**A data-driven approach to predict fracture intensity using machine learning for pre-salt carbonate reservoirs**

Eberton Rodrigues de Oliveira Neto\*1, Fábio Júnior Damasceno Fernandes1, Tuany Younis Abdul Fatah1, Raquel Macedo Dias1, Antonio Fernando Menezes Freire1,2, and Wagner Moreira Lupinacci1,2 1 GIECAR – UFF, 2 INTC-GP

Copyright 2023, SBGf - Sociedade Brasileira de Geofísica

This paper was prepared for presentation during the 18th International Congress of the Brazilian Geophysical Society held in Rio de Janeiro, Brazil, 16-19 October 2023.

Contents of this paper were reviewed by the Technical Committee of the 18th International Congress of the Brazilian Geophysical Society and do not necessarily represent any position of the SBGf, its officers or members. Electronic reproduction or storage of any part of this paper for commercial purposes without the written consent of the Brazilian Geophysical Society is prohibited.

# Abstract

The prediction of fracture intensity is essential for optimizing reservoir production and mitigating drilling risks in the Brazilian pre-salt layer. We propose a data-driven approach that utilizes well logs and seismic attributes (Curvature, Distance to Fault) to predict fracture intensity. To achieve this, we evaluate the effectiveness of two tree-ensemble models. Results show curvature as the most important feature, followed by distance to fault and Young’s Modulus (or P-Impedance). Also, XGBoost achieved better results compared to Random Forest. The approach was validated with rock sampling information and can be applied to other wells in nearby fields. The study provides a valuable tool for quantitatively estimating fracture intensity in pre-salt reservoirs. Future studies may use this study to estimate fracture intensity in a seismic volume.

# Introduction

Recent studies on the Brazilian pre-salt have demonstrated the significance of adequately characterizing the fracture system to obtain more reliable dynamic reservoir models (Correa et al., 2019; Tanaka et al., 2022). This fact arises from the intrinsic relationship between faulted and fractured zones and high permeabilities identified in many pre-salt fields (Fernández-Ibáñez et al., 2022a; Fernández-Ibáñez et al., 2022b). Therefore, correctly identifying faults in seismic data and estimating fracture intensity, especially in anomalous permeability zones, can add significant value to reservoir characterization workflows. Additionally, as Fernández-Ibáñez et al. (2022a) pointed out, these zones can be associated with drilling accidents, and their identification can help mitigate drilling risks.

Estimating fracture intensity on the pre-salt reservoir is a challenging task. Previous studies addressing this challenge relied heavily on conceptual models and did not fully use machine learning techniques to handle complex and nonlinear problems. A common issue when working with faults and fractured zones in the pre-salt layer is the low bandwidth of seismic data due to dispersion, attenuation effects, and poor illumination, which diminish the seismic resolution. This problem is particularly acute for high-frequency content, crucial for defining faults and fractured zones (Botter et al., 2014; Chen et al., 2020). Consequently, faults are rarely properly defined in raw seismic data of the pre-salt. Francisco et al. (2023) demonstrated that seismic preconditioning aimed at increasing spectral bandwidth and fault sharpness can significantly improve fault identification.

Traditional fracture modeling in the pre-salt still requires manual fault interpretation to either extrapolate fracture properties along fault zones (e.g., Tanaka et al., 2022) or as labels in machine learning approaches for automatic fault identification (Babasafari et al., 2022). Consequently, these approaches are highly susceptible to human bias and error.

To overcome these challenges, we propose a novel approach that uses machine learning to identify faults in seismic data and to predict fracture intensity. This study aims to validate this approach in the 1D domain before applying it to a seismic volume.

# Method

The present workflow shares similarities with the studies developed by Correa et al. (2019) and Babasafari et al. (2022), which also used machine learning. Correa et al. (2019) obtained a fracture intensity attribute (P32) using Hampson et al. (2001) approach, while Babasafari et al. (2022) used Multi-Layer Perceptron to create a fault attribute and calculated fracture density using statistical analysis. Another similarity between Babasafari et al. (2022) work and our paper is the preconditioning of the seismic amplitude. Although, they only aimed to attenuate noise and did not produce any enhancement in the seismic resolution.

In this work, we apply inverse Q filtering (Braga and Moraes, 2013) to enhance the contribution of the high-frequency content and consequently improve the definition of seismic amplitude discontinuities associated with faults and fractured zones. The faults were identified from a pre-trained 3D UNet fault predictor. The pre-trained network did not demand any human/expert labeling, preventing the process from human bias. Although, a preliminary result on this step was considered unsatisfactory, probably due to the very different aspect of our seismic dataset from the training data set of the UNet model (Wu et al., 2019). Thus, we applied the edge-preserving smoothing filter (EPS) (Luo et al., 2002) to further enhance the sharpness of the faults.

We combined concepts from brittleness studies with curvature analysis and fault-damage zones to select the input features for the 1D fracture intensity prediction model. Silicified zones have a greater fracture intensity due to their greater brittleness and are commonly associated with highly fractured intervals in the pre-salt (Correa et al., 2019; Fernandez-Ibanez et al., 2022a; Fernandez-Ibanez et al., 2022b, Lupinacci et al., 2023, Wennberg et al., 2023), while muddy facies are less likely to develop fractures (Correa et al., 2019). Highly deformed regions tend to have a higher curvature. Areas close to seismic scale faults will have a higher fracture intensity towards the faults, which will generally decay exponentially away from them (Savage and Brodsky, 2011).

Based on those principles, we chose as input features Young’s Modulus (E), Poisson’s Ratio (v), Silica Content (Si), Most-Positive (k1) Curvature, and a Distance to Fault attribute for predicting fracture intensity. The first three inputs are from the upscaled well logs, while the last two we extracted from seismic data at well position. We compute fault distance attribute from the fault probability volume estimated by the UNet model. The probability value 0.7 was the threshold for calculating the distance to a fault.

We evaluated two types of tree-based ensemble predictors to predict fracture intensity: XGBoost (Chen and Guestrin, 2016) and Random Forest (Breiman, 2001). Tree-based ensembles are famous for overperforming neural networks and other supervised methods in tabular datasets (Grinsztajne et al., 2022). Previous geoscience studies also came to the same conclusion (Bhattacharya et al., 2019; Man et al., 2021; Mandal et al., 2022; Nguyen et al., 2022).

We used five wells for this study (Figure 1) and excluded intervals with igneous rocks. The number of samples used was 224. We randomly separate the dataset to evaluate the models using the 80/20 rule, 80% for training and 20% for testing.

