UNIVERSIDADE FEDERAL DE JUIZ DE FORA
FACULDADE DE ENGENHARIA
ENGENHARIA COMPUTACIONAL

Rodrigo Oliveira Silva

Abordagem evolutiva de aprendizado de máquina para caracterização litológica de poços de exploração de petróleo.

Juiz de Fora
2022

Rodrigo Oliveira Silva

Abordagem evolutiva de aprendizado de máquina para caracterização litológica de poços de exploração de petróleo.

Trabalho de conclusão de curso apresentado á Faculdade de Engenhariada Universidade Federal de Juiz de Fora como requisito parcial à obtenção do grau de bacharel em Engenharia Computacional.

Orientador: Prof. Dr. Leonardo Goliatt da Fonseca

Juiz de Fora

2022

Ficha catalográfica elaborada através do Modelo Latex do CDC da UFJF
com os dados fornecidos pelo(a) autor(a)

Silva, Rodrigo Oliveira.
Abordagem evolutiva de aprendizado de máquina para caracterização litológica de poços de exploração de petróleo. / Rodrigo Oliveira Silva . – 2022.
42 f. : il.

Orientador: Leonardo Goliatt da Fonseca
Trabalho de Conclusão de Curso (graduação) – Universidade Federal de Juiz de Fora, Faculdade de Engenharia. Engenharia Computacional, 2022.

1. Carbono Orgânico Total. 2. Aprendizado de Máquina. 3. Redes Neurais Convolucionais. I. Fonseca, Leonardo Goliatt da, orient.

Rodrigo Oliveira Silva

Abordagem evolutiva de aprendizado de máquina para caracterização litológica de poços de exploração de petróleo.

Trabalho de conclusão de curso apresentado á Faculdade de Engenhariada Universidade Federal de Juiz de Fora como requisito parcial à obtenção do grau de bacharel em Engenharia Computacional.

Aprovada em 05 de Agosto de 2022

BANCA EXAMINADORA

Prof. Dr. Leonardo Goliatt da Fonseca - Orientador
Universidade Federal de Juiz de Fora

Prof^a Dra. Camila Martins Saporetti
Universidade do Estado de Minas Gerais

Prof. Me. Samuel da Costa Alves Basílio
Centro Federal de ensino tecnologico de leopoldina

Dedico este trabalho aos meus pais, irmão e amigos, que me ajudaram e apoiaram em todos os momentos de minha jornada neste curso de graduação.

AGRADECIMENTOS

Agradeço primeiramente a meus pais, Dirce Oliveira Silva e Robson José Silva pelo amor, incentivo e apoio incondicional. Por nunca terem medido esforços para ajudar e aconselhar e me proporcionar um ensino de qualidade. Obrigado por tudo e pela paciência em todos esses anos.

Agradeço a meu irmão, Rafael Oliveira Silva, pelo companheirismo, amizade e pelo apoio em todos os momentos difíceis de minha jornada acadêmica e de minha vida. Obrigado por todos os conselhos, bem como palavras de conforto e motivação.

Agradeço a meu orientador, Leonardo Goliatt da Fonseca, por esta e muitas outras oportunidades e apoio na elaboração deste trabalho. Agradeço sua confiança, incentivos e dedicação inabalável. Agradeço a todo os professores e professoras da Universidade Federal de Juiz de Fora que fizeram parte desta minha jornada acadêmica.

Agradeço a meus queridos amigos com quem convivi intensamente durante os últimos anos, pelo apoio, força, companheirismo, risadas e assistência inabalável.

Agradeço à Universidade Federal de Juiz de Fora por toda a infraestrutura que permitiu chegar a minha formação. Agradeço ao Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq) por apoiar e proporcionar o desenvolvimento deste trabalho.

RESUMO

O nível de carbono orgânico total das amostras de rochas é a principal medida quantitativa e qualitativa da quantidade de matéria orgânica presente em uma bacia. Muitas vezes é estimado manualmente através do estudo de amostras de rochas de origem. Esta abordagem, no entanto, requer tempo e dinheiro porque depende de amostras retiradas de vários intervalos de poços em rochas geradoras. Como resultado, tentativas de pesquisa têm sido feitas para auxiliar nessa empreitada. Algoritmos de aprendizado de máquina surgem como uma alternativa para fornecer estimativas de carbono orgânico total com base em registros de poços de dados e análise estratigráfica. Considerando esse cenário, a pesquisa atual sugere que a estimativa de carbono orgânico total seja automatizada usando abordagens de aprendizado de máquina. Para adicionar flexibilidade ao modelo, a seleção dos parâmetros dos modelos foi realizada usando uma abordagem com meta-heurísticas emparelhadas com validação cruzada. Esta técnica computacional permite a identificação de modelos com maior potencial de generalização. Foram usados Redes Neurais Convolucionais (RNC), Extreme Learning Machine, Elastic Net Linear Model e Extreme Gradient Boost. A abordagem sugerida foi validada usando amostras de algumas bacias sedimentares. Em várias métricas estudadas, a técnica RNC se sobressai às demais abordagens, demonstrando capacidade em auxiliar geólogos na previsão de concentrações de carbono orgânico total.

Palavras-chave: COT. Redes Neurais Convolucionais. Aprendizado de Máquina.

ABSTRACT

The total organic carbon of rock samples is the main detailed and qualitative measure of the organic matter present in a basin. It is often manually calculated by studying source samples. This approach, however, requires time and money because it relies on withdrawals from various well intervals in source rocks. As a result, research has been carried out to assist in this endeavor. Machine learning algorithms are an alternative to providing estimates of total organic carbon based on data well records and stratigraphic analysis. In this scenario, the current one suggests an estimate of total organic carbon that is thought using the machine learning approach. To add flexibility to the model, the selection of the cross parameters was performed using a standard heuristic approach with validation. This computational technique allows the identification of models with greater generalization potential. Convolutional Neural Networks (CNN), Extreme Learning Machine, Elastic Net Linear Model and Extreme Gradient Boost were used. The suggested approach was validated using samples from some sedimentary basins. In research capacity, the RNC technique excels approaches, demonstrating in helping the prediction of total organic carbon.

Keywords: TOC. Convolutional Neural Networks. Machine Learning.

LISTA DE ILUSTRAÇÕES

LISTA DE TABELAS

LISTA DE ABREVIATURAS E SIGLAS

CE	Computação Evolutiva
COT	Carbono Orgânico Total
Conv2D	Rede Neural Convolucional desnvolvida no trabalho
DB	Base de Dados (*Data Base*)
ED	Evolução Diferencial
ELM	*Extreme Machine Learning*
EN	*Elastic Linear Net*
GloRiSe	*Global River Sediments*
IEEE	*Institute of Electrical and Electronics Engineers*
MAE	*Mean Absolut Error*
MARE	*Mean Absolute Relative Error*
ML	Aprendizado de Máquina (*Machine Learning*)
R^2	Coeficiente de Determinação
RMSE	*Root-Mean-Square Deviation*
RNC	Redes Neurais Convolucionais
RNR	Redes Neurais Recorrentes
UFJF	Universidade Federal de Juiz de Fora
WI	*Index of Agreement*
XGBoost	*eXtreme Gradient Boosting*

LISTA DE SÍMBOLOS

$\sum$	Somatório
$\in$	Pertence
$\leq$	Menor ou igual
$\|x\|$	Módulo
$\|\|x\|\|$	Norma
$\sqrt{x}$	Raiz quadrada
$\mathbb{N}$	Números Naturais
$\mathbb{R}$	Números Reais
α	Letra grega: Alfa minúsculo
β	Letra grega: Beta minúsculo
η	Letra grega: Eta minúsculo
ϕ	Letra grega: Fi minúsculo
Φ	Letra grega: Fi maiúsculo
γ	Letra grega: Gama minúsculo
λ	Letra grega: Lambda minúsculo

SUMÁRIO

1 INTRODUÇÃO

Os recursos naturais são substâncias encontradas na natureza que são essenciais ao ser humano para sua existência e conforto, bem como para o avanço da civilização. O avanço da tecnologia é o principal responsável por disponibilizar uma parte desses recursos à humanidade para uso generalizado. Esses recursos são extraídos para uso como materiais ou fontes de energia para os seres humanos. A energia natural, que pode ser classificada em formas renováveis e não renováveis, é o recurso que as pessoas e os processos industriais mais utilizam. A energia não renovável é responsável por mais de 80% do uso global de energia. A fonte de energia mais utilizada, o petróleo serve de base para a geração de eletricidade, bem como o combustível para veículos e equipamentos industriais, apesar de várias novas fontes de energia estarem constantemente chamando a atenção (1).

