딥러닝 논문 발표

김대영 교수님

2025.04.28

김주현

ViT-Based Image Regression Model for Shear-Strength Prediction of Transparent Soil

2024.3 (China)

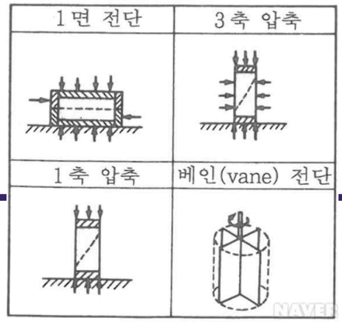
Buildings (IF 3.1)

2025.04.28

김주현

Introduction

- Many scholars use transparent soil to study ***the stability of slopes*** [11,12], ***the shear deformation of soil*** [13], and ***the surface uplift*** in model experiments.
- **Traditional direct-shear tests or triaxial direct-shear tests** require multiple repetitions to be performed to obtain the average value of each set of test results [20,21,22], and the experimental procedures are relatively cumbersome.
- The shear strength of transparent soil may be connected to ***the distribution of optical spots in the images***.
- The distribution of optical spots and the transparency differ between ***different transparent soils*** with ***different levels of shear strength***. However, quantifying these differences is difficult, and traditional image processing methods are not effective for analyzing them.
- **The current mainstream feature extractors** employed are **convolutional neural networks (CNNs)** [31,32,33] and **vision transformers (ViTs)** [34,35].
- Finding features related to shear strength in the speckle images of transparent soil is a challenging task.
- It requires a model that not only has excellent local feature-capturing capabilities, but also understands the global features of the image.



Introduction (Cont.)

- As shown in Figure 1, unlike existing image classification tasks, the features in the speckle image of transparent soil are distributed globally rather than concentrated in specific areas.
- Additionally, the speckles do not have significant boundaries, making it difficult for existing methods to capture their features.
- Based on previous research, **CNN**-based feature-processing modules are **able to perceive details well**, but remain limited in their ability to capture global features [38].
- A **ViT** has **excellent global feature-capturing capabilities** [39]; however, its image preprocessing method, which involves the segmentation of image blocks, does not convert the detailed information of the image into high-level features and is not suitable for processing transparent-soil images.
- Furthermore, predicting the shear strength of transparent soil based on images is, essentially, a regression task.
- In image processing problems, deep learning does not perform as well in regression tasks as it does in classification tasks.

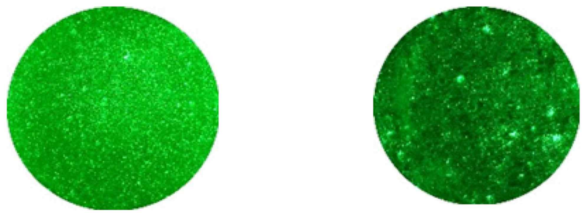


Figure 1. Transparent-soil speckle images with different parameters.

Introduction (Cont.)

- Therefore, it is necessary to develop a regression module that is able to facilitate the transfer of information between the feature extraction of transparent soil and shear-strength prediction processes.
- In this paper, a novel image regression model called the **ViT-based image regression model (VIRM)** is proposed.
- This model aims to improve the poor performance of existing methods in transparent-soil image feature extraction tasks.
- The input images are preprocessed using a CNN module, and the segmented image patches are replaced with feature maps to enter the transformer encoder.