We evaluate the performance of the models with regression analysis. To understand the importance of each input feature, we tested some feature importance methods. Among several methods of feature importance evaluated in the study, we selected the permutation importance (Altmann et al. 2010) since this method better deals with collinearity between the features as in the case of k1 curvature and distance to fault. The concept of permutation importance is to randomly shuffle each feature a determined number of times, run the prediction with this shuffled dataset and measure the impact in a given metric. Figure 2 illustrates a resume of the workflow followed in this work.

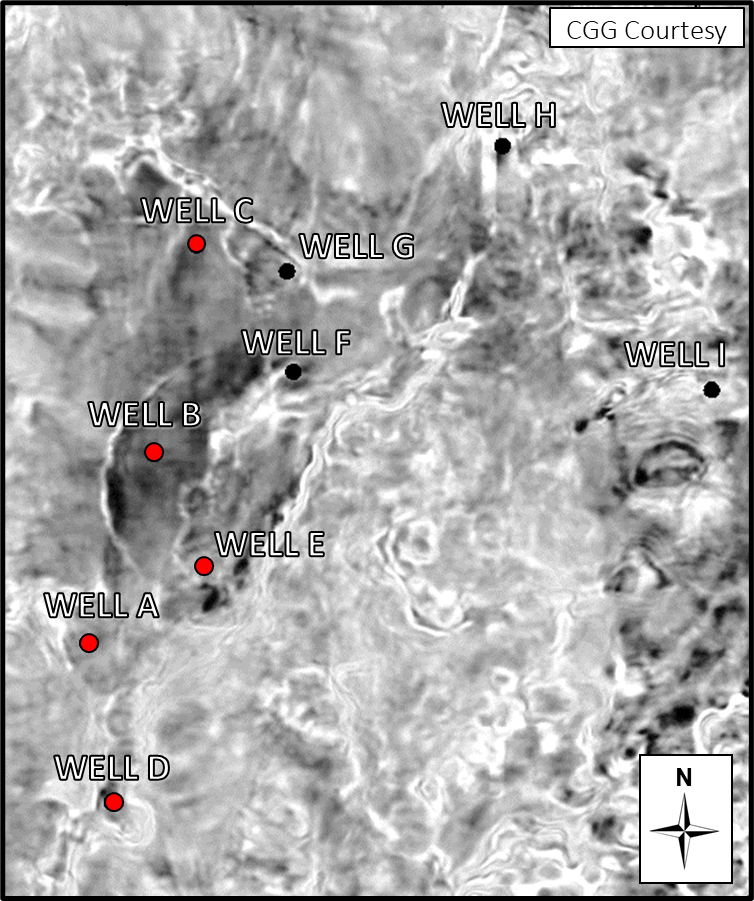


Figure : Base map of the study area with the location of the wells available and the seismic sections presented. The wells available and used to estimate the fracture intensity log are highlighted in red.

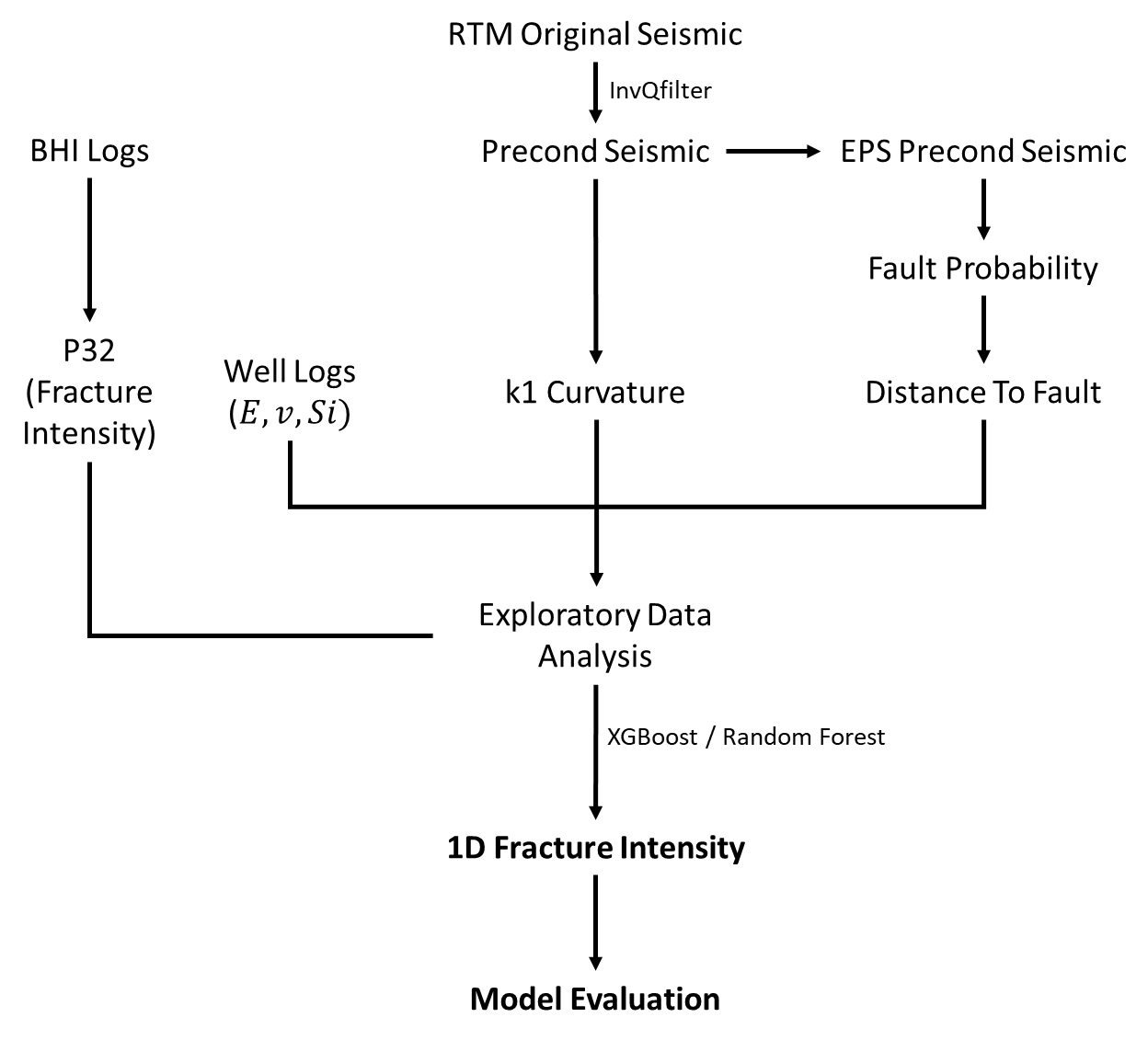


Figure 2: Illustration of the workflow to perform a feasibility study for fracture intensity modeling using machine learning.