Mapear e determinar as propriedades da matéria orgânica no processo de formação rochosa é necessário para avaliar o potencial dos reservatórios para a produção de petróleo e gás. Uma forma é através do cálculo do Carbono Orgânico Total (COT). O COT define a percentagem de compostos químicos nas rochas sedimentares, ou seja, o seu potencial de geração (2). As rochas geradoras têm um mínimo de 0,5 a 1% do teor de carbono orgânico total, chegando a 10%. Vale ressaltar que as rochas não geradoras também possuem matéria orgânica, mas seu teor costuma ser inferior a 1% (3).

A determinação do COT é feita principalmente por meio de análises geoquímicas laboratoriais, que dependem de fragmentos da rocha ou mesmo de dados de poços, no entanto, este procedimento comumente usado não é prático, pois depende de amostras obtidas de vários intervalos de poços em rochas geradoras, o que leva a um aumento nos gastos. Pesquisadores e especialistas têm divulgado cada vez mais estudos propondo metodologias para estimar o COT a partir de registros de poços.

Métodos inteligentes de análise e interpretação de dados, como redes neurais, raciocínio difuso e computação evolutiva, estão se tornando cada vez mais importantes instrumentos para coletar dados e transformá-los em conhecimento.

Essas abordagens inteligentes podem ser aplicadas no setor de petróleo e gás para descoberta de conhecimento de muitos tipos de dados, incluindo sísmica 3D, dados geológicos, perfil de poço e dados de produção, bem como para análise de incerteza, avaliação de risco, fusão e mineração de dados. Esses métodos também podem ser essenciais para encontrar e extrair quaisquer depósitos residuais de petróleo e gás. As técnicas podem ser empregadas como ferramenta para: diminuir o risco associado à exploração; reduzir o custo de exploração e produção; aumentar a recuperação através de uma produção mais eficaz; e prolongar a vida útil dos poços produtores (4).

2 REVISÃO BIBLIOGRÁFICA

O presente capítulo destina-se a apresentar trabalhos disponíveis na literatura que contribuem para o entendimento do contexto da presente pesquisa, assim como referências que sustentam a justificativa de avanços em estudos sobre a predição do carbono orgânico total de forma computacional.

Primeiramente apresentam-se trabalhos que exemplificam o desenvolvimento de métodos para a predição do carbono ogânico total com métodos de aprendizado de máquina, sem o ajuste ou busca computacional de seus parâmetros.Tais estudos são os mais clássicos na literatura, porém abrem mão da capacidade de métodos evolutivos nas buscas de melhores parâmetros e consequentemente, melhores modelos.

Tan, Liu e Zhang(5) aplicam uma rede neural RBF (*radial basis function*) para a previsão do carbono ogânico total, a rede neural contém três camadas, incluindo camada de entrada, camada oculta e a camada de saída.Campo de xisto orgânico foram estudados para a previsão de COT, e os resultados da previsão usando o método RBF foram comparados com os do método $\Delta\ log\ R$. Segundo os autores a rede RBF é uma rede *feed-forward* com bom desempenho, possuindo propriedades de aproximação global e melhor desempenho de aproximação, já que a análise de erro entre os resultados de previsão e COT medido em laboratório em alguns exemplos indicou que a nova abordagem é mais precisa do que um único método de regressão empírica e mais flexível do que o $\Delta\ log\ R$.

Lee, Wood e Phrampus(6) utilizam um algoritmo de aprendizado de máquina não paramétrico (ou seja, orientado por dados), especificamente k-vizinhos mais próximos (*k-nearest neighbors* - kNN), para estimar a distribuição global do COT do fundo do mar. Esta aplicação conseguiu resultados para o coeficiente de correlação $R = 0.78$, utilizando-se *cross-validation*. Este trabalho também traz uma comparação de métodos com o GMT, que resultou em um coeficiente de correlação $R = 0.39O$. Os autores afirmam que a aplicação do algoritmo de aprendizado de máquina de k-vizinhos mais próximos resulta em uma previsão de carbono orgânico total consistente e facilmente atualizável em todos os pontos do fundo do mar.

Yu et al.(7) dizem que o estudo propõe, um novo método de previsão de COT baseado em Regressão de Processos Gaussiana (*Gaussian Process Regression* - GPR) ligando geoestatística e técnicas de aprendizado de máquina. Os autores afirmam que o método utiliza uma abordagem de regressão não paramétrica em previsões de COT que possui uma melhor capacidade de adaptação e generalização em comparação com outros métodos de aprendizado de máquina. Este trabalho compara o método descrito com sete funções de *kernel* diferentes e utiliza-se do $RMSE$ para a validação, encontrando um valor de $RMSE = 0.344$ para um dos *kernels* descritos.

Shi et al.(8) aplicam dois métodos de aprendizado de máquina, redes neurais

artificiais (ANN) e *Extreme Learning Machine* (ELM). Os dados utilizados para a previsão de COT contém apenas dados petrofísicos, contendo dados como raios gama, porosidade e resistividade. Os autores utilizaram de diferentes métricas para validarem os resultados dos modelos, tais como *RMSE*, *MAE*, *VAF* e R^2.

Handhal et al.(9) avaliam o uso de três modelos de aprendizado de máquina, *Random Forest*(RF), *rotation forest* (rF), *k-nearest neighbors* (KNN) para estimar o COT com base em dados de poços convencionais. Os dados do poço contém resultados de raios gama, acústica, densidade, nêutrons e resistividade profunda. Além disso, os resultados dos modelos discutidos neste trabalho foram comparados com mais dois métodos de aprendizado de máquina, SVR (*Support Vector Machine*) e BPANN (*Backpropagation Artificial Artificial Neural Network*). Os autores deste trabalho também validaram a performance dos resultados usando as métricas *RMSE*, *MAE* e R^2 encontrando valores de R^2 próximos a 0.95 com o algoritmo *Random Forest* na fase de testes.

Podemos perceber que a maioria dos trabalhos comentados anteriormente utilizam de métodos de *ML* juntamente com dados petrográficos dos poços examinados, poucos trabalhos utilizam dados mineralógicos, que são as composições minerais das rochas extraídas.

A seguir, são apresentados uma sequencia de trabalhos deu utilizaram alguma técnica de ajuste e otimização dos parâmetros dos métodos de aprendizado de máquina. São trabalhos que apresentam otimizações simples, como a busca em grade (*Grid Search*) a buscas mais robustas como abordagens evolutivas (*GWO*, *PSO* e *SaDE*).

Elkatatny(10) utilizam de uma abordagem evolutiva para poder encontrar os melhores parâmetros para uma rede ANN (*Artificial Neural Network*) para a previsão de carbono orgânico total. A tecnica descrita é a abordagem evolução diferencial autoadaptativa (*Self adaptive Differential Evolution* - SaDE) onde os autores buscaram otimizar somente a quantidade de neurônios presentes na rede. Segundo os autores, a abordagem teve uma boa performance quando comparada à métrica de erro AAPE (*Average Absolute Percentage Error)* com um valor de 6%.

Mahmoud, Elkatatny e Al-AbdulJabbar(11) faz uso de três algoritmos de aprendizado de máquina, ANN (*Artificial Neural Network*), ANFIS (*Adaptive Neuro-Fuzzy Inference System*) e FNN (*Functional Neural Networks*) para a previsão do carbono orgânico total. Para a otimização dos parâmetros os autores utilizaram-se de um laço com diferentes parâmetros dos algoritmos, afim de encontrar os melhores modelos. Os autores afirmam ter encontrado uma acuracia próxima a 98%.