Methodology

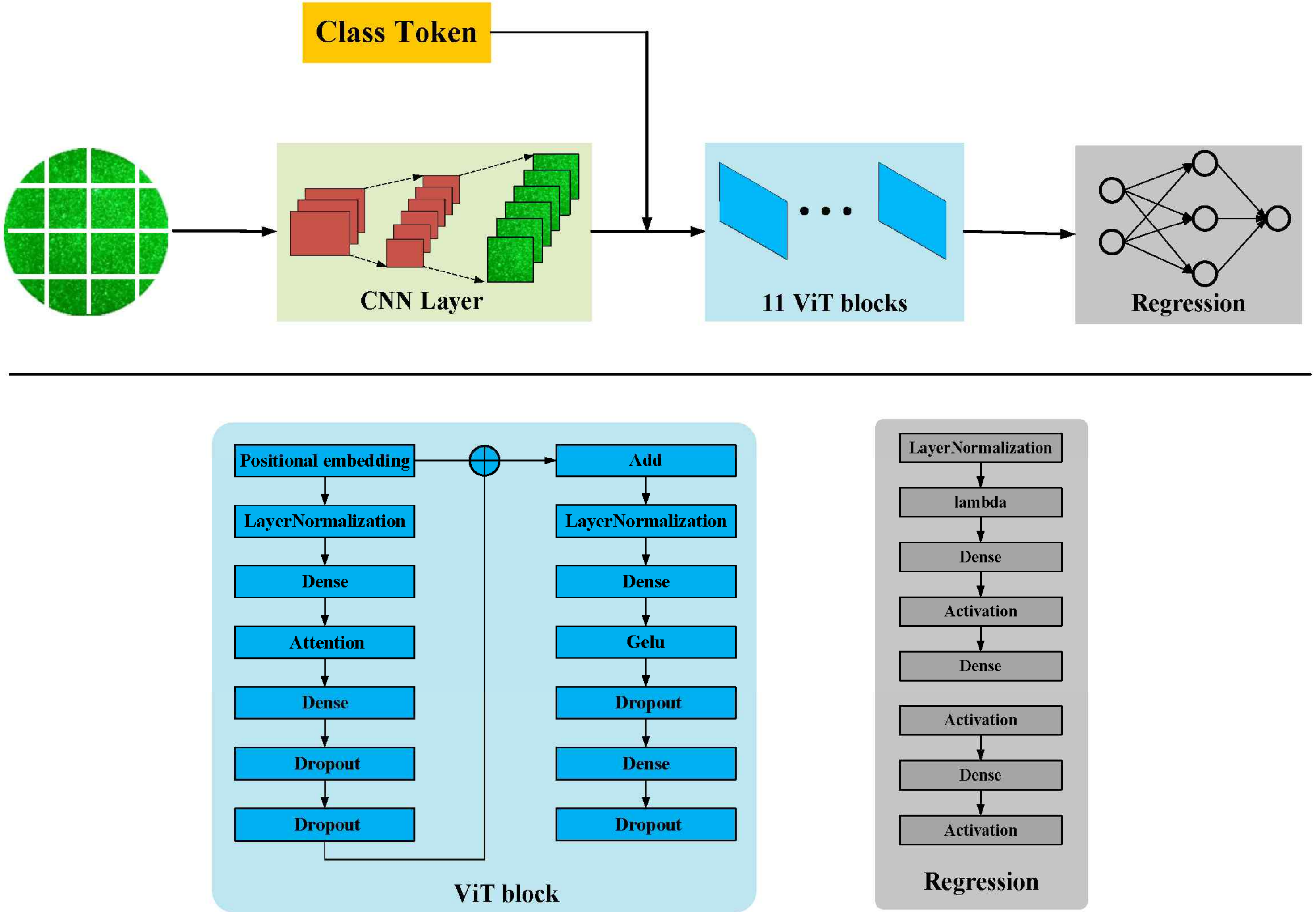


Figure 2. Flow chart of the proposed method.

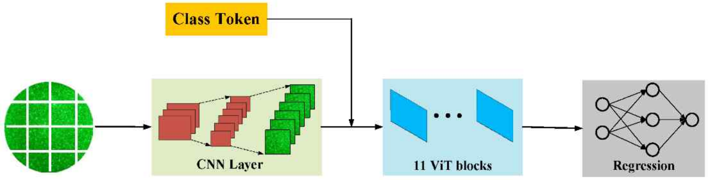
AI for Medical Imaging

Methodology (Cont.)

