exceptionally imbalanced, bringing the risk of over-fitting or under-
96 fitting to the test results. Hence, a DCNN method (i.e. H-ResNet-34) employing a residual module (He et al.,
97 2016) as the backbone framework and a hierarchical classification structure (Seo and Shin, 2019) as the
98 multiple level classification structures is proposed to enhance the efficiency and accuracy over that of the
99 imbalanced datasets. A resized image (229×229 pixels) dataset is then created, trained, validated, and tested
100 to generate the optimal target model for water inflow classification. The water inflow classification
101 performances of the proposed H-ResNet-34 method and the original ResNet-34 method are systematically
102 assessed with regards to the evaluation metrics and visualization methods.

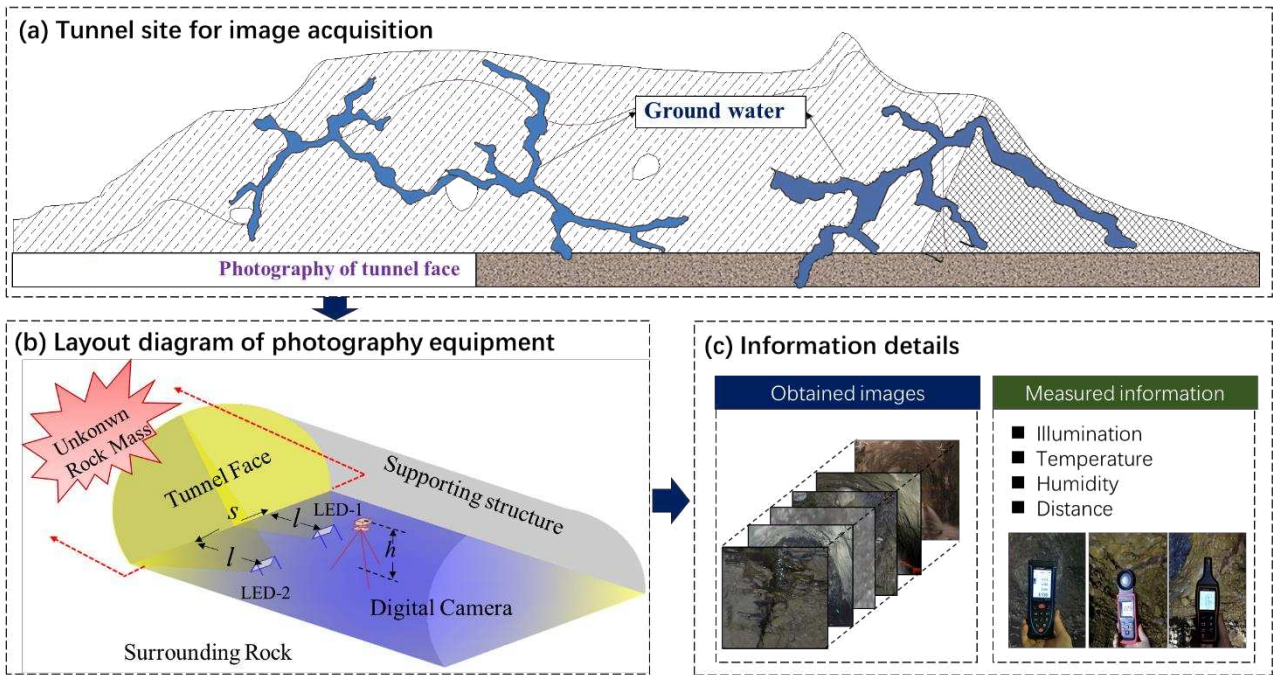
103

104 **2. Water Inflow ImageNet**

105 *2.1. Image database collection*

106 Inspired by the establishment and employment of the target ImageNet in DCNN (Deng et al., 2009), a
107 Water Inflow ImageNet (WIIN) was built consisting of rock tunnel face images relevant to five different water
108 inflow status: complete dry (CD), wet state (WS), dripping state (DS), flowing state (FS) and gushing state
109 (GS). The WIIN database is used for identification and detection of water inflow problems in rock tunnel
110 projects. To construct such a database, a digital photograph method (as shown in Fig.1, consisting of a digital
111 camera, tapeline, tripod, light source, and measuring equipment namely laser rangefinder, thermo hygrometer,
112 and illuminometer) is proposed for image acquisition, which can cover a variety of different rock tunnel faces.
113 The limited lighting conditions and the cramped surrounding environment in the rock tunnel raise significant
114 challenges for the acquisition of quality images. For improving the quality of images, two adjustable power
115 drop LED lamps were used to increase the illumination for photographing. Additional tunnel face images were
116 also collected from search engines like Baidu and Google to increase the image samples in the dataset, which

117 account for approximately 20% of the total samples.



118

119 **Fig. 1.** Schematic diagram of the digital photograph method, including: (a) tunnel site for image acquisition,
 120 (b) layout diagram of photography equipment, and (c) information details.
 121

122

2.2. Imbalanced image problem

123

124

125

126

127

128

129

130

131

132

133

134

135

136

137

138

139

140

141

142

In this study, 4012 water inflow images from various rock tunnel faces were classified manually into five categories. Sample images of the five categories of water inflow, namely CD, WS, DS, FS, and GS, are shown in Fig. 2. Because of the complex and changeable geological conditions in the highway rock tunnel, the damage texture and grey-scale distribution differences of water inflow status of various categories do not differ significantly, leading to inefficiency in the manual classification. The number of original water inflow images in each category is listed in Table 2, where the number distribution of the images is extremely imbalanced. The raw images include 2,102 non-damage (i.e. complete dry) images and 1,910 images with water inflow damage. Approximately 52% of the original images are complete dry state images without water inflow, which is three times larger than the second-largest sample size, WS, and twenty-one times larger than the smallest sample size, GS. Furthermore, the imbalance issue also exists among the images with water inflow damage. The number of WS images is seven times that of GS. Nevertheless, the issues become more challenging for multi-class classification tasks due to the existence of several minority classes. The proposed method has to balance between guaranteeing the classification rate for water inflow with typical distinctive sample scales and avoiding overfitting the minority categories. In most cases, the imbalance between different categories influences both the convergence of the training and validation processes and the generalization of the pre-trained framework on the test dataset. Hence, the imbalance problem invariably leads high classification accuracy for the majority categories, and results in low classification accuracy for the minority categories (Huang et al., 2016; Khan et al., 2017; Koopialipoor et al., 2019a; López et al., 2013).

Table 2. Water inflow categories and image dataset statistics.

