



UNIVERSIDADE FEDERAL DE PERNAMBUCO  
CENTRO DE TECNOLOGIA E GEOCIÊNCIAS  
DEPARTAMENTO DE ENGENHARIA CIVIL E AMBIENTAL  
PROGRAMA DE PÓS-GRADUAÇÃO EM ENGENHARIA CIVIL

ANNY VIRGÍNIA SOUZA DE LIMA

**APLICAÇÕES DE TÉCNICAS DE MACHINE LEARNING E ANÁLISE DE  
SENSIBILIDADE DE PRESSÕES DE INJEÇÃO EM CENÁRIOS DE REATIVAÇÃO  
DE ZONAS DE FALHAS GEOLÓGICAS**

Recife

2023

ANNY VIRGÍNIA SOUZA DE LIMA

**APLICAÇÕES DE TÉCNICAS DE MACHINE LEARNING E ANÁLISE DE  
SENSIBILIDADE DE PRESSÕES DE INJEÇÃO EM CENÁRIOS DE REATIVAÇÃO  
DE ZONAS DE FALHAS GEOLÓGICAS**

Dissertação apresentada ao Programa de Pós-Graduação em Engenharia Civil da Universidade Federal de Pernambuco, como requisito parcial para a obtenção do título de Mestre em Engenharia Civil.

Área de concentração: Simulação e Gerenciamento de Reservatórios de Petróleo.

Orientador: Prof. Dr. Leonardo José do Nascimento Guimarães.

Coorientadora: Profa. Dra. Leila Brunet de Sá Beserra.

Recife

2023

Catálogo na fonte:  
Bibliotecária Sandra Maria Neri Santiago, CRB-4 / 1267

L732a

Lima, Anny Virgínia Souza de.

Aplicações de técnicas de machine learning e análise de sensibilidade de pressões de injeção em cenários de reativação de zonas de falhas geológicas / Anny Virgínia Souza de Lima. – 2023.

81 f.: il., fig., tab., e siglas.

Orientador: Prof. Dr. Leonardo José do Nascimento Guimarães.

Coorientadora: Profa. Dra. Leila Brunet de Sá Beserra.

Dissertação (Mestrado) – Universidade Federal de Pernambuco. CTG. Programa de Pós-Graduação em Engenharia Civil. Recife, 2023.

Inclui referências.

1. Engenharia civil. 2. Reativação de falhas. 3. Análise de sensibilidade. 4. Pressão de injeção. 5. Análise discriminante linear. 6. Redes neurais artificiais. I. Guimarães, Leonardo José do Nascimento (Orientador). II. Beserra, Leila Brunet de Sá (Coorientadora). III. Título.

UFPE

624 CDD (22. ed.)

BCTG/2023-39

ANNY VIRGÍNIA SOUZA DE LIMA

**APLICAÇÕES DE TÉCNICAS DE MACHINE LEARNING E ANÁLISE DE  
SENSIBILIDADE DE PRESSÕES DE INJEÇÃO EM CENÁRIOS DE REATIVAÇÃO  
DE ZONAS DE FALHAS GEOLÓGICAS**

Dissertação apresentada ao Programa de Pós-Graduação em Engenharia Civil da Universidade Federal de Pernambuco, Centro de Tecnologia e Geociências, como requisito parcial para a obtenção do título de Mestre em Engenharia Civil. Área de concentração: Simulação e Gerenciamento de Reservatórios de Petróleo.

Aprovada em: 15/02/2023.

**BANCA EXAMINADORA**

Participação por videoconferência  
Prof. Dr. Bernardo Horowitz (Examinador Interno)  
Universidade Federal de Pernambuco

Participação por videoconferência  
Prof. Dr. Manoel Porfírio Cordão Neto (Examinador Externo)  
Universidade de Brasília

Participação por videoconferência  
Prof. Dr. Jonathan da Cunha Teixeira (Examinador Externo)  
Universidade Federal de Alagoas

## **AGRADECIMENTOS**

“Espera no Senhor, anima-te, e ele fortalecerá o teu coração; espera, pois, no Senhor.”, Salmos 27:14.

Agradeço a Deus por todas as bênçãos que permitiram a realização deste trabalho, e à intercessão de Nossa Senhora em toda a minha caminhada.

A minha família, em especial meu pai, Alfredo, meu irmão, Vagner, que sempre me apoiaram e incentivaram em todas as etapas da minha vida, e minha mãe (in memoriam), que deixou grandes ensinamentos e eternas saudades.

A minha tia Vitória e meu tio José Audair, pelo carinho, presença constante, apoio e incentivo.

Aos meus avós Augusto (in memoriam) e Maria Júlia, que além de me educarem, não mediram esforços para a concretização deste sonho, sendo meus segundos pais na terra.

Ao meu orientador professor Leonardo José do Nascimento Guimarães e à minha coorientadora professora Leila Brunet de Sá Beserra pela confiança depositada no meu trabalho, assim como pela oportunidade concedida e disponibilidade em meio às minhas dúvidas.

Aos meus amigos, agradeço por todo o apoio, amizade e companheirismo.

Este trabalho foi realizado com o apoio da Coordenação de Aperfeiçoamento de Pessoal de Nível Superior – Brasil (CAPES) – Código de Financiamento 001.

## RESUMO

Falhas geológicas são estruturas comuns em reservatórios de petróleo que podem atuar como canais facilitadores de escoamento ou como barreiras selantes. Durante a exploração de hidrocarbonetos, a pressão no interior do reservatório varia, o que provoca uma resposta mecânica no meio e pode levar ao fenômeno de reativação das falhas. Isso acontece quando as deformações da rocha reabrem a falha, aumentando sua permeabilidade e permitindo o fluxo de fluidos. Neste trabalho, foi realizada uma análise de sensibilidade para estudar a influência de diferentes cenários de injeção na deformação, tensão cisalhante, permeabilidade, pressão dos líquidos e fluxo de fluidos em zonas de falhas geológicas. Para isso, foi utilizado o software CODE\_BRIGHT, um simulador baseado no método dos elementos finitos, onde o acoplamento hidromecânico é feito de maneira implícita e o modelo elastoplástico utilizado para modelar o comportamento mecânico da falha foi o de Drucker-Prager. Além disso, foi possível implementar algoritmos de aprendizado de máquina, como Análise Discriminante Linear (LDA) e Rede Neural Artificial (RNA), usando o modelo analítico de Mohr-Coulomb para classificar cenários de reativação de falhas geológicas. Esses modelos foram implementados em linguagem de programação Python com bibliotecas já consagradas na literatura para os métodos mencionados. Os resultados da análise de sensibilidade mostram que os incrementos na pressão de injeção são responsáveis pela reativação da falha, associados a deformações plásticas crescentes e ao surgimento de pressões e fluxos de fluidos na extensão da falha. Foi encontrado um limite para essa pressão de injeção, que não reativaria a falha, correspondente a pressão de injeção de 19,7 MPa. Pressões de injeção superiores podem tornar a falha um conduto que pode levar à perda de fluidos e à despressurização do reservatório, entre outros problemas relacionados à reativação da falha. Os modelos de classificação de reativação de falhas tiveram um desempenho excelente tanto para o LDA quanto para o RNA e podem ser uma abordagem preliminar para avaliar os cenários de reativação. Eles também podem incorporar soluções numéricas e dados laboratoriais de caracterização geomecânica para aumentar a complexidade e generalidade dessas técnicas.

Palavras-chave: reativação de falhas; análise de sensibilidade; pressão de injeção; análise discriminante linear; redes neurais artificiais.

## ABSTRACT

Geological faults are common structures in oil reservoirs that can act as facilitators of fluid flow channels or sealing barriers. During hydrocarbon exploration, the pressure inside the reservoir varies, which causes a mechanical response in the medium and can lead to the phenomenon of fault reactivation. This occurs when rock deformations reopen the fault, increasing its permeability and allowing fluid flow. In this work, a sensitivity analysis was performed to study the influence of different injection scenarios on deformation, shear stress, permeability, liquid pressure, and fluid flow in geological fault zones. For this, the CODE\_BRIGTH software, a simulator based on the finite element method, was used, where the hydro-mechanical coupling is done implicitly and the elastoplastic model used to model the mechanical behavior of the fault was the Drucker-Prager model. In addition, machine learning algorithms such as Linear Discriminant Analysis (LDA) and Artificial Neural Network (ANN) were implemented using the Mohr-Coulomb analytical model to classify geological fault reactivation scenarios. These models were implemented in the Python programming language with established libraries in the literature for the mentioned methods. The sensitivity analysis results show that injection pressure increments are responsible for fault reactivation, associated with increasing plastic deformations and the emergence of pressure and fluid flows in the fault extension. A limit for this injection pressure was found, which would not reactivate the fault, corresponding to an injection pressure of 19,7 MPa. Higher injection pressures can make the fault a conduit that can lead to fluid loss and reservoir depressurization, among other problems related to fault reactivation. The fault reactivation classification models had excellent performance for both LDA and ANN and can be a preliminary approach to evaluate reactivation scenarios. They can also incorporate numerical solutions and geomechanical characterization laboratory data to increase the complexity and generality of these techniques.

Keywords: fault reactivation; sensitivity analysis; injection pressure; linear discriminant analysis; artificial neural networks.

## LISTA DE FIGURAS

Figura 1 –	Zona de falha composta por núcleo, zonas de dano e protólito.....	23
Figura 2 –	Zona de falha dividida em núcleo, zonas de dano e protólito.....	23
Figura 3 –	(a) Diagrama esquemático da estrutura através de uma zona de falha, (b) modelo conceitual da distribuição de permeabilidade, (c) módulo de Young.....	24
Figura 4 –	Modelo esquemático de variação da permeabilidade ao longo da zona de falha.....	25
Figura 5 –	Distribuição do Módulo de Young ao longo da extensão da Zona de Falha.....	26
Figura 6 –	Teoria das falhas de Anderson relacionando falhas com as tensões principais: a) Falha Normal, b) Falha Transcorrente e c) Falha Reversa.....	27
Figura 7 –	Fluxograma de Etapas da Análise de Sensibilidade do Processo de Reativação de Falhas Associado ao Parâmetro Pressão de Injeção .....	39
Figura 8 –	Fluxograma de Atividades Para os Modelos de Classificação (LDA e RNA) de Cenários de Reativação de Falhas .....	41
Figura 9 –	Seção Geológica Bidimensional Típica .....	44
Figura 10 –	Esquemática da Geometria do Modelo .....	45
Figura 11 –	Geometria do Modelo e Malha de Elementos Finitos .....	46
Figura 12 –	Critério de Ruptura de Mohr Coulomb.....	49
Figura 13 –	Seção esquemática do reservatório para o modelo analítico .....	50
Figura 14 –	a) Amostras de duas classes (decratas em vermelho e azul) juntamente com os histogramas resultantes da projeção na linha que une as médias das classes. b) Projeção correspondente baseada no discriminante linear.....	52
Figura 15 –	Modelo de arquitetura de RNA.....	57
Figura 16 –	Fluxograma da Rede Neural.....	61
Figura 17 –	Distribuição das variáveis de entrada ( $P$ , $c$ , $\phi$ , $\sigma$ , $\alpha$ , $k_0$ ) .....	61
Figura 18 –	Distribuição das variáveis após a normalização, referente ao conjunto de treino.....	62

Figura 19 –	Distribuição das variáveis após a normalização, referente ao conjunto de teste.....	63
Figura 20 –	Gráfico de evolução da deformação plástica para os diferentes cenários de pressões de injeção (Elemento A) .....	64
Figura 21 –	Distribuição das deformações plásticas para as pressões de injeção de 19,7 MPa, 19,8 MPa, 21,0 MPa e 22,0 MPa no tempo final de simulação.....	65
Figura 22 –	Distribuição das tensões cisalhantes para as pressões de injeção de 19,7 MPa, 19,8 MPa, 21,0 MPa e 22,0 MPa no tempo final de simulação.....	66
Figura 23 –	Distribuição das permeabilidades para as pressões de injeção de 19,7 MPa, 19,8 MPa, 21,0 MPa e 22,0 MPa no tempo final de simulação.....	66
Figura 24 –	Distribuição das pressões de líquidos para as pressões de injeção de 19,7 MPa, 19,8 MPa, 21,0 MPa e 22,0 MPa no tempo final de simulação.....	67
Figura 25 –	Distribuição dos vetores de fluxo para as pressões de injeção de 19,7 MPa,19,8 MPa,21,0 MPa e 22,0 MPa no tempo final de simulação.....	67
Figura 26 –	Matriz de Confusão Para Modelo de LDA.....	68
Figura 27 –	Evolução da Acurácia para os dados de treino e teste ao longo das épocas.....	70
Figura 28 –	Evolução da perda para os dados de treino e teste ao longo das épocas.....	71
Figura 29 –	Desempenho do modelo de classificação de RNA através da matriz de confusão.....	72

## LISTA DE TABELAS

Tabela 1 –	Pressões de injeção analisadas.....	45
Tabela 2 –	Propriedades Para os Materiais do Modelo .....	47
Tabela 3 –	Dados Geomecânicos Para Modelo Analítico .....	50
Tabela 4 –	Matriz de Confusão.....	58
Tabela 5 –	Métricas de Avaliação Para o Modelo de LDA.....	69
Tabela 6 –	Métricas de Avaliação do Modelo de RNA .....	72

## LISTA DE SIGLAS

AUC	Area Under Curve
CODE_BRIGTH	COupled DEformation, BRIne, Gas and Heat Transport
FN	Falsos Negativos
FP	Falsos Positivos
LDA	Linear Discriminant Analysis
ML	Machine Learning
PCA	Principal Component Analysis
QDA	Quadratic Discriminant Analysis
RNA	Rede Neural Artificial
ROP	Rate Of Penetration
SVM	Support Vector Machines
TOB	Torque-On-Bit
VN	Verdadeiros Negativos
VP	Verdadeiros Positivos
ZDE	Zona de Dano Externa
ZDI	Zona de Dano Interna

## LISTA DE SÍMBOLOS

$c$	Coesão
$\varphi$	Ângulo de atrito
$\sigma'_1$	Tensão Efetiva Principal Máxima
$\sigma'_3$	Tensão Efetiva Principal Mínima
$\sigma_1$	Tensão Principal Máxima
$\sigma_2$	Tensão Principal Intermediária
$\sigma_3$	Tensão Principal Mínima
$\sigma_v$	Tensão Vertical
$\sigma_H$	Tensão Horizontal Máxima
$\sigma_h$	Tensão Horizontal Mínima
$\nabla$	Operador Nabla
$\sigma$	Tensor de Tensões Totais
$\mathbf{b}$	Vetor de Forças de Corpo
$\phi$	Porosidade
$\rho_l$	Densidade do Líquido
$P_l$	Pressão de Líquido (Fluido)
$\mathbf{q}_l$	Fluxo de Darcy
$\mathbf{u}$	Velocidade da Fase Sólida
$\rho_s$	Densidade da Fase Sólida
$\varepsilon_v$	Deformação Volumétrica Total
$\sigma'$	Tensor de Tensões Efetivas
$\sigma$	Tensor de Tensões Totais
$I$	Tensor Identidade
$\alpha$	Constante de Biot
$d\sigma'$	Incremento do Tensor de Tensões Efetivas
$D^{ep}$	Tensor Constitutivo Elastoplástico
$d\varepsilon$	Incremento do Tensor de Deformações
$F(\sigma, k)$	Função de Fluência
$J$	Tensão Desviadora
$p'$	Tensão Efetiva Média
$c'$	Coesão Efetiva

$\varphi'$	Ângulo de Atrito Efetivo
$\theta$	Ângulo de Lode
$e_d^p$	Deformação Plástica Cisalhante
$e_d^p/máx$	Deformação Plástica Cisalhante Máxima
$\mathbf{k}_i$	Tensor de Permeabilidade Intrínseca Inicial
$\mathbf{k}$	Tensor de Permeabilidade Intrínseca
$\Delta P$	Incremento de Pressão
$E$	Módulo de Elasticidade
$k$	Permeabilidade
$\Phi$	Porosidade
$\sigma'_v$	Tensão Vertical Efetiva
$\sigma'$	Tensão Efetiva Vertical
$k_0$	Coefficiente de Empuxo
$P$	Poropressão
$p$	Número de Variáveis
$X$	Variáveis
$g$	Quantidade de Populações
$\Sigma$	Matrizes de Variâncias e Covariâncias
$B$	Soma de Produtos Cruzados Entre Grupos
$\bar{\mu}_i$	Vetor de Médias Amostrais no Grupo $i$
$\bar{\mu}$	Vetor de Médias Amostrais
$W$	Soma de Produtos Cruzados Intra-Grupos
$n_i$	Número de Observações Grupo $i$
$S_i$	Variância no Grupo $i$
$\mu_{ij}$	Amostra $j$ que pertence ao grupo $i$
$\hat{a}$	Matriz de Projeção da Variação Entre Classes e a Variação Dentro da Classe
$\lambda_s$	Autovalores
$\hat{e}_1$	Autovetores Normais
$y$	Saída da Rede
$i$	Número de Neurônios na Camada
$l$	Número de Camadas
$N$	Número de Neurônios (Entradas)

$w_i$	Peso da Entrada $i$
$x_i$	Entrada $i$ da Rede
$b$	Viés
$z_i^l$	Soma Ponderada dos $i$ Neurônios na Camada $l$
$W$	Matriz de Peso da Rede Neural
$y_i^l$	Vetor de Saídas da Rede Neural
$b^l$	Vetor dos Vieses da Rede Neural
$C(W)$	Função de Custo da Rede Neural
$t_i$	Vetor de Valores Alvo da Rede Neural
$y_i$	Vetor de Predições/Saídas da Rede Neural

## SUMÁRIO

<b>1</b>	<b>INTRODUÇÃO .....</b>	<b>16</b>
1.1	OBJETIVO GERAL .....	20
1.2	OBJETIVOS ESPECÍFICOS .....	20
1.3	ESTRUTURA DO TRABALHO.....	20
<b>2</b>	<b>REVISÃO BIBLIOGRÁFICA.....</b>	<b>22</b>
2.1	ZONAS DE FALHAS GEOLÓGICAS.....	22
2.2	REATIVAÇÃO DE FALHAS .....	27
2.3	APLICAÇÕES DE TÉCNICAS DE ML EM RESERVATÓRIOS DE PETRÓLEO.....	32
<b>3</b>	<b>METODOLOGIA .....</b>	<b>38</b>
3.1	ANÁLISE DE SENSIBILIDADE .....	41
<b>3.1.1</b>	<b>Formulação Matemática .....</b>	<b>41</b>
<b>3.1.1</b>	<b>Modelo Numérico .....</b>	<b>43</b>
3.2	APLICAÇÕES DE MODELOS DE ML EM CENÁRIOS DE REATIVAÇÃO DE FALHAS .....	47
<b>3.2.1</b>	<b>Dados Para o Método Analítico de Análise de Reativação de Falhas.....</b>	<b>47</b>
<b>3.2.2</b>	<b>Aprendizado de Máquina (Machine learning (ML)) .....</b>	<b>50</b>
<b>3.2.3</b>	<b>Análise Discriminante Linear (Linear Discriminant Analysis (LDA)) .....</b>	<b>51</b>
3.2.3.1	Formulação Matemática de LDA .....	53
<b>3.2.4</b>	<b>Redes Neurais Artificiais (RNA).....</b>	<b>54</b>
3.2.4.1	Formulação Matemática de RNA .....	55
<b>3.2.5</b>	<b>Métricas de desempenho da classificação.....</b>	<b>57</b>
3.2.5.1	Matriz de Confusão .....	57
3.2.5.2	Acurácia.....	58
3.2.5.3	Precisão.....	59
3.2.5.4	Sensibilidade ( <i>Recall</i> ).....	59
3.2.5.4	F1-score.....	59
<b>3.2.6</b>	<b>Implementação dos Modelos Baseados em Aprendizado de Máquina.....</b>	<b>60</b>

3.2.6.1	Normalização dos dados de entrada .....	61
<b>4</b>	<b>RESULTADOS E DISCUSSÕES .....</b>	<b>64</b>
4.1	ANÁLISE DE SENSIBILIDADE DAS PRESSÕES DE INJEÇÃO.....	64
4.2	LDA .....	68
4.3	RNA.....	69
<b>5</b>	<b>CONCLUSÕES .....</b>	<b>74</b>
5.1	SUGESTÕES DE TRABALHOS FUTUROS.....	75
	<b>RERERÊNCIAS.....</b>	<b>76</b>

## 1 INTRODUÇÃO

Um reservatório de petróleo consiste em um meio poroso e deformável, onde diversos fenômenos de fluxo e deformação ocorrem e interagem uns com os outros. O entendimento deste problema é um grande desafio, no qual se busca formular matematicamente as equações que regem esses fenômenos para posteriormente embasar ferramentas numéricas que resolvem problemas reais de fluxo com acoplamento geomecânico nas escalas de poço, reservatório e bacia. No âmbito da Engenharia de Petróleo, compreender como os eventos ocorrem e evoluem é imprescindível para determinar o comportamento do reservatório e dos poços durante as fases de desenvolvimento, produção e abandono de um campo, visando obter eficácia e segurança nos processos. Portanto, este entendimento é fundamental para a redução de riscos e tomada de decisões durante todo o período de operação dos reservatórios.

Pode-se definir um reservatório de petróleo como um meio poroso, deformável e multifásico, de modo que se apresenta saturado por hidrocarbonetos e água. Conforme estudado por Zoback (2007), este meio encontra-se submetido a um determinado estado de tensões iniciais e à medida que esse estado de tensão é alterado, desenvolvem-se deformações, sendo as falhas e fraturas resultados dessas deformações.

As falhas geológicas são estruturas comuns em reservatórios de petróleo, podendo atuar como canais que facilitam o escoamento ou como barreiras (falhas selantes). Essas estruturas são imprescindíveis para o correto entendimento do comportamento geomecânico de um reservatório, uma vez que, quando ocorre sua reativação, tornam-se caminhos preferenciais para o fluxo, alterando a dinâmica da produção e podendo gerar graves acidentes ((Wiprut e Zoback, 2000), (Wiprut e Zoback, 2002)).

Durante a exploração de hidrocarbonetos, a pressão no interior do reservatório varia, caracterizando uma resposta mecânica em todo o meio, seja na própria rocha reservatório, bem como nas suas rochas vizinhas (*overburden*, *underburden*, *siderburden*). Como resposta comum da rocha reservatório, durante o processo de recuperação secundária, tem-se o fenômeno de reativação de falhas selantes, de maneira que as deformações da rocha induzem a reabertura da falha, associado ao

aumento de sua permeabilidade e o fluxo de fluidos através dela. Essa reativação está associada ao cisalhamento, e a inerente dilatância.

