Imperial College London

Department of Earth Science and Engineering

MSc Geo-Energy with Machine Learning and Data Science

Arcadia Project Report

Deep and Machine Learning Strategies for Geological Data Interpretation in Arcadia

Team: PoroPermeables

South Kensington Campus, London, UK.

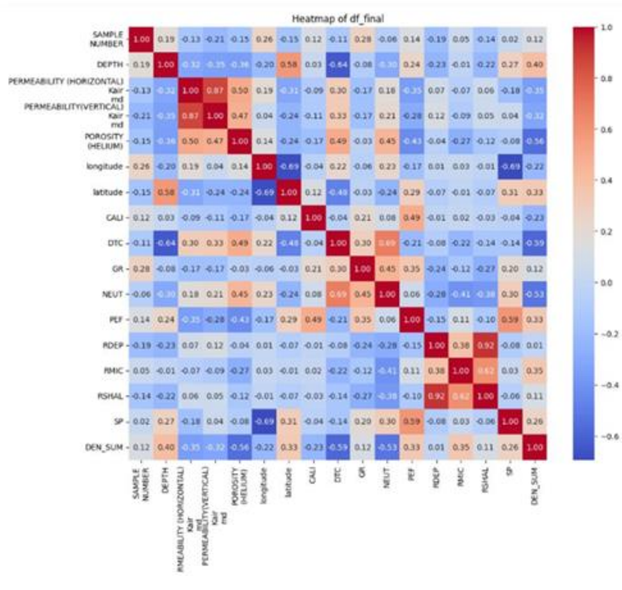
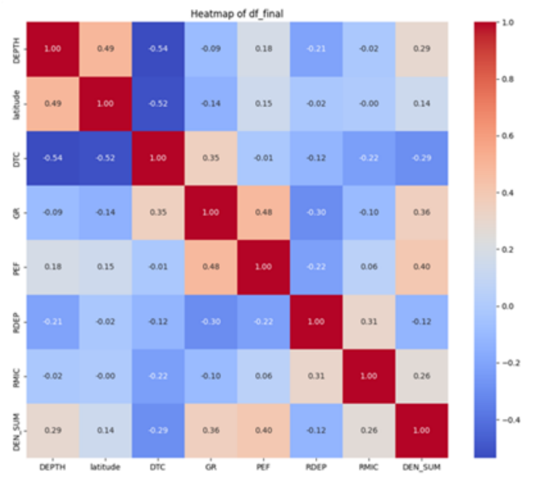
2024

Data Processer

This project employs the Mask-RCNN machine learning model for segmenting and aligning geological core sample images by depth, utilizing the CoreSegmenter API. This innovation significantly enhances geological data analysis and efficiency for researchers by streamlining workflows and improving the precision of geological assessments. The introduction of the CoreColumn data structure facilitates improved data manipulation and access, emphasizing the critical role of depth registration in analyzing stratigraphic sequences and mineral compositions. By leveraging a pre-trained Mask-RCNN model, the project avoids the need for retraining, saving resources while ensuring accuracy and consistency in data processing. The primary goal is to automate the segmentation and depth-alignment of core samples, thus advancing geological evaluations. Data is sourced from eleven wells in the Avalon Basin, a key area for oil exploration. The project addresses challenges by refining image processing techniques, thereby enhancing the reliability of geological interpretations.