Water inflow category	CD	WS	DS	FS	FS
Number of images	2,102	775	733	304	98

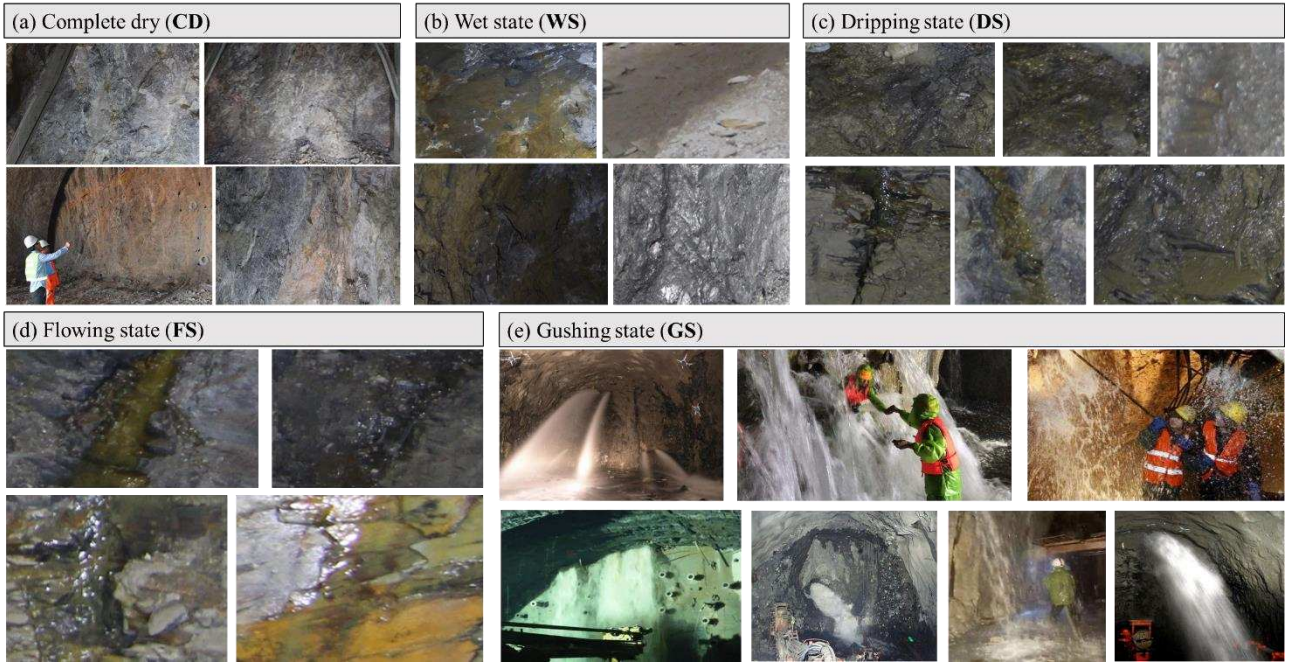


Fig. 2. Sample images for different water inflow categories: (a) complete dry (CD), (b) wet state (WS), (c) dripping state (DS), (d) flowing state (FS), and (e) gushing state (GS)

2.3. Database establishment

For relieving the imbalance issue between the images of different water inflow categories, the images with minority damage were oversampled rather than under-sampled since the over-sampling processes are proved to have better performance (López et al., 2013). In this section, some data augmentation techniques such as random blur, local amplification, random horizontal flip, Gaussian sampling, and channel scaling were selected on the raw images. Moreover, samples of minority damages were oversampled multiple times to match the balance of the proposed model. In total, 8,000 images were generated, as shown in Table 3, where the statistics of each water inflow status in the mentioned datasets are listed. To simplify analysis, the number of damaged category samples is consistent with each other, and the total number of damaged samples is equal to the non-damaged samples. Meanwhile, the proportion of training, validation, and testing dataset in each label is approximately adjusted to 60%, 25%, and 15%.

Table 3. The image numbers of training, validation, and testing datasets for different water inflow categories

Water inflow category	Training	Validation	Testing	Total number
CD	2,400	1,000	600	4,000
WS	600	250	150	1,000
DS	600	250	150	1,000
FS	600	250	150	1,000
GS	600	250	150	1,000
Total number	4,800	2,000	1,200	8,000

3 The Proposed DCNN Method

The DCNN methods have achieved excellent performance in image classification tasks, such as VGG (Chen et al., 2021c; Simonyan and Zisserman, 2014), Inception (Szegedy et al., 2015), etc. However, as the

165 network depth increases, the performance of DCNNs gradually becomes saturated or even declines rapidly,
 166 which is known as the degradation problem of a network. To address these issues, a residual learning network
 167 (ResNet) (He et al., 2016) has shown considerably superior performance on detection rate and accuracy from
 168 the increased network depth. Meanwhile, it is much easier to optimize and modify the framework using a
 169 residual learning network instead of an unreferenced network. This study proposes a residual learning network
 170 with 34 layers (ResNet-34) as the typical backbone network to classify the water inflow images.

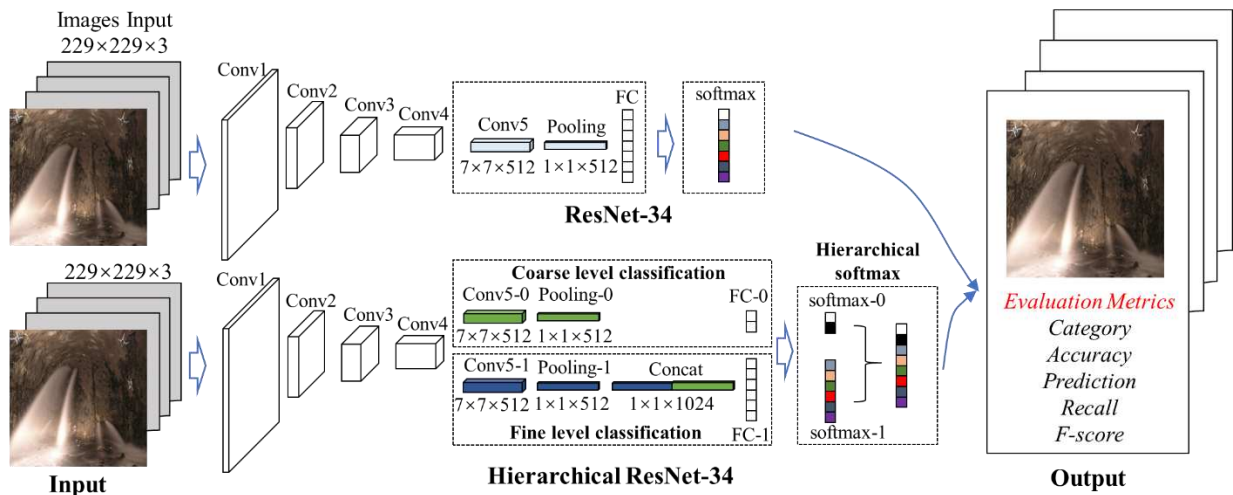
171 Furthermore, a DCNN with a hierarchical classification structure, named as H-ResNet-34, is proposed in
 172 this study to handle the afore-mentioned imbalanced water inflow image issues. Remarkably, the original
 173 ResNet-34 was modified in this study to handle hierarchical identification at both coarse-level and fine-level,
 174 that is, the H-ResNet-34 model. The coarse-level task of the framework belongs to a simple binary
 175 classification issue that classifies the damage samples from the non-damage datasets. The fine-level task then
 176 predicts the probability of each damage category by assuming at the images have damages. The hypotheses
 177 of the proposed H-ResNet-34 are that it is more efficient to classify a few categories within one category than
 178 all of the categories, and it is more efficient to train a model with balanced samples than imbalanced samples.
 179 The coarse prediction can enhance the prediction by providing the obtained fine-level features. Finally, the
 180 proposed hierarchical classifier is used to integrate the multi-level predictions to produce the final prediction
 181 results.