2.1. *Image Feature Extraction*

- The scattered image of the transparent soil first travels through a CNN layer, whose main purpose is to convert the scattered image into a multi-dimensional feature map that contains the multi-level features of the scattered image.

- The feature map is then passed into the ViT module.



- The conventional **ViT** module splits the complete image into N blocks of the same size, and then the N blocks are converted into N high-dimensional image block feature vectors via a linear mapping layer; however, such methods are limited in their ability to deal with the positional relationships between image blocks [40].

- Since the **CNN** has the property of inductive bias and **the transformer** has the ability to perform strong global inductive modeling, better results can be obtained when using the hybrid CNN + transformer model; therefore, instead of image segmentation, the CNN is used in this paper.

- The core component of a **ViT** is the attention mechanism, which forces the model to focus on more important feature maps [41,42].

- Because different feature maps contribute differently to the prediction task, weights are added to each feature map as an indication of the importance of the feature map.

AI for Medical Imaging

Methodology (Cont.)

2.2. *Regression Module*

- The regression module is a linear layer that connects the feature extraction component to the labels, and mainly consists of activation functions and dense layers.

- The feature extraction component converts the image features into feature vectors, which are first normalized and then passed through several activation and dense layers, which increase the nonlinearity of the regression module;
  - Finally, the image features are connected to the labels.

Table 1. The structure of the feature extraction component.

	Layer Name	Input Size	Output Size
CNN	CNN	(224, 224, 3)	(197, 768)
ViT	LayerNormalization	(197, 768)	(197, 768)
	Dense	(197, 768)	(197, 2304)
	Attention	(197, 2304)	(197, 768)
	Dense	(197, 768)	(197, 768)
	Dropout	(197, 768)	(197, 768)
	Dropout	(197, 768)	(197, 768)
	Add	(197, 768), (197, 768)	(197, 768)
	LayerNormalization	(197, 768)	(197, 768)
	Dense	(197, 768)	(197, 3072)
	Gelu	(197, 3072)	(197, 3072)
	Dropout	(197, 3072)	(197, 3072)
	Dense	(197, 3072)	(197, 768)
	Dropout	(197, 768)	(197, 768)

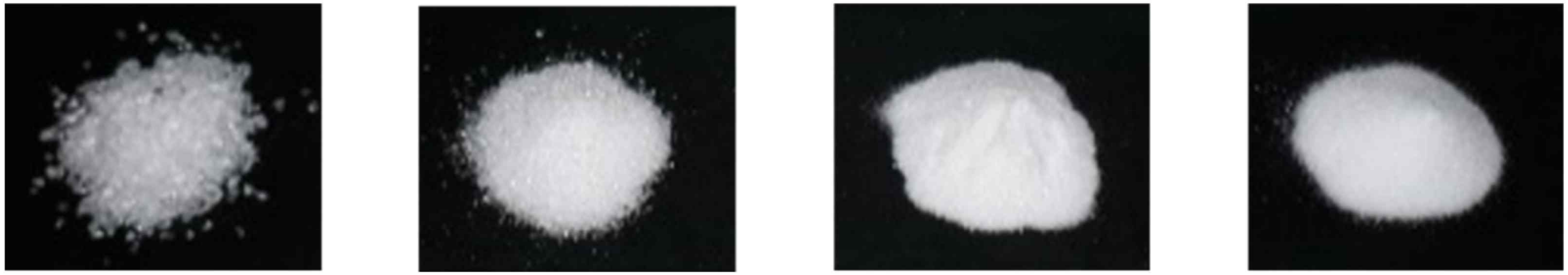
Table 2. Structure of the regression module.