Siddig, Ibrahim e Elkatatny(12) aplicam diferentes métodos de aprendizado de máquina com dados de três diferentes poços, contendo dados petrofísicos. Dentre os métodos utilizados estão *Support vector machine* - (SVM), *Random Forest* (RF)e *decision tree* (DT). Os autores fazem uso de *Grid-Search* (busca em grade) para poder encontrar os

melhores parâmetros dos algoritmos utilizados. Além disto, os autores também afirma que o modelo RF foi o que teve melhor performance encontrando valores para o coeficiente de determinação $R = 0.93$ a $R = 0.99$ para oa diferentes bases de dados.

Wang, Peng e He(13) utilizaram um modelo LS-SVM (*Least-Squares Support-Vector Machine*). Para a otimização dos parâmetros do método de aprendizado de máquina, os autores utilizaram uma abordagem PSO (*Particle Swarm Optimization*), conseguindo encontrar modelos que retornam bons resultados de previsão de COT. Dados presentes neste trabalho contém dados mineralógico e petrográficos.

Safaei-Farouji e Kadkhodaie(14) utilizaram a quantidade de carbono organico total, juntamente com outras variáveis, para pode estimar quantidade de hidrogênio (HI - *Hydrogen Index*) e oxigênio (HI - *Oxigen Index*) das amostras. Para isto, empregam vários métodos de aprendizado de máquina (*Radial Basis Function* - RBF, *Multi-Layer Perceptron* - MLP, *Random Forest* - RF, *Support Vector Machine* - SVM, *Decision Tree* - DT) juntamente com uma otimização de parâmetros para o MLP via GWO (*Grey Wolf Optimization*), GA (*Genetic Algorithm*) e PSO (*Particle Swarm Optimization*). Os autores afirmam que os resultados obtidos chegam próximos a $R^2_{MLP} = 0.7153$, $R^2_{MLP+PSO} = 0.7115$, $R^2_{MLP+GA} = 0.7028$ e $R^2_{MLP+GWO} = 0.7753$.

Por fim, é apresentado um trabalhos de revisão que apresenta a aplicação do método de redes neurais convolucionais em uma dimensão. O intuito é justificar que apesar de existirem trabalhos que utilizam este tipo de rede, pouquíssimos trabalhos fizeram o uso de redes convolucionais em duas dimensões.

Asante-Okyere, Ziggah e Marfo(15) apresentam uma abordagem para a previsão do carbono orgânico total utilizando-se redes neurais convolucionais 1D, a rede possui três camadas de convolução seguida de apenas uma camada conectada contendo 64 neurônios. Os autores fazem uso de dois tipos de redes, uma recebe dados mineralógicos e petrofísicos (perfilagem), e a outra apenas os dados petrofísicos (perfilagem). Os autores afirmam que a performance do primeiro modelos foi melhor do que a do segundo. Nenhum tipo de otimização de parâmetros foi feito neste trabalho

Desse forma, no capítulo a seguir, serão apresentadas as descrições dos materiais e métodos utilizados para o desenvolvimento da presente pesquisa, o que nos possibilita identificar onde o presente trabalho está inserido no contexto proposto.

3 MATERIAIS E MÉTODOS

3.1 ÁREA DE ESTUDO E CONJUNTO DE DADOS

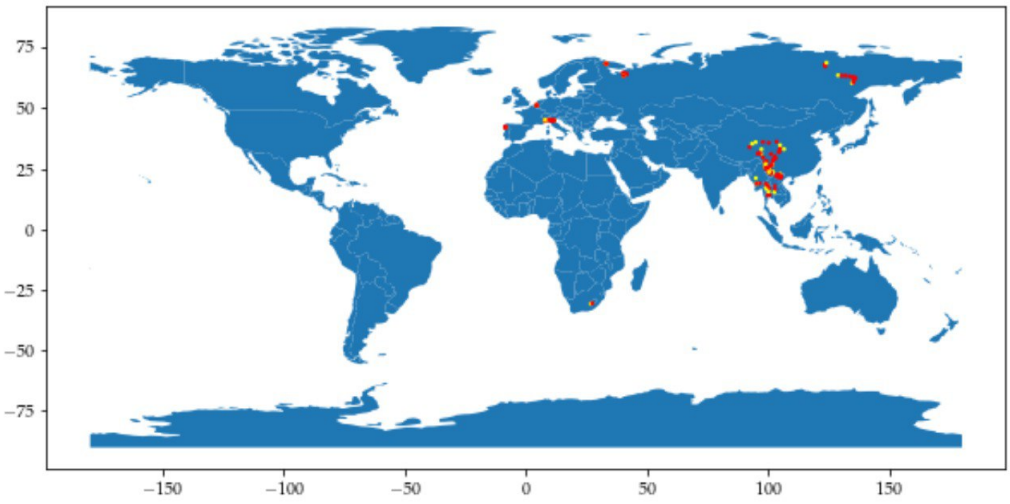
Neste trabalho foram utilizadas duas bases de dados para poder avaliar o desempenho e capacidade de previsão dos métodos implementados.

3.1.1 Global River Sediments (GloRiSe)

GloRiSe é um conjunto de medições de dados mineralógicos e petrográficos realizadas em bacia sedimentares de rios ao redor do globo. A tabela 1 traz as localidades de onde foram extraídos os dados utilizados neste trabalho.

Os dados desta base estão exemplificados na tabela 2, mais informações sobre a base de dados podem ser encontradas em (16).

Figura 1 – Mapa das bacias presentes na base de dados GloRiSe



Fonte: Disponível em (16)

3.1.2 Chuanzhong uplift

Como podemos observar na imagem 2, Chuanzhong Uplift está localizada no coração da Bacia de Sichuan. O estudo baseou-se em dados de três poços que visam Formação Siluriana Longmaxi inferior de Chuanzhong Uplift (15). A formação Longmaxi é um depósito de gás de xisto marinho rico em matéria orgânica. A bacia de xisto de Longmaxi inclui argila, quartzo, feldspato, calcita, dolomita e pirita. A tabela 3 apresenta o conjunto de dados usado neste trabalho.

Tabela 1 – Regiões de extração de dados GloRiSe

País	**Região**
Africa do Sul	South Africa
Bélgica	Confluência de Rupel e Scheldt na Bélgica
China	Upper Changjian Upper Huanghe
Espanha	Galiza
Itália	Nordeste da Itália Norte da Itália
Mianmar	Centro de Mianmar Lower Salween Sul de Mianmar
Russia	Kola Lower Lena White Sea
Tailândia	Centro da Tailândia Sul da Tailândia
Vietnã	Lower Hong He Lower MekongSul do Vietnã

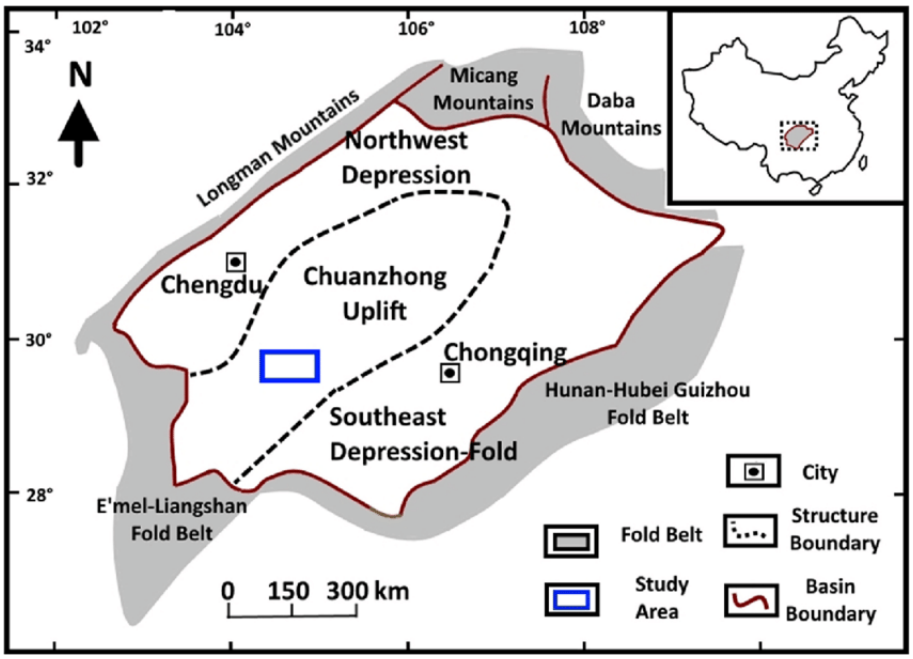
Tabela 2 – Composição mineralógica percentual e TOC das amostras da base GloRiSe