182

183 3.1. Base ResNet-34 model

184 As displayed in Fig. 3, the original ResNet-34 framework consists of five main convolution modules,
 185 which include a total of 33 convolution layers, an average pooling layer, and a fully connected (FC) layer. In
 186 order to promote the image processing efficiency, the raw images were cropped from their original sizes of
 187 3968×2240 pixels into smaller resized images of 229×229 pixels. The first convolution module consists of
 188 a single convolution layer with 64 filters of 7×7 , a stride of 2, and a padding of 3. The remaining convolution
 189 module consists of two basic residual modules.

190



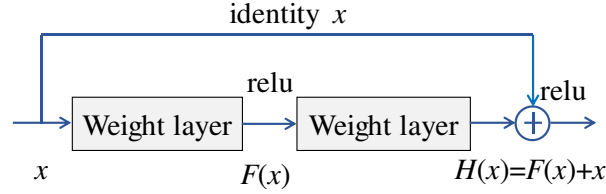
191

192 **Fig. 3.** The main structure of the proposed DCNN for water inflow classification, including the original
 193 ResNet-34 (on the top), and the H-ResNet-34 (at the bottom).
 194

195

196 The residual module shown in Fig. 4 consists of two convolution layers with the same number of 3×3
 filters. Each convolution module can decrease the size of the images by half and increase the feature scale in

197 a specific range. Unlike traditional CNNs, each pair of 3×3 filters adds shortcut connection, which is applied
 198 to skip particular layers and pass raw data directly to the next layer. These new shortcut connections will not
 199 increase the parameters and the complexity of the original model, moreover the whole model can still be trained
 200 using an end-to-end approach. The configuration of each convolution module is listed in Table 4, where the
 201 final output size is 512 dimensional 1×1 filter.



202
203 **Fig. 4.** Residual learning: a building module.

204
205 **Table 4.** H-ResNet-34 layers and configurations

Layer	Cov1	Cov2_x	Cov3_x	Cov4_x	Cov5_x	Pooling
Output size	112×112×64	56×56×56	28×28×128	14×14×256	7×7×512	1×1×512
Filters	7×7,64 stride 2	$\begin{bmatrix} 3 \times 3, 64 \\ 3 \times 3, 64 \end{bmatrix} \times 3$	$\begin{bmatrix} 3 \times 3, 128 \\ 3 \times 3, 128 \end{bmatrix} \times 4$	$\begin{bmatrix} 3 \times 3, 256 \\ 3 \times 3, 256 \end{bmatrix} \times 6$	$\begin{bmatrix} 3 \times 3, 512 \\ 3 \times 3, 512 \end{bmatrix} \times 3$	Average

206
207 Although DCNNs have achieved significant success on established benchmark datasets, they are still
 208 unable to get rid of the negative impacts caused by the imbalanced image dataset. The solution to the
 209 imbalanced issue can be divided into data-level and classifier-level methods (Buda et al., 2018; Gordan et al.,
 210 2016; Huang et al., 2016; Momeni et al., 2015). Among these, the primary process used in data-level methods
 211 is resampling, which attempts to balance the images in each category by under-sampling the main category or
 212 over-sampling the minority category. On the other hand, the classifier-level method aims to handle tasks by
 213 specific algorithms, including thresholding and ensemble learning (Buda et al., 2018; Hajihassani et al., 2015;
 214 Khan et al., 2017; Zhang et al., 2021). Therefore, a hierarchical DCNN framework, in combination with a
 215 resampling process, is adopted in this study.

216 Previously, several tentative hierarchical recognition algorithms (Yan et al., 2015) have shown excellent
 217 performance in visual detection on non-damage datasets. According to the hypothesis mentioned above,
 218 solving a relatively balanced binary identification issue that detects damage from a non-damage dataset is more
 219 efficient than directly detecting each individual type of damage. Undoubtedly, identifying images containing
 220 water inflow from a dataset consists of tunnel face images with water inflow damage is more efficient
 221 compared to detecting them from a dataset where both non-damage and damage images coexist. Thus, as
 222 illustrated in Fig. 3, a branch module similar to ‘convolution 5’ was built following ‘convolution 4’ to construct
 223 a hierarchical module in the original ResNet-34 framework.

224 The proposed identical modules in the model, named ‘convolution 5-0’ and ‘convolution 5-1’, were used
 225 for the coarse-level identification (binary classification) and the fine-level in recognition (classification of
 226 specific water inflow categories), respectively. Furthermore, the low-level features obtained by the first four
 227 convolution modules were shared and then employed as a significant basis for damage detection with the
 228 ‘convolution 5’ module. Thus, the features acquired from the comparatively balanced coarse-level
 229 identification were employed for the further fine-level task. Finally, the obtained coarse-level information after
 230 the average pooling layer was combined with the fine-level information to determine the final damage
 231 recognition.

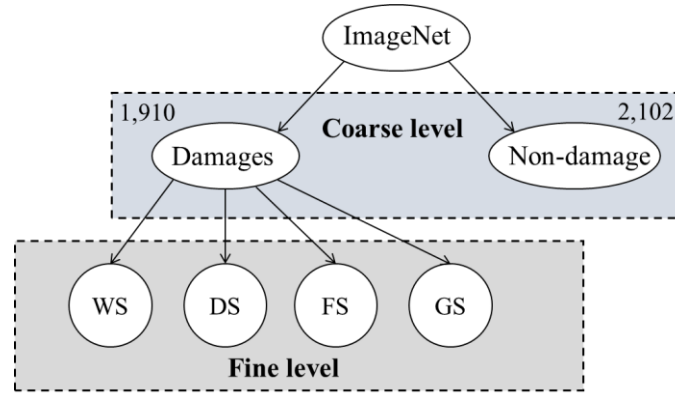
232

233 3.2. Hierarchical ResNet-34 model (H-ResNet-34)

