

Peer Review

Review of: "Uncertainty-Driven Modeling of Microporosity and Permeability in Clastic Reservoirs Using Random Forest"

Tarek Ganat¹

1. Petroleum and Chemical department, Sultan Qaboos University, Oman

Please find below the comments and suggestions for improvement to the paper:

1. Regarding novelty clarification, the integration of uncertainty analysis with Random Forest is underemphasized.
2. The expansion from 41 samples to 61,600 via resampling lacks statistical justification. Cite literature supporting this methodology to validate its reliability.
6. The $\pm 4\%$ porosity and ± 10 mD permeability ranges in Table 1 lack citations. Reference prior studies or field data justifying these thresholds.
7. Clarify whether resampling used bootstrapping, Monte Carlo, or another method. Specify how correlation between variables (e.g., grain size and permeability) was handled.
8. The absence of k-fold cross-validation or train-test split details raises reproducibility concerns. Include validation metrics beyond accuracy (e.g., F1-score, ROC-AUC).
9. The model's reliance on parameters like the ThK ratio is unclear. Add a feature importance plot or SHAP analysis to identify critical predictors.
10. Missing evaluation of how input variability (e.g., $\pm 2\%$ clay content) impacts predictions. Include tornado plots or partial dependence plots.
11. No benchmarks against traditional methods (e.g., regression models) or other ML algorithms (e.g., XGBoost). Add a comparative table.
12. High accuracy (93%) may indicate overfitting. Report precision, recall, and confusion matrices for both microporosity and permeability.
13. Results lack discussion of how regional factors (e.g., Borneo's tectonics) influence model generalizability to other basins.

14. Figure 2 Resolution. Low-quality workflow diagram. Redraw with vector graphics for clarity.
15. Figure 3 Context. The stratigraphic column lacks scale bars or age labels. Add detailed annotations.
18. No discussion of model limitations (e.g., applicability to carbonate reservoirs or ultra-tight rocks). Address these gaps.
19. Underdeveloped section on how the model reduces exploration costs. Quantify potential savings (e.g., vs. MICP/SEM costs).
20. Clay diagenesis (e.g., illitization) is mentioned but not linked to model predictions. Integrate diagenetic thresholds into the uncertainty framework.
21. No mention of code or data accessibility. Share a GitHub repository or supplementary materials for reproducibility.
22. Omitted Random Forest hyperparameters (e.g., tree depth, estimators). Specify tuning methods (e.g., grid search).
23. Uncertainty analysis does not quantify error accumulation from resampling. Include error bars in results.
24. Introduction and Methodology sections redundantly describe data collection. Condense for brevity.
25. The conclusion is generic. Restate key findings (e.g., accuracy metrics, cost reductions) and future directions (e.g., real-time field applications).

These revisions would enhance clarity, rigor, and impact, positioning the study for successful journal submission.

Declarations

Potential competing interests: No potential competing interests to declare.