# Results

The k1 curvature shows a higher correlation with fracture intensity, followed by silica content and distance to fault (Figure 3). Young’s Modulus and Poisson’s Ratio showed little to no correlation. However, this does not mean that these features cannot be useful for the model or that their importance follows the same order. There are more robust and specific methods for evaluating feature importance, which we discuss later.

In addition to the correlation matrix evaluation, we observe a considerable correlation (>0.5) between k1 curvature and distance to fault, although not high enough to classify them as redundant information. This is confirmed by the cross-plot analysis (Figure 4) between k1 curvature and distance to fault. The two variables have a reasonable correlation with little scattering at the upper-left corner of the plot, which is related to samples of low curvature and higher distances to fault. Towards lower levels of distances to fault, the dispersion increases, with those levels being associated with both low and high values of k1 curvature. That is explained by the fact that distance to fault only measures the proximity of a sample to the nearest fault and does not indicate any information about the amount of deformation associated with this particular fault. On the other hand, the curvature can be seen as a proxy for the amount of strain. Therefore, despite being somewhat correlated, distance to fault and k1 curvature are not redundant features and can be used together in the model.

Additionally, the cross-plot analysis confirms that the highest levels of fracture intensity are found in samples with higher k1 curvature, closer to faults, and with higher silica content.

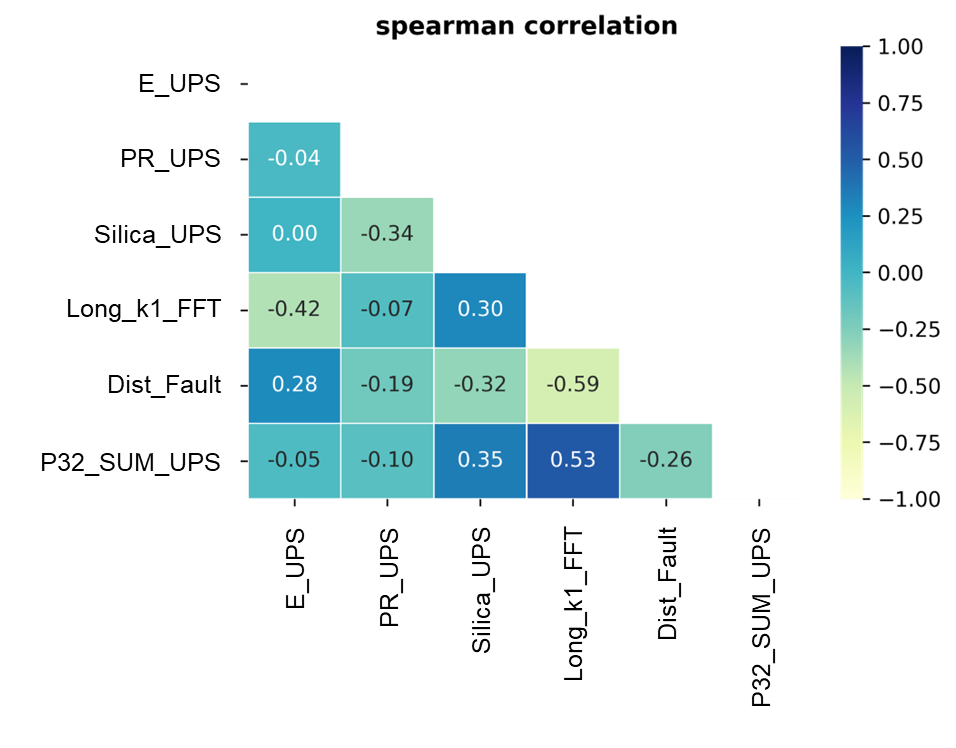


Figure 3: Spearman correlation matrix showing the measured nonlinear correlation between the input and target features.

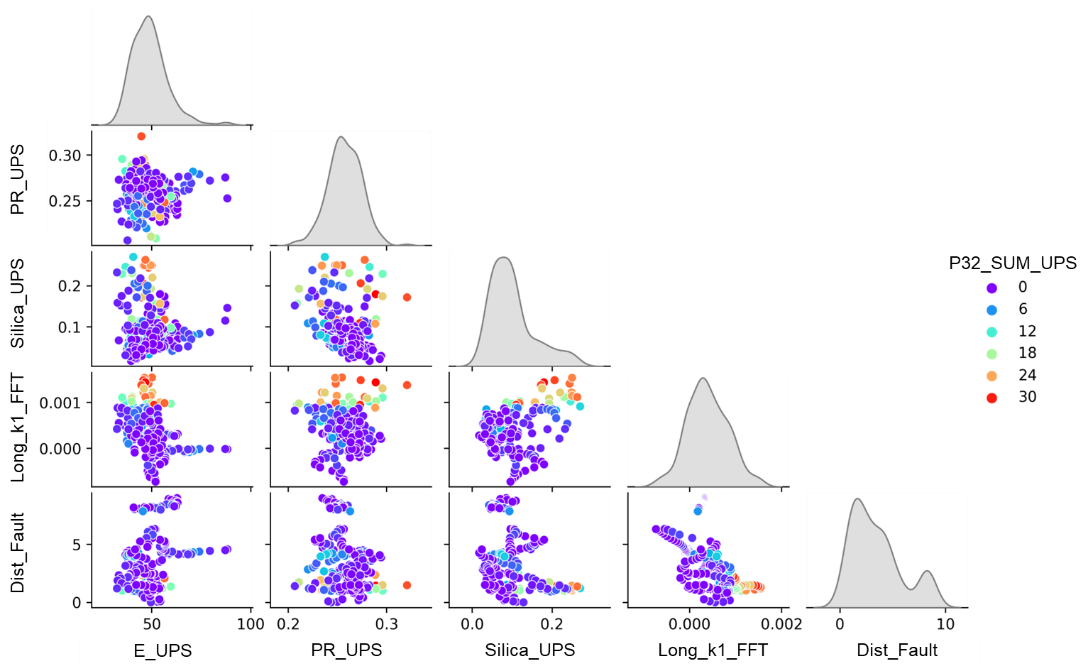


Figure 4: Joint cross-plot analysis between the input features colored by the levels of fracture intensity.