	K2O(%)	Fe2O3T(%)	SiO2(%)	Al2O3(%)	P2O5(%)	TOC(%)
0	1.90	6.30	58.60	8.70	1.40	4.20
1	1.90	6.20	60.30	8.10	1.70	6.30
2	1.80	6.90	49.00	8.60	1.90	6.50
3	1.90	6.00	55.80	7.70	1.60	6.60
4	2.10	6.20	54.60	8.90	1.60	7.00
5	1.90	7.00	49.40	8.90	1.60	7.40
6	1.80	7.30	49.40	8.50	1.90	8.70
7	1.70	6.30	49.20	7.90	2.00	10.10
8	1.19	28.06	32.39	11.31	0.18	0.05
9	2.75	1.96	73.67	8.60	0.07	0.07
10	1.31	4.09	49.67	4.95	0.17	0.09
⋮	⋮	⋮	⋮	⋮	⋮	⋮
148	2.14	3.41	76.82	11.45	0.09	0.52
149	3.06	5.82	62.05	16.39	0.18	1.44

3.2 MÉTODOS DE APRENDIZADO DE MÁQUINA

Este trabalho tem como objetivo principal o desenvolvimento de uma Rede Neural Convolucional (RNC), a qual tem como entrada dados em duas dimensões, para que se possa realizar a previsão do conteúdo de carbono orgânico em rochas geradoras, e comparar seu desempenho com outros algoritmos já consolidados na literatura. O desempenho da RNC desenvolvida será comparado com os algoritmos Extreme Machine Learning (EML), Elastic Linear Net (EN) e Extreme Gradient Boosting (XGBoost). A seguir temos uma

Figura 2 – Mapa da bacia de Chuanzhong uplift



Fonte: Adaptado de Asante-Okyere *et al.* (15)

Tabela 3 – Composição mineralógica percentual e TOC das amostras da bacia Chuanzhong uplift

	Cl(%)	Qz(%)	KFd(%)	Ca(%)	Do(%)	Py(%)	CFp(%)	TOC(%)
0	48.0	35.0	4.0	4.0	1.0	7.0	1.0	0.8
1	50.0	34.0	3.0	7.0	1.0	4.0	1.0	1.3
2	36.0	36.0	2.0	7.0	15.0	4.0	0.0	2.6
3	26.0	26.0	3.0	27.0	13.0	5.0	0.0	2.8
4	32.0	35.0	4.0	19.0	5.0	5.0	0.0	2.4
5	28.0	41.0	1.0	6.0	20.0	4.0	0.0	4.3
6	28.0	39.0	2.0	14.0	12.0	5.0	0.0	3.1
7	54.0	35.0	4.0	2.0	1.0	3.0	1.0	1.0
8	50.0	34.0	4.0	3.0	3.0	5.0	1.0	1.0
9	48.0	37.0	4.0	4.0	2.0	4.0	1.0	1.3
10	50.0	30.0	3.0	4.0	3.0	9.0	1.0	0.9
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
89	34.6	27.4	3.0	18.6	12.0	3.5	0.9	2.6
90	42.2	25.5	8.7	5.7	9.1	6.3	2.5	4.0

explicação sobre os métodos utilizados neste trabalho.

3.2.1 Redes Neurais Convolucionais (RNC)

Existem dois tipos comuns de arquiteturas de redes neurais: redes neurais convolucionais (RNC) e redes neurais recorrentes (RNR) (17). RNCs são usadas para discernir padrões visuais diretamente de imagens enquanto as RNRs destinam-se a detectar padrões

em séries temporais compostas de símbolos ou formas de onda de áudio ou fala. As redes neurais convolucionais são um tipo especializado de redes neurais artificiais e são uma classe de rede neural artificial do tipo *feed-forward*, que vem sendo aplicada com sucesso no processamento e análise de imagens digitais (18).

Uma rede neural convolucional é composta por uma camada de entrada, uma camada convolucional, uma camada de pooling, uma camada totalmente conectada e uma camada de saída, conforme ilustrado na figura 5. Geralmente, a camada de entrada recebe apenas o sinal de entrada e transfere para a camada convolucional onde o cálculo principal é realizado.

A camada de convolução é a camada básica mais simples, mas também a mais crucial em uma RNC. Ele essencialmente envolve ou multiplica a matriz de pixels criada para a imagem ou objeto fornecido para construir um mapa de ativação para a entrada fornecida (19).

Essa camada realiza um produto escalar entre duas matrizes, onde uma matriz é o conjunto de parâmetros que podem ser aprendidos, também conhecido como kernel, e a outra matriz é a entrada dos dados. O kernel é menor que a entrada.

Durante a passagem para frente, o kernel desliza pela altura e largura da imagem, produzindo a representação da imagem daquela região receptiva. Isso produz uma representação bidimensional da imagem,o mapa de ativação, que fornece a resposta do kernel em cada posição espacial da imagem. O tamanho de deslizamento do kernel é chamado de passo (20).

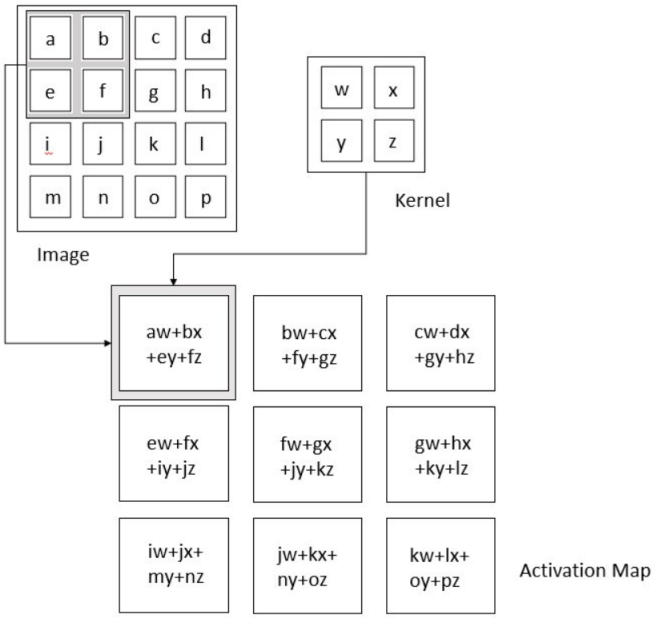
A vantagem fundamental dos mapas de ativação é que eles armazenam todas as qualidades diferenciadoras de uma determinada imagem enquanto minimizam a quantidade de dados que precisam ser processados. Os dados são convoluídos usando uma matriz que é um detector de características (kernel), que é essencialmente um conjunto de valores com os quais a máquina é compatível (19, 20). A figura 3 exemplifica o processo de convolução.

A convolução preserva valores nas mesmas posições do mapa e dos dados, ou seja, valores com valor 1 ou maior que 1, enquanto remove o restante.

Cada camada de convolução consiste em vários canais de convolução (também conhecidos como profundidade ou filtros). Na prática, eles são um número na forma 2^n. Isso é igual ao número de canais na saída de uma camada convolucional. Ou seja, o filtro é a quantidade de kernels que serão utilizados na operação de convolução.