234 In conventional machine learning methods, predicting the probability value P_j that one target image
 235 belongs to a specific class j , requires the calculation of the normalized value for each category. The value of
 236 P_j is computed by using a certain softmax function to obtain the unnormalized value Z_j for the corresponding
 237 category j . Then, the traditional models employ the softmax function to the last layer with T categories, as
 238 shown in Eq. (1):

239
$$P_c = \frac{e^{Z_c}}{\sum_{j=1}^T e^{Z_j}} \quad (c \in 1, 2, \dots, T) \quad (1)$$

240 In the early developments, the hierarchical structure was proposed as a natural language categorization
 241 task for the original objective to lessen the calculation cost for forecasting using large vocabularies (Morin and
 242 Bengio, 2005). It has also shown excellent performance in the classification of uncommon words. This study
 243 employed the hierarchical classification structure to solve the issue of imbalanced images. Thus, the water
 244 inflow, as shown in Fig. 5, is classified at two levels: the coarse-level and the fine-level. In the coarse level,
 245 the pre-trained images are categorized as normal images versus damage images to detect the potential damage.
 246 In the fine level, the selected images with potential damage are further classified by learning the details of
 247 different types of water inflow damage.



248
 249 **Fig. 5.** Hierarchical classification of water inflow with a balanced coarse level and fine level dataset.

250
 251 The probability values of each image with water inflow damage P_1 or not P_0 are obtained from the softmax-
 252 0 algorithm in the coarse-level recognition. As for the fine-level, the conditional probability $P(d_j|1)$ of
 253 individual damage d_j with softmax-1 is predicted by assuming that the images in the dataset all contain the
 254 water inflow damage. Finally, the probability values of each type of water inflow damage are computed using
 255 Eq. (2):

256
$$P_{d_j} = P(d_j|1) \cdot P_1 \quad (2)$$

257 whereas P_1 equal to 1 when it is recognized as water inflow damage in coarse recognition. By this approach,
 258 the hierarchical module can classify the water inflow category, where the final-level prediction is determined
 259 by both the coarse-level and fine-level prediction (as shown in Table 5).

260
 261 **Table 5.** Definition of prediction probability of hierarchical structure for water inflow classification.

Prediction & Classifiers		Coarse-level	Fine-level	Final-level
		softmax-0	softmax-1	H-softmax
Categories	CD	P_0		P_0

WS		P_{WS}	$P_1 P_{WS}$
DS	P_1	P_{DS}	$P_1 P_{DS}$
FS		P_{FS}	$P_1 P_{FS}$
GS		P_{GS}	$P_1 P_{GS}$

262

263

264 4. Experiment and results

265

266

267

268

269

270

271

272

273

274

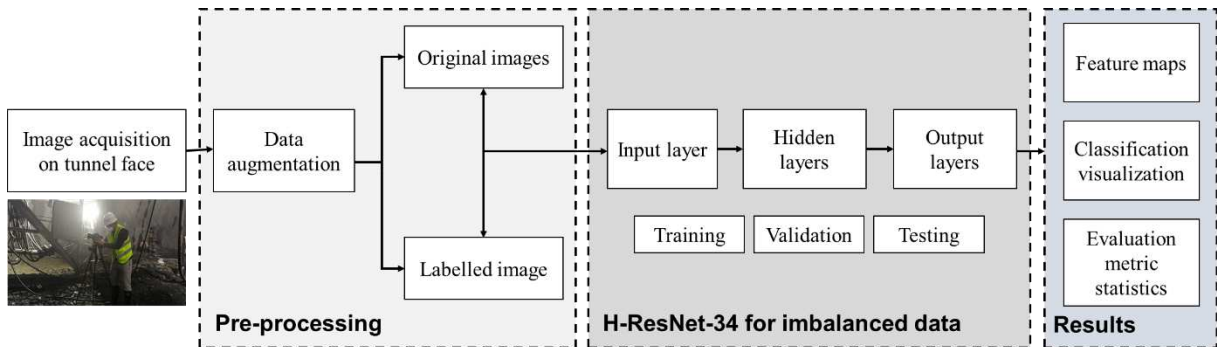
275

276

277

The proposed H-ResNet-34 framework was trained with Tensorflow (a deep learning engine specialized in CNN methods) in this study. A workstation implemented with Intel Core i7-8700 processor @3.70GHz, Nvidia GTX 1080 Ti 11GB GPU, Windows 10 operating system was employed in this research. The original ResNet-34 framework was also computed as a baseline to inspect the superiority and practicability of the proposed network. For the two DCNN methods, the epoch was terminated at 50 times, the initial learning rate was used as 0.0001, and the momentum of stochastic gradient descent (SGD) is set to 0.9.

Fig. 6 shows the main process experiment of the water inflow classification, that is, the full-cycle from data acquisition to visualization results. In the aspect of dataset preparation, it mainly contains four processes: field acquisition, image filtering and clipping, data augment, and manual classification of samples. Then the dataset goes through the training, validation, and testing processes of the two DCNN methods. Meanwhile, evaluation metrics are computed to evaluate the two DCNN methods. Finally, the two methods are visualized through feature maps, random selected image classification, and confusion matrix to demonstrate the corresponding performances.



278

279 **Fig. 6.** The main process experiment of the water inflow classification

280

281 4.1 Evaluation metrics of experiment

282

283

284

285

286

287

The evaluation metrics, namely total loss, accuracy, precision, recall, and F-score are frequently applied in the training, validation, and testing processes to evaluate the performance of the proposed classification methods. The correlation between the evaluation metrics and the essential metrics (i.e., true positive (TP), true negative (TN), false positive (FP), and false negative (FN)) are presented in Eqs. (3-6).

The accuracy metric is the proportion of accurately classified samples in all tasks, which can be computed as follows:

288

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \quad (3)$$

289

290

The precision metric is the proportion of true positives in all the samples marked as positive. It is calculated as follows:

291

$$Precision = \frac{TP}{TP + FP} \quad (4)$$

292 The recall metric presents the proportion of true positive in all positive samples, that is, the reflection of
 293 samples correctly labelled as positive in the target:

$$294 \quad \text{Recall} = \frac{TP}{TP + FN} \quad (5)$$

295 The F-score metric is a comprehensive indicator calculated by a specific way between recall and precision:

$$296 \quad F_{\alpha} = \frac{(\alpha^2 + 1)Precision \times Recall}{\alpha^2(Precision \times Recall)} \quad (6)$$

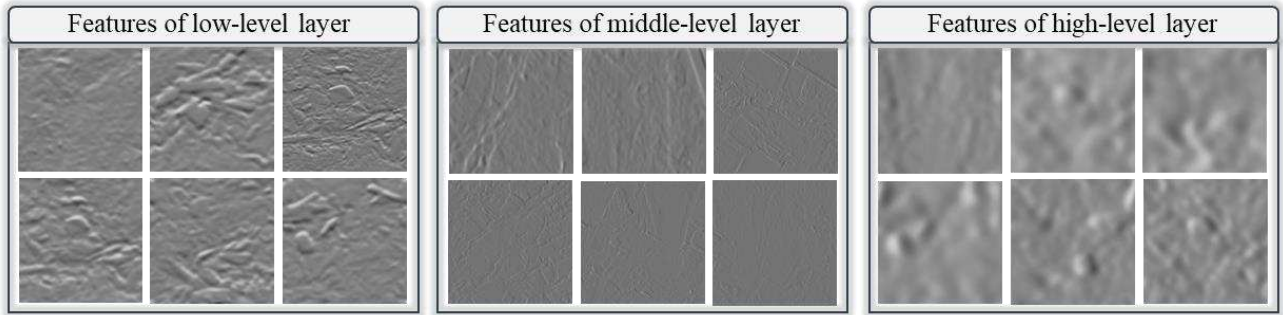
297 where α value is set as 1 in this study, which reflects the significance of recall and precision is the same.

