

**Application of Time-Series and Advanced  
Machine Learning Techniques for  
Breakdown Pressure Estimation**

BY  
**Subhan Aliyev**

A Thesis Presented to the  
DEANSHIP OF GRADUATE STUDIES

**KING FAHD UNIVERSITY OF PETROLEUM & MINERALS**

DHAHRAN, SAUDI ARABIA

In Partial Fulfillment of the  
Requirements for the Degree of

**MASTER OF SCIENCE**

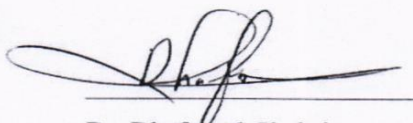
In

**PETROLEUM ENGINEERING**

**February 2025**

KING FAHD UNIVERSITY OF PETROLEUM & MINERALS  
DHAHRAN- 31261, SAUDI ARABIA  
DEANSHIP OF GRADUATE STUDIES

This thesis, written by **Subhan Aliyev** under the direction of his thesis advisor and approved by his thesis committee, has been presented and accepted by the Dean of Graduate Studies, in partial fulfillment of the requirements for the degree of **MASTER OF SCIENCE IN PETROLEUM ENGINEERING**.



Dr. Dhafar Al-Shehri  
Department Chairman



Prof. Suliman Salem Al-Homidan  
Dean of Graduate Studies

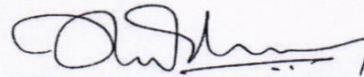
02.25.2025  
Date



Dr. Talal Al Shafloot  
(Advisor)



Dr. Murtada Al-Jawad  
(Co-advisor)



Dr. Abdulazeez Abdulraheem  
(Member)



© Subhan Aliyev

2025

*To my family, whose endless encouragement, boundless love, and belief in me have  
been my constant source of strength and inspiration.*

## ACKNOWLEDGMENTS

I would like to express my deepest gratitude to my thesis advisor, **Dr. Talal Al Shafloot**, for his invaluable guidance, support, and encouragement throughout my MSc journey. His expertise, insightful feedback, and dedication have been instrumental in shaping the direction of my research and ensuring its successful completion. His patience and continuous motivation have greatly contributed to my academic and professional growth.

I am also sincerely grateful to my committee members, **Dr. Murtada Al-Jawad** and **Dr. Abdulazez Abdulraheem**, for their valuable feedback, constructive criticism, and unwavering support. Their expertise and insightful recommendations have significantly strengthened the quality of this research, allowing me to refine my approach and broaden my understanding of the subject matter.

Finally, I am deeply thankful to my peers, colleagues, and everyone who has supported me throughout this journey. Their encouragement, cooperation, and shared knowledge have played a vital role in the successful completion of this work. I am truly grateful to all those who contributed in any way to this thesis, making this experience both fulfilling and enriching.

# TABLE OF CONTENTS

|                                                                                           |             |
|-------------------------------------------------------------------------------------------|-------------|
| <b>ACKNOWLEDGMENTS .....</b>                                                              | <b>V</b>    |
| <b>TABLE OF CONTENTS .....</b>                                                            | <b>VI</b>   |
| <b>LIST OF TABLES.....</b>                                                                | <b>IX</b>   |
| <b>LIST OF FIGURES.....</b>                                                               | <b>X</b>    |
| <b>LIST OF ABBREVIATIONS.....</b>                                                         | <b>XIV</b>  |
| <b>ABSTRACT .....</b>                                                                     | <b>XV</b>   |
| <b>خلاصة.....</b>                                                                         | <b>XVII</b> |
| <b>CHAPTER 1 INTRODUCTION AND LITERATURE REVIEW.....</b>                                  | <b>1</b>    |
| <b>1.1 General information about hydraulic fracturing .....</b>                           | <b>1</b>    |
| 1.1.1 Clarification of Key Pressure Terminologies .....                                   | 2           |
| 1.1.2 Idealized Pressure-Volume Relationship .....                                        | 4           |
| <b>1.2 Literature review.....</b>                                                         | <b>6</b>    |
| 1.2.1 Traditional Methods to Determine Breakdown Pressure .....                           | 6           |
| 1.2.2 Application of Traditional Methods in Field and Experimental Studies.....           | 14          |
| 1.2.3 Factors Influencing Formation Breakdown Pressure .....                              | 15          |
| 1.2.4 Field Factors Influencing Breakdown Pressure .....                                  | 22          |
| 1.2.5 Simulation and Machine Learning Applications for Breakdown Pressure Prediction..... | 25          |
| <b>1.3 Problem statement and Research Objectives .....</b>                                | <b>28</b>   |
| <b>CHAPTER 2 METHODOLOGY.....</b>                                                         | <b>31</b>   |
| <b>2.1 Database Summary .....</b>                                                         | <b>33</b>   |
| 2.1.1 Field Case Data Collection and Analysis.....                                        | 33          |
| 2.1.2 Experimental Case Data <i>Data Collection from the Literature:</i> .....            | 35          |
| 2.1.3 Time-Series Data for Breakdown Pressure Forecasting.....                            | 43          |
| <b>2.2 Simulation and Machine Learning Model Development .....</b>                        | <b>47</b>   |
| 2.2.1 Numerical Simulation for Breakdown Pressure Estimation.....                         | 47          |
| 2.2.2 Machine Learning Models for Field and Experimental Case.....                        | 50          |
| 2.2.3 Time-Series Models for Breakdown Pressure Forecasting .....                         | 57          |

|                                                     |           |
|-----------------------------------------------------|-----------|
| <b>2.3 Evaluation and Sensitivity Analysis.....</b> | <b>66</b> |
| 2.3.1 Evaluation Metrics .....                      | 66        |
| 2.3.2 Sensitivity Analysis Metrics.....             | 68        |

## **CHAPTER 3 RESULTS AND DISCUSSIONS ..... 72**

|                                                                              |            |
|------------------------------------------------------------------------------|------------|
| <b>3.1 Simulation Results .....</b>                                          | <b>72</b>  |
| 3.1.1 Performance of the Kirsch Model.....                                   | 72         |
| 3.1.2 Performance of the Hubbert and Willis Model .....                      | 73         |
| 3.1.3 Performance of Haimson and Fairhurst Model .....                       | 74         |
| 3.1.4 Performance of Rummel models.....                                      | 75         |
| 3.1.5 Performance of Schmitt and Zoback model .....                          | 78         |
| 3.1.6 Performance of Ito and Hayashi model .....                             | 79         |
| 3.1.7 Performance of Aadnoy and Balayneh models .....                        | 80         |
| 3.1.8 Evaluation Metrics of Simulation Models.....                           | 83         |
| <b>3.2 Machine Learning Results for Experimental Case .....</b>              | <b>84</b>  |
| 3.2.1 One-Hot Encoding.....                                                  | 84         |
| 3.2.2 Correlation Analysis.....                                              | 85         |
| 3.2.3 Model Training and Cross-Validation .....                              | 86         |
| 3.2.4 Ranking and Sensitivity Analysis .....                                 | 93         |
| 3.2.5 Comparison of Correlation Analysis and Sensitivity Analysis .....      | 96         |
| 3.2.6 Re-Training with Top Features.....                                     | 97         |
| 3.2.7 Pressurization Rate Analysis .....                                     | 98         |
| <b>3.3 Machine Learning Results for Field Case .....</b>                     | <b>101</b> |
| 3.3.1 Model Training and Cross Validation.....                               | 101        |
| 3.3.2 Sensitivity Analysis and Ranking .....                                 | 107        |
| 3.3.3 Re-Training with Top Features.....                                     | 110        |
| <b>3.4 Time Series Forecasting of Breakdown Pressure .....</b>               | <b>112</b> |
| 3.4.1 Model Performance.....                                                 | 112        |
| 3.4.2 MAE and MSE .....                                                      | 117        |
| 3.4.3 Time .....                                                             | 118        |
| <b>3.5 Discussions .....</b>                                                 | <b>120</b> |
| 3.5.1 Accuracy and Model Performance Across Approaches .....                 | 120        |
| 3.5.2 Discrepancy in Influential Parameters and the Need for Upscaling ..... | 121        |
| 3.5.3 Correlation vs. Sensitivity Analysis .....                             | 121        |
| 3.5.4 Time Series Application for Breakdown Pressure Forecasting .....       | 122        |

## **CHAPTER 4 CONCLUSIONS, FUTURE WORK AND RECOMMENDATIONS ..... 123**

|                                                  |            |
|--------------------------------------------------|------------|
| <b>4.1 Conclusions.....</b>                      | <b>123</b> |
| <b>4.2 Future Work and Recommendations .....</b> | <b>125</b> |

|                        |            |
|------------------------|------------|
| <b>REFERENCES.....</b> | <b>127</b> |
| <b>VITAE .....</b>     | <b>140</b> |

## LIST OF TABLES

|                                                                                                                                        |     |
|----------------------------------------------------------------------------------------------------------------------------------------|-----|
| <b>Table 1.</b> Database Summary .....                                                                                                 | 34  |
| <b>Table 2.</b> Summary of Dataset Gathered from Various Literature Sources on Breakdown<br>Pressure Across Different Rock Types ..... | 38  |
| <b>Table 3.</b> Dataset Gathered from Various Literature Sources on Pressurization Rate .....                                          | 39  |
| <b>Table 4.</b> Statistical Analysis of the Total Dataset .....                                                                        | 40  |
| <b>Table 5.</b> Summary of Applied Simulation Models .....                                                                             | 48  |
| <b>Table 6.</b> Hyperparameters for Machine Learning and Statistical Models.....                                                       | 64  |
| <b>Table 7.</b> Performance Comparison of Formation Breakdown Pressure Simulation Models<br>.....                                      | 83  |
| <b>Table 8.</b> Optimized Hyperparameter Ranges Utilized for Training Machine Learning<br>Algorithms .....                             | 90  |
| <b>Table 9.</b> Optimized Hyperparameter Ranges Utilized for Training Machine Learning<br>Algorithms .....                             | 103 |
| <b>Table 10.</b> Metrics comparison of Time-Series models .....                                                                        | 116 |
| <b>Table 11.</b> Time Comparison of Time-Series models .....                                                                           | 117 |

# LIST OF FIGURES

|                                                                                                                                                                                                             |    |
|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|----|
| <b>Figure 1.</b> Conceptual diagram depicting the hydraulic fracturing process (Zeng et al. 2020) .....                                                                                                     | 2  |
| <b>Figure 2.</b> Schematic of Pressure and Fluid Volume Interactions during an XLOT or Minifrac(Gaarenstroom et al. 1993) .....                                                                             | 6  |
| <b>Figure 3.</b> Wellbore Exhibiting Symmetrical Double Cracks (Zhang et al. 2018a).....                                                                                                                    | 10 |
| <b>Figure 4.</b> Pressure-time graph showing fracturing tests performed with various fluids and samples under identical stress conditions. (a) shale samples; (b) sandstone samples (Wu et al. 2021b) ..... | 21 |
| <b>Figure 5.</b> Schema of ML workflow applied in that study (Tariq et al. 2022) .....                                                                                                                      | 27 |
| <b>Figure 6.</b> Schematic representation of research phases and associated experiments .....                                                                                                               | 33 |
| <b>Figure 7.</b> Pearson and Spearman Correlation of Field Features .....                                                                                                                                   | 34 |
| <b>Figure 8.</b> The Box-Plot Distribution of Entire Dataset .....                                                                                                                                          | 41 |
| <b>Figure 9.</b> Histogram of 16 parameters, illustrating the diverse distribution types for each feature .....                                                                                             | 44 |
| <b>Figure 10.</b> Pearson Correlation Coefficients Plot .....                                                                                                                                               | 45 |
| <b>Figure 11.</b> Spearman Correlation Coefficient Plot.....                                                                                                                                                | 46 |
| <b>Figure 12.</b> Overall Structure of Catboost Model (Pandey et al. 2023) .....                                                                                                                            | 52 |
| <b>Figure 13.</b> Flowchart of TabNet Model Architecture (Arik and Pfister 2021) .....                                                                                                                      | 55 |
| <b>Figure 14.</b> A schematic representation of an LSTM cell showing the interaction between gates and the cell state (Yu et al. 2019) .....                                                                | 58 |
| <b>Figure 15.</b> A schematic representation of a Bi-LSTM network, showing the flow of information in both directions (Saxena et al., 2013).....                                                            | 60 |

|                                                                                                                                                  |    |
|--------------------------------------------------------------------------------------------------------------------------------------------------|----|
| <b>Figure 16.</b> Comparison of breakdown pressure estimated using the Kirsch approach with the actual measured pressure .....                   | 72 |
| <b>Figure 17.</b> Comparison of breakdown pressure estimated using the Hubbert and Willis approach with the actual measured pressure.....        | 73 |
| <b>Figure 18.</b> Comparison of breakdown pressure predicted by the Haimson and Fairhurst approach with the actual pressure values.....          | 74 |
| <b>Figure 19.</b> Evaluation of the breakdown pressure estimated through the Rummel 1 approach against actual pressure data. ....                | 75 |
| <b>Figure 20.</b> Assessment of breakdown pressure computed using the Rummel 2 approach compared to the actual pressure.....                     | 76 |
| <b>Figure 21.</b> Difference between the Formation Breakdown Pressure estimated by the Rummel 3 approach and the actual pressure.....            | 76 |
| <b>Figure 22.</b> Analysis of breakdown pressure determined by the Rummel 4 approach relative to the actual pressure values .....                | 77 |
| <b>Figure 23.</b> Comparison between breakdown pressure derived from the Schmitt and Zoback approach and the actual pressure readings.....       | 78 |
| <b>Figure 24.</b> Comparison of breakdown pressure calculated through the Ito and Hayashi approach with observed pressure values.....            | 79 |
| <b>Figure 25.</b> Assessment of breakdown pressure estimated using the Aadnoy and Balyaneh 1 approach against the actual measured pressure. .... | 80 |
| <b>Figure 26.</b> Assessment of breakdown pressure estimated using the Aadnoy and Balyaneh 2 approach against the actual measured pressure. .... | 81 |
| <b>Figure 27.</b> The transformation of the categorical variables to numerical .....                                                             | 84 |

|                                                                                                                           |     |
|---------------------------------------------------------------------------------------------------------------------------|-----|
| <b>Figure 28.</b> Heatmap of Spearman and Pearson Correlations for Breakdown Pressure<br>Features .....                   | 85  |
| <b>Figure 29.</b> A 5-Fold Cross-Validation approach applied in the research.....                                         | 86  |
| <b>Figure 30.</b> Estimated optimal parameters for Ensemble and TabNet models.....                                        | 87  |
| <b>Figure 31.</b> Neuron Number Sensitivity For Kolmogorov-Arnold Network .....                                           | 88  |
| <b>Figure 32.</b> Final Structure of Kolmogorov-Arnold Network.....                                                       | 89  |
| <b>Figure 33.</b> Metrics comparison for trained models .....                                                             | 91  |
| <b>Figure 34.</b> Training and Testing prediction for Kolmorov-Arnold Network.....                                        | 92  |
| <b>Figure 35.</b> SHAP Analysis of Feature Contributions for XGBoost, LightGBM, and<br>CatBoost Models.....               | 93  |
| <b>Figure 36.</b> TabNet Attention Score Analysis for BP Features .....                                                   | 94  |
| <b>Figure 37.</b> Distribution of Breakdown Pressure Features according to KAN weights ....                               | 95  |
| <b>Figure 38.</b> Top Chosen Features for Re-Training .....                                                               | 96  |
| <b>Figure 39.</b> Evaluation Metrics Comparison for Re-Trained Models .....                                               | 97  |
| <b>Figure 40.</b> Application of Frish-Waugh-Lovell Analysis for Correlation Correction .....                             | 98  |
| <b>Figure 41.</b> Modeled Relationship Between Pressurization Rate and Breakdown Pressure<br>for Various Rock Types ..... | 99  |
| <b>Figure 42.</b> Cross-validation for Field Data .....                                                                   | 101 |
| <b>Figure 43.</b> Neuron Sensitivity for Field Case .....                                                                 | 102 |
| <b>Figure 44.</b> Training Results for Field Case .....                                                                   | 104 |
| <b>Figure 45.</b> Kolmogorov-Arnold Training and Testing Results for Field Case .....                                     | 105 |
| <b>Figure 46.</b> SHAP Analysis of Field Data .....                                                                       | 108 |
| <b>Figure 47.</b> Distribution of Weights according to TabNet for Field Case.....                                         | 108 |

|                                                                                        |     |
|----------------------------------------------------------------------------------------|-----|
| <b>Figure 48.</b> Distribution of Features regarding KAN for Field Case .....          | 109 |
| <b>Figure 49.</b> Top Chosen Features for Re-Training (Field Case) .....               | 110 |
| <b>Figure 50.</b> Re-Training Results for Field Case .....                             | 111 |
| <b>Figure 51.</b> Forecasted BP versus Actual BP for LSTM.....                         | 112 |
| <b>Figure 52.</b> Forecasted BP versus Actual BP for BI-LSTM case.....                 | 113 |
| <b>Figure 53.</b> Forecasted BP versus Actual BP for SARIMA case .....                 | 114 |
| <b>Figure 54.</b> Forecasted BP versus Actual BP for VARMA.....                        | 115 |
| <b>Figure 55.</b> Representation of MSE and MAE for different Time-Series models ..... | 119 |
| <b>Figure 56.</b> Representation of Time differences for Time-Series models.....       | 119 |

## LIST OF ABBREVIATIONS

|                         |   |                                                   |
|-------------------------|---|---------------------------------------------------|
| <b>BP</b>               | : | Breakdown Pressure                                |
| <b>ML</b>               | : | Machine Learning                                  |
| <b>TabNet</b>           | : | Tabular Neural Network                            |
| <b>KAN</b>              | : | Kolmogorov-Arnold Network                         |
| <b>LightGBM</b>         | : | Light Gradient Boosting Machine                   |
| <b>CatBoost</b>         | : | Categorical Boosting Algorithm                    |
| <b>XGBoost</b>          | : | Extreme Gradient Boosting                         |
| <b>LSTM</b>             | : | Long Short-Term Memory                            |
| <b>Bi-LSTM</b>          | : | Bidirectional Long Short-Term Memory              |
| <b>SARIMA</b>           | : | Seasonal AutoRegressive Integrated Moving Average |
| <b>VARMA</b>            | : | Vector AutoRegressive Moving Average              |
| <b>MSE</b>              | : | Mean Squared Error                                |
| <b>MAE</b>              | : | Mean Absolute Error                               |
| <b>RMSE</b>             | : | Root Mean Squared Error                           |
| <b><math>R^2</math></b> | : | Coefficient of Determination                      |
| <b>AAPE</b>             | : | Average Absolute Percentage Error                 |
| <b>SD</b>               | : | Standard Deviation                                |
| <b>FWL</b>              | : | Frisch-Waugh-Lovell Theorem                       |
| <b>SHAP</b>             | : | SHapley Additive exPlanations                     |

## **ABSTRACT**

Full Name : Subhan Aliyev

Thesis Title : Application of Time-Series and Advanced Machine Learning  
Techniques for Breakdown Pressure Estimation

Major Field : Petroleum Engineering

Date of Degree : February 2025

Accurately predicting formation breakdown pressure (FBP) is essential for optimizing hydraulic fracturing operations, particularly in sandstone formations subjected to various injection fluids such as water, supercritical CO<sub>2</sub>, and nitrogen. This research develops advanced predictive models to address the limitations of traditional numerical simulations and improve the accuracy of FBP estimation through machine learning and time-series forecasting techniques. By integrating field case data, experimentally collected datasets, and numerical simulation outputs, this study provides a comprehensive approach to understanding the complex geomechanical interactions influencing breakdown pressure.

The study involved analyzing key operational and geological parameters, including injection rate, viscosity, confining pressure, minimum horizontal stress, tensile strength, permeability, and porosity. These variables were used to train and evaluate machine learning models, including TabNet, CatBoost, LightGBM, Kolmogorov-Arnold Network

(KAN), and XGBoost. Additionally, long short-term memory (LSTM) and Bi-LSTM models were employed to capture temporal variations in breakdown pressure, enabling real-time forecasting. The models were assessed using metrics such as mean squared error (MSE), mean absolute error (MAE), and  $R^2$ , demonstrating that TabNet and CatBoost achieved the highest predictive accuracy, outperforming traditional numerical models.

A key finding of this research is the discrepancy in influential parameters across experimental and field-scale datasets. While previous studies suggested that pressurization rate, sample size, and hole radius significantly affect breakdown pressure, these factors exhibited minimal impact in field-scale models. This highlights the challenge of scaling experimental insights to real-world applications and emphasizes the need for refined methodologies to improve the transferability of laboratory-derived findings. The study also evaluates the limitations of existing numerical simulations in capturing field-scale complexity, reinforcing the importance of integrating data-driven models for more reliable predictions.

The integration of time-series forecasting further enhances breakdown pressure estimation, offering a dynamic approach for real-time monitoring. LSTM-based models effectively captured depth-varying breakdown pressure trends, demonstrating their potential for hydraulic fracturing optimization. By leveraging numerical simulations, field case data, and machine learning techniques, this research provides practical insights for improving wellbore stability, fracture propagation, and overall efficiency in hydraulic fracturing operations.

## خلاصة

الاسم الكامل: سبحان علييف

عنوان الرسالة: تطبيقات نماذج السلاسل الزمنية وتقنيات التعلم الآلي المتقدمة للتنبؤ بضغط تكسر الصخور

التخصص: هندسة البترول

تاريخ الدرجة العلمية: فبراير 2025

أمرًا ضروريًا لتحسين عمليات التكسير الهيدروليكي، خاصة في (FBP) يُعد التنبؤ الدقيق بضغط الانهيار التكويني التكوينات الرملية التي تخضع لحقن سوائل مختلفة مثل الماء، وثنائي أكسيد الكربون فوق الحرج، والنيتروجين. تهدف هذه الدراسة إلى تطوير نماذج تنبؤية متقدمة لمعالجة قيود المحاكاة العددية التقليدية وتحسين دقة تقدير ضغط الانهيار من خلال تقنيات التعلم الآلي والتنبؤ بالسلاسل الزمنية. من خلال دمج بيانات الحقول، والمجموعات التجريبية، ونماذج المحاكاة العددية، توفر هذه الدراسة نهجًا شاملاً لفهم التفاعلات الجيوميكانيكية المعقدة التي تؤثر على ضغط الانهيار.

شملت الدراسة تحليل المعلمات التشغيلية والجيولوجية الرئيسية، مثل معدل الحقن، واللزوجة، وضغط الاحتواء، والإجهاد الأفقي الأدنى، وقوة الشد، والنفاذية، والمسامية. تم استخدام هذه البيانات لتدريب وتقييم نماذج التعلم الآلي، Kolmogorov-Arnold Network (KAN) و LightGBM و CatBoost و TabNet بما في ذلك والنماذج ثنائية الاتجاه (LSTM) بالإضافة إلى ذلك، تم استخدام نماذج الذاكرة طويلة وقصيرة المدى. XGBoost و لالتقاط التغيرات الزمنية في ضغط الانهيار، مما يتيح التنبؤ في الوقت الفعلي. تم تقييم أداء النماذج (Bi-LSTM) ، حيث  $R^2$  ومعامل التحديد (MAE) ومتوسط الخطأ المطلق (MSE) باستخدام معايير مثل متوسط مربع الخطأ أعلى دقة تنبؤية، متفوقة على النماذج العددية التقليدية CatBoost و TabNet أظهرت نماذج

أحد النتائج الرئيسية لهذه الدراسة هو التباين في تأثير العوامل المختلفة بين البيانات التجريبية وبيانات الحقول. في حين أشارت الدراسات السابقة إلى أن معدل الضغط، وحجم العينة، ونصف قطر البئر لها تأثير كبير على ضغط

الانهيار، فقد أظهرت هذه العوامل تأثيرًا ضئيلاً في النماذج القائمة على بيانات الحقول. يبرز هذا التباين التحديات المرتبطة بتوسيع نطاق البيانات التجريبية لتطبيقات الحقول الحقيقية، ويؤكد على الحاجة إلى تطوير منهجيات محسنة لنقل نتائج المختبر إلى البيئات الميدانية. كما تقيم الدراسة قيود المحاكاة العددية التقليدية في تمثيل التعقيدات الميدانية، مما يعزز أهمية دمج النماذج المستندة إلى البيانات للحصول على تنبؤات أكثر دقة.

يعمل دمج نماذج التنبؤ بالسلاسل الزمنية على تعزيز دقة تقدير ضغط الانهيار، مما يوفر نهجاً ديناميكياً للمراقبة في قدرتها على التقاط الاتجاهات المتغيرة لضغط الانهيار مع LSTM الوقت الحقيقي. أظهرت النماذج المستندة إلى العمق، مما يؤكد إمكاناتها في تحسين عمليات التكسير الهيدروليكي. من خلال الجمع بين المحاكاة العددية، وبيانات الحقول، وتقنيات التعلم الآلي، توفر هذه الدراسة رؤية عملية لتحسين استقرار الآبار، ومسارات التكسير، والكفاءة التشغيلية العامة في عمليات التكسير الهيدروليكي.

# **CHAPTER 1**

## **INTRODUCTION AND LITERATURE REVIEW**

### **1.1 General information about hydraulic fracturing**

Hydraulic fracturing, often referred to as fracking, is a crucial completion technique widely used in the petroleum industry to enhance the production of hydrocarbons from reservoirs with low permeability. This technique requires the injection of high-pressure fluids into a wellbore to generate fractures in the surrounding rock formations. These fractures serve as pathways that allow hydrocarbons to move more freely from the rock pores toward the production well, significantly boosting the production rates of oil and gas wells

The process of hydraulic fracturing has transformed the energy landscape by unlocking vast reserves of previously inaccessible hydrocarbons, particularly in shale and tight sand formations. The success of hydraulic fracturing in unconventional reservoirs has been a key driver in the boom of shale gas and oil production in regions such as North America. Despite its widespread use and economic benefits, hydraulic fracturing is a complex operation that requires a deep understanding of subsurface geology, fluid dynamics, and rock mechanics to be carried out effectively and safely.

High-pressure fluid, usually a mixture of water, proppant (such as sand), and chemical additives, is injected into the wellbore. The fluid pressure creates fractures in the rock, and the proppant keeps these fractures open once the injection stops. As the fractures propagate

away from the wellbore into the surrounding rock formation, permeability is enhanced, allowing hydrocarbons to flow more easily to the well.

After the fracturing process, the pressure is released, and some of the injected fluid flows back to the surface. The well is then put into production, and hydrocarbons flow from the reservoir to the wellbore and up to the surface. This entire process requires careful monitoring and precise control to ensure that the fractures are created in the desired locations and that they propagate effectively to enhance hydrocarbon recovery (Figure 1).

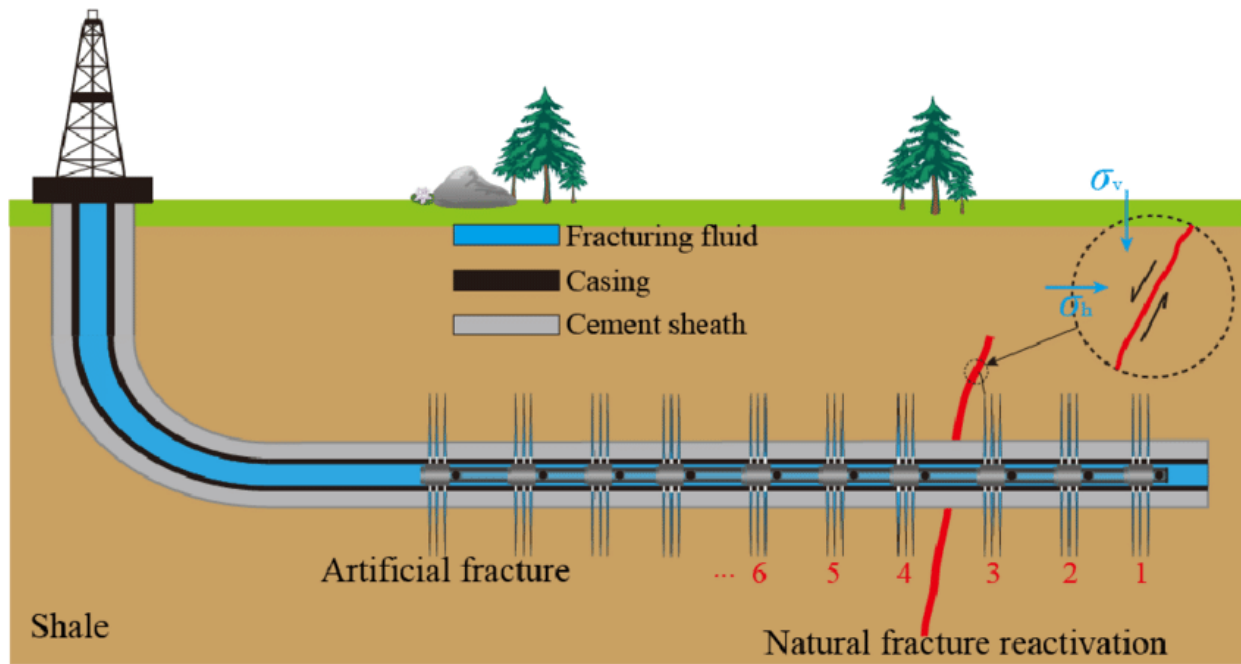


Figure 1. Conceptual diagram depicting the hydraulic fracturing process (Zeng et al. 2020)

### 1.1.1 Clarification of Key Pressure Terminologies

In hydraulic fracturing, two critical pressure points are frequently misunderstood: Fracture Initiation Pressure and Formation Breakdown Pressure. For eliminating any confusion, the specific denotation utilized in this research are outlined below.

**Fracture Initiation Pressure (FIP):** Fracture Initiation Pressure (FIP) is the minimum pressure required to create a new fracture at the wellbore wall. These fractures typically form at the wall and extend only a short distance, usually a fraction of an inch. While this pressure is important for starting the fracture process, it is less critical than the Formation Breakdown Pressure (FBP), which determines when fractures occur, enabling the flow of hydrocarbons and making it crucial for the overall success of hydraulic fracturing operations (Hubbert and D. G. Willis. 1957).

**Formation Breakdown Pressure (FBP):** Formation Breakdown Pressure (FBP) is a critical parameter in hydraulic fracturing, representing the pressure required to create and propagate fractures from the wellbore into the surrounding rock formation. Understanding and accurately predicting FBP is essential for designing effective fracturing treatments.

The complexity of predicting FBP arises from its dependency on numerous aspects, including the stress concentration near the wellbore, the rock mechanical properties, the properties of the fracturing fluid and the complexity of the fracture network. Traditional models for predicting FBP, such as those developed by (Kirsch E.G. 1898) and (Hubbert and D. G. Willis., 1957), have provided foundational insights but often require simplifications that limit their accuracy in complex geological settings.

Accurate prediction of Formation Breakdown Pressure (FBP) is essential not only from an economic standpoint but also for environmental sustainability. From the economic perspective, precise FBP prediction helps optimize the volume of fluids required during hydraulic fracturing, ensuring that the process is efficient. If the breakdown pressure is overestimated, more fluid may be injected than necessary, increasing operational costs and

wasting resources. Conversely, underestimating FBP can lead to unsuccessful fracturing attempts, necessitating additional interventions, which also raises expenses.

On the environmental side, the prediction of FBP plays a crucial role in the management and disposal of fracturing fluids. Excess fluid injection can lead to fluid migration beyond the target formation, increasing the risk of contamination in surrounding aquifers and ecosystems. Poor disposal of these fluids after the fracturing process can lead to long-term environmental hazards, including groundwater contamination and induced seismicity. Thus, accurate FBP prediction helps minimize these risks by ensuring only the necessary volume of fluids is used and properly managed. This balance between efficient use of resources and environmental protection underscores the critical need for advanced, reliable FBP prediction methods (Zhang et al. 2017), (Zhao et al. 2016).

### **1.1.2 Idealized Pressure-Volume Relationship**

Illustrated by (Gaarenstroom et al. 1993), the relationship between pressure and fluid volume during minifrac is explored as illustrated in Figure 2 . Idealized Pressure-Volume Relationship in Minifrac

As fluid is pumped, pressure initially rises linearly due to fluid compressibility. When this pressure surpasses the rock's tensile strength, a new fracture initiates at the wellbore wall—this is known as Fracture Initiation Pressure (FIP). The initiation depends on stress concentration around the wellbore and the rock's tensile strength.

With continued pumping, pressure reaches its peak at Formation Breakdown Pressure (FBP), where the fracture fully opens and extends. FBP is the highest pressure point, influenced by fracture network complexity, rock properties, and fluid dynamics. Accurately estimating this value is crucial to avoid unintended formation damage or inefficient fracture design.

As the fracture extends further, pressure stabilizes at Fracture Propagation Pressure (FPP), maintaining fracture growth at a relatively steady level. This stage is influenced by frictional losses, rock permeability, and the least principal stress in the region.

Once pumping stops, pressure drops rapidly to Instantaneous Shut-In Pressure (ISIP), followed by a gradual decline until Fracture Closure Pressure (FCP) is reached, signaling fracture closure as the least principal stress overcomes fluid pressure.

Understanding FBP as the peak pressure helps optimize hydraulic fracturing treatments, ensuring safe and efficient fracture propagation without excessive pressure buildup.

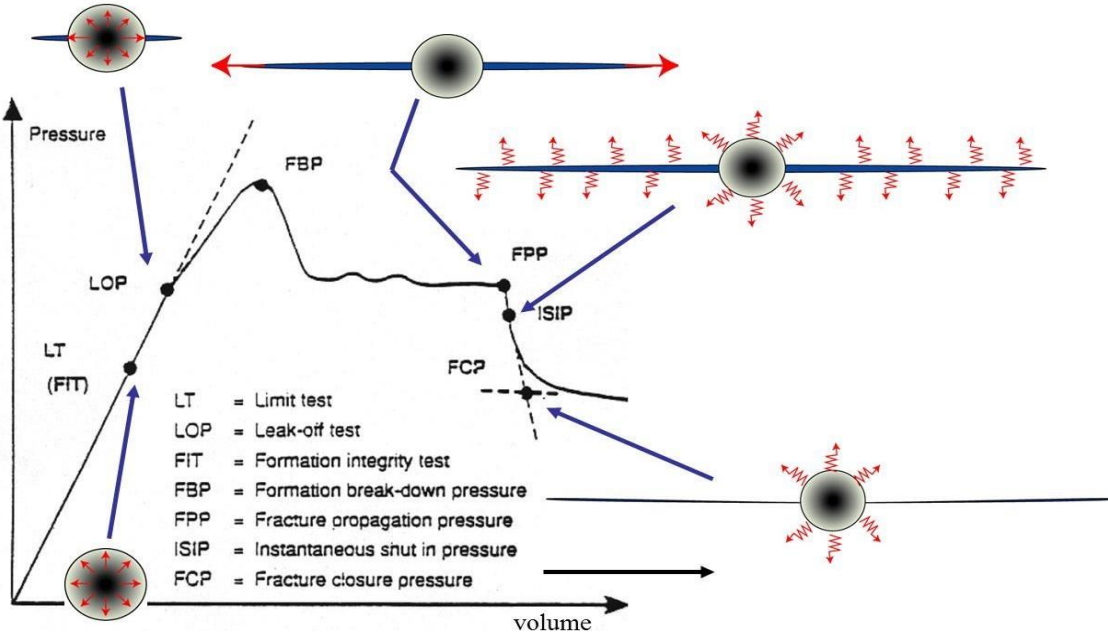


Figure 2. Schematic of Pressure and Fluid Volume Interactions during an XLOT or Minifrac (Gaarenstroom et al. 1993)

## 1.2 Literature review

### 1.2.1 Traditional Methods to Determine Breakdown Pressure

**Elastic Model by Kirsch (1898):** Originally developed by Kirsch (Zhang et al. 2019) for analyzing stress around circular openings, this method has been adapted for use in oil well drilling to calculate Formation Breakdown Pressure (FBP). The model presupposes a balanced borehole condition where the pressure exerted by the drilling fluid matches the lithostatic pressure, thereby neutralizing deviatoric stress. The formula given is:

$$FBP = 2S_{H,min} - P_o \quad (1)$$

Here, **FBP** denotes the formation breakdown pressure,  $S_{h,min}$  is the minimum horizontal stress in the far field, and  $P_o$  represents the pore pressure. Kirsch's solution is known for

its simplicity and reliance on easily accessible parameters, which facilitates accurate FBP predictions when the pore pressure and minimum horizontal stress are investigated.

**Linear Elastic Model by Hubbert and Willis (1957):** This model by (Hubbert and D. G. Willis, 1957) portrays the rock surrounding the wellbore as a homogeneous, isotropic, and linear elastic medium, impervious to the injected fluid. They describe the breakdown pressure with the equation:

$$FBP = P_{co} + 3S_{H,min} - S_{H,max} - P_o \quad (2)$$

where  $P_{co}$  is defined as the zero breakdown pressure (absence of initial pore pressure and far-field stresses), FBP denotes the breakdown pressure,  $S_{h,max}$  and  $S_{H,min}$  represent the maximum and minimum far-field horizontal stresses, and  $P_o$  stands for the pore pressure. This model asserts that  $P_{co}$  matches the rock's tensile strength (TS). In this equation, the model assumes that the maximum tensile strength criterion is met at the borehole wall, making  $P_{co}$  equivalent to the rock's tensile strength (TS). Thus, the model posits that the zero breakdown pressure is a constant material property of the rock.

**Poroelastic Model by Haimson and Fairhurst (1967):** This theory (Haimson and Fairhurst, 1967) addresses the stress alterations due to the fracturing fluid's penetration, incorporating the Biot poroelastic coefficient ( $\alpha$ ) and Poisson's ratio, indicative of the rock's compressibility and matrix properties. They defined the breakdown pressure using:

$$FBP = P_{co} + (3S_{H,min} - S_{H,max} - P_o)/A \quad (3)$$

where  $A$  is calculated as:

$$A = 2 - \alpha(1 - 2\nu)/(1 - \nu) \quad (4)$$

and  $\alpha$  is determined by:

$$\alpha = 1 - C_r/C_b \quad (5)$$

In this formula,  $C_b$  and  $C_r$  are the compressibilities of the bulk rock and rock matrix, correspondingly. The model presumes  $P_{co}$  as a basic property of the rock.

***Linear Elastic Fracture Mechanics Model by Rummels (1987):*** Rummel (1987)

introduced four models based on linear elastic fracture mechanics, focusing on the initiation and growth of fractures while considering material fracture toughness and stress intensity factors. The essential breakdown pressure formula is:

$$FBP = P_{co} + k_1 S_{H,min} + k_2 S_{H,max} \quad (6)$$

Here,  $K_{IC}$  is the fracture toughness,  $f(d,a)$  is a function related to borehole diameter  $d$  and half crack length  $a$ , and  $k_1$  and  $k_2$  are coefficients influenced by these dimensions. Rummel suggests that in the absence of visible fractures, typical values are  $k_1=3$  and  $k_2=-1$ , emphasizing that propagation occurs when the stress intensity surpasses the fracture toughness.

***Rummel 1:***

- Focused on understanding fracture initiation in homogeneous materials.
- The model provided insights into how the size and shape of initial flaws could affect breakdown pressure.

***Rummel 2:***

- Extended the model to account for the presence of natural fractures in the rock.
- Showed that pre-existing fractures could significantly lower the breakdown pressure required for hydraulic fracturing.