Pooling é um passo fundamental para reduzir ainda mais as dimensões do mapa de ativação, retendo apenas os elementos importantes e minimizando a invariância espacial (19). Como resultado, os recursos que podem ser aprendidos do modelo são reduzidos. Isso contribui para a resolução do problema de *overfitting* (15, 20). O processo de pooling pode ser classificado em várias categorias, incluindo pooling máximo, pooling médio, pooling

Figura 3 – Exemplificação da camada de convolução

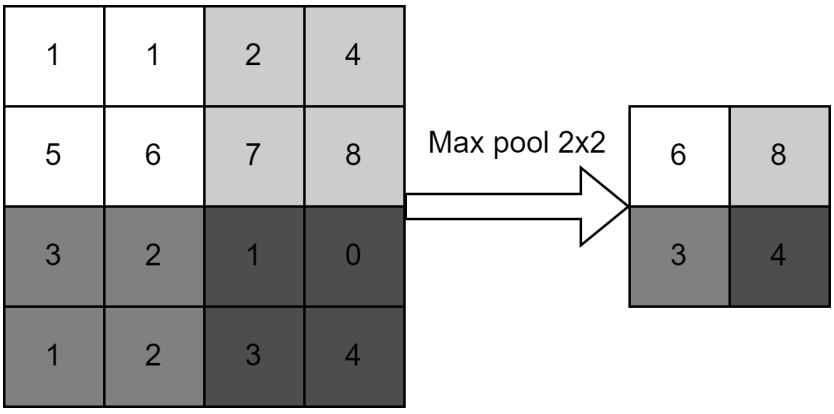


Fonte: Disponível em (21)

estocástico e pooling de pirâmide espacial. O mais proeminente deles é o pooling máximo

O Max Pooling pega o maior valor de cada submatriz do mapa de ativação e cria uma nova matriz a partir dela. Isso garante que o número de variáveis permaneça pequeno, mantendo simultaneamente os elementos importantes (20). Este processo está exemplificado na imagem 4.

Figura 4 – Max pooling com filtros 2x2



Fonte: Produzida pelo autor (2022).

Nas camadas totalmente conectadas os neurônios têm conectividade total com todos os neurônios da camada anterior e seguinte, como visto em uma RNR. Esta é a última camada que é alimentada na rede neural (15). Em geral, as matrizes são achatadas antes de serem transmitidas aos neurônios. É difícil acompanhar os dados após este ponto devido à presença de várias camadas ocultas com pesos variados para a saída de cada

neurônio. Todo o raciocínio e computação de dados é feito aqui (19).

Para o propósito deste estudo, uma biblioteca de rede neural de código aberto codificada em Python Tensor Flow conhecida como KERAS foi usada para desenvolver o modelo. O modelo RNC desenvolvido foi composto por uma camada de entrada, duas camadas convolucionais e pooling, uma camada Flattening, que é uma camada de achatamento onde os dados são transformados para apenas uma dimensão, e duas camadas totalmente conectadas conforme descrito na figura 5.

Entre as camadas conectadas, há procedimentos de dropout. Dropout é simplesmente desconsiderar unidades (ou seja, neurônios) durante a fase de treinamento de uma coleção aleatória de neurônios. Ignorar indica que essas unidades não são levadas em consideração durante uma determinada passagem da rede. Nós individuais são retirados da rede com probabilidade $1 - p$ ou retidos com probabilidade p, resultando em uma rede menor; arestas de entrada e saída para um nó descartado também são excluídas.

3.2.2 Extreme Learning Machine (ELM)

Extreme Machine Learning (ELM) é um algoritmo que emprega uma rede feedforward de camada única, no qual os pesos da camada oculta podem ser inicializados arbitrariamente, necessitando apenas da otimização dos pesos da camada de saída (22).

Um inverso generalizado de Moore-Penrose pode ser usado para realizar essa otimização. Como resultado, o ELM proporciona uma redução no tempo de computação necessário para o ajuste de parâmetros. Abordagens de descida de gradiente, por exemplo, ou métodos de busca global, para citar alguns, levam muito mais tempo (23).

Ao contrário dos algoritmos de aprendizado típicos, o ELM visa tanto o erro de treinamento mínimo quanto a menor norma de pesos de saída, portanto, para redes neurais feedforward com menos erros de treinamento, quanto menores forem as normas de peso, melhor será o desempenho de generalização das redes (24).

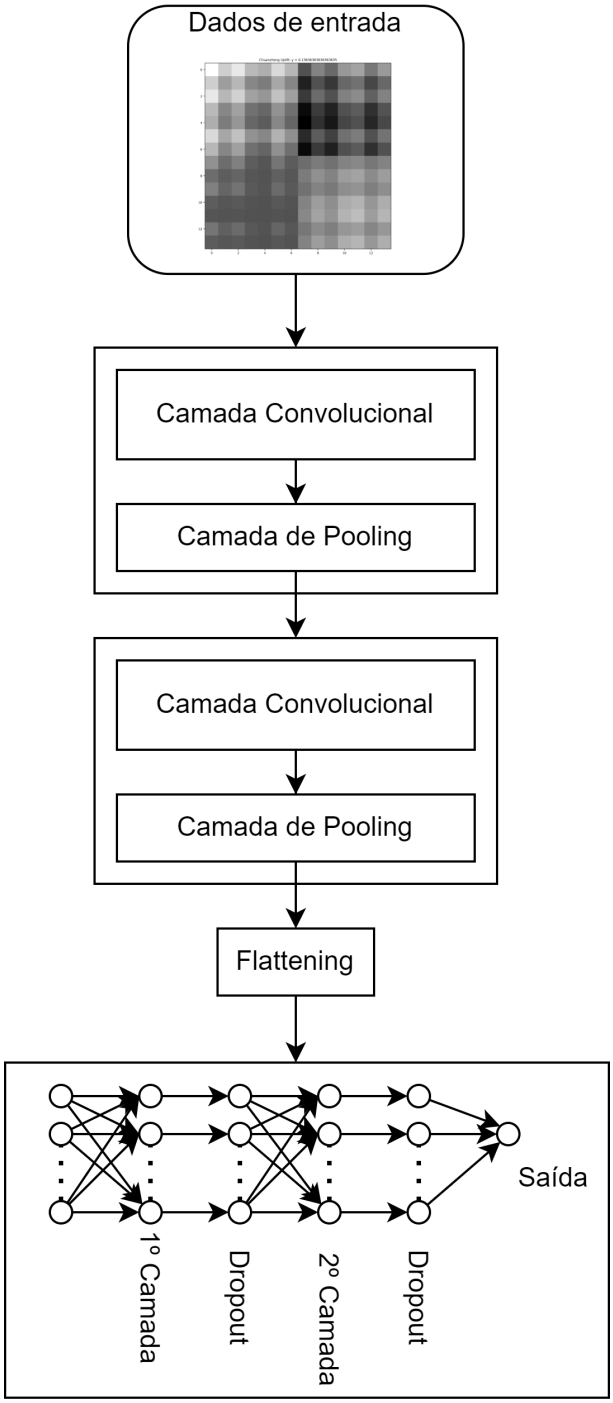
Informações mais completas sobre o método podem ser encontradas em Haung *et al.* (22) e (24).

3.2.3 Elastic Net Linear Model (EN)

Elastic Net (EN) foi apresentada por Zou e Hastie (25) como uma nova abordagem de regularização e seleção de variáveis. EN é um modelo linear regularizado que, devido à sua formulação matemática, realiza seleção de variáveis durante a construção do modelo. A regularização visa melhorar as previsões, reduzindo os coeficientes dos parâmetros para zero. A rede Elastic Net apresenta as características das regressões Lasso e Ridge (26).

O Elastic Net é uma extensão da regressão linear que adiciona penalidades de regularização à função de perda durante o treinamento (25). É um modelo que assume

Figura 5 – Arquitetura do modelo Convolucional



Fonte: Produzida pelo autor (2022).

uma conexão linear entre as variáveis de entrada e a variável de destino. Os coeficientes do modelo são determinados usando um procedimento de otimização que tenta minimizar o erro quadrático total entre as previsões $\hat{y}$ e os valores alvo previstos y (27, 28, 25).