298

299 4.2. Training and validation results

300 In both DCNN methods, 4,800 and 2,000 water inflow images are selected for the processes of training
 301 and validation, respectively. To further reveal the feature extraction process during the training process, the
 302 feature maps of water inflow images are shown in Fig. 7. Among them, basic features, such as corner, texture,
 303 and edge, are extracted through the low-level features. The middle-layer then checks the motifs by observing
 304 the arrangements of the low-level image features. As a result, more feature combinations can be assembled
 305 from the motifs by the high-level layer, and the water inflow categories can then be classified.

306

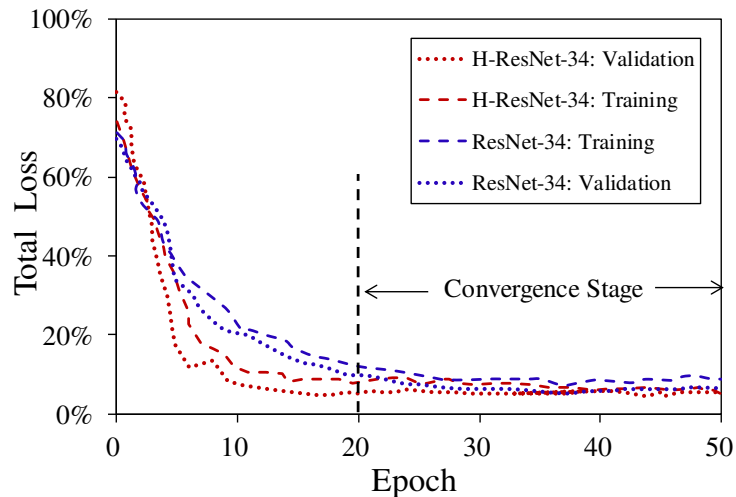


307

308 **Fig. 7.** Feature maps of water inflow images during the training process generated by the DCNNs.

309

310 The total loss curves are plotted in Fig. 8 to quantitatively measure the fitting degree and convergence of
 311 the target DCNNs (ResNet-34 and H-ResNet-34). It is reported in Fig. 8 that both DCNN methods converge
 312 in the training and validation processes, and the total loss curve of H-ResNet-34 converges faster than ResNet-
 313 34 in both the training and validation processes.



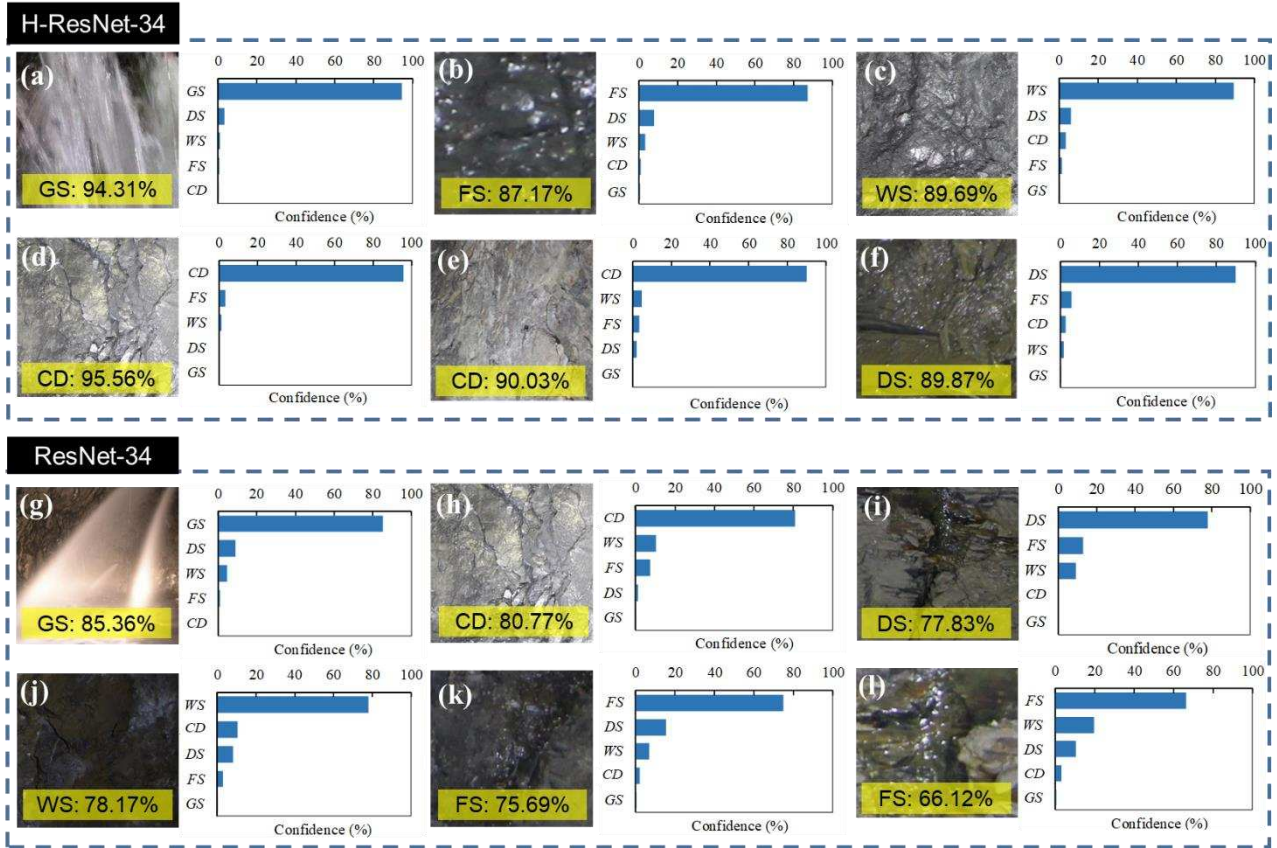
314

315 **Fig. 8.** Performance of both DCNN methods for total loss in training and validation processes.