Layer Name	Input Size	Output Size
LayerNormalization	(197, 768)	(197, 768)
Lambda	(197, 768)	(None, 768)
Dense	(None, 768)	(None, 128)
Activation	(None, 128)	(None, 128)
Dense	(None, 128)	(None, 256)
Activation	(None, 256)	(None, 256)
Dense	(None, 256)	(None, 1)
Activation	(None, 1)	(None, 1)

imaging

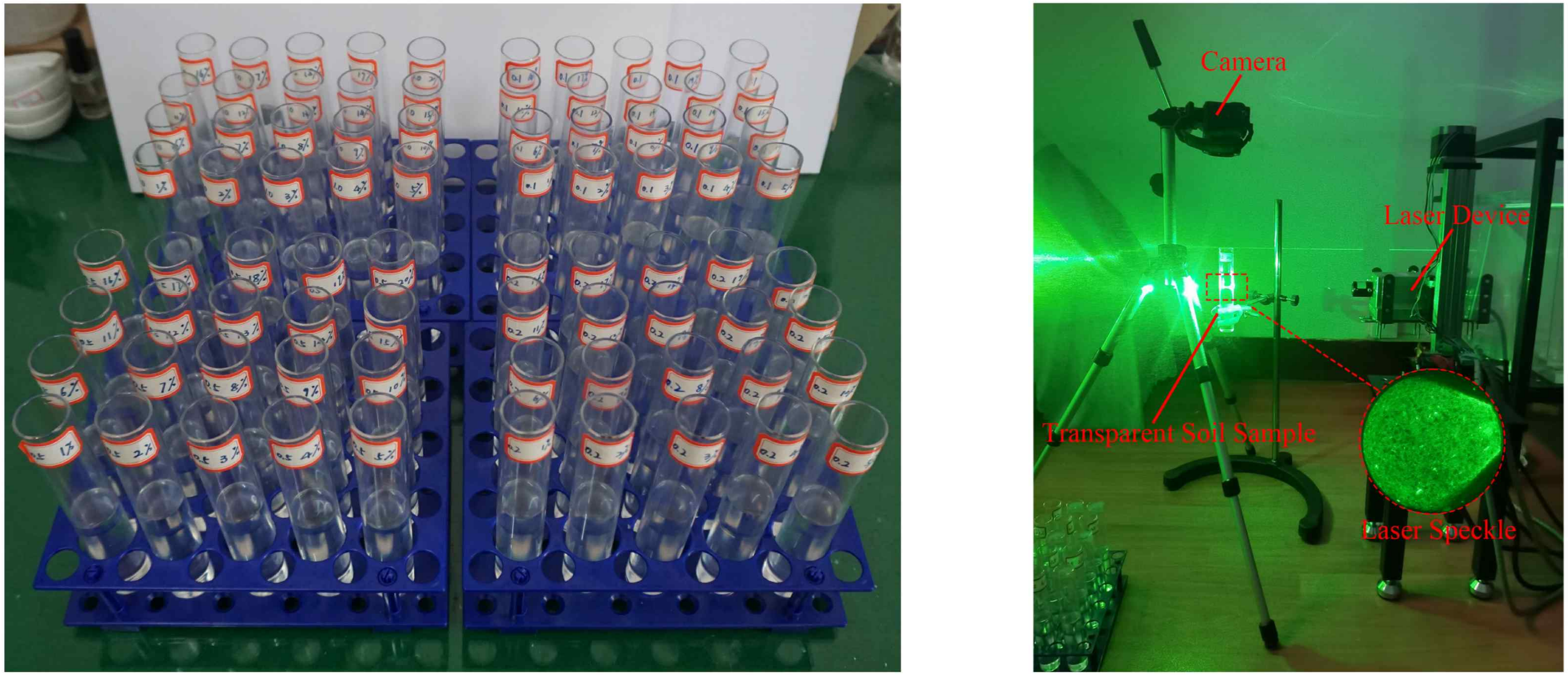
Transparent-Soil Straight-Shear Experiment

3.1. Transparent Cemented Soil Preparation



(**a**) 1.0–3.0 (**b**) 0.5–1.0 (**c**) 0.2–0.5 (**d**) 0.1–0.2

Figure 3. Fused quartz sand.