We evaluate the performance of each model using residual plots (Figure 5 and Figure 6) and regression metrics (Table 1) for fitting with the testing set. The residual plots indicate that XGBoost had slightly better performance than Random Forest, particularly in lower levels of fracture intensity. This is represented by a median and mode closer to zero compared to Random Forest. Regression metrics also support this analysis, with XGBoost outperforming Random Forest in all three evaluated metrics.

As the permutation importance method is stochastic, one way to analyze its results is by using boxplots (Figure 7). We ran the permutation importance on the test dataset to evaluate each feature’s importance for the model and its real predictive power. The results show that k1 curvature has the most significant impact on the R² score, followed by distance to fault and, surprisingly, Young’s Modulus. Silica content ranked fourth, while Poisson’s ratio is the least important feature, with the potential for a null-to-negative impact on the model.

These analyses support using seismic data to predict fracture intensity away from wells. In the present study, of the three most important variables, it is not possible to estimate the Young’s Modulus, as the seismic data is post-stacked. So, we use acoustic impedance (P-impedance) instead. The workflow is repeated with k1 curvature, distance to fault, and upscale P-impedance log.

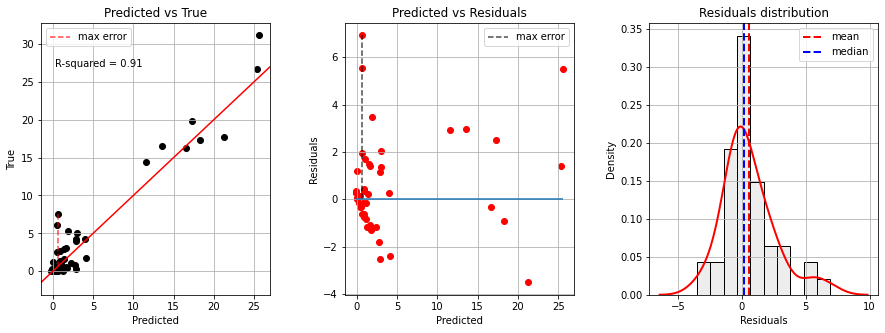


Figure 5: Residual plots of the predicted fracture intensity from the XGBoost model for the test dataset.

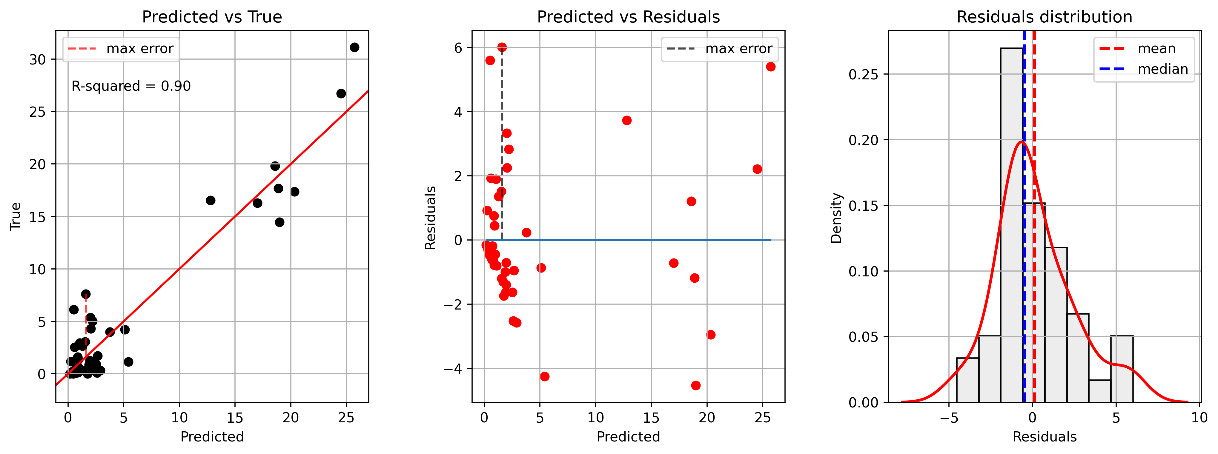


Figure 6: Residual plots of the predicted fracture intensity from the Random Forest model for the test dataset.

Table 1: Regression metrics for the testing set from both models, XGBoost and Random Forest. MAE: Mean Absolute Error; RMSE: Root Mean Squared Error; R2: R2 score.

|  |  |  |
| --- | --- | --- |
| **Metrics** | **Model** | |
|  | **XGBoost** | **Random Forest** |
| MAE | 1.47 | 1.73 |
| RMSE | 2.13 | 2.29 |
| R2 | 0.91 | 0.90 |

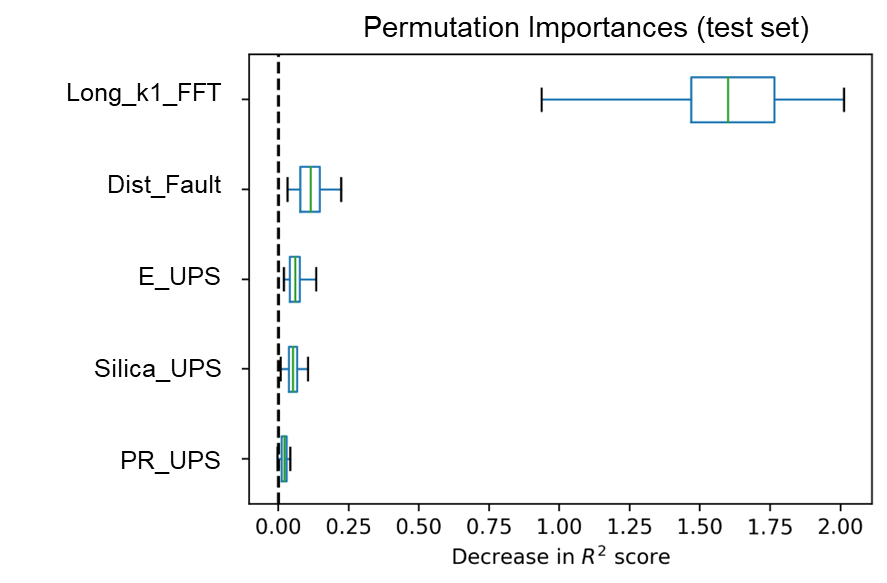


Figure 7: Boxplot analysis from permutation feature importance evaluated in the test set from the XGBoost fitted model. Curvature, distance to fault, and Young’s Modulus (E) have the most impact on the R² score between the predictions and the actual value of the target.