Os coeficientes previstos do modelo podem se tornar enormes na regressão linear, tornando o modelo sensível às entradas e potencialmente instável (28, 25) . Isso é

especialmente verdadeiro para situações com um pequeno número de observações (amostras) ou um número maior de amostras n do que preditores de entrada p ou variáveis (25).

Um método para abordar a estabilidade do modelo de regressão é modificar a função de perda para incorporar custos adicionais para modelos com grandes coeficientes. A regressão linear penalizada refere-se a modelos de regressão linear que aplicam essas funções de perda ajustadas durante o treinamento. EN é um modelo de regressão linear penalizado que treina com as penalidades L1 e L2 (28).

A regressão Elastic Net sempre visa minimizar a seguinte função de perda

$$\frac{\sum_{i=1}^{n}(y_i - x_i^J \hat{\beta})^2}{2n} + \lambda \left(\frac{1-\alpha}{2} \sum_{j=1}^{m} \hat{\beta}_j^2 + \alpha \sum_{j=1}^{m} |\hat{\beta}_j| \right)$$

Além disso, o Elastic Net nos permite ajustar o parâmetro α, com $\alpha = 0$ correspondendo à regressão Ridge e $\alpha = 1$ correspondendo à regressão Lasso. Quando $\alpha = 0$, a função de penalidade é reduzida para regularização L1(ridge) e quando $\alpha = 1$, é reduzida para regularização L2(lasso). Como resultado, podemos otimizar o Elastic Net selecionando um valor α entre 0 e 1, o que diminuirá certos coeficientes e definirá outros como nulos para seleção esparsa. O hiperparâmetro λ na regressão EN depende em grande parte e altamente do hiperparâmetro α (28, 25, 27).

A descrição completa do algoritmo EN pode ser encontrada com mais detalhas em (25) e (29). Detalhes sobre a regressão Lasso e Ridge podem ser encontrados respectivamente em (30) e (26).

3.2.4 Extreme Gradient Boost (XGBoost)

Extreme Gradient Boosting (XGBoost) é uma versão baseada em árvore de decisão do Gradient Boosting na qual os erros são reduzidos por meio de gradiente descendente. Ele tem alto desempenho e é significativamente mais rápido do que as implementações anteriores de gradient boosting usadas para aprendizado supervisionado (31).

O funcionamento do XGBoost pode ser resumido nas seguintes operações: Considere um conjunto de dados com m características e n amostras $(x_1, y_1), \ldots, (x_n, y_n)$, onde $x_i \in \mathbb{R}$ e $y_i \in \mathbb{R}$. Seja $\hat{y}_i$ representar a saída prevista de uma abordagem de árvore de conjunto com base nas equações:

$$\hat{y}_i = \Phi_M(x_i) = \sum_{k=1}^{M} f_k(x_i)$$

onde M é o número de árvores no modelo e fk é a k-ésima árvore de decisão A profundidade da árvore de decisão fk é menor ou igual a $m_{profundidade}$.

No método de boosting aditivo a aproximação é cada vez mais construída como

$$\Phi_M(x) = \Phi_{M-1}(x_i) + \eta f_n(x)$$

onde η é a taxa de aprendizado e $f_n(xi)$ uma árvore de decisão ajustada para minimizar a função de perda Ln

$$L(\Phi) = \sum_i l(y_i, \Phi_{M-1}(x_i)) + \frac{1}{2}\lambda\|w\|^2$$

sendo $l = |\hat{y}_i - y_i|$ a função de perda, $\hat{y}_i$ a saída prevista e y_i a saída real, T é o número de folhas da árvore e w é o peso de cada folha, e λ é o termo de regularização L_2 sobre pesos.

Uma descrição completa do método pode ser encontrada em (32) e (33).

4 SELEÇÃO DE MODELO USANDO VALIDAÇÃO CRUZADA E ABORDAGEM EVOLUTIVA

Com base em ideias da genética e na teoria da seleção natural de Darwin, a computação evolutiva (CE) é a evolução de programas de computador utilizando analogias com muitos dos métodos utilizados pela evolução biológica natural. Esses programas, definidos como uma série de instruções, são considerados como possíveis respostas para um determinado problema.

A CE serve como uma ferramenta para resolver problemas de engenharia e como um modelo científico simplificado para processos naturais. Esses são os dois principais objetivos da CE. Mesmo que os modelos CE tenham sido usados para resolver muitos problemas de engenharia diferentes, não há muitos que prevejam como a caracterização litológica será determinada.

Os indivíduos da população são programas de computador armazenados na forma de árvores fictícias na computação evolutiva. Esses programas são candidatos para a solução do problema proposto. A recombinação ocorre na computação evolutiva através da troca de subárvores entre dois indivíduos candidatos à solução.

O processo evolutivo pode chegar ao fim de duas maneiras: ou quando o número atribuído de gerações é alcançado, ou quando somos capazes de produzir uma solução com um erro menor que o limite especificado pelo especialista.

Para avaliar o desempenhos dos modelos as métricas mostradas na tabela 4 serão utilizadas.

Tabela 4 – Métricas

Métrica	**Expressão**
R^2	$\sum_{i=1}^{N}(y_i - \hat{y}_i)/\sum_{i=1}^{N}(y_i - \overline{y}_i)$
MAE	$\frac{1}{N}\sum_{t=1}^{N}\lvert(y_i - \hat{y}_i)\rvert$
RSME	$\frac{1}{N}\sqrt{\sum_{i=1}^{N}(y_i - \hat{y}_i)^2}$
MARE	$\frac{\sum_{i=1}^{N}\lvert y_i - \hat{y}_i\rvert}{\sum_{i=1}^{N} y_i}$
WI	$1 - \frac{\sum_{i=1}^{N}(y_i - \hat{y}_i)^2}{\sum_{i=1}^{N}(\lvert\hat{y}_i - \overline{y}_i\rvert + \lvert y_i - \overline{y}_i\rvert)^2}$

Onde R^2 é o coeficiente de determinação, RSME (Root Mean Squared Error) é o erro quadrático médio, MAE (Mean Absolute Error) é o erro médio absoluto, MARE (Mean Absolute Relative Error) é o erro relativo médio absoluto, WI (Index of Agreement) (34) é a razão entre o erro quadrático médio e o erro potencial variando de 0 a 1, y_i representa os dados observados e $\hat{y}_i$ os valores previstos, $\overline{y}$ é a média dos valores observados.

4.1 EVOLUÇÃO DIFERENCIAL

A evolução diferencial é um dos vários algoritmos baseados em populações que podem ser extremamente eficazes na resolução de desafios de otimização. ED é um método de busca meta-heurística baseado em população que otimiza um problema melhorando iterativamente uma solução candidata por meio de um processo evolutivo. Esses algoritmos fazem pouca ou nenhuma suposição sobre o problema de otimização subjacente e são capazes de explorar rapidamente espaços de design muito vastos (35). ED otimiza o problema retendo uma população de soluções candidatas e construindo novas, mesclando as antigas de acordo com a fórmula mais simples, mantendo então quais soluções candidatas têm a melhor avaliação ou adequação para o problema de otimização em mãos.

A primeira publicação sobre ED, conceituada por Storn (36), foi na forma de relatório técnico. O desempenho da ED foi exibido um ano depois no Primeiro Concurso Internacional de Otimização Evolucionária em maio de 1996, que foi realizado em conexão com a Conferência Internacional IEEE de Computação Evolutiva (CEC) (37).

O ED é dividido em duas etapas: inicialização e evolução. A população é criada aleatoriamente na primeira fase, e a população formada passa por processos de mutação, cruzamento e seleção na segunda fase, que se repete até que uma condição de término seja alcançada (38).