***Rummel 3:***

- Considered the effect of varying fracture toughness in heterogeneous materials.
- Demonstrated that zones with lower fracture toughness would fail first, influencing the fracture propagation path and pressure.

***Rummel 4:***

- Examined the impact of fluid infiltration on fracture propagation.
- Highlighted that the presence of fluid at the fracture tip could reduce the effective stress intensity factor, thereby lowering the breakdown pressure.

Figure 3 illustrates a wellbore with symmetrical pre-existing fractures aligned parallel to the maximum horizontal stress present in the rock. It is evident that the fluid pressure within the wellbore also exerts force on these fractures.

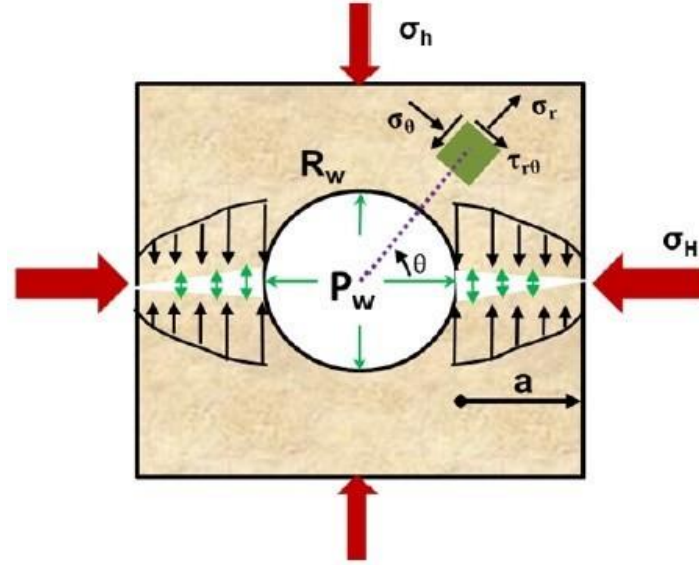


Figure 3. Wellbore Exhibiting Symmetrical Double Cracks (Zhang et al. 2018a)

**Poroelastic Model by Schmitt and Zoback (1989):** (Schmitt and Zoback, 1989) refined the poroelastic model by including additional parameters to better predict breakdown pressure in complex geological settings. Their model considers the anisotropy of the stress field and the mechanical properties of the formation. The breakdown pressure (FBP) can be calculated using the following equation:

$$FBP = (P_{co} - S_{H,max} + 3S_{H,min} - \eta P_o) / (1 + \beta - \eta) \quad (7)$$

where  $\beta$  is a porosity-dependent parameter (ranging from 0 to 1), and  $\eta$  is a material property defined by:

$$\eta = \alpha(1 - 2\nu) / (1 - \nu) \quad (8)$$

In this context,  $P_{co}$  represents the breakdown pressure when the initial pore pressure is zero, which equates to the tensile strength of the rock.  $S_{h,min}$  and  $S_{H,max}$  are the minimum and maximum horizontal far-field stresses, respectively,  $P_o$  and is the pore pressure.

The primary technical challenge with the above equation is determining the porosity-dependent parameter  $\beta$ , given the constantly and rapidly changing stress field around a pressurized wellbore.

**Point Stress Model by Ito and Hayashi (1990):** Ito and Hayashi (1990) introduced the point stress model, which considers the localized stress concentration at the tip of a pre-existing fracture or flaw. This model is particularly useful for understanding the initiation of fractures in heterogeneous formations. The formula is:

$$FBP = P_{co} + S_{H,max}(1 - 3I^2) + \frac{S_{H,min} \left(1 + \frac{2}{I^2} + 3I^2\right)}{2} \quad (9)$$

$$I = d / (d + 2d') \quad (10)$$

and  $P_{co} = TS / I^2 \quad (11)$

**Load History Model by Aadnoy and Belayneh (2008):** Aadnoy and Belayneh (2008) developed a model to improve the prediction of breakdown pressure by incorporating the influence of Poisson's ratio and anisotropic stress conditions. Unlike the classical Kirsch equation, this approach includes a Poisson scaling factor  $C$ , which adjusts the pressure prediction for more accurate results in scenarios where stress anisotropy plays a significant role. The equation for estimating breakdown pressure (FBP) is presented as:

$$FBP = S_{H,max} + C \cdot \left( \frac{3}{2} S_{H,min} - S_{H,max} - Po \right) + Po \quad (12)$$

In this expression:

- $S_{H,min}$  denote the maximum and minimum horizontal stresses, respectively,
- $Po$  is the initial pore pressure,
- $C$  represents the Poisson's scaling factor, which is calculated using the following relation:

$$C = \frac{(1 - \nu)(1 - \nu^2)}{3\nu(1 - 2\nu) + (1 + \nu)^2} \quad (13)$$

Here,  $\nu$  indicates Poisson's ratio. In cases where Poisson's ratio is zero, the scaling factor  $C$  equals one, reducing the model to the classical Kirsch solution. This adjustment enhances the accuracy of fracturing pressure predictions under various stress conditions by factoring in the rock's deformation behavior.

Furthermore, Aadnoy and Belayneh accounted for the impact of temperature changes on fracturing pressure. Since temperature fluctuations in the wellbore influence hoop stress due to thermal expansion or contraction of the rock, a correction term was introduced:

$$K \cdot EK(T - T_{init})K \quad (14)$$

Where:

- $E$  represents Young's modulus,
- $K$  denotes the linear coefficient of thermal expansion,
- $T_{init}$  is the initial temperature at in-situ conditions,
- $T$  is the current wellbore temperature.

**Complete Model by Aadnoy and Belayneh (2008):** Building on their earlier work, Aadnoy and Belayneh (Aadnoy and Belayneh 2009) presented a comprehensive model that integrates stress anisotropy, temperature effects, and borehole conditions to better predict the formation breakdown pressure. This model extends the previous load history model by adding a term that accounts for the yield strength of the filter cake and its thickness. The complete equation is given as:

$$FBP = S_{H,max} + \frac{2(1+\nu)(1-\nu^2)}{3\nu(1-2\nu)+(1+\nu)^2} \left( \frac{3}{2} S_{H,min} - S_{H,max} - Po \right) + Po + EK(T - T_{init}) + \frac{2Y}{\sqrt{3}} \ln \left( 1 + \frac{t}{r_w} \right)$$

(15)

Where:

- $Y$  refers to the yield strength of the filter cake particles,
- $t$  is the thickness of the filter cake,
- $r_w$  denotes the radius of the borehole.

This equation provides a holistic approach by combining multiple factors that influence breakdown pressure, including stress anisotropy, thermal effects, and borehole-related

properties. The inclusion of the filter cake parameters helps in predicting the pressure more accurately in cases where drilling mud forms a substantial filter cake on the borehole walls.

### **1.2.2 Application of Traditional Methods in Field and Experimental Studies**

The Kirsch model, while foundational in stress analysis around boreholes, has been applied extensively in field operations due to its simplicity and its effectiveness is particularly noted in wells with stable, homogenous likely formations where minimal tectonic activity is present (Rasouli et al., 2011). However, its assumption of uniform stress distribution limits its accuracy in complex geological settings with significant anisotropy, often leading to underestimation of breakdown pressure (BP) in such conditions (Aadnøy and Belayneh 2004).

Hubbert and Willis's linear elastic model has proven valuable in predicting BP in tight sandstone reservoirs(Weijermars 2011). Field data suggest its predictions align well with observed fracture initiation pressures under controlled drilling environments.

Nevertheless, this model struggles in heterogeneous formations where fluid invasion alters stress distributions, leading to discrepancies between predicted and actual BP (Vernik and Zoback 1992).

Rummel's linear elastic fracture mechanics models have demonstrated accuracy in predicting BP in materials with known fracture toughness (Rummel et al., 1995).

However, their performance diminishes in natural formations with irregular fracture networks, as they do not fully capture the complexities of real-world fracture geometries (Mostafavi et al. 2011).

Schmitt and Zoback's poroelastic refinement has been applied in unconventional reservoirs, providing insights into BP under varying stress regimes (Schmitt and Zoback 1992). Its strength lies in addressing anisotropic conditions, but it tends to overpredict BP in formations with high natural fracture densities, where stress redistribution significantly influences fracture initiation (Schmitt and Zoback 1993).

Aadnøy and Belayneh's models, incorporating load history and thermal effects, have shown promise in deep well scenarios, particularly in regions with significant thermal gradients (Elsworth et al., 2013). Field applications highlight their improved accuracy over traditional models, especially when accounting for thermal stresses (Aadnøy et al. 2008). However, their complexity and the need for detailed thermal and mechanical property data can limit practical implementation, especially in data-sparse environments (Hareland et al., 2009).

### **1.2.3 Factors Influencing Formation Breakdown Pressure**

Breakdown pressure (BP) is primarily influenced by confining pressure (Li et al. 2022; Lockner and Byerlee 1977; Wu et al. 2021a). Fjar et al. (2008) found that increasing confining pressure raises BP by applying additional stress to the rock, making fracture initiation more difficult. Laboratory studies (Almani et al., 2023; Zhou and He, 2015) demonstrate that high permeability formations tend to have lower BP due to greater fluid leak-off, which reduces pressure near the fracture tip. Ahmed et al. (2019) demonstrated that optimizing entry points could significantly impact fracture propagation and the

pressures required to initiate fractures. They found that wellbore size and confining pressure were critical factors affecting breakdown pressure, with larger wellbores and higher confining pressures requiring higher breakdown pressures.

Similarly, high porosity rocks exhibit lower BP (Gray et al., 1981; Ito 2008) , because they absorb more fluid before fracture initiation, reducing resistance to fracturing. Siddiqui et al. (2020) discovered that increased porosity and permeability within the rock matrix significantly lowered the breakdown pressure needed to initiate fractures, thereby facilitating easier and more efficient fracture propagation throughout the formation. Mathur et al. (2024) explored the effects of water block on permeability and breakdown pressure. Their study highlighted how the interaction between fracturing fluids and rock formations could impact productivity and the pressures needed for fracturing. They found that higher permeability in the rock matrix resulted in lower breakdown pressures, whereas the presence of water block increased the pressures required for fracture initiation.

Injection rate also plays a crucial role in BP , as higher injection rates increase fluid pressure buildup within the rock, leading to higher BP (AlTammar et al. 2020; Bungler, Lakirouhani, and Detournay 2010a). Additionally, high injection rates can create more intense stress concentrations near the fracture tip, promoting faster fracture initiation but requiring greater pressure to overcome the rock's mechanical resistance ( Miskimins et al., 2015).

Generally, higher Poisson's ratio increases ductility, which enhances a rock's ability to deform under stress, often leading to higher BP (Kong et al. 2021) . However, in some cases, increased ductility can lead to lower BP because the rock deforms more easily

under stress, dissipating energy that would otherwise contribute to fracture initiation (Bai, 2016). Huang et al. (2024) studied the microscale mechanical changes in shale due to fracturing fluid contact. Their research provided detailed insights into how fracturing fluids interact with shale at the microscale, impacting breakdown pressure and fracture propagation. They found that fluid-rock interactions could significantly alter the rock's Young's modulus and Poisson's ratio, leading to variations in breakdown pressure.

Yang et al. (2022) conducted a study to investigate the effects of Young's modulus and Poisson's ratio on breakdown pressure in tight sandstone formations. Their findings indicate that an increase in Young's modulus enhances the rock's stiffness, which in turn raises the breakdown pressure, making it more resistant to fracturing. Conversely, low Young's modulus indicates that the stress from fluid injection is absorbed more through deformation rather than fracture, leading to lower BP as the rock yields more easily under stress (Morgenstern et al., 1993).

The tensile strength of the rock is a critical factor in determining the amount of pressure needed to initiate a fracture. Rocks with higher tensile strength generally require a higher breakdown pressure (BP) because they can withstand greater stress before failing.

Research by Zhao et al. (2016) and Hubbert and D. G. Willis. (1957) indicates that tensile strength remains a key predictor of BP, with stronger rocks resisting fracture propagation more effectively.

The stress field around the wellbore significantly influences BP. Elevated stress concentrations increase the required BP because the localized stress amplifies the rock's inherent resistance to fracture initiation. Haimson and Fairhurst (1969) demonstrated that

regions with high in-situ stresses exhibit higher BP, as the differential stress near the wellbore creates conditions where greater fluid pressure is necessary to initiate and propagate fractures. This effect is more pronounced in formations with anisotropic stress fields, where stress distribution varies with orientation (Sato et al., 1997).

The geometry and connectivity of fractures play a vital role in determining BP. Complex fracture networks, particularly in heterogeneous reservoirs, can alter the pressure required for effective fracture propagation. Wang et al. (2014) found that pre-existing fractures and natural faults often reduce BP because they provide pathways of least resistance, facilitating easier fracture propagation. Conversely, in formations with fewer natural fractures, higher pressures are needed to initiate new fractures (Tan et al. 2019).

The properties of the injected fluids, such as viscosity and density, directly impact BP. Viscous fluids tend to increase frictional losses during injection, requiring higher BP to overcome the resistance to flow within the rock matrix. Fallahzadeh et al. (2016). emphasized the role of fluid viscosity in both fracture propagation and fluid-rock interactions. Higher viscosity fluids create greater pressure gradients, leading to elevated BP, while lower viscosity fluids can penetrate the formation more easily, reducing the pressure needed for fracturing(Ishida et al. 2012). Schmid et al. (2024) represented that fluid viscosity and rock matrix composition were key factors influencing breakdown pressure, with higher viscosity fluids and more calcite content requiring higher pressures.

Additionally, wellbore geometry significantly affects BP, as studies have determined that larger wellbore diameters result in lower BP, likely due to the increased surface area reducing localized stress concentration at the fracture initiation point (Jiang et al. 2025;

Wang et al. 2025). Similarly, fluid properties can influence BP; an experiment by Wang et al. (2018) showed that higher viscosity fluids increase BP by creating greater resistance to flow and reducing fluid penetration into the surrounding rock.

Zhang (2017) conducted laboratory studies on the performance of high-temperature carbonate reservoirs. His research highlighted the impact of high temperatures on rock mechanical properties and breakdown pressure. The study found that elevated temperatures significantly weakened the rock structure, reducing its strength and increasing the pressure required for fracture initiation.

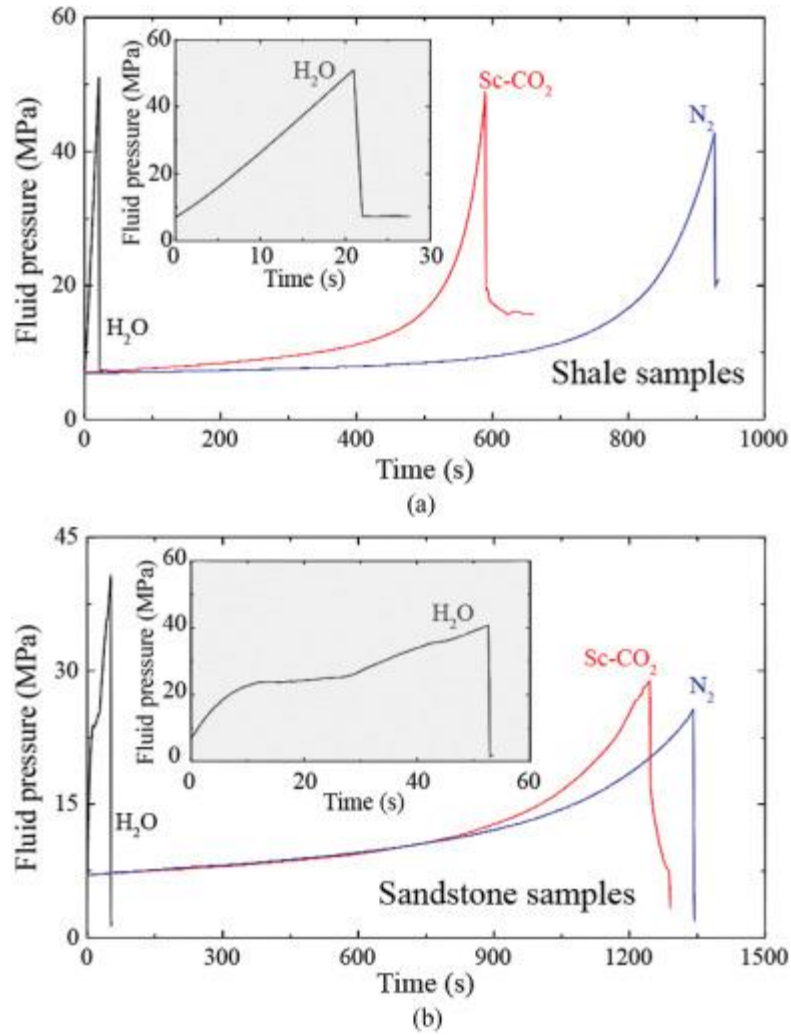
Hamdi and Boulvain (2018) performed geochemical and mineralogical characterizations of Silurian shales, emphasizing the influence of mineral composition on breakdown pressure and reservoir potential. Their study highlighted that higher mineral content and lower porosity in shales increased the breakdown pressure required for hydraulic fracturing. Furthermore, Tuzingila et al. (2024) reviewed experimental techniques used to study the effects of mineral content on the geomechanical properties of shale. Their comprehensive review provided insights into how different minerals influence rock strength and breakdown pressure. They found that rocks with higher quartz content had lower breakdown pressures due to their higher brittleness, while higher clay content increased the breakdown pressure due to greater ductility.

André et al., (2010) explored how supercritical CO<sub>2</sub> injection influences breakdown pressure in deep saline aquifers. Their findings indicated that CO<sub>2</sub> injection alters the pore structure, reducing the rock's effective stress and consequently lowering breakdown pressure. Furthermore, the presence of supercritical CO<sub>2</sub> resulted in an increase in pore

pressure, which in turn diminished the effective confining pressure and ultimately reduced the breakdown pressure needed to initiate fractures (Chen et al. 2024).

Yang et al. (2021) employed X-ray computed tomography (CT) to investigate the impact of CO<sub>2</sub> on enhanced oil recovery (EOR) in a carbonate oil reservoir. Their study revealed that CO<sub>2</sub> injection significantly improves pore network connectivity, which directly affects breakdown pressure. This process also led to mineral precipitation that sealed certain pores, requiring higher breakdown pressures for fracturing.

Wu et al. (2021) examined the effect of nitrogen injection on the breakdown pressure of shale formations. Since nitrogen is less dense than CO<sub>2</sub>, it exhibited a distinct interaction with the rock matrix. Injecting nitrogen gas into shale formations reduced pore pressure less effectively than CO<sub>2</sub>, thereby leading to higher breakdown pressure. The study emphasized the importance of selecting the most suitable gas based on specific reservoir conditions to optimize fracturing efficiency ( Figure 4).



**Figure 4. Pressure-time graph showing fracturing tests performed with various fluids and samples under identical stress conditions. (a) shale samples; (b) sandstone samples (Wu et al. 2021b)**

Lamy-Chappuis (2020) investigated how the injection of  $N_2$ -enriched brine affected the mechanical properties of calcite-bearing sandstone. Their findings indicated that the mechanical characteristics, including compressive strength and tensile strength, were significantly reduced after exposure to  $N_2$ -enriched brine. This reduction in rock strength led to lower breakdown pressures required for fracturing, making  $N_2$  a potential alternative for hydraulic fracturing in specific formations.

#### **1.2.4 Field Factors Influencing Breakdown Pressure**

Field observations provide valuable insights into the various geological and operational parameters influencing breakdown pressure (BP) in hydraulic fracturing. Unlike controlled laboratory settings, field conditions introduce additional complexities as examined by Morita et al. (1996), including stress anisotropy, wellbore conditions, formation heterogeneity, and fluid-rock interactions. A detailed understanding of these factors is crucial for optimizing fracturing strategies and improving operational efficiency.

The magnitude and orientation of in-situ stresses play a critical role in determining BP. Hou et al. (2024) conducted a field-based study demonstrating that formations with higher differential stress between the maximum and minimum horizontal stresses exhibit elevated BP values, as the increased stress contrast necessitates greater force to overcome the rock's inherent resistance. Wanniarachchi et al. (2017) highlighted that wells drilled parallel to the minimum horizontal stress direction tend to exhibit lower breakdown pressure, as this orientation reduces stress concentration around the borehole. Conversely, wells aligned with the maximum horizontal stress direction require higher breakdown pressure due to increased resistance to tensile failure.

Field studies consistently indicate that deeper formations require higher BP due to increasing lithostatic stress. Field breakdown pressure data from deep sandstone reservoirs and observed that formations located beyond 3,000 meters in depth required pressures exceeding 60 MPa for fracture initiation. This trend is attributed to the increased compaction and diagenetic strengthening of deep-seated rocks. Similarly, research on carbonate formations by Zhang et al. (2024) showed that deeper reservoirs tend to have

higher BP due to stronger intergranular bonding and mineral cementation, which increase the rock's overall mechanical strength.

The mechanical behavior of reservoir rocks, including Young's modulus, tensile strength, and Poisson's ratio, significantly impacts BP. For instance, a study by Hou et al. (2024) demonstrated that formations with higher Young's modulus exhibited increased BP due to their greater resistance to deformation under stress. Similarly, Lei et al. (2024) observed that stiffer rock formations required significantly higher pressures to initiate fractures, aligning with trends identified in geomechanical modeling studies.

The presence of natural fractures and pre-existing weaknesses further influences BP. Powlay et al. (2024) conducted field experiments showing that reservoirs with pre-existing fractures experience significantly reduced BP, as these weak planes serve as stress concentrators that facilitate fracture initiation. Similarly, Wu et al. (2023) analyzed faulted reservoirs and reported that wells drilled near active fault zones displayed highly variable BP values, as fault reactivation can either assist or hinder fracture propagation depending on stress conditions.

Borehole size and condition significantly affect BP Gan et al. (2015) . Larger boreholes generally require lower BP, as they create a greater surface area for fluid infiltration, reducing the localized stress concentration. However, irregular borehole shapes and borehole collapse can introduce unpredictable BP variations, complicating fracture initiation that has been shown by Detournay and Cheng (1988).

Berchenko and Detournay (1997) conducted experiments demonstrating that higher fluid injection rates result in increased BP, as rapid pore pressure buildup does not allow

sufficient time for stress redistribution, leading to elevated fracture initiation pressures. Conversely, field studies on cyclic injection techniques showed that alternating between pressure build-up and relaxation lowers BP by inducing fatigue damage in the rock matrix, making it more susceptible to fracturing (Jung et al. 2021).

Temperature variations within the wellbore and surrounding formation significantly influence BP by altering the stress state. Barree and Miskimins (2015b) modeled thermal effects on hydraulic fracturing and demonstrated that cooling the wellbore prior to injection reduces BP, as thermal contraction induces tensile stress, making fractures easier to initiate. Conversely, high-temperature reservoirs tend to increase BP, as rock expansion creates additional resistance to fracture propagation.

Wellbore geometry and integrity also play a crucial role. Research by Tammar et al. (2019) indicated that larger wellbores exhibit lower BP, as the increased borehole surface area reduces hoop stress concentration. However, irregular borehole shapes and borehole collapse can introduce unpredictable BP variations, emphasizing the importance of borehole stability monitoring to ensure reliable fracturing operations.

Higher pore pressures within the reservoir generally result in lower BP, as the increased fluid pressure reduces effective stress on the rock matrix. Field observations Song and Haimson (2001) confirmed that reservoirs with high initial pore pressure require lower external pressure for fracturing, facilitating easier stimulation of the formation.

Additionally, the fracture toughness of the rock is a fundamental property influencing BP. Studies on fracture mechanics Zhang et al. (2018) indicate that formations with high fracture toughness require greater pressures to propagate fractures, particularly when pre-

existing flaws or bedding planes are absent. Accurate BP prediction thus necessitates incorporating rock toughness characteristics into fracturing models.

### **1.2.5 Simulation and Machine Learning Applications for Breakdown Pressure Prediction**

Numerical simulations have been extensively used to predict breakdown pressure in hydraulic fracturing, offering a computational approach to understanding stress distribution, fracture initiation, and propagation. Among these, Finite Element Method (FEM) has been widely applied due to its ability to model complex rock mechanics and stress-strain behavior under varying conditions (Ikeda et al. 2024). FEM simulations allow for detailed analysis of stress redistribution around the wellbore, fracture propagation patterns, and the influence of mechanical rock properties on breakdown pressure. However, despite its accuracy, FEM is highly computationally demanding, limiting its feasibility for real-time or large-scale field applications. Additionally, FEM models require extensive calibration with experimental and field data to ensure accuracy, making their application resource-intensive (Fatahi et al. 2016).

Beyond FEM, Finite Difference Method (FDM) has also been utilized for breakdown pressure prediction (Groenenboom and Falk, 2000)

FDM discretizes the rock medium into a grid structure, allowing for the numerical solution of differential equations governing stress and fracture mechanics. Compared to FEM, FDM is computationally less intensive and provides faster solutions, making it suitable for real-time simulations. However, its accuracy can be limited when dealing with complex rock heterogeneity and irregular fracture geometries (Lee et al., 2000). Another widely adopted approach is the Discrete Element Method (DEM) as pointed by Azevedo and Lemos

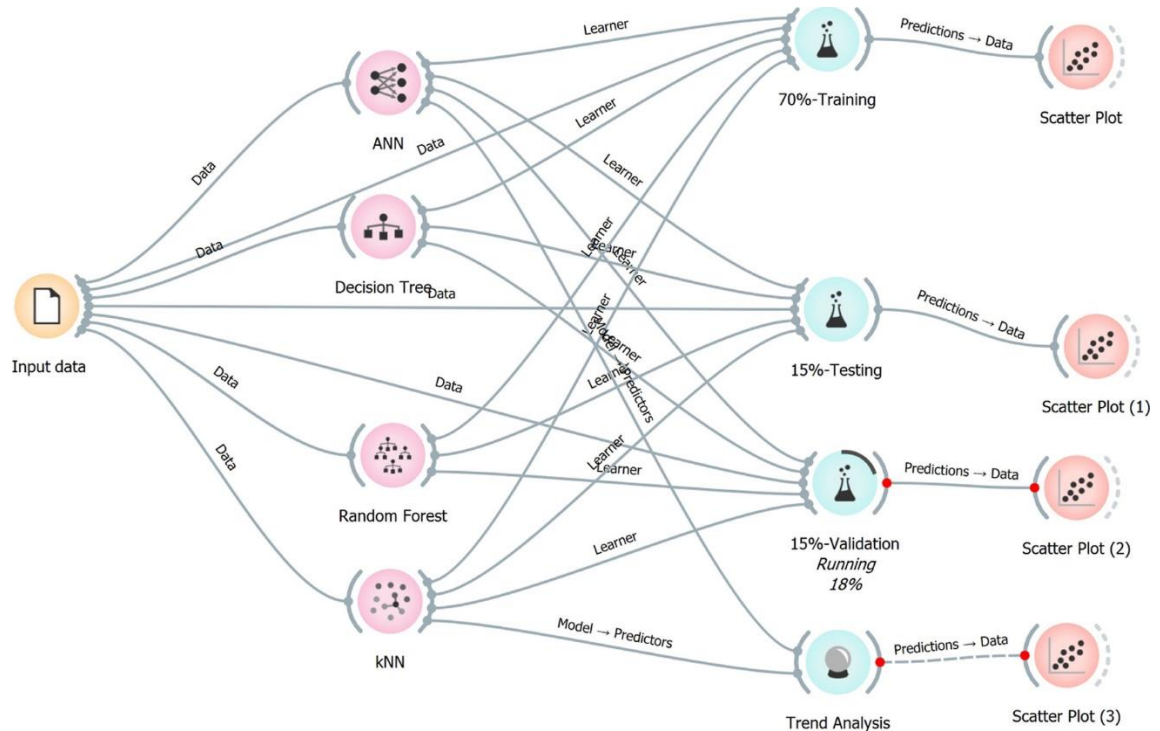
(2006), which models rock formations as an assembly of individual particles rather than a continuous medium. DEM excels in capturing rock failure mechanisms, fracture coalescence, and microcrack evolution, making it particularly useful for analyzing naturally fractured formations and unconventional reservoirs. Despite its advantages, DEM requires significant computational power and detailed input parameters regarding rock microstructure, limiting its widespread use in field-scale applications (Scholtès and Donzé, 2012).

Recent advancements in hybrid simulation techniques have combined numerical methods with data-driven approaches to improve breakdown pressure predictions. GOHFER (Grid-Oriented Hydraulic Fracture Extension Replicator), a widely used hydraulic fracturing simulator, integrates finite element modeling with field-based empirical data to enhance fracture propagation predictions (Mukhamedzianova 2017). GOHFER accounts for various operational parameters such as fluid viscosity, rock permeability, and in-situ stress distributions, providing a comprehensive framework for breakdown pressure estimation in complex reservoirs. The hybridization of empirical models with numerical simulations significantly improves prediction accuracy, but limitations remain in handling real-time field variability and non-stationary rock behavior (Zhan et al. 2016).

The application of machine learning (ML) for breakdown pressure prediction is emerging as a transformative approach, providing improved accuracy, efficiency, and adaptability to diverse reservoir conditions (Otchere et al. 2021). ML algorithms can analyze vast datasets, uncover complex non-linear relationships among geological, petrophysical, and operational parameters, and provide near-instantaneous predictions of breakdown pressure (Gong et al. 2019). Unlike conventional numerical models that rely on deterministic

equations, ML models learn patterns from historical data, making them highly adaptive to new formations and operational scenarios.

(Tariq et al. 2022) applied machine learning techniques to predict breakdown pressure in various unconventional rock types. Their study highlighted the capability of machine learning to accelerate the prediction process and improve accuracy by analyzing large datasets of rock properties and fracturing parameters. This approach represents a noteworthy advancement in the field, providing more reliable and efficient breakdown pressure predictions ( Figure 5).



**Figure 5. Schema of ML workflow applied in that study (Tariq et al. 2022)**

Machine learning models such as ANNs, SVMs, random forests, and deep learning approaches (Almani and Khan 2023) have advanced the prediction of breakdown pressure by capturing complex, non-linear relationships among geological, petrophysical, and operational variables. However, their effectiveness is limited by the availability of high-quality, extensive datasets and the difficulty in interpreting the results. Additionally, many models fail to account for all key parameters influencing breakdown pressure, such as rock brittleness, fluid viscosity, and in-situ stress conditions. While hybrid models combining numerical simulations and machine learning have improved accuracy (Yan B et al., 2022), they remain computationally intensive and are still limited by incomplete datasets and the absence of crucial variables affecting breakdown pressure.

### **1.3 Problem statement and Research Objectives**

While numerical simulations have long been used to estimate breakdown pressure in hydraulic fracturing, their accuracy remains limited due to the simplifying assumptions that fail to capture real-world geological complexity. Many of these models struggle to account for rock heterogeneity, anisotropic stress conditions, and dynamic interactions between fracturing fluids and reservoir formations. As a result, simulation-based predictions often diverge from field observations, making it necessary to reassess their effectiveness in practical applications. Machine learning offers a promising alternative, providing data-driven insights that traditional models fail to capture. However, the existing applications of machine learning in breakdown pressure prediction remain largely confined to either experimental or field-scale datasets, with limited efforts to integrate both for more robust and scalable models.

Another critical challenge lies in the lack of research on less-explored parameters that significantly influence breakdown pressure. Factors such as pressurization rate, sample size, cylindricity, and hole radius can alter fracture initiation mechanics but are often overlooked in conventional models. A more detailed investigation of these parameters is necessary to refine existing breakdown pressure prediction methodologies. Furthermore, scaling experimental results to field applications remains an ongoing issue, as laboratory conditions do not always replicate the complexities of real subsurface environments. A systematic comparison of experimental and field cases is essential to develop improved upscaling techniques that enhance the reliability of breakdown pressure estimations.

In addition to these advancements, there is an opportunity to introduce time-series forecasting as a novel approach for predicting breakdown pressure. Current models predominantly rely on static input-output relationships, failing to capture the temporal variations that influence fracturing dynamics. Implementing a time-series approach could provide a cost-effective, real-time prediction framework, improving operational decision-making in hydraulic fracturing applications.

Therefore, the primary objective of this study is to critically evaluate the effectiveness of simulation models, refine machine learning approaches for breakdown pressure prediction, assess the role of less-explored parameters, improve experimental-to-field upscaling methodologies, and develop a time-series forecasting model for real-time pressure estimation. The specific objectives of this study are as follows:

- Assess the effectiveness of simulation models in predicting breakdown pressure by examining their limitations and comparing their results with both experimental and field data.
- Develop machine learning models for predicting breakdown pressure using datasets from both laboratory and field cases while ensuring their ability to generalize across different formations.
- Examine the influence of less investigated parameters such as pressurization rate, sample size, cylindricity, and hole radius on stress distribution and fracture initiation to address existing gaps in breakdown pressure prediction.
- Compare experimental and field cases to establish better upscaling methodologies that improve the reliability of lab-derived predictions for field applications.
- Apply time-series forecasting methods to breakdown pressure prediction, allowing for a dynamic and cost-effective approach to real-time fracture monitoring.

This research has the potential to bridge the existing gaps between traditional simulation methods, machine learning approaches, and real-world hydraulic fracturing conditions. By incorporating underexplored parameters and integrating time-series forecasting into prediction models, this study aims to improve the accuracy, efficiency, and scalability of breakdown pressure estimation for diverse geological formations.

## **CHAPTER 2**

### **METHODOLOGY**

This research was conducted in four structured phases to systematically evaluate simulation models, develop machine learning-based prediction frameworks, incorporate time-series forecasting, and establish effective upscaling strategies for breakdown pressure prediction in hydraulic fracturing. The first phase focused on conducting GOHFER software simulations to assess the performance of existing numerical models across different formations. The objective was to analyze how well these simulations predict breakdown pressure under varying geological conditions, including differences in stress anisotropy, rock heterogeneity, and fluid interactions. By comparing the outputs from different formations, this phase provided insights into the accuracy and limitations of GOHFER in predicting breakdown pressure across diverse reservoir settings.

The second phase involved the building of machine learning models and time-series forecasting techniques to enhance breakdown pressure prediction. The study utilized TabNet, Kolmogorov-Arnold Networks (KAN), and ensemble-based models, trained on both experimentally collected and field datasets. These models were designed to capture non-linear relationships between key fracturing parameters, rock properties, and breakdown pressure. In parallel, time-series forecasting models, including statistical approaches and Long Short-Term Memory (LSTM) networks, were introduced to analyze breakdown pressure variations dynamically over time. The combination of machine learning and time-series forecasting aimed to create a robust, data-driven predictive

framework that integrates static and temporal dependencies for improved real-time applications.

The third phase focused on the evaluation of model performance and sensitivity analysis to determine the reliability and accuracy of the predictive frameworks. The outputs from GOHFER simulations, machine learning models, and time-series forecasts were systematically assessed against experimentally collected and field data. Sensitivity analyses were conducted on the machine learning models to evaluate how key parameters—including pressurization rate, sample size, cylindricity, and hole radius—influence breakdown pressure predictions. This phase provided a deeper understanding of the relative importance of various parameters in determining breakdown pressure, allowing for adjustments to improve the robustness of predictive models.

The final phase involved comparing experimentally collected data with field-scale observations to propose effective upscaling strategies. The differences between experimental and real-world breakdown pressure values were analyzed to develop a systematic approach for translating experimental findings into field applications. Based on the insights gained from simulation assessments, machine learning predictions, and time-series forecasting, recommendations were formulated to refine existing predictive models, improve scalability, and enhance the practical implementation of these methods in hydraulic fracturing operations. The overall workflow of this research is outlined in Figure 6.

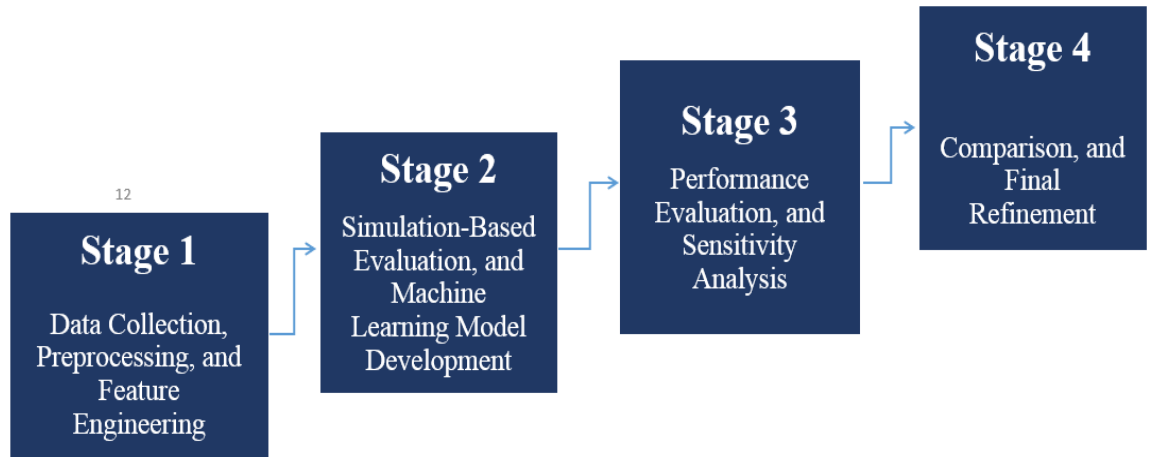


Figure 6. Schematic representation of research phases and associated experiments

## 2.1 Database Summary

### 2.1.1 Field Case Data Collection and Analysis

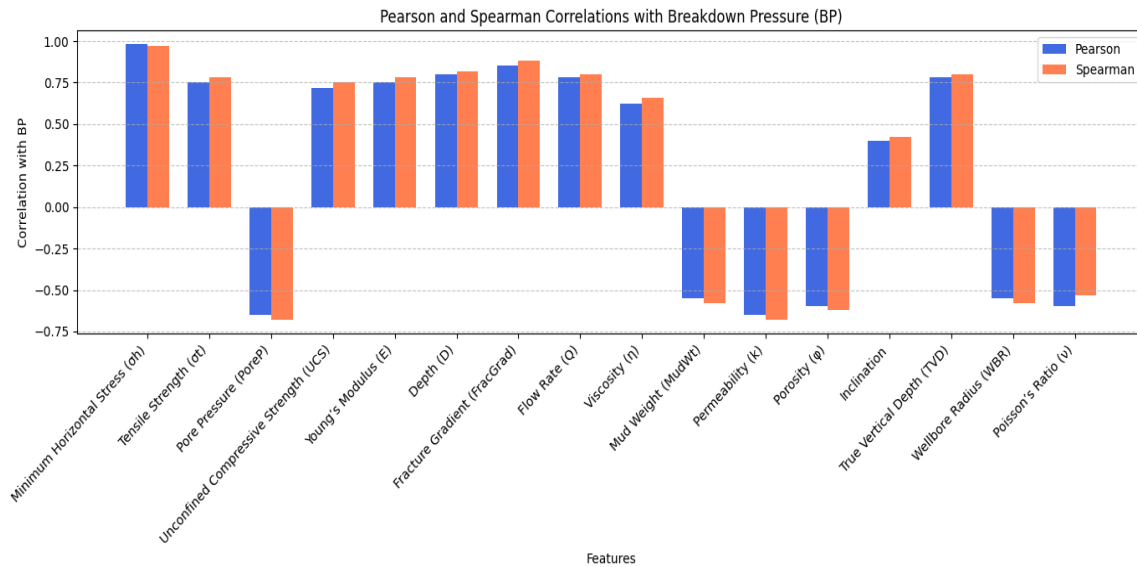
**Data collection:** A comprehensive dataset was compiled from 11 wells, predominantly located in the Eastern Hemisphere, specifically within Azerbaijan and the Caspian Sea region. The collected data also includes results from laboratory experiments conducted in the Western Hemisphere, ensuring a broad representation of geological and operational conditions. The dataset encompasses breakdown pressure measurements for water and supercritical CO<sub>2</sub> injections. The wells from which data were collected are distributed across several key fields.

