

For research article

Response to Reviewer 1

Comments

1. Summary

Thank you very much for taking the time to review this manuscript. Please find the detailed responses below and the corresponding revisions/corrections highlighted/in track changes in the re-submitted files.

2. Questions for General Reviewer's Evaluation	Response and Revisions
Does the introduction provide sufficient background and include all relevant references?	Can be improved Agree
Are all the cited references relevant to the research?	Yes
Is the research design appropriate?	Yes
Are the methods adequately described?	Can be improved Agree
Are the results clearly presented?	Can be improved Agree
Are the conclusions supported by the results?	Can be improved Agree

3. Point-by-point response to Comments and Suggestions for Authors

Comments 1: *The validation R^2 of 0.99 is unrealistically high for rock mass prediction and indicates likely overfitting. The authors must provide learning curves and test the model on independent tunnel projects with different geological conditions. Regularization techniques and their effects should be explicitly reported.*

Response 1: Thank you for pointing this out. Your suggestion is very helpful. In Section 3, the reported R^2 values are indeed above 0.99. These numbers are from the test sets-(page 17, and line444), which can raise concerns about overfitting. We have revised the manuscript to clarify this point. We also added regularization details to mitigate overfitting (weight decay). Following your second comment, we re-evaluated the model with 10-fold cross-validation. We trained on 9 folds and tested on the held-out fold in each round. We repeated the experiments and report the learning curves. We now provide 95% confidence intervals for all evaluation metrics. These

additions are included in the revised manuscript- (page 13, paragraph3, and line348-line364).

Comments 2: *sing only an 80/20 split for 512 samples is insufficient. Implement k-fold cross-validation (k=5 or 10) and report mean and standard deviation of all metrics. Include confidence intervals for all reported performance measures.*

Response 2: Thank you for the helpful feedback. We revised the manuscript and added supplementary experiments to bolster our findings-(page 13, paragraph3, and line354).

Comments 3: *"Empirical methods" for hyperparameter selection is inadequate. Document the search space and use grid search or Bayesian optimization. Explain why ResNet50 was chosen over other depths through ablation studies.*

Response 3: Thank you for raising this point.

Hyperparameter selection is critical in machine learning.

The search space is complex.

Manual exploration can easily get stuck in local optima.

We therefore used an exhaustive enumeration (grid search).

Within a predefined search space, we trained all combinations and selected the best by validation performance.

The search space was:

- Learning rate: {0.001, 0.0001, 0.00001}
- Batch size: {8, 16, 32}
- Optimizer: {Adam}

Before tuning, we built a lightweight subset (50% of the data).

We ran multiple trials on this subset to compare all hyperparameter combinations.

We have revised the manuscript and now describe the procedure as “grid search (exhaustive enumeration)” rather than an “empirical method.”-(page 12, paragraph3, and line313-line318).

Regarding the CNN architecture, our primary goal is to show that deep learning enables efficient and accurate inverse analysis during excavation.

Before the main study, we ran a small pilot to compare common ResNet backbones on a ~500-sample dataset:

Model	Training time (min)	Accuracy within 5% error
ResNet18	20	92.5%
ResNet34	20	94.1%
ResNet50	25	95.8%
ResNet101	30	95.9%

The differences in training time and accuracy were modest at this scale.

Hence, backbone-induced variance was not the focus of this work.

We therefore fixed a standard ResNet backbone for the subsequent experiments and concentrated on the inverse-analysis task itself.

Comments 4: *Model descriptions, training procedures, and metric definitions are repeated multiple times. Consolidate methodology into one section, merge similar tables (5-6 and 8-9), and define R² and RMSE only once. Section 3 should compare models while Section 4 focuses on temporal analysis.*

Response 4: Thank you for the valuable question.

Yes, Section 3.1 compares deep neural networks that model multiple features induced by excavation.

Section 3.2 compares simple machine-learning models using a single feature point.

Section 3.3 summarizes the findings of the first two subsections.

Section 4 analyzes the efficiency of the deep models. We emphasize this point in the summary title.- (page 17, line435)

We believe that an error of less than 5 % is a more accurate result evaluation index, so we would like to show it separately, and thank you for your suggestion.

We will strengthen this explanation to make the paper's structure clear and the conclusions of each section explicit.

4. Response to Comments on the Quality of English Language

Point 1: The manuscript contains grammatical errors and inconsistent article usage throughout. Verb tenses shift between past and present within sections. Many sentences are too long and should be divided for clarity. Technical terms like "back analysis" need consistent hyphenation. The abstract contains run-on sentences that obscure key findings. Passive voice is overused where active voice would be clearer. Prepositions are occasionally incorrect, particularly "in" versus "on" for data inputs.

Response 1: Thank you for pointing this out. We agree with this comment. Therefore, we conducted a language clean-up focusing on preposition accuracy and splitting long sentences, and we standardized key terminology and usage across the manuscript. Specifically:

Prepositions: corrected inaccurate or awkward uses (e.g., divide based on → divide by; shown on Figure → shown in Figure; values of monitoring points → values at monitoring points; trained with/on the test set where appropriate; application to back-analysis; correlation with the target variable).-

Long sentences: split overly long or multi-clause sentences into shorter units to improve readability while preserving meaning.

Consistency (light touch in this round): kept terms consistent when they occurred in the edited passages (e.g., back-analysis)

(page 4, line132-134, line145-147)、(page 5, line132-134, line152-155, 161)、(page 6, line171-175)、(page 7, line207-207)、(page 8, line226-229)、(page 12, line313-318)、(page 13, line324-331, 342-347)、(page 15, line388-392)、(page 18, line461-473, 495)、(page 21, line516-521)、(page 22, line546-566)