O problema de reativação de falhas e seu entendimento é de grande importância dentro da geomecânica de reservatórios, onde diversos autores tentam entender esse processo e prevê-lo. A análise de predição e prevenção do problema de reativação consiste na definição da máxima pressão de injeção durante a produção para evitar a reativação da falha e conseqüentemente, perdas de fluidos do reservatório. Como exemplos, podem ser citados diversos trabalhos envolvendo modelagem analítica e numérica em problemas de reativação de falhas, como Soltanzadeh e Hawkes (2009), Neves, Paiva e Luis (2009), Nacht et al. (2010), Guimarães, Gomes e Valadares (2009), Guimarães et al. (2010), Gomes et al. (2016), Rodrigues (2019), De Luca e Oliveira (2014), Siler, Hinz e Faulds (2018), Haddad e Eichhubl (2020), Lesueur, Poulet e Veveakis (2020), Khan, Khulief e Al-Shuhail (2020), Urpi et al.(2020), Taghipour et al. (2020), dentre outros.

Associado à reativação de falhas, deve-se considerar a incerteza das variáveis que configuram o modelo, dada a enorme variação dos parâmetros geomecânicos e petrofísicos na zona de falha, que é altamente heterogênea. Neste problema, é necessário lidar com a ausência de dados experimentais confiáveis associados aos materiais que compõem as zonas de falha, conforme apontam Pereira et al. (2014) e Pereira (2015). Portanto, para resolver o problema de reativação de falha geológica, é imprescindível entender os parâmetros que melhor representam o cenário de reativação e ainda buscar alternativas que tratem as incertezas inerentes ao processo.

Para lidar com as incertezas em vários aspectos da engenharia de petróleo, o uso de técnicas de Aprendizado de Máquina (*Machine Learning* (ML)) está em ascensão, conforme citado por Motamedi & Farchad (2020), Abdelghany et al. (2022), Jafarizadeh, Motamedi e Ahmadi (2021). Os algoritmos baseados em ML são modelos que detectam padrões e produzem relações com base em conjuntos de amostras, conhecidos como dados de treinamento. Existem vários tipos de algoritmos de ML. Para um problema de classificação, quando temos uma variável alvo categórica, como o caso de reativação de falhas, modelos de ML como Análise Discriminante Linear (*Linear Discriminant Analysis*(LDA)) e Redes Neurais Artificiais (RNA) podem resolver essa categorização. Exemplos de aplicações promissoras em engenharia de reservatórios, operações de perfuração, caracterização de reservatórios e

interpretações petrofísicas e sísmicas são explorados nos estudos de Motamedi e Farchad (2020), Miah (2020), Jafarizadeh, Motamedi e Ahmadi (2021), Farsi et al. (2021), Hamzeh (2021), Behesht Abad et al. (2021), Abad et al. (2022), Abdelghany et al.(2022), Banerjee e Chatterjee (2021), entre outros.

Apesar da vasta ascensão da utilização de ML em pesquisas envolvendo reservatórios de petróleo, pouco se relata a sua aplicação com ênfase na análise de reativação de falhas, ainda que existam aplicações na área, como o estudo apresentado por Ahmadi Goltapeh et al. (2022). Estes autores avaliaram a confiabilidade estrutural da rocha capeadora baseado em rede neural para análise de local de injeção de CO<sub>2</sub> utilizando um exemplo de cenário do Mar do Norte. Como os dados de caracterização da rocha capeadora eram escassos, os autores aplicaram o algoritmo de Monte Carlo para gerar parâmetros estocásticos a partir dos valores de média e desvio padrão disponíveis. Os conjuntos de treino do modelo correspondem a parâmetros de resistência, coesão ( $c$ ) e ângulo de atrito ( $\phi$ ), assim como dados de estado de tensão inicial (geostática), referentes as tensões efetivas principais máxima e mínima ( $\sigma'_1$  e  $\sigma'_3$ ). O procedimento tinha como objetivo classificar reativações de falhas presentes na rocha capeadora, fundamentado na determinação de fatores de segurança avaliados pelo critério de resistência da rocha, com o modelo de Mohr Coulomb. As simulações de Monte Carlo produziram amostras com tamanhos distintos: 1000, 10000, 50000, 100000 para cada uma das variáveis, mescladas para fornecerem entradas das redes neurais. Os dados foram divididos em conjuntos de treino e teste, de maneira que 75% dos valores foram usados para o treinamento da rede correspondente ao ajuste do modelo, enquanto os 25% restantes foram usados para teste, necessários para validação e avaliação do modelo. Como os parâmetros continham variáveis com escalas de unidades diferentes, foi necessário o pré-processamento de normalização dos dados, obtendo um aprendizado mais efetivo e minimizando a complexidade do modelo.

O modelo da arquitetura das redes desenvolvidas no trabalho de Ahmadi Goltapeh et al. (2022) foi baseado em tentativas para minimizar o tempo de aprendizado e obter acurácia ao modelo. Na etapa de definição da arquitetura das redes neurais, três camadas foram suficientes para criação das redes, sendo a primeira camada composta por 14 neurônios, a segunda camada com 10 neurônios e a última camada com apenas um neurônio. A última camada foi ativada usando a

função sigmoide, enquanto as demais camadas utilizaram a função de ativação da Unidade Linear Retificada (ReLU) (Rumelhart et al., 1986), tratando-se de uma classificação binária. Portanto, o modelo baseado em redes neurais produzido pelos autores, consiste em um conjunto de entrada contendo quatro atributos ( $\varphi$ ,  $c$ ,  $\sigma'_1$  e  $\sigma'_3$ ), usados para obtenção da classificação que pode indicar 1 (indicando não-falha) ou 0 (mostrando falha). Com base nos resultados, observa-se que ao aumentar o tamanho dos dados de entrada, as redes neurais convergem em épocas mais curtas, aumentando também a acurácia do modelo que se estabiliza após 50000 amostras. Os autores ainda relatam necessidade de investigações de campo e experimentais para melhor definir o comportamento da rocha capeadora, visando aprofundamento e melhora na implementação do modelo de redes neurais. Alguns benefícios das redes, conforme os autores expõem nas suas conclusões, além do tratamento da incerteza dos parâmetros, citam vantagens dos algoritmos de ML em comparação com os métodos dos elementos finitos nos aspectos de curto tempo de ajuste, alta precisão, flexibilidade e possibilidade de generalização a casos similares.

Neste contexto, o presente trabalho busca utilizar a análise numérica de sensibilidade associada à pressão de injeção de fundo de poço, considerando a reativação sobre a deformação elastoplástica, as tensões cisalhantes, a permeabilidade, as pressões e os fluxos de fluidos do meio, com o objetivo de quantificar a influência dos valores de pressurização no processo de reativação. Além disso, serão utilizados dados de casos referentes ao modelo analítico de Mohr Coulomb para analisar o processo de reativação de falhas. Posteriormente, serão realizadas implementações de modelos de ML, RNA e LDA, utilizando avaliações analíticas para treinar e validar os modelos. Dessa forma, com a criação de modelos de aprendizado de máquina treinados a partir dos dados analíticos, é possível ter uma forma abrangente de avaliar diversos cenários de reativação, com precisão, generalização e redução do esforço computacional, a partir das características de entrada dos modelos. Portanto, o trabalho contribui para uma melhor compreensão do processo de reativação de falhas geológicas em reservatórios de petróleo e apresenta abordagens inovadoras para a avaliação de cenários de reativação.

## 1.1 OBJETIVO GERAL

Contribuir para o entendimento de zonas de falhas e dos parâmetros que influenciam o processo de reativação de falhas por meio de métodos baseados em análise de sensibilidade e implementação de modelos de *machine learning*.

## 1.2 OBJETIVOS ESPECÍFICOS

- Realizar análise de sensibilidade da pressão de injeção de fundo do poço e avaliar sua influência sobre a deformação elastoplástica, tensões cisalhantes, permeabilidade, poropressão e fluxo de fluidos em zonas de falhas geológicas durante o processo de reativação.
- Quantificar a influência dos valores de pressurização no processo de reativação;
- Desenvolvimento de metodologias baseadas em aprendizado de máquina, redes neurais e LDA, para verificação de casos de reativação de falhas, capazes de representar com precisão o fenômeno de reativação;
- Simulação de casos testes que promovam validação dos modelos;
- Redução de esforço computacional para avaliação do fenômeno de reativação de falhas.

## 1.3 ESTRUTURA DO TRABALHO

O presente estudo inicia-se com a introdução do tema, contextualizando a reativação de falhas associada às mudanças do estado de tensões *in-situ* causadas pela injeção e produção de fluidos em reservatórios de petróleo. É destacada a incerteza associada ao processo de reativação e os modelos que auxiliam na identificação do fenômeno. Em seguida, é justificada a escolha do tema e apresentados os objetivos gerais e específicos do estudo. No segundo capítulo, realiza-se uma revisão bibliográfica básica sobre zonas de falhas geológicas, reativação de falhas e aplicações de técnicas de ML na indústria do petróleo.

No terceiro capítulo, são explicadas as etapas da pesquisa através de fluxogramas relacionados à análise de sensibilidade e à implementação dos modelos

de classificação baseados em aprendizado de máquina. Na análise de sensibilidade, baseada em simulação numérica, é descrita a formulação matemática do problema acoplado hidromecânico, assim como a descrição do modelo numérico e as propriedades dos materiais estabelecidas. Para os modelos de classificação, LDA e RNA, baseados na abordagem analítica de Mohr-Coulomb, são detalhadamente descritos a obtenção dos dados de treino e teste para os modelos de LDA e RNA e suas respectivas formulações matemáticas e métricas de avaliação de desempenho de classificação. Também é apresentada a abordagem de implementação das técnicas.

Os resultados obtidos são discutidos no quarto capítulo, apresentando a relação entre a análise de sensibilidade do parâmetro pressão de injeção e as deformações plásticas, tensões cisalhantes, permeabilidades, pressões e fluxo de líquidos. Além disso, através dos resultados dos métodos de ML aplicados aos dados analíticos, é observado o desempenho da LDA e da RNA nas classificações dos cenários de reativação de falhas.

Por fim, no quinto capítulo são apresentadas as conclusões obtidas a partir dos estudos de análise de sensibilidade e dos modelos de classificação e seus resultados. Por fim, são apresentadas as considerações finais, conclusões, possíveis melhorias e trabalhos futuros sobre o tema em estudo.

## 2 REVISÃO BIBLIOGRÁFICA

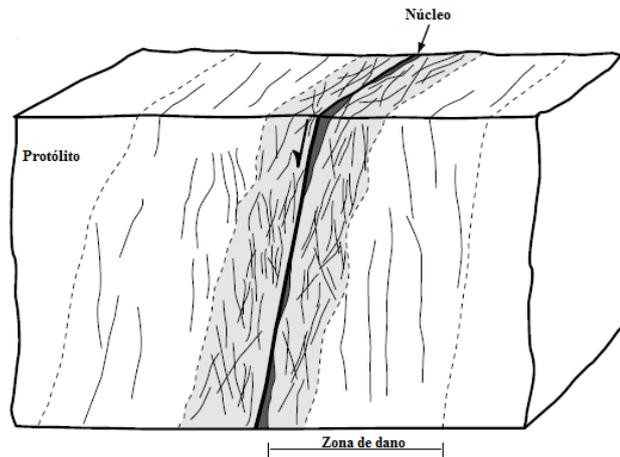
Neste capítulo está apresentada uma breve revisão da literatura sobre geomecânica de reativação de falhas geológicas e aplicações de técnicas de aprendizado de máquina em reservatórios de petróleo. Foram coletados diversos estudos e apresentados para fundamentação teórica da pesquisa, como norte da presente dissertação.

### 2.1 ZONAS DE FALHAS GEOLÓGICAS

O entendimento das zonas de falhas geológicas é essencial para a construção de modelos de reservatórios de petróleo. As falhas podem ser definidas como superfícies ou zonas estreitas onde as rochas estão cisalhadas e deslocadas, correspondendo a descontinuidades comuns em reservatórios de petróleo. Essas estruturas têm origem a partir do dinamismo litológico em conjunto com o acúmulo de tensões em subsuperfície. Enquanto selada, a falha funciona como uma barreira impermeável para o sistema, confinando todo o fluido presente no reservatório

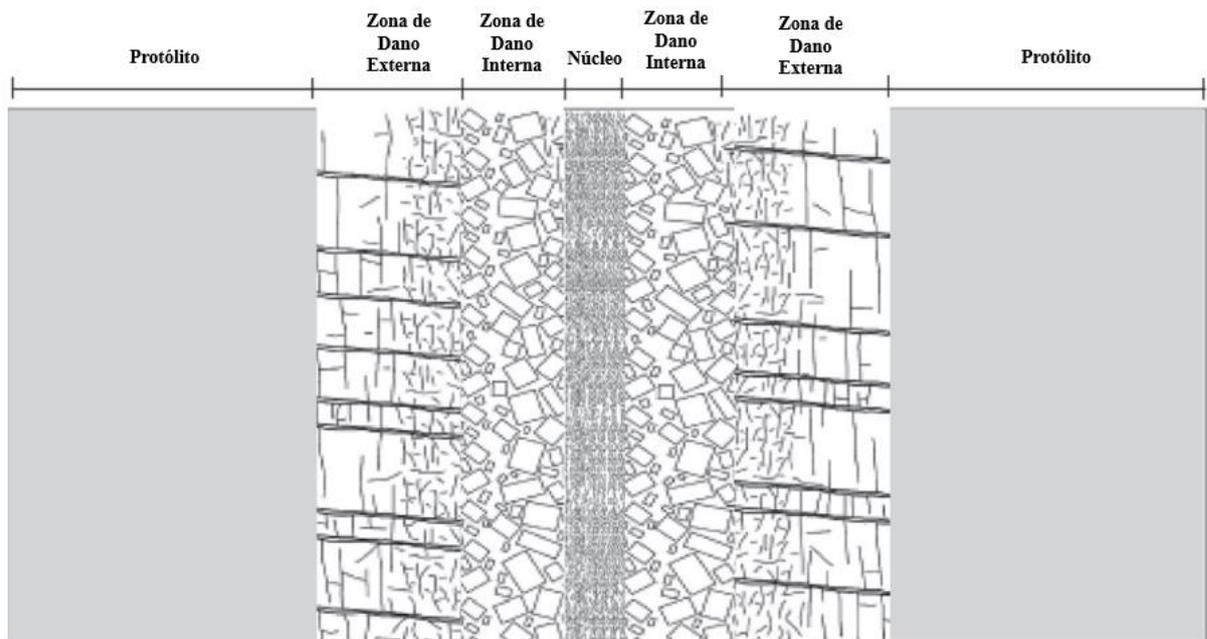
Caine, Evans e Forster (1996) caracterizam as zonas de falha como descontinuidades litológicas heterogêneas e estruturalmente anisotrópicas presentes na crosta terrestre. Os autores ainda relatam que a estrutura da zona de falha é formada por núcleo da falha, zona de dano e protólito (Figura 1 e Figura 2). Diversos autores, como Gudmundsson (2004), Cappa e Rutqvist (2011), Seyedi, Vu e Pouya (2015), Fossen (2016), Wetzel, Kempka, Kuhn (2018) e Celestino et al. (2020), se referem à zona de falha subdividida em núcleo da falha e zonas de dano, sendo objeto de estudo de pesquisas recentes.

Figura 1- Zona de falha composta por núcleo, zonas de dano e protólito.



Fonte: Adaptado de Micarelli, Benedicto e Wibberley (2006).

Figura 2-Zona de falha dividida em núcleo, zonas de dano e protólito.



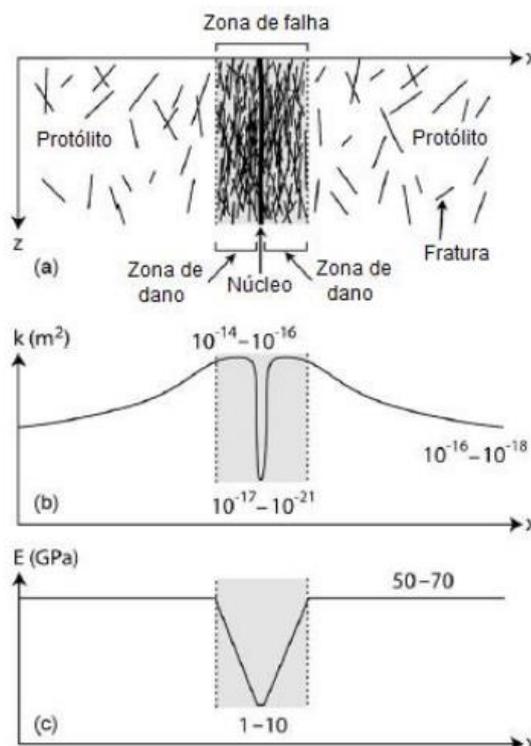
Fonte: Adaptado de Pereira et al. (2014)

O núcleo da falha compreende uma porção estrutural e morfológica da falha onde se concentram deslocamentos mais expressivos, comumente apresentando redes de fraturas e com uma extensão de cerca de metros a dezenas de metros, conforme ressalta Gudmundsson (2004). A redução do tamanho do grão e/ou a precipitação mineral presentes no núcleo levam a menores valores de porosidade e permeabilidade nessa região.

A zona de dano pode apresentar até quilômetros de extensão, compartilhando uma rede de estruturas secundárias que aumentam de intensidade à medida que se aproximam do núcleo, responsável por uma maior permeabilidade em relação ao núcleo. Essas estruturas secundárias incluem redes de falhas menores, cadeias de fraturas, veios, brechas e dobras que conferem heterogeneidade e anisotropia ao longo da extensão da zona de falha. Caine, Evans e Forster (1996) também relatam que uma ampla zona de dano pode indicar múltiplos episódios de deslizamento e a sobreposição de sucessivos eventos deformacionais.

As zonas de falhas são altamente heterogêneas e anisotrópicas ao longo de sua extensão, sendo objeto de inúmeras pesquisas o entendimento dessas propriedades ao longo da zona. Por exemplo, Cappa e Rutqvist (2011) discutem que o núcleo da falha é uma zona de baixa permeabilidade com pequena porosidade intergranular, enquanto a zona de dano é uma região mais permeável, como consequência de sua rede de fraturas macroscópica. Propriedades como permeabilidade e módulo de Young podem variar drasticamente ao longo da zona de falha, conforme mostra a Figura 3.

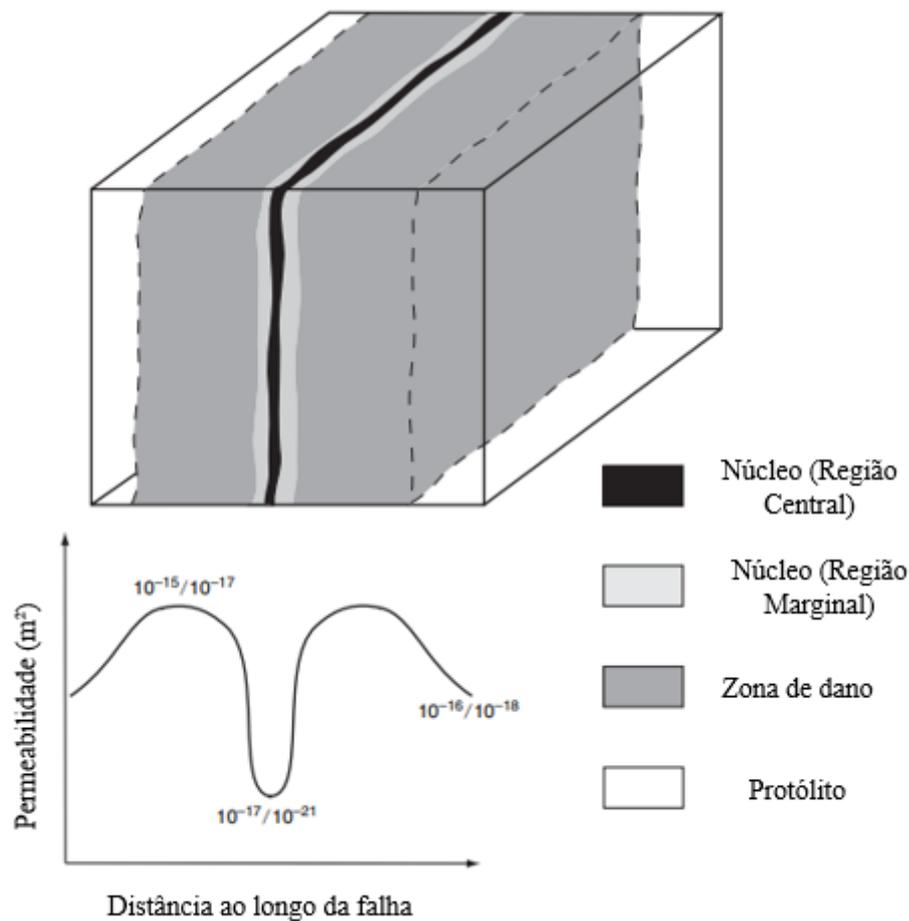
Figura 3-(a) Diagrama esquemático da estrutura através de uma zona de falha, (b) modelo conceitual da distribuição de permeabilidade, (c) módulo de Young.



Fonte: Adaptado de Cappa e Rutqvist (2011)

No trabalho de Boutareaud et al. (2008), os autores investigaram como a permeabilidade varia ao longo da zona de falha. Para isso, coletaram amostras de rocha da zona de falha e realizaram medições de permeabilidade em laboratório. Os resultados indicam que a permeabilidade varia ao longo da zona de falha, com algumas áreas mais permeáveis do que outras. Além disso, observaram que a permeabilidade é afetada pela estrutura da falha, como a presença de fraturas e estratificação da rocha. Uma seção esquemática da variação da permeabilidade apresentada pelos autores é mostrada na Figura 4.

Figura 4- Modelo esquemático de variação da permeabilidade ao longo da zona de falha.

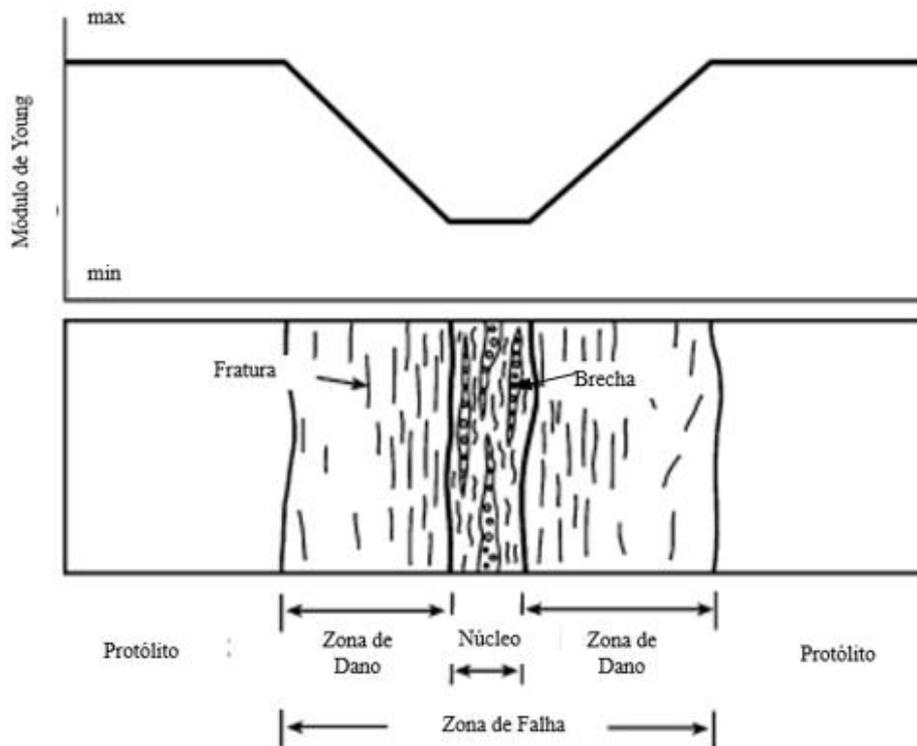


Fonte: Adaptado de Boutareaud et al. (2008)

Através do estudo de Gudmundsson (2004), também é possível observar como o desenvolvimento do núcleo da falha e da zona de dano podem afetar as propriedades mecânicas, em particular o foco do estudo dedicado ao estudo da variação do módulo de Young ao longo da extensão da zona de falha, na região onde

ocorre o deslocamento da falha. A distribuição do módulo de Young é representada desde a rocha hospedeira (protólito) até o núcleo da falha, conforme mostrado na Figura 5.