Table 1 provides a summary of the origin of the wells included in the database.

**Table 1. Database Summary**

| Region      | Field   | Well Names |
|-------------|---------|------------|
| Azerbaijan  | Field A | Well 1-3   |
|             | Field B | Well 4-7   |
| Caspian Sea | Field C | Well 7-8   |
|             | Field D | Well 9-11  |

**Exploratory Data Analysis:** To assess the relationships between breakdown pressure and key geological, mechanic. Figure 7 illustrates the correlation coefficients between breakdown pressure and several critical parameters.



**Figure 7. Pearson and Spearman Correlation of Field Features**

The minimum horizontal stress and tensile strength exhibited the highest positive correlation with breakdown pressure, indicating their direct influence on fracture initiation. Conversely, parameters like pore pressure and Poisson's ratio showed a negative correlation, suggesting that increased pore pressure reduces the effective stress, thereby lowering the breakdown pressure.

Certain variables, such as permeability and porosity, displayed weak to moderate negative correlations, reinforcing the concept that more permeable formations require lower pressures for fracture propagation. Meanwhile, operational parameters like mud weight and flow rate also showed varying levels of influence, suggesting that fluid properties play a key role in determining fracture initiation pressures.

These findings highlight the importance of integrating both mechanical rock properties and operational factors when developing predictive models for breakdown pressure. The dataset provides a strong foundation for the subsequent machine learning model training and simulation validation phases, ensuring that predictive frameworks are based on well-established geological and operational patterns.

### **2.1.2 Experimental Case Data**

#### ***Data Collection from the Literature:***

In this study, the dataset for breakdown pressure (BP) was curated from a combination of published literature and data obtained from laboratory experiments. The assembled dataset encompasses unique data points from each study regarding the fracturing fluid properties, rock properties, and experimental conditions, providing a comprehensive overview of factors influencing BP predictions. It is worth clarifying that a unique data

point refers to a single instance derived from a dataset of experiments, helping to eliminate inconsistencies and enhance the reliability of predictions by ensuring a diverse representation of influencing factors. Furthermore, this method helps mitigate class imbalance by ensuring that the dataset contains a well-distributed representation of various conditions (Ghosh et al. 2024; Niaz, Shahariar, and Patwary 2022), preventing the model from being overly influenced by dominant classes and improving overall predictive performance.

The fracturing fluid properties include fluid phase, fluid viscosity ( $\mu$ ) and pH levels. These parameters are critical in defining the flow behavior and pressure dynamics during hydraulic fracturing. The rock properties collected include the type of rock samples (e.g., sandstone, shale), unconfined compressive strength (UCS), tensile strength ( $\sigma_t$ ), Poisson's ratio ( $\nu$ ), Young's modulus ( $E$ ), porosity ( $\phi$ ) and permeability ( $k$ ), and confine pressure ( $p_c$ ). Additionally, parameters such as fracture geometry ratio and cylindricity were included to capture the geometric variations and mechanical characteristics of the rock samples.

The experimental conditions comprise essential factors like the injection rate ( $Q$ ), confining pressure, and hole radius ( $r$ ). These features provide a realistic representation of the fracturing environment. Furthermore, sample size ( $s$ ) and specific properties of the borehole structure were considered to account for the variations of BP and to ensure a comprehensive dataset. This holistic dataset was utilized to develop and validate the machine learning models.

Table 2 summarizes some of the key references from which the BP data were sourced:

| Rock Types            | Injected Fluids                                 | Core Domension (mm)                                        | Fluid Injection Rates (cm <sup>3</sup> /min) | Permeability (mD) | References                           |
|-----------------------|-------------------------------------------------|------------------------------------------------------------|----------------------------------------------|-------------------|--------------------------------------|
| Scioto sandstone      | Supercritical Co <sub>2</sub> , water           | $D = 50.5$ ,<br>$H = 50.5$ ,<br>$d = 6$ ,<br>$h = 20$      | 5                                            | —                 | (Muqtadir et al. 2018)               |
| Sandstone             | Water                                           | $D = 35$ ,<br>$H = 75.1$ ,<br>$d = 4.8$ ,<br>$h = 26.4$    | 2.5–10                                       | —                 | (Farid Ibrahim and Nasr-El-Din 2018) |
| Carbonate (Limestone) | Supercritical CO <sub>2</sub> , water           | $D = 44$ ,<br>$H = 80.4$<br>$d = 10.2$ ,<br>$h = 55$       | 0.5–5                                        | 0.1–1             | (Mathur et al. 2024)                 |
| Sandstone             | Liquid CO <sub>2</sub> , water, nitrogen        | $D = 38.1$ ,<br>$H = 76.2$ ,<br>$d = 6.35$ ,<br>$h = 25.4$ | 0.1–10                                       | 100–200           | (Zou et al. 2018)                    |
| Carbonate (Dolomite)  | Supercritical CO <sub>2</sub> , water           | $D = 50$ ,<br>$H = 100$ ,<br>$d = 10$ ,<br>$h = 50$        | 0.5–5                                        | 0.01–0.1          | (Nicholson 2001)                     |
| Shale (Marcellus)     | Nitrogen, water                                 | $D = 20.8$ ,<br>$H = 55.3$ ,<br>$d = 4.18$ ,<br>$h = 21.2$ | 0.05–1                                       | 0.001–0.01        | (Marcon et al. 2017)                 |
| Sandstone             | Supercritical CO <sub>2</sub> , water, nitrogen | $D = 30.1$ ,<br>$H = 66.4$ ,<br>$d = 5.54$ ,<br>$h = 23.2$ | 0.1–10                                       | 100–200           | (Siggins 2006)                       |
| Carbonate (Limestone) | Supercritical CO <sub>2</sub> , water           | $D = 50$ ,<br>$H = 100$ ,<br>$d = 10$ ,<br>$h = 50$        | 0.5–5                                        | 0.1–1             | (Chatterji and Christensen 1979)     |

**Table 2. Summary of Dataset Gathered from Various Literature Sources on Breakdown Pressure Across Different Rock Types**

| Rock Types            | Injected Fluids                                 | Core Domension (mm)                                     | Fluid Injection Rates (cm <sup>3</sup> /min) | Permeability (mD) | References                           |
|-----------------------|-------------------------------------------------|---------------------------------------------------------|----------------------------------------------|-------------------|--------------------------------------|
| Scioto sandstone      | Supercritical Co <sub>2</sub> , water           | $D = 50.5,$<br>$H = 50.5,$<br>$d = 6,$<br>$h = 20$      | 5                                            | —                 | (Muqtadir et al. 2018)               |
| Sandstone             | Water                                           | $D = 35,$<br>$H = 75.1,$<br>$d = 4.8,$<br>$h = 26.4$    | 2.5–10                                       | —                 | (Farid Ibrahim and Nasr-El-Din 2018) |
| Carbonate (Limestone) | Supercritical CO <sub>2</sub> , water           | $D = 44,$<br>$H = 80.4$<br>$d = 10.2,$<br>$h = 55$      | 0.5–5                                        | 0.1–1             | (Mathur et al. 2024)                 |
| Sandstone             | Liquid CO <sub>2</sub> , water, nitrogen        | $D = 38.1,$<br>$H = 76.2,$<br>$d = 6.35,$<br>$h = 25.4$ | 0.1–10                                       | 100–200           | (Zou et al. 2018)                    |
| Carbonate (Dolomite)  | Supercritical CO <sub>2</sub> , water           | $D = 50,$<br>$H = 100,$<br>$d = 10,$<br>$h = 50$        | 0.5–5                                        | 0.01–0.1          | (Nicholson 2001)                     |
| Shale (Marcellus)     | Nitrogen, water                                 | $D = 20.8,$<br>$H = 55.3,$<br>$d = 4.18,$<br>$h = 21.2$ | 0.05–1                                       | 0.001–0.01        | (Marcon et al. 2017)                 |
| Sandstone             | Supercritical CO <sub>2</sub> , water, nitrogen | $D = 30.1,$<br>$H = 66.4,$<br>$d = 5.54,$<br>$h = 23.2$ | 0.1–10                                       | 100–200           | (Siggins 2006)                       |
| Carbonate (Limestone) | Supercritical CO <sub>2</sub> , water           | $D = 50,$<br>$H = 100,$<br>$d = 10,$<br>$h = 50$        | 0.5–5                                        | 0.1–1             | (Chatterji and Christensen 1979)     |

Where the diameter of the rock specimen is denoted as  $D$ , while the height of the rock sample is represented as  $H$ . The diameter of the borehole is indicated by  $d$ , and the length of the borehole is represented as  $h$ .

**Data collection for Pressurization Rate:** In hydraulic fracturing, accurately determining the pressurization rate is essential because it directly influences the breakdown pressure (BP) and fracture initiation dynamics (Haimson and Fairhurst 1969). A recent study (Bunger et al., 2010) has highlighted that relying on flow rate alone can lead to inaccurate predictions of BP, especially when dealing with compressible fluids. This oversight stems from the fact that flow rate does not account for how quickly pressure builds up in the wellbore (Song et al. 2001a), particularly in cases where fluid compressibility significantly affects pressure propagation. Therefore, measuring and analyzing the pressurization rate, which describes the rate of pressure increase in the wellbore prior to breakdown, provides a more accurate representation of the conditions leading to fracture initiation.

In this study, we separately collected 32 data points of pressurization rates immediately preceding the breakdown pressure during hydraulic fracturing experiments. Analyzing pressurization rates prior to breakdown is crucial as it helps assess the relationship between pressure increase and rock fracturing. The cited study by Wu et al. (2020) demonstrated that higher pressurization rates reduce the distance from the fracture initiation point to the borehole wall—approximately 34 mm at 0.01 MPa/s and 1.74 mm at 5 MPa/s—resulting in increased stress concentration. This highlights the need to

control pressurization rates to ensure safe and effective fracture propagation. Table 3 summarizes some of the collected data points, showing variations in pressurization rates leading up to the breakdown pressure.

**Table 3. Dataset Gathered from Various Literature Sources on Pressurization Rate**

| <b>Pressurization Rate (MPa/s)</b> | <b>Reference</b>              |
|------------------------------------|-------------------------------|
| 0.01                               | Garagash and Detournay (1997) |
| 0.1                                | AlTammar and Sharma (2019b)   |
| 0.5                                | Song et al. (2001b)           |
| 1                                  | Ha and Yun (2020)             |
| 2                                  | Song and Haimson (2001)       |

**Exploratory Data Analysis:** The collected dataset for this study incorporates key parameters such as injection rate ( $Q$ ), fluid viscosity ( $\eta$ ), confining pressure ( $p_c$ ), rock tensile strength ( $\sigma_t$ ), unconfined compressive strength (UCS), Young's modulus ( $E$ ), Poisson's ratio ( $\nu$ ), permeability ( $k$ ), porosity ( $\phi$ ), sample size ( $s$ ), hole radius ( $r$ ), fracture geometry ratio (FGR), and pressurization rate (PR). The feature Cylindricity was excluded from the statistical analysis due to its categorical nature, as it represents a binary condition and does not contribute numerically to the model inputs. The statistical measures include mean, standard deviation (SD), minimum, 25th percentile (25%), median (50%), 75th percentile (75%), and maximum values, which provide insights into the distribution of the input features used for model training as shown in Table 4.

**Table 4. Statistical Analysis of the Total Dataset**

| <b>Parameter</b> | <b>Mean</b> | <b>SD</b> | <b>Min</b> | <b>25%</b> | <b>50%</b> | <b>75%</b> | <b>Max</b> |
|------------------|-------------|-----------|------------|------------|------------|------------|------------|
| $Q$ , mL/min     | 6.2         | 4.8       | 0.1        | 4          | 6          | 8          | 30         |
| $\eta$ , cP      | 1.1         | 215       | 0.02       | 1          | 1.1        | 1.2        | 1.5        |
| $p_c$ , psi      | 1300        | 2100      | 15         | 0          | 600        | 1600       | 8500       |
| $\sigma_t$ , MPa | 7.2         | 2.5       | 3.5        | 5.6        | 7          | 8.2        | 14.5       |
| UCS, MPa         | 120         | 65        | 28         | 55         | 85         | 170        | 220        |
| $E$ , GPa        | 29.5        | 16.2      | 12.5       | 20         | 27         | 47         | 59         |
| $\nu$            | 0.26        | 0.03      | 0.195      | 0.235      | 0.265      | 0.275      | 0.305      |
| $k$ , mD         | 1.1         | 28.31     | 0.0001     | 0.015      | 0.5        | 4.5        | 50.25      |
| $\phi$ , %       | 7.2         | 6.1       | 0.6        | 1.2        | 7          | 13         | 19         |
| $s$ , mm         | 25          | 5         | 15         | 20         | 25         | 30         | 40         |
| $R$ , mm         | 5.5         | 2         | 2.5        | 4          | 5.5        | 7          | 10         |
| FGR              | 0.8         | 0.15      | 0.4        | 0.7        | 0.8        | 0.9        | 1.2        |
| PR, MPa/s        | 1           | 1.5       | 0.01       | 0.1        | 0.5        | 2          | 5          |
| BP, psi          | 2900        | 2600      | 120        | 1100       | 1800       | 4200       | 15800      |

Figure 8 illustrates the box plots for the non-categorical variables in the dataset, providing a concise summary of their statistical distributions. Each box plot highlights key metrics, including the median, interquartile range (IQR), and potential outliers. The central box represents the IQR, which captures the range between the 25th (Q1) and 75th (Q3) percentiles, while the whiskers extend to show the variability outside this range. Any points beyond the whiskers indicate outliers, emphasizing deviations from the overall distribution. Box plots are particularly useful for comparing variables side-by-side, as they provide insights into the symmetry, spread, and presence of skewness in the data (Mishra and Datta-Gupta, 2017). For instance, wider IQRs suggest greater variability, whereas shorter whiskers imply concentrated values.

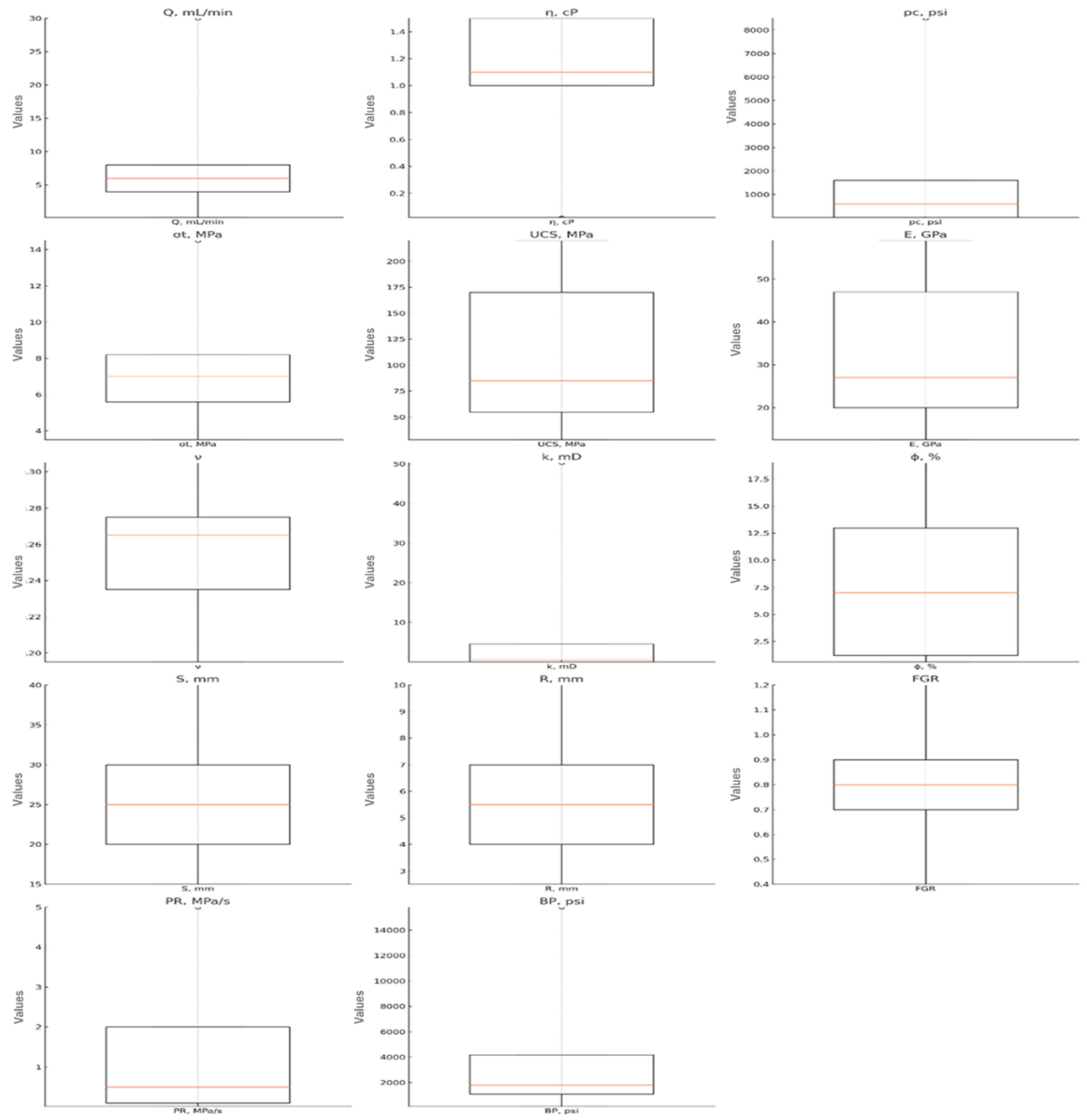


Figure 8. The Box-Plot Distribution of Entire Dataset

### 2.1.3 Time-Series Data for Breakdown Pressure Forecasting

**Data Acquisition:** This study undertakes a detailed analysis of well-log data collected from a total of 10 wells, with 9 wells designated for training purposes and 1 well reserved for forecasting and validation. The dataset comprises approximately 158 data points, carefully selected based on depth and the availability of breakdown pressure and associated parameters. A total of 16 parameters, including well-log data, mechanical properties, and breakdown pressure values, were incorporated into the research to ensure a broad comprehension of the factors influencing breakdown pressure.

**Data Preprocessing:** Preprocessing played a critical role in ensuring data consistency and reliability, essential for accurate modeling and forecasting. Steps were taken to address any missing or inconsistent values within the dataset, including the use of mean imputation to fill gaps in continuous data, and nearest neighbor technique to estimate missing values based on surrounding data points. The breakdown pressure was observed to exhibit a distribution consistent with field conditions, facilitating its use in predictive modeling. This step ensured that the dataset was suitable for exploring the relationships between breakdown pressure and the selected input parameters.

#### ***Visualization and Parameter Selection:***

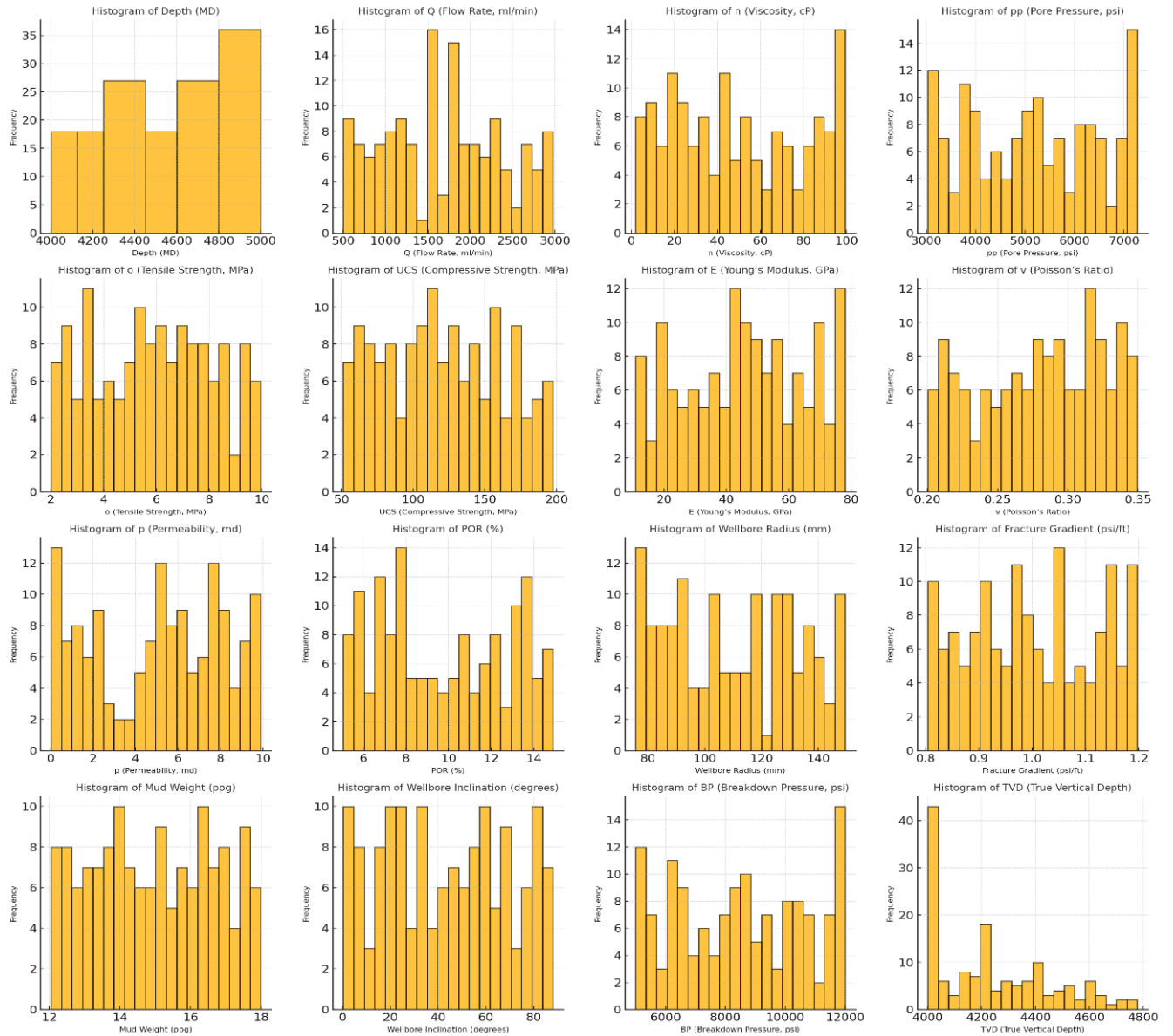
Visual exploration was employed to identify key parameters influencing breakdown pressure. Histogram plots were used to visualize the distribution between input parameters, revealing significant correlations with certain mechanical properties, depth, and pore pressure, as illustrated in Figure 9.

To capture a broad understanding of these relationships, all 16 parameters were initially included in the analysis. Although reducing the dimensionality of the model is important for computational efficiency, this initial comprehensive approach provided insights into parameter interactions and their contributions to the predictive accuracy of the model. This allowed for a strategic reduction of parameters to optimize performance without compromising accuracy.

Correlation coefficient analysis, presented in Figure 10 and Figure 11, further highlighted these relationships. Heatmaps generated using Python's seaborn library depicted both Pearson's and Spearman's correlation coefficients. Figure 10 showcased Pearson's correlation coefficients, emphasizing linear relationships, while Figure 11 highlighted Spearman's correlation coefficients to account for nonlinearities and the effects of outliers.

***Insights from Correlation Analysis:*** The combined analysis of Pearson's and Spearman's correlation coefficients provided deeper insights(into the relationships between input parameters and breakdown pressure (De Winter et al., 2016). Pearson's correlation coefficients revealed the strength and direction of linear relationships, where high positive values indicated strong positive relationships and high negative values reflected strong inverse relationships. Spearman's correlation coefficients, on the other hand, captured monotonic relationships, making them robust to outliers and nonlinear interactions.

In this study, the distinction between Pearson's and Spearman's methods was vital. Certain relationships displayed increased strength in the Spearman analysis due to its ability to capture nonlinear dependencies. This dual approach ensured a more comprehensive understanding of parameter interactions and their relevance to the model.



**Figure 9. Histogram of 16 parameters, illustrating the diverse distribution types for each feature**

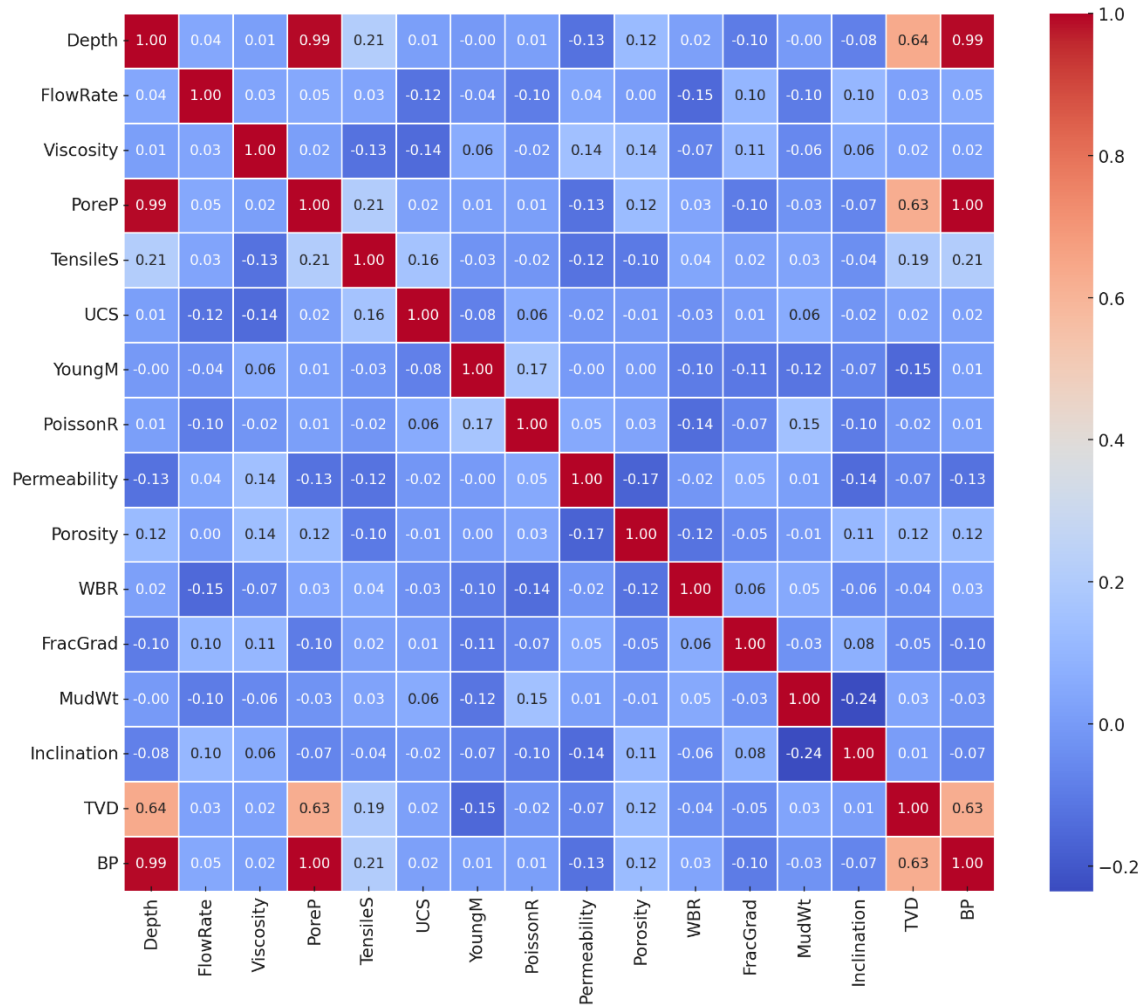


Figure 10. Pearson Correlation Coefficients Plot

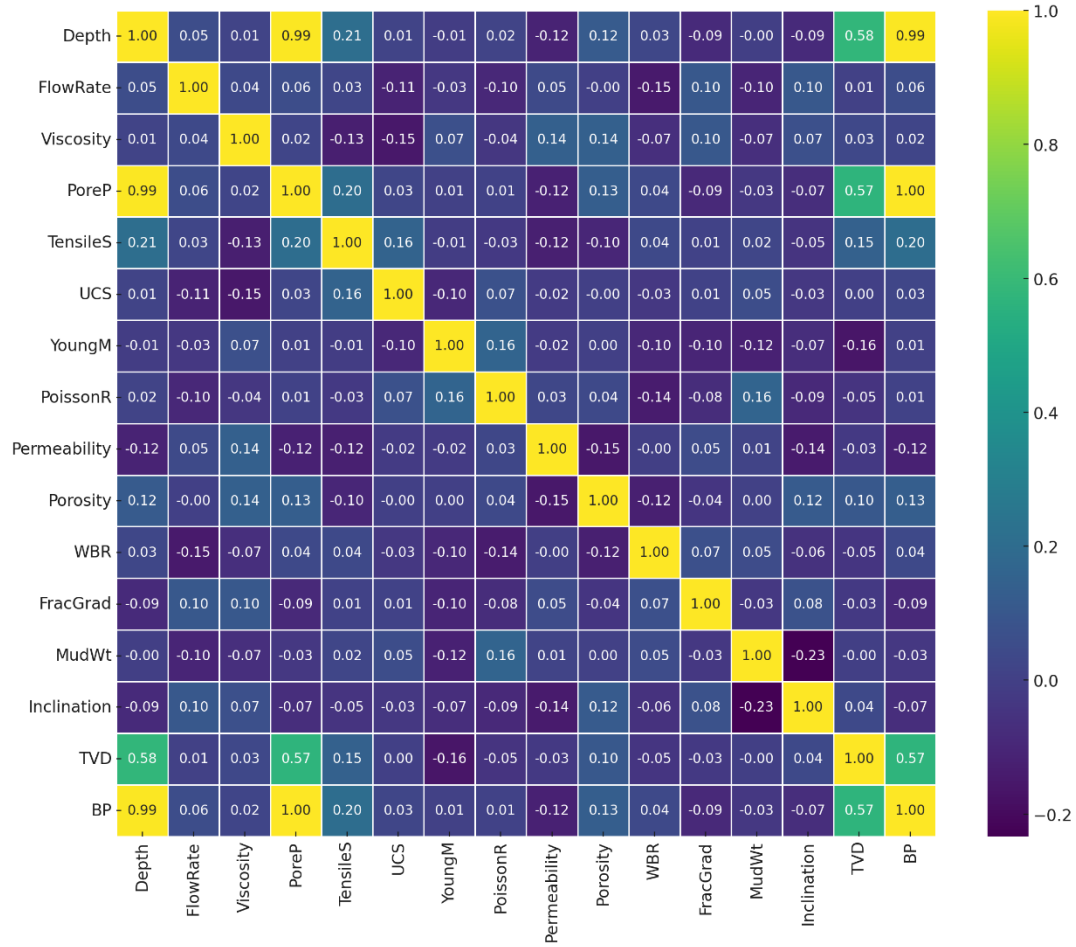


Figure 11. Spearman Correlation Coefficient Plot

## 2.2 Simulation and Machine Learning Model Development

### 2.2.1 Numerical Simulation for Breakdown Pressure Estimation

To evaluate the predictive capabilities of existing numerical models, GOHFER software was utilized to simulate breakdown pressure across different geological formations. These simulations were conducted using field and experimental datasets, incorporating key parameters such as in-situ stresses, tensile strength, permeability, porosity, fluid properties, and wellbore geometry. The goal was to assess how well GOHFER can predict breakdown pressure under varying reservoir conditions and to identify its limitations in capturing

complex fracture initiation mechanisms. The simulation techniques utilized in this stage are following:

- Kirsch
- Haimson and Fairhurst
- Hubbert and Willis
- Rummel 1
- Rummel 2
- Rummel 3
- Rummel 4
- Ito and Hayashi
- Schmitt and Zoback
- Aadnoy and Belayneh 1
- Aadnoy and Belayneh 2

These models vary in their underlying assumptions, ranging from elastic and poroelastic formulations to fracture mechanics-based approaches. Table 5 provides a summary of the key models, detailing their authors, year of publication, methodology, description, and fundamental assumptions.

**Table 5. Summary of Applied Simulation Models**

| Author(s)             | Year | Method                            | Description                                                                                                                | Assumptions                                                                                              |
|-----------------------|------|-----------------------------------|----------------------------------------------------------------------------------------------------------------------------|----------------------------------------------------------------------------------------------------------|
| Kirsch                | 1898 | Elastic model                     | One of the earliest models for stress distribution around a circular borehole, commonly used in fracture simulations.      | Assumes a circular borehole with a balanced drilling fluid pressure equal to lithostatic stress.         |
| Hubbert and Willis    | 1957 | Linear elastic model              | An extension of Kirsch's model, incorporating additional stress factors to refine FBP predictions.                         | Assumes a homogeneous, isotropic, and impermeable rock, with $P_{co}$ as a constant mechanical property. |
| Haimson and Fairhurst | 1967 | Poroelastic model                 | Introduces the effects of fluid penetration into the rock matrix to refine stress estimations.                             | Assumes $P_{co}$ remains a constant mechanical property of the rock.                                     |
| Rummel 1              | 1987 | Linear elastic fracture mechanics | Considers the influence of pre-existing symmetrical cracks, ignoring initial pore pressure.                                | Assumes the pressure within the crack equals the pressure in the wellbore.                               |
| Rummel 2              | 1987 | Linear elastic fracture mechanics | Similar to Rummel 1 but assumes a different pressure distribution within the crack.                                        | Crack pressure is assumed to be 75% of the wellbore pressure.                                            |
| Rummel 3              | 1987 | Linear elastic fracture mechanics | Considers a linear pressure drop along a pre-existing symmetrical crack.                                                   | Ignores initial pore pressure effects.                                                                   |
| Rummel 4              | 1987 | Linear elastic fracture mechanics | Extends Rummel 3 by incorporating a quadratic pressure drop along the crack.                                               | Assumes a varying pressure distribution within the crack.                                                |
| Schmitt and Zoback    | 1989 | Poroelastic model                 | Accounts for the influence of fluid injection and includes porosity-dependent corrections.                                 | Correction parameter is porosity-dependent and requires laboratory validation.                           |
| Ito and Hayashi       | 1990 | Point stress model                | Proposes that a finite region of material, rather than a single point, must reach critical stress for fracture initiation. | Neglects initial pore pressure; $P_{co}$ depends on material size.                                       |
| Aadnoy and Belayneh 1 | 2008 | Thermoporoelastic model           | Incorporates the historical load variations experienced by the wellbore.                                                   | Includes Poisson's effect, considering tangential and axial stress responses.                            |
| Aadnoy and Belayneh 2 | 2008 | Thermoporoelastic model           | Extends the previous model by integrating temperature effects along with load history.                                     | Accounts for Poisson's effect under both thermal and mechanical loading conditions.                      |

These models serve as the foundation for numerical simulations conducted in GOHFER software, providing a comparative framework for evaluating breakdown pressure predictions across different formations. By testing multiple models under varying geological and operational conditions, this study aims to determine the most reliable methodologies for predicting breakdown pressure and identify the limitations of existing approaches.

### 2.2.2 Machine Learning Models for Field and Experimental Case

**LightGBM:** Light Gradient Boosting Machine (LightGBM) is a highly efficient and scalable gradient boosting framework designed to tackle large-scale data with complex feature interactions. Introduced by (Ke et al. 2017), LightGBM employs decision trees as base learners, training them sequentially to minimize residual errors from previous models. Unlike traditional gradient boosting frameworks, which grow trees level-wise, LightGBM utilizes a leaf-wise growth strategy. This method splits the leaf with the highest loss reduction, leading to more optimized trees. Such an approach reduces computational complexity and improves model accuracy, especially when working with high-dimensional datasets typical in geomechanical contexts such as hydraulic fracturing.

The LightGBM framework optimizes the following objective function for regression tasks given by eq 16:

$$L(\theta) = \sum_{i=1}^n l(y_i, f(x_i; \theta)) + \Omega(f) \quad (16)$$

where  $L(\theta)$  represents the total loss,  $l(y_i, f(x_i; \theta))$  is the loss function, typically the mean squared error (MSE), and  $\Omega(f)$  is a regularization term that penalizes model complexity. This structure encourages the development of models that balance accuracy with generalization, thereby reducing overfitting.

In the context of hydraulic fracturing, where predicting breakdown pressure (BP) requires the integration of numerous geological and operational parameters—such as confining pressure, rock properties, permeability, and fluid injection rates—LightGBM offers several advantages. The histogram-based technique employed by LightGBM discretizes continuous features into bins, reducing both memory consumption and training time. This is particularly beneficial for handling large datasets efficiently, as seen in geomechanical modeling, where the number of features and the volume of data are substantial. The ability to handle both categorical and continuous variables within the same model makes it an ideal choice for geologically diverse datasets (Anghel et al. 2018).

In this study, LightGBM was selected over Random Forest primarily due to its computational efficiency and superior handling of large datasets. Random Forest, while robust and easy to implement, tends to suffer from slower training times and higher memory consumption when dealing with large-scale data (Choudhury et al., 2024).

**CatBoost:** CatBoost is a powerful gradient boosting algorithm known for its ability to efficiently process categorical features without requiring extensive preprocessing.

Developed by Yandex, CatBoost utilizes ordered boosting and target-based encoding techniques specifically designed to mitigate overfitting when handling categorical data (Prokhorenkova et al. 2018). The algorithm operates by constructing an ensemble of

decision trees, where each tree is trained to minimize the residual errors of its predecessor. The primary objective function for regression tasks in CatBoost is defined as follows (Equation 17):

$$L(\theta) = \sum_{i=1}^n l(y_i, f(x_i; \theta)) + \lambda \sum_{j=1}^T ||w_j||^2 \quad (17)$$

where  $L(\theta)$  is the total loss,  $l(y_i, f(x_i; \theta))$  represents the loss function (usually mean squared error),  $\lambda$  is a regularization parameter to avoid overfitting,  $T$  represents the number of trees,  $w_j$  and are the weights for the  $j$ -th tree. This architecture reduces the risk of overfitting by employing ordered boosting, a technique that prevents data leakage by ensuring that only past observations are used to predict the current one.

Figure 12 provides an illustration of CatBoost's decision tree structure, showing how each tree in the sequence is built to correct the errors of the previous one while processing categorical features natively. This figure demonstrates how CatBoost handles data in a way that is distinct from other gradient boosting algorithms.