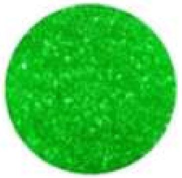
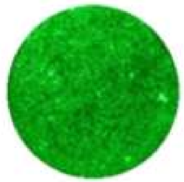
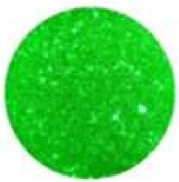


(**a**) Transparent-soil specimens (**b**) Transparent-soil photography equipment

Figure 4. Transparent–soil specimens and transparent–soil photograph equipment.

3.1. Transparent Cemented Soil Preparation

Table 3. Speckle image of transparent soil with different parameters.

Content of Fumed Silica Powder	Particle Size of Fused Quartz Sand/mm			
	0.1–0.2	0.2–0.5	0.5–1.0	1.0–3.0
0%				
5%				
10%				
15%				
20%				

Transparent-Soil Straight-Shear Experiment

3.3.1. Dataset 1: Cohesion

- Cohesion refers to the force generated by the cementation and electrostatic gravitational force between particles

- The percentage of hydrophobic fumed silica powder was the main factor affecting the cohesion of the specimen.

- In this dataset, the proposed ViT model was trained using the captured transparent-soil images as the input images and the cohesion as the label; the dataset contained a total of 2000 pre-processed samples.

3.3.2. Dataset 2: Friction Angle

- The friction angle is an important index used to describe the soil friction strength.

- In this paper, the content of quartz sand was the main factor affecting the friction angle of transparent cemented soil.

- In this dataset, 2000 transparent-soil cross-section images were used as the input of the dataset, and the corresponding friction angles were used as labels to train the proposed ViT.

Experiments and Results

4.1. Experiment Configuration

- The network model used in this paper was trained on a high-performance GPU NVIDIA GeForce RTX2080.
- The software environment was CUDA version bit 10.2, the python version was 3.6.3, the system version was WIN10, and the network used Tensorflow version 1.13.

4.2. Experimental Results

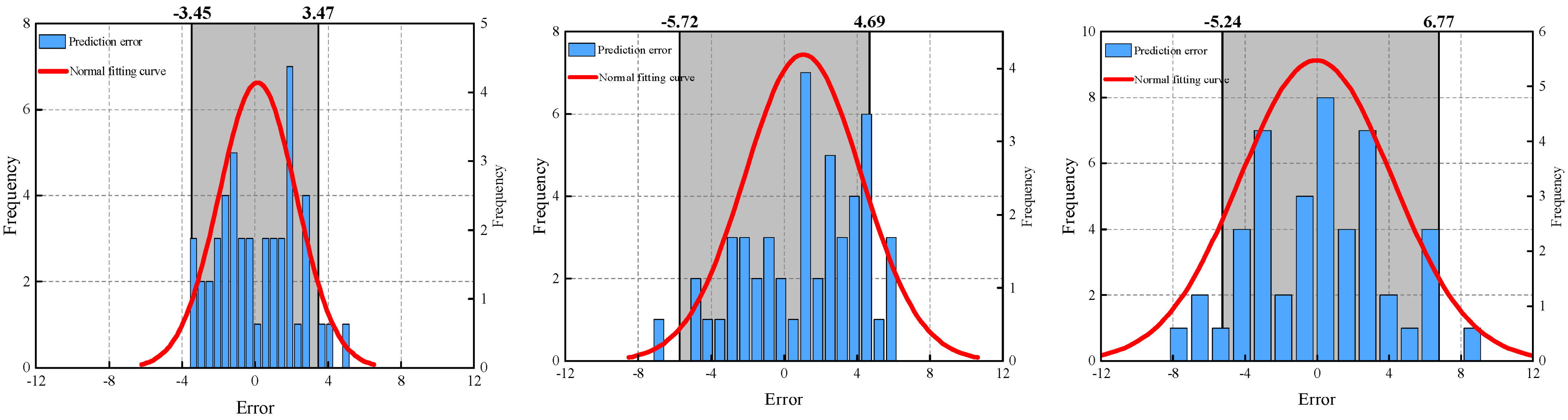
- The backbone networks that were compared are VGG and ResNet.
- The **VGG network** is characterized by the use of very small convolutional kernels (typically 3 × 3) and a very deep network structure. → This design enables the VGG network to use a smaller number of parameters.
- **ResNet** addresses the gradient problem by introducing shortcut connections that span across network layers. Shortcut connections directly pass the input information (i.e., intermediate features) around various layers to the subsequent layers, allowing the neural network to better optimize residual information.