Repeating the exploratory data analysis, we confirm the P-Impedance has no direct correlation with the fracture intensity (Figure 8). However, it is essential to the model once the samples with the highest fracture intensity are in a limited impedance range (Figure 9). The P-impedance reasonably impacts the permutation importance evaluation, with a contribution close to the distance to fault attribute (Figure 10).

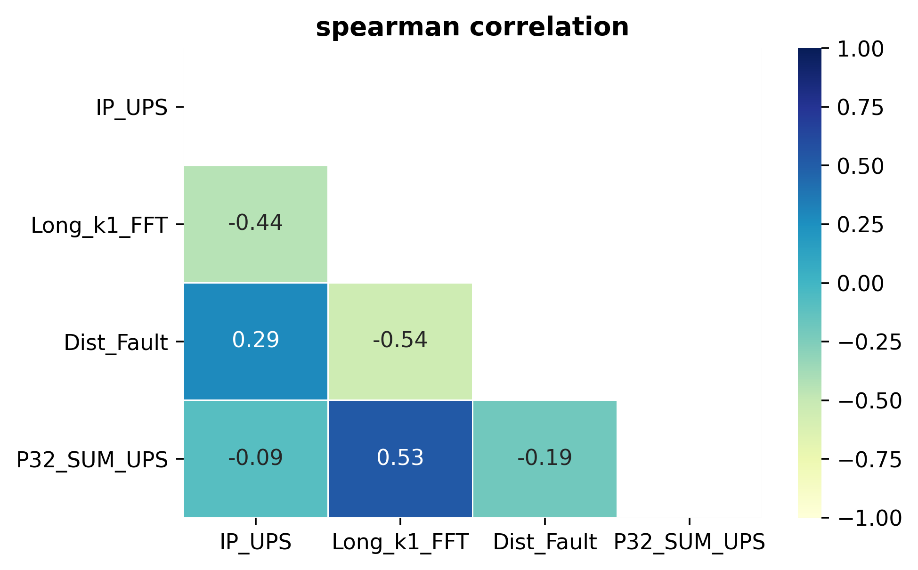


Figure 8: Spearman correlation matrix showing the measured nonlinear correlation between the input and target features.

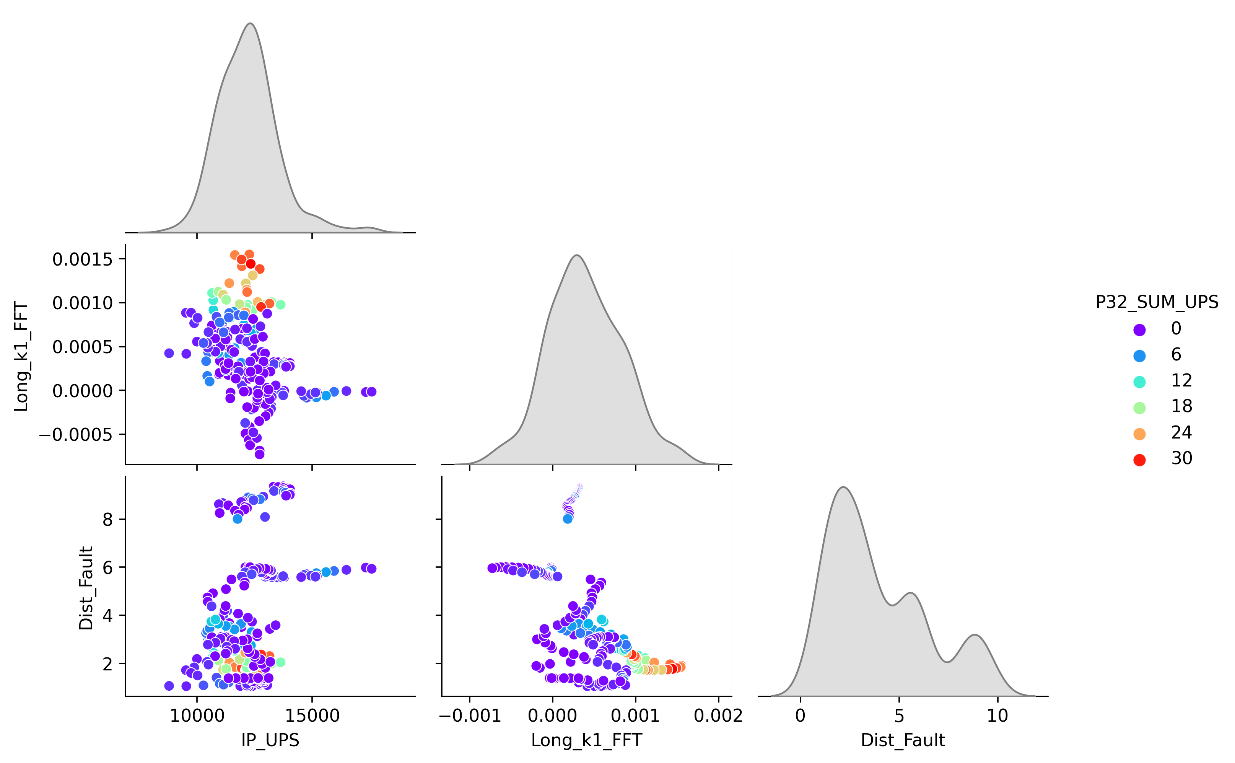


Figure 9: Joint cross-plot analysis between the input features colored by the levels of fracture intensity.