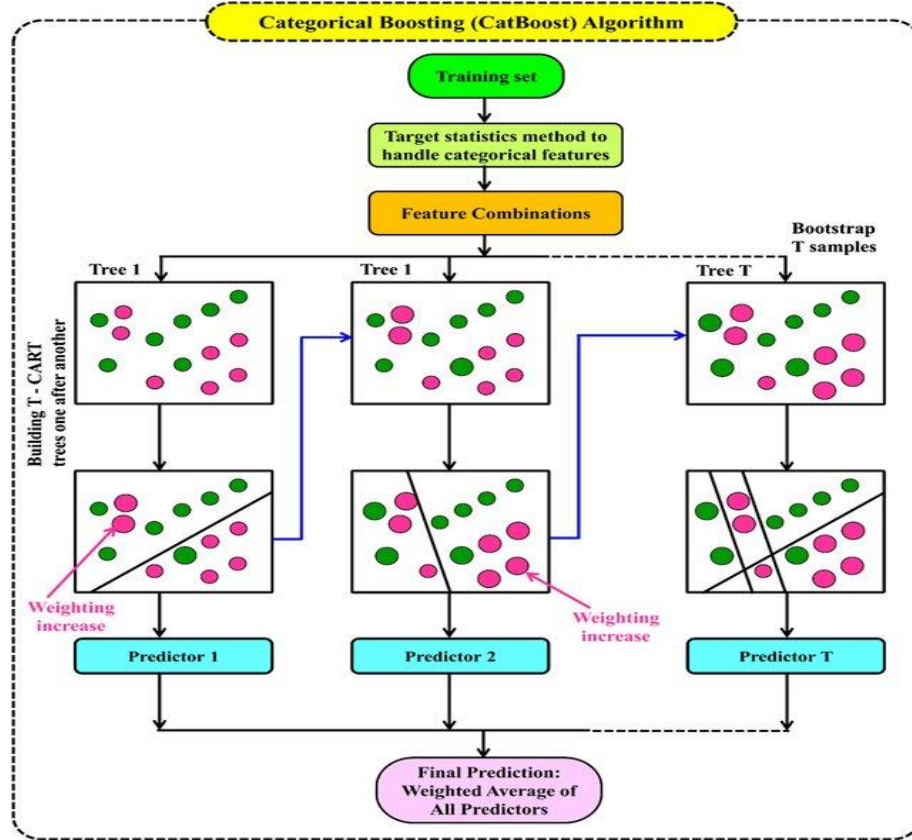


Figure 12. Overall Structure of Catboost Model (Pandey et al. 2023)

**XGBoost (Extreme Gradient Boosting):** Developed by Chen and Guestrin (Chen and Guestrin, 2016), XGBoost is a highly regarded gradient boosting framework that sequentially builds decision trees, with each tree correcting the residual inaccuracies of the preceding trees. XGBoost incorporates various enhancements such as regularization, efficient handling of missing data, and the ability to parallelize tree construction, making it highly efficient for large-scale and high-dimensional datasets.

The objective function for XGBoost in regression is given by:

$$L(\theta) = \sum_{i=1}^n l(y_i, \hat{y}_i) + \sum_{j=1}^T \Omega(f_j) \quad (17)$$

where:

- $L(\theta)$  is the total loss,
- $l(y_i, \hat{y}_i)$  is the loss function (e.g., mean squared error),
- $\Omega(f_j)$  is the regularization term for each of the  $T$  trees to control overfitting,
- $f_j$  represents the individual trees in the ensemble.

XGBoost is particularly known for its ability to reduce overfitting by applying  $L_1$  (Lasso) and  $L_2$  (Ridge) regularization techniques to control the complexity of the model.

Additionally, XGBoost optimizes both gradient boosting and tree pruning techniques by using a greedy algorithm that selects the best split points based on maximum loss reduction. Furthermore, one key advantage of XGBoost is its ability to parallelize tree construction, which significantly improves computational efficiency, particularly on large datasets. Its ability to optimize across multiple cores makes it one of the fastest gradient boosting implementations.

**TabNet:** TabNet is a deep learning framework specifically designed for processing tabular data, introduced by (Arik and Pfister 2021). Unlike conventional machine learning models such as gradient boosting, which rely on sequential learning through decision trees, TabNet utilizes a combination of attention mechanisms and sequential decision-making steps to handle tabular data more adaptively and interpretably.

A key innovation of TabNet is its use of self-supervised learning along with soft feature selection. Through an attention mechanism, the model dynamically determines which features to prioritize at each decision stage.

TabNet's objective function can be expressed as eq 18:

$$L(\theta) = \sum_{i=1}^n l(y_i, \hat{y}_i) + \sum_{t=1}^T \mathbb{E}(\text{Mask}_t) \quad (18)$$

Where:

- $L$  is the total loss,
- $l(y_i, \hat{y}_i)$  is the loss function (typically cross-entropy for classification and MSE for regression),
- $\text{Mask}_t$  represents the feature mask at decision step  $t$ , which controls which features are attended to during learning,
- $\lambda$  is a regularization term to enforce sparsity.

TabNet processes data in a sequential manner, where each decision step uses a soft attention mechanism to create a feature mask that determines which features will be selected and passed to the next layer. This attention-based mechanism makes the model more interpretable, as it highlights which features were critical to making each decision. TabNet also includes a sparse regularization term in its objective function, which penalizes the model for attending to too many features, encouraging it to focus only on the most relevant ones.

Figure 13 demonstrates the architecture of TabNet, showcasing the sequential decision steps and the soft feature selection mechanism that allows the model to adaptively attend to the most important features.



$$L(\theta) = \sum_{i=1}^n l(y_i, \hat{y}_i) + \lambda \cdot \Omega(\Phi, \psi) \quad (19)$$

Where:

- $L(\theta)$  is the total loss,
- $l(y_i, \hat{y}_i)$  is the loss function (commonly mean squared error for regression tasks),
- $\lambda$  is a regularization coefficient,
- $\Omega(\Phi, \psi)$  is a regularization term that enforces smoothness or sparsity on the univariate functions  $\Phi_q$  (outer functions) and  $\psi_{pq}$  (inner functions).

K-Nets decompose the input data into univariate inner functions  $\psi_{pq}(x_p)$  that are aggregated and transformed by outer functions  $\Phi_q$ , resulting in an interpretable and flexible representation of complex multivariate functions. This structured approach, derived from theoretical principles(Hou and Zhang, 2024), enhances generalization and reduces overfitting risks. The inclusion of regularization ensures that the learned functions are smooth and stable, further improving their performance in practical applications.

### 2.2.3 Time-Series Models for Breakdown Pressure Forecasting

Time-series models might be highly applicable for breakdown pressure forecasting due to their ability to capture temporal dependencies and sequential patterns in field data. Since breakdown pressure typically increases with depth due to rising overburden and in-situ stresses, these models can effectively be deployed to forecast breakdown pressure as a function of depth. Additionally, these models offer cost-effectiveness by reducing the

need for extensive physical testing and simulations, enabling efficient data-driven predictions that save both time and resources in operational planning

**Long Short-Term Memory (LSTM) Networks:** Long Short-Term Memory (LSTM) networks are a specialized type of recurrent neural network (RNN) developed to overcome the limitations of traditional RNNs, particularly their difficulty in retaining information across long sequences due to the vanishing gradient problem.

LSTMs were initially proposed by Hochreiter (1997) and have since been widely adopted across multiple domains, including time-series forecasting, natural language processing, and speech recognition. Their architecture includes memory cells and gating mechanisms that allow selective information retention and updating, making them highly suitable for sequential data modeling (Huang et al. 2022).

An LSTM cell consists of three main gates:

1. **Forget Gate:** Determines which information from the previous cell state should be discarded.
2. **Input Gate:** Decides which new information is to be added to the cell state.
3. **Output Gate:** Filters the output based on the current cell state.

The mathematical formulation of an LSTM cell is as follows:

$$f_t = \sigma(W_f \cdot [h_{t-1}, x_t] + b_f) \quad (20)$$

$$i_t = \sigma(W_i \cdot [h_{t-1}, x_t] + b_i) \quad (21)$$

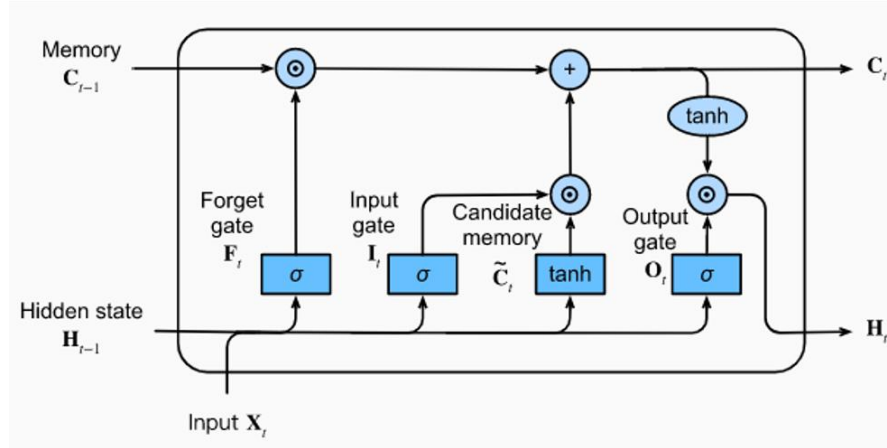
$$c_t = \tanh(W_c \cdot [h_{t-1}, x_t] + b_c) \quad (22)$$

$$C_t = f_t \cdot C_{t-1} + i_t \cdot \zeta_t \quad (23)$$

$$o_t = \sigma(W_o \cdot [h_{t-1}, x_t] + b_o) \quad (24)$$

$$h_t = o_t \cdot \tanh(C_t) \quad (25)$$

where  $f_t, i_t, \zeta_t, C_t, o_t, h_t$  serve as process functions (matrix), while  $W_f, W_i, W_c$  and  $W_o$  are the associated weights, and  $b_f, b_i, b_c$ , and  $b_o$  are biases. Equation (1) is associated with the stage of selecting relevant information. As data moves from one moment to the next, not all of it proves necessary and irrelevant data is eliminated by the forget gate, which is the initial gate. Equations (2) through (4) outline the stages of processing this information. Within the unit's state, valuable information is continually updated through the second gate (the input gate), until it meets a specific criterion. Finally, Equations (5) and (6) cover the stages of outputting information, where the refined data is released following proper assessment by the third gate (the output gate). The Figure 14 illustrated overall architecture of aforementioned equations:



**Figure 14.** A schematic representation of an LSTM cell showing the interaction between gates and the cell state (Yu et al. 2019)

To compute predictions with an LSTM, the model updates its cell and hidden states iteratively using the above equations for each time step in the input sequence. The final hidden state can be passed through a fully connected layer to obtain the forecasted output values.

LSTMs have been extensively applied in forecasting tasks that involve non-stationary data. For instance, they have demonstrated superior performance in energy load forecasting. Their ability to capture non-linear temporal patterns makes them particularly valuable in scenarios like breakdown pressure forecasting, where complex, evolving dependencies exist.

***Bidirectional Long Short-Term Memory (BiLSTM) Networks:*** The Bidirectional LSTM (Bi-LSTM) enhances the capabilities of traditional LSTM by processing input sequences in both forward and backward directions. This bidirectional mechanism enables the model to leverage information from both preceding and succeeding time steps, allowing for a more holistic understanding of the temporal dependencies within the data sequence.

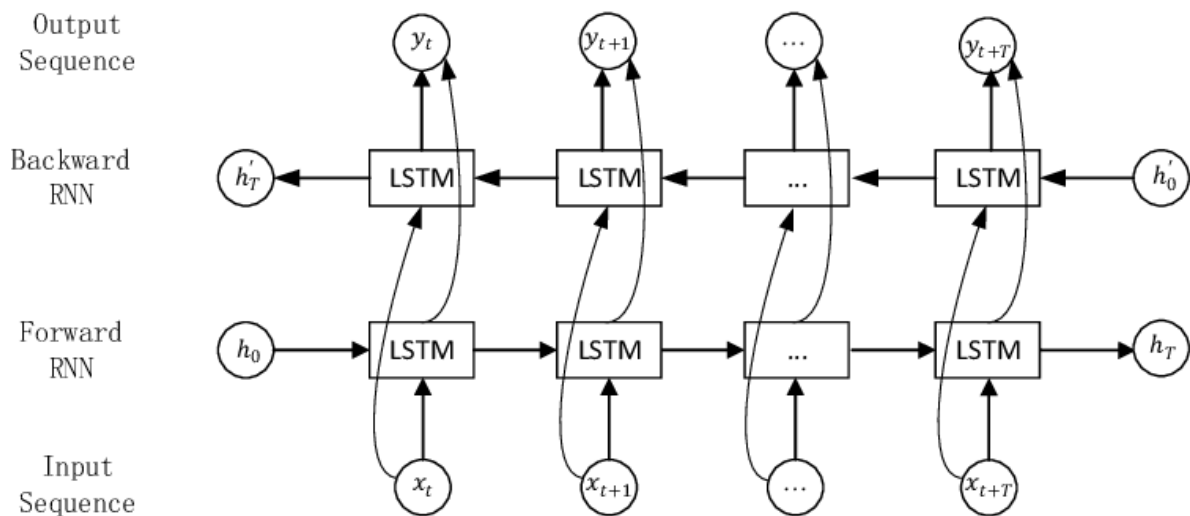
The model utilizes information from both past and future time steps, enabling a more comprehensive understanding of temporal dependencies within the data sequence.

BiLSTM networks are structured with two parallel LSTM layers:

- **Forward layer:** Processes the sequence in a forward direction, from start to finish.
- **Backward layer:** Processes the sequence in reverse, from end to beginning.

The outputs from these layers are usually combined, either through concatenation or averaging, to generate the final result, thereby enhancing the model's predictive capabilities. This bidirectional processing approach has been demonstrated to improve accuracy in time-series forecasting tasks, particularly for patterns that depend on contextual relationships (Kocoglu, Gorell, and McElroy 2021).

For forecasting breakdown pressure, the ability of BiLSTMs to integrate information from both directions can be highly advantageous. This dual-context mechanism enables the model to better capture transitions and anomalies within the data, offering a more refined prediction framework. The equations governing the BiLSTM operations are similar to LSTM but applied bidirectionally as shown in Figure 15:



**Figure 15.** A schematic representation of a Bi-LSTM network, showing the flow of information in both directions (Saxena et al., 2013)

**Vector Autoregressive Moving-Average (VARMA) Model:** The VARMA model extends the ARMA framework to a multivariate setting, enabling the capture of linear

interdependencies across multiple time series. It integrates the Vector Autoregressive (VAR) and Vector Moving Average (VMA) components to model both the relationships between variables and the impact of past error terms on current observations (Lütkepohl 2006).

The VARMA model is represented as:

$$\Phi(L)Y_t = \theta(L)\epsilon_t \quad (26)$$

where:

- $\Phi(L)$  is the matrix polynomial for the VAR component,
- $\theta(L)$  is the matrix polynomial for the VMA component,
- $Y_t$  is the vector of time series values at time  $t$ ,
- $\epsilon_t$  is the vector of error terms.

The VARMA model computes forecasts by solving the above equation iteratively, using historical data and error terms (Mbunge et al. 2022) (Chatfield 1988).

**Seasonal Autoregressive Integrated Moving Average (SARIMA) Model:** SARIMA model extends the ARIMA framework to handle seasonal data patterns, making it particularly effective for data with periodic fluctuations (Vagropoulos et al. 2016). It incorporates seasonal components into the autoregressive (AR), differencing (I), and moving average (MA) terms (Dubey et al. 2021). The general form of the SARIMA model is:

$$\Phi_p(L)\Phi_p(L^S)(1-L)^d(1-L^S)^DY_t = \theta_q(L)\theta_q(L^S)\epsilon_t \quad (27)$$

where:

- $\Phi_p(L)$  and  $\Phi_p(L^s)$  are polynomials for the non-seasonal and seasonal AR components,
- $\theta_q(L)$  and  $\theta_q(L^s)$  are polynomials for the non-seasonal and seasonal MA components,
- $L$  is the lag operator,
- $\epsilon_t$  is the error term.

The SARIMA model's ability to handle seasonality is one of its key advantages, particularly in applications such as weather forecasting, sales prediction, and reservoir management where periodic trends are prevalent (Milenković et al. 2018). Its reliance on explicit mathematical formulations ensures interpretability, enabling users to understand how the seasonal and non-seasonal components influence the forecast. Moreover, SARIMA's structured approach is computationally efficient (Ma et al. 2021), making it a practical choice for smaller datasets or applications where computation time is critical.

Despite its strengths, the SARIMA model also has notable limitations. Its reliance on stationarity assumptions means that pre-processing steps like differencing and detrending are necessary to ensure stable predictions (Chen et al. 2018). Additionally, SARIMA struggles with datasets that exhibit complex nonlinear relationships or sudden changes in patterns, as its framework is inherently linear.

This study investigates a combination of machine learning and statistical models, specifically LSTM, Bidirectional LSTM (Bi-LSTM), VARMA, and SARIMA, for

forecasting breakdown pressure in well-log data. These models utilize depth and other well-log parameters as sequential time series inputs to capture temporal dependencies, enabling accurate predictions crucial for wellbore integrity analysis. The key characteristics of these models are summarized below:

- LSTM is well-suited for learning long-term dependencies in sequential data, making it effective for time series problems.
- Bidirectional LSTM enhances performance by processing data in both forward and backward directions, allowing the model to leverage information from both past and future time steps.
- VARMA (Vector Autoregressive Moving Average) and SARIMA (Seasonal ARIMA) are statistical models that use historical data and autoregressive relationships to predict future values, with SARIMA specifically addressing seasonality in the time series.

The experiments were conducted using Python-based frameworks, including TensorFlow for LSTM models and stats models for VARMA and SARIMA. Given the limited number of data points (**158**) and the need for efficient training, we designed the models to balance computational complexity and predictive accuracy.

The models were designed, as shown in Table 6, to ensure a fair comparison while accounting for the sequential nature of the well-log data and time series characteristics. To achieve optimal performance, we employed key techniques such as "ReduceLROnPlateau" for dynamic learning rate adjustment and ModelCheckpoint to save the models at the minimum loss. The following hyperparameters were used:

**Table 6. Hyperparameters for Machine Learning and Statistical Models**

| <b>Hyperparameter</b>      | <b>LSTM / Bi-LSTM</b>                 | <b>VARMA</b>               | <b>SARIMA</b>                     | <b>Explanation</b>                                                            |
|----------------------------|---------------------------------------|----------------------------|-----------------------------------|-------------------------------------------------------------------------------|
| <b>Number of layers</b>    | 2                                     | N/A                        | N/A                               | Two layers allow the model to learn complex temporal patterns.                |
| <b>Number of units</b>     | 64 per layer                          | N/A                        | N/A                               | Determines the number of neurons in each layer for capturing dependencies.    |
| <b>Activation function</b> | ReLU                                  | N/A                        | N/A                               | Handles non-linearities effectively.                                          |
| <b>Input shape</b>         | (look-back steps, number of features) | Time Series (Multivariate) | Time Series (Depth + Seasonality) | Defines the input dimensions for sequential data.                             |
| <b>Dense layer units</b>   | 32                                    | N/A                        | N/A                               | Fully connected layer balances model complexity and performance.              |
| <b>Optimizer</b>           | Adam                                  | N/A                        | N/A                               | Adaptive learning rate optimizer.                                             |
| <b>Learning Rate</b>       | Reduce LR On Plateau                  | N/A                        | N/A                               | Reduces LR when loss plateaus to improve convergence.                         |
| <b>Loss function</b>       | Mean Absolute Error (MAE)             | N/A                        | N/A                               | Commonly used for regression tasks to minimize absolute prediction errors.    |
| <b>Evaluation Metrics</b>  | MAE, MSE                              | MAE, MSE                   | MAE, MSE                          | Used to measure forecasting accuracy by comparing predictions with actuals.   |
| <b>Epochs</b>              | 1000                                  | N/A                        | N/A                               | Adequate training iterations for a dataset with 158 data points.              |
| <b>Order (p, d, q)</b>     | N/A                                   | (1, 1)                     | (1, 1, 1) (1, 0, 1)               | VARMA and SARIMA orders to capture autoregressive and moving average effects. |

## 2.3 Evaluation and Sensitivity Analysis

### 2.3.1 Evaluation Metrics

Various standard error metrics were applied to measure the predictive performance of machine learning models. These include Mean Absolute Error (MAE), Mean Squared Error (MSE), Root Mean Square Error (RMSE), Average Absolute Percentage Error (AAPE), Standard Deviation (SD), and the Coefficient of Determination ( $R^2$ ). Each of these metrics provides insight into different aspects of model accuracy and error. Below are their definitions and formulas:

**Mean Absolute Error (MAE):** MAE measures the average error between the predicted and actual breakdown pressure values. It ignores the direction of the errors, focusing purely on their magnitude as expressed in eq 28:

$$MAE = \frac{1}{n} \sum_{i=1}^n |BP_{measured} - BP_{predicted}| \quad (28)$$

Here:

- $BP_{measured}$  is the observed value of breakdown pressure,
- $BP_{predicted}$  is the model's estimated value,
- $n$  is the total number of samples.

A smaller MAE value indicates that the model's predictions are close to the actual observations.

**Mean Squared Error (MSE):** MSE measures the mean of the squared discrepancies between predicted and observed values, emphasizing significant errors as a result of the squaring operation described in eq 29:

$$MSE = \frac{\sum (BP_{measured} - BP_{predicted})^2}{n} \quad (29)$$

A lower MSE value suggests fewer large deviations from the true values.

**Root Mean Square Error (RMSE):** RMSE is a frequently used evaluation metric that measures the standard deviation of prediction errors by taking the square root of the Mean Squared Error (MSE). It gives an indication of the typical error magnitude and estimated by eq 30:

$$RMSE = \sqrt{\frac{\sum (BP_{measured} - BP_{predicted})^2}{n}} \quad (30)$$

A lower RMSE value indicates better model performance, as it reflects a lower average error.

**Average Absolute Percentage Error (AAPE):** AAPE measures the percentage error between the predicted and actual values given by eq 31, making it useful for comparing errors across different scales:

$$AAPE = \frac{1}{n} \sum \left| \frac{BP_{measured} - BP_{predicted}}{BP_{measured}} \right| \times 100 \quad (31)$$

Lower AAPE values indicate higher prediction accuracy.

**Standard Deviation (SD):** SD helps assess the spread of the prediction errors. A high SD indicates significant variability, while a low SD suggests that the predictions are consistently close to the actual values. The expression given by eq 32:

$$SD = \sqrt{\frac{1}{n-1} \sum \left( \frac{X_{measured} - X_{predicted}}{X_{measured}} \right)^2} \quad (32)$$

**Coefficient of Determination ( $R^2$ ):**  $R^2$  is used to evaluate the proportion of variance in the observed data explained by the model. It varies between 0 and 1, with values nearer to one representing a stronger fit, and given by the following eq 33:

$$R^2 = 1 - \frac{\sum (BP_{measured} - BP_{predicted})^2}{\sum (BP_{measured} - \text{mean } BP_{predicted})^2} \quad (33)$$

Higher  $R^2$  values indicate that the model explains a larger portion of the variance in the data.

### 2.3.2 Sensitivity Analysis Metrics

In this study, a comprehensive sensitivity analysis has been conducted to assess the impact of various input features on the prediction of Breakdown Pressure (BP). Several methods were employed, including preliminary correlation analysis, advanced neural network feature weight extraction, deep learning attention mechanisms, correction using the First-Waugh-Lovell theorem, and SHapley Additive exPlanations (SHAP) for ensemble models.

**Pearson Correlation Coefficient:** This coefficient measures the linear relationship between two variables  $x$  and  $y$ , defined as eq 34:

$$r = \frac{\sum(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum(x_i - \bar{x})^2} \sqrt{\sum(y_i - \bar{y})^2}} \quad (34)$$

Here:

- $x_i$  and  $y_i$  are the values of the two variables.
- $\bar{x}$  and  $\bar{y}$  are the mean values of  $x$  and  $y$ .

**Spearman Rank Correlation Coefficient:** This non-parametric metric evaluates the rank correlation (monotonic relationship) between features given by eq 35:

$$\rho = 1 - \frac{6 \sum d_i^2}{n(n^2 - 1)} \quad (35)$$

where:

- $d_i$  is the difference between the ranks of  $x_i$  and  $y_i$ .
- $n$  is the number of data points.

**Kolmogorov-Arnold Neural Network (KAN) Weights:** The Kolmogorov-Arnold Neural Network approximates complex functions and allows us to extract feature importance based on its learned weights. The optimal neuron configuration was determined through validation. Feature importance is calculated by eq 36:

$$W_j = \frac{|\sum_{i=1}^n w_{ij}|}{\sum_{j=1}^m |\sum_{i=1}^n w_{ij}|} \quad (36)$$

Where:

- $W_{ij}$  is the importance weight for feature  $j$ .

- $w_{ij}$  represents the connection weight from input  $i$  to the neuron related to feature  $j$ .
- $n$  is the number of neurons in the hidden layer, and  $m$  is the total number of features.

This method highlighted permeability and injection rate as the most influential features.

**TabNet Deep Learning Sensitivity Analysis:** TabNet uses an attentive feature selection mechanism, leveraging masks that learn to prioritize different features during training.

The importance of each feature is quantified by its attention score eq 37:

$$A_j = \frac{\exp(\text{Mask}_j X)}{\sum_{j=1}^m \exp(\text{Mask}_j X)} \quad (37)$$

Here:

- $A_j$  is the attention score for feature  $j$ .
- $\text{Mask}_j$  is the learned mask weight for feature  $j$ .
- $X$  is the input feature vector.

**First-Waugh-Lovell (FWL) Theorem for Pressurization Rate Adjustment:** The FWL theorem is used to control for the impact of pressurization rate while accounting for its interaction with other features. This correction ensures a more accurate estimation of feature effects expressed in eq 38:

$$Y_{adj} = Y - \beta_{rate}X_{rate} \quad (38)$$

Where:

- $Y_{adj}$  is the adjusted target (BP).
- $\beta_{rate}$  is the regression coefficient for the pressurization rate.
- $X_{rate}$  is the pressurization rate feature.

This adjustment process reduces the bias introduced by the pressurization rate, allowing for a clearer interpretation of the effects of other variables.

**SHapley Additive exPlanations (SHAP) for Ensemble Models:** SHAP values employ a game-theoretic approach to breakdown and attribute the effect of each feature on the outcome generated by the model. This method was applied to ensemble models (XGBoost, LightGBM, CatBoost) to interpret their complex decisions.

SHAP Value estimated by following eq 39:

$$\varphi_j = \sum_{S \subseteq N \setminus \{j\}} \frac{|S|!(|N| - |S| - 1)!}{|N|!} [f(S \cup \{j\}) - f(S)] \quad (39)$$

Where:

- $\varphi_j$  is the SHAP value for feature  $j$ .
- $N$  is the set of all features.
- $S$  is any subset of features excluding  $j$ .
- $f(S)$  is the model output for subset  $S$ .

## **CHAPTER 3**

### **RESULTS AND DISCUSSIONS**

#### **3.1 Simulation Results**

##### **3.1.1 Performance of the Kirsch Model**

Figure 16 represents the results of the Kirsch model for estimating breakdown pressure for field case. While the data points are somewhat scattered, a notable portion aligns with the perfect correlation line, indicating that the model captures some general trends.

However, the  $R^2$  value of 0.57 suggests only moderate accuracy. This means that while the model provides a reasonable first estimate, it does not fully account for the complex geological variations and fluid interactions that influence breakdown pressure in real-world formations. The spread of data points highlights the model's limitations, particularly in handling stress anisotropy and heterogeneous rock properties.

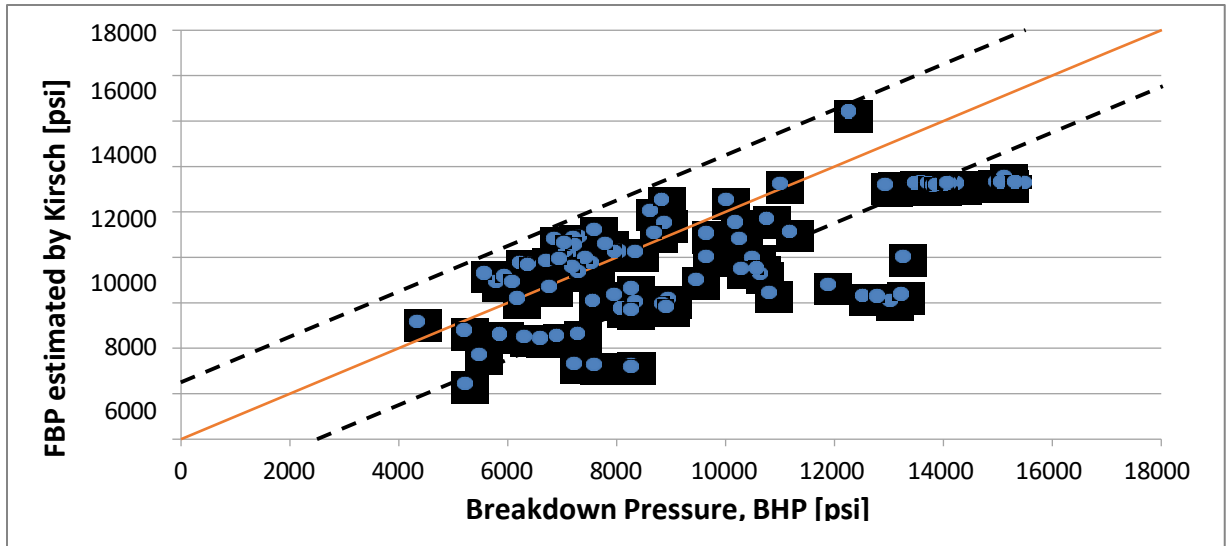
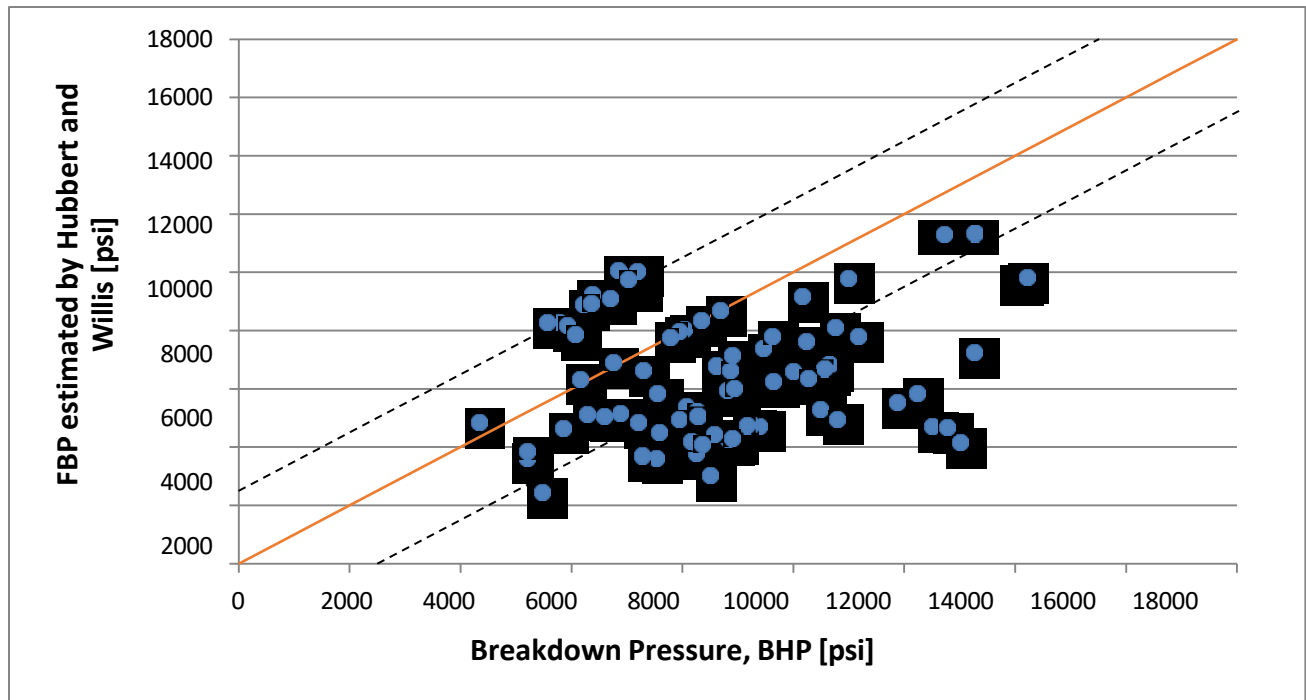


Figure 16. Comparison of breakdown pressure estimated using the Kirsch approach with the actual measured pressure

### 3.1.2 Performance of the Hubbert and Willis Model

Figure 17 presents the results for the Hubbert and Willis model, which builds upon the Kirsch model but incorporates additional stress factors. The results indicate a wider scatter of data points, with many falling above the perfect correlation line, meaning the model overestimates breakdown pressure in several cases. The  $R^2$  value of 0.49 suggests even lower accuracy compared to the Kirsch model. This could be due to the model's assumption of an idealized stress distribution, which does not always hold true in complex formations. Additionally, the inability to accurately capture fluid penetration effects and varying rock permeability likely contributes to the deviations from actual field values.



**Figure 17. Comparison of breakdown pressure estimated using the Hubbert and Willis approach with the actual measured pressure.**

### 3.1.3 Performance of Haimson and Fairhurst Model

Figure 18 illustrates the performance of the Haimson and Fairhurst poroelastic model, which accounts for fluid penetration into the rock matrix. Unlike the previous models, this one tends to underestimate breakdown pressure, as most data points fall below the perfect correlation line. The  $R^2$  value of 0.43 further confirms that the model struggles to fully capture the range of observed breakdown pressures. This could be due to variations in pore pressure diffusion, rock permeability, and other mechanical factors that the model does not explicitly consider. While this model is an improvement in incorporating fluid effects, it still lacks the precision needed for accurate field predictions.

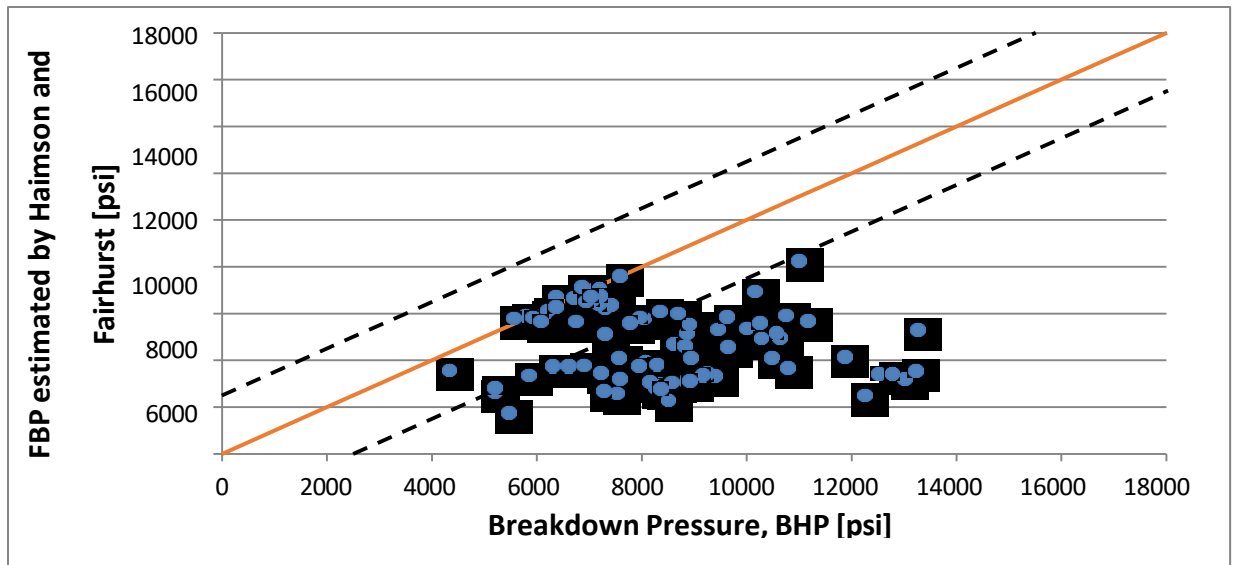


Figure 18. Comparison of breakdown pressure predicted by the Haimson and Fairhurst approach with the actual pressure values.

### 3.1.4 Performance of Rummel models

Figure 19 presents the results of the Rummel 1 model, which achieved an  $R^2$  value of 0.51, indicating moderate predictive accuracy. The scatter plot shows a reasonable alignment with the perfect correlation line, though it tends to underestimate breakdown pressure in lower ranges, suggesting limitations in capturing stress redistribution effects.

Figure 20 illustrates the performance of the Rummel 2 model, which had a lower  $R^2$  value of 0.44, reflecting weaker predictive reliability. The data points display a wider scatter and a tendency to underestimate breakdown pressure across all ranges, likely due to the model's assumption that pressure in the crack is only 75% of the wellbore pressure.

Figure 21 **Error! Reference source not found.** shows the Rummel 3 model, which performed slightly better with an  $R^2$  value of 0.48. The scatter plot reveals that while the model works well for mid-range breakdown pressures, it overestimates higher-pressure

values, likely due to its assumption of linear pressure drop along the fracture. Finally, Figure 22 presents the Rummel 4 model, which achieved the highest accuracy among the four models with an  $R^2$  value of 0.53. The results indicate that while this model better captures higher breakdown pressures, it still deviates in lower ranges, suggesting that the quadratic pressure drop assumption improves predictions but does not fully account for real-world variations.

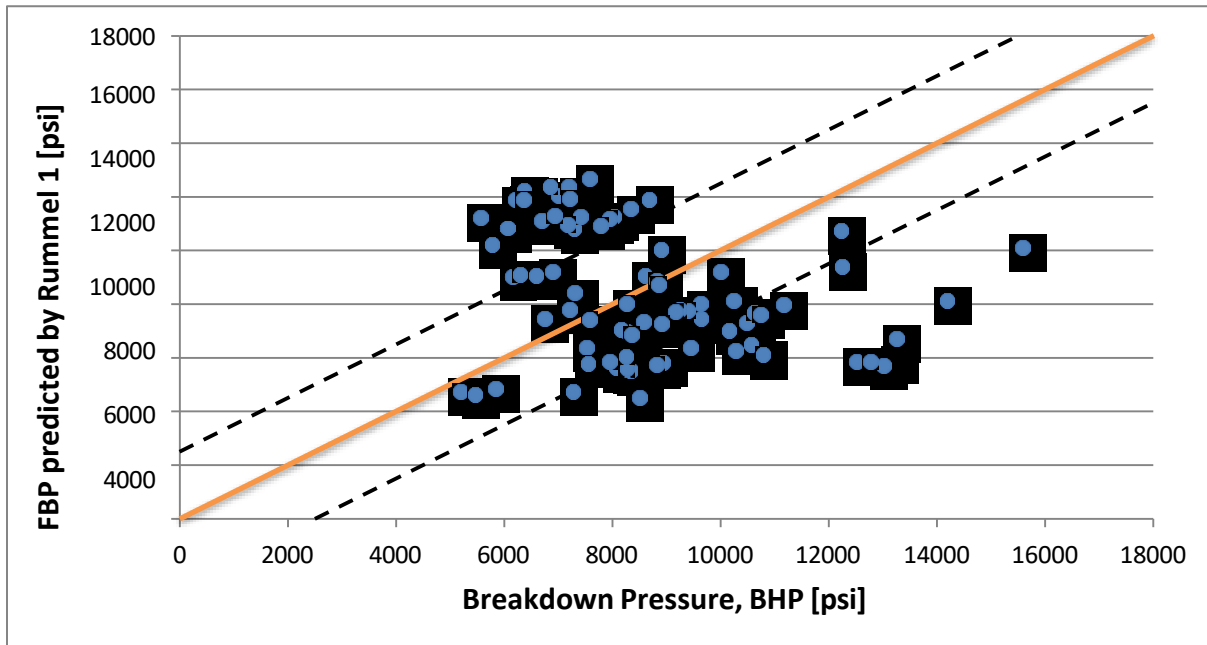


Figure 19. Evaluation of the breakdown pressure estimated through the Rummel 1 approach against actual pressure data.