Experiments and Results

Table 4. Training times and average prediction errors for different methods

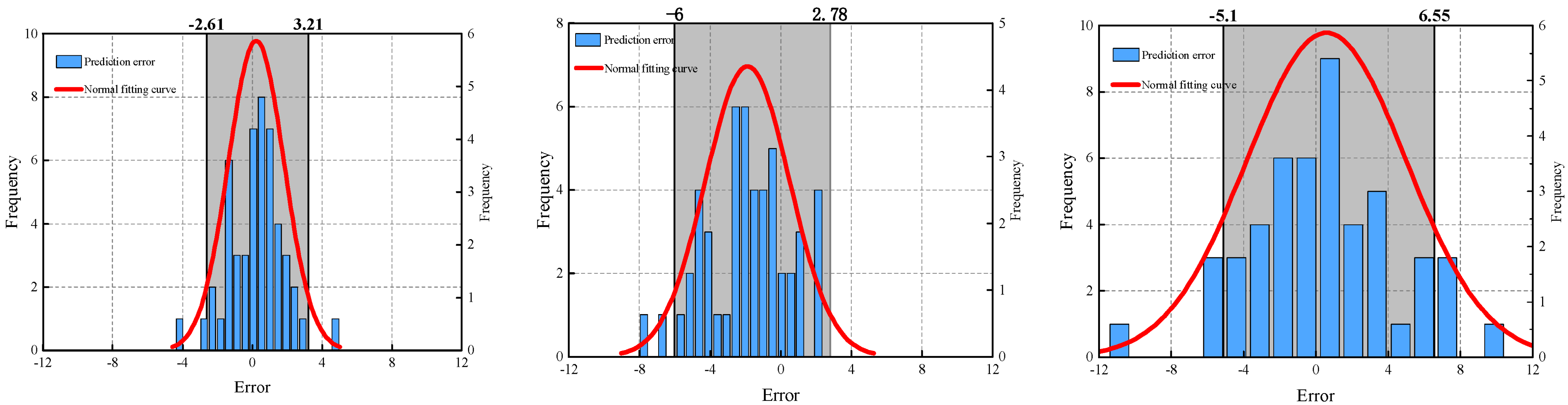
Method	Training Time	Cohesion Average Error	Friction Angle Average Error
Vit	124 min	1.64	0.73
Resnet	78 min	3.95	2.33
VGG	55 min	5.27	2.26

Experiments and Results



(**a**) ViT (**b**) Resnet (**c**) VGG

Figure 6. Error statistics for cohesion prediction tasks with different backbone networks.