Figura 5- Distribuição do Módulo de Young ao longo da extensão da Zona de Falha.



Fonte: Adaptado de Gudmundsson (2004)

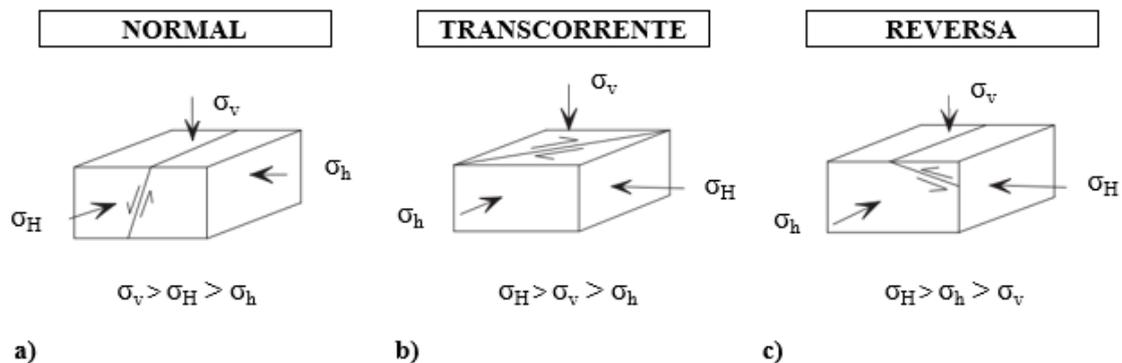
Para correta caracterização das zonas de falhas, é necessário também o entendimento dos tipos de falhas, que dependem do estado de tensão desse elemento. As tensões representam uma força aplicada sobre determinada área, capaz de induzir deformações, conforme relata Zoback (2007). Para categorizar o estado de tensão, tem-se a seguinte representação:  $\sigma_1$ ,  $\sigma_2$  e  $\sigma_3$ , como componentes da tensão em cada eixo, onde, geralmente,  $\sigma_1 > \sigma_2 > \sigma_3$ , ou  $\sigma_v$ ,  $\sigma_H$  e  $\sigma_h$ , como respectivamente tensão vertical, tensão horizontal máxima e tensão horizontal mínima.

Para classificar o tipo de falha, utiliza-se o modelo clássico proposto por Anderson (1951), onde as falhas podem ser rotuladas como: normais, reversas ou transcorrentes. Conforme o modelo, uma falha normal ocorre quando a tensão atuante principal é a tensão vertical e obtém-se o movimento descendente de um bloco em relação ao outro. Para o caso em que a tensão vertical está entre a horizontal máxima

e mínima, a falha é dita transcorrente e o movimento dos blocos é paralelo. Caso a tensão vertical seja menor que as horizontais, a falha é classificada como reversa, ou inversa, e observa-se o movimento ascendente relativo entre os blocos de rocha.

A classificação de Anderson (1951) baseia-se no estado de tensão dado pelas tensões principais. Para tanto as tensões principais horizontais são estabelecidas com base na tensão vertical. A tensão vertical,  $\sigma_v$ , é a tensão principal máxima ( $\sigma_1$ ) em regimes de falha normais, a tensão principal intermediária ( $\sigma_2$ ) em regimes de transcorrência e a mínima tensão principal ( $\sigma_3$ ) em regimes de falha reversa. O esquema com os respectivos tipos de falhas, podem ser vistos na Figura 6.

Figura 6- Teoria das falhas de Anderson relacionando falhas com as tensões principais: a) Falha Normal, b) Falha Transcorrente e c) Falha Reversa.



Fonte: Adaptado de Zoback (2007)

## 2.2 REATIVAÇÃO DE FALHAS

A reativação de falhas geológicas é um fenômeno em que falhas são novamente ativadas devido às mudanças de pressão e tensões *in situ* associadas à produção e injeção de fluidos em reservatórios, como descrito por estudos como os de Guimarães et al. (2010), Rutqvist (2012), Pereira et al. (2014), Pereira (2015), Gomes et al. (2016) e Lesueur, Poulet e Veveakis (2020), Abbassi et al. (2022), entre outros.

O processo de reativação de falhas é caracterizado por deformações cisalhantes que reduzem a capacidade selante da falha. A dilatância, que ocorre em conjunto com o cisalhamento, leva ao aumento do volume do material componente da falha e ao surgimento de permeabilidade. Como resultado, estruturas que eram

selantes passam a ser condutoras de fluidos, podendo também causar deformações superficiais, danos em tubulações, subsidência e exsudação de fluidos na superfície do fundo do mar.

Para minimizar esses problemas, é importante considerar a tendência de reativação das falhas durante as fases de planejamento e monitoramento dos campos. Isso pode ajudar a melhor definir a localização de poços, pressões operacionais e outras variáveis. Além disso, durante a fase de produção e recuperação, é importante estimar a pressão máxima de injeção para evitar a reativação de falhas e garantir a integridade das armadilhas de hidrocarbonetos.

O deslizamento e a reativação de falhas foram identificados como uma possível fonte de fluxo de fluido em áreas com falhas por Sibson (1994) e são controlados pelo critério de Mohr-Coulomb. Este fenômeno tem sido amplamente estudado e uma revisão completa pode ser encontrada em Zoback (2007). Muitos autores usam abordagens analíticas, semi-analíticas e/ou numéricas para representar o fenômeno de reativação de falhas geológicas (NACHT ET AL. (2010), PARK ET AL. (2020)).

As abordagens analíticas e semi-analíticas simplificam o problema para reservatórios homogêneos e de forma regular submetidos a pressão de poros constante, utilizando soluções fechadas que vão desde relações de falha de Mohr-Coulomb até soluções fundamentais de inclusões elásticas embutidas em um semiespaço (Zoback, 2007; Soltanzadeh e Hawkes, 2008; Soltanzadeh e Hawkes, 2009). Embora simplifiquem o acoplamento pressão-tensão dos poros em algum grau no reservatório e/ou na falha (Hillis, 2001), os procedimentos analíticos mais simples são as opções mais populares para uma análise inicial. O estudo do processo de reativação de falhas por meio de soluções analíticas e semi-analíticas tem sido amplamente discutido na literatura, incluindo trabalhos como Soltanzadeh e Hawkes (2008), Soltanzadeh e Hawkes (2009), Neves, Paiva e Luis (2009), Nacht et al. (2010), Siler, Hinz e Faulds (2018), Urpi et al. (2020) e Taghipour et al. (2020).

As abordagens numéricas podem melhor aproximar a geometria e a heterogeneidade de propriedades do campo e da falha. A falha é representada nos modelos discretos como uma superfície descontínua com deslocamentos relativos ou como uma banda contínua com propriedades enfraquecidas. As aplicações utilizam uma variedade de técnicas numéricas, como o método dos elementos finitos e o método das diferenças finitas. Muitos trabalhos utilizam simulação numérica para avaliar o fenômeno de reativação, incluindo os de Nacht et al. (2010), Guimarães et

al. (2010), Pereira et al. (2014), Gomes et al. (2016), Urpi et al.(2020) e Taghipour et al. (2020).

Soltanzadeh e Hawkes (2008) utilizaram a teoria das inclusões para encontrar os campos de tensão dentro e ao redor de reservatórios horizontais e inclinados no plano da seção transversal, com geometrias de seção transversal elíptica e retangular, em um semi-espaço plano de deformação. As equações fornecidas incluem soluções para reservatórios de qualquer profundidade ou dimensão, desde que suas geometrias sejam semelhantes às geometrias do modelo idealizado. As taxas de arqueamento de tensão normalizadas calculadas usando essas equações foram mapeadas para diferentes geometrias e profundidades de reservatórios usando parâmetros adimensionais. Os gráficos obtidos podem ser usados para estimar as mudanças de tensão induzidas em reservatórios. Além disso, foram apresentadas equações para avaliar a tendência de reativação de falhas resultante de mudanças de tensão induzidas usando o método Mudança de Tensão de Falha de Coulomb. Foi demonstrado que o potencial de reativação de falhas depende da geometria do reservatório e do ângulo de mergulho, sendo importante considerar a geometria real do reservatório na análise da sismicidade induzida e da integridade da rocha capeadora. Neves, Paiva e Luis (2009) desenvolveram uma ferramenta gráfica para realizar a análise de reativação de falhas.

Soltanzadeh e Hawkes (2009) aprofundaram seus estudos para investigar a tendência de reativação de falhas em reservatórios durante a injeção ou produção de fluidos. A análise de tensão induzida é realizada usando a teoria de inclusões de Eshelby para um material poroelástico, e o conceito de mudança de tensão avaliado pelo Critério de Coulomb é utilizado para determinar a possibilidade de reativação de falha. Os resultados mostram que, durante a produção de um reservatório em regime normal de tensão de falha, é provável que ocorra reativação de falha dentro do reservatório e nas áreas adjacentes aos seus flancos. Análises de sensibilidade mostram que as localizações dos contornos que definem essas regiões de tendência de reativação (ou estabilização) não são altamente sensíveis ao coeficiente de atrito da falha.

Nacht et al. (2010) avaliaram a reativação e abertura de falhas sob um regime de tensão normal, aplicando uma abordagem analítica simplificada correspondente ao critério de Mohr-Coulomb, assim como simulações numéricas através de elementos finitos, utilizando o modelo elastoplástico de Mohr-Coulomb para estabelecer a

relação tensão-deformação. Realizaram uma análise paramétrica para avaliar a influência do ângulo de mergulho da falha, espessura e profundidade do reservatório no incremento de pressão de poro necessário para induzir o fenômeno de reativação. A diferença entre os resultados do modelo analítico e numérico levou os autores a sugerir que a deformação causada pelo acoplamento pressão de poro-tensão seria imprescindível para avaliar os cenários, não podendo ser desconsiderada. Os autores recomendam a abordagem analítica utilizada como uma ferramenta poderosa para uma análise preliminar, que pode ser usada em conjunto com um simulador de fluxo para identificar cenários críticos em um campo. De acordo com as análises, o parâmetro espessura do reservatório tem uma pequena influência quando comparado com as variáveis ângulo de mergulho da falha e profundidade do reservatório.

Guimarães et al. (2010) simularam a injeção de CO<sub>2</sub> em um aquífero salino atravessado por um sistema de impermeabilização geológica, considerando o acoplamento hidromecânico. Dois modelos constitutivos mecânicos alternativos foram usados para representar o comportamento do material da falha: o modelo elastoplástico de Mohr-Coulomb (com dilatação considerada) e o modelo de dano isotrópico contínuo (com microfissuração considerada), o que levou a diferentes respostas do sistema de falha de reservatório. Os autores observaram que, quando o modelo de dano contínuo é usado, a reativação é mais intensa (todas as falhas foram reativadas), porém mais lenta. Eles também avaliaram a máxima pressão de injeção para evitar a abertura das falhas e sugerem que a definição dessa variável deve ser realizada por meio de uma análise hidromecânica acoplada, que representa de forma realista o comportamento constitutivo dos materiais.

Pereira et al. (2014) apresentaram um estudo de reativação de falha em um cenário de reservatório de petróleo usando uma estrutura integrada baseada em simulação numérica hidromecânica acoplada e quantificação de incerteza para propriedades de materiais de zona de falha aplicando a teoria da evidência. Os autores consideraram a heterogeneidade presente na zona de falha e estimaram as pressões de injeção máximas permitidas para evitar a reativação da falha.

Gomes et al. (2016) realizaram simulação hidromecânica e usaram o método de elementos finitos para modelar o escoamento de fluido em um reservatório deformável atravessado por falha geológica, representando um sistema geológico real. O esquema de acoplamento foi realizado de forma totalmente acoplada e implícita, considerando as zonas de falhas compostas por um único material e

divididas em núcleo e zonas de dano. O modelo elastoplástico de Mohr Coulomb foi utilizado para modelar o comportamento tensão/deformação dos materiais componentes da falha. Assim, através de simulação numérica, os autores avaliaram o fenômeno e determinaram as pressões máximas para evitar a reativação. Eles também ressaltaram a importância de considerar a variação das propriedades da zona de falha no processo de reativação.

Siler, Hinz e Faulds (2018) avaliaram a influência do deslizamento de falhas no desenvolvimento da permeabilidade e fluxo de fluido em descontinuidades estruturais, bem como as escalas de tempo relativas de recorrência de terremotos e o processo de mineralização hidrotermal. Para isso, modelaram mudanças de tensão através do modelo analítico de Coulomb como resultado do deslizamento ao longo de zonas de falhas, que controlam os sistemas geotérmicos em Salt Wells e Bradys, Nevada. Os resultados indicam que o fluxo de fluido geotérmico está focado nas mesmas áreas onde se encontram as regiões mais tensionadas. O deslizamento sobre essas falhas e fraturas secundárias gera permeabilidade por deformação cinemática e permite a transmissão de fluidos. No entanto, ainda deve-se considerar a mineralização que reduz a permeabilidade ao longo do tempo.

Em Urpi et al. (2020), realizaram a modelagem de reativação de falhas hidromecânicas baseada na teoria da elasto-plasticidade. Para verificar o submodelo mecânico, os dados foram comparados com a solução analítica. Já o comportamento hidromecânico acoplado foi validado com dados de campo e incluiu a validação do modelo e código criados. O modelo reproduziu a presença de um poço que injetava diretamente em uma estrutura de falha de baixa permeabilidade embutida em um maciço rochoso impermeável e intacto. A partir dos dados das simulações após a ruptura, observou-se um aumento súbito de pressão no ponto de observação, resultante do aumento da permeabilidade devido ao cisalhamento, que permitiu o fluxo. Os deslocamentos finais mostraram discrepâncias em relação aos dados de campo, que se devem tanto ao deslocamento de cisalhamento quanto às diferentes condições de pressão. As discrepâncias entre os resultados numéricos decorrem de incertezas nas propriedades da falha (rigidez normal e de cisalhamento, coesão, comportamento da deformação) e na caracterização da estrutura da falha, bem como da negligência da anisotropia na resistência e rigidez da matriz de argila. A abordagem apresentada permite a solução monolítica de problemas poromecânicos 3D na presença de falhas e fraturas de grande escala.

Taghipour et al. (2020) utilizaram dados como registros de poços, estimativas de pressões de poros e resultados de testes mecânicos de rochas para simular numericamente e verificar analiticamente o reservatório de Asmari no campo petrolífero de Gachsaran, no sudoeste do Irã. Para isso, o campo de tensões foi determinado usando módulos elásticos de rochas reservatório e resultados de testes de integridade de formação através de relações estabelecidas. Em seguida, aplicaram métodos analíticos, como os diagramas de Mohr Coulomb, para analisar a possibilidade de reativação de quatro falhas (F1, F2, F3 e F4) no campo devido à injeção de gás e para estimar as pressões máximas de poros sustentáveis. Usando simulações numéricas no software ABAQUS, investigaram o caminho do fluxo do gás injetado, o potencial de vazamento através da rocha de cobertura, a possível reativação de falhas devido à injeção de gás e o acúmulo de tensões de cisalhamento e o desenvolvimento de deformações plásticas em diferentes partes do reservatório. Os resultados dos diagramas de Mohr mostraram que todas as falhas estão estáveis, e a falha F2 tem potencial de sustentar uma pressão máxima de poros de 55 MPa no campo. Já a falha F3 é mais suscetível à reativação (em relação à orientação da tensão horizontal máxima). Os resultados das simulações numéricas sugeriram que uma pressão de 30 MPa não causaria novas fraturas ou reativação de falhas em um período de 5 anos após a injeção. Também foi demonstrado que, aplicando uma pressão de injeção de 60 MPa, o deslizamento ocorreria na falha F4 após 10 dias de injeção. Além disso, afirma-se que uma das principais incertezas relacionadas aos métodos analíticos são as propriedades do sistema de rocha reservatório-cobre, como a permeabilidade, que desempenha um papel significativo no padrão de fluxo do gás injetado e determina a localização das deformações plásticas acumuladas.

### 2.3 APLICAÇÕES DE TÉCNICAS DE ML EM RESERVATÓRIOS DE PETRÓLEO

O aprendizado de máquina pode ser aplicado em várias etapas do processo de exploração e produção de petróleo, tais como a identificação de novas reservas, a avaliação de campos existentes, a otimização da produção de óleo e a previsão de parâmetros operacionais, entre outras possibilidades. Além disso, de acordo com Motamedi e Farchad (2020), Abdelghany et al. (2022), Jafarizadeh, Motamedi e Ahmadi (2021) e Abbassi et al. (2022), o uso de técnicas de ML tem se tornado cada

vez mais comum para lidar com incertezas em vários problemas da engenharia de petróleo.

Através da revisão bibliográfica, é possível encontrar inúmeras aplicações de modelos ML em reservatórios de petróleo. Alguns exemplos incluem o uso para estimativa de parâmetros de reservatórios (Alaei e Alaei, 2013; Banerjee e Chatterjee, 2021), no processo de injeção de fluidos e sismicidade associada (He, Li e Li, 2020), na otimização de perfuração de poços (Oyedere e Gray, 2020), na detecção e interpretação de falhas (Hamzeh, 2021; Yuan et al., 2019; Gao et al., 2021; Zhu et al., 2021), bem como em aplicações relacionadas à reativação de falhas (Ahmadi Goltapeh et al., 2022; Abbassi et al., 2022).

No estudo de Alaei e Alaei (2013), é observado o uso de técnicas de análise de dados e redes neurais para estimar parâmetros de reservatórios de petróleo. Os autores buscaram prever os parâmetros do reservatório de Marun, localizado em Ahwaz, Irã, utilizando dados geofísicos disponíveis. Para isso, os autores descreveram um método integrado que combina a Análise de Componentes Principais (Principal Component Analysis (PCA)), a LDA e redes neurais para estimar propriedades do reservatório com base em dados geofísicos. Eles argumentam que este método é superior aos individuais, pois a PCA e a LDA são usadas para selecionar as melhores variáveis de entrada para a rede neural e para reduzir a dimensionalidade dos dados. Os autores testaram o método em um conjunto de dados de reservatório de petróleo e compararam os resultados obtidos com os previstos usando apenas uma RNA ou apenas a PCA e a LDA. Eles concluíram que o método integrado produziu resultados mais precisos do que os métodos individuais.

No trabalho de Motamedi e Farchad (2020), observa-se uma revisão das vantagens, desvantagens, aplicações e desafios das técnicas de ML mais utilizadas em aplicações na Engenharia de Petróleo. São citadas diversas aplicações, incluindo operação de perfuração de poços, caracterização de reservatórios e interpretações petrofísicas. Portanto, a partir da revisão, destaca-se a importância dos métodos de análise de dados na maximização de lucro, avaliação de incertezas e minimização de riscos em uso complementar com as práticas operacionais, servindo como suporte baseado em *Big Data*, que fornecem subsídios para as decisões futuras, através do ajuste histórico dos dados.

A pesquisa de Jafarizadeh, Motamedi e Ahmadi (2021) demonstra a viabilidade de aplicação de algoritmos de aprendizado de máquina orientados a dados para

gerenciamento integrado de reservatórios. O objetivo do artigo é desenvolver uma relação com a aplicação da abordagem de análise de dados na geomecânica de reservatórios. Para isso, os autores utilizam o caso da análise da injeção de fluidos de recuperação aprimorada de petróleo e analisam a aplicação de técnicas baseadas em ML para determinação da escolha da técnica com base em dados anteriores, sugerindo a utilização de Redes Bayesianas para esse processo, justificado por sua implementação associada com incertezas.

Em He, Li e Li (2020), é observada uma visão geral sobre o progresso atual da pesquisa sobre sismicidade induzida relacionada à injeção de fluidos e é proposta uma estrutura básica de uso de ML para melhorar o *Traffic Light System*, que é comumente utilizado para avaliar cenários de sismicidade induzida. Ao comparar com abordagens estatísticas, o modelo baseado em ML, com uma grande quantidade de dados, pode refletir de forma mais abrangente a complexa relação entre parâmetros operacionais e sísmicos. Além disso, ele pode ser aplicado a vários terremotos induzidos por injeção, coletando dados característicos de diferentes atividades sísmicas. No entanto, os autores sugerem que a coleta completa de dados relevantes é o principal desafio.

No estudo de Oyedere e Gray (2020), é tratada a otimização da perfuração de poços por meio de previsões da taxa de penetração (*Rate of penetration* (ROP)) e do torque na broca (*Torque-on-bit* (TOB)). ROP e TOB são modelados com base no peso da broca, taxa de fluxo, velocidade rotativa e resistência à compressão não confinada. Neste artigo, foram implementados cinco algoritmos diferentes de aprendizado de máquina de classificação: regressão logística, LDA, análise discriminante quadrática (*Quadratic Discriminant Analysis* (QDA)), máquinas de vetores de suporte (*Support Vector Machines* (SVM)) e floresta aleatória, para desenvolver o modelo de classificação. Usando a área sob a curva (*Area Under Curve* (AUC)) como a métrica de desempenho de classificação, os resultados das simulações mostraram que o melhor classificador deve ser escolhido com base no tipo de formação, obtendo altas taxas de classificação com AUC variando de 0,85 a 1,00 para os métodos avaliados. Além disso, foi desenvolvida uma ferramenta gráfica para aplicação prática desta abordagem na previsão de ROP e TOB, que pode ajudar o perfurador a tomar decisões sobre as combinações de valores de velocidade rotativa e peso da broca que resultariam nas regiões desejadas de ROP e TOB.

Banerjee e Chatterjee (2021) realizaram modelagem da pressão de poros e determinação da tensão *in situ* na bacia de Raniganj, Índia. Uma nova abordagem é executada para mapear a pressão de poros subsuperficial através da implementação de arquitetura de Rede Neural, integrando dados sísmicos e de perfil de poço. Subsequentemente, a direção da tensão é analisada a partir do perfil de imagem e o regime de falha predominante é estimado usando a teoria de falha. O método utilizado para obtenção da pressão de poros baseado em Rede Neural apresenta boa concordância com os valores de medições diretas para os poços analisados. Além disso, a análise de tensão indica um regime de tensão de falha normal, onde as falhas se encontram estáveis com base na teoria de falha por atrito utilizada na análise. O resultado deste estudo ajuda a entender a distribuição de poropressão e perfis de tensão, mitiga o risco de perfuração associado, prevê o limite de pressão para a reativação de falhas existentes durante a injeção de fluido em tratamentos de fraturamento hidráulico e também fornece insumos vitais para futuros estudos geomecânicos.