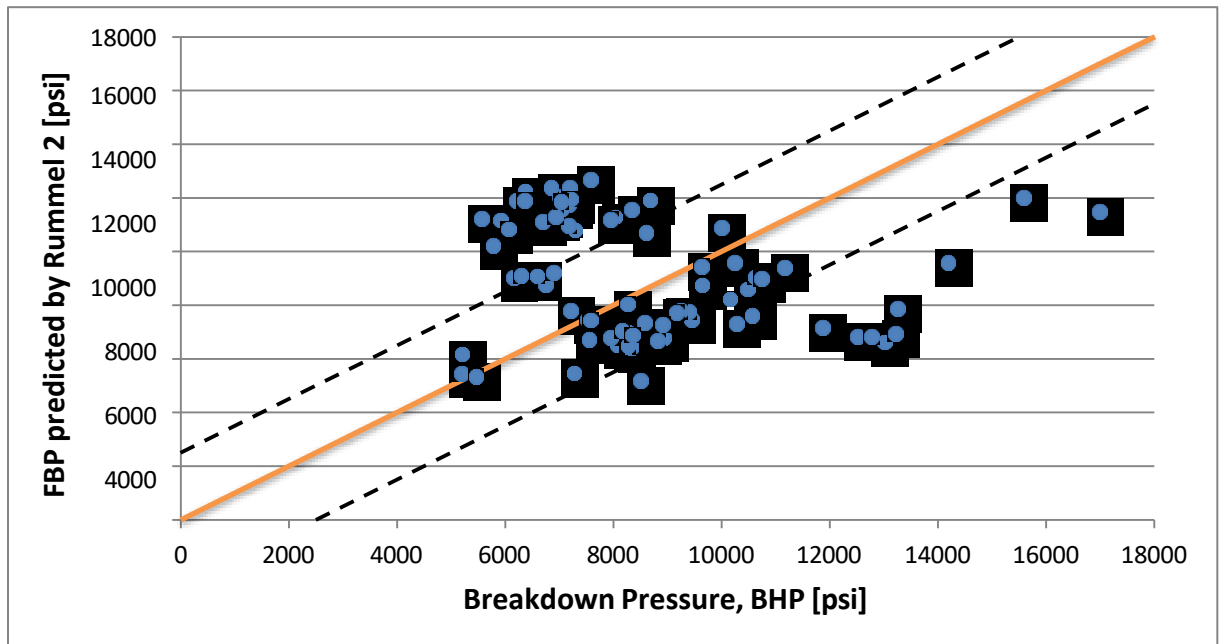


Figure 20. Assessment of breakdown pressure computed using the Rummel 2 approach compared to the actual pressure.

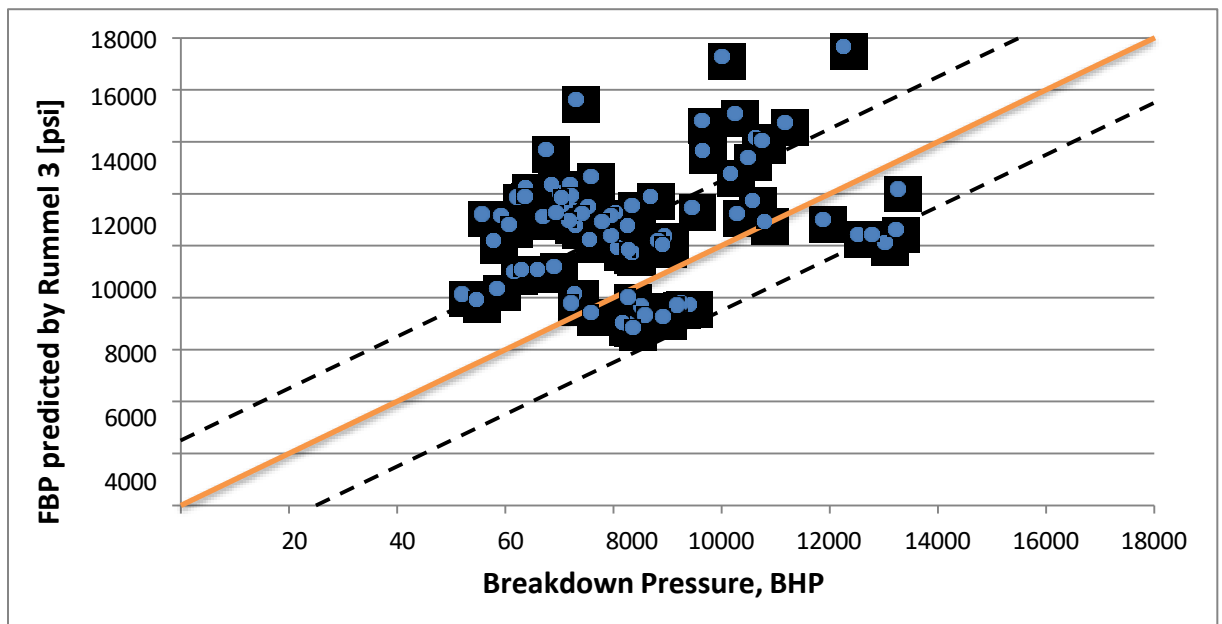


Figure 21. Difference between the Formation Breakdown Pressure estimated by the Rummel 3 approach and the actual pressure

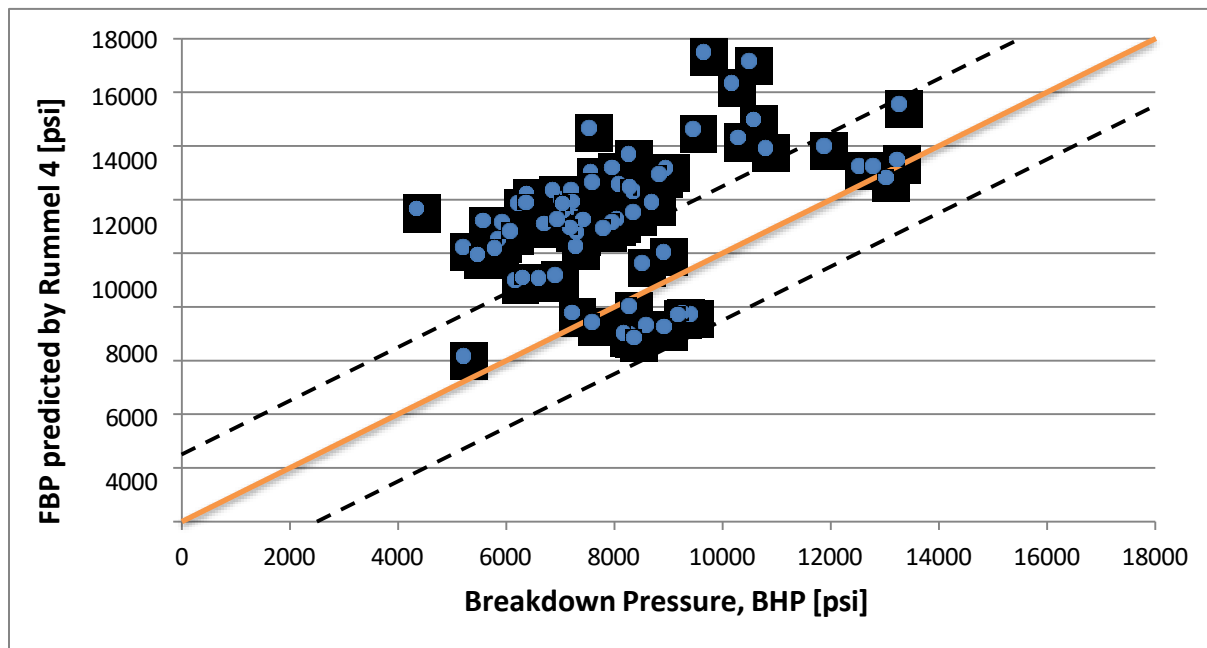


Figure 22. Analysis of breakdown pressure determined by the Rummel 4 approach relative to the actual pressure values

### 3.1.5 Performance of Schmitt and Zoback model

Figure 23 illustrates the performance of the Schmitt and Zoback model in predicting breakdown pressure, comparing its estimates with actual field data. The scatter plot shows a significant spread of data points, with many deviating from the perfect correlation line, indicating that the model struggles with prediction accuracy. The  $R^2$  value for this model is relatively low, reflecting its limited ability to fully capture variations in rock properties, stress anisotropy, and pore pressure effects. The distribution of points suggests a noticeable skew towards underestimation in higher breakdown pressure values, meaning that the model tends to predict lower pressures than those

observed in the field. This could be attributed to its poroelastic assumptions, which might not fully account for complex fluid interactions and formation heterogeneity

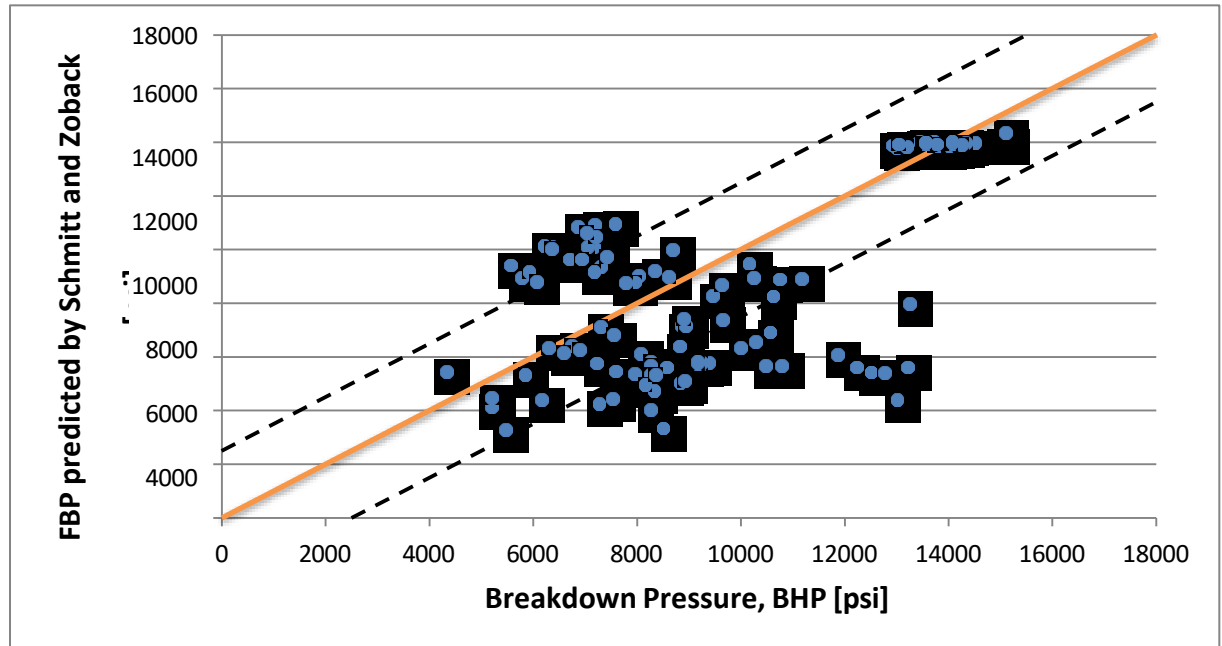


Figure 23. Comparison between breakdown pressure derived from the Schmitt and Zoback approach and the actual pressure readings.

### 3.1.6 Performance of Ito and Hayashi model

Figure 24 presents the performance of a breakdown pressure prediction model, comparing its estimates to actual field data. The scatter plot shows a noticeable skew to the left, indicating that the model underestimates breakdown pressure at higher values while maintaining better alignment in the mid-range pressure levels. The  $R^2$  value is relatively low, suggesting that the model struggles with accuracy, particularly in capturing variations at both high and low pressure extremes. The clustering of points below the perfect correlation line implies that the model does not fully account for stress

redistribution and formation heterogeneity, leading to consistent underprediction in high-pressure scenarios

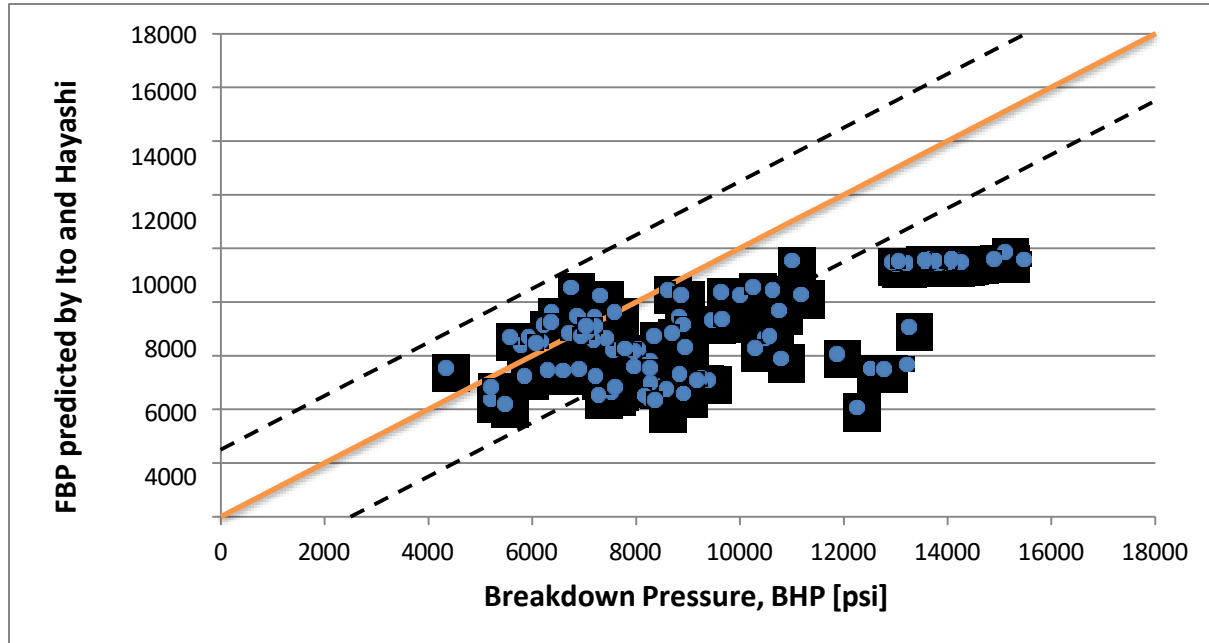


Figure 24. Comparison of breakdown pressure calculated through the Ito and Hayashi approach with observed pressure values.

### 3.1.7 Performance of Aadnoy and Balayneh models

Figure 25 and Figure 26 illustrate the predictive performance of the Aadnoy and Balayneh 1 and 2 models, with  $R^2$  values of 0.68 and 0.72, respectively, making them the most accurate among the evaluated models. These models outperform others by incorporating thermoporoelastic effects, accounting for both load history and temperature variations, which are critical in real-world wellbore conditions. The scatter plots show that predictions are better aligned with the perfect correlation line, indicating improved accuracy across different breakdown pressure ranges. The inclusion of Poisson's scaling

factor and temperature corrections in these models enhances their ability to simulate stress redistribution and fluid-rock interactions more effectively than purely elastic or poroelastic models. However, limitations remain, as minor deviations at higher pressures suggest potential inaccuracies in capturing complex non-linear fracture mechanics or cases where anisotropic stress conditions dominate.

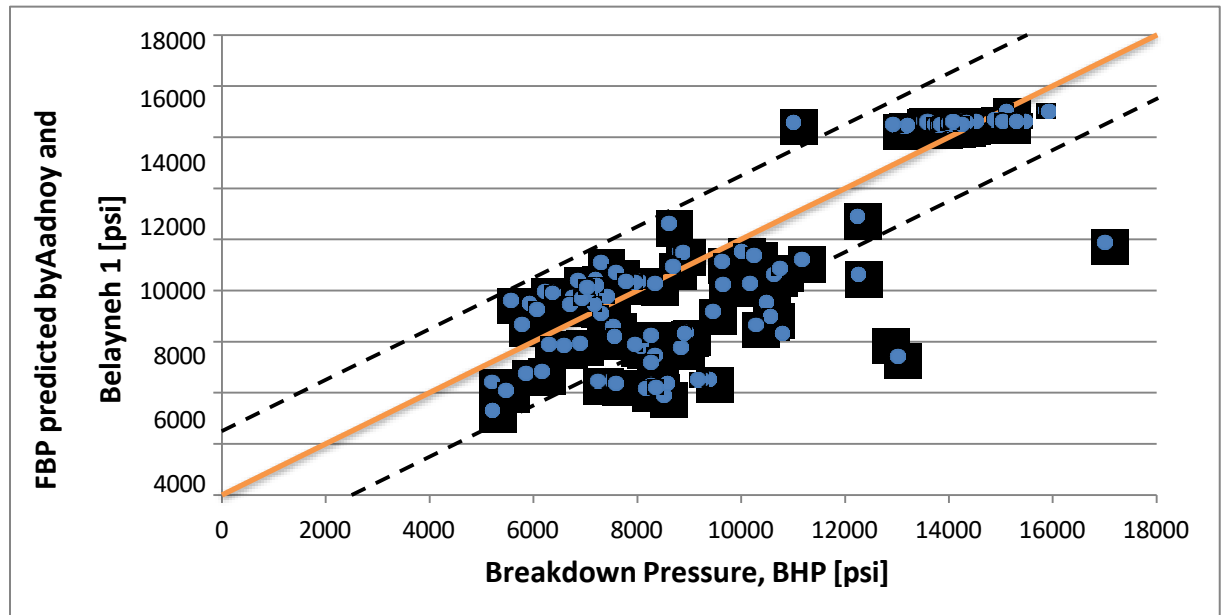


Figure 25. Assessment of breakdown pressure estimated using the Aadnoy and Balyaneh 1 approach against the actual measured pressure.

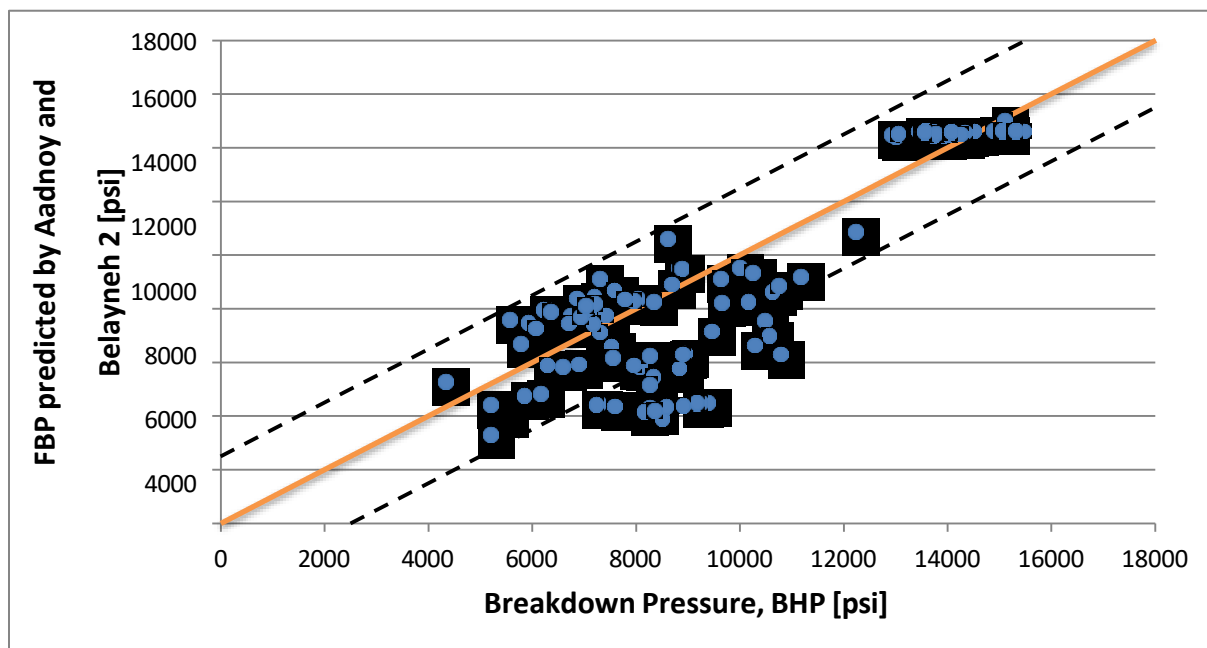


Figure 26. Assessment of breakdown pressure estimated using the Aadnoy and Balyaneh 2 approach against the actual measured pressure.

### 3.1.8 Evaluation Metrics of Simulation Models

The following Table 7 discusses the evaluation metrics for various breakdown pressure prediction models, ranked from the most accurate to the least reliable based on their  $R^2$  values. The Aadnoy & Belayneh 2 model demonstrated the highest accuracy ( $R^2 = 0.72$ ), with the lowest RMSE and MAE, indicating its ability to closely estimate breakdown pressure by incorporating thermoporoelastic effects and wellbore load history. Similarly, Aadnoy & Belayneh 1 ( $R^2 = 0.68$ ) also performed well, further validating the importance of temperature corrections in improving prediction accuracy.

Among the Rummel models, Rummel 4 ( $R^2 = 0.53$ ) provided the best results, benefiting from a quadratic pressure drop assumption, which better captures stress variations along fractures. The Kirsch model ( $R^2 = 0.57$ ), despite being a basic elastic model, showed better performance than some poroelastic approaches, suggesting that simpler models can still be useful in certain conditions.

**Table 7. Performance Comparison of Formation Breakdown Pressure Simulation Models**

| <b>Model</b>                   | <b>R<sup>2</sup></b> | <b>RMSE (psi)</b> | <b>MSE (psi<sup>2</sup>)</b> | <b>MAE (psi)</b> | <b>SD (psi)</b> |
|--------------------------------|----------------------|-------------------|------------------------------|------------------|-----------------|
| <b>Aadnoy &amp; Belayneh 2</b> | 0.72                 | 1384.62           | 1916902.91                   | 1107.69          | 1343.15         |
| <b>Aadnoy &amp; Belayneh 1</b> | 0.68                 | 1486.75           | 2209813.06                   | 1189.4           | 1442.78         |
| <b>Rummel 4</b>                | 0.53                 | 1759.21           | 3091672.98                   | 1407.37          | 1702.68         |
| <b>Kirsch</b>                  | 0.57                 | 1720.45           | 2958900.2                    | 1376.36          | 1648.31         |
| <b>Rummel 1</b>                | 0.51                 | 1802.33           | 3248393.43                   | 1441.86          | 1756.24         |
| <b>Hubbert &amp; Willis</b>    | 0.49                 | 1835.72           | 3369402.71                   | 1468.58          | 1784.43         |
| <b>Rummel 3</b>                | 0.48                 | 1859.76           | 3465271.3                    | 1487.81          | 1807.65         |
| <b>Schmitt &amp; Zoback</b>    | 0.46                 | 1901.52           | 3615773.51                   | 1521.22          | 1860.29         |
| <b>Rummel 2</b>                | 0.44                 | 1932.47           | 3733634.5                    | 1545.98          | 1890.34         |
| <b>Haimson &amp; Fairhurst</b> | 0.43                 | 1958.14           | 3831404.82                   | 1566.51          | 1912.57         |

On the lower end, Haimson & Fairhurst ( $R^2 = 0.43$ ) and Rummel 2 ( $R^2 = 0.44$ ) exhibited the highest RMSE and MAE values, indicating significant deviations from actual breakdown pressure measurements. These models tended to underestimate pressure in high-stress conditions, making them less reliable for complex formations.

## **3.2 Machine Learning Results for Experimental Case**

### **3.2.1 One-Hot Encoding**

The analysis commenced by splitting the dataset of 144 unique data points into training and testing sets, using a 75:25 ratio to ensure sufficient data for model training. This split was conducted in a stratified manner, meaning that the proportions of categorical features were maintained in both the training and testing sets. Stratification ensures that the model is exposed to a balanced distribution of data during training and testing, preventing biases and improving generalization. To further avoid data leakage and enhance model

reliability, the testing data points were carefully selected to fall within the range of the training data.

One-hot encoding, illustrated in Figure 27, was applied for categorical features. Rock types (sandstone, shale, and carbonate) were transformed using a binary system (0 for absence, 1 for presence). Additionally, the feature "Cylindricity" was encoded similarly (0 for False, 1 for True) since it represents a binary categorical variable. However, for models like CatBoost and TabNet, which can handle categorical features natively, explicit encoding was not necessary.

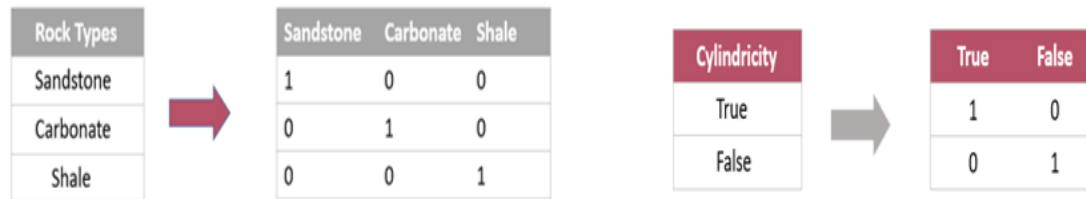


Figure 27. The transformation of the categorical variables to numerical

### 3.2.2 Correlation Analysis

Pearson and Spearman correlation coefficients were initially used to evaluate feature importance (Figure 28). Pearson correlation assesses linear relationships, while Spearman correlation examines monotonic trends. The results highlighted Confining Pressure ( $p_c$ ), Injection Rate ( $Q$ ), Rock Tensile Strength ( $\sigma_t$ ), Unconfined Compressive Strength (UCS), and Young's Modulus ( $E$ ) as the top five influential features. However, features like

Poisson’s Ratio ( $\nu$ ) exhibited low correlation, indicating negligible impact on Breakdown Pressure (BP), despite some studies suggesting otherwise.

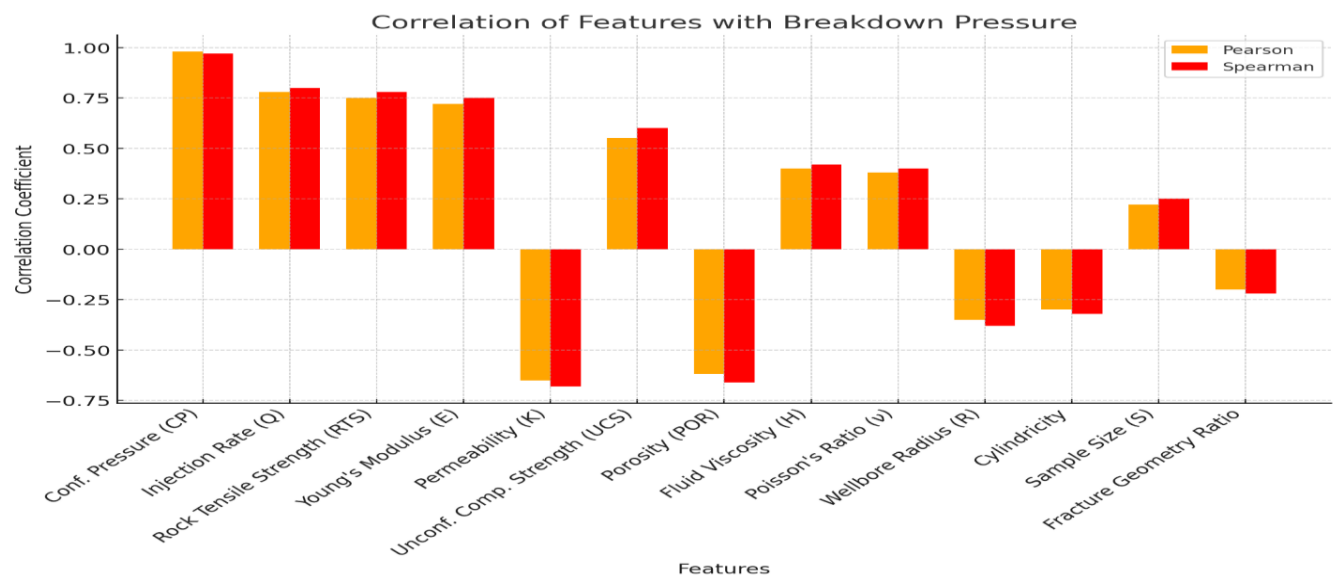
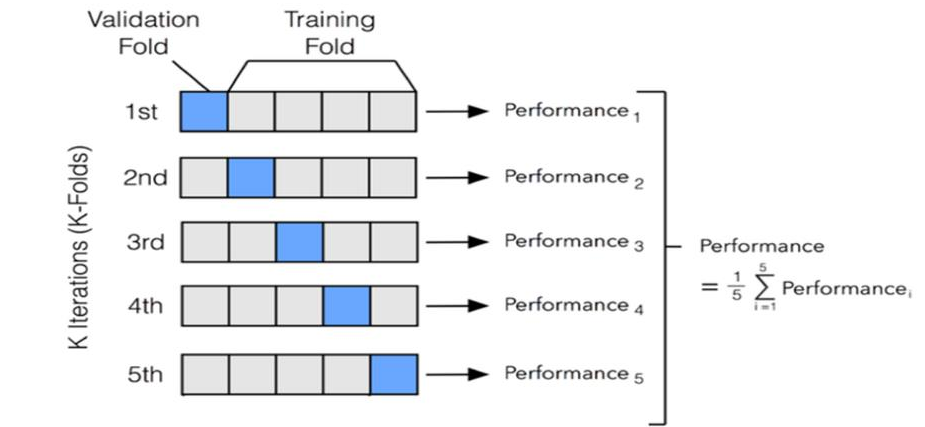


Figure 28. Heatmap of Spearman and Pearson Correlations for Breakdown Pressure Features

3.2.3 Model Training and Cross-Validation

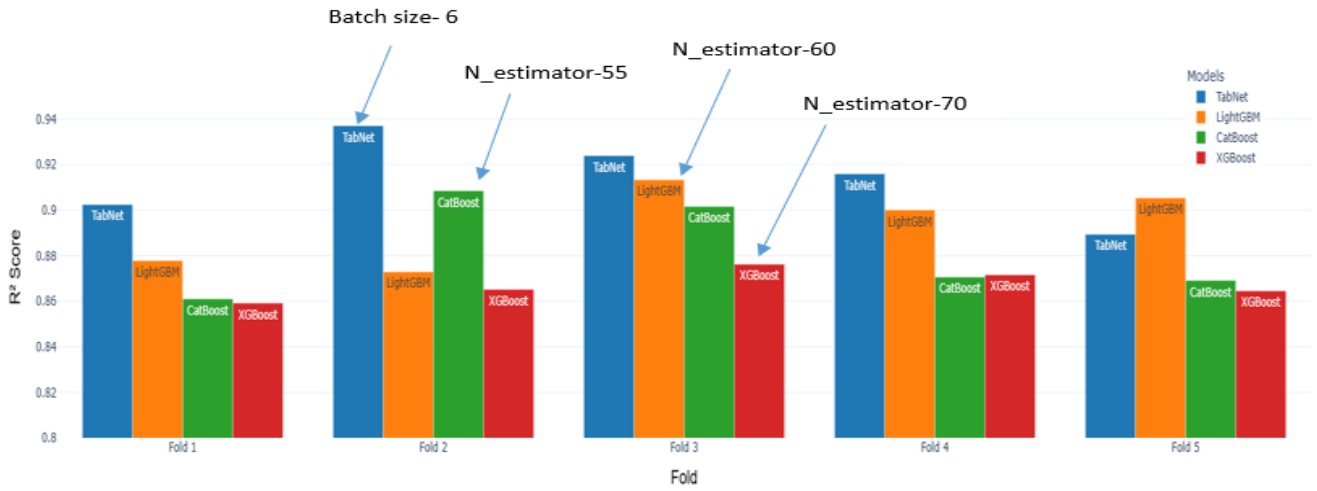
A comprehensive 5-fold cross-validation (Figure 29) was employed for each model to ensure robust hyperparameter optimization.



**Figure 29. A 5-Fold Cross-Validation approach applied in the research**

The best configurations identified through grid search (Figure 30) include the following:

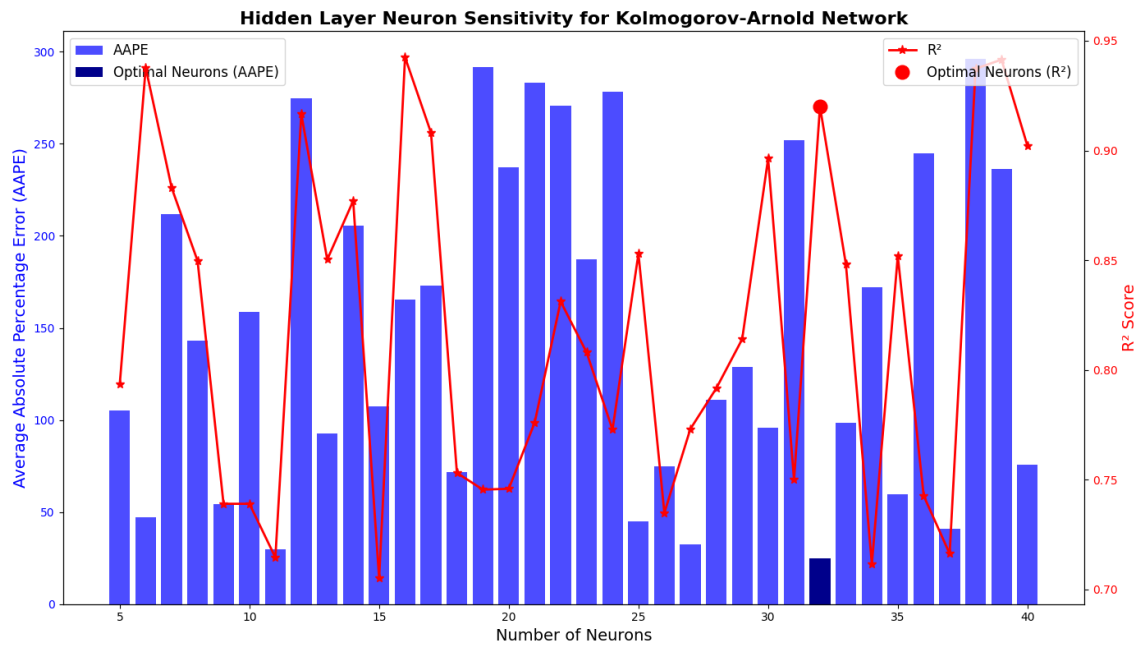
- TabNet: Batch size of 6, learning rate optimized during cross-validation.
- LightGBM: N\_estimators set to 60, with maximum depth adjusted based on validation metrics.
- CatBoost: N\_estimators set to 70, leveraging its native support for categorical features.
- XGBoost: N\_estimators tuned to 55 with a finely adjusted learning rate.



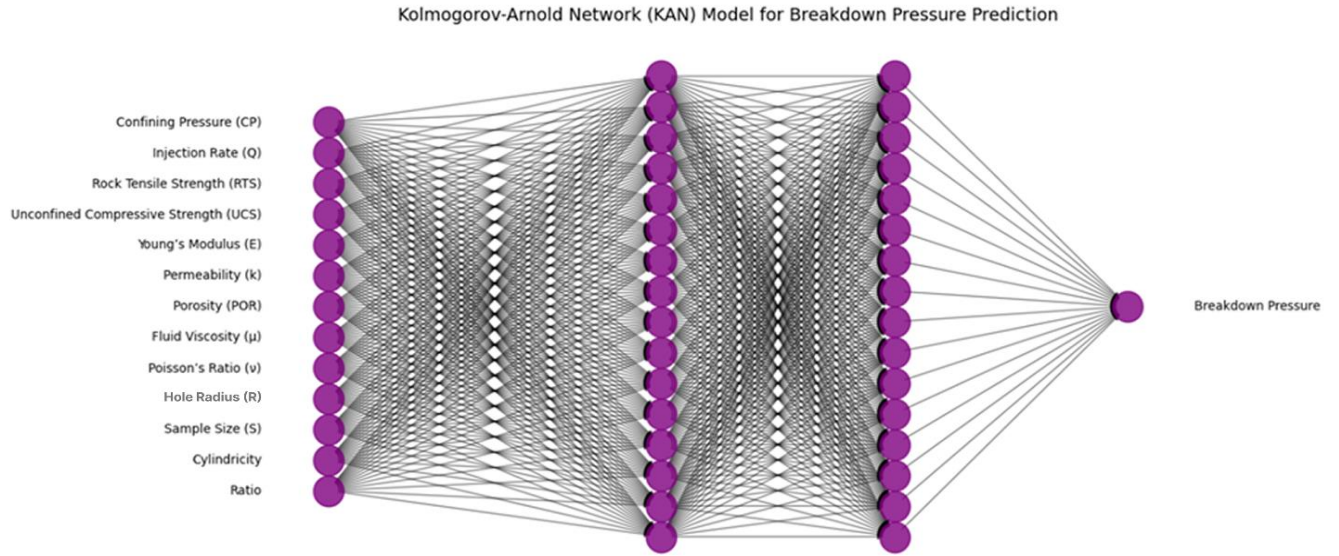
**Figure 30. Estimated optimal parameters for Ensemble and TabNet models**

The parameter  $N\_estimators$  represents the number of trees or boosting iterations in ensemble models such as LightGBM, CatBoost, and XGBoost. Increasing this value generally enhances model performance by reducing bias. However, excessively high values may lead to overfitting, making the model overly complex and limiting its ability to generalize effectively to new data. Batch size, on the other hand, determines the number of samples processed simultaneously in each training iteration, particularly in deep learning models like TabNet. Smaller batch sizes result in more precise updates to the model's parameters, though they require longer training times. Conversely, larger batch sizes expedite the training process but may lead to less stable convergence due to increased noise in updates.

The Kolmogorov-Arnold Network (KAN) model was subjected to a neuron sensitivity assessment, identifying 32 as the optimal number of neurons based on minimized Average Absolute Percentage Error (AAPE) and maximized  $R^2$  score (Figure 31). The optimal neuron configuration effectively balanced complexity and predictive power. The structure of KAN model with 2 hidden layers and 14 input features is shown in Figure 32.