316

317 4.3 Testing results

318 By the allocated dataset distribution, the testing dataset consists of 1,200 randomly selected images. The
319 testing images are classified through the trained and validated DCNN methods (ResNet-34 and H-ResNet-34).
320 Six images out of the 1,200 testing images for each DCNN method are chosen, and the corresponding
321 classification confidences are presented in Fig. 9. Overall, H-ResNet-34 presents a better performance than
322 ResNet-34. The classification confidence of both DCNN methods for the five water inflow categories is highest
323 for the GS category and followed by the CD category.

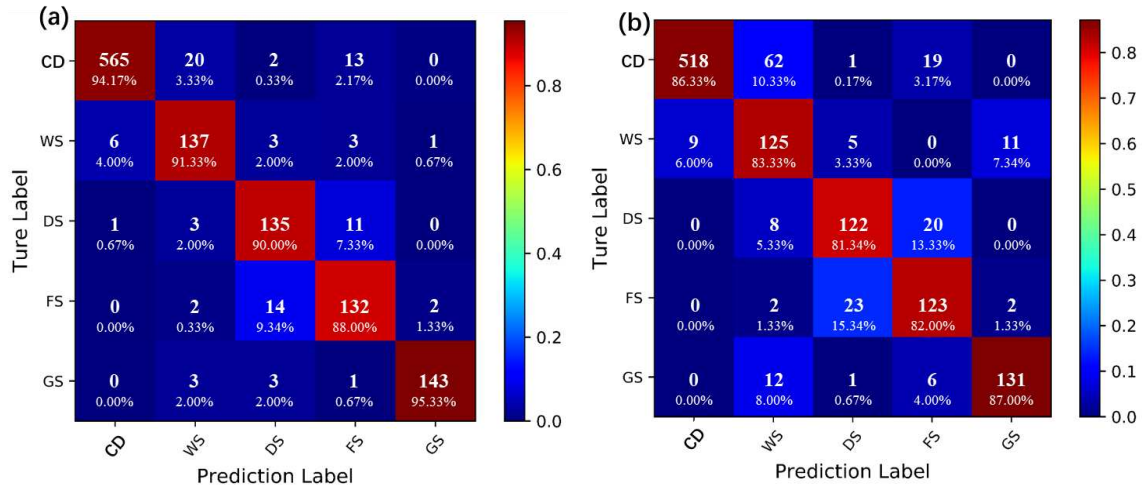


324

325 **Fig. 9.** Classification results of randomly selected water inflow images: (a) to (f) with H-ResNet-34, (g) to (l)
326 with ResNet-34.

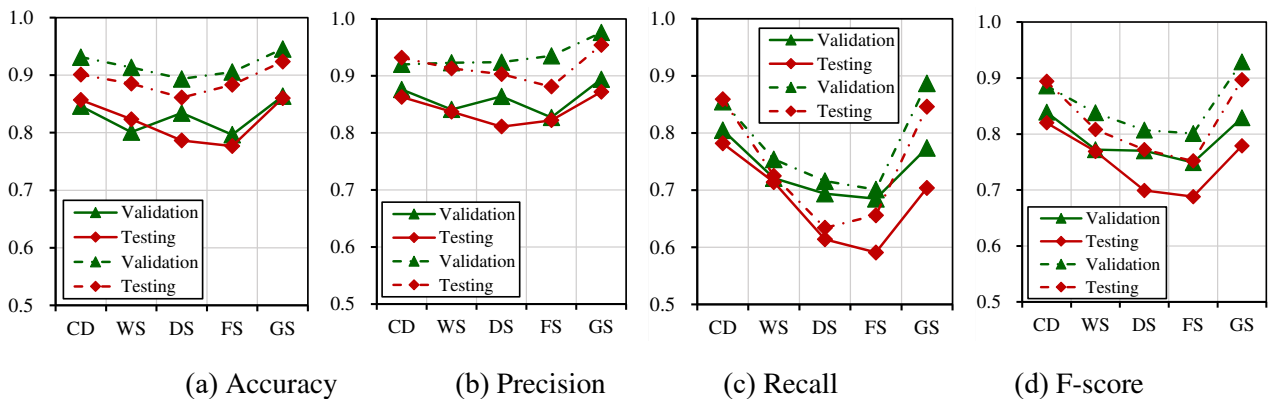
327

328 To present an informative comparison, a confusion matrix is computed in this study (shown in Fig. 10).
329 A confusion matrix is a standard form of matrix form with n rows and n columns to express classification
330 performance, where the n is assumed as the total number of categories. By quantitatively comparing the
331 confusion matrixes between two DCNN methods, the proposed H-Resnet-34 shows excellent improvement in
332 the classification of the water inflow categories. The GS category has the highest true positive probability of
333 95.33%, followed by CD, WS, DS, and FS categories with probabilities of 94.17%, 91.33%, 90.00%, and
334 88.00%, respectively. By investigating the misclassification between different water inflow categories, DS has
335 a 9.34% probability of being misclassified as FS, 0 and FS has a 7.33% probability of being misjudged as DS.
336 The relatively high misjudgment probabilities between DS and FS are due to similar subtle textures and grey
337 values in the sample images of the DS and FS categories. Thus, it is urgent to increase the DS and FS image
338 dataset and enhance the texture morphology identification in the training process.



339
 340 **Fig. 10.** Confusion matrixes of the testing dataset classification results: (a) with H-ResNet-34, (b) with
 341 ResNet-34.
 342

343 For further quantitative analysis, Fig 11 plots the comparison between the original ResNet-34 and the H-
 344 Resnet-34 in terms of accuracy, precision, recall value, and F-score for both validation and testing dataset. All
 345 the evaluation metrics of both DCNN methods present a similar trend. The values of the metrics from the
 346 highest to the lowest always follow the same order: GS, CD, WS, DS, and FS. Since the texture features and
 347 distinct appearance of the GS images, it makes the DCNN methods corresponding more prominent in GS
 348 identification. The classification of DS and FS images suggest relatively poor performance, as the evaluation
 349 metrics of these two categories present rather low values. The potential reason is that the DS and FS images in
 350 the dataset do not have distinct features compared with the other three water in flow categories. By computing
 351 the mean values of precision, recall, and F-score, the corresponding values of H-Resnet-34 suggest 7.5%, 4.6%,
 352 and 6.7% higher than those of the original ResNet-34 method, respectively. Overall, adding a hierarchical
 353 structure to the original ResNet-34 can improve the classification accuracy on an imbalanced image dataset.
 354 Moreover, in the deep learning classification process, relatively low accuracy occurs in two categories with
 355 similar textures (e.g., DS and FS).
 356



358
 359 **Fig. 10.** Performance of both DCNN methods in validation and testing processes (solid line for ResNet-34,
 360 dotted line for H-ResNet-34): (a) accuracy, (b) precision, (c) recall, and (d) F-score.
 361

362 5. Conclusion

363 A vision-based automated method for classification of water inflow damage from imbalanced images of

364 under-construction tunnel faces was proposed, employing a deep convolutional neural network. For solving
365 the issue of extremely imbalanced image datasets, the ResNet-34 framework was employed as the backbone
366 network and was then modified with a hierarchical classification structure to classify damage at two different
367 levels: coarse-level and fine-level. The coarse-level identification, which could be simplified to a binary
368 problem, was used to distinguish the dataset of images with water inflow damage from non-damage images.
369 Then the fine-level identification was applied to compute the occurrence probability of each type of damage.
370 The two branches were finally gathered to acquire the ultimate results of each type of damage based on the
371 defined conditional probability.