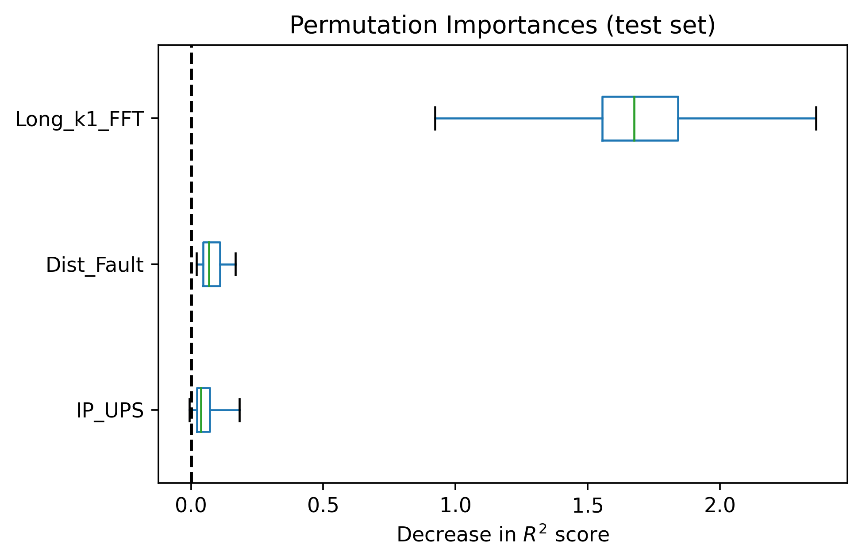


Figure 10: Boxplot analysis from permutation feature importance evaluated in the test set from the XGBoost fitted model. The rank of feature importance is k1 curvature, distance to fault, and P-Impedance, respectively.

The changes in model performance are minimal, with reasonable residual behavior and similar metrics. The residual plots (Figure 11) and regression metrics (Table 2) confirm the robustness of the model, even excluding the silica content and switching Young’s Modulus for the P-Impedance. We do not evaluate the use of Random Forest in this stage.

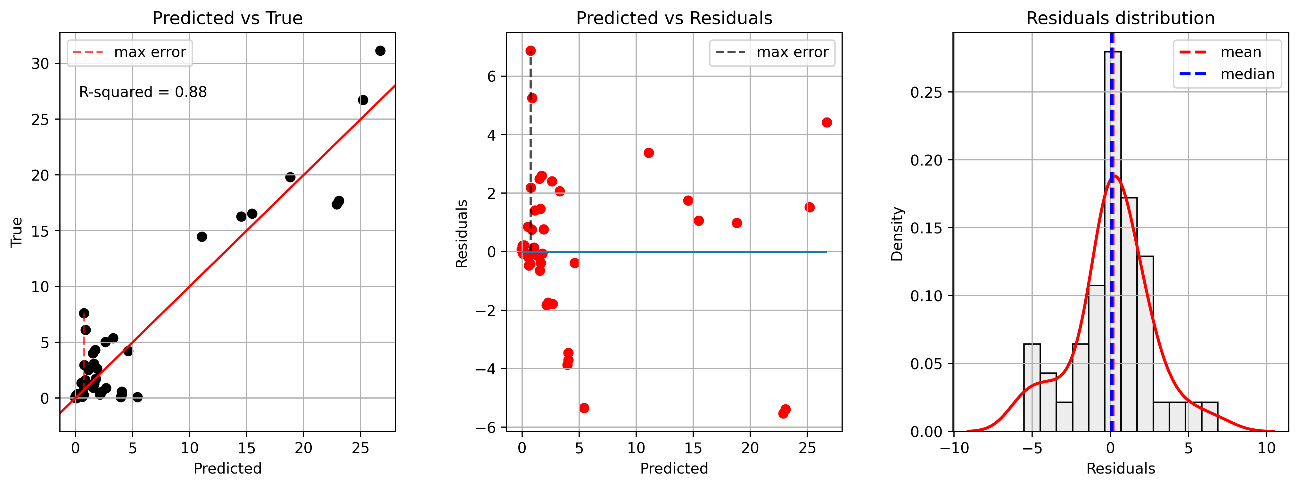


Figure 11: Residual plots of the predicted fracture intensity from the Random Forest model for the test dataset.

Table 2: Regression metrics for the testing set from both models, XGBoost and Random Forest. MAE: Mean Absolute Error; RMSE: Root Mean Squared Error; R2: R2 score.

|  |  |
| --- | --- |
| **Metrics** | **Model** |
|  | **XGBoost** |
| MAE | 1.74 |
| RMSE | 2.52 |
| R2 | 0.88 |

A blind-well test is not performed, as the five wells analyzed each have a fracture intensity and input features relationship pattern. An alternative to evaluate the model is to predict the fracture intensity in well H. This well does not have BHI logs. Therefore, it is not possible to measure the fracture intensity.

However, it does have the other necessary information for the prediction (P-Impedance, k1 curvature, and distance to fault). It also has a continuous sampling of rock data. By comparing our model’s predictions with the rock sample, we gained insight into the overall performance of our model (Figure 12). The results show that our predicted fracture intensity aligned with the rock sample, with higher values of predicted fracture intensity corresponding to higher, more fractured samples and the lowest levels of predicted fracture intensity associated with fine-grained carbonate rocks without any evidence of fractures.

If the model prediction is correct, at least in identifying trends, there is a thick continuous interval of a fractured reservoir with more than 100m. In this case, this interval can potentially have an anomalous production behavior associated with the so-called excess permeability zones, as defined by Fernandes-Ibanez et al. (2022a). It is worth noting that we excluded igneous samples from the training phase, and therefore, we cannot evaluate the model’s performance in those intervals.

The above analysis shows a potential applicability of our study. When a reasonable number of wells with BHI logs is available to train a machine learning model, one can use it to run predictions of fracture intensity in other wells without BHI logs, at least to be aware of zones more or less fractured, and to have a quantitative measurement of this potential.