**Figure 31. Neuron Number Sensitivity For Kolmogorov-Arnold Network**



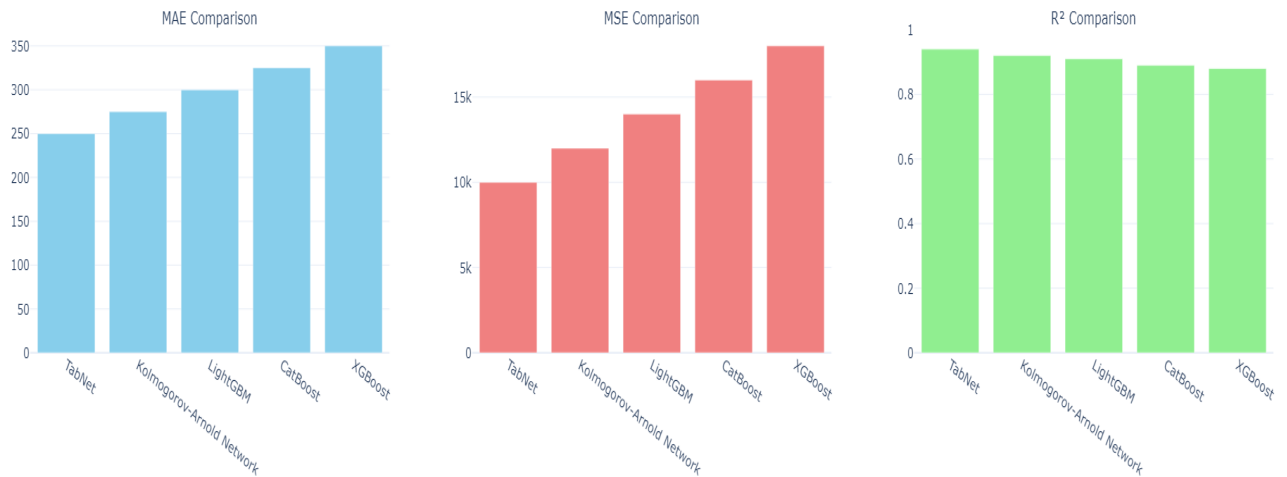
**Figure 32. Final Structure of Kolmogorov-Arnold Network**

The hyperparameters of these models were optimized, and the optimal configurations are shown in Table 8

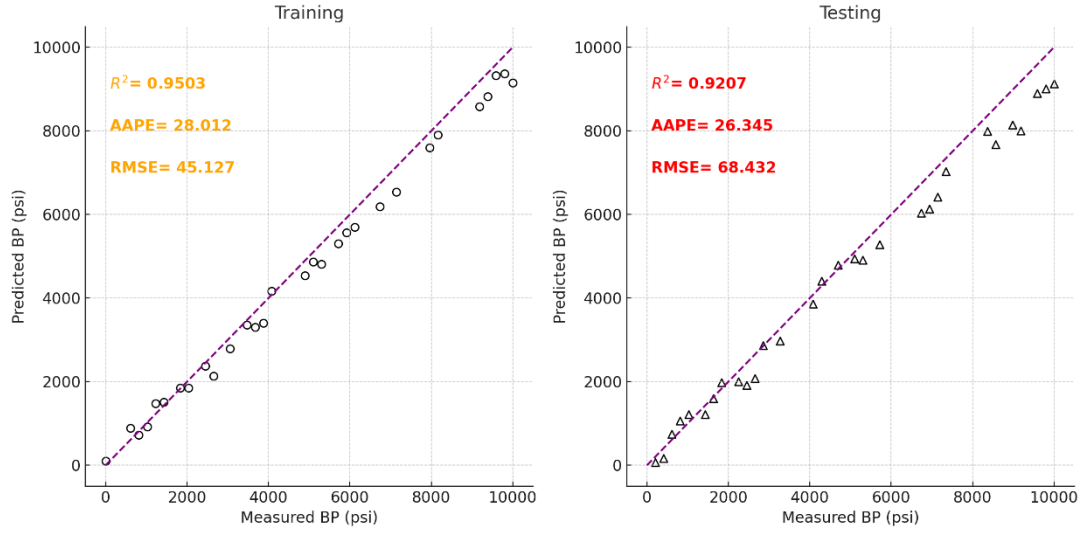
**Table 8. Optimized Hyperparameter Ranges Utilized for Training Machine Learning Algorithms**

| <b>Model</b>                           | <b>Hyperparameters</b>              | <b>Ranges Explored</b>          | <b>Selected Values</b> |
|----------------------------------------|-------------------------------------|---------------------------------|------------------------|
| <b>TabNet</b>                          | Virtual Batch Size                  | 4–10                            | 6                      |
|                                        | Number of Decision Steps            | 3–10                            | 5                      |
|                                        | Learning Rate                       | 0.01–0.5                        | 0.2                    |
|                                        | Batch Size                          | 10–50                           | 16                     |
|                                        | Number of Epochs                    | 50–300                          | 150                    |
| <b>LightGBM</b>                        | Number of Estimators (n_estimators) | 50–100                          | 60                     |
|                                        | Maximum Depth (max_depth)           | 5–15                            | 10                     |
|                                        | Learning Rate                       | 0.01–0.3                        | 0.1                    |
|                                        | Number of Leaves                    | 20–50                           | 30                     |
| <b>CatBoost</b>                        | Number of Estimators (n_estimators) | 50–100                          | 55                     |
|                                        | Maximum Depth                       | 5–10                            | 8                      |
|                                        | Learning Rate                       | 0.01–0.2                        | 0.15                   |
|                                        | L2 Regularization                   | 1–10                            | 3                      |
| <b>XGBoost</b>                         | Number of Estimators (n_estimators) | 50–100                          | 70                     |
|                                        | Maximum Depth                       | 5–10                            | 6                      |
|                                        | Learning Rate                       | 0.01–0.3                        | 0.1                    |
|                                        | Subsample Ratio                     | 0.5–1.0                         | 0.8                    |
| <b>Kolmogorov-Arnold Network (KAN)</b> | Number of Neurons (Hidden Layer)    | 5–40                            | 32                     |
|                                        | Activation Function                 | Relu, Tansig, Sigmoid           | Relu                   |
|                                        | Learning Rate                       | 0.1–0.5                         | 0.15                   |
|                                        | Optimization Algorithm              | Gradient Descent, Adam, RMSProp | Adam                   |

The results for model training and testing are shown in Figure 33. TabNet achieved the highest  $R^2$  scores, with 0.94, demonstrating excellent generalization capabilities. This highlights TabNet's robust performance in capturing complex patterns within the dataset while avoiding overfitting. The Kolmogorov-Arnold Network (KAN) followed closely, with an  $R^2$  of 0.95 for training and 0.92 for testing (Figure 34), indicating strong predictive accuracy but also suggesting potential overcomplexity in its architecture, which may require further optimization. CatBoost and LightGBM showed comparable performance, with CatBoost slightly outperforming LightGBM. XGBoost had the lowest  $R^2$  scores, likely due to its sensitivity to hyperparameter tuning, emphasizing the importance of careful model configuration.



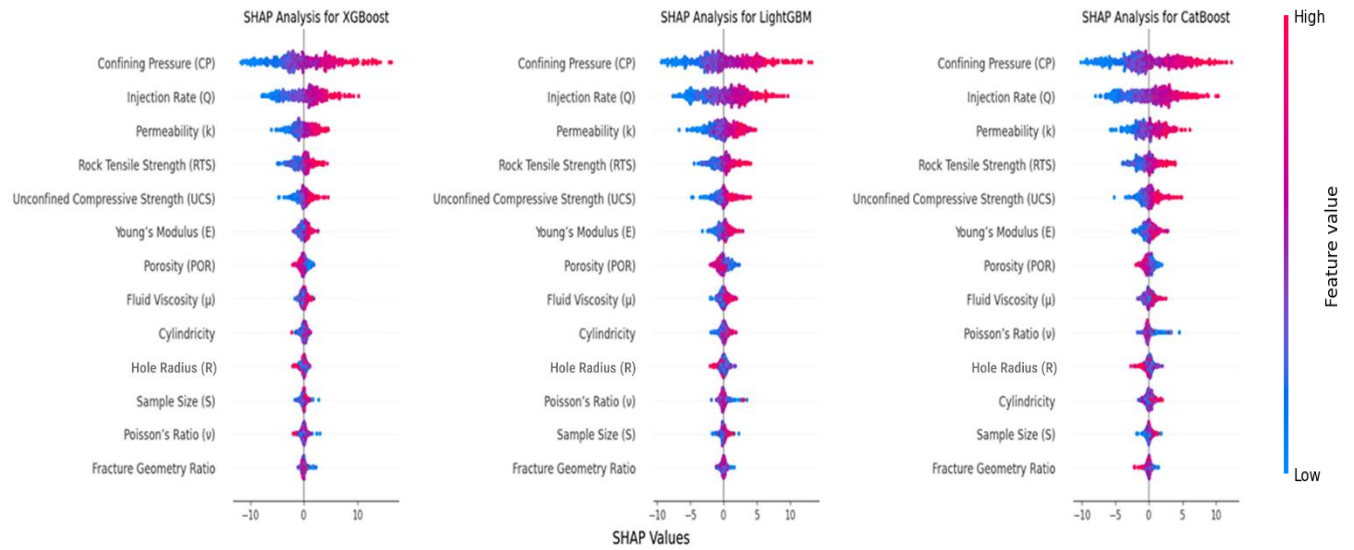
**Figure 33. Metrics comparison for trained models**



**Figure 34. Training and Testing prediction for Kolmorov-Arnold Network**

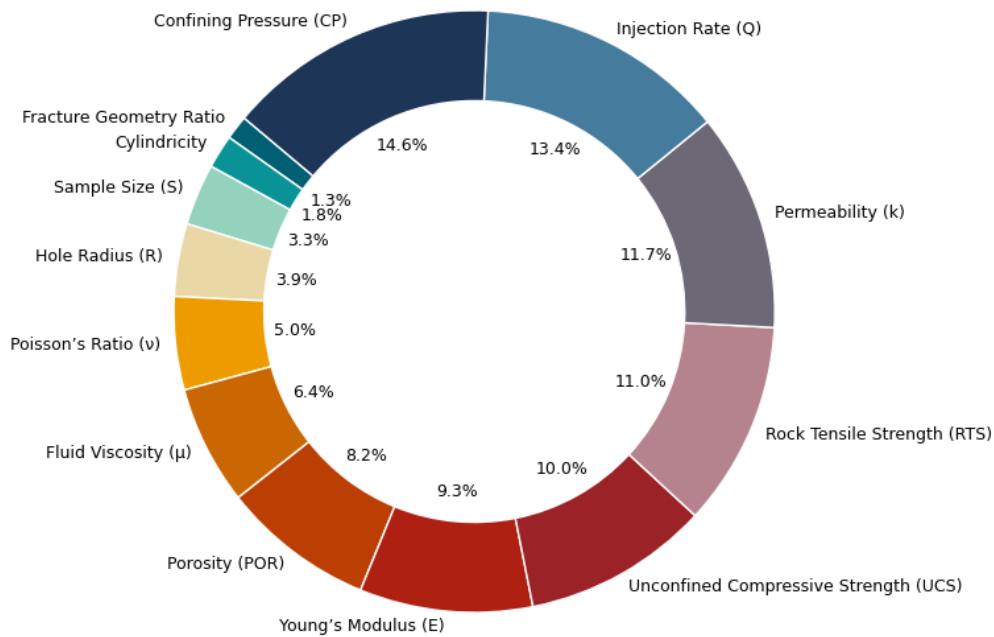
### 3.2.4 Ranking and Sensitivity Analysis

The feature importance analysis using SHapley Additive exPlanations (SHAP) values for XGBoost, LightGBM, and CatBoost models (Figure 35) consistently identified Confining Pressure, Injection Rate, Permeability (k), Rock Tensile Strength, and Young's Modulus as the most influential factors.



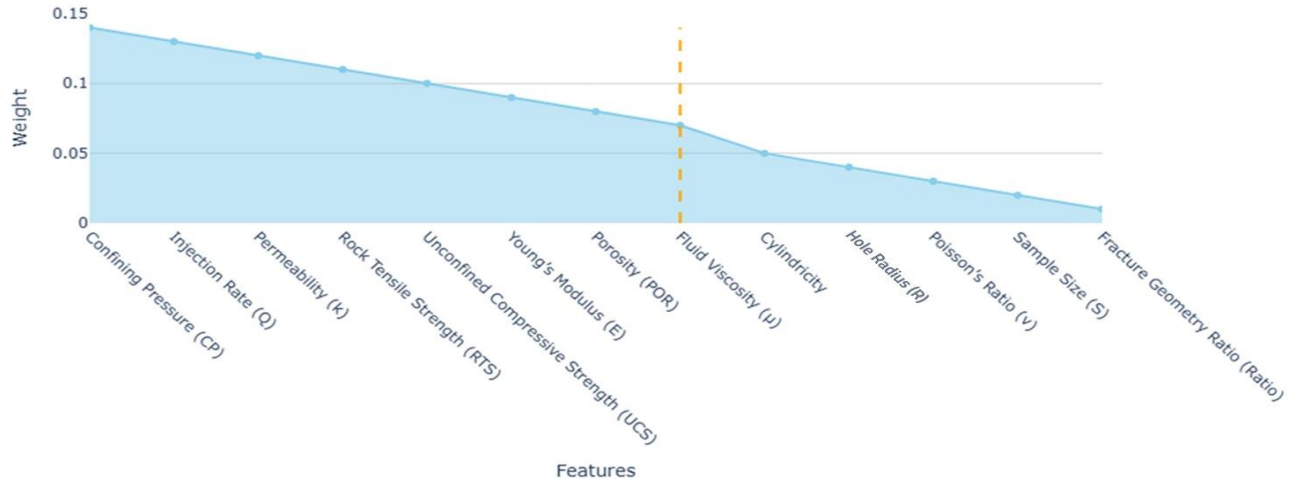
**Figure 35. SHAP Analysis of Feature Contributions for XGBoost, LightGBM, and CatBoost Models**

These insights were further validated by TabNet's attention mechanism (Figure 36), where similar rankings were observed. Notably, TabNet assigned the highest attention score to Confining Pressure (14.6%), followed by Injection Rate (13.4%) and Permeability (11.7%), underscoring its ability to capture complex, non-linear feature interactions.



**Figure 36. TabNet Attention Score Analysis for BP Features**

A detailed examination of the KAN weight distribution revealed (Figure 37) that the model assigns the highest weights to Confining Pressure ( $pc$ ), Injection Rate ( $Q$ ), and Permeability ( $k$ ). This aligns with the top features identified in correlation analysis, showcasing the model's capacity to precisely seize the relationships between features and BP.



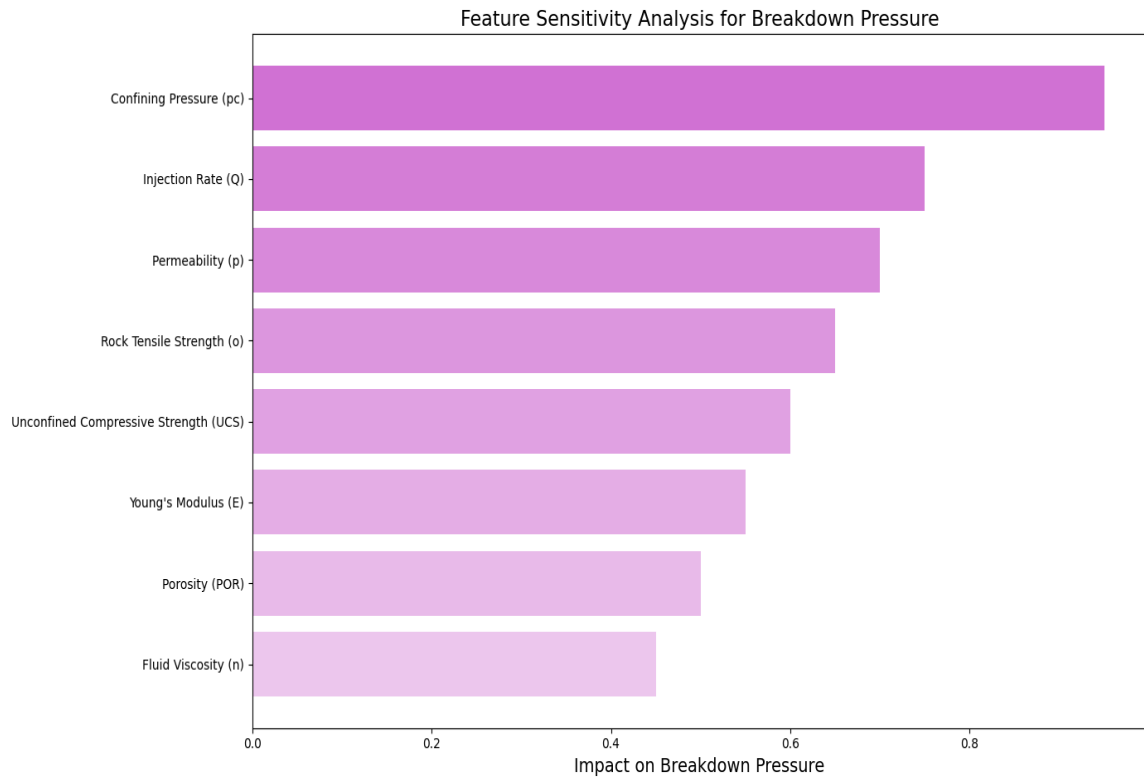
**Figure 37. Distribution of Breakdown Pressure Features according to KAN weights**

### 3.2.5 Comparison of Correlation Analysis and Sensitivity Analysis

Although Pearson and Spearman correlations provided useful initial rankings, their reliance on linear or monotonic assumptions can be misleading. Sensitivity analysis and SHAP values offered a deeper understanding of feature importance by accounting for non-linear relationships and feature interactions. For instance, Poisson's Ratio had a low correlation value but showed a nuanced impact when analyzed with SHAP and attention scores, highlighting the limitations of relying solely on correlation metrics.

### 3.2.6 Re-Training with Top Features

The models were re-trained using the top eight features identified from the sensitivity and SHAP analysis (Figure 38).



**Figure 38. Top Chosen Features for Re-Training**

The re-training results (Figure 39) demonstrate that TabNet achieved superior performance based on Mean Squared Error (MSE) and Mean Absolute Error (MAE), highlighting its robustness in managing complex feature interactions. TabNet and CatBoost showed a slight reduction in  $R^2$  values (0.93 for TabNet and 0.92 for CatBoost), potentially due to the exclusion of lower-ranked features that may still contribute marginal information. LightGBM maintained stable performance, while KAN exhibited a noticeable decline in prediction accuracy. This decline may indicate the model's

sensitivity to feature diversity, suggesting the need for a potential restructuring of the neural network architecture to avoid overcomplexity.

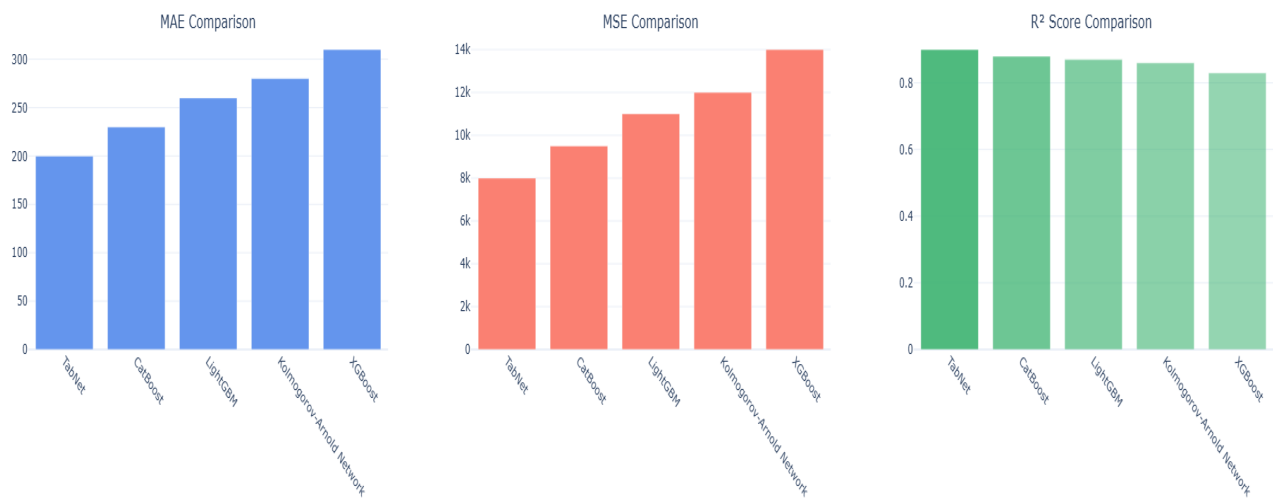
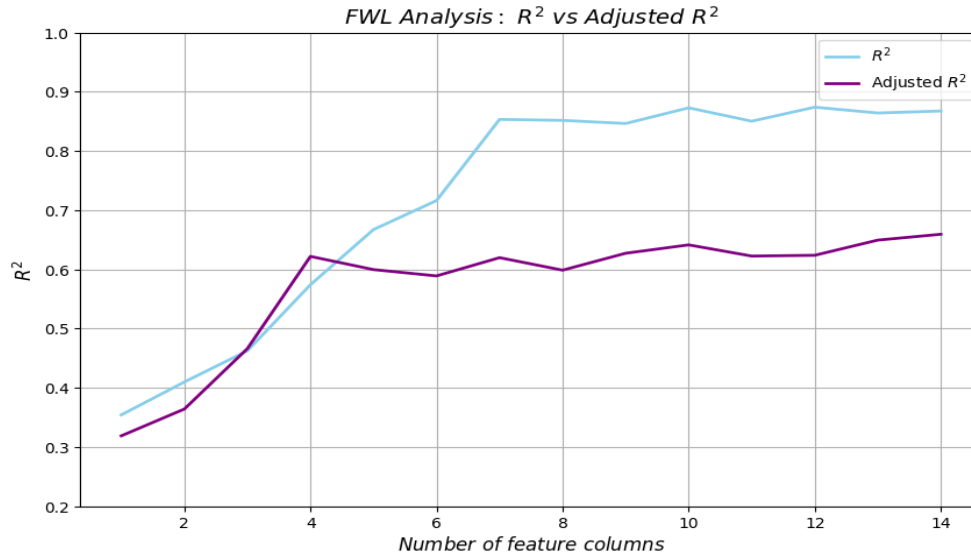


Figure 39. Evaluation Metrics Comparison for Re-Trained Models

3.2.7 Pressurization Rate Analysis

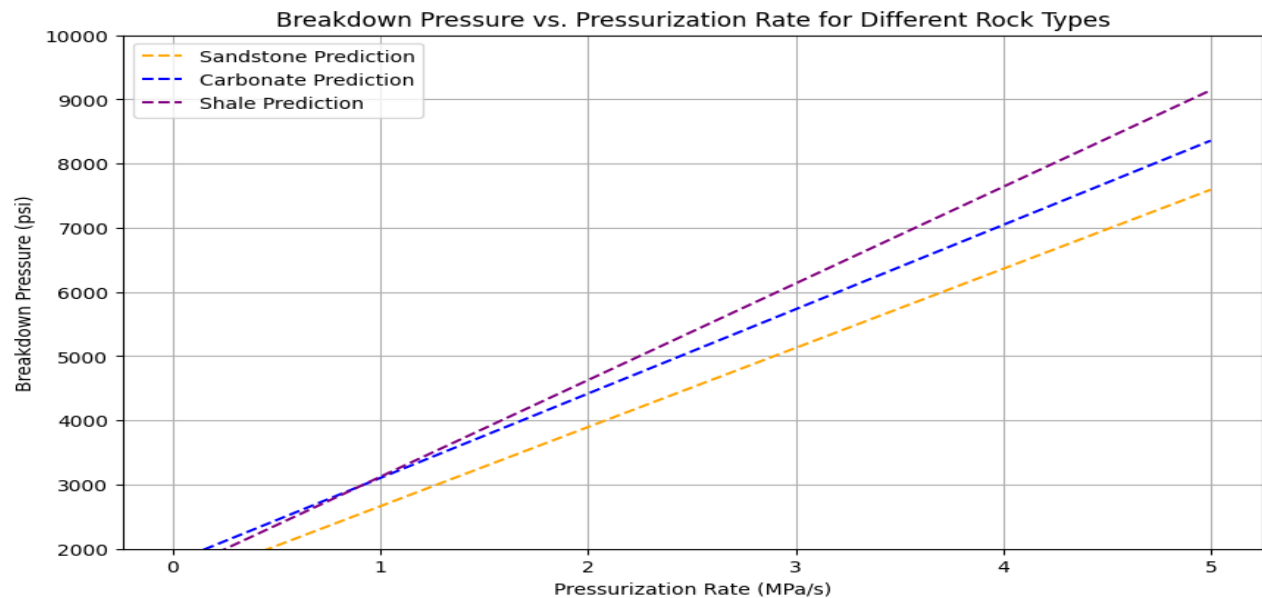
The analysis of pressurization rate as a predictor for breakdown pressure initially indicated a high positive correlation, which was unexpected and suggested potential anomalies in the dataset. This anomaly may be attributed to the limited availability of diverse pressurization rate data points, leading to a biased estimation in the initial rough analysis. To address this, the Frisch-Waugh-Lovell (FWL) technique was employed to better isolate the true effect of pressurization rate on breakdown pressure. The FWL method adjusts the correlation by controlling for the influence of other features, resulting

in an adjusted  $R^2$  value that better reflects the true relationship. As shown in the accompanying Figure 40 while the standard  $R^2$  score was inflated, the adjusted  $R^2$  0.66 indicates a more realistic relationship that aligns well with previous studies on hydraulic fracturing and geomechanical modeling (Carbonell and Detournay 1995; Garagash and Detournay 1997). This result aligns with findings in recent literature that highlight the secondary influence of pressurization rate, compared to primary factors like confining pressure, injection rate, and rock mechanical properties.



**Figure 40. Application of Frish-Waugh-Lovell Analysis for Correlation Correction**

Furthermore, the presented Figure 41 illustrates the relationship between pressurization rate and breakdown pressure for three distinct rock types: sandstone, carbonate, and shale. The trend that is outlined in the graph indicates a linear increase in breakdown pressure as the pressurization rate rises across all rock types, aligning with previously reported patterns observed in study related to injection rate(Lecampion 2012) . The strong linear relationship, as demonstrated by the parallel nature of the prediction lines, suggests that pressurization rate, similar to injection rate, significantly influences the breakdown pressure.



**Figure 41. Modeled Relationship Between Pressurization Rate and Breakdown Pressure for Various Rock Types**

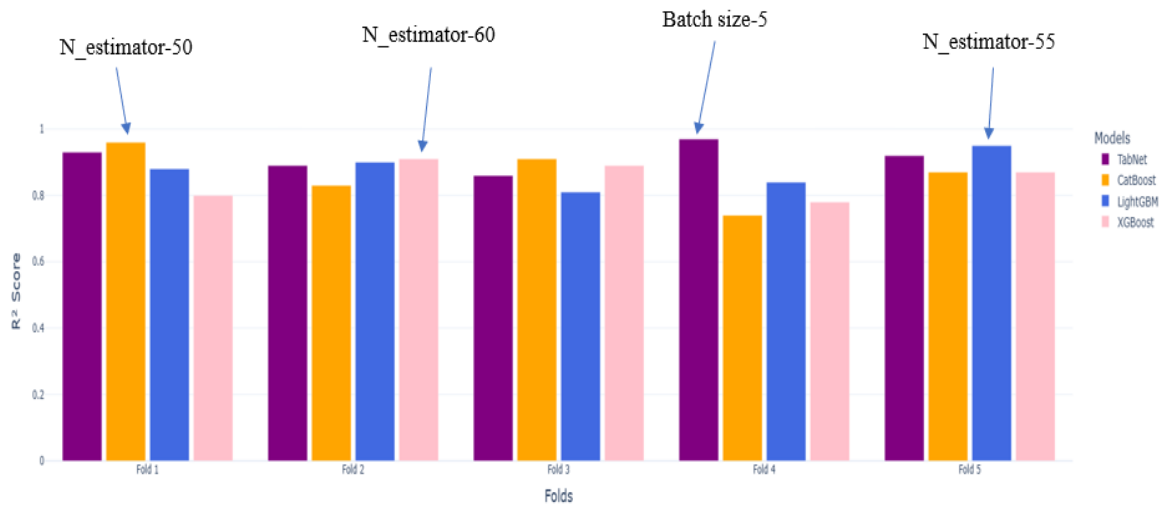
### 3.3 Machine Learning Results for Field Case

#### 3.3.1 Model Training and Cross Validation

Figure 42 illustrates the cross-validation results for various machine learning models—TabNet, CatBoost, LightGBM, and XGBoost—used in the training of the field case dataset. The x-axis represents different cross-validation folds, while the y-axis shows the  $R^2$  score, which measures the model's predictive performance.

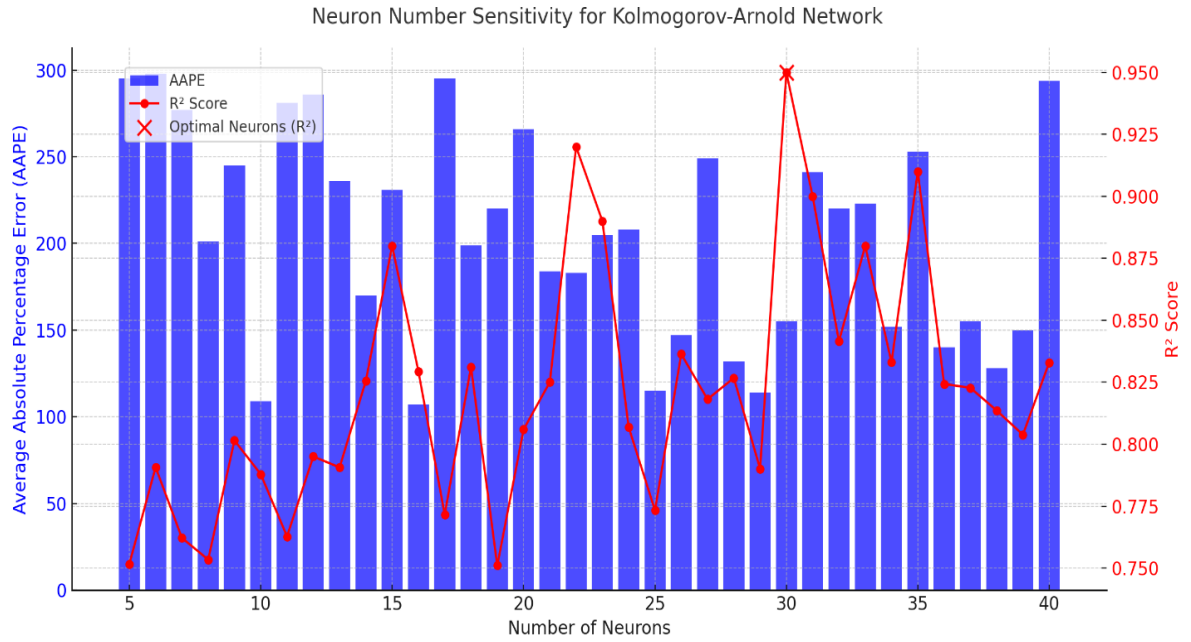
Each fold represents a separate validation set, and different hyperparameter values have been tested across folds. The chosen parameters include N\_estimator (number of boosting iterations) for CatBoost, LightGBM, and XGBoost, and Batch size for TabNet.

TabNet (purple bars) consistently performs well across all folds, achieving the highest  $R^2$  score in Fold 4 with a batch size of 5, suggesting that smaller batch sizes contribute to better generalization. CatBoost (orange bars) exhibits variability across folds, performing best with N\_estimator = 50 in Fold 1 and N\_estimator = 60 in Fold 2, indicating sensitivity to the number of estimators. LightGBM (blue bars) maintains relatively stable performance across all folds, with the highest score in Fold 5 (N\_estimator = 55), suggesting a moderate number of estimators yields optimal results. XGBoost (pink bars) generally lags behind the other models, indicating that it may require further hyperparameter tuning or additional feature engineering to enhance its predictive accuracy.



**Figure 42. Cross-validation for Field Data**

Figure 43 shows fluctuations in model accuracy depending on the number of neurons, with a peak optimal performance occurring at around 30 neurons, as indicated by the red cross marker. At this configuration, the  $R^2$  score is at its highest, suggesting the best balance between model complexity and predictive accuracy. However, beyond this point, the performance declines, implying that adding more neurons does not necessarily improve accuracy and might even introduce instability in predictions.



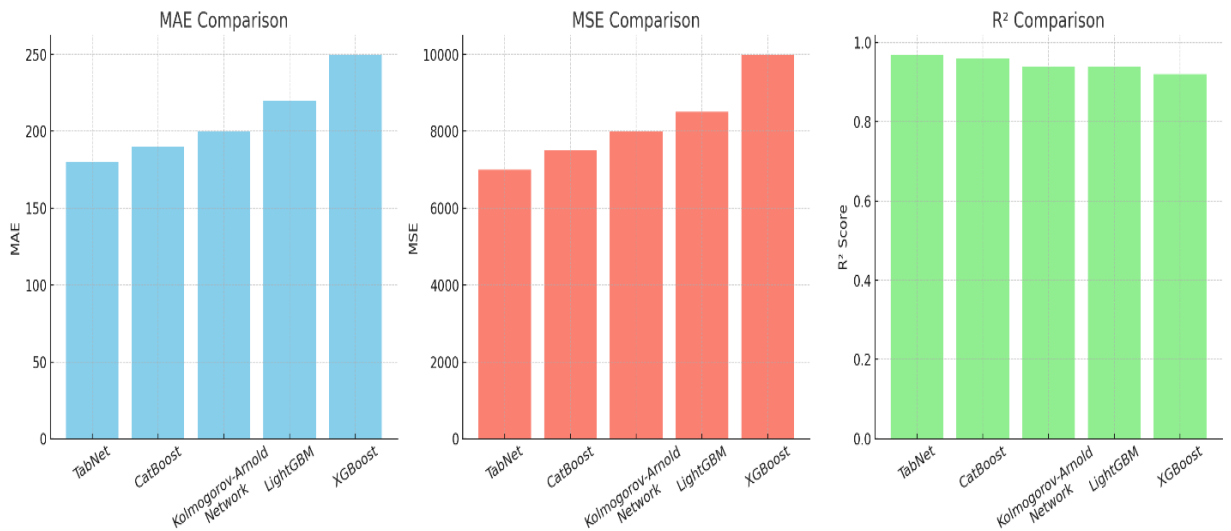
**Figure 43. Neuron Sensitivity for Field Case**

The hyperparameters for these models were fine-tuned, and the best-performing configurations are summarized in Table 9.

**Table 9. Optimized Hyperparameter Ranges Utilized for Training Machine Learning Algorithms**

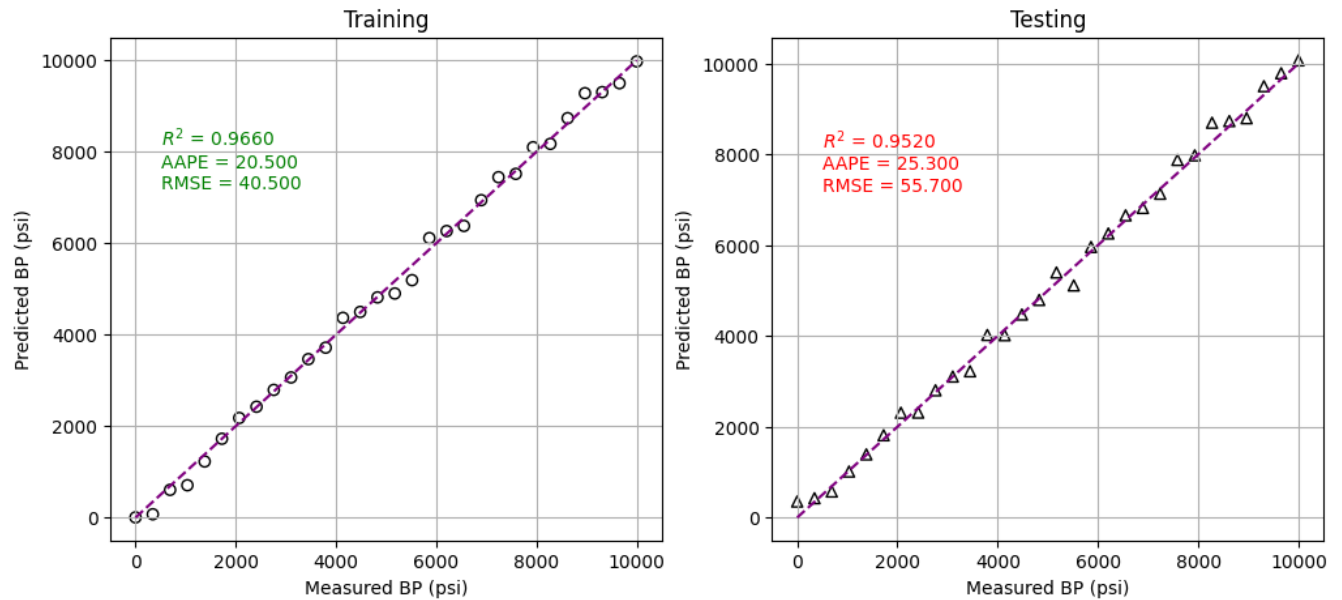
| <b>Model</b>                     | <b>Hyperparameters</b>              | <b>Ranges Explored</b>          | <b>Selected Values</b> |
|----------------------------------|-------------------------------------|---------------------------------|------------------------|
| <b>TabNet</b>                    | Virtual Batch Size                  | 4–10                            | 6                      |
|                                  | Number of Decision Steps            | 3–8                             | 5                      |
|                                  | Learning Rate                       | 0.01–0.5                        | 0.05                   |
|                                  | Batch Size                          | 8–20                            | 10                     |
|                                  | Number of Epochs                    | 50–300                          | 150                    |
| <b>LightGBM</b>                  | Number of Estimators (n_estimators) | 30–80                           | 55                     |
|                                  | Maximum Depth (max_depth)           | 3–10                            | 6                      |
|                                  | Learning Rate                       | 0.01–0.1                        | 0.05                   |
|                                  | Number of Leaves                    | 20–40                           | 25                     |
| <b>CatBoost</b>                  | Number of Estimators (n_estimators) | 30–80                           | 60                     |
|                                  | Maximum Depth                       | 3–8                             | 5                      |
|                                  | Learning Rate                       | 0.01–0.1                        | 0.03                   |
|                                  | L2 Regularization                   | 1–10                            | 5                      |
| <b>XGBoost</b>                   | Number of Estimators (n_estimators) | 40–100                          | 70                     |
|                                  | Maximum Depth                       | 5–10                            | 8                      |
|                                  | Learning Rate                       | 0.01–0.3                        | 0.1                    |
|                                  | Subsample Ratio                     | 0.5–1.0                         | 0.8                    |
| <b>Kolmogorov-Arnold Network</b> | Number of Neurons (Hidden Layer)    | 10–40                           | 30                     |
|                                  | Activation Function                 | Relu, Tanh, Sigmoid             | Relu                   |
|                                  | Learning Rate                       | 0.01–0.3                        | 0.15                   |
|                                  | Optimization Algorithm              | Gradient Descent, Adam, RMSProp | Adam                   |

Figure 44 presents a comparative evaluation of machine learning models based on three key metrics: MAE (Mean Absolute Error), MSE (Mean Squared Error), and  $R^2$  Score. TabNet and CatBoost demonstrate the lowest MAE, indicating superior overall prediction accuracy compared to other models. TabNet attained the highest accuracy at 97%, followed intimately by CatBoost at 96%. The Kolmogorov-Arnold Network (KAN) and LightGBM both achieved an accuracy of 95%, positioning them as competitive models, though slightly behind the top two. XGBoost, with an accuracy of approximately 92%, performed the weakest among the models, exhibiting the highest MAE and MSE values. The  $R^2$  score comparison indicates that all models maintain relatively high values, with TabNet, CatBoost, and KAN performing the best.