Permeability Predictor

During our data preprocessing, we incorporated latitude and longitude of well cores, aligned depths with wire log data, and selected crucial features including 'DEPTH', 'POROSITY (HELIUM)', 'DEN_SUM' (combining 'DENS' and 'DENC'), 'DTC', 'log_GR', 'PEF', 'log_RDEP', and 'latitude'. 'RDEP', which measures resistivity at greater depths, was chosen for its significance in predicting water saturation and permeability. Employing a Stacking ensemble method with RandomForest and XGBoost as base models and Linear Regression as the meta-model, we achieved promising results with a validation RMSE of 371.77 and R^2 of 0.548, and a test RMSE of 305.18 and R^2 of 0.623. Challenges in preprocessing, including handling missing values and special characters, were addressed, and hyperparameter tuning improved base model performance, leading to the selection of a stacked model for the permeability predictor

ALL FEACTURES	RELEVANT FEACTURES
	

Correlation matrix of all features and relevant features [image 1]

LITHOFACIES CLASSIFICACION MODEL:

In this stage of the project, we applied a Convolutional Neural Network (CNN) to predict lithofacies in core image data. We used 0.01 m windows and focused on the central part of the images, achieving a robust accuracy of 63%. However, we encountered challenges related to incomplete data and classification errors. To prevent overfitting, we implemented measures such as dropout layers and batch normalization. A key improvement was the correction of classification errors, resulting in high accuracy in predicting the 'nc' lithofacies. Consequently, our model generates images with overlaid colours representing the lithofacies on the base image, providing an accurate visual representation of geological data.