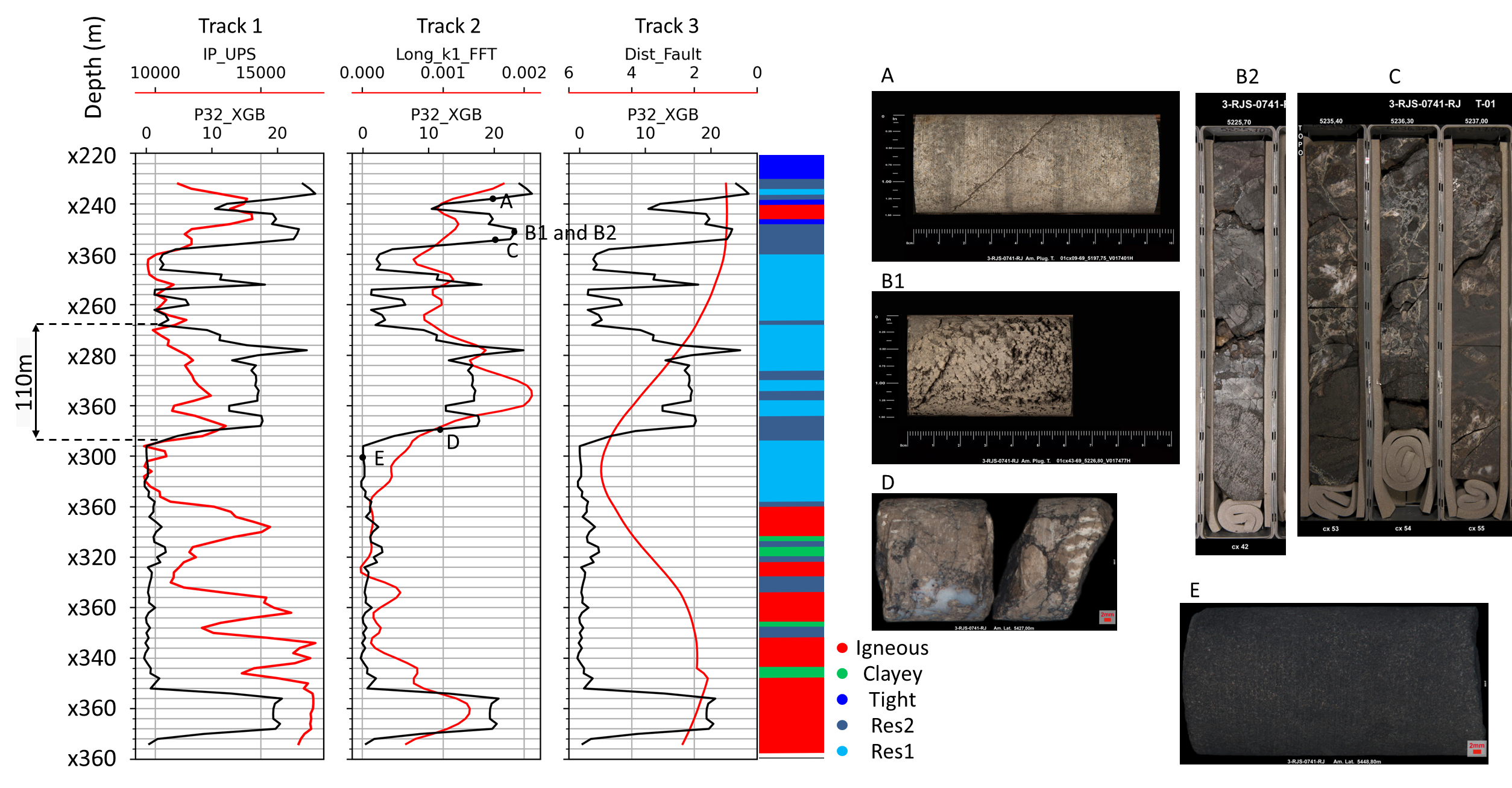


Figure 12: Well log plot from Well H with the 1D fracture intensity predicted by the XGBoost model and the input features, P-Impedance, k1 curvature, and distance to fault (measured in amount of samples). Sidetrack with the upscaled eletrofacies. Points A to E are examples of rock samples (sidewall and well cores) that corroborate the response of the predicted fracture intensity. Samples A to D have a higher degree of fracture intensity with evidence of karstification (B1), silica precipitation (C and D), and dissolution-enhanced fractures (C). Sample E is a fine-grained carbonate with the absence of fractures.

# Discussions

To improve the accuracy of our model, we need to obtain more data. We may not have captured the full range of fracture patterns with only five wells. Additionally, we had to exclude igneous samples due to their anomalous mechanical behavior and the poor quality of BHI logs in those intervals. With more wells, at least one could have a proper recording of BHI logs in the igneous facies. In that case, to deal with the different mechanical behaviors from very distinct mechanical facies, we could add as input features an electrofacies log. Going further into it, we could also add another input feature, the bed thickness, which is another property that controls the mechanical behavior of the stratigraphic unit (McQuillan, 1973; Wennberg et al., 2006; Ferrill and Morris, 2008).

Finally, Fine-tuning the models using K-Fold Cross Validation and Grid Search can also improve their generalization and prevent overfitting.

# Conclusions

Our data-driven approach has shown great potential in dealing with the complex relationship between seismic scale strain, brittleness, and fracture intensity as measured in well logs. The permutation importance results reveal that curvature is the most important feature, followed by the distance to fault attribute and Young’s Modulus. While Young's Modulus is typically used as an elastic curve to indicate brittleness, our results suggest that P-Impedance can replace it in cases where pre-stack or partial-stack seismic data is unavailable.

Both tree-ensemble models have good performance and successfully predict fracture intensity in both training and test datasets. XGBoost performs better than Random Forest. Further improvements are possible with additional features and fine-tuning of the model.

Our approach has successfully predicted the fracture intensity of a hundred-meter-thick carbonate mound of the Barra-Velha Formation. The prediction agrees with rock sampling information. To continue validating our approach, we plan to run the model on other nearby wells and the entire seismic volume using the three seismic attributes.

# Acknowledgments

The authors thank Equinor for financing the R&D project and the Agência Nacional do Petróleo (ANP) and CGG for providing the data used in this research.

# References

Altmann, A., Toloşi, L., Sander, O., Lengauer, T., 2010. Permutation importance: A corrected feature importance measure. Bioinformatics 26, 1340–1347. https://doi.org/10.1093/bioinformatics/btq134

Babasafari, A.A., Chinelatto, G.F., Vidal, A.C., 2022. Fault and fracture study by incorporating borehole image logs and supervised neural network applied to the 3D seismic attributes: a case study of pre-salt carbonate reservoir, Santos Basin, Brazil. Petroleum Science and Technology 0, 1–20. https://doi.org/10.1080/10916466.2021.2025072

Bhattacharya, S., Ghahfarokhi, P.K., Carr, T.R., Pantaleone, S., 2019. Application of predictive data analytics to model daily hydrocarbon production using petrophysical, geomechanical, fiber-optic, completions, and surface data: A case study from the Marcellus Shale, North America. Journal of Petroleum Science and Engineering 176, 702–715. https://doi.org/10.1016/j.petrol.2019.01.013

Botter, C., Cardozo, N., Hardy, S., Lecomte, I., Escalona, A., 2014. From mechanical modeling to seismic imaging of faults: A synthetic workflow to study the impact of faults on seismic. Marine and Petroleum Geology 57, 187–207. https://doi.org/10.1016/j.marpetgeo.2014.05.013

Braga, I.L.S., Moraes, F.S., 2013. High-resolution gathers by inverse Q filtering in the wavelet domain. Geophysics 78, V53–V61. https://doi.org/https://doi.org/10.1190/geo2011-0508.1