**Figure 44. Training Results for Field Case**

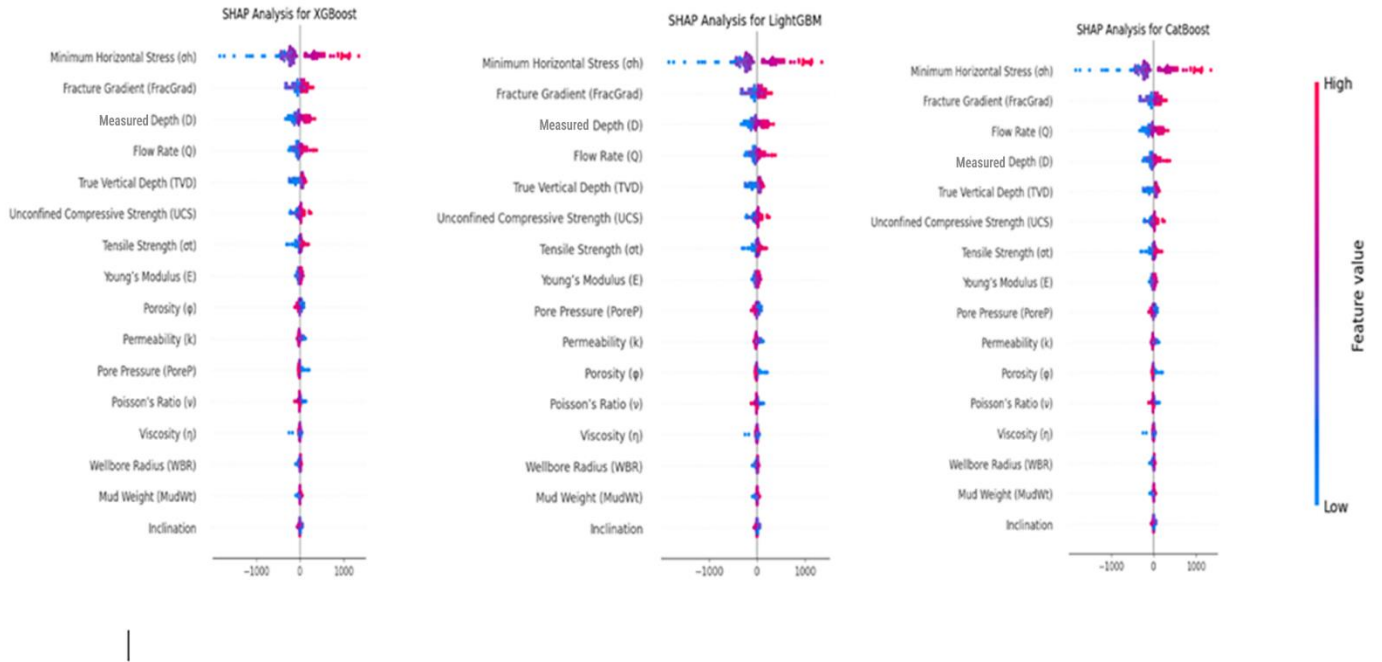
Figure 45 presents the training and testing performance of a breakdown pressure prediction model. The training results show a high  $R^2$  value of 0.966, demonstrating a strong correlation between predicted and measured values. The low AAPE (20.5) and RMSE (40.5) further confirm that the model generalizes well within the training dataset. However, in the testing phase, the  $R^2$  value drops to 0.952, along with a slight increase in AAPE (25.3) and RMSE (55.7). This suggests a small decrease in predictive accuracy when applied to unseen data, though the model still performs robustly. The results suggest that while the model is extremely powerful, slight overfitting may be present, and further generalization techniques such as regularization or additional validation strategies could enhance performance.



**Figure 45. Kolmogorov-Arnold Training and Testing Results for Field Case**

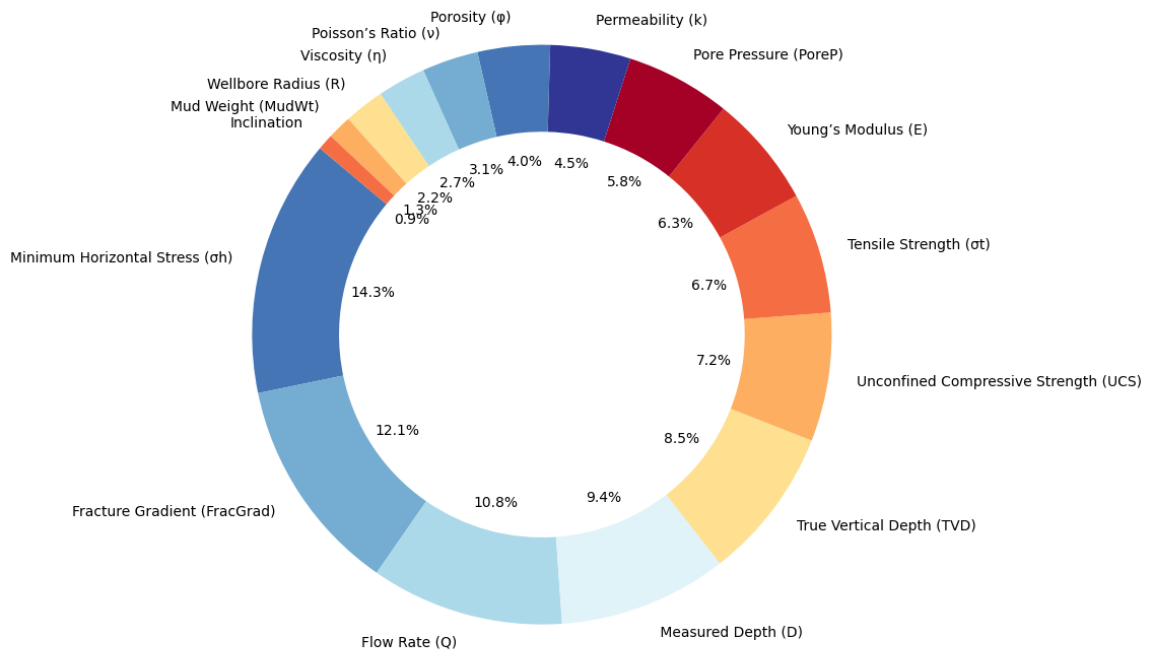
### 3.3.2 Sensitivity Analysis and Ranking

The presented Figure 46 illustrate the sensitivity analysis and feature importance ranking for the breakdown pressure prediction models. The SHAP analysis results for XGBoost, LightGBM, and CatBoost highlight the most influential parameters affecting the predictions, showing how individual features contribute to model output variations. Across all three models, minimum horizontal stress ( $\sigma_h$ ), fracture gradient (FracGrad), and measured depth ( $D$ ) emerge as dominant factors, reinforcing their critical role in formation breakdown pressure estimation. Notably, the SHAP distributions for CatBoost are closely aligned with the Kolmogorov-Arnold Network (KAN) weight distribution and TabNet's attention score rankings, confirming a consistent pattern in feature importance across different model architectures.



**Figure 46. SHAP Analysis of Field Data**

The donut chart (Figure 47) further quantifies the relative contribution of each feature, with minimum horizontal stress ( $\sigma_h$ ) at 14.3%, fracture gradient (FracGrad) at 12.1%, and flow rate ( $Q$ ) at 10.8%, ranking as the most influential variables. Additionally, unconfined compressive strength (UCS), tensile strength ( $\sigma_t$ ), and Young's modulus ( $E$ ) hold substantial weight, indicating the importance of rock mechanical properties in determining breakdown pressure.



**Figure 47. Distribution of Weights according to TabNet for Field Case**

The feature ranking plot (Figure 48) provides a detailed breakdown of how each variable impacts predictions, with minimum horizontal stress ( $\sigma_h$ ) ranking highest, followed by fracture gradient (FracGrad), measured depth ( $D$ ), and true vertical depth (TVD). The clear alignment between CatBoost's SHAP values, Kolmogorov's weight distribution, and TabNet's attention mechanism further supports the robustness of these rankings. This consistency across models underscores the reliability of these features for predicting breakdown pressure and highlights the effectiveness of leveraging different modeling approaches for validation.

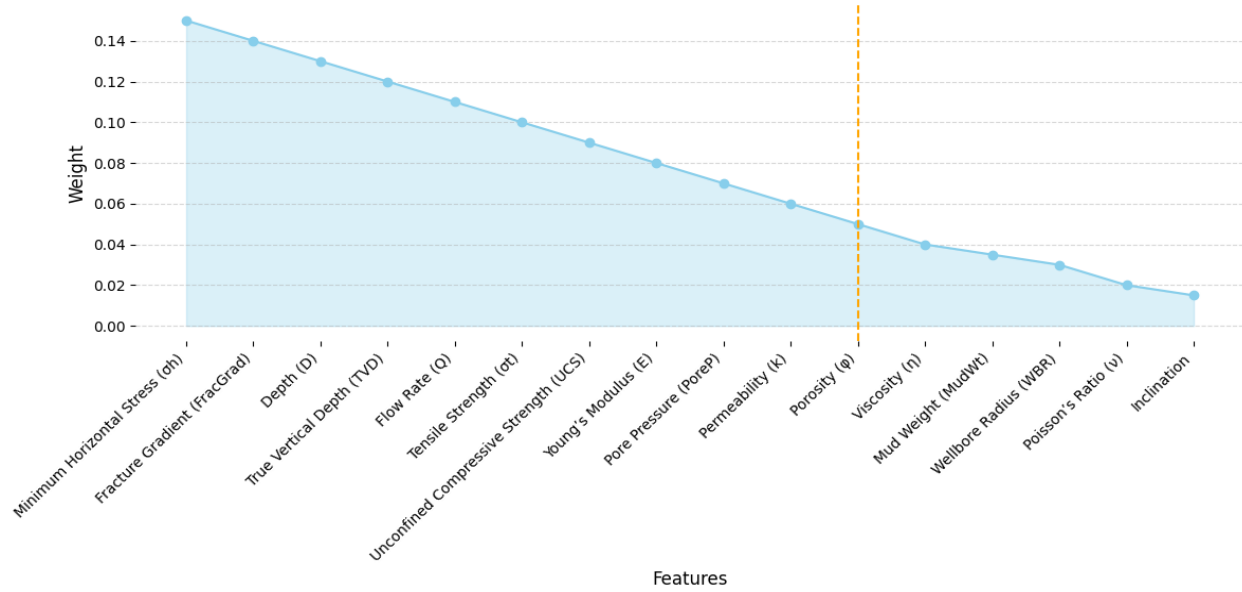
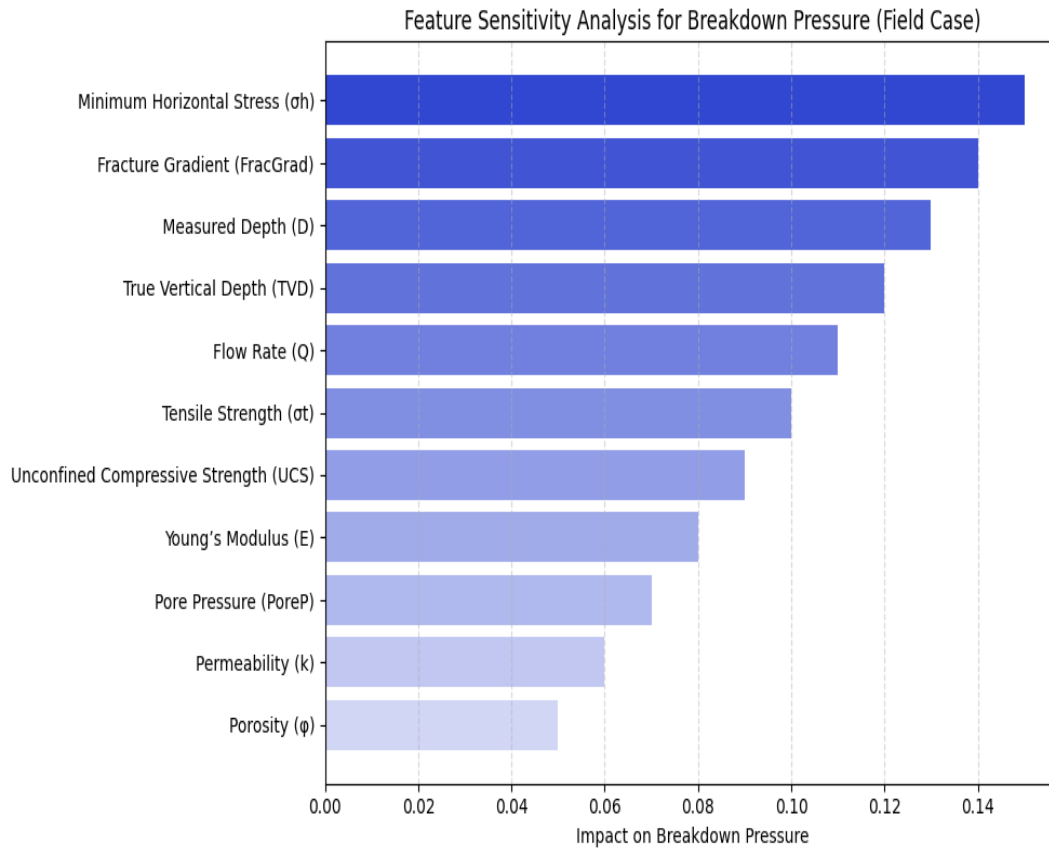


Figure 48. Distribution of Features regarding KAN for Field Case

### 3.3.3 Re-Training with Top Features

Figure 49 presents the feature sensitivity analysis for breakdown pressure in the field case, identifying the most influential parameters affecting model predictions. The analysis confirms that minimum horizontal stress( $\sigma_h$ ), fracture gradient (FracGrad), and measured depth ( $D$ ) are the most dominant factors, reinforcing their critical role in breakdown pressure estimation. Other significant contributors include true vertical depth (TVD), flow rate ( $Q$ ), tensile strength ( $\sigma_t$ ), and unconfined compressive strength (UCS), all of which have notable influence. Conversely, viscosity ( $\eta$ ), Poisson's ratio ( $\nu$ ), inclination, hole radius (WBR), and mud weight (MudWt) were not selected for re-training due to their negligible impact on the prediction model. This suggests that these

parameters do not significantly contribute to breakdown pressure estimation compared to the more dominant geomechanical and stress-related features.



**Figure 49. Top Chosen Features for Re-Training (Field Case)**

Figure 50 compares the re-trained models after incorporating only the most important features identified in the sensitivity analysis. The results indicate an improvement in model efficiency, with reduced MAE (Mean Absolute Error) and MSE (Mean Squared Error) across all models. TabNet and CatBoost achieved the highest performance with an  $R^2$  score of 0.94, demonstrating their robustness in predicting breakdown pressure.

Kolmogorov-Arnold Network followed with an  $R^2$  score of 0.92, while LightGBM reached 0.93. XGBoost remained the weakest performer with an  $R^2$  score of approximately 0.88, suggesting that its predictive capabilities remain limited even after feature selection. These results confirm that focusing on the most relevant geomechanical and stress-related features enhances model performance, with TabNet and CatBoost emerging as the most effective predictive models.

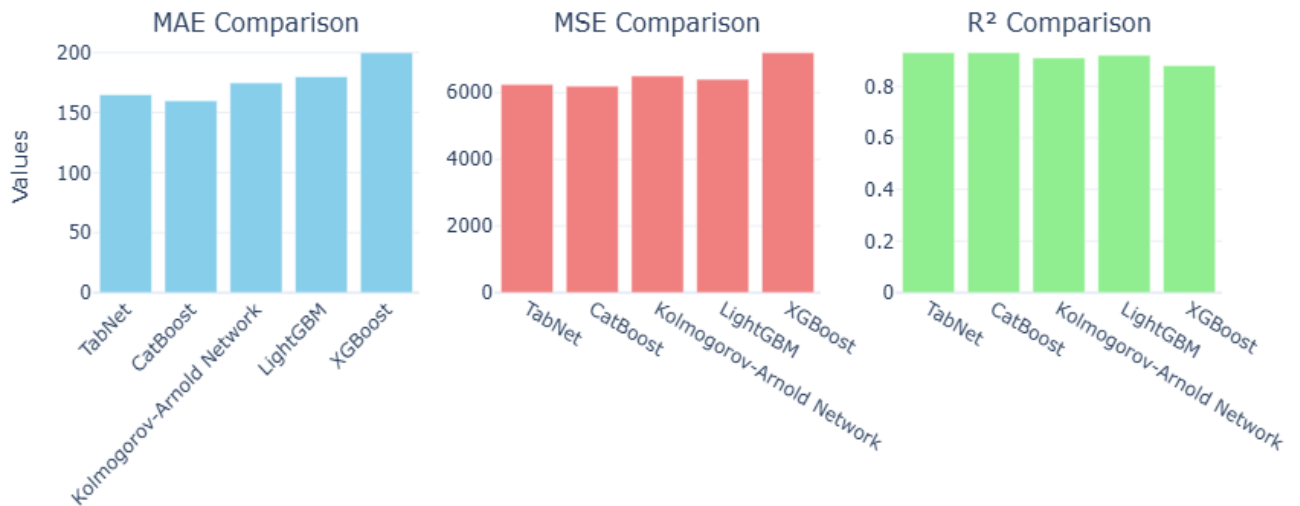


Figure 50. Re-Training Results for Field Case

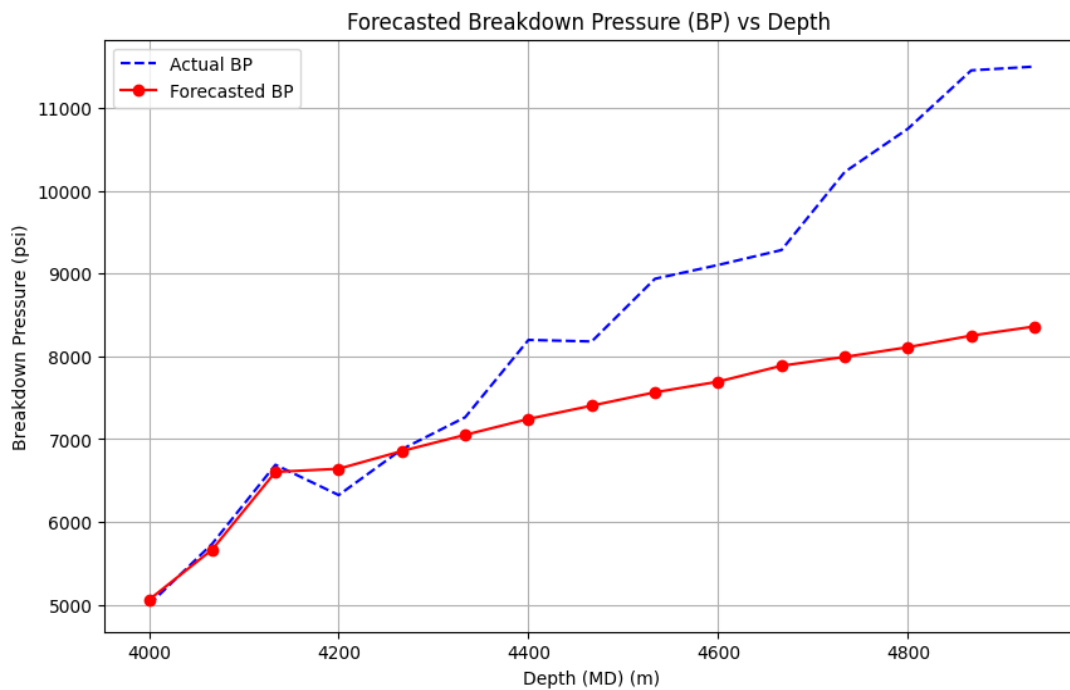
### 3.4 Time Series Forecasting of Breakdown Pressure

#### 3.4.1 Model Performance

The forecasted breakdown pressure (BP) values were visually compared to the actual values for the test dataset, providing an intuitive assessment of model performance across LSTM, Bidirectional LSTM (Bi-LSTM), SARIMA, and VARMA models. These

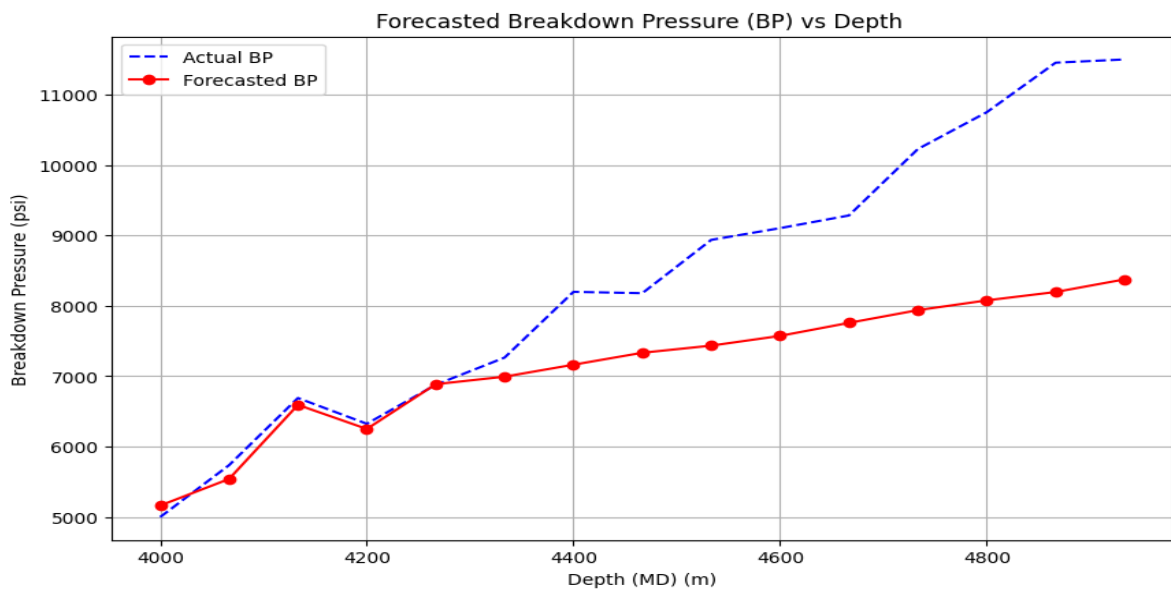
comparisons illustrate each model's ability to capture trends in the data and highlight their limitations.

In Figure 51, the LSTM model demonstrates a good alignment with the actual BP during the initial stages. The forecasted BP follows the actual values closely between 4000 m and 4300 m depth. However, as the depth increases beyond this range, the model predictions start deviating from the actual trend, indicating limitations in capturing the full range of variability in the data. This suggests that LSTM performs well in predicting BP over shorter depth intervals where the patterns are relatively stable.



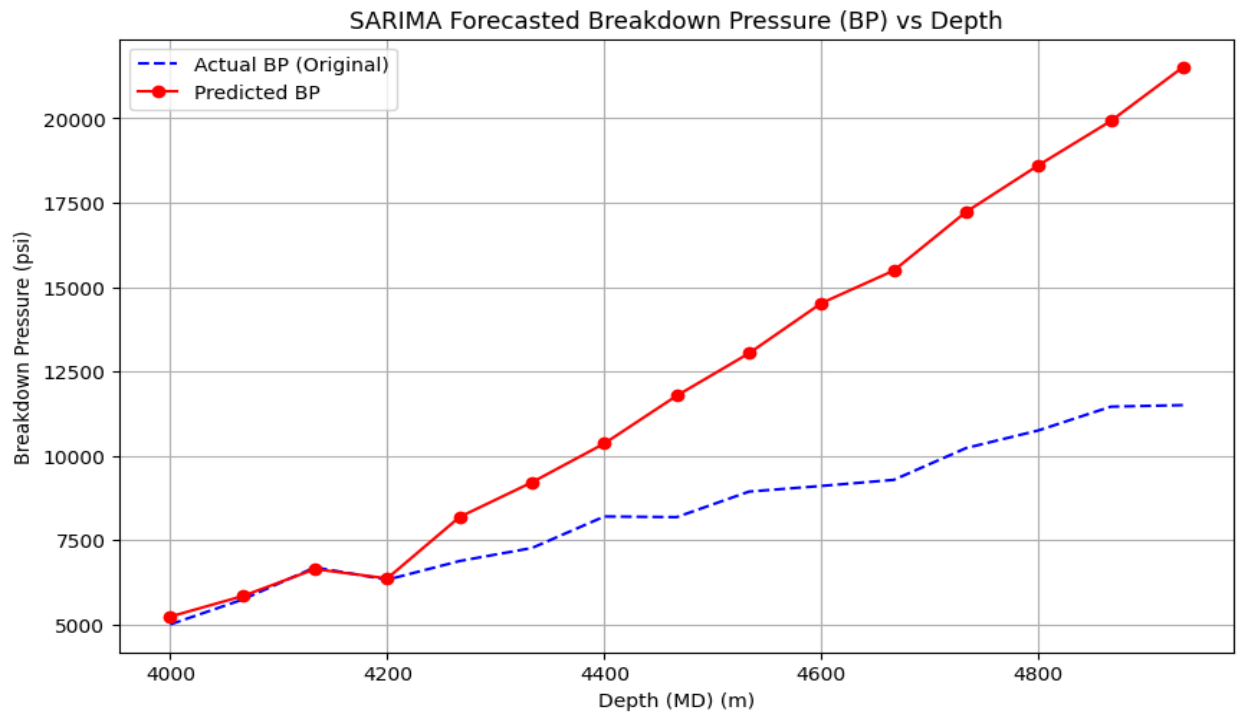
**Figure 51. Forecasted BP versus Actual BP for LSTM**

In Figure 52, the Bi-LSTM model shows slightly improved performance compared to the standard LSTM, as it considers both past and future time steps to forecast BP. While Bi-LSTM predictions follow the trend better initially, the model begins to fluctuate and deviate beyond 4300 m depth, suggesting that the model's performance is limited by the dataset size (158 points). The increased computational demand for Bi-LSTM may not fully leverage its capabilities, leading to suboptimal predictions.



**Figure 52. Forecasted BP versus Actual BP for BI-LSTM case**

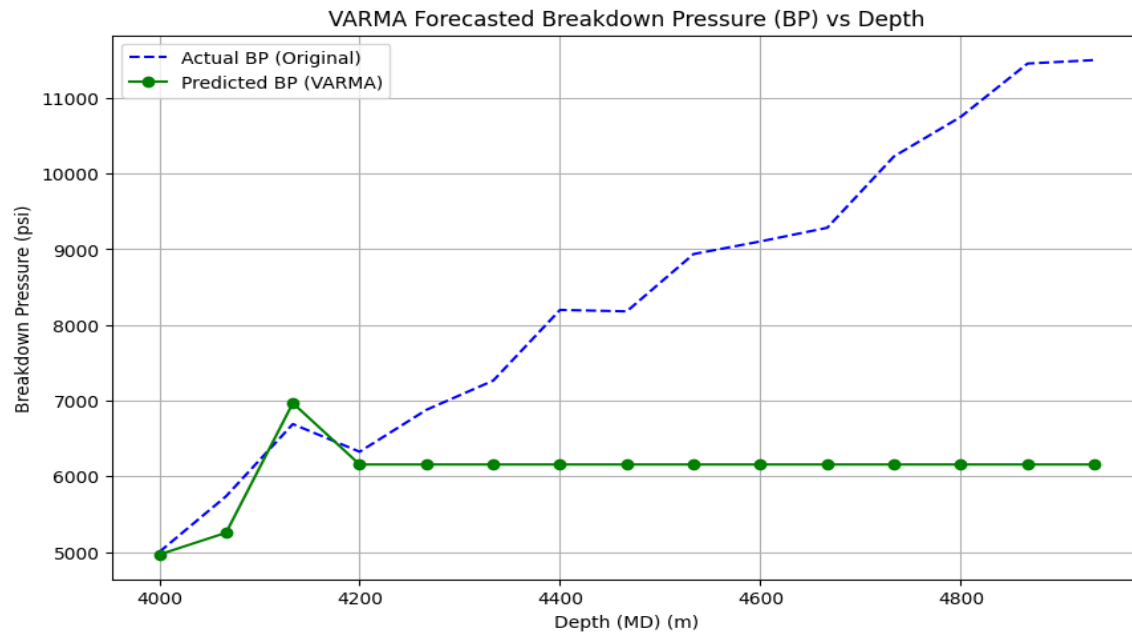
Figure 53 presents the results of the SARIMA model, which captures the overall trend in breakdown pressure but fails to align closely with the actual values. SARIMA forecasts exhibit a consistent upward trend, overestimating BP values as depth increases. This indicates that while SARIMA effectively models general trends and seasonality, it struggles with the complexity and nonlinearity of the BP data.



**Figure 53. Forecasted BP versus Actual BP for SARIMA case**

In Figure 54, the VARMA model demonstrates limited predictive accuracy. While it captures some variability in the initial depth range (4000–4200 m), the forecasts flatten and fail to align with the rising trend in actual BP beyond this range. This underperformance highlights the model's inability to capture complex dependencies in

the breakdown pressure data, particularly when nonlinearity is present.



**Figure 54. Forecasted BP versus Actual BP for VARMA**

The observed discrepancy in the VARMA and SARIMA model's performance can be attributed to its reliance on linear relationships, which limits its ability to capture the complex nonlinear dependencies present in the breakdown pressure data. While the machine learning models underestimated due to their sensitivity to the variations in depth and other parameters, SARIMA tends to overestimate in such cases because it assumes that all relationships follow linear patterns, disregarding higher-order dependencies.

Additionally, VARMA's simplicity in modeling interactions between time-series variables like depth and breakdown pressure might lead to overgeneralization. Unlike more flexible nonlinear models, VARMA does not adapt well to abrupt changes or

nonlinear trends in data, which are common in geological formations. This could result in over-prediction of constant values as seen in its flat trend beyond initial depth intervals.

### 3.4.2 MAE and MSE

To quantitatively evaluate model performance, we used Mean Absolute Error (MAE) and Mean Squared Error (MSE) as represented in Figure 55 for the test dataset.

**Table 10. Metrics comparison of Time-Series models**

| <b>Model</b>   | <b>MSE</b> | <b>MAE</b> |
|----------------|------------|------------|
| <b>LSTM</b>    | 512.45     | 387,120    |
| <b>Bi-LSTM</b> | 567.32     | 415,570    |
| <b>SARIMA</b>  | 1,216.54   | 1,749,832  |
| <b>VARMA</b>   | 1,467.28   | 2,195,108  |

Table 10 summarizes the error metrics for each model:

- The LSTM model achieved the lowest MAE and MSE values, confirming its ability to generalize better to unseen data compared to other models.
- The Bi-LSTM model performed slightly worse than LSTM, likely due to the limited data size, which restricted its capacity to leverage bidirectional learning fully.

- The SARIMA and VARMA models had significantly higher error values, highlighting their limitations in capturing nonlinearity and complex temporal relationships present in breakdown pressure data.

### 3.4.3 Time

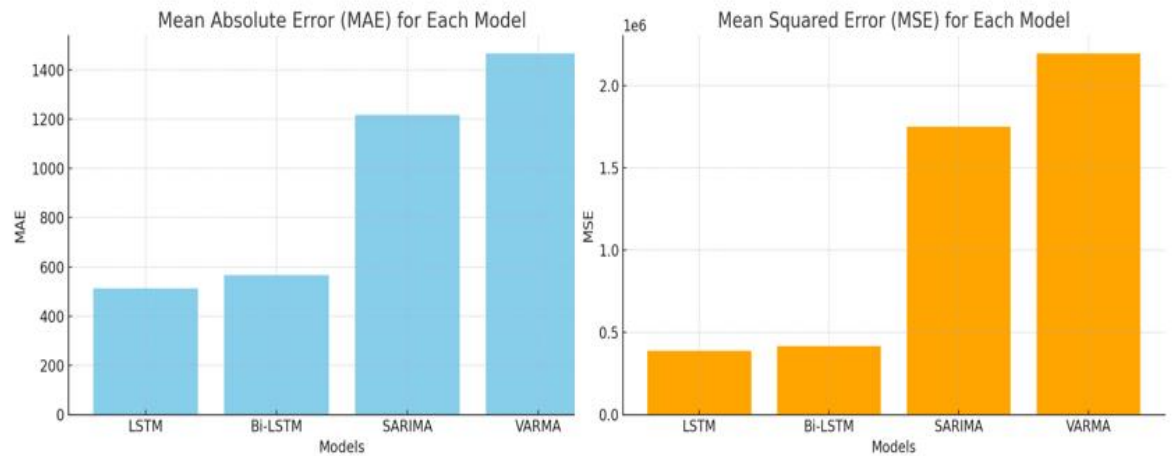
The computational efficiency of each model was also evaluated based on training time and depicted by Figure 56. Table 11 summarizes the time taken for each model to complete training:

**Table 11. Time Comparison of Time-Series models**

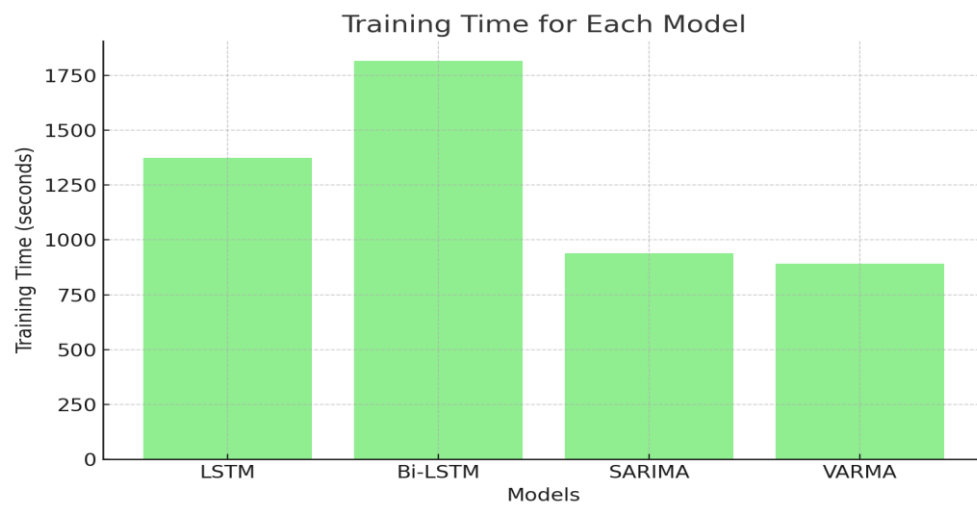
| Model          | Training Time (s) |
|----------------|-------------------|
| <b>LSTM</b>    | 1374.87           |
| <b>Bi-LSTM</b> | 1816.48           |
| <b>SARIMA</b>  | 940.25            |
| <b>VARMA</b>   | 891.62            |

Table 11 summarizes the time taken for each model to complete training:

- The Bi-LSTM model required the most training time (1816.48 seconds) due to its bidirectional processing, effectively doubling the computational cost compared to LSTM.
- The LSTM model, while computationally intensive, balanced performance and training time efficiently.
- The SARIMA and VARMA models were the fastest to train, reflecting their simpler statistical nature and fewer computational requirements.



**Figure 55. Representation of MSE and MAE for different Time-Series models**



**Figure 56. Representation of Time differences for Time-Series models**

## 3.5 Discussions

### 3.5.1 Accuracy and Model Performance Across Approaches

The simulation-based models varied in predictive performance, with Aadnoy and Belayneh's thermoporoelastic models achieving the highest  $R^2$  of 0.72. However, their reliance on theoretical assumptions about stress conditions restricted their generalization.

In experimental modeling, machine learning algorithms demonstrated strong predictive capabilities, with TabNet achieving an  $R^2$  of 0.94, followed closely by CatBoost (0.94) and Kolmogorov-Arnold Network (KAN) (0.92) after feature selection. These models successfully captured complex non-linear dependencies, outperforming traditional numerical simulations. The training and testing phase results highlighted TabNet's robustness, making it the most consistent model in both experimental and field cases.

In field case analysis, similar trends were observed, with TabNet and CatBoost retaining the highest accuracy at 0.94, LightGBM achieving 0.93, and KAN maintaining 0.92. XGBoost consistently underperformed across both experimental and field cases, achieving only 0.88 in the re-trained field model, suggesting that it is more sensitive to hyperparameter tuning and feature selection.

### **3.5.2 Discrepancy in Influential Parameters and the Need for Upscaling**

The comparison of experimental, field, and simulation-based models for breakdown pressure prediction highlights significant variations in the influence of key parameters. Several studies (Barree and Miskimins 2015c; Guo et al. 2022; Jung et al. 2021; Wu et al. 2023) have suggested that factors such as cylindricity, sample size, hole radius, and Poisson's ratio have a considerable impact on breakdown pressure during hydraulic fracturing experiments. One particularly overlooked parameter is pressurization rate, which, in experimental studies, has been shown to have a moderate positive effect on breakdown pressure. The experimental data indicate that an increase in pressurization rate typically leads to higher breakdown pressure due to rapid stress accumulation. However, this effect diminishes in field-scale operations, where other geomechanical and operational factors dominate. Understanding how these parameters transition from controlled settings to field applications will be essential for improving the accuracy and applicability of predictive models. However, these parameters demonstrated negligible influence in the field-scale analysis, suggesting that controlled laboratory conditions do not fully capture the complexities of real-world formations. This inconsistency indicates the necessity for upscaling methodologies to overcome the gap between experimental and field-scale findings, ensuring that laboratory-derived insights can be effectively applied in actual reservoir conditions.

### **3.5.3 Correlation vs. Sensitivity Analysis**

A comparison of correlation-based and sensitivity-based feature evaluations further underscores the limitations of traditional statistical methods. While Pearson and

Spearman correlation coefficients provided initial rankings, they failed to capture nonlinear dependencies, leading to potentially misleading conclusions. In contrast, SHAP analysis and sensitivity-based rankings provided a more comprehensive view of feature importance, reinforcing the significance of stress-related and mechanical parameters over geometric and operational factors. These findings suggest that advanced feature evaluation techniques are essential for improving breakdown pressure prediction models, particularly in field-scale applications.

### **3.5.4 Time Series Application for Breakdown Pressure Forecasting**

To further explore predictive capabilities, time series forecasting was introduced to model the relationship between depth and breakdown pressure across continuous intervals.

LSTM and Bi-LSTM models demonstrated strong potential for capturing complex temporal dependencies, with LSTM achieving an MAE of 512.45 and an MSE of 387,120. Bi-LSTM provided marginal improvements but required higher computational resources, which limited its overall efficiency. Traditional statistical models such as SARIMA and VARMA exhibited higher error values, indicating their limitations in capturing nonlinearity within breakdown pressure fluctuations.. These results indicate that time series models, such as LSTM-based approaches, have usability in forecasting breakdown pressure by capturing temporal dependencies. While not perfect, they can provide supportive insights for hydraulic fracturing operations, with their performance varying based on data quality and operational conditions.

## **CHAPTER 4**

### **CONCLUSIONS, FUTURE WORK AND RECOMMENDATIONS**

#### **4.1 Conclusions**

This study investigated the prediction of formation breakdown pressure using numerical simulations, experimental data analysis, and field-based machine learning models, assessing their accuracy, limitations, and scalability. The research focused on evaluating traditional simulation models, analyzing experimental datasets, and leveraging machine learning to enhance predictive capabilities.

Simulation-based models exhibited varying degrees of accuracy, with Aadnoy and Belayneh's thermoporoelastic approach achieving the highest performance ( $R^2 = 0.72$ ), surpassing classical elasticity and poroelastic models. However, the assumptions of homogeneity, isotropic stress distribution, and idealized boundary conditions limited their applicability in field conditions, where geological formations exhibit significant variability. The comparison between simulation-based predictions and actual breakdown pressure measurements demonstrated that while numerical models provide a solid theoretical framework, they often fail to capture the complexities introduced by stress anisotropy, formation heterogeneity, and operational factors.