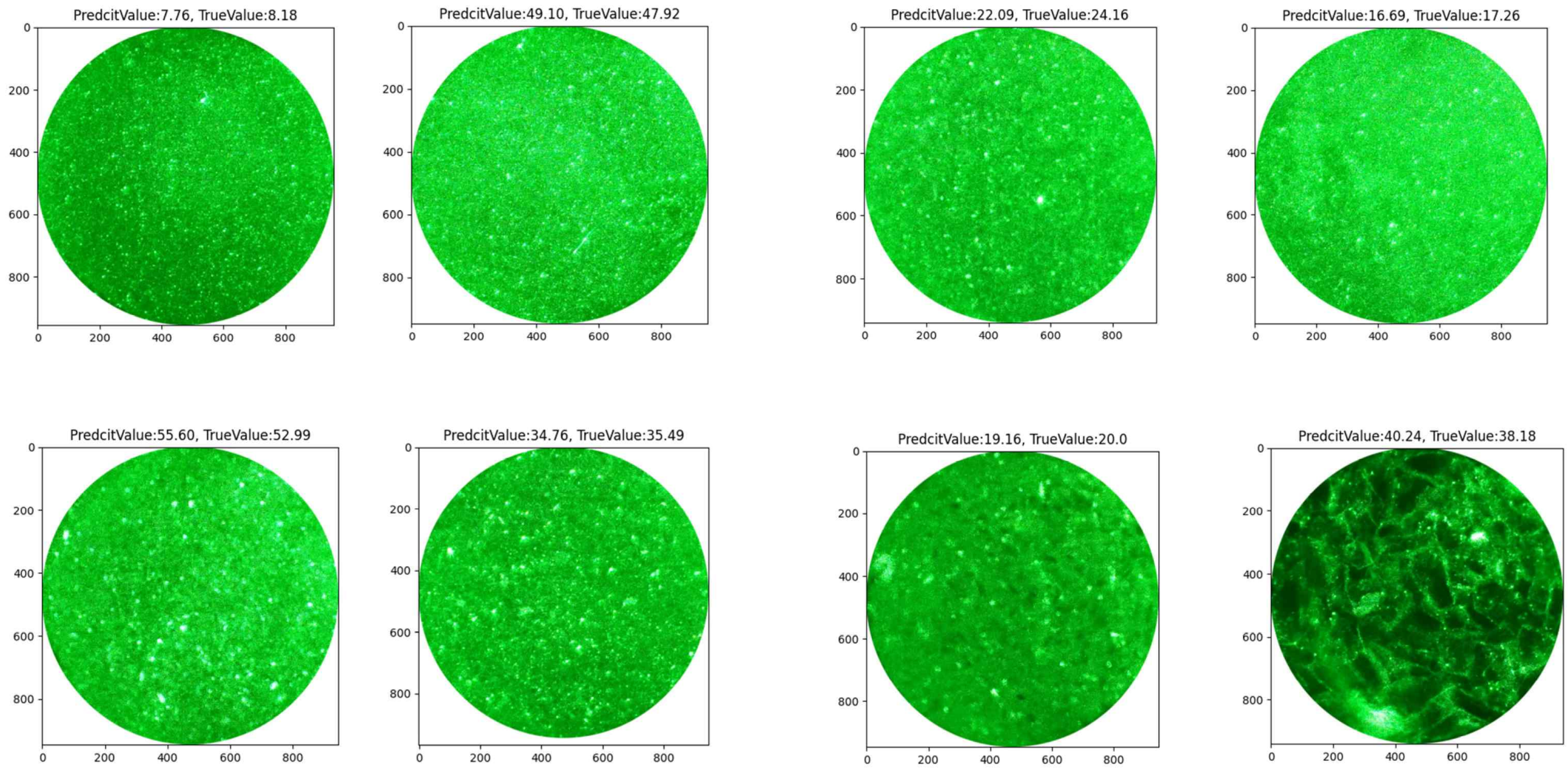


(**a**) ViT (**b**) Resnet (**c**) VGG

Figure 7. Error statistics for friction angle prediction tasks with different backbone networks.