372 The proposed H-ResNet-34 model consists of five convolution modules, an average pooling layer, and
373 two hierarchical classification modules. It was trained, validated, and tested on the resized images captured by
374 a digital photography method in highway tunnels under construction in Yunnan, China. The images within
375 minority categories were first oversampled, and then the dataset was increased using augmentation techniques.
376 The experimental results revealed that the proposed classifier could significantly improve the final accuracy
377 of water inflow classification.

378 The misjudgment rate between DS and FS is relatively high due to similar subtle textures and grey values
379 in the sample images of these two categories. Thus, it is urgent to strengthen the robustness of the imbalanced
380 dataset to handle the typical error identification that mainly occurs in two categories with similar textures. In
381 order to achieve higher accuracy and efficiency in the minority categories, increasing the number of samples
382 within each category and identifying more texture morphology are both required for future research
383

384 **Acknowledgments**

385 The research presented in this paper is supported by Key innovation team program of innovation talents
386 promotion plan by MOST of China (No. 2016RA4059), Natural Science Foundation Committee Program of
387 China (Grant No. 1538009, Grant No. 51778474), Science and Technology Project of Yunnan Provincial
388 Transportation Department (No. 25 of 2018), and the Fundamental Research Funds for the Central Universities
389 of China (Grant No. 0200219129).

391 **References**

- 392 Annan, A., Cosway, S., Redman, J., 1991. Water table detection with ground-penetrating radar, SEG Technical Program Expanded
393 Abstracts 1991. Society of Exploration Geophysicists, pp. 494-496.
- 394 Bieniawski, Z., 1988. The rock mass rating (RMR) system (geomechanics classification) in engineering practice, Rock Classification
395 Systems for Engineering Purposes. ASTM International.
- 396 Buda, M., Maki, A., Mazurowski, M.A., 2018. A systematic study of the class imbalance problem in convolutional neural networks.
397 *Neural Networks*. 106, 249-259.
- 398 Cai, W., Zhu, H., Liang, W., 2022. Three-dimensional tunnel face extrusion and reinforcement effects of underground excavations in
399 deep rock masses. *International Journal of Rock Mechanics and Mining Sciences*. 150, 104999.
- 400 Cai, W., Zhu, H., Liang, W., Zhang, L., Wu, W., 2021. A New Version of the Generalized Zhang–Zhu Strength Criterion and a
401 Discussion on Its Smoothness and Convexity. *Rock Mechanics and Rock Engineering*. 1-17.
- 402 Chen, J., Huang, H., Cohn, A.G., Zhang, D., Zhou, M., 2021a. Machine learning-based classification of rock discontinuity trace:
403 SMOTE oversampling integrated with GBT ensemble learning. *International Journal of Mining Science and Technology*.
- 404 Chen, J., Yang, T., Zhang, D., Huang, H., Tian, Y., 2021b. Deep learning based classification of rock structure of tunnel face. *Geoscience
405 Frontiers*. 12(1), 395-404.
- 406 Chen, J., Zhang, D., Huang, H., Shadabfar, M., Zhou, M., Yang, T., 2020. Image-based segmentation and quantification of weak
407 interlayers in rock tunnel face via deep learning. *Automation in Construction*. 120, 103371.
- 408 Chen, J., Zhou, M., Huang, H., Zhang, D., Peng, Z., 2021c. Automated extraction and evaluation of fracture trace maps from rock

409 tunnel face images via deep learning. *International Journal of Rock Mechanics and Mining Sciences*. 142, 104745.

410 Chen, J., Zhou, M., Zhang, D., Huang, H., Zhang, F., 2021d. Quantification of water inflow in rock tunnel faces via convolutional
411 neural network approach. *Automation in Construction*. 123, 103526.

412 Chen, J., Zhou, M., Zhang, D., Huang, H., Zhang, F., 2021e. Towards automatic quantification of water inflow in rock tunnel faces via
413 a novel CNN approach. *Automation in Construction*.

414 Deng, J., Dong, W., Socher, R., Li, L.-J., Li, K., Fei-Fei, L., 2009. Imagenet: A large-scale hierarchical image database, 2009 IEEE
415 conference on computer vision and pattern recognition. *Ieee*, pp. 248-255.

416 Fang, W., Ding, L., Zhong, B., Love, P.E.D., Luo, H., 2018. Automated detection of workers and heavy equipment on construction
417 sites: A convolutional neural network approach. *Advanced Engineering Informatics*. 37, 139-149.

418 Fernandez, G., Moon, J., 2010. Excavation-induced hydraulic conductivity reduction around a tunnel–Part 1: Guideline for estimate of
419 ground water inflow rate. *Tunnelling and Underground Space Technology*. 25(5), 560-566.

420 Gordan, B., Armaghani, D.J., Hajihassani, M., Monjezi, M., 2016. Prediction of seismic slope stability through combination of particle
421 swarm optimization and neural network. *Engineering with Computers*. 32(1), 85-97.

422 Guo, H., Zhou, J., Koopialipour, M., Armaghani, D.J., Tahir, M., 2019. Deep neural network and whale optimization algorithm to
423 assess flyrock induced by blasting. *Engineering with Computers*. 1-14.

424 Hajihassani, M., Armaghani, D.J., Marto, A., Mohamad, E.T., 2015. Ground vibration prediction in quarry blasting through an artificial
425 neural network optimized by imperialist competitive algorithm. *Bulletin of Engineering Geology and the Environment*. 74(3),
426 873-886.

427 He, K., Zhang, X., Ren, S., Sun, J., 2016. Deep residual learning for image recognition, *Proceedings of the IEEE conference on*
428 *computer vision and pattern recognition*, pp. 770-778.

429 Huang, C., Li, Y., Change Loy, C., Tang, X., 2016. Learning deep representation for imbalanced classification, *Proceedings of the IEEE*
430 *conference on computer vision and pattern recognition*, pp. 5375-5384.