Siler et al. (2021) utilizaram de técnicas de ML não supervisionado para identificar os principais fatores geológicos que contribuem para a produção geotérmica no campo geotérmico de Brady, Nevada. O campo de Brady é um sistema hidrotermal no noroeste de Nevada que suporta tanto a produção de eletricidade quanto o uso direto de fluidos hidrotermais. Assim, o estudo analisa dados geológicos juntamente com métodos de ML para desvendar os controles geológicos locais nos caminhos de fluxo. O método de ML, fatoração matricial não negativa com agrupamento *k-means*, é aplicado a uma biblioteca de 14 características geológicas 3D hipotetizadas para controlar a circulação hidrotermal no campo geotérmico de Brady. A partir dos resultados, observa-se que as falhas de macroescala e um *step-over* local no sistema de falhas ocorrem preferencialmente ao longo de poços produtores quando comparados aos poços injetores e a poços não produtivos, sendo características que controlam as vias de transmissão hidrotermal em Brady. Os resultados demonstram controles geológicos no sistema hidrotermal de Brady, eficácia do emparelhamento de técnicas de ML com caracterização geológica 3D para melhorar a compreensão dos processos em subsuperfície.

As aplicações associadas à detecção e interpretação de falhas são discutidas no trabalho de Hamzeh (2021), que é uma revisão bibliográfica com o objetivo de fornecer uma atualização sobre o estado atual das técnicas de aprendizado de

máquina na detecção de falhas sísmicas. Para isso, a revisão inclui a discussão sobre o uso de técnicas de ML para detecção e interpretação de falhas, revisando 79 artigos que demonstram a importância e a eficácia dessas técnicas para a interpretação de falhas. É possível notar a predominância do uso de redes neurais convolucionais para interpretação de falhas. As pesquisas de Yuan et al. (2019), Gao et al. (2021) e Zhu et al. (2021) também se dedicam à interpretação de falhas através do uso de aprendizado de máquina.

Apesar do amplo crescimento da utilização de aprendizado de máquina em pesquisas envolvendo reservatórios de petróleo, há pouca informação sobre sua aplicação específica na análise de reativação de falhas, apesar de haver algumas aplicações nesta área, como os estudos apresentados por Ahmadi Goltapeh et al. (2022) e Abbassi et al. (2022).

Ahmadi Goltapeh et al. (2022) avaliaram a confiabilidade estrutural da rocha capeadora usando uma rede neural para analisar o local de injeção de CO<sub>2</sub> em um cenário exemplo do Mar do Norte. O conjunto de treinamento do modelo inclui parâmetros de resistência, coesão ( $c$ ) e ângulo de atrito ( $\varphi$ ), bem como dados de tensão de estado inicial (geostática), relacionados às tensões efetivas máxima e mínima principais ( $\sigma'_1$  e  $\sigma'_3$ ). O objetivo do procedimento foi classificar as reativações de falhas presentes na rocha capeadora com base na determinação de fatores de segurança avaliados pelo critério de resistência da rocha utilizando o modelo de Mohr Coulomb. A arquitetura da rede desenvolvida no trabalho foi baseada em tentativas de minimizar o tempo de aprendizado e aumentar a precisão do modelo. Na etapa de definição da arquitetura da rede neural, três camadas foram suficientes para criar a rede, sendo a primeira camada composta por 14 neurônios, a segunda camada com 10 neurônios e a última camada com apenas um neurônio. A última camada foi ativada usando a função sigmoide, enquanto as outras camadas usaram a função de ativação da Unidade Linear Retificada (ReLU) (Rumelhart et al., 1986), tratando-se de uma classificação binária. Portanto, o modelo de rede neural produzido pelos autores consiste em um conjunto de entrada contendo quatro atributos ( $\varphi$ ,  $c$ ,  $\sigma'_1$  e  $\sigma'_3$ ) usados para obter a função de classificação  $g(x)$  que pode indicar 1 (indicando não-falha) ou 0 (indicando falha). De acordo com os resultados, observa-se que ao aumentar o tamanho dos dados de entrada, as redes neurais convergem em épocas mais curtas, aumentando também a precisão do modelo. Os autores também relatam a

necessidade de investigações de campo e experimentais para melhor definir o comportamento da rocha capeadora, visando aprofundamento e melhoria na implementação do modelo de redes neurais

No estudo de Abbassi et al. (2022), investigou-se a reativação de falhas pré-existentes devido à injeção de fluidos em reservatórios de hidrocarbonetos sob diferentes pressões e temperaturas. Para isso, foi utilizado um modelo 3D com uma falha normal contínua que divide o domínio em dois compartimentos. O modelo constitutivo foi definido pelos autores e baseou-se na mecânica do dano contínuo implementada como uma sub-rotina em Fortran. Foi realizada uma análise paramétrica para examinar a influência de parâmetros geométricos, como o ângulo de inclinação da falha, as características do reservatório e os parâmetros de injeção de fluido. Além disso, foi utilizado um método de aprendizado de máquina baseado em RNA para prever a recuperação aprimorada do petróleo usando a injeção de fluido. Os resultados previstos pela RNA foram posteriormente confirmados pela modelagem numérica. A análise paramétrica também investigou os efeitos do ângulo de inclinação da falha, da permeabilidade do reservatório, da localização do injetor e do extrator e da magnitude da velocidade de injeção de fluido na velocidade de extração de petróleo e no potencial de reativação. Os resultados indicaram que a permeabilidade da falha tem uma influência significativa na reativação da falha. Além disso, foi mostrado que o ângulo de inclinação da falha determina diretamente o regime de tensão compressiva *in situ* através da magnitude da velocidade de injeção de fluido e controla a reativação da falha.

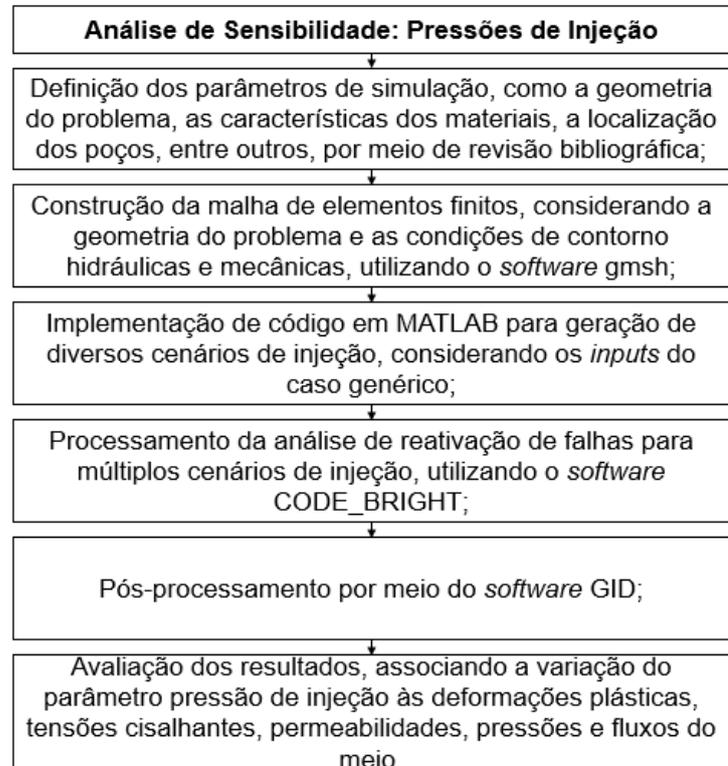
Em Abbassi et al. (2022), os autores afirmam que a RNA oferece uma abordagem nova e alternativa para investigar a reativação de falhas pré-existentes durante a injeção de fluido em reservatórios de hidrocarbonetos. Essa abordagem tem a vantagem de ser eficaz em termos de custo, pois o tempo necessário é limitado em comparação com a simulação de elementos finitos, sem comprometer a precisão. O estudo demonstra que a RNA pode ser usada para prever a ativação de falhas com a entrada de parâmetros adequados que influenciam o processo e pode atuar como uma abordagem prática e robusta de generalização.

### 3 METODOLOGIA

As etapas correspondentes à análise de sensibilidade do parâmetro pressão de injeção no processo de reativação de falhas são descritas na Figura 7. As simulações numéricas hidromecânicas associadas à variação do campo de pressão no interior do reservatório são divididas em pré-processo, processo e pós-processo. Um reservatório arenítico consolidado, conforme descrito por Pereira et al. (2014), teve seu pré-processo definido, incluindo a geometria, os parâmetros dos materiais e as condições de contorno, utilizando o software gmsk (Geuzaine e Remacle, 2009). É importante destacar que o modelo constitutivo de Drucker-Prager foi utilizado para modelar o comportamento tensão-deformação dos materiais que compõem a falha.

Por meio de um código em MATLAB, o parâmetro pressão de injeção foi alterado nos arquivos de *inputs*, gerando múltiplas análises de reativação com variação desse parâmetro. Os arquivos de *inputs* alimentam o código de elementos finitos e, após o processamento com o programa *in house* de elementos finitos CODE\_BRIGHT (COupled DEformation, BRine, Gas and Heat Transport) (Olivella et al., 1994; Olivella et al., 1996), os arquivos de saída são gerados, contendo a variação das incógnitas (deformação plástica, fluxo e pressão de fluido) e outros parâmetros de interesse (tensões, permeabilidade, etc). Os resultados são visualizados através do pós-processamento com o software GID (Vaunat e Olivella, 2002).

Figura 7- Fluxograma de Etapas da Análise de Sensibilidade do Processo de Reativação de Falhas Associado ao Parâmetro Pressão de Injeção.



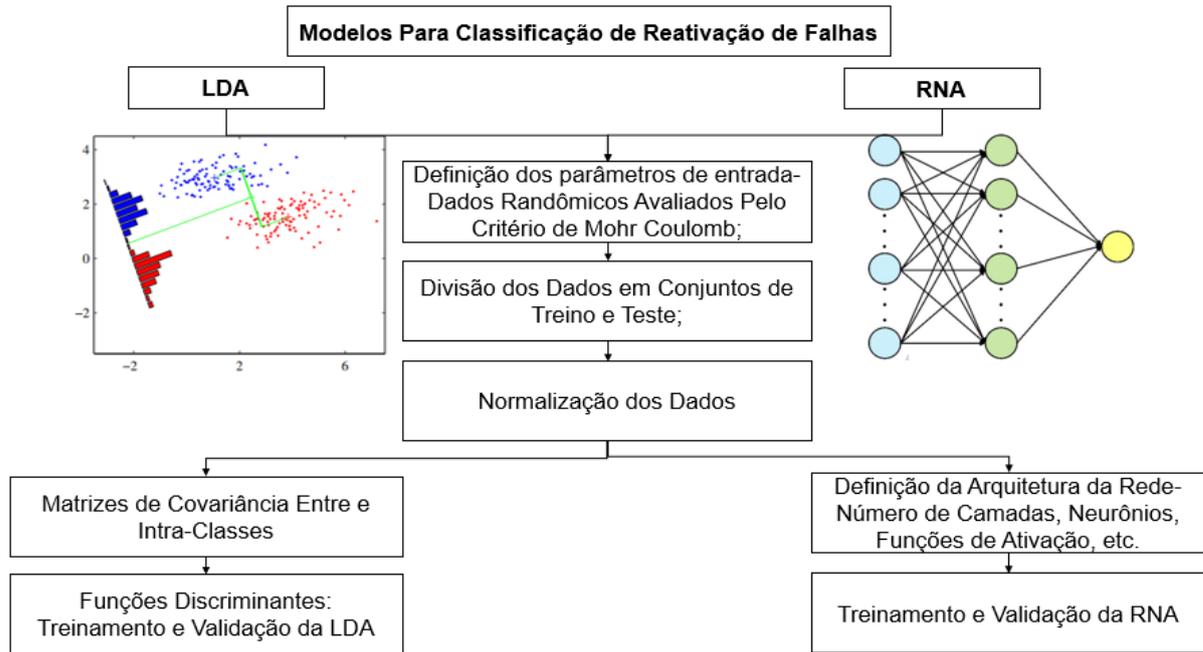
Fonte: A Autora (2022)

O presente trabalho também apresenta aplicações de modelos de classificação LDA e RNA para avaliar possíveis cenários de reativação de falhas geológicas. Ambas as técnicas foram aplicadas com o uso de códigos em Python, utilizando bibliotecas amplamente utilizadas, como Keras e Sklearn. Keras é uma biblioteca de aprendizado de máquina de código aberto escrita em Python, desenvolvida por François Chollet, um engenheiro de aprendizado de máquina da Google. Ela permite que os usuários construam e treinem modelos de aprendizado profundo de forma rápida e fácil, sendo amplamente utilizada em contextos como visão computacional, processamento de linguagem natural e reconhecimento de fala. Sklearn, por outro lado, é também uma biblioteca de aprendizado de máquina de código aberto escrita em Python, desenvolvida por vários cientistas de dados da INRIA (Instituto Nacional de Pesquisa em Tecnologia da Informação e Automática) e outras instituições de pesquisa (Buitinck et al. (2013)). Ela fornece uma variedade de algoritmos de aprendizado supervisionado e não supervisionado, incluindo regressão e classificação, e é amplamente utilizada em análise de dados, mineração de dados e ciência de dados.

Devido à falta de dados experimentais, os parâmetros necessários para treinar e validar os modelos foram gerados a partir de dados gerados aleatoriamente dentro de intervalos específicos de suas propriedades. Posteriormente, esses parâmetros foram avaliados por um modelo analítico que atende ao critério de Mohr Coulomb e representam diversos cenários de reativação de falhas. Para ambas as técnicas, é necessário realizar etapas prévias à sua implementação, como dividir os dados em conjuntos de treino e teste, pois são modelos de aprendizado de máquina que aprendem com os dados. Além disso, foi realizada a etapa de pré-processamento dos dados, incluindo a normalização dos parâmetros, para facilitar a criação de exemplos de aprendizado de máquina.

Para o LDA, o processo consiste em determinar parâmetros estatísticos dos dados, como as matrizes de covariância entre e intra-classes, que permitem definir funções discriminantes. O modelo é então validado com dados de teste e tem classificação baseada na função mencionada. Já para a RNA, o processo segue com a definição da arquitetura da rede, incluindo o número de camadas, neurônios e funções de ativação. Em seguida, a rede é treinada com os dados de treinamento e seus hiperparâmetros (número de camada, tipo de camadas, número de neurônios nas respectivas camadas, funções de ativação, etc) são otimizados para um desempenho mais eficaz. Finalmente, a validação é realizada com dados de teste que permitem validar o modelo de RNA. Ambas as técnicas são detalhadamente descritas na Figura 8.

Figura 8 - Fluxograma de Atividades Para os Modelos de Classificação (LDA e RNA) de Cenários de Reativação de Falhas



Fonte: A Autora (2022)

### 3.1 ANÁLISE DE SENSIBILIDADE

Nesta seção, detalha-se a análise de sensibilidade, baseada em simulação numérica. É descrita a formulação matemática do problema acoplado hidromecânico, bem como a descrição do modelo numérico e as propriedades dos materiais estabelecidas.

#### 3.1.1 Formulação Matemática

A formulação matemática que embasa o problema de reativação de falhas é descrita de maneira simplificada, mas é possível observar uma formulação detalhada através dos trabalhos de Pereira (2015) e Cisneros (2016).

O problema mecânico é definido pela equação de equilíbrio (Equação 1):

$$\nabla \cdot \boldsymbol{\sigma} + \mathbf{b} = \mathbf{0} \quad (1)$$

Onde  $\nabla$  é o operador nabla,  $\boldsymbol{\sigma}$  é o tensor de tensões totais e  $\mathbf{b}$  o vetor de forças de corpo.

Considerando o meio poroso saturado apenas com um único fluido, neste trabalho corresponde ao fluxo monofásico de água, o balanço de massa de líquidos é dado por (Equação 2):

$$\frac{\partial(\phi\rho_l)}{\partial t} = \nabla \cdot (\rho_l \mathbf{q}_l + \phi\rho_l \dot{\mathbf{u}}) = 0 \quad (2)$$

Onde  $\phi$  é a porosidade e  $\rho_l$  é a densidade do líquido. O líquido é considerado levemente compressível e sua densidade depende da pressão do líquido ( $P_l$ ). O fluxo de Darcy é definido por  $\mathbf{q}_l$  e  $\dot{\mathbf{u}}$  é a velocidade da fase sólida.

Partindo da equação da conservação de massa de sólidos e aplicando o conceito de derivada material, conforme Olivella, *et al.*, (1996), a variação na porosidade é expressa como (Equação 3):

$$\frac{D\phi}{Dt} = \frac{(1-\phi)}{\rho_s} \frac{D\rho_s}{Dt} + (1-\phi)\varepsilon_v \quad (3)$$

onde  $\rho_s$  é a densidade da fase sólida, que depende da compressibilidade dos grãos dos sólidos que compõe a rocha, e  $\varepsilon_v$  é a deformação volumétrica total.

As de tensões efetivas ( $\boldsymbol{\sigma}'$ ) são responsáveis pela deformação do meio, como determinado pelo princípio de tensão efetiva de Biot, segundo a Equação 4:

$$\boldsymbol{\sigma}' = \boldsymbol{\sigma} - \alpha I P_l \quad (4)$$

onde  $\boldsymbol{\sigma}$  é o do tensor de tensões totais e  $I$  o tensor identidade, no qual os componentes na diagonal principal são 1 e as demais componentes são 0,  $\alpha$  é a constante de Biot, que leva em conta a compressibilidade da fase sólida, e  $P_l$  corresponde a pressão de fluido.

A relação entre as variações do tensor de tensões efetivas e o tensor de deformações são calculados conforme a Equação 5:

$$d\boldsymbol{\sigma}' = \mathbf{D}^{ep} \cdot d\boldsymbol{\varepsilon} \quad (5)$$

Onde  $d\boldsymbol{\sigma}'$  corresponde ao incremento do tensor de tensões efetivas,  $\mathbf{D}^{ep}$  ao tensor constitutivo elastoplástico e  $d\boldsymbol{\varepsilon}$  o incremento do tensor de deformações.

O modelo constitutivo adotado no trabalho é baseado no critério de Drucker-Prager, onde o conjunto de equações (Equações 6 e 7) para a função de fluência são apresentadas a seguir:

$$F(\sigma, k) = J - \left( \frac{c'}{\tan \varphi'} + p' \right) G(\theta) \leq 0 \quad (6)$$

$$G(\theta) = \frac{\sin \varphi'}{\cos \theta + \frac{\sin \theta \cdot \sin \varphi'}{\sqrt{3}}} \quad (7)$$

Tem-se que  $F(\sigma, k)$  corresponde a função de fluência para o critério de ruptura de Drucker-Prager,  $J$  é a tensão desviadora,  $p'$  a tensão efetiva média,  $c'$  a coesão efetiva,  $\varphi'$  é o ângulo de atrito efetivo e  $\theta$  é o ângulo de Lode.

Para o problema hidráulico, este trabalho assume que a permeabilidade intrínseca do meio pode variar em função da deformação plástica cisalhante através da lei exponencial a seguir (Equação 8):

$$\mathbf{k} = \mathbf{k}_i e^{c \frac{e_d^p}{e_d^p/máx}} \quad (8)$$

Onde  $e_d^p$  corresponde a deformação plástica cisalhante,  $e_d^p/máx$  a deformação plástica cisalhante máxima (0,00005),  $\mathbf{k}_i$  o tensor de permeabilidade intrínseca inicial. O parâmetro  $c$  controla o efeito da deformação plástica cisalhante sobre a permeabilidade, neste trabalho adota-se  $c=4$ .

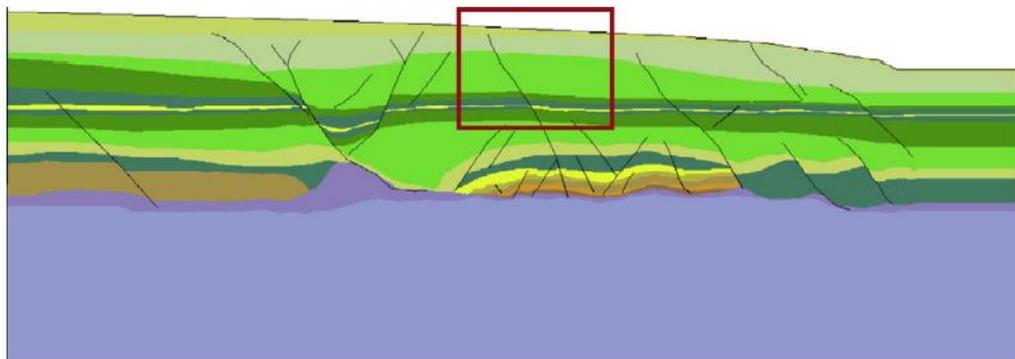
### 3.1.2 Modelo Numérico

Para analisar a reativação de falhas, foi proposto um modelo numérico para simulações computacionais no software CODE\_BRIGTH. O programa de elementos finitos CODE\_BRIGTH resolve o problema de fluxo de fluidos em meios porosos sensíveis ao estado de tensões de maneira acoplada. As equações do problema de fluxo são resolvidas conjuntamente com a equação de equilíbrio do problema mecânico em um esquema totalmente acoplado, onde, a cada iteração de Newton-Rapshon, é solucionado um único sistema de equações para toda a malha, nas quais

as incógnitas são a pressão (problema hidráulico) e o deslocamento (problema mecânico) (OLIVELLA ET AL, 1996). É importante destacar que o modelo constitutivo elastoplástico de Drucker-Prager foi utilizado para modelar o comportamento tensão-deformação dos materiais que compõem a falha.

As análises numéricas concentram-se no comportamento de uma zona de falha. Na figura 9, tem-se a representação de uma seção geológica típica de um reservatório de petróleo. O retângulo presente na figura delimita a seção de estudo, conforme compilado de Pereira et al. (2014).