Breiman, L., 2001. Random Forests. Machine Learning 45, 5–32. https://doi.org/https://doi.org/10.1007/9781441993267\_5

Chen, T., Guestrin, C., 2016. XGBoost: A Scalable Tree Boosting System. https://doi.org/10.1145/2939672.2939785

Chen, Y., Liao, Z., Fu, L.Y., Zhou, G., Xu, L., Marfurt, K.J., Mu, X., Zou, H., 2020. Effect of main frequencies on characterizing fault damage-zones using forward modeling and attribute of variance. Interpretation 8. https://doi.org/10.1190/int-2020-0017.1

Correa, R.S.M., Pereira, C.E.L., Cruz, F.A.S., Lisboa, S.N.D., Junior, M.P.A., Carvalho, B.R.B.M., Souza, V.H.P., Rocha, C.H.A., Araujo, F.G., 2019. Integrated Seismic-Log-Core-Test Fracture Characterization, Barra Velha Formation, Pre-salt of Santos Basin. Search and Discovery. https://doi.org/10.1306/42425Correa2019

Fernández-Ibáñez, F., Bowen, M., Jones, G., Mimoun, J., 2022. Excess permeability in the Brazil pre-salt: Non-matrix types and diagnostic indicators. AAPG Bulletin 1–38. https://doi.org/10.3997/2214-4609.201982006

Fernandez-Ibanez, F., Nolting, A., Breithaupt, C.I., Darby, B., Mimoun, J., Henares, S., 2022. The properties of faults in the Brazil pre-salt: A reservoir characterization perspective. Marine and Petroleum Geology 146, 105955. https://doi.org/10.1016/J.MARPETGEO.2022.105955

Ferrill, D.A., Morris, A.P., 2008. Fault zone deformation controlled by carbonate mechanical stratigraphy, Balcones fault system, Texas. AAPG Bulletin 92, 359–380. https://doi.org/10.1306/10290707066

Francisco, J., Morales, V., Abbas, A., Furlan, G., Ricardo, C., David, C., Campane, A., 2023. Impact of seismic data conditioning on the identification of structural elements: A case of study from the pre-salt reservoir, Santos Basin, Brazil. Journal of Applied Geophysics 211, 104980. https://doi.org/10.1016/j.jappgeo.2023.104980

Grinsztajn, L., Oyallon, E., Varoquaux, G., 2022. Why do tree-based models still outperform deep learning on tabular data? 36th Conference on Neural Information Processing Systems (NeurIPS 2022).

H. McQuillan, 1973. Small-Scale Fracture Density in Asmari Formation of Southwest Iran and its Relation to Bed Thickness and Structural Setting. AAPG Bulletin 57, 2367–2385. https://doi.org/10.1306/83d9131c-16c7-11d7-8645000102c1865d

Hampson, D.P., Schuelke, J.S., Quirein, J.A., 2001. Use of multiattribute transforms to predict log properties from seismic data. Geophysics 66, 220–236. https://doi.org/10.1190/1.1444899

Luo, Y., Marhoon, M., Dossary, S. Al, Alfaraj, M., 2002. Edge-preserving smoothing and applications. The Leading Edge 21, 136–158. https://doi.org/https://doi.org/10.1190/1.1452603

Lupinacci, W.M., Fatah, T.Y.A., Carmo, M.C. do, Freire, A.F.M., Gamboa, L.A.P., 2023. Controls of fracturing on porosity in pre-salt carbonate reservoirs. Energy Geoscience 4, 100146. https://doi.org/10.1016/j.engeos.2022.100146

Man, H.Q., Hien, D.H., Thong, K.D., Dung, B.V., Hoa, N.M., Hoa, T.K., Kieu, N. Van, Ngoc, P.Q., 2021. Hydraulic flow unit classification and prediction using machine learning techniques: A case study from the nam con son basin, offshore vietnam. Energies 14. https://doi.org/10.3390/en14227714

Mandal, P.P., Rezaee, R., 2019. Facies classification with different machine learning algorithm – An efficient artificial intelligence technique for improved classification. Exploration Geophysics 2019. https://doi.org/10.1080/22020586.2019.12072918

Nguyen, N.T., Tran, N.T.H., Hoang, K.S., Tran, V.T., 2022. Supervised machine learning application of lithofacies classification for a hydrodynamically complex gas - condensate reservoir in Nam Con Son basin. Petrovietnam Journal 6, 27–35. https://doi.org/10.47800/pvj.2022.06-03

Savage, H.M., Brodsky, E.E., 2011. Collateral damage: Evolution with displacement of fracture distribution and secondary fault strands in fault damage zones. Journal of Geophysical Research: Solid Earth 116. https://doi.org/10.1029/2010JB007665

Tanaka, A.P.B., Gomes Borges, J.P., Correa de Matos, G., Relvas Campos, M.T., Cunha, B.M., Brandi de Souza, R., Natan de Moraes Caldeira, J., Soares de Oliveira, T.A., Marçon, D.R., Martins Lima, A.P., 2022. Fault-related fracture modeling in a pre-salt lacustrine carbonate reservoir from Santos Basin, offshore Brazil: Predicting preferential fluid flow paths using 3D geological and flow simulation models. Marine and Petroleum Geology 135, 105392. https://doi.org/10.1016/j.marpetgeo.2021.105392

Wennberg, O.P., De Oliveira Ramalho, F., Virgolino Mafia, M., Lapponi, F., Chandler, A.S., Gomis Cartesio, L.E., Hunt, D.W., 2023. The characteristics of natural open fractures in acoustic borehole image logs from the pre-salt Barra Velha formation, Santos Basin, Brazil. Journal of Structural Geology 167. https://doi.org/10.1016/j.jsg.2023.104794

Wennberg, O.P., Svånå, T., Azizzadeh, M., Aqrawi, A.M.M., Brockbank, P., Lyslo, K.B., Ogilvie, S., 2006. Fracture intensity vs. mechanical stratigraphy in platform top carbonates: The Aquitanian of the Asmari Formation, Khaviz Anticline, Zagros, SW Iran. Petroleum Geoscience 12, 235–245. https://doi.org/10.1144/1354-079305-675

Wu, X., Liang, L., Shi, Y., Fomel, S., 2019. FaultSeg3D: Using synthetic data sets to train an end-to-end convolutional neural network for 3D seismic fault segmentation. Geophysics 84, IM35–IM45. https://doi.org/10.1190/geo2018-0646.1