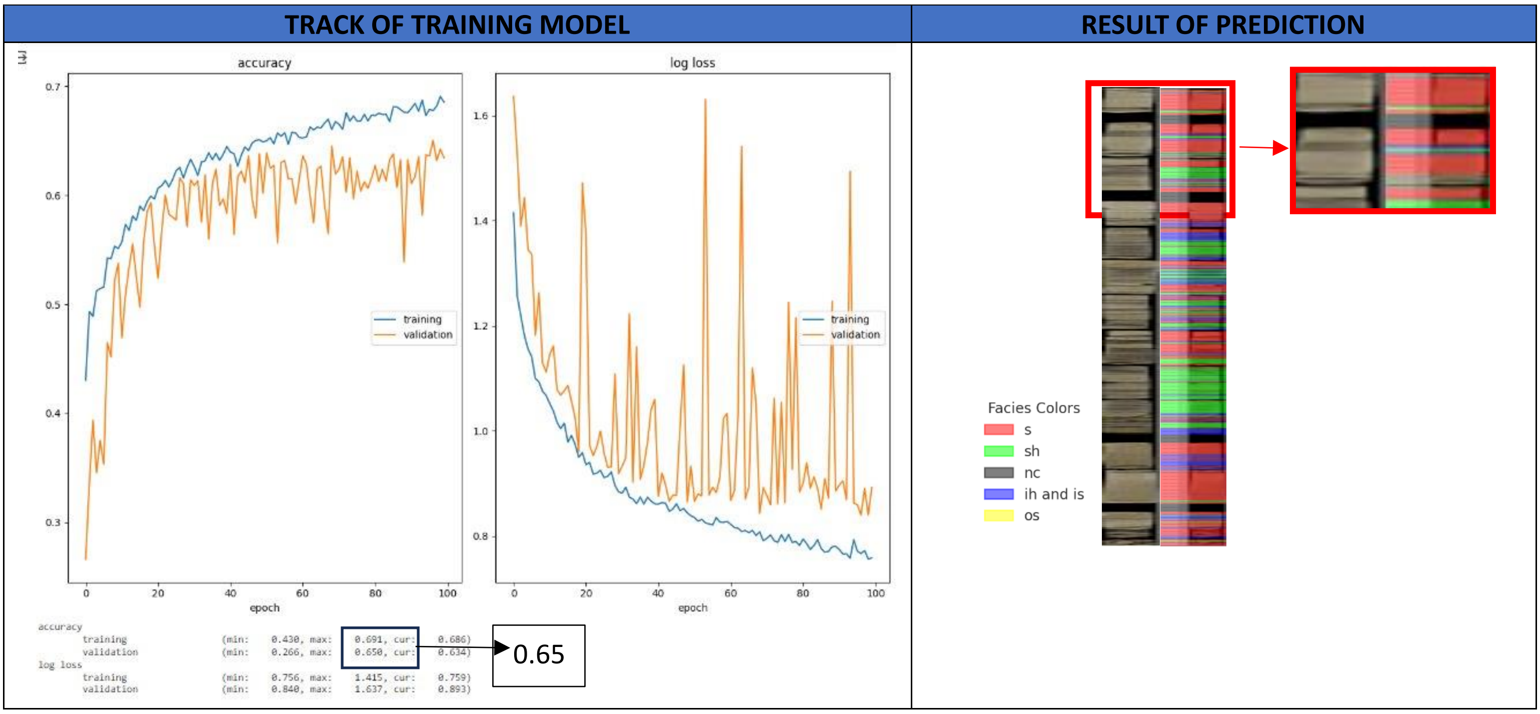
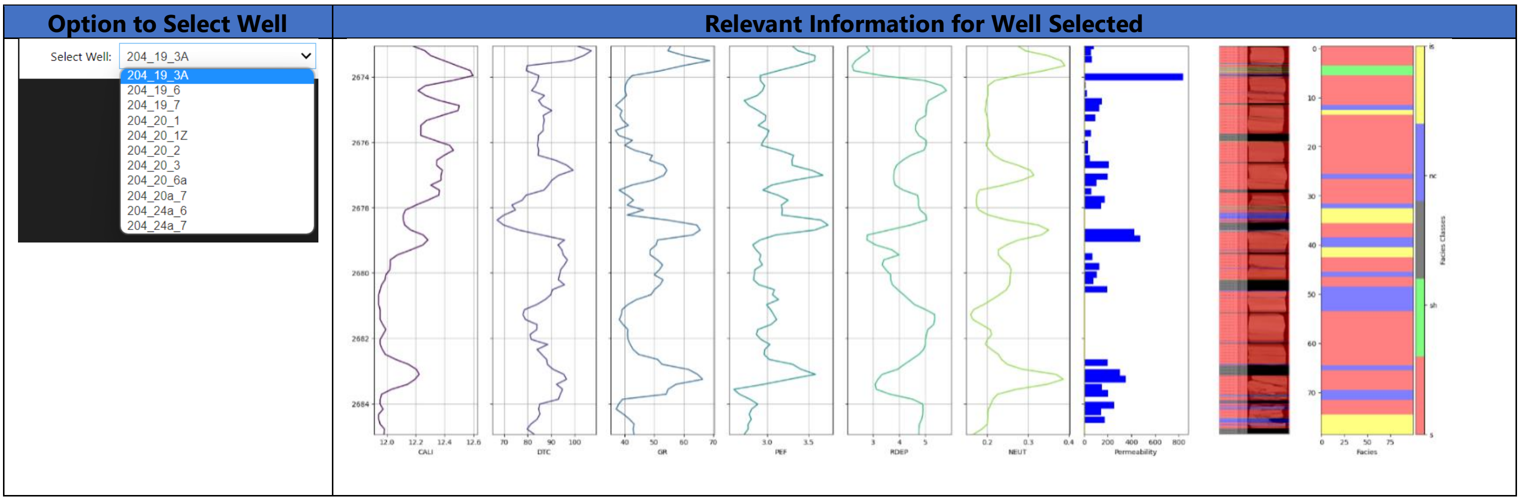


Image of the core well 204_19_6 and Image of the prediction for lithofacies using our Lithofacies model [image 2]

VISUALIZATION TOOL

The Python-coded tool significantly enhances the interpretation and analysis of geological data. The use of interactive widgets enables dynamic exploration of characteristics and measurements from various wells, as shown in Image 3. This visualization includes data ranging from borehole diameter and gamma radioactivity to permeability predictions and lithofacies classifications. Predictive models overlaid on actual core images in Image 3 provide a direct comparison between the observed data and the model's deductions, highlighting the code's role in integrating detailed analysis with informed geological decisions.



Visualization tool User interface [image 3]