Durante a inicialização, a seguinte coleção de populações uniformemente dispersas é gerada: Seja $S^G = \{X_j^G : j = 1, 2, \ldots, \mathrm{NP}\}$ a população em qualquer geração G, NP denota o tamanho da população. X_j^G denota um vetor D-dimensional como $X_j^G = \{x_1^G, x_2^G, \ldots, x_{D,j}^G\}$. X_j^G é gerado usando números aleatórios distribuídos uniformemente entre zero e um (37).

$$X_j^G = X_l + (X_u - X_l) \times rand(0, 1)$$

onde X_l, X_u são os limites inferior e superior do espaço de busca S^G (37).

A segunda fase da ED, a evolução, envolve ações de mutação, cruzamento e seleção. Na mutação geramos um vetor mutante V_j^G para cada vetor alvo X_j^G na geração G como

$$V_j^G = X_{r1}^G + F \times (X_{r2}^G - X_{r3}^G)$$

onde F é o fator de escala e o valor de F varia entre 1 a 0 e $r1, r2, r3 \in \{1, 2, \ldots, NP\}$ são vetores mutuamente diferentes, escolhidos aleatoriamente.

Após a mutação, o cruzamento é feito para gerar um novo vetor chamado vetor de tentativa denotado como $U_j^G = \{u_{1,j}^G, u_{2,j}^G, \ldots, u_{D,j}^G\}$. O cruzamento é realizado entre o vetor de destino $X_j^G = \{x_{1,j}^G, x_{2,j}^G, \ldots, x_{D,j}^G\}$ e o vetor mutante $V_j^G = \{v_{1,j}^G, v_{2,j}^G, \ldots, v_{D,j}^G\}$ usando uma probabilidade de cruzamento Cr cujo valor está entre 0 e 1. U_j^G é gerado como

$$u_{i,j}^G = \begin{cases} v_{i,j}^G, & \text{se } rand_j \leq Cr \\ x_{i,j}^G, & \text{caso contrário} \end{cases}$$

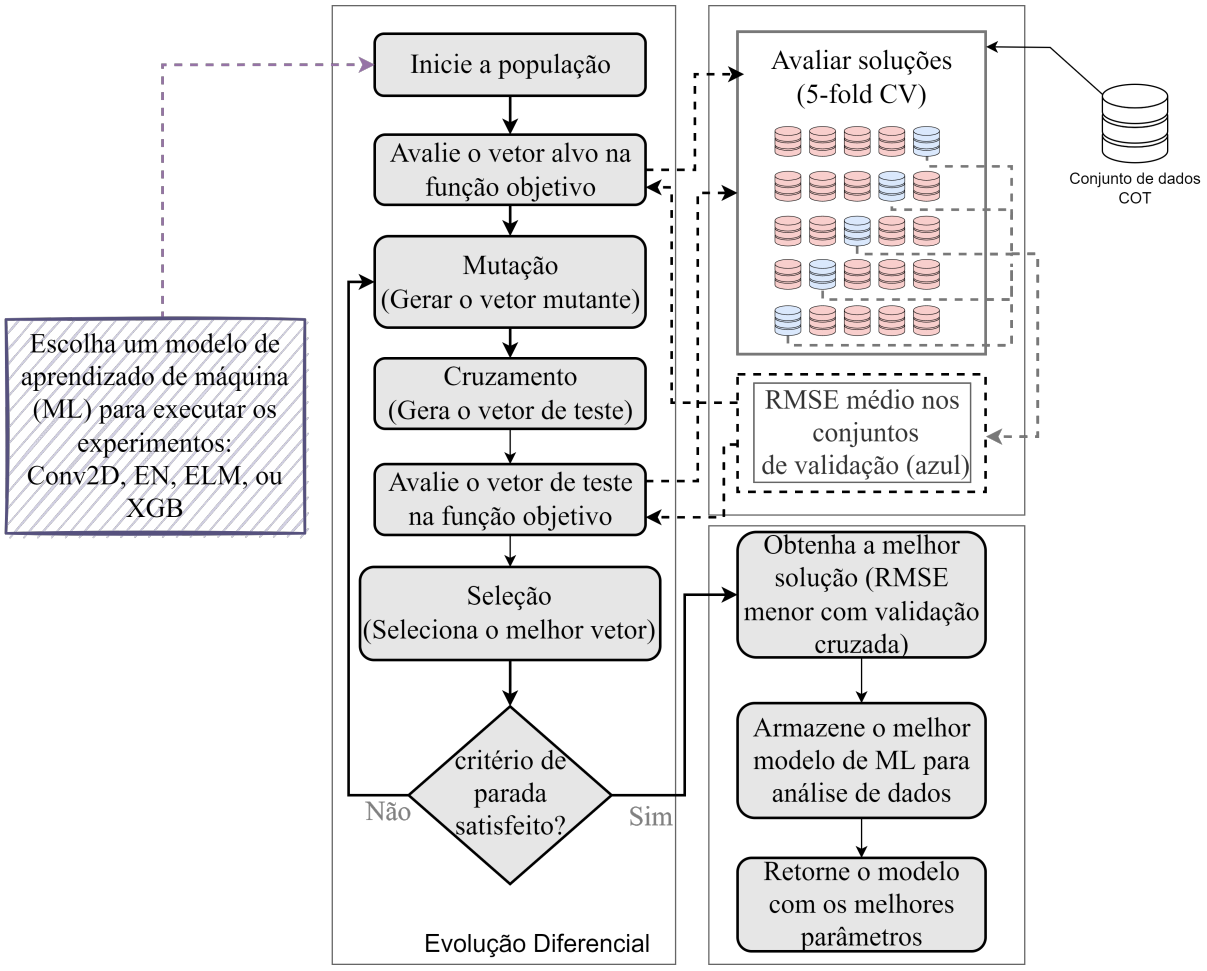
onde $i \in \{1, 2, \ldots, D\}$ and $Cr \in [0, 1]$.

Durante o procedimento de seleção, o vetor alvo e o vetor de teste são comparados com base em seu valor de aptidão. O mais apto sobrevive para a geração seguinte. Este procedimento é realizado da seguinte forma:

$$X_j^{G+1} = \begin{cases} U_j^G, & \text{se } f(U_j^G) \leq f(X_j^G) \\ X_{i,j}^G, & \text{caso contrário} \end{cases}$$

Mutação, cruzamento e seleção da fase de evolução são repetidos até que um critério de terminação pré-definido seja satisfeito. Na figura 6 temos um *flowchart* exemplificando o processo da Evolução Diferencial na escolha dos parâmetros para os modelos de aprendizado de máquinas apresentados anteriormente, juntamente com a validação cruzada utilizada neste trabalho. A tabela 5 trás os parâmetros pesquisados de cada algoritmo de ML e os seus respectivos intervalos.

Figura 6 – Flowchart



Fonte: Produzida pelo autor (2022).

Tabela 5 – Grade de parâmetros.