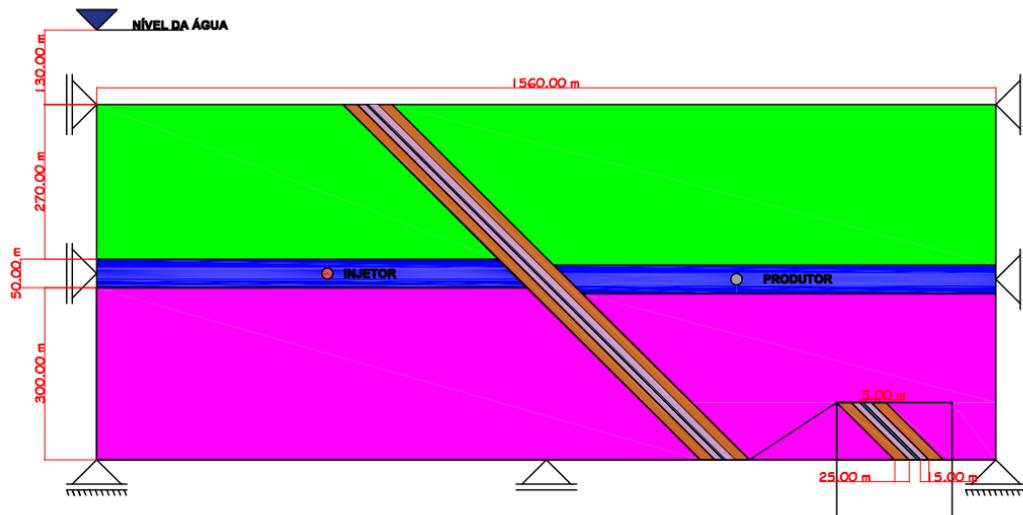
Figura 9- Seção Geológica Bidimensional Típica



Fonte: Pereira et al. (2014)

O reservatório consiste em uma camada central, a rocha superior ao reservatório denomina-se *overburden* e a rocha inferior como *underburden*. Introduziu-se dois poços ao sistema, um poço injetor na porção esquerda do reservatório e um poço produtor na direita. O topo da camada *overburden*, está a 130 metros de profundidade do nível da água. O bloco possui dimensões de 1560 metros de largura por 620 metros de altura. A rochas de topo e a base possuem, respectivamente, 270 metros e 300 metros de espessura e o reservatório 50 metros. A falha encontra-se dividida em seus componentes, que consiste em núcleo com 5 metros de espessura, zona de dano interna com 15 metros de espessura, zona de dano externa com 25 metros de espessura, que totalizam 85 metros de extensão para seguinte zona de falha (Figura 10). O modelo geológico, compilado de Pereira et al.(2014), trata-se de dados reais da Bacia de Campos, Rio de Janeiro, caracterizando o reservatório um arenito consolidado de 50 m de espessura, coberto por um depósito de argila de 270 m de espessura.

Figura 10-Esquemática da Geometria do Modelo.



Fonte: A Autora (2022)

Para analisar a sensibilidade, foi desenvolvida uma rotina utilizando o software MATLAB para alterar o parâmetro de sensibilidade estudado, que corresponde à pressão de injeção. Os valores de pressão de injeção de fundo de poço variam entre 19,0 MPa e 22,0 MPa, com um incremento de 0,1 MPa no intervalo, conforme apresentado na Tabela 1. Dessa forma, o parâmetro de pressão de injeção é associado à sensibilidade das alterações da deformação plástica, tensão cisalhante, permeabilidade, bem como à pressão e fluxo de fluidos do meio.

Tabela 1 – Pressões de injeção analisadas

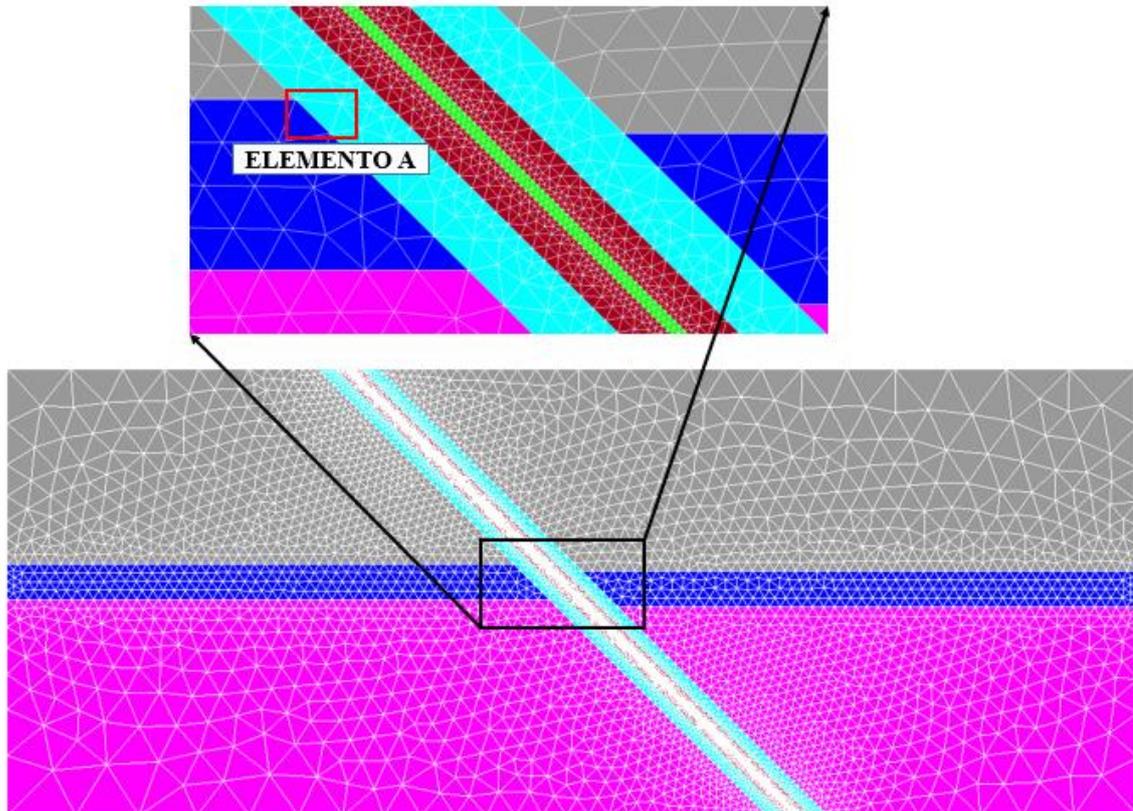
Pressão de Injeção	(MPa)
	22,0
	(...)
Pressão Mínima para reativação da falha	<b>19,8</b>
Pressão Máxima para não reativação da falha	<b>19,7</b>
	(...)
	19,0

Fonte: A Autora (2022).

De acordo com a descrição do modelo, a malha criada para análise acoplada hidro-mecânica consiste em 7077 nós e 13992 elementos. Na malha, encontra-se

identificado o elemento A, localizado na zona de dano externa da falha, conforme mostrado na Figura 11.

Figura 11 – Geometria do Modelo e Malha de Elementos Finitos



Fonte: A Autora (2022)

Como referência para os valores das propriedades dos materiais (Tabela1), utilizou-se o trabalho de Pereira *et al.* (2014). As propriedades utilizadas para cada material são o módulo de elasticidade  $E$ , coesão  $c'$  e ângulo de atrito  $\phi'$  efetivos, permeabilidade  $k$  e porosidade  $\Phi$ . Denomina-se Zona de Dano Externa (ZDE) e Zona de Dano Interna (ZDI).

Tabela 2 – Propriedades Para os Materiais do Modelo

Camada	$E$ (MPa)	$c'$ (MPa)	$\phi'$ (°)	$k$ (m <sup>2</sup> )	$\Phi$
Overburden	42000	1000	35	$9,87e^{-21}$	0,01
Underburden	26000	1000	35	$9,87e^{-21}$	0,01
Reservatório	30000	1000	35	$4,93e^{-14}$	0,20
Falha(Núcleo)	8000	1	27	$9,87e^{-21}$	0,10
Falha(ZDI)	6000	0,8	27	$9,87e^{-21}$	0,30
Falha(ZDE)	8000	0,8	27	$9,87e^{-21}$	0,20

Fonte: Adaptado de Pereira et al. (2014)

### 3.2 APLICAÇÕES DE MODELOS DE ML EM CENÁRIOS DE REATIVAÇÃO DE FALHAS

Esta seção discute os princípios teóricos utilizados para implementações de modelos de ML em cenários de reativação de falhas geológicas com base em casos analíticos. A seção 3.2.1 descreve o algoritmo de amostragem em linguagem Python para obtenção de casos analíticos a serem avaliados pelo critério de Mohr-Coulomb. Através da seção 3.3.2, temos o aprofundamento de conceitos de ML, na 3.2.3, da LDA e a RNA na seção 3.2.4, relativos à classificação binária para o caso de reativação de falhas geológicas. Na seção 3.2.5 são explicitadas métricas de avaliação de modelos de classificação. E na seção 3.2.6, detalha-se a implementação dos modelos de aprendizado de máquina.

#### 3.2.1 Dados Para o Método Analítico de Análise de Reativação de Falhas

Sibson (1994) introduziu a possibilidade de reativação de falhas como um caminho de fluxo em áreas com falhas, sendo o processo geralmente controlado pelo critério de falha de Coulomb. As três principais abordagens para avaliar este fenômeno são: analíticas, semi-analíticas e numéricas (Nacht et al. 2010). O modelo analítico de Mohr-Coulomb é amplamente utilizado na literatura para avaliação de

falhas, podendo ser citadas suas aplicações nos trabalhos de Huyghe e Mugnier (1992), Moraes (2004), Nacht et al. (2010), Pereira (2015), Skurtveit et al. (2018), Taghipour et al. (2020), entre outros. Através deste modelo, avaliam-se as alterações induzidas pelas tensões decorrentes da pressurização do reservatório, o que fica evidente no processo de injeção de fluidos para permitir a produção em reservatórios de petróleo.

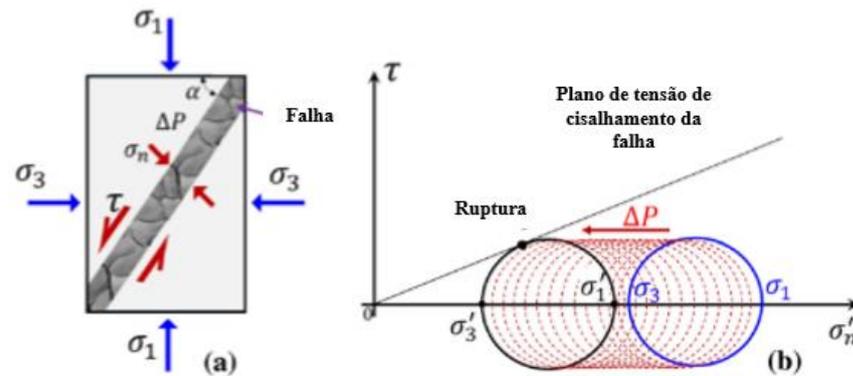
Sabe-se que o comportamento mecânico do reservatório é controlado pelas tensões efetivas, conforme Equação 4. O modelo constitutivo mecânico adotado neste trabalho se baseia no critério de Mohr Coulomb, no qual a superfície de fluência é expressa como (Equação 9):

$$F = \sigma'_1 - \sigma'_3 - 2c' \cos \varphi - (\sigma'_1 + \sigma'_3) \sin \varphi' \quad (9)$$

Onde  $\sigma'_1$  e  $\sigma'_3$  são as tensões principais, máxima e mínima, respectivamente;  $c'$  e  $\varphi'$  definem a coesão efetiva e ângulo de atrito efetivo da rocha, respectivamente. Ou seja, para avaliar a condição de reativação, é necessário o conhecimento do estado de tensão atuante, dado pelas tensões principais, e das propriedades mecânicas da rocha componente da falha, dados pelos parâmetros de coesão e ângulo de atrito, entre outros.

O aumento da pressão do fluido induz um movimento progressivo do círculo de Mohr para a esquerda, podendo provocar a ruptura do material quando o círculo toca a superfície de fluência (Figura 12), caracterizando a reativação de falhas pré-existentes. É importante destacar que a hipótese adotada neste trabalho considera que o plano de falha tem os mesmos parâmetros de resistência do reservatório.

Figura 12- Critério de Ruptura de Mohr Coulomb



Fonte: Adaptado de Taghipour et al. (2020)

Na figura 13, tem-se uma seção esquemática de um reservatório cruzado por uma falha geológica para avaliação por meio do modelo analítico. Como mencionado anteriormente, para avaliar a reativação pelo critério de Coulomb, é necessário determinar o estado de tensão, dado pelas tensões efetivas principais, e as propriedades da falha, como ângulo de atrito, coesão, coeficiente de empuxo e coeficiente de Biot. Devido à escassez de dados experimentais, as variáveis geomecânicas foram determinadas através da geração de dados randômicos seguindo distribuição normal com intervalos específicos das propriedades, utilizando o programa Python.

As variáveis geradas e os correspondentes possíveis cenários de reativação, posteriormente avaliados pelo critério de Mohr Coulomb (Equação 9), simulam casos de reativação de falhas normais, conforme Anderson (1951). É considerado que a tensão principal máxima é, portanto, equivalente à tensão vertical efetiva (Equação 10):

$$\sigma'_1 = \sigma'_v \quad (10)$$

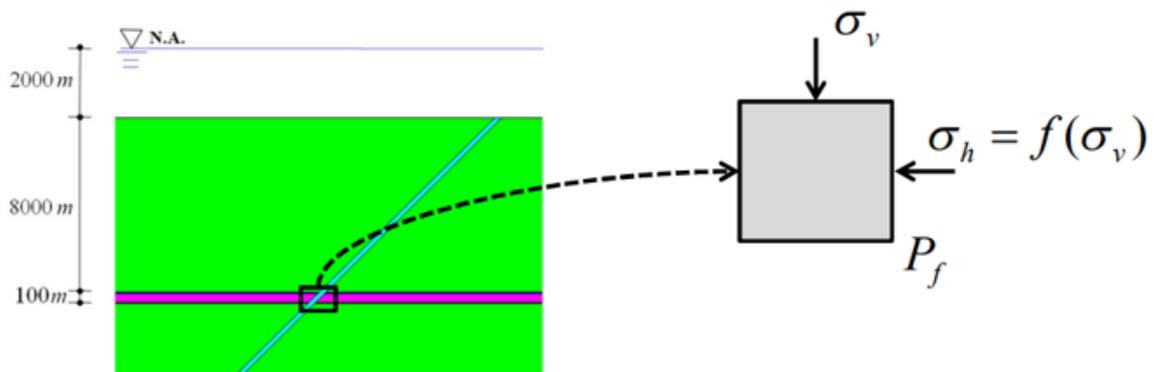
E a tensão principal mínima equivalente à (Equação 11):

$$\sigma'_3 = k'_0 \sigma'_1 \quad (11)$$

Ressalta-se ainda que a geração de cenários de falhas por meio de variáveis aleatórias pode contribuir para a generalização desses cenários, pois o processo de reativação de falhas contém inúmeras incertezas associadas e a estratégia utilizada destaca a variabilidade e a generalização no processo, não sendo totalmente

determinística. Dessa forma, por meio das variáveis aleatórias, torna-se possível avaliar múltiplos cenários de reativação, fundamentais para a utilização de modelos de aprendizado de máquina. A Figura 13 representa uma seção esquemática de um reservatório para o modelo analítico, enquanto a Tabela 3 mostra a faixa de valores utilizados para a geração das variáveis aleatórias.

Figura 13-Seção esquemática do reservatório para o modelo analítico.



Fonte: A Autora (2022)

Tabela 3- Dados Geomecânicos Para Modelo Analítico

Parâmetro	Faixa de Valores
Coesão ( $c'$ )	8-30 MPa
Coefficiente de Biot ( $\alpha$ )	0-1
Ângulo de Atrito ( $\varphi'$ )	15-40 °
Tensão Efetiva Vertical ( $\sigma'$ )	150-350 MPa
Coefficiente de Empuxo ( $k_0$ )	0-1
Poropressão (P)	80-150 MPa

Fonte: A Autora (2022)

### 3.2.2 Aprendizado de Máquina (Machine learning (ML))

O aprendizado de máquina é uma área da inteligência artificial que permite que os computadores aprendam com dados, sem intervenção humana. Existem dois tipos principais de técnicas de ML: o aprendizado supervisionado, em que o conjunto de

dados é rotulado, e o aprendizado não supervisionado, em que o conjunto de dados não é rotulado. Para criar um modelo matemático de ML, é preciso um conjunto de dados de treinamento e um conjunto de dados de teste. O conjunto de dados de treinamento é usado para treinar e ajustar a rede, enquanto o conjunto de dados de teste é usado para avaliar o desempenho do modelo. No aprendizado supervisionado, o modelo pode prever valores categóricos (classificação) ou contínuos (regressão). Exemplos de algoritmos de ML incluem Redes Bayesianas, Regressão Linear, Redes Neurais Artificiais, Análises Discriminantes Lineares e Árvores de Decisão (Bishop e Nasrabadi, 2006; He, Li e Li, 2020; Ray, 2019; Oyedere e Gray, 2020).

Neste trabalho, utilizamos de RNA's e LDA's para obter correlações matemáticas entre cenários de reativação de falhas considerando os dados de entrada do modelo analítico de Mohr Coulomb. Para a rede construída com dados analíticos, obteremos uma correlação matemática entre a classificação e os parâmetros de entrada ( $P$ ,  $c$ ,  $\varphi$ ,  $\sigma$ ,  $\alpha$ ,  $k_0$ ) com base nos algoritmos de ML, utilizando dados randômicos para o treinamento. Após o treinamento da rede, introduziremos os dados de teste para fazer previsões e validar os modelos.

### **3.2.3 Análise Discriminante Linear (Linear Discriminant Analysis (LDA))**

A Análise Discriminante Linear (LDA) é um método de classificação usado em ML para prever a classe de um dado exemplo com base em um conjunto de características ou variáveis independentes, de acordo com Bishop e Nasrabadi (2006). Ela é baseada na hipótese de normalidade e na suposição de que as classes possuem distribuições normais com a mesma covariância (HASTIE et al., 2009).

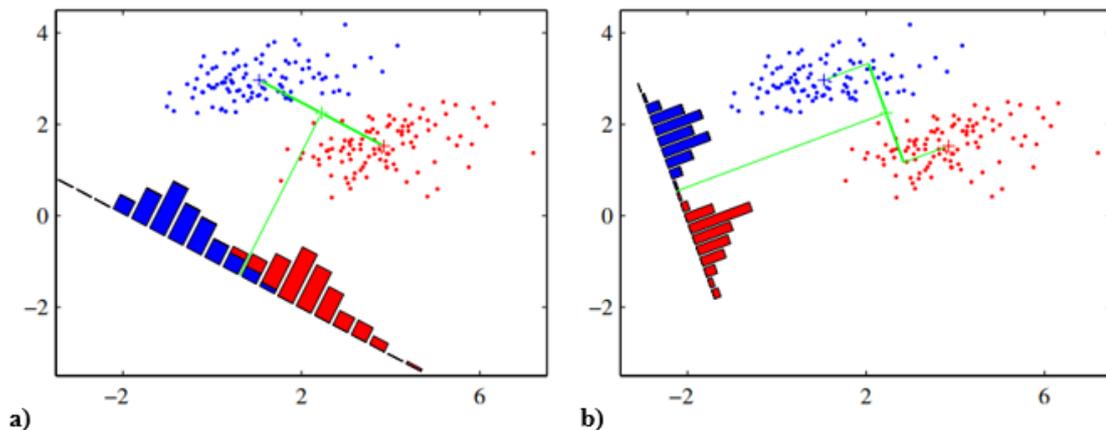
LDA é também um algoritmo supervisionado, o que significa que é necessário ter um conjunto de dados de treinamento rotulado previamente para treinar o modelo. Durante o treinamento, o modelo estima os parâmetros de cada classe (como a média e a variância) a partir dos exemplos rotulados. Depois, quando recebe um novo exemplo, o modelo usa esses parâmetros para calcular a probabilidade de esse exemplo pertencer a cada uma das classes. A classe com a maior probabilidade é escolhida como a classe prevista para o exemplo.

Essa técnica é chamada de linear porque assume que as diferenças entre as classes são lineares em relação às variáveis preditoras. Portanto, uma das principais vantagens da LDA é que ela é fácil de implementar e entender, pois utiliza uma

equação linear para separar as classes. Além disso, ela é eficiente em termos de tempo de processamento e é capaz de lidar com um número relativamente grande de variáveis independentes.

Uma maneira de visualizar o modelo de classificação linear fornecido pela técnica de LDA é em termos de redução de dimensionalidade. No entanto, ainda há um problema com essa abordagem, como ilustrado na Figura 14. Ela mostra duas classes bem separadas no espaço bidimensional original ( $x_1, x_2$ ), mas que têm uma sobreposição considerável quando projetadas na linha que une seus valores médios. Esse problema surge devido às covariâncias fortemente não diagonais das distribuições de classe. A ideia proposta por Fisher (1936) é maximizar uma função que dê uma grande separação entre as médias das classes projetadas, enquanto também proporciona uma pequena variância dentro de cada classe, minimizando assim a sobreposição entre as classes. O objetivo mencionado pode ser observado através da significativa melhora na separação de classes na Figura 14. (BISHOP e NASRABADI, 2006).

Figura 14-a) Amostras de duas classes (decratas em vermelho e azul) juntamente com os histogramas resultantes da projeção na linha que une as médias das classes. b) Projeção correspondente baseada no discriminante linear.



Fonte: Adaptado de Bishop e Nasrabadi (2006)

### 3.2.3.1 Formulação Matemática de LDA

A proposta inicial de Fisher (1936) para discriminar e classificar entre dois ou mais grupos era transformar as observações multivariadas, através de combinações lineares das variáveis, em observações univariadas de forma a minimizar a classificação equivocada de um indivíduo em uma população à qual ele não pertence.

Considerando que  $p$  variáveis  $X$  foram observadas em elementos amostrais de  $g$  populações distintas, em que não é possível assumir a normalidade dos dados, mas é razoável assumir que as matrizes de variâncias e covariâncias ( $\Sigma$ ) das  $g$  populações são iguais (Equação 12):

$$\Sigma_1 = \Sigma_2 = \dots = \Sigma_g = \Sigma \quad (12)$$

Tomando a soma de produtos cruzados entre grupos (Equação 13)

$$B = \sum_i^g (\bar{\mu}_i - \bar{\mu}) (\bar{\mu}_i - \bar{\mu})^T \quad (13)$$

Onde  $\bar{\mu}_i$  é o vetor de médias amostrais no grupo  $i$  e  $\bar{\mu}$  é o vetor de médias amostrais.

E a soma de produtos cruzados intra-grupos (Equação 14):

$$W = \sum_{i=1}^g (n_i - 1) S_i \quad (14)$$

Onde  $n_i$  e  $S_i$  são, respectivamente, o número de observações e a variância no grupo  $i$ .

A Equação 3.14 pode ser rescrita da seguinte forma (Equação 15):

$$W = \sum_{i=1}^g \sum_{j=1}^{n_i} (\mu_{ij} - \bar{\mu}_i) (\mu_{ij} - \bar{\mu}_i)^T \quad (15)$$

Onde  $\mu_{ij}$  é uma amostra  $j$  que pertence ao grupo  $i$ .

O objetivo do LDA é encontrar uma matriz de projeção  $\hat{a}$  que maximiza a razão entre a variação entre classes e a variação dentro da classe (Equação 16):

$$\hat{a} = \operatorname{argmax}(\hat{a}) \left( \frac{B}{W} \right) \quad (16)$$

Sejam  $\widehat{\lambda}_1 > \dots > \widehat{\lambda}_s > 0$  os autovalores de  $\mathbf{W}^{-1}\mathbf{B}$  e  $\widehat{e}_1, \dots, \widehat{e}_s$  os autovetores ortonormais correspondentes. Então a matriz de projeção  $\widehat{\mathbf{a}}$  que maximiza a razão, dada pela equação acima, é  $\widehat{\mathbf{a}} = \widehat{\mathbf{e}}$ .