Imaging

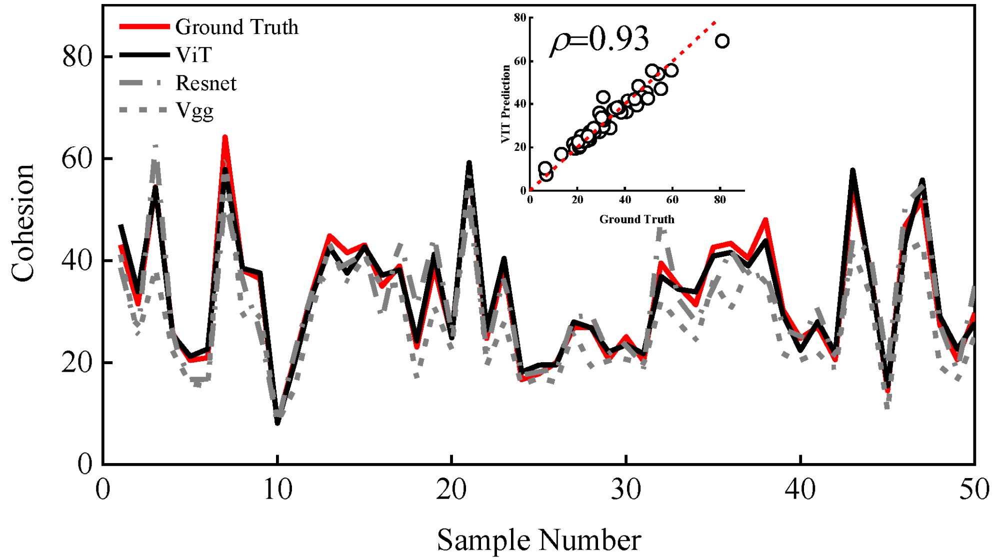
Experiments and Results



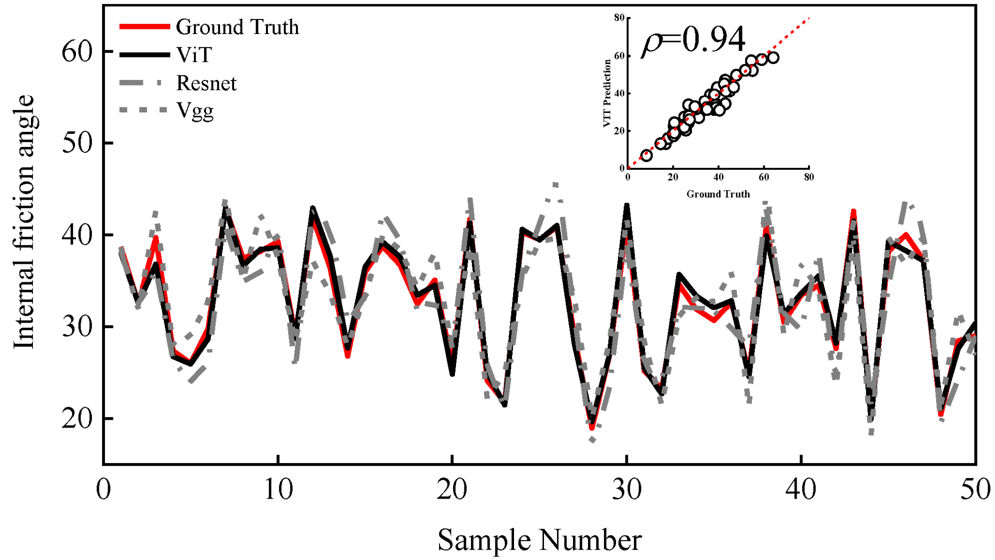
(**a**) Results predicted for cohesion (**b**) Results predicted for friction angle

Figure 8. Results predicted for cohesion and friction angle.

Experiments and Results



(a) Cohesion



(b) Friction angle

Figure 9. Prediction results for 50 samples and correlation statistics.

AI for Medical Imaging

Conclusions

- In order to predict the shear strength of transparent soil based on images, a VIRM image regression prediction model is proposed in this paper.

- To address the issue of unclear feature boundaries in transparent-soil images and the characteristics of their global distribution, a combination of CNN and ViT is proposed; this overcomes the limitations of traditional ViT models, which rely on the input of segmented soil blocks.

(1) To demonstrate the effectiveness of the proposed feature extraction module, the module was replaced with classical Resnet and VGG models for comparison.

- The results showed that in both datasets, the feature extraction module based on CNN + ViT was more suitable for predicting the shear strength of transparent soil.
- The shear-strength prediction method based on this module achieved a smaller error distribution and higher prediction accuracy.

(2) Fifty samples were randomly selected for prediction, and the correlation coefficients between the predicted values and the true values were calculated.

- The results showed that the proposed method achieved correlation coefficients of 0.93 and 0.94 in the two datasets, indicating a high level of reliability.

Thank you