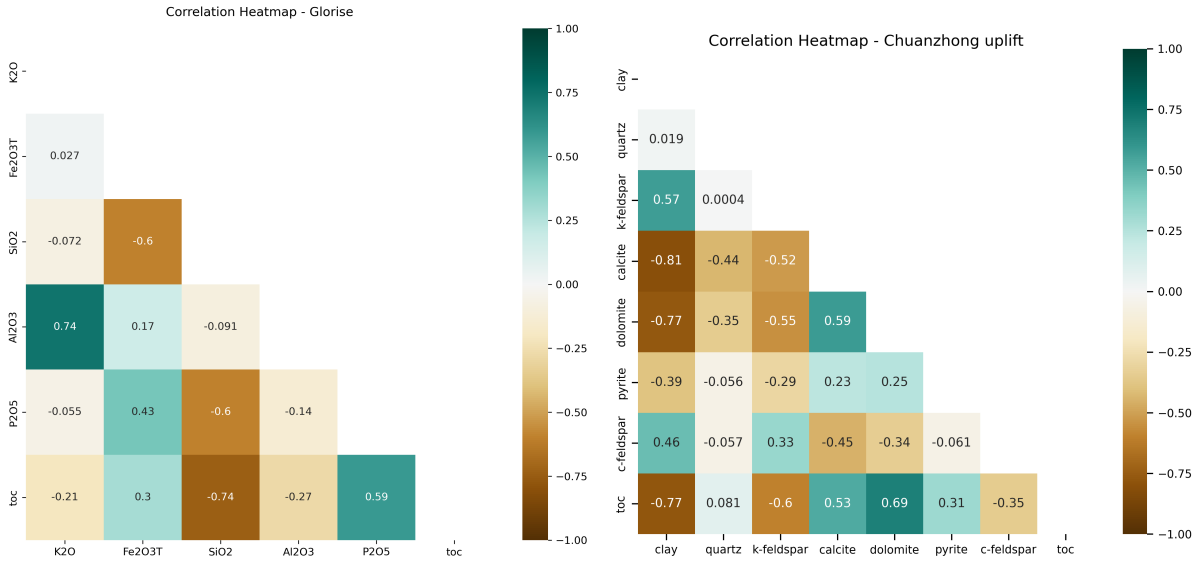
Estimador	Parâmetro	Descrição	Configurações/Intervalo
Conv2D	Filters	Dimensionalidade do espaço de saída	2^n ($n \in \mathbb{N} : 1 \leq n \leq 6$)
	Kernel Size	Dimensão da janela de convolução 2D	[3, 5, 7]
	φ	Função de ativação	linear, relu
	Pool size	Tamanho da janela sobre a qual se deve obter o máximo.	[1, 2, 3]
	Dense units	Número de neurônios.	2^n ($n \in \mathbb{N} : 1 \leq n \leq 8$)
	dropout rate	Taxa de desconsideração de neurônios	De 0 a 0.1
EN	α	Constante que multiplica os termos de penalidade.	1×10^{-6} a 2
	L1 ratio	O parâmetro de mixagem ElasticNet	De 0 a 1
ELM	C_2	Penalidade L2	De 0 a 10^4
	HL	Número de Neurônios	De 1 a 3×10^2
	G	Função de ativação	identity, relu, swish, sigmoid, gaussian, multiquadric, inv multiquadric
	γ	*rbf width*	De 1 a 10
XGB	LR	Taxa de aprendizagem	De 1×10^{-6} a 1
	No. Estimators	N. estimadores fracos	De 10 a 100
	m_{depth}	Profundidade máxima	De 1 - 20
	λ	Regularização L2	De 0 a 100

4.2 PRÉ-PROCESSAMENTO DE DADOS

Para determinar se há correlação entre minerais e COT, foi feito um gráfico de análise de correlação para cada base de dados. Um bom modelo pode ter melhor desempenho se as variáveis de entrada e saída tiverem um vínculo forte.

O coeficiente de correlação obtido a partir dos gráficos COT contra o tipo de mineral são mostrados nas figuras 7a e 7b.

Figura 7 – Matrizes de correlação



(a) Correlação GloRiSe (b) Correlação Chuanzhong uplift

Fonte: Produzida pelo autor (2022).

Para que se obtenha um resultado mais preciso, onde todas as variáveis de entrada da rede tenham o mesmo peso, os dados foram normalizados em uma mesma escala, a escala mais utilizada é a que varia entre os valores zero e um, mesma utilizada neste trabalho.

4.3 TRANSFORMAÇÃO DOS DADOS EM MATRIZES

Ao se utilizar o método de convolução em duas dimensões, precisamos transformar os dados em uma matriz de duas dimensões. Para tal efeito, transforma-se cada linha da entrada de dados em uma matriz que pode ser interpretada como uma imagem em escala de cinza, tal matriz é composta por quatro outras matrizes que serão unidas com o intuito de formar a matriz principal.

As matrizes são montadas a partir de operações realizadas em cada linha onde é feita uma operação relacionando cada variável de entrada com as outras variáveis, sendo assim a posição m_{ij} da matriz m é dada por uma operação $f(c_i, c_j)$, onde c são as variáveis de entrada. As operações utilizadas na criação das matrizes são soma, subtração, multiplicação e exponenciação $\exp(-a_i) \times \exp(-a_j)$. A operação de divisão não foi utilizada para evitar divisões pelo número zero, provenientes da normalização dos dados. Deste modo criamos matrizes de acordo com a quantidade de linhas de entrada dos dados.

A matriz final de cada linha de entrada é composta pelas quatro matrizes iniciais, sendo que a parte superior esquerda é composta pela matriz de soma, a parte superior direita pela matriz de subtração, a parte inferior esquerda pela matriz de multiplicação e a

parte inferior direita pela exponenciação, a figura 8 exemplifica o processo. As figuras 9a e 9b são um exemplo da transformação de dados das duas bases estudadas. Estas matrizes serão utilizadas como entrada na rede de convolução e serão entendidas como imagens pela rede.

Para exemplificar o processo, suponhamos que uma base de dados tenha três variáveis como característica (a, b, c) e n amostras. Cada linha das n entradas passará pelos quatro processos da formação das submatrizes:

- Soma

$$Sum(a_i, b_i, c_i) = \begin{bmatrix} a_i + a_i & a_i + b_i & a_i + c_i \\ b_i + a_i & b_i + b_i & b_i + c_i \\ c_i + a_i & c_i + b_i & c_i + c_i \end{bmatrix}$$

- Subtração

$$Sub(a_i, b_i, c_i) = \begin{bmatrix} a_i - a_i & a_i - b_i & a_i - c_i \\ b_i - a_i & b_i - b_i & b_i - c_i \\ c_i - a_i & c_i - b_i & c_i - c_i \end{bmatrix}$$

- Multiplicação

$$Mult(a_i, b_i, c_i) = \begin{bmatrix} a_i \times a_i & a_i \times b_i & a_i \times c_i \\ b_i \times a_i & b_i \times b_i & b_i \times c_i \\ c_i \times a_i & c_i \times b_i & c_i \times c_i \end{bmatrix}$$

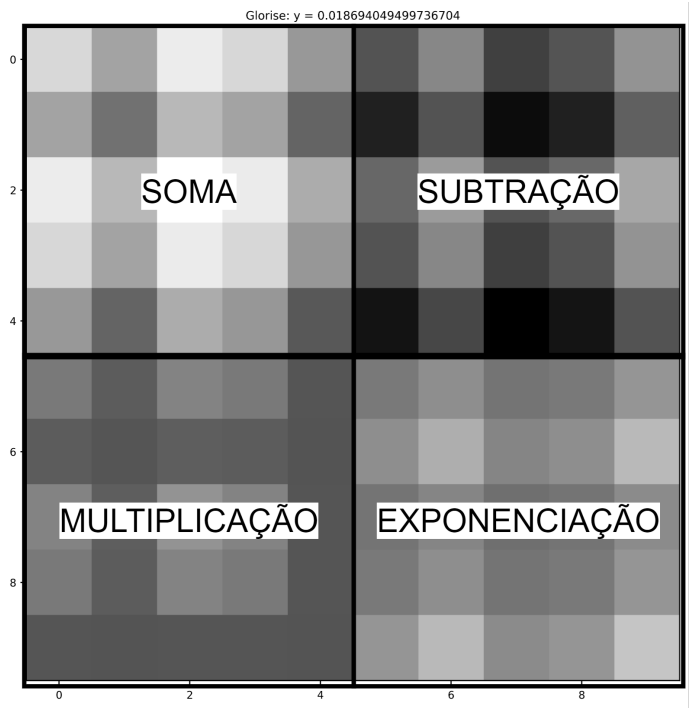
- Exponenciação

$$E(a_i, b_i, c_i) = \begin{bmatrix} e^{-a_i} \times e^{-a_i} & e^{-a_i} \times e^{-b_i} & e^{-a_i} \times e^{-c_i} \\ e^{-b_i} \times e^{-a_i} & e^{-b_i} \times e^{-b_i} & e^{-b_i} \times e^{-c_i} \\ e^{-c_i} \times e^{-a_i} & e^{-c_i} \times e^{-b_i} & e^{-c_i} \times e^{-c_i} \end{bmatrix}$$

- Matriz Final i

