Qeios

Peer Review

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

Victor Mensah[1]

1. Universidad Politécnica de Madrid, Spain

General Overview

This manuscript addresses an important technical challenge: how to predict microporosity and permeability in clastic reservoirs using machine learning (Random Forest) and an uncertainty augmentation strategy. The authors gathered valuable field and laboratory data from Labuan Island, Malaysia, and performed machine learning classification to predict key reservoir properties cost-effectively.

The study is practically significant and relevant to exploration geoscience. However, the manuscript **overstates its novelty**, mischaracterizing the use of a standard algorithm (Random Forest) as the "development" of a machine learning model. Further, although the methodology is sound overall, several improvements in scientific clarity, terminology precision, and technical validation are needed.

Major Comments

Mischaracterization of Work Scope

The authors claim they “developed” a machine learning model. In reality, they **implemented and trained an existing Random Forest algorithm** with uncertainty-augmented data. There is no modification to the Random Forest architecture or algorithmic innovation.

Required Action: Correct the language throughout the manuscript: “implementation,” “application,” and “training” should replace “development.”

Strength and Limits of Dataset

doi.org/10.32388/9J11D2

The original dataset is very small (41 samples expanded synthetically). Although uncertainty augmentation produced a large training set, synthetic variability **cannot fully replace true geological heterogeneity**.

Suggestion: A more critical discussion of dataset limitations and the risks of training bias is needed in the Discussion section.

Absence of Feature Importance Analysis

Random Forest offers a natural feature importance ranking. It is crucial to **show which input features** (e.g., porosity, Th/K ratio, grain size) most influenced predictions.

Required Action: Add a feature importance plot or table to strengthen interpretability and credibility.

Validation Strategy and External Generalization

The 5-fold cross-validation and internal holdout validation are good practices.

However, **external validation** (e.g., using samples from different formations not seen in training) was not performed.

Suggestion: Acknowledge this as a limitation explicitly and suggest future testing on broader datasets.

Microporosity Classification Simplicity

Reducing microporosity to a binary "macroporosity dominated" vs. "microporosity dominated" may oversimplify complex pore systems.

Suggestion: Propose multi-class classification or regression for future work.

Minor Comments

Permeability Categories in Table 2:

The table is inconsistent: permeability >50 mD should correspond to "Good to Very Good." Correct the typo.

Figures and Confusion Matrices:

Label ROC curves and axes more clearly (e.g., "True Positive Rate," "False Positive Rate").

Writing and Grammar:

Minor typos due to PDF encoding errors ("dif□cult" → "difficult", etc.) — needs careful proofreading.

Inconsistent use of active and passive voice — recommend standardizing depending on journal guidelines.

References:

Some references seem automatically formatted and could benefit from minor manual editing (e.g., missing page numbers, line breaks).

Recommendation

Major Revisions Required.

After revisions, the manuscript would be strong enough for publication.

Declarations

Potential competing interests: No potential competing interests to declare.