Assim, uma vez encontrada a matriz de projeção  $\widehat{\mathbf{a}}$ , ela pode ser utilizada para projetar as amostras em um espaço de dimensão menor, dado pelas funções discriminantes lineares, como representado pela Equação 17.

$$Y(x) = \widehat{\mathbf{a}}^T x \quad (17)$$

### 3.2.4 Redes Neurais Artificiais (RNA)

O modelo de neurônios artificiais foi proposto inicialmente por McCulloch e Pitts (McCulloch e Pitts, 1943) para compreender o processamento de sinais no cérebro, sendo posteriormente aprimorado por outros. Tomando como base as redes neurais biológicas no cérebro humano, compostas por bilhões de neurônios que comunicam entre si, transmitindo sinais elétricos, surge-se o conceito de neurônios artificiais. Cada neurônio coleta sinais de recepção para produzir uma saída, que ultrapassam o limite de ativação, caso o limite não seja ultrapassado o neurônio permanece inativo. Assim, considerando o comportamento mencionado, desenvolveu-se o modelo matemático básico relativo aos neurônios artificiais (Ahmadi Goltapeh et al., 2022). Com base na Equação 18 tem-se a representação da saída de um neurônio, relativo a soma ponderada  $x_1, \dots, x_n$  entradas de N neurônios:

$$y = f(\sum_{i=1}^N w_i x_i + b) \quad (18)$$

Uma Rede Neural Artificial (RNA) consiste em camadas de neurônios interligadas, com base no sistema neural biológico, os neurônios interagem uns com os outros transmitindo informações através das camadas (Anderson, 1995). A arquitetura de uma RNA consiste em uma camada de entrada, uma camada de saída e camadas intermediárias, também denominadas camadas ocultas (Heaton, 2020). Cada camada possui um número arbitrário de neurônios e a conexão estabelecida entre dois neurônios é o peso. Os pesos são números reais multiplicados pelas

respectivas entradas criando uma soma ponderada aplicada a função de ativação e sendo transmitido da camada anterior para a posterior, ou seja, transmitindo informações entre as camadas. São inúmeras as funções de ativações, e seu uso depende das camadas e objetivos da rede, podendo ser citadas as funções ReLu, Softmax, TanH, dentre outras (SHARMA; SHARMA, ATHAIYA, 2017), (Heaton, 2020). As redes neurais também apresentam vieses que auxiliam no processo de aprendizagem.

#### 2.1.1.1 Formulação Matemática de RNA

Implementando uma Rede Neural, a Equação 19 reflete a soma ponderada da primeira camada oculta:

$$y_i^1 = f(z_i^1) = f\left(\sum_{j=1}^{N_0} w_{ij}^1 x_j + b_i^1\right) \quad (19)$$

Onde  $y_i^1$  é a saída da primeira camada,  $z_i^1$  é a soma ponderada,  $f$  é a função de ativação,  $i$  é o número de neurônios na primeira camada,  $b$  é o viés,  $N_0$  é o número de neurônios na primeira camada, e o sobrescrito 1 mostra a primeira camada oculta. Para  $l$ -ésima camada oculta, temos (Equação 20):

$$y_i^l = f^l(z_i^l) = f\left(\sum_{j=1}^{N_{l-1}} w_{ij}^l y_j^{l-1} + b_i^l\right) \quad (20)$$

Na Equação 20, o sobrescrito  $l$  refere-se a  $l$ -ésima camada, e mostra que diferentes funções de ativação  $f^l(z_i^l)$  que podem ser aplicadas as camadas. Tomando como base o algoritmo de programação direta (*forward propagation*), observa-se que os pesos são somados com as entradas dadas pela camada anterior, e após são combinados com a função de ativação, avançado para suas respectivas saídas.

O algoritmo de programação direta (*forward propagation*) pode ainda ser escrito para multicamadas em formato matricial, como mostrado na Equação 21:

$$y_i^l = f^l(W^l y^{l-1} + b^l) \quad (21)$$

Onde  $W$  é uma matriz de peso com dimensões de  $N_{l-1} \times N_l$ , enquanto  $y_i^l$  e  $b^l$  são respectivamente, vetor de saída e dos vieses com dimensões de  $N_l \times 1$ , mostrados em camadas. Através da equação observa-se que as variáveis desconhecidas são os pesos e os vieses, que necessitam serem ajustados para diminuir o erro do modelo. O algoritmo de retropropagação (*backward propagation*) (Rumelhart, Hinton, Williams, 1986) (Heaton, 2020) permite o ajuste dos pesos e vieses através da minimização da função de custo, também denominada função de erro.

$$C(W) = \frac{1}{2} \sum_{i=1}^n (y_i - t_i)^2 \quad (22)$$

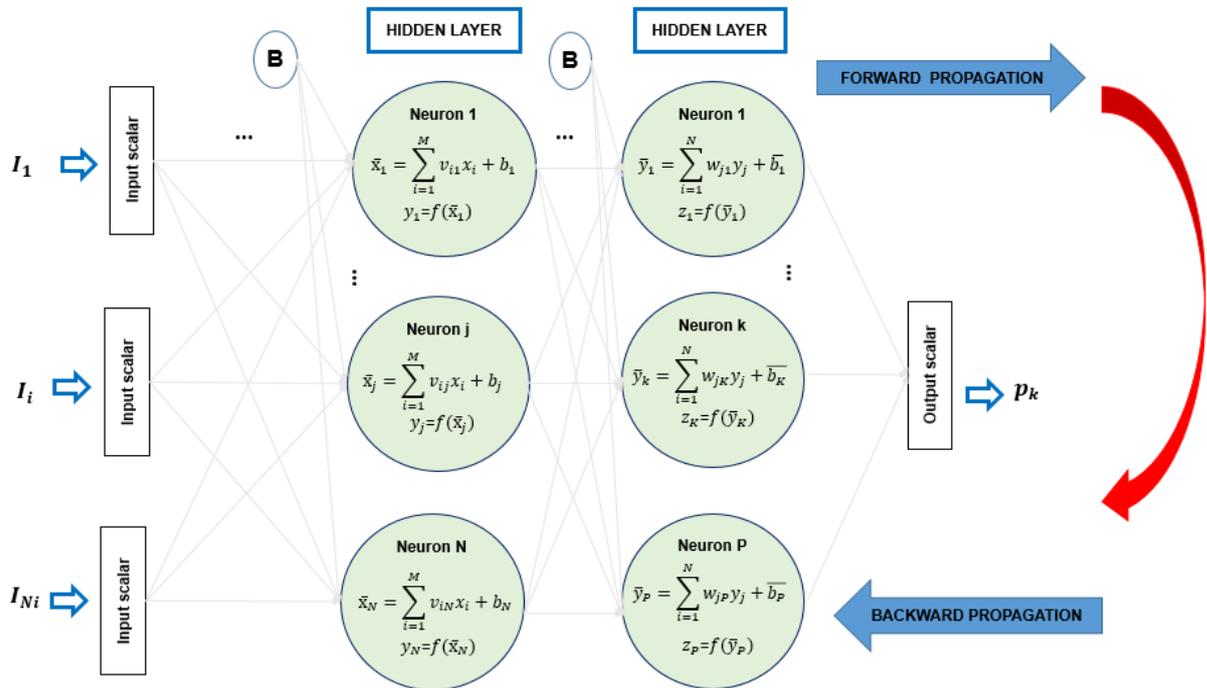
A equação 22 representa a função de custo, onde  $t_i$  é o vetor de valores alvo e  $y_i$  consiste no vetor de predições/saídas da rede neural. Assim, a programação direta de  $x$  fornece  $y$ , através da rede neural. A equação 23 introduz o algoritmo de retropropagação, que calcula  $a^{l-1} = f(z^l)$  e então conecta a última camada oculta como uma entrada.

$$z^l = (W^l)^T a^{l-1} + b^l \quad (23)$$

Na equação o argumento ( $z^l$ ) é atualizado pelas modificações dadas por  $(W^l)^T$  e  $b^l$ . Assim, o valor da função de custo (Equação 22), é otimizado pela propagação da saída para camada de entrada por pesos e vieses modificados. Ou seja, o processo de aprendizagem consiste no ajuste dos pesos e vieses para minimizar a função de custo e aumentar a precisão do modelo. A rede neural desenvolve sua própria lógica a partir do treinamento da rede.

Para implementação da rede se faz necessário a definição de hiperparâmetros como número de camada, tipo de camadas, número de neurônios nas respectivas camadas e funções de ativação. A rede é então treinada com os dados de treino. Após o treinamento, a arquitetura do modelo é salva com os pesos e vieses finais. E assim, o modelo treinado pode ser aplicado a dados não vistos, referente ao conjunto de dados de teste, tornando possível avaliar a eficiência do modelo. Através da Figura 15 tem-se uma arquitetura geral de RNA de classificação, incluindo a programação para frente e a retropropagação.

Figura 15-Modelo de arquitetura de RNA



Fonte: A Autora (2022)

### 3.2.5 Métricas de desempenho da classificação

Como os algoritmos de aprendizado de máquina são construídos com base em dados de treinamento e utilizados para fazer previsões em dados de teste, é fundamental ter métodos para avaliar o desempenho dos algoritmos. Os modelos também têm erros associados a eles, portanto, nesta seção, discutiremos as medidas de desempenho comumente utilizadas para algoritmos de aprendizado de máquina de classificação, conforme orientam os estudos de Grandini, Bagli e Visani (2020), Oyedere e Gray (2020), entre outros.

#### 3.2.5.1 Matriz de Confusão

A matriz de confusão é uma tabela que é usada para avaliar o desempenho de um classificador em relação a um conjunto de dados de teste. Ela é muito útil para visualizar a performance do classificador e identificar onde ele está tendo dificuldades para fazer previsões corretas.

Uma matriz de confusão é composta por quatro quadrantes:

- Verdadeiros positivos (VP): são aqueles casos em que o classificador previu corretamente a classe positiva.
- Falsos positivos (FP): são aqueles casos em que o classificador previu incorretamente a classe positiva.
- Falsos negativos (FN): são aqueles casos em que o classificador previu incorretamente a classe negativa.
- Verdadeiros negativos (VN): são aqueles casos em que o classificador previu corretamente a classe negativa.

A matriz de confusão (Tabela 4) pode ser usada para calcular várias métricas de desempenho do classificador, como a acurácia, precisão, a sensibilidade (*recall*) e a F1-score.

Tabela 4- Matriz de Confusão

	Previsão Positiva	Previsão Negativa
Classe Positiva	VP	FN
Classe Negativa	FP	VN

Fonte: A Autora (2022)

### 3.2.5.2 Acurácia

A acurácia é a métrica de desempenho mais básica para modelos de classificação, que mede a fração de pontos de dados corretamente classificados pelo modelo. Em um classificador de duas classes, a acurácia é expressa da seguinte maneira (Equação 24):

$$Acurácia = \frac{VP+VN}{VP+FP+VN+FN} \quad (24)$$

A acurácia deve ser usada com cautela, pois só funciona bem quando as classes estão aproximadamente balanceadas. Portanto, é importante ter cuidado ao utilizar a acurácia como medida de desempenho em problemas de classificação em que as classes podem estar desbalanceadas. Nestes casos, é recomendado utilizar

outras métricas, como a sensibilidade (*recall*) ou a *F1-score*, que são mais robustas ao desbalanceamento de classes, para resolver este problema.

### 3.2.5.3 Precisão

A precisão mede a fração de pontos de dados corretamente classificados em relação a todos os pontos de dados previstos com esse valor de rótulo, portanto é sensível ao número de pontos de dados corretamente classificados. A precisão é expressa como (Equação 25):

$$Precisão = \frac{VP}{VP+FP} \quad (25)$$

### 3.2.5.4 Sensibilidade (*Recall*)

A sensibilidade (*recall*) mede a fração de pontos de dados corretamente classificados para um determinado valor de rótulo em relação a todos os pontos de dados que possuem esse valor de rótulo. A sensibilidade (*recall*) é expressa como (Equação 26):

$$Recall = \frac{VP}{VP+FN} \quad (26)$$

### 3.2.5.5 F1-score

A *F1-score* combina as métricas de precisão e sensibilidade (*recall*) para fornecer uma única métrica que pode ser usada para avaliar um algoritmo de classificação. Ela calcula a média harmônica da precisão e do *recall* e é expressa da seguinte forma (Equação 27):

$$F1 - score = 2 \times \frac{Recall \times Precisão}{Precisão + Recall} \quad (27)$$

### 3.2.6 Implementação dos Modelos Baseados em Aprendizado de Máquina

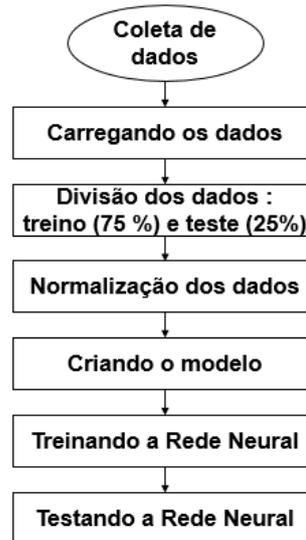
Devido à escassez de dados experimentais geomecânicos, determinou-se as variáveis geomecânicas através da geração de dados randômicos, conforme já mencionado. Tal processo tende a contribuir para uma análise não determinística do problema, sendo conveniente para lidar com as incertezas do processo.

A geração dos dados geomecânicos randômicos foi realizada por código em linguagem Python. Assim, foram gerados 100000 pontos de dados para cada uma das variáveis ( $P$ ,  $c$ ,  $\varphi$ ,  $\sigma$ ,  $\alpha$ ,  $k_0$ ). Os dados foram então mesclados e avaliados pelo critério de Mohr Coulomb, consistindo assim em 100000 cenários de avaliações de reativação de falhas.

O processo de construção da arquitetura das redes neurais compreende um processo de tentativas que buscam minimizar o erro e o tempo de treinamento, assim como aumentar a acurácia do modelo. Para construção do modelo analítico baseado em redes neurais, três camadas foram suficientes para obter um desempenho satisfatório, correspondentes a camada de entrada com 6 neurônios (relativos ao vetor de características do modelo tomadas como entrada da rede), camada oculta com 25 neurônios e camada de saída com apenas um neurônio, relativo ao problema de classificação binária. A camada de saída foi ativada usando a função sigmóide, clássica em problemas relativos à classificação binária (Hashemi et al. (2019)), enquanto as demais camadas utilizaram a função de ativação tangente hiperbólica ( $\tan H$ ) pelo seu melhor desempenho quando comparado a função de ativação de unidade linear retificada (ReLU). Portanto, no modelo cada conjunto de dados de entrada contém seis atributos (*features*) ( $P$ ,  $c$ ,  $\varphi$ ,  $\sigma$ ,  $\alpha$ ,  $k_0$ ), bem como um alvo  $g(x)$  que corresponde a classificação em um (indicando cenário de reativação de falha) ou zero (indicando cenário de não reativação de falha).

A Figura 16 ilustra o fluxograma da rede neural considerando a fundamentação da rede neural com o conjunto de dados de treino e, em seguida, validação com o conjunto de dados de teste. O conjunto de dados gerados são divididos em dados de treino e teste, sendo os dados de treino responsáveis pelo ajuste do modelo, enquanto os dados de teste são utilizados para fazer previsões e avaliar a acurácia do modelo. A divisão consiste em 75% dos dados para treinamento e 25% para teste. Ressalta-se que no link: [https://github.com/AnnyVirginia/MestradoAnny\\_LDA\\_RNA](https://github.com/AnnyVirginia/MestradoAnny_LDA_RNA), é possível encontrar códigos de implementação com informações detalhadas dos processos.

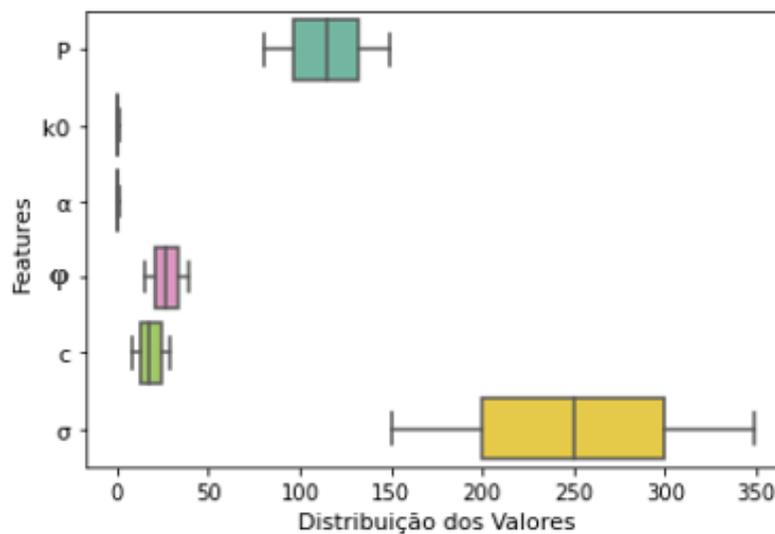
Figura 16-Fluxograma da Rede Neural



Fonte: A Autora (2022)

### 3.2.6.1 Normalização dos dados de entrada

A Figura 17 apresenta um gráfico de diagrama de caixa (*boxplot*) com a distribuição dos atributos (*features*) representadas pelas variáveis de entrada ( $P$ ,  $c$ ,  $\varphi$ ,  $\sigma$ ,  $\alpha$ ,  $k_0$ ) para os 100000 casos referentes ao modelo analítico.

Figura 17-Distribuição das variáveis de entrada ( $P$ ,  $c$ ,  $\varphi$ ,  $\sigma$ ,  $\alpha$ ,  $k_0$ )

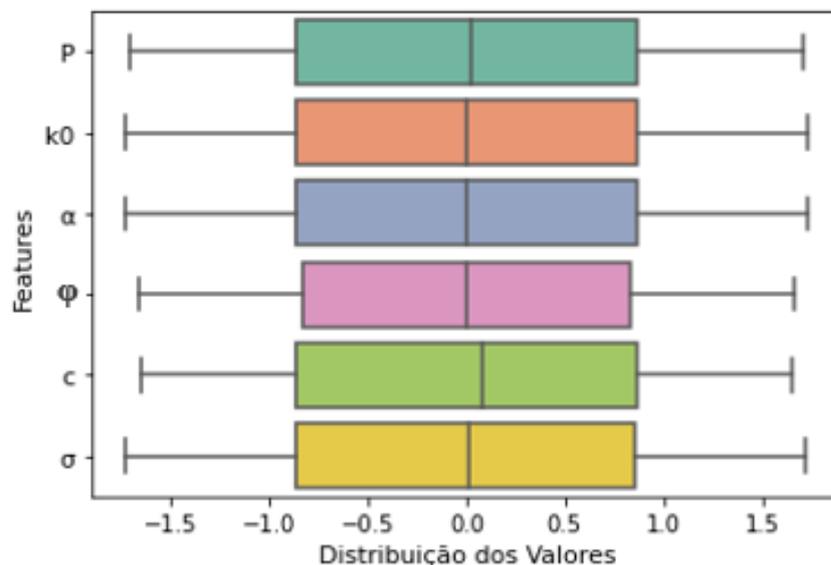
Fonte: A Autora (2022)

Através da Figura 17 observa-se uma distribuição não normalizada das variáveis, de modo que os atributos (*features*) tem distribuições diferentes entre si. Associado a essa distribuição necessita-se da normalização dos dados dos atributos,

correspondente a uma tarefa de pré-processamento das redes neurais, de maneira a garantir que o algoritmo não seja enganado pelo fato das variáveis serem altamente dispersas. O processo de pré-processamento mencionado não apenas auxilia na determinação das variáveis em um intervalo específico, mas também acelera o processo de aprendizagem do algoritmo. Dada a complexidade referente as diferentes escalas apresentadas pelas variáveis, o modelo torna-se complexo, dado que grandes valores de entrada podem levar a grandes pesos no modelo. Um peso com valor elevado é instável em função dos resultados negativos referentes ao treinamento, como também aumenta a sensibilidade dos valores de entrada, podendo causar erros significativos e lentidão no processo de aprendizagem (Ahmadi Goltapeh et al., 2022).

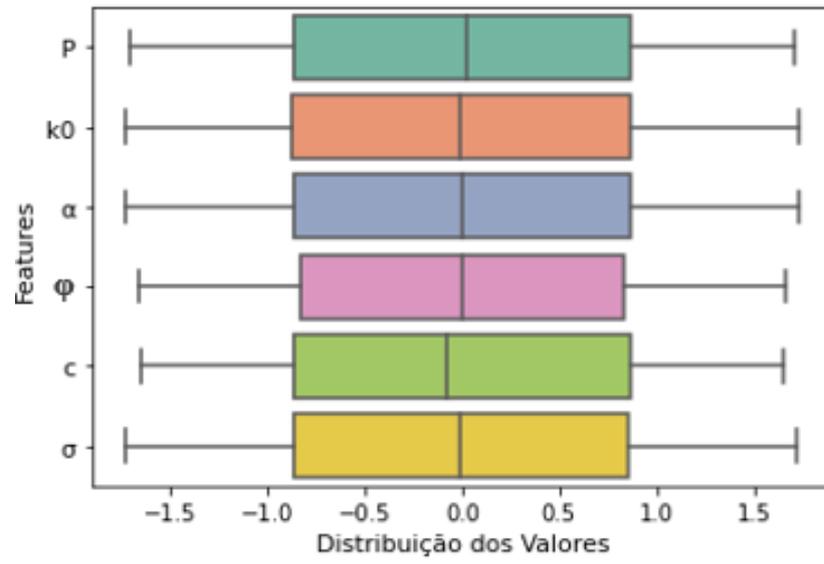
Através da Figura 18 e 19 tem-se a representação da distribuição das variáveis após a normalização, respectivamente relativas aos dados de treino e teste. A normalização consiste em subtrair da média e dividir pelo desvio padrão de cada uma das variáveis. Observa-se que após a normalização os dados apresentam uma média zero e um desvio padrão de 1, otimizando assim o processo de aprendizado da RNA e da LDA.