431 Huang, H., Cheng, W., Zhou, M., Chen, J., Zhao, S., 2020. Towards Automated 3D Inspection of Water Leakages in Shield Tunnel
432 Linings Using Mobile Laser Scanning Data. *Sensors*. 20(22), 6669.

433 Hwang, J.-H., Lu, C.-C., 2007. A semi-analytical method for analyzing the tunnel water inflow. *Tunnelling and Underground Space*
434 *Technology*. 22(1), 39-46.

435 Khan, S.H., Hayat, M., Bennamoun, M., Sohel, F.A., Togneri, R., 2017. Cost-sensitive learning of deep feature representations from
436 imbalanced data. *IEEE transactions on neural networks and learning systems*. 29(8), 3573-3587.

437 Koopialipour, M., Fahimifar, A., Ghaleini, E.N., Momenzadeh, M., Armaghani, D.J., 2019a. Development of a new hybrid ANN for
438 solving a geotechnical problem related to tunnel boring machine performance. *Engineering with Computers*. 1-13.

439 Koopialipour, M., Ghaleini, E.N., Haghghi, M., Kanagarajan, S., Maarefvand, P., Mohamad, E.T., 2019b. Overbreak prediction and
440 optimization in tunnel using neural network and bee colony techniques. *Engineering with Computers*. 35(4), 1191-1202.

441 Kumar, S.S., Abraham, D.M., Jahanshahi, M.R., Iseley, T., Starr, J., 2018. Automated defect classification in sewer closed circuit
442 television inspections using deep convolutional neural networks. *Automation in Construction*. 91, 273-283.

443 Li, S., Liu, B., Xu, X., Nie, L., Liu, Z., Song, J., Sun, H., Chen, L., Fan, K., 2017. An overview of ahead geological prospecting in
444 tunneling. *Tunnelling and Underground Space Technology*. 63, 69-94.

445 Liu, S., Wu, J., Zhou, J., Zeng, Z., 2010. Numerical simulations of borehole radar detection for metal ore. *IEEE Geoscience and Remote*
446 *Sensing Letters*. 8(2), 308-312.

447 López, V., Fernández, A., García, S., Palade, V., Herrera, F., 2013. An insight into classification with imbalanced data: Empirical results
448 and current trends on using data intrinsic characteristics. *Information sciences*. 250, 113-141.

449 Lü, Q., Xiao, Z.-P., Ji, J., Zheng, J., Shang, Y.-Q., 2017. Moving least squares method for reliability assessment of rock tunnel
450 excavation considering ground-support interaction. *Computers and Geotechnics*. 84, 88-100.

451 Man, J., Huang, H., Ai, Z., Chen, J., 2022. Analytical model for tunnel face stability in longitudinally inclined layered rock masses
452 with weak interlayer. *Computers and Geotechnics*. 143, 104608.

453 Marjoribanks, R., 2010. *Geological methods in mineral exploration and mining*. Springer Science & Business Media.

454 Martinez-Murcia, F.J., Gorriz, J.M., Ramirez, J., Ortiz, A., 2018. Convolutional Neural Networks for Neuroimaging in Parkinson's
455 Disease: Is Preprocessing Needed? *Int J Neural Syst.* 28(10), 1850035.

456 Momeni, E., Armaghani, D.J., Hajihassani, M., Amin, M.F.M., 2015. Prediction of uniaxial compressive strength of rock samples using
457 hybrid particle swarm optimization-based artificial neural networks. *Measurement.* 60, 50-63.

458 Morin, F., Bengio, Y., 2005. Hierarchical probabilistic neural network language model, *Aistats.* Citeseer, pp. 246-252.

459 Nhat-Duc, H., Nguyen, Q.-L., Tran, V.-D., 2018. Automatic recognition of asphalt pavement cracks using metaheuristic optimized edge
460 detection algorithms and convolution neural network. *Automation in Construction.* 94, 203-213.

461 Rálek, P., Hokr, M., 2013. Methods of Water Inflow Measurement in the Bedřichov Tunnel. *EGRSE. Explor. Geophys., Remote Sens.*
462 *Environment.* 2, 30-39.

463 Santos, V., da Silva, P., Brito, M., 2018. Estimating RMR Values for Underground Excavations in a Rock Mass. *Minerals.* 8(3), 78.

464 Seo, Y., Shin, K.-s., 2019. Hierarchical convolutional neural networks for fashion image classification. *Expert Systems with*
465 *Applications.* 116, 328-339.

466 Simonyan, K., Zisserman, A., 2014. Very deep convolutional networks for large-scale image recognition. *arXiv preprint*
467 *arXiv:1409.1556.*

468 Sou-Sen Leu, A.T.J.W., 2011. Probabilistic prediction of tunnel geology using a Hybrid Neural-HMM. *Engineering Applications of*
469 *Artificial Intelligence.* 24(4), 658-665.

470 Szegedy, C., Liu, W., Jia, Y., Sermanet, P., Reed, S., Anguelov, D., Erhan, D., Vanhoucke, V., Rabinovich, A., 2015. Going deeper with
471 convolutions, *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 1-9.

472 Warren, S.N., Kallu, R.R., Barnard, C.K., 2016. Correlation of the Rock Mass Rating (RMR) System with the Unified Soil
473 Classification System (USCS): Introduction of the Weak Rock Mass Rating System (W-RMR). *Rock Mechanics and Rock*
474 *Engineering.* 49(11), 4507-4518.

475 Yan, Z., Zhang, H., Piramuthu, R., Jagadeesh, V., DeCoste, D., Di, W., Yu, Y., 2015. HD-CNN: hierarchical deep convolutional neural
476 networks for large scale visual recognition, *Proceedings of the IEEE international conference on computer vision*, pp. 2740-2748.

477 Zarei, H., Uromeihy, A., Sharifzadeh, M., 2013. A new tunnel inflow classification (TIC) system through sedimentary rock masses.
478 *Tunnelling and underground space technology.* 34, 1-12.

479 Zarei, H.R., Uromeihy, A., Sharifzadeh, M., 2012. Identifying geological hazards related to tunneling in carbonate karstic rocks-Zagros,
480 Iran. *Arabian Journal of Geosciences.* 5(3), 457-464.

481 Zhang, J.-Z., Phoon, K.K., Zhang, D.-M., Huang, H.-W., Tang, C., 2021. Novel approach to estimate vertical scale of fluctuation based
482 on CPT data using convolutional neural networks. *Engineering Geology.* 294, 106342.

483 Zhao, S., Zhang, D., Xue, Y., Zhou, M., Huang, H., 2021. A deep learning-based approach for refined crack evaluation from shield
484 tunnel lining images. *Automation in Construction.* 132, 103934.

485 Zhou, M., Cheng, W., Huang, H., Chen, J., 2021. A novel approach to automated 3d spalling defects inspection in railway tunnel linings
486 using laser intensity and depth information. *Sensors.* 21(17), 5725.

487