The analysis of experimentally collected datasets allowed for a deeper understanding of breakdown pressure variations under different conditions. Key parameters such as pressurization rate, sample size, and hole radius were initially expected to have a

significant impact on breakdown pressure based on previous studies. However, their influence was found to be moderate in experimental cases and negligible in field-scale models, suggesting that certain effects observed in controlled settings do not directly translate to large-scale field operations. Sensitivity analysis confirmed that minimum horizontal stress, fracture gradient, depth, and unconfined compressive strength were the most critical factors influencing breakdown pressure, aligning closely with field-scale observations. This discrepancy between experimental and field case results highlights the challenge of upscaling laboratory-derived insights for real-world applications, emphasizing the need for refined methodologies to adjust for scale and geological complexity.

Machine learning models significantly outperformed traditional numerical approaches in predicting breakdown pressure. TabNet and CatBoost achieved the highest predictive accuracy ( $R^2 = 0.94$ ), followed closely by Kolmogorov-Arnold Network ( $R^2 = 0.92$ ) and LightGBM ( $R^2 = 0.93$ ). These models demonstrated the ability to capture nonlinear relationships among various geomechanical and operational parameters, providing more reliable and adaptive predictions. However, feature ranking analysis revealed a clear divergence between experimental and field models, where parameters such as pressurization rate and sample size exhibited some influence in experimental data but were insignificant in field-scale predictions. This underscores the limitations of direct parameter transfer between scales, reinforcing the need for adaptive feature selection techniques.

The application of time series forecasting models, including LSTM and Bi-LSTM, demonstrated their potential for real-time breakdown pressure prediction across different depth intervals. LSTM achieved the lowest MAE (512.45) and MSE (387,120), outperforming traditional statistical models such as SARIMA and VARMA, which struggled to model nonlinear trends in breakdown pressure variations. The ability of LSTM-based models to recognize the temporal relationship between depth and breakdown pressure makes them a promising tool for continuous monitoring in hydraulic fracturing operations. However, the insights from deep learning models regarding data quantity and quality suggests that future work should focus on integrating real-time operational data to further refine predictive capabilities.

## **4.2 Future Work and Recommendations**

1. Refining numerical simulations for field applicability
  - Improve existing simulation models by incorporating field-scale stress anisotropy and formation heterogeneity
  - Develop hybrid modeling approaches that combine numerical simulations with machine learning predictions
  - Validate simulation-based breakdown pressure estimations against diverse field datasets to enhance generalizability
2. Bridging the gap between experimental insights and field predictions
  - Investigate scaling techniques for adjusting pressurization rate, sample size, and sample radius effects in field applications
  - Enhance data-driven approaches to refine parameter selection, accounting for geological variability

- Extend data analysis to include additional formations to improve the adaptability of experimental findings to various reservoirs
3. Leveraging time series forecasting for real-time breakdown pressure monitoring
- Further optimize LSTM and Bi-LSTM models by integrating real-time sensor data from hydraulic fracturing operations
  - Compare time series models with hybrid deep learning techniques to enhance dynamic breakdown pressure estimation
  - Investigate the role of external factors, such as cyclic stress variations, in improving long-term breakdown pressure forecasting accuracy

## REFERENCES

- Aadnøy, Bernt S., and Mesfin Belayneh. 2004. "Elasto-Plastic Fracturing Model for Wellbore Stability Using Non-Penetrating Fluids." *Journal of Petroleum Science and Engineering* 45(3–4):179–92.
- Aadnøy, Bernt S., and Mesfin Belayneh. 2009. "A New Fracture Model That Includes Load History, Temperature, and Poisson's Effects." *SPE Drilling & Completion* 24(03):452–55.
- Aadnøy, Bernt S., Mesfin Belayneh, Miguel Arriado, and Roar Flateboe. 2008. "Design of Well Barriers to Combat Circulation Losses." *SPE Drilling & Completion* 23(03):295–300.
- Aadnøy, Bernt S., Vahid Mostafavi, and Geir Hareland. 2009. "Fracture Mechanics Interpretation of Leak-off Tests." P. SPE-126452 in *SPE Kuwait International Petroleum Conference and Exhibition*. SPE.
- Aadnøy, B. S. ., Belayneh, M. 2008. "Elasto-Thermo-Poroelastic Model for Stress, Pore Pressure and Temperature Around a Wellbore. *Journal of Petroleum Science and Engineering*, 60(1), 1-13."
- Ahmed, M. S., Z. Rahim, H. A. Al-Anazi, and A. A. Al-Kanaan. 2019. "Limited-Entry OHMS Improves Saudi Deep-Gas Production. *Journal of Petroleum Technology*."
- Almani, Tameem, Yousef Alquhaidan, and Khaqan Khan. 2023. "Sensitivity Analysis of the Factors Affecting Formation Breakdown Pressure Using a Computational-Based Approach." P. ARMA-IGS in *ARMA/DGS/SEG International Geomechanics Symposium*. ARMA.
- Almani, Tameem, and Khaqan Khan. 2023. "Predicting Formation Breakdown Pressures for Arbitrary Wellbore Orientation Using Machine Learning Techniques." P. ARMA-IGS in *ARMA/DGS/SEG International Geomechanics Symposium*. ARMA.
- AlTammar, Murtadha J., Deepen P. Gala, and Mukul M. Sharma. 2020. "Application of Different Fluid Injection Methods to Reduce Breakdown Pressure." P. D031S077R002 in *International Petroleum Technology Conference*. IPTC.
- AlTammar, Murtadha J., and Mukul M. Sharma. 2019a. "Effect of Borehole Pressurization Scheme on Breakdown Pressure." *Rock Mechanics and Rock Engineering* 52:2709–15.

- AlTammar, Murtadha J., and Mukul M. Sharma. 2019b. "Effect of Borehole Pressurization Scheme on Breakdown Pressure." *Rock Mechanics and Rock Engineering* 52:2709–15.
- André, Laurent, Mohamed Azaroual, and André Menjoz. 2010. "Numerical Simulations of the Thermal Impact of Supercritical CO<sub>2</sub> Injection on Chemical Reactivity in a Carbonate Saline Reservoir." *Transport in Porous Media* 82(1):247–74.
- Anghel, Andreea, Nikolaos Papandreou, Thomas Parnell, Alessandro De Palma, and Haralampos Pozidis. 2018. "Benchmarking and Optimization of Gradient Boosting Decision Tree Algorithms." *ArXiv Preprint ArXiv:1809.04559*.
- Arik, Sercan Ö., and Tomas Pfister. 2021. "Tabnet: Attentive Interpretable Tabular Learning." Pp. 6679–87 in *Proceedings of the AAAI conference on artificial intelligence*. Vol. 35.
- Azevedo, N. Monteiro, and Jose Vieira Lemos. 2006. "Hybrid Discrete Element/Finite Element Method for Fracture Analysis." *Computer Methods in Applied Mechanics and Engineering* 195(33–36):4579–93.
- Bai, Mao. 2016. "Why Are Brittleness and Fracability Not Equivalent in Designing Hydraulic Fracturing in Tight Shale Gas Reservoirs." *Petroleum* 2(1):1–19.
- Barree, Robert D., and Jennifer L. Miskimins. 2015a. "Calculation and Implications of Breakdown Pressures in Directional Wellbore Stimulation." P. D011S001R007 in *SPE Hydraulic Fracturing Technology Conference and Exhibition*. SPE.
- Barree, Robert D., and Jennifer L. Miskimins. 2015b. "Calculation and Implications of Breakdown Pressures in Directional Wellbore Stimulation." P. D011S001R007 in *SPE Hydraulic Fracturing Technology Conference and Exhibition*. SPE.
- Barree, Robert D., and Jennifer L. Miskimins. 2015c. "Calculation and Implications of Breakdown Pressures in Directional Wellbore Stimulation." P. D011S001R007 in *SPE Hydraulic Fracturing Technology Conference and Exhibition*. SPE.
- Berchenko, I., and E. Detournay. 1997. "Deviation of Hydraulic Fractures through Poroelastic Stress Changes Induced by Fluid Injection and Pumping." *International Journal of Rock Mechanics and Mining Sciences* 34(6):1009–19.
- Bunger, A. P., Ali Lakirouhani, and Emmanuel Detournay. 2010a. "Modelling the Effect of Injection System Compressibility and Viscous Fluid Flow on Hydraulic Fracture Breakdown Pressure." P. ISRM-ISRS in *ISRM international symposium on in-situ rock stress*. ISRM.

- Bunger, A. P., Ali Lakirouhani, and Emmanuel Detournay. 2010b. “Modelling the Effect of Injection System Compressibility and Viscous Fluid Flow on Hydraulic Fracture Breakdown Pressure.” P. ISRM-ISRS in *ISRM international symposium on in-situ rock stress*. ISRM.
- Carbonell, Roberto, and Emmanuel Detournay. 1995. “Modeling Fracture Initiation and Propagation from a Pressurized Hole: A Dislocation-Based Approach.” P. ARMA-95 in *ARMA US Rock Mechanics/Geomechanics Symposium*. ARMA.
- Chatfield, Chris. 1988. “What Is the ‘Best’ Method of Forecasting?” *Journal of Applied Statistics* 15(1):19–38.
- Chatterji, S., and P. Christensen. 1979. “A Mechanism of Breakdown of Limestone Nodules in a Freeze-Thaw Environment.” *Cement and Concrete Research* 9(6):741–46.
- Chen, Bowen, Qi Li, Yongsheng Tan, Tao Yu, Xiaying Li, and Xiaochun Li. 2024. “Experimental Measurements and Characterization Models of Caprock Breakthrough Pressure for CO<sub>2</sub> Geological Storage.” *Earth-Science Reviews* 104732.
- Chen, Peng, Aichen Niu, Duanyang Liu, Wei Jiang, and Bin Ma. 2018. “Time Series Forecasting of Temperatures Using SARIMA: An Example from Nanjing.” P. 052024 in *IOP conference series: materials science and engineering*. Vol. 394. IOP Publishing.
- Chen, Tianqi, and Carlos Guestrin. 2016. “Xgboost: A Scalable Tree Boosting System.” Pp. 785–94 in *Proceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining*.
- Choudhury, Arghya, Arpita Mondal, and Subhadeep Sarkar. 2024. “Searches for the BSM Scenarios at the LHC Using Decision Tree-Based Machine Learning Algorithms: A Comparative Study and Review of Random Forest, AdaBoost, XGBoost and LightGBM Frameworks.” *The European Physical Journal Special Topics* 1–39.
- Detournay, E., and AH-D. Cheng. 1988. “Poroelastic Response of a Borehole in a Non-Hydrostatic Stress Field.” Pp. 171–82 in *International Journal of Rock Mechanics and Mining Sciences & Geomechanics Abstracts*. Vol. 25. Elsevier.
- Dubey, Ashutosh Kumar, Abhishek Kumar, Vicente García-Díaz, Arpit Kumar Sharma, and Kishan Kanhaiya. 2021. “Study and Analysis of SARIMA and LSTM in Forecasting Time Series Data.” *Sustainable Energy Technologies and Assessments* 47:101474.

- Fallahzadeh, S. H. 2016. "The Influence of Fracturing Fluid Properties on Hydraulic Fracture Propagation. *Journal of Petroleum Science and Engineering*, 147, 94-105."
- Farid Ibrahim, Ahmed, and Hisham Nasr-El-Din. 2018. "Evaluation of the Breakdown Pressure to Initiate Hydraulic Fractures of Tight Sandstone and Shale Formations." P. D011S020R002 in *SPE Trinidad and Tobago Section Energy Resources Conference? SPE*.
- Fatahi, Hassan, Md Mofazzal Hossain, Seyed Hassan Fallahzadeh, and Masood Mostofi. 2016. "Numerical Simulation for the Determination of Hydraulic Fracture Initiation and Breakdown Pressure Using Distinct Element Method." *Journal of Natural Gas Science and Engineering* 33:1219–32.
- Fjar, E., R. M. Holt, A. M. Raaen, R. Risnes, and P. Horsrud. 2008. "Petroleum Related Rock Mechanics. Elsevier."
- Gaarenstroom, L., R. A. J. Tromp, M. C. De Jong, and A. M. Brandenburg. 1993. "Overpressures in the Central North Sea: Implications for Trap Integrity and Drilling Safety." Pp. 1305–13 in *Geological Society, London, Petroleum Geology Conference Series*. Vol. 4. The Geological Society of London.
- Gan, Quan, Derek Elsworth, J. S. Alpern, Chris Marone, and Peter Connolly. 2015. "Breakdown Pressures Due to Infiltration and Exclusion in Finite Length Boreholes." *Journal of Petroleum Science and Engineering* 127:329–37.
- Garagash, D., and E. Detournay. 1997. "An Analysis of the Influence of the Pressurization Rate on the Borehole Breakdown Pressure." *International Journal of Solids and Structures* 34(24):3099–3118.
- Ghosh, Kushankur, Colin Bellinger, Roberto Corizzo, Paula Branco, Bartosz Krawczyk, and Nathalie Japkowicz. 2024. "The Class Imbalance Problem in Deep Learning." *Machine Learning* 113(7):4845–4901.
- Gong\*, Yiwen, Mohamed Mehana, Ilham El-Monier, Feng Xu, and Fengyang Xiong. 2019. "Machine Learning for Estimating Rock Mechanical Properties beyond Traditional Considerations." Pp. 466–80 in *Unconventional Resources Technology Conference, Denver, Colorado, 22-24 July 2019*. Unconventional Resources Technology Conference (URTeC); Society of ....
- Gray, K. E., C. M. Kim, and B. J. Hughes. 1981. "Stress State, Porosity, Permeability And Breakdown Pressure Around A Borehole During Fluid Injection." P. ARMA-81 in *ARMA US Rock Mechanics/Geomechanics Symposium*. ARMA.

- Groenenboom, Jeroen, and Joachim Falk. 2000. "Scattering by Hydraulic Fractures: Finite-Difference Modeling and Laboratory Data." *Geophysics* 65(2):612–22.
- Guo, F., N. R. Morgenstern, and J. D. Scott. 1993. "Interpretation of Hydraulic Fracturing Breakdown Pressure." Pp. 617–26 in *International Journal of Rock Mechanics and Mining Sciences & Geomechanics Abstracts*. Vol. 30. Elsevier.
- Guo, Zhaoquan, Shouceng Tian, Qingling Liu, Luyao Ma, Yuning Yong, and Ruiyue Yang. 2022. "Experimental Investigation on the Breakdown Pressure and Fracture Propagation of Radial Borehole Fracturing." *Journal of Petroleum Science and Engineering* 208:109169.
- Haimson, Bezalel, and Charles Fairhurst. 1969. "Hydraulic Fracturing in Porous-Permeable Materials." *Journal of Petroleum Technology* 21(07):811–17.
- Haimson, B. ., & Fairhurst, C. n.d. "Initiation and Extension of Hydraulic Fractures in Rocks. Society of Petroleum Engineers Journal, 7(3), 310-318." 1967.
- Hamdi, O., and F. Boulvain. n.d. "Geochemical and Mineralogical Characterizations of Silurian Hot Shales: Implications for Shale Gas/Oil Reservoir Potential in Jeffara Basin-Southeastern Tunisia, North Africa."
- Ha, Seong Jun, and Tae Sup Yun. 2020. "Experimental and Numerical Investigation of Time-Delayed Pressurization for the Manipulation of Breakdown Pressure and Acoustic Emissions during Hydraulic Stimulation." *Engineering Fracture Mechanics* 231:107005.
- Hayashi, Kazuo, Akira Sato, and Takatoshi Ito. 1997. "In Situ Stress Measurements by Hydraulic Fracturing for a Rock Mass with Many Planes of Weakness." *International Journal of Rock Mechanics and Mining Sciences* 34(1):45–58.
- Hochreiter, S. 1997. "Long Short-Term Memory." *Neural Computation MIT-Press*.
- Hou, Lei, Linbo Zhou, Derek Elsworth, Sen Wang, and Wendong Wang. 2024. "Prediction of Fracturing Pressure and Parameter Evaluations at Field Practical Scales." *Rock Mechanics and Rock Engineering* 57(4):2567–80.
- Hou, Yuntian, and Di Zhang. 2024. "A Comprehensive Survey on Kolmogorov Arnold Networks (Kan)." *ArXiv Preprint ArXiv:2407.11075*.
- Huang, Hai, Wentong Zhang, Huaqiang Shi, Jun Ni, Li Ding, Bing Yang, Yong Zheng, and Xinru Li. 2024. "Experimental Investigation of Microscale Mechanical Alterations in Shale Induced by Fracturing Fluid Contact." *Gas Science and Engineering* 124:205264.

- Huang, Ruijie, Chenji Wei, Baohua Wang, Jian Yang, Xin Xu, Suwei Wu, and Suqi Huang. 2022. “Well Performance Prediction Based on Long Short-Term Memory (LSTM) Neural Network.” *Journal of Petroleum Science and Engineering* 208:109686.
- Hubbert, M. K. ,., and D. G. Willis. 1957. “‘Mechanics of Hydraulic Fracturing.’ *Journal of Petroleum Technology*, Vol. 9, No. 6, 1957, Pp. 153-168.”
- Ikeda, S., S. Ogata, H. Yasuhara, and K. Kishida. 2024. “Numerical Simulation of Hydraulic Fracturing Considering Grain Boundary Weakness and Fluid-Rock Heat Transfer.” P. D022S024R016 in *ARMA US Rock Mechanics/Geomechanics Symposium*. ARMA.
- Ishida, Tsuyoshi, Kazuhei Aoyagi, Tomoya Niwa, Youqing Chen, Sumihiko Murata, Qu Chen, and Yoshiki Nakayama. 2012. “Acoustic Emission Monitoring of Hydraulic Fracturing Laboratory Experiment with Supercritical and Liquid CO<sub>2</sub>.” *Geophysical Research Letters* 39(16).
- Ito, Takatoshi. 2008. “Effect of Pore Pressure Gradient on Fracture Initiation in Fluid Saturated Porous Media: Rock.” *Engineering Fracture Mechanics* 75(7):1753–62.
- Ito, T. ,. Hayashi, K. 1990. “Hydraulic Fracturing Simulation by a Point Stress Model. *International Journal of Rock Mechanics and Mining Sciences & Geomechanics Abstracts*, 27(6), 491-499.”
- Jiang, Changbao, Yunpeng Lei, Bozhi Deng, Xiangyu Fan, Minghui Li, and Mingyang Wu. 2025. “Effect of Mesoscopic Heterogeneity on Hydraulic Fracturing Characteristics of Granite under High Temperature.” *Geoenergy Science and Engineering* 246:213626. doi: <https://doi.org/10.1016/j.geoen.2024.213626>.
- Jung, SungGyu, Melvin B. Diaz, Kwang Yeom Kim, Hannes Hofmann, and Günter Zimmermann. 2021. “Fatigue Behavior of Granite Subjected to Cyclic Hydraulic Fracturing and Observations on Pressure for Fracture Growth.” *Rock Mechanics and Rock Engineering* 1–14.
- Ke, Guolin, Qi Meng, Thomas Finley, Taifeng Wang, Wei Chen, Weidong Ma, Qiwei Ye, and Tie-Yan Liu. 2017. “Lightgbm: A Highly Efficient Gradient Boosting Decision Tree.” *Advances in Neural Information Processing Systems* 30.
- Kirsch E.G. 1898. “The Theory of Elasticity and the Requirements of Strength Theory.”
- Kocoglu, Yildirim, Sheldon Gorell, and P. McElroy. 2021. “Application of Bayesian Optimized Deep Bi-LSTM Neural Networks for Production Forecasting of Gas Wells in Unconventional Shale Gas Reservoirs.” Pp. 2176–96 in *Unconventional*

- Resources Technology Conference, 26–28 July 2021. Unconventional Resources Technology Conference (URTeC).*
- Kong, Mingwei, Zhaopeng Zhang, Chunyan Zhao, Huasheng Chen, Xinfang Ma, and Yushi Zou. 2021. “Rock Mechanical Properties and Breakdown Pressure of High-Temperature and High-Pressure Reservoirs in the Southern Margin of Junggar Basin.” *Geofluids* 2021(1):1116136.
- Lamy-Chappuis, F. ., et al. 2020. “ Effects of N<sub>2</sub>-Enriched Brine Injection on Mechanical Properties of Calcite-Bearing Sandstone.”
- Lecampion, Brice. 2012. “Hydraulic Fracture Initiation from an Open-Hole: Wellbore Size, Pressurization Rate and Fluid-Solid Coupling Effects.” P. ARMA-2012 in *ARMA US Rock Mechanics/Geomechanics Symposium*. ARMA.
- Lee, S. H., C. L. Jensen, and M. F. Lough. 2000. “Efficient Finite-Difference Model for Flow in a Reservoir with Multiple Length-Scale Fractures.” *SPE Journal* 5(03):268–75.
- Li, Ning, Heping Xie, Ziqi Gao, and Cunbao Li. 2022. “Study on the Hydraulic Fracturing Failure Behaviour of Granite and Its Comparison with Gas Fracturing.” *Sustainability* 14(21):14593.
- Liu, Ziming, Yixuan Wang, Sachin Vaidya, Fabian Ruehle, James Halverson, Marin Soljačić, Thomas Y. Hou, and Max Tegmark. 2024. “Kan: Kolmogorov-Arnold Networks.” *ArXiv Preprint ArXiv:2404.19756*.
- Lockner, D., and J. D. Byerlee. 1977. “Hydrofracture in Weber Sandstone at High Confining Pressure and Differential Stress.” *Journal of Geophysical Research* 82(14):2018–26.
- Lütkepohl, Helmut. 2006. “Forecasting with VARMA Models.” *Handbook of Economic Forecasting* 1:287–325.
- Lu, Y. L., D. Elsworth, and L. G. Wang. 2013. “Microcrack-Based Coupled Damage and Flow Modeling of Fracturing Evolution in Permeable Brittle Rocks.” *Computers and Geotechnics* 49:226–44.
- Marcon, Virginia, Craig Joseph, Kimberly E. Carter, Sheila W. Hedges, Christina L. Lopano, George D. Guthrie, and J. Alexandra Hakala. 2017. “Experimental Insights into Geochemical Changes in Hydraulically Fractured Marcellus Shale.” *Applied Geochemistry* 76:36–50.

- Ma, Shuqi, Qianyi Liu, and Yudong Zhang. 2021. "A Prediction Method of Fire Frequency: Based on the Optimization of SARIMA Model." *PLoS One* 16(8):e0255857.
- Mathur, A., S. Ali, S. Eichmann, A. Gupta, M. Ali, R. Vaidya, R. Mesdour, M. Albassam, M. Cordes, and W. D. Von Gonten. 2024. "Crude Steady-State Permeability Measurements on Carbonate Source Rocks and Effect of Water Block from Fracturing Fluid on Productivity." P. IPTC-24536 in *International Petroleum Technology Conference*. IPTC.
- Mbunge, Elliot, John Batani, Racheal Mafumbate, Caroline Gurajena, Stephen Fashoto, Talent Rugube, Boluwaji Akinuwaesi, and Andile Metfula. 2022. "Predicting Student Dropout in Massive Open Online Courses Using Deep Learning Models - A Systematic Review." Pp. 212–31 in.
- Milenković, Miloš, Libor Švadlenka, Vlastimil Melichar, Nebojša Bojović, and Zoran Avramović. 2018. "SARIMA Modelling Approach for Railway Passenger Flow Forecasting." *Transport* 33(5):1113–20.
- Mishra, Srikanta, and Akhil Datta-Gupta. 2017. *Applied Statistical Modeling and Data Analytics: A Practical Guide for the Petroleum Geosciences*. Elsevier.
- Morita, N., A. D. Black, and G. F. Fuh. 1996. "Borehole Breakdown Pressure with Drilling Fluids—I. Empirical Results." Pp. 39–51 in *International journal of rock mechanics and mining sciences & geomechanics abstracts*. Vol. 33. Elsevier.
- Mostafavi, V., G. Hareland, M. Belayneh, and B. S. Aadnøy. 2011. "Experimental and Mechanistic Modeling of Fracture Sealing Resistance with Respect to Fluid and Fracture Properties." P. ARMA-11 in *ARMA US Rock Mechanics/Geomechanics Symposium*. ARMA.
- Mukhamedzianova, Albina. 2017. "GOHFER® Software Evaluation."
- Muqtadir, Arqam, S. M. Elkatatny, M. A. Mahmoud, Abdulazeez Abdulraheem, and A. Gomaa. 2018. "Effect of Saturating Fluid on the Geomechanical Properties of Low Permeability Scioto Sandstone Rocks." P. ARMA-2018 in *ARMA US Rock Mechanics/Geomechanics Symposium*. ARMA.
- Niaz, Nazim Uddin, K. M. Nadim Shahariar, and Muhammed J. A. Patwary. 2022. "Class Imbalance Problems in Machine Learning: A Review of Methods and Future Challenges." Pp. 485–90 in *Proceedings of the 2nd International Conference on Computing Advancements*.

- Nicholson, Dawn T. 2001. "Pore Properties as Indicators of Breakdown Mechanisms in Experimentally Weathered Limestones." *Earth Surface Processes and Landforms: The Journal of the British Geomorphological Research Group* 26(8):819–38.
- Otchere, Daniel Asante, Tarek Omar Arbi Ganat, Raoof Gholami, and Syahrir Ridha. 2021. "Application of Supervised Machine Learning Paradigms in the Prediction of Petroleum Reservoir Properties: Comparative Analysis of ANN and SVM Models." *Journal of Petroleum Science and Engineering* 200:108182.
- Pandey, Manish, Masoud Karbasi, Mehdi Jamei, Anurag Malik, and Jaan H. Pu. 2023. "A Comprehensive Experimental and Computational Investigation on Estimation of Scour Depth at Bridge Abutment: Emerging Ensemble Intelligent Systems." *Water Resources Management* 37(9):3745–67.
- Powlay, Ben, Murat Karakus, Khalid Amrouch, and Chris Chester. 2024. "Effects of Notches on Breakdown Pressures and Fracture Evolution in Hydraulic Fracturing." *Rock Mechanics and Rock Engineering* 1–14.
- Prokhorenkova, Liudmila, Gleb Gusev, Aleksandr Vorobev, Anna Veronika Dorogush, and Andrey Gulin. 2018. "CatBoost: Unbiased Boosting with Categorical Features." *Advances in Neural Information Processing Systems* 31.
- Rasouli, Vamegh, Zachariah John Pallikathakathil, and Elike Mawuli. 2011. "The Influence of Perturbed Stresses near Faults on Drilling Strategy: A Case Study in Blacktip Field, North Australia." *Journal of Petroleum Science and Engineering* 76(1–2):37–50.
- Rummel, F. 1987. "Fracture Mechanics Approach to Hydraulic Fracturing Stress Measurements." *Frac Mechan Rock* 217:217–40.
- Rummel, F., G. Klee, and U. Weber. 1995. "Hydraulic vs Pneumatic Fracturing for in Situ Stress Determination in Rock Salt." Pp. 337–42 in *International journal of rock mechanics and mining sciences & geomechanics abstracts*. Vol. 32. Pergamon.
- Saxena, Abhishek, Neeta Verma, and Krishna Chandra Tripathi. 2013. "A Review Study of Weather Forecasting Using Artificial Neural Network Approach." *Int. J. Eng. Res. Technol* 2(11):2029–36.
- Schmid, J., B. Aiken, F. El Yaakobi, and L. Colmenares. 2024. "Mitigating Downhole Calcite and Barite Deposition in the Montney: A Successful Scale Squeeze Program." P. D021S021R004 in *SPE Canadian Energy Technology Conference*. SPE.

- Schmitt, Douglas R., and Mark D. Zoback. 1992. "Diminished Pore Pressure in Low-porosity Crystalline Rock under Tensional Failure: Apparent Strengthening by Dilatancy." *Journal of Geophysical Research: Solid Earth* 97(B1):273–88.
- Schmitt, D. R., and M. D. Zoback. 1989. "Poroelastic Effects in the Determination of the Maximum Horizontal Principal Stress in Hydraulic Fracturing Tests—A Proposed Breakdown Equation Employing a Modified Effective Stress Relation for Tensile Failure." Pp. 499–506 in *International Journal of Rock Mechanics and Mining Sciences & Geomechanics Abstracts*. Vol. 26. Elsevier.
- Schmitt, D. R., and M. D. Zoback. 1993. "Infiltration Effects in the Tensile Rupture of Thin Walled Cylinders of Glass and Granite: Implications for the Hydraulic Fracturing Breakdown Equation." Pp. 289–303 in *International journal of rock mechanics and mining sciences & geomechanics abstracts*. Vol. 30. Elsevier.
- Scholtès, L. U. C., and Frédéric-Victor Donzé. 2012. "Modelling Progressive Failure in Fractured Rock Masses Using a 3D Discrete Element Method." *International Journal of Rock Mechanics and Mining Sciences* 52:18–30.
- Siddiqui, H. A., M. S. Iqbal, H. Rafique, and A. Hasnain. 2020. "Successful Integration of Channel Fracturing Unveils Promising Prospects in a Southern Region of Pakistan."
- Siggins, Anthony F. 2006. "Velocity-Effective Stress Response of CO<sub>2</sub>-Saturated Sandstones." *Exploration Geophysics* 37(1):60–66.
- Song, I., and B. C. Haimson. 2001. "Effect of Pressurization Rate and Initial Pore Pressure on the Magnitude of Hydrofracturing Breakdown Pressure in Tablerock Sandstone." P. ARMA-01 in *ARMA US Rock Mechanics/Geomechanics Symposium*. ARMA.
- Song, Insun, Mancheol Suh, Kyoung Sik Won, and Bezalel Haimson. 2001a. "A Laboratory Study of Hydraulic Fracturing Breakdown Pressure in Tablerock Sandstone." *Geosciences Journal* 5:263–71.
- Song, Insun, Mancheol Suh, Kyoung Sik Won, and Bezalel Haimson. 2001b. "A Laboratory Study of Hydraulic Fracturing Breakdown Pressure in Tablerock Sandstone." *Geosciences Journal* 5:263–71.
- Sun, S. ., et al. 2020. "Hybrid Model Combining Finite Element Simulations with Machine Learning Predictions for Breakdown Pressure."
- Tan, Peng, Yan Jin, Liang Yuan, Zhen-Yu Xiong, Bing Hou, Mian Chen, and Li-Ming Wan. 2019. "Understanding Hydraulic Fracture Propagation Behavior in Tight

- Sandstone–Coal Interbedded Formations: An Experimental Investigation.” *Petroleum Science* 16:148–60.
- Tariq, Zeeshan, Bicheng Yan, Shuyu Sun, Manojkumar Gudala, Murtada Saleh Aljawad, Mobeen Murtaza, and Mohamed Mahmoud. 2022. “Machine Learning-Based Accelerated Approaches to Infer Breakdown Pressure of Several Unconventional Rock Types.” *ACS Omega* 7(45):41314–30.
- Tuzingila, Romulus Mawa, Lingyun Kong, and Ruben Koy Kasongo. 2024. “A Review on Experimental Techniques and Their Applications in the Effects of Mineral Content on Geomechanical Properties of Reservoir Shale Rock.” *Rock Mechanics Bulletin* 100110.
- Vagropoulos, Stylianos I., G. I. Chouliaras, Evaggelos G. Kardakos, Christos K. Simoglou, and Anastasios G. Bakirtzis. 2016. “Comparison of SARIMAX, SARIMA, Modified SARIMA and ANN-Based Models for Short-Term PV Generation Forecasting.” Pp. 1–6 in *2016 IEEE international energy conference (ENERGYCON)*. IEEE.
- Vernik, Lev, and Mark D. Zoback. 1992. “Estimation of Maximum Horizontal Principal Stress Magnitude from Stress-induced Well Bore Breakouts in the Cajon Pass Scientific Research Borehole.” *Journal of Geophysical Research: Solid Earth* 97(B4):5109–19.
- Wang, H. ., et al. 2014. “Fracture Network Development in Hydraulic Fracturing. SPE Hydraulic Fracturing Technology Conference, Paper SPE-168589-MS.”
- Wang, Jiehao, Derek Elsworth, Yu Wu, Jishan Liu, Wancheng Zhu, and Yu Liu. 2018. “The Influence of Fracturing Fluids on Fracturing Processes: A Comparison between Water, Oil and SC-CO 2.” *Rock Mechanics and Rock Engineering* 51:299–313.
- Wang, Suifeng, Yanhui Han, Wanrui Hu, Xianyu Zhao, Liping Zhang, and Tao Wang. 2025. “Modeling Interactions between Hydraulic Fracture and Pre-Existing Microcracks in Crystalline Rocks Using Hydro-Grain-Texture Model.” *Geoenergy Science and Engineering* 244:213459. doi: <https://doi.org/10.1016/j.geoen.2024.213459>.
- Wanniarachchi, Wanniarachchillage Ayal Maneth, Ranjith Pathegama Gamage, Mandadige Samintha Anne Perera, Tharaka Dilanka Rathnaweera, Mingzhong Gao, and Eswaran Padmanabhan. 2017. “Investigation of Depth and Injection Pressure Effects on Breakdown Pressure and Fracture Permeability of Shale Reservoirs: An Experimental Study.” *Applied Sciences* 7(7):664.

- Weijermars, Rudy. 2011. "Analytical Stress Functions Applied to Hydraulic Fracturing: Scaling the Interaction of Tectonic Stress and Unbalanced Borehole Pressures." P. ARMA-11 in *ARMA US Rock Mechanics/Geomechanics Symposium*. ARMA.
- De Winter, Joost C. F., Samuel D. Gosling, and Jeff Potter. 2016. "Comparing the Pearson and Spearman Correlation Coefficients across Distributions and Sample Sizes: A Tutorial Using Simulations and Empirical Data." *Psychological Methods* 21(3):273.
- Wu, Feipeng, Xianzhang Fan, Jing Liu, and Xiaojun Li. 2020. "Analytical Interpretation of Hydraulic Fracturing Initiation Pressure and Breakdown Pressure." *Journal of Natural Gas Science and Engineering* 76:103185.
- Wu, Xiaolong, Yintong Guo, Xin Chang, Zhenhui Bi, Guokai Zhao, Hanzhi Yang, and Wuhao Guo. 2023. "Experimental Study on Cyclic Hydraulic Fracturing of Tight Sandstone under In-Situ Stress." *Processes* 11(3):875.
- Wu, Yu, Jing Tao, Jiehao Wang, Yan Zhang, and Shuhua Peng. 2021a. "Experimental Investigation of Shale Breakdown Pressure under Liquid Nitrogen Pre-Conditioning before Nitrogen Fracturing." *International Journal of Mining Science and Technology* 31(4):611–20.
- Wu, Yu, Jing Tao, Jiehao Wang, Yan Zhang, and Shuhua Peng. 2021b. "Experimental Investigation of Shale Breakdown Pressure under Liquid Nitrogen Pre-Conditioning before Nitrogen Fracturing." *International Journal of Mining Science and Technology* 31(4):611–20.
- Yang, Bing, Haizhu Wang, Zhonghou Shen, Olufemi Olorode, Bin Wang, Yong Zheng, Wanjuan Yan, and Zhaoyang Jia. 2021. "Full-Sample X-Ray Microcomputed Tomography Analysis of Supercritical CO<sub>2</sub> Fracturing in Tight Sandstone: Effect of Stress on Fracture Dynamics." *Energy & Fuels* 35(2):1308–21.
- Yang, Huohai, Binghong Xie, Xuanyu Liu, Xiangshu Chu, Jingxin Ruan, Yanxu Luo, and Jie Yue. 2022. "Breakdown Pressure Prediction of Tight Sandstone Horizontal Wells Based on the Mechanism Model and Multiple Linear Regression Model." *Energies* 15(19):6944.
- Yu, Yong, Xiaosheng Si, Changhua Hu, and Jianxun Zhang. 2019. "A Review of Recurrent Neural Networks: LSTM Cells and Network Architectures." *Neural Computation* 31(7):1235–70.
- Zeng, Y. J., Kui Liu, S. D. Ding, S. M. Zhou, Q. Tao, J. L. Du, and X. Y. Du. 2020. "Effect of the Hydraulic Fracturing-Induced Slippage of Regional Natural Fracture on Casing Deformation in Horizontal Shale Gas Wells: A Case Study on the

- Weiyuan Block.” P. 022011 in *IOP Conference Series: Earth and Environmental Science*. Vol. 570. IOP Publishing.
- Zhang, Hua, Shunde Yin, and Bernt S. Aadnoy. 2019. “Numerical Investigation of the Impacts of Borehole Breakouts on Breakdown Pressure.” *Energies* 12(5):888.
- Zhang, J. ., et al. 2017. “Effects of High Temperatures on Rock Mechanical Properties and Breakdown Pressure in Carbonate Reservoirs. *Journal of Geophysics and Engineering*, 14(4), 907-919.”
- Zhang, Xiufeng, Min Zhang, Shuyuan Liu, and Heyang Liu. 2024. “Prediction of Breakdown Pressure Using a Multi-Layer Neural Network Based on Supercritical CO2 Fracturing Data.” *Applied Sciences* 14(22):10545.
- Zhang, Yushuai, Jincal Zhang, Bin Yuan, and Shangxian Yin. 2018a. “In-Situ Stresses Controlling Hydraulic Fracture Propagation and Fracture Breakdown Pressure.” *Journal of Petroleum Science and Engineering* 164:164–73.
- Zhang, Yushuai, Jincal Zhang, Bin Yuan, and Shangxian Yin. 2018b. “In-Situ Stresses Controlling Hydraulic Fracture Propagation and Fracture Breakdown Pressure.” *Journal of Petroleum Science and Engineering* 164:164–73.
- Zhan, J., S. Seetahal, J. Cao, H. Hejazi, D. Alexander, R. He, K. Zhang, and Z. Chen. 2016. “An Integrated Method to Characterize Shale Gas Reservoir Performance.” P. SPE-180884 in *SPE Trinidad and Tobago Section Energy Resources Conference?* SPE.
- Zhao, X. 2016. “Mechanical Behavior of Rocks under Complex Stresses and High Temperatures: Effects on Breakdown Pressure.” *International Journal of Rock Mechanics and Mining Sciences*, 86, 90-102.
- Zhou, Desheng, and Pei He. 2015. “Major Factors Affecting Simultaneous Frac Results.” P. SPE-173633 in *SPE Oklahoma City Oil and Gas Symposium/Production and Operations Symposium*. SPE.
- Zou, Yushi, Ning Li, Xinfang Ma, Shicheng Zhang, and Sihai Li. 2018. “Experimental Study on the Growth Behavior of Supercritical CO2-Induced Fractures in a Layered Tight Sandstone Formation.” *Journal of Natural Gas Science and Engineering* 49:145–56.

## VITAE

**Name:** Subhan Aliyev

**Nationality:** Azerbaijani

**Date of Birth:** 09/09/2000

**Email:** subhan.aliyev.agali@gmail.com

**Address:** Azerbaijan, Baku, Nizami District, Ilham wMuzaffaroglu street 54A, P.O. Box:

AZ 1002

**Academic Background:** Petroleum Engineering

**Education:** Petroleum Engineering MSc., KFUPM 2025

Petroleum Engineering BSc., ASOIU 2021

### **Publications:**

Mohamed, Badr A., Aliyev, Subhan, Ibrahim, Ahmed Farid, and Talal Al-Shafloot.

"Real-Time Forecasting of Subsurface Porosity During Drilling Using Advanced Time Series Models." Paper presented at the ADIPEC, Abu Dhabi, UAE, November 2024.

doi: <https://doi.org/10.2118/222460-MS>