Figura 18-Distribuição das variáveis após a normalização, referente ao conjunto de treino



Fonte: A Autora (2022)

Figura 19-Distribuição das variáveis após a normalização, referente ao conjunto de teste



Fonte: A Autora (2022)

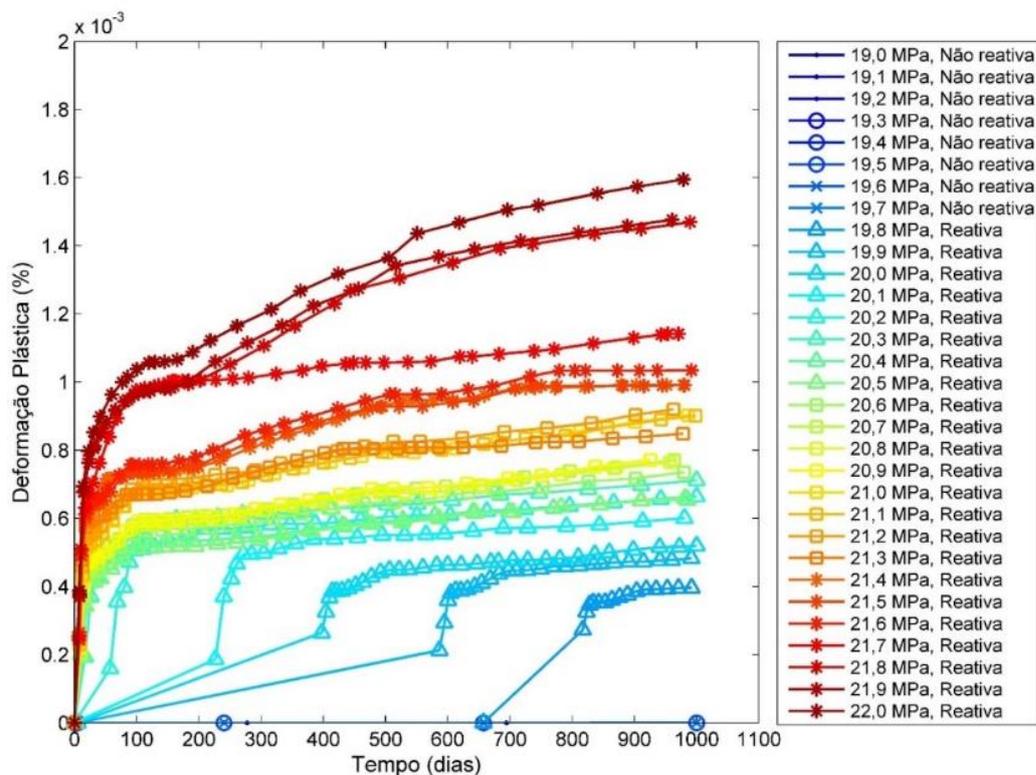
## 4 RESULTADOS E DISCUSSÕES

Nesta seção, discutem-se os resultados obtidos, apresentando a relação entre a análise de sensibilidade do parâmetro de pressão de injeção e as deformações plásticas, tensões cisalhantes, permeabilidades, pressões e fluxos de líquidos. Além disso, são observados os desempenhos da LDA e da RNA na classificação dos cenários de reativação de falhas, por meio dos resultados dos métodos de ML aplicados aos dados analíticos.

### 4.1 ANÁLISE DE SENSIBILIDADE DAS PRESSÕES DE INJEÇÃO

Na figura 20 são apresentados os resultados da evolução da deformação plástica para os diferentes cenários de pressões de injeção:

Figura 20-Gráfico de evolução da deformação plástica para os diferentes cenários de pressões de injeção (Elemento A).



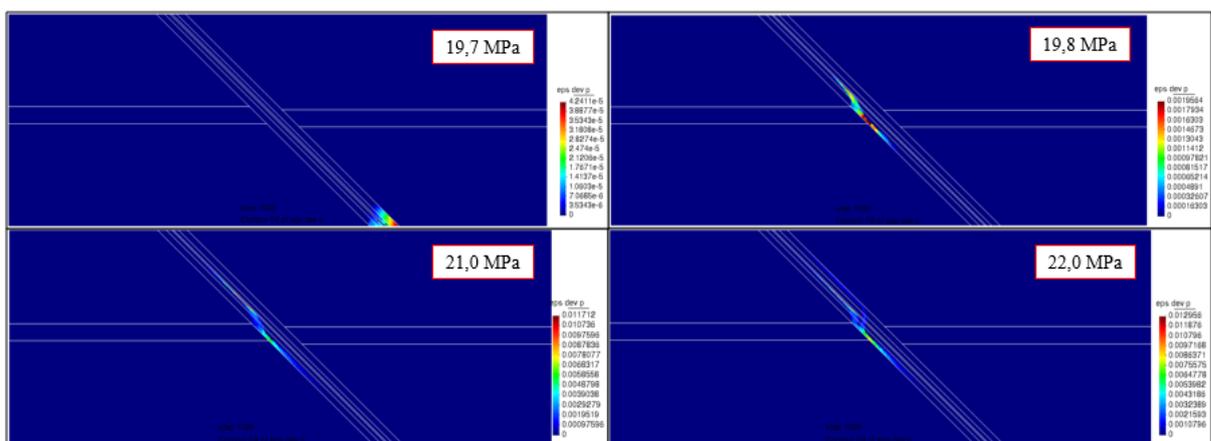
Fonte: A Autora (2022)

Através da figura 20 observa-se um comportamento crescente da deformação quando se tem um incremento no valor de pressão de injeção. O indicador do

processo de reativação da falha consiste na deformação plástica de cisalhamento (modelo elastoplástico). Portanto, para os cenários correspondentes às pressões de 19,0 MPa-19,7 MPa, a falha encontra-se selante. Contudo, a partir do valor de 19,8 MPa, ocorre a reativação da falha. Com isso é possível determinar que a pressão mínima para que haja a reativação da falha situa-se entre os valores de 19,7 MPa e 19,8 MPa. Além disso, é possível observar uma relação entre o aumento da pressão de injeção e o aumento da deformação plástica do local (Elemento A) em um mesmo intervalo de tempo. Com isso é possível concluir que quanto maior a vazão de injeção, mais rápida e mais severa é a plastificação do local. Por exemplo, com 19,8 MPa são necessários 658 dias para ocorrer a reativação da falha, enquanto que a 20,4 MPa são necessários apenas 12 dias.

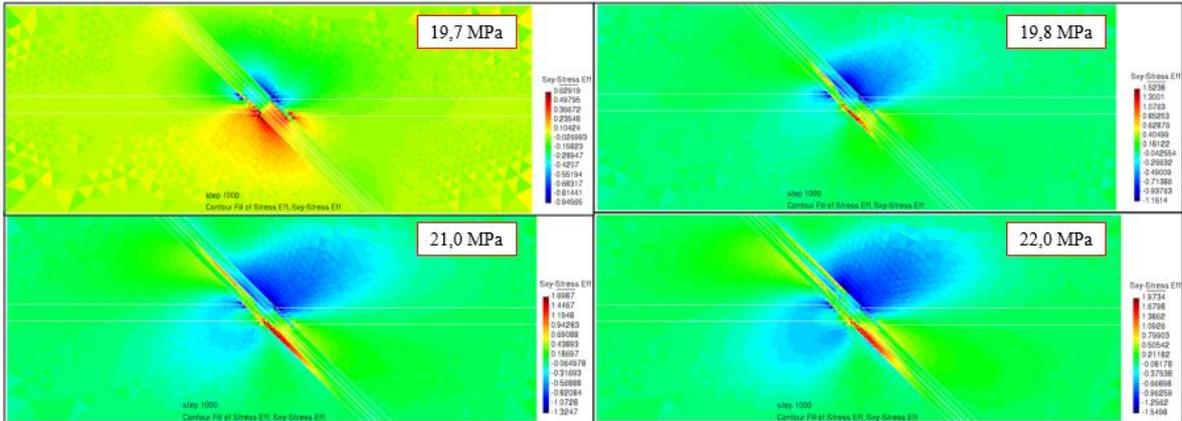
As deformações plásticas e tensões cisalhantes correspondentes às pressões de injeção de 19,7 MPa, 19,8 MPa, 21,0 MPa e 22,0 MPa são apresentadas nas Figuras 21 e 22 no final da simulação. A pressurização da falha durante a reativação, associada à alteração do estado de tensões, fica evidente pela presença de deformações plásticas ao longo da zona de falha (Figura 21). Essas deformações são originadas pelo desenvolvimento de tensões cisalhantes, que se concentram na região da zona de falha e aumentam sua intensidade com o aumento da pressão de injeção (Figura 22).

Figura 21- Distribuição das deformações plásticas para as pressões de injeção de 19,7 MPa, 19,8 MPa, 21,0 MPa e 22,0 MPa no tempo final de simulação.



Fonte: A Autora (2022)

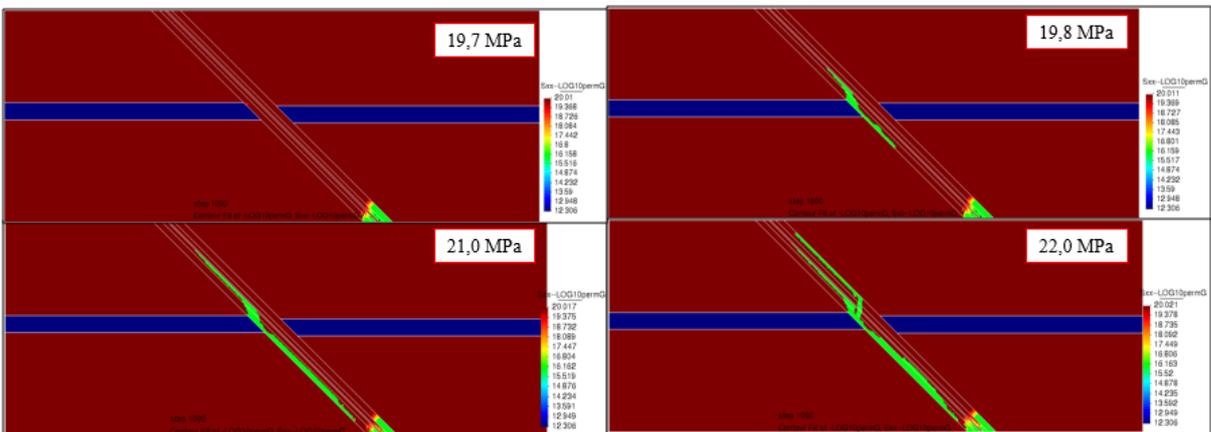
Figura 22- Distribuição das tensões cisalhantes para as pressões de injeção de 19,7 MPa, 19,8 MPa, 21,0 MPa e 22,0 MPa no tempo final de simulação.



Fonte: A Autora (2022)

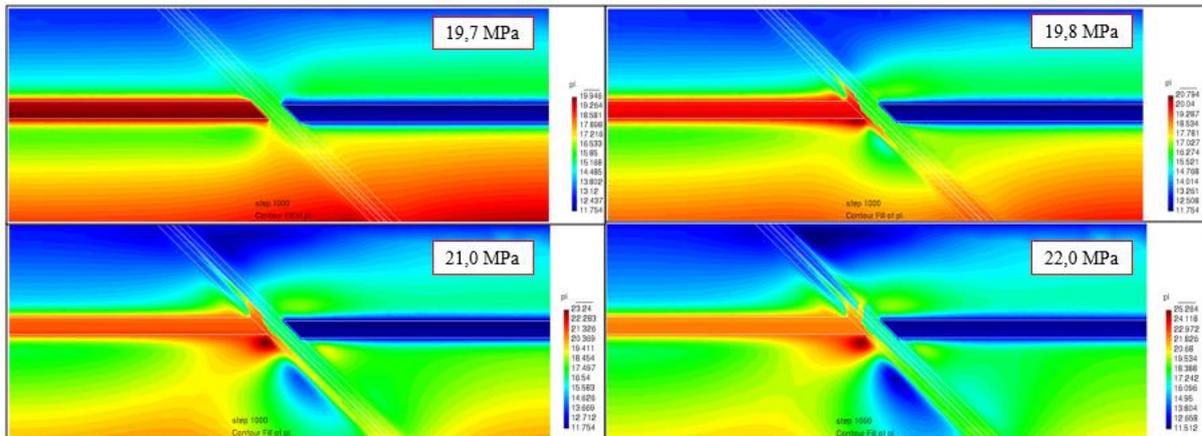
A permeabilidade, pressões e fluxos de líquidos correspondentes às pressões de injeção de 19,7 MPa, 19,8 MPa, 21,0 MPa e 22,0 MPa são apresentados, respectivamente, nas Figuras 23, 24 e 25 no final da simulação. Para a pressão de injeção de 19,7 MPa, não há plastificação da falha, o que leva à compartimentalização do reservatório, ou seja, os lados não se conectam. A dilatância inerente ao processo de cisalhamento leva à plastificação da falha e promove um aumento da permeabilidade da falha, conforme mostrado na Figura 23. Com uma maior pressurização, a falha se torna permeável ao longo de toda sua extensão. Esse aumento da permeabilidade leva ao surgimento de um gradiente de pressão (Figura 24) e fluxo de fluidos (Figura 25) ao longo da zona de falha.

Figura 23- Distribuição das permeabilidades para as pressões de injeção de 19,7 MPa, 19,8 MPa, 21,0 MPa e 22,0 MPa no tempo final de simulação.



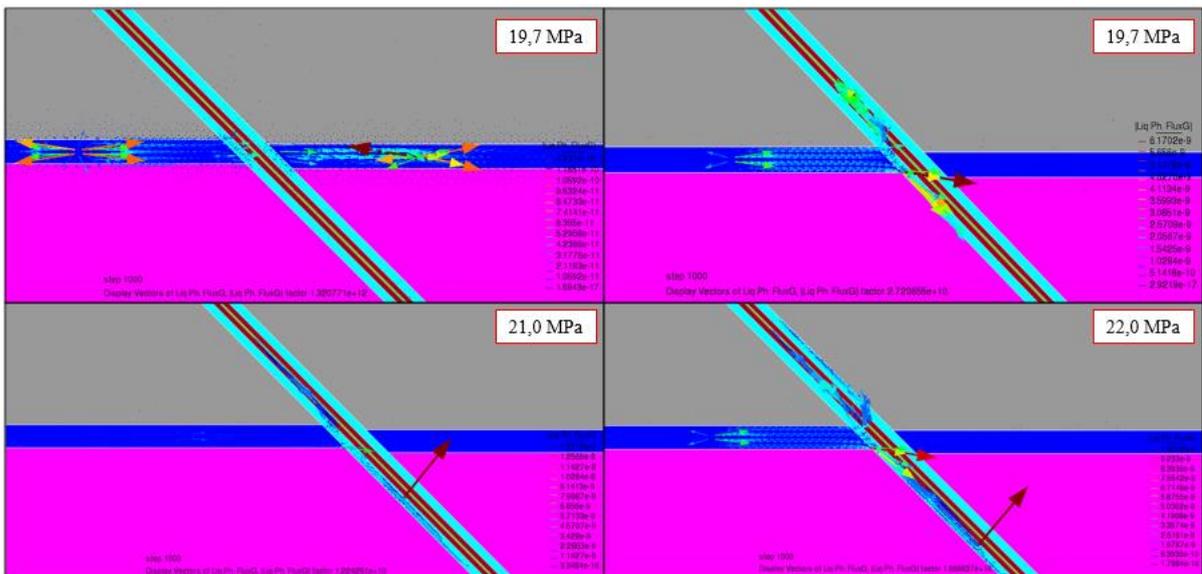
Fonte: A Autora (2022)

Figura 24-Distribuição das pressões de líquidos para as pressões de injeção de 19,7 MPa, 19,8 MPa, 21,0 MPa e 22,0 MPa no tempo final de simulação.



Fonte: A Autora (2022)

Figura 25-Distribuição dos vetores de fluxo para as pressões de injeção de 19,7 MPa, 19,8 MPa, 21,0 MPa e 22,0 MPa no tempo final de simulação.



Fonte: A Autora (2022)

Assim, a partir da pressão de injeção de 19,8 MPa, a falha plastifica e pressões e fluxos de líquidos se tornam presentes na região da falha. À medida que o parâmetro, pressão de injeção, tem valores maiores, as pressões e os fluxos de líquidos correspondentes se estendem ao longo do comprimento da zona da falha. A falha torna-se cada vez mais pressurizada e passa a funcionar como conduto para o fluxo dos fluidos (evidente pelos vetores do fluxo). Com a falha pressurizada, o

reservatório sofre consequente despressurização e há a possibilidade de perda de fluido, entre outros prejuízos decorrentes da reativação da falha.

Observa-se que a pressão de injeção de 19,7 MPa corresponde à máxima pressão de injeção para que não haja reativação, enquanto que a pressão de 19,8 MPa corresponde à pressão mínima de reativação. Já as pressões de injeção de 21,0 MPa e 22,0 MPa representam a progressão da reativação com a variável pressão de injeção, comprovada pelos resultados mencionados de deformação plástica, tensões cisalhantes, permeabilidade, pressão de líquidos e fluxo de fluidos. Isso mostra a importância da variável pressão de injeção no processo de reativação de falhas geológicas.

#### 4.2 LDA

Para quantificar o desempenho da classificação, utiliza-se uma função que mede a proporção de previsões corretas em relação a todas as previsões possíveis. Isso permite obter quatro tipos de previsões: verdadeiro positivo, falso positivo, verdadeiro negativo e falso negativo. As frequências dessas classificações para o modelo são mostradas na matriz de confusão do modelo LDA, apresentada na Figura 26. Observa-se um bom ajuste pelas frequências presentes na matriz.

Figura 26- Matriz de Confusão Para Modelo de LDA



Fonte: A Autora (2022)

A representação positiva do modelo LDA para avaliação de cenários de reativação pode ser reforçada pela alta acurácia do modelo, que é de 0,97. Outras métricas, como precisão, sensibilidade (*recall*) e F1-score, também podem ser usadas para destacar o excelente desempenho do modelo, permitindo a análise das métricas obtidas para cada classe. A tabela 5 mostra o desempenho dessas métricas para o modelo LDA.

Tabela 5- Métricas de Avaliação Para o Modelo de LDA

Classe	Precisão	<i>Recal l</i>	<i>F1- Score</i>
Não Reativa (0)	0,99	0,96	0,98
Reativa (1)	0,89	0,99	0,94

Fonte: A Autora (2022)

#### 4.3 RNA

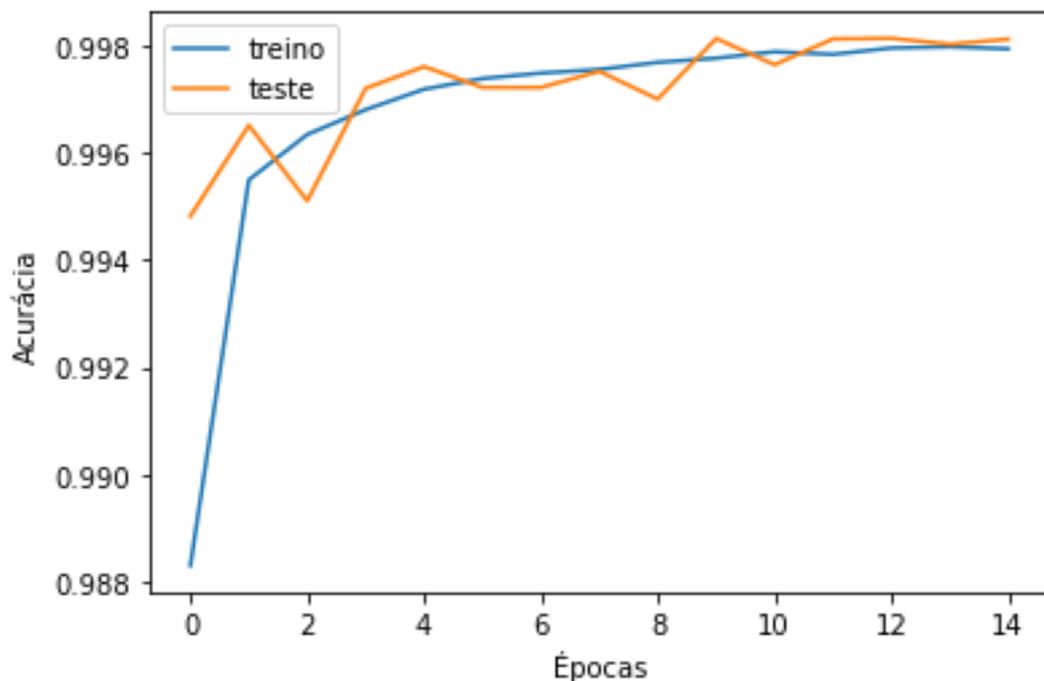
De acordo com as recomendações de Ahmadi Goltapeh et al. (2022), o otimizador Adam foi utilizado para monitorar o processo de convergência do modelo de rede neural. Isso se deve ao fato de que escolher a taxa de aprendizado não é uma tarefa fácil, uma vez que a rede tende a se aproximar de um mínimo local com taxas de aprendizado altas, o que pode causar instabilidade no processo de treinamento. Além disso, taxas de aprendizado baixas são computacionalmente caras, pois prolongam o processo de aprendizado. No entanto, o otimizador Adam possui taxas de aprendizado adaptáveis que otimizam funções objetivo estocásticas atualizando repetidamente os pesos da rede. A função de perda especificada foi a entropia cruzada binária, que define o objetivo da rede durante o treinamento. A entropia cruzada binária foi utilizada para estimar a função de perda e minimizar o *gap* entre as probabilidades. Além disso, a métrica de acurácia foi selecionada junto com a entropia cruzada para acompanhar o progresso de convergência da função de perda.

Para tornar o processo de aprendizado da rede mais eficiente, utilizamos o recurso de Parada Antecipada (*Early Stopping*), o que garante que o processo de treinamento continue por determinadas épocas enquanto a acurácia nos dados de

teste é mantida. Isso significa que um dos hiperparâmetros da rede, o número de épocas de treinamento, é definido, diminuindo assim o risco de sobreajuste (*overfitting*) da rede. Dessa forma, o aprendizado da rede é eficaz e não apenas específico para os dados de treinamento, permitindo que a rede generalize para dados que não foram vistos durante o treinamento.

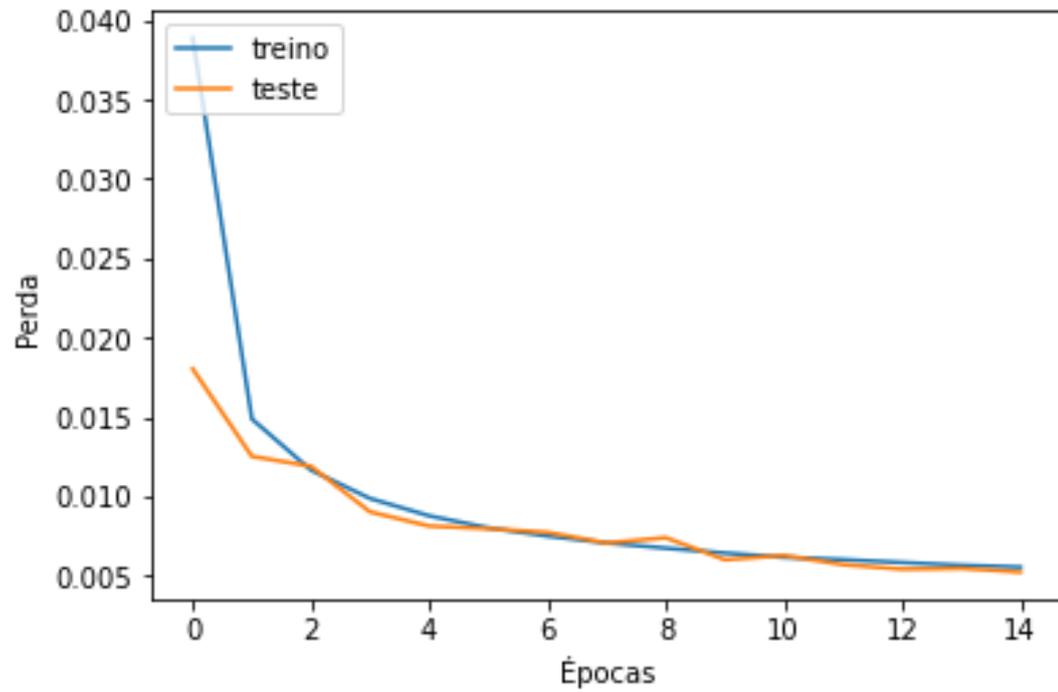
As Figuras 27 e 28 representam, respectivamente, as tendências das funções de acurácia e perda para os dados de treinamento e teste. O perfil ascendente da acurácia (Figura 27) e o perfil descendente da perda (Figura 28) indicam que a rede treinada está aprendendo adequadamente. O fato de o número de épocas de treinamento ser pequeno sugere que o aprendizado sobre os dados é relativamente simples. No entanto, se o modelo fosse mais complexo, por exemplo, se fossem utilizados dados numéricos como entrada com outras características, provavelmente haveria mais complexidade no modelo e, conseqüentemente, na forma como ele aprende.

Figura 27-Evolução da Acurácia para os dados de treino e teste ao longo das épocas.



Fonte: A Autora (2022)

Figura 28-Evolução da perda para os dados de treino e teste ao longo das épocas.



Fonte: A Autora (2022)

Para quantificar o desempenho da classificação, são mostradas na matriz de confusão as frequências de classificações para o modelo, apresentada na Figura 29. Observamos um excelente ajuste pelas frequências mostradas na matriz, que também pode ser destacada pela alta acurácia do modelo de 0,99.

Figura 29-Desempenho do modelo de classificação de RNA através da matriz de confusão.



Fonte: A Autora (2022)

A tabela 6 mostra o desempenho das métricas de classificação para o modelo de RNA.

Tabela 6- Métricas de Avaliação do Modelo de RNA

Classe	Precisão	Recall	F1-Score
Não Reativa (0)	0,99	0,99	0,99
Reativa (1)	0,99	0,99	0,99

Fonte: A Autora (2022)

Os resultados apresentados nas tabelas 5 e 6 sugerem que tanto a LDA quanto a RNA foram muito eficientes na classificação de cenários de reativação de falhas. A acurácia da rede neural é de 0,99, o que significa que o modelo foi capaz de prever corretamente 99% das classes de cenários de reativação de falhas. Já o LDA apresenta uma acurácia de 0,97, o que significa que o modelo foi capaz de prever corretamente 97% das classes de cenários de reativação de falhas. A precisão, o recall e o F1-score da classe "Reativa (0)" para a rede neural são de 0,99, 0,99 e 0,99, respectivamente, enquanto para o LDA são de 0,99, 0,96 e 0,98, respectivamente. A

precisão, o recall e o F1-score da classe "Reativa (1)" para a rede neural são de 0,99, 0,99 e 0,99, respectivamente, enquanto para o LDA são de 0,89, 0,99 e 0,94, respectivamente. Isso sugere que a RNA foi mais precisa em prever cenários de reativação de falhas do que o LDA.

Ambos os modelos foram muito eficientes na classificação de cenários de falhas de reativação, mas a rede neural apresentou um desempenho levemente superior em relação à precisão e ao recall para a classe "Reativação (1)". No entanto, é importante levar em consideração o contexto em que os modelos foram treinados e avaliados, pois isso pode afetar o desempenho de cada um. Além disso, é importante avaliar o desempenho dos modelos com diferentes conjuntos de dados e métricas para ter uma visão mais abrangente do desempenho de cada técnica.

As técnicas de LDA e RNA também podem ser usadas em conjunto para classificação. Enquanto a LDA é responsável por reduzir a dimensionalidade do problema, a RNA é específica para classificação, o que pode melhorar o desempenho do modelo. Além disso, a caracterização laboratorial e de campo das propriedades geomecânicas dos materiais, juntamente com análises numéricas para identificar a reativação, também podem aumentar o desempenho. Análises de sensibilidade podem ser realizadas para definir os parâmetros que melhor representam o fenômeno de reativação e uma rede com maior complexidade e capacidade de generalização pode ser implementada. No entanto, ainda é importante destacar que a RNA baseada em dados analíticos é uma boa abordagem inicial para avaliar a reativação, principalmente devido à sua simplicidade e ampla utilização.

## 5 CONCLUSÕES

Conforme observado pelos resultados da análise de sensibilidade, a pressão de injeção é uma importante variável para o processo de reativação de falhas geológicas. Na análise de sensibilidade em questão, a pressão de injeção de 19,8 MPa é responsável pela reativação da estrutura geológica: para pressões abaixo desse valor, a falha é selante, enquanto que para pressões acima, a falha continua a reativar. A falha é reativada devido à plastificação resultante da alteração do estado de tensões, o que leva a deformações plásticas e ao aumento da permeabilidade. Além disso, quanto maior o valor da variável de sensibilidade (pressão de injeção), menor o tempo necessário para a plastificação ocorrer, maiores as deformações, tensões cisalhantes e o aumento de permeabilidade, pressões e fluxos de fluidos ao longo do comprimento da falha. A pressurização da falha leva à consequente despressurização do reservatório, o que pode afetar a produção de petróleo e causar inúmeros prejuízos que precisam ser controlados.

Considerando os modelos de classificação utilizados para prever cenários de reativação baseados em casos analíticos, obtém-se um excelente desempenho tanto para a técnica de RNA quanto para a de LDA. Ou seja, ambos os modelos podem ser utilizados de maneira eficiente para avaliar cenários de reativação. Ao comparar o LDA e a RNA, observa-se um desempenho superior da classificação baseada na RNA. Este fato está associado ao processo de treinamento da RNA e à sua própria lógica.

As técnicas de LDA e RNA podem ainda ser utilizadas em conjunto. O LDA permitindo a redução da dimensionalidade do problema enquanto a RNA seria especialmente útil para as classificações de ativações de falhas. Para garantir um melhor desempenho, complexidade e generalização da classificação, são ainda necessários dados de caracterização das propriedades geomecânicas, análises numéricas e definição dos parâmetros que caracterizem o fenômeno de reativação. No entanto, um grande problema na construção e implementação da RNA é a necessidade de dados que representem o fenômeno de forma fiel e em quantidade suficiente para que possa ocorrer o aprendizado efetivo.

O modelo de classificação também pode incorporar a regressão e, através da RNA, fornecer pressões operacionais adequadas e otimizar a locação dos poços,

entre outros parâmetros que garantam a segurança no processo de análise de reativação, considerando a implementação de um modelo mais robusto.

Portanto, os modelos implementados oferecem uma abordagem preliminar, eficiente e baseada em soluções analíticas amplamente utilizadas para avaliar o fenômeno de reativação de falhas. Eles possuem simplicidade e possibilidade de generalização para classificar outros cenários de reativação com suas próprias características. É possível utilizar os ajustes mencionados para desenvolver classificações mais completas e complexas.

## 5.1 SUGESTÕES DE TRABALHOS FUTUROS

Como sugestões para trabalhos futuros para continuidade desta linha de pesquisa, são propostas:

- Utilizar diferentes modelos constitutivos para representar o comportamento mecânico do material nas zonas de falha;
- Estudar a influência da mudança de parâmetros como coesão, ângulo de atrito, espessura do reservatório, porosidade, no processo de reativação de falhas geológicas;
- Definir área de estudo com base em afloramentos;
- Caracterizar as propriedades dos materiais utilizando ensaios laboratoriais e de campo;
- Utilizar de simulações de Monte Carlo para gerar os parâmetros de entradas dos modelos de classificação.
- Realizar um estudo de sensibilidade e aplicar técnicas estatísticas para definir os melhores parâmetros de entrada para a RNA e LDA.
- Utilizar modelagens numéricas para treinar e validar RNA e LDA com parâmetros de entrada que descrevam o fenômeno mais adequadamente;
- Determinar a máxima pressão de injeção associado ao processo de reativação.

## REFERÊNCIAS

ABAD, A. R. B. et al. Robust hybrid machine learning algorithms for gas flow rates prediction through wellhead chokes in gas condensate fields. **Fuel**, v. 308, p. 121872, jan. 2022.

ABDELGHANY, W. K. et al. Implications of machine learning on geomechanical characterization and sand management: a case study from Hilal field, Gulf of Suez, Egypt. **Journal of Petroleum Exploration and Production Technology**, v. 13, n. 1, p. 297–312, 6 ago. 2022.

ABBASSI, F. et al. Poromechanics of Fractured/Faulted Reservoirs During Fluid Injection Based on Continuum Damage Modeling and Machine Learning. **Natural Resources Research**, v. 32, n. 1, p. 413–430, 30 out. 2022.

AHMADI GOLTAPPEH, S. et al. Artificial neural network-based Caprock structural reliability analysis for CO<sub>2</sub> injection site—an example from northern North Sea. **Energies**, v. 15, n. 9, p. 3365, 2022.

ALAEI, H. K.; ALAEI, H. K. Estimation of Petroleum Reservoir Parameters Using an Integrated Approach Neural Network, Principal Component Analysis and Fisher Discriminant Analysis. **Petroleum Science and Technology**, v. 31, n. 5, p. 530–539, mar. 2013.

ANDERSON, E. M. The dynamics of faulting and dyke formation with applications to Britain: Hafner Pub. Co., **Edinburgh**, v. 660, 1951.

ANDERSON, J. A. **An introduction to neural networks**. MIT Press: Cambridge, MA, USA; London, UK, 1995.

BANERJEE, A.; CHATTERJEE, R. Pore pressure modeling and in situ stress determination in Raniganj basin, India. **Bulletin of Engineering Geology and the Environment**, v. 81, n. 1, 28 dez. 2021.

BEHESHT ABAD, A. R. et al. Hybrid machine learning algorithms to predict condensate viscosity in the near wellbore regions of gas condensate reservoirs. **Journal of Natural Gas Science and Engineering**, v. 95, p. 104210, nov. 2021.

BISHOP, C. M., & NASRABADI, N. M. **Pattern recognition and machine learning**. v. 4, n. 4, p. 738. New York: Springer, 2006.

BOUTAREAUD, S. et al. Permeability structure and co-seismic thermal pressurization on fault branches: insights from the Usukidani fault, Japan. **Geological Society, London, Special Publications**, v. 299, n. 1, p. 341–361, jan. 2008.

BUITINCK, Lars et al. API design for machine learning software: experiences from the scikit-learn project. **arXiv preprint arXiv:1309.0238**, 2013. DOI: <https://doi.org/10.48550/arXiv.1309.0238>.

CAINE, J. S.; EVANS, J. P.; FORSTER, C. B. Fault zone architecture and permeability structure. **Geology**, v. 24, n. 11, p. 1025, 1996.

CAPPA, F.; RUTQVIST, J. Modeling of coupled deformation and permeability evolution during fault reactivation induced by deep underground injection of CO<sub>2</sub>. **International Journal of Greenhouse Gas Control**, v. 5, n. 2, p. 336–346, mar. 2011.

CELESTINO, M. A. L. et al. Fault damage zones width: Implications for the tectonic evolution of the northern border of the Araripe Basin, Brazil, NE Brazil. **Journal of Structural Geology**, v. 138, p. 104116, set. 2020.

Cisneros, O. S. M. CISNEROS, Oscar Simón Melgar. **Comparação de esquemas de acoplamento hidro-mecânico via elementos finitos em problemas de fluxo e simulação de reservatórios de petróleo**.2016.55f. Dissertação (Mestrado em Engenharia Civil) - Universidade Federal de Pernambuco, Recife, 2016.

DE LUCA, S.M.; OLIVEIRA, R.F. (2014) Fault Reactivation as Mechanism of Early Water Production in Unconsolidated Sandstones Reservoirs. In: **SPE Annual Technical Conference and Exhibition**, OnePetro, Amsterdam, Holanda, 2014.

FARSI, M. et al. Predicting Formation Pore-Pressure from Well-Log Data with Hybrid Machine-Learning Optimization Algorithms. **Natural Resources Research**, v. 30, n. 5, p. 3455–3481, 1 abr. 2021.

FISHER, R. A. THE USE OF MULTIPLE MEASUREMENTS IN TAXONOMIC PROBLEMS. **Annals of Eugenics**, v. 7, n. 2, p. 179–188, set. 1936.

Fossen, H. *Structural Geology*. 2nd ed., Cambridge: Cambridge University Press, 2016.

GAO, K. et al. Delineating faults at the soda lake geothermal field using machine learning. In: **Proceedings of the 46th Workshop on Geothermal Reservoir Engineering, Stanford University, Stanford, CA, USA**. 2021. p. 16-18.

GEUZAINÉ, C.; REMACLE, J.-F. Gmsh: A 3-D finite element mesh generator with built-in pre- and post-processing facilities. **International Journal for Numerical Methods in Engineering**, v. 79, n. 11, p. 1309–1331, 7 maio 2009.

GOMES, I.F. et al. Coupled Hydro-Mechanical Modeling of Fault Reactivation Problem in Petroleum Reservoir Considering the Fault Divided into Core and Damage Zones. In: **AAPG/SEG International Conference & Exhibition**, Cancun, Mexico, 2016.

GRANDINI, M.; BAGLI, E.; VISANI, G. Metrics for multi-class classification: an overview. **arXiv preprint arXiv:2008.05756**, 2020.

GUDMUNDSSON, A. Effects of Young's modulus on fault displacement. **Comptes Rendus Geoscience**, v. 336, n. 1, p. 85–92, jan. 2004.

GUIMARÃES L. J. N.; I. F. GOMES; J. P. V. VALADARES. Influence of Mechanical Constitutive Model on the Coupled Hydro-Geomechanical Analysis of Fault Reactivation. In: **Reservoir Simulation Symposium - Society of Petroleum Engineers SPE**, The Woodlands, 2-4 February 2009, SPE-119168-PP. Texas: EUA.

GUIMARÃES L. J. N et al. Numerical modelling of CO<sub>2</sub> flow through deformable aquifers with the possibility of fault reactivation. In: **Sixth International Congress on Environmental Geotechnics**. New Delhi, India, 2010.

HADDAD, M.; EICHHUBL, P. Poroelastic models for fault reactivation in response to concurrent injection and production in stacked reservoirs. **Geomechanics for Energy and the Environment**, v. 24, p. 100181, dez. 2020.

HAMZEH, A. **Machine Learning Methods in Fault Detection**. Dissertation (Master's Degree in Petroleum and Mining Engineering)- Politecnico di Torino, Torino, Italy, 2021.

HASHEMI, S. R. et al. Exclusive independent probability estimation using deep 3D fully convolutional DenseNets: Application to Isointense infant brain MRI segmentation. In: **International Conference on Medical Imaging with Deep Learning**. PMLR, 2019. p. 260-272.

HASTIE, Trevor et al. **The elements of statistical learning: data mining, inference, and prediction**. v. 2, pp. 1-758. New York: Springer, 2009.

HE, M.; LI, Q.; LI, X. Injection-Induced Seismic Risk Management Using Machine Learning Methodology – A Perspective Study. **Frontiers in Earth Science**, v. 8, 22 jun. 2020.

HEATON, Jeff. Applications of Deep Neural Networks with Keras. **arXiv e-prints**, p. arXiv: 2009.05673, 2020.

HILLIS, R. R. Coupled changes in pore pressure and stress in oil fields and sedimentary basins. **Petroleum Geoscience**, v. 7, n. 4, p. 419–425, dez. 2001.

HUYGHE, P.; MUGNIER, J.-L. The influence of depth on reactivation in normal faulting. **Journal of Structural Geology**, v. 14, n. 8-9, p. 991–998, ago. 1992.

JAFARIZADEH, F., MOTAMEDI, M.H., AHMADI, M. Application of machine learning approaches in reservoir integrated management regarding geomechanical analysis. In: **3rd International Congress on Engineering, Tehnology & Innovation**, Darmstadt, Germany, 2021.

KHAN, S.; KHULIEF, Y. A.; AL-SHUHAIL, A. A. Effects of reservoir size and boundary conditions on pore-pressure buildup and fault reactivation during CO<sub>2</sub>

injection in deep geological reservoirs. **Environmental Earth Sciences**, v. 79, n. 12, jun. 2020.

LESUEUR, M.; POULET, T.; VEVEAKIS, M. Fault reactivation during fluid production, modelled as a multi-physics multi-scale instability. **E3S Web of Conferences**, v. 205, p. 03002, 2020.

MCCULLOCH, W. S.; PITTS, W. A logical calculus of the ideas immanent in nervous activity. **The Bulletin of Mathematical Biophysics**, v. 5, n. 4, p. 115–133, dez. 1943.

MIAH, M. I. Predictive models and feature ranking in reservoir geomechanics: A critical review and research guidelines. **Journal of Natural Gas Science and Engineering**, v. 82, p. 103493, out. 2020.

MICARELLI, L.; BENEDICTO, A.; WIBBERLEY, C. A. J. Structural evolution and permeability of normal fault zones in highly porous carbonate rocks. **Journal of Structural Geology**, v. 28, n. 7, p. 1214–1227, jul. 2006.

Moraes, A. **Comportamento Mecânico de Zonas de Falhas**.2004.320f. Tese (Doutorado em Geologia) - Universidade Federal do Rio de Janeiro, Rio de Janeiro, 2004.

MOTAMEDI, M.; FARSHAD, F. J. An Overview on Applications of Machine learning in Petroleum Engineering. In: **3rd International Congress on Science and Engineering, Hamburg**. 2020.

NACHT, P. K. et al. Investigation of Geological Fault Reactivation and Opening. **Mecânica Computacional**, v. 29, n. 89, p. 8687–8697, 2010.

NEVES, M. C.; PAIVA, L. T.; LUIS, J. Software for slip-tendency analysis in 3D: A plug-in for Coulomb. **Computers & Geosciences**, v. 35, n. 12, p. 2345–2352, dez. 2009.

OLIVELLA, S. et al. Nonisothermal multiphase flow of brine and gas through saline media. **Transport in Porous Media**, v. 15, n. 3, p. 271–293, jun. 1994.

OLIVELLA, S. et al. Numerical formulation for a simulator (CODE\_BRIGHT) for the coupled analysis of saline media. **Engineering Computations**, v. 13, n. 7, p. 87–112, 1 nov. 1996.

OYEDERE, M.; GRAY, K. ROP and TOB optimization using machine learning classification algorithms. **Journal of Natural Gas Science and Engineering**, v. 77, p. 103230, maio 2020.

PARK, J.-W. et al. Modeling of fluid injection-induced fault reactivation using coupled fluid flow and mechanical interface model. **International Journal of Rock Mechanics and Mining Sciences**, v. 132, p. 104373, ago. 2020.

PEREIRA, L. C. et al. Coupled hydro-mechanical fault reactivation analysis incorporating evidence theory for uncertainty quantification. **Computers and Geotechnics**, v. 56, p. 202–215, mar. 2014.

Pereira, L. C. **Quantificação de Incertezas Aplicada à Geomecânica de Reservatórios**. 2015. 195f. Tese (Doutorado em Engenharia Civil) - Universidade Federal de Pernambuco, Recife, 2015.

RAY, S. A quick review of machine learning algorithms. In: **2019 International conference on machine learning, big data, cloud and parallel computing (COMITCon)**. IEEE, 2019. p. 35-39.

RODRIGUES, T. **Análise Numérica de reativação de zonas de falhas geológicas devido à produção de petróleo**. 2019. 66f. Dissertação (Mestrado em Engenharia Civil) - Universidade Federal de Pernambuco, Recife, 2019.

RUMELHART, D. E. et al. A general framework for parallel distributed processing. **Parallel distributed processing: Explorations in the microstructure of cognition**, v. 1, n. 45-76, p. 26, 1986.

RUMELHART, D. E.; HINTON, G. E.; WILLIAMS, R. J. Learning representations by back-propagating errors. **Nature**, v. 323, n. 6088, p. 533–536, out. 1986.

RUTQVIST, J. The Geomechanics of CO2 Storage in Deep Sedimentary Formations. **Geotechnical and Geological Engineering**, v. 30, n. 3, p. 525–551, 12 jan. 2012.

SEYEDI, D. M.; VU, M.N.; POUYA, A. A two-scale hydromechanical model for fault zones accounting for their heterogeneous structure. **Computers and Geotechnics**, v. 68, p. 8–16, jul. 2015.

SHARMA, S.; SHARMA, S.; ATHAIYA, A. Activation functions in neural networks. **Towards Data Sci**, v. 6, n. 12, p. 310-316, 2017.

SIBSON, R. H. Crustal stress, faulting and fluid flow. **Geological Society, London, Special Publications**, v. 78, n. 1, p. 69–84, 1994.

SILER, D. L.; HINZ, N. H.; FAULDS, J. E. Stress concentrations at structural discontinuities in active fault zones in the western United States: Implications for permeability and fluid flow in geothermal fields. **Geological Society of America bulletin**, v. 130, n. 7–8, p. 1273–1288, 2018.

SILER, D. L. et al. Machine learning to identify geologic factors associated with production in geothermal fields: a case-study using 3D geologic data, Brady geothermal field, Nevada. **Geothermal Energy**, v. 9, n. 1, 30 jun. 2021.

SKURTVEIT, E. et al. 3D fault integrity screening for Smeaheia CO2 injection site. In: **14th Greenhouse Gas Control Technologies Conference Melbourne**. 2018. p. 21-26.

SOLTANZADEH, H.; HAWKES, C. D. Semi-analytical models for stress change and fault reactivation induced by reservoir production and injection. **Journal of Petroleum Science and Engineering**, v. 60, n. 2, p. 71–85, fev. 2008.

SOLTANZADEH, H.; HAWKES, C. D. Assessing fault reactivation tendency within and surrounding porous reservoirs during fluid production or injection. **International Journal of Rock Mechanics and Mining Sciences**, v. 46, n. 1, p. 1–7, jan. 2009.

TAGHIPOUR, M. et al. A Geomechanical Evaluation of Fault Reactivation Using Analytical Methods and Numerical Simulation. **Rock Mechanics and Rock Engineering**, v. 54, n. 2, p. 695–719, 25 nov. 2020.

URPI, L. et al. Hydro-mechanical fault reactivation modeling based on elasto-plasticity with embedded weakness planes. **Journal of Rock Mechanics and Geotechnical Engineering**, v. 12, n. 4, p. 877–885, ago. 2020.

VAUNAT, J.; OLIVELLA, S. Code\_Bright/GiD: A 3-D program for thermo-hydro-mechanical analysis in geological media. In: **1st Conference on Advances and Applications of GiD. Publicació CIMNE N°-210**. CIMNE. 2002. p. 91-94.

WETZEL, M.; KEMPKA, T.; KÜHN, M. Quantifying Rock Weakening Due to Decreasing Calcite Mineral Content by Numerical Simulations. **Materials**, v. 11, n. 4, p. 542, 1 abr. 2018.

WIPRUT, D.; ZOBACK, M. D. Fault reactivation and fluid flow along a previously dormant normal fault in the northern North Sea. **Geology**, v. 28, n. 7, p. 595–598, 2000.

WIPRUT, D.; ZOBACK, M. D. Fault reactivation, leakage potential, and hydrocarbon column heights in the northern North Sea. In: **Norwegian Petroleum Society Special Publications**. Elsevier, 2002. p. 203-219.

YUAN, Z. et al. An enhanced fault-detection method based on adaptive spectral decomposition and super-resolution deep learning. **Interpretation**, v. 7, n. 3, p. T713–T725, 1 ago. 2019.

ZHU, D. et al. Fault Detection by Using Instance Segmentation. In: **International Petroleum Technology Conference**. OnePetro, 2021.

ZOBACK, M. D. **Reservoir Geomechanics**. Cambridge, England: Cambridge University Press, 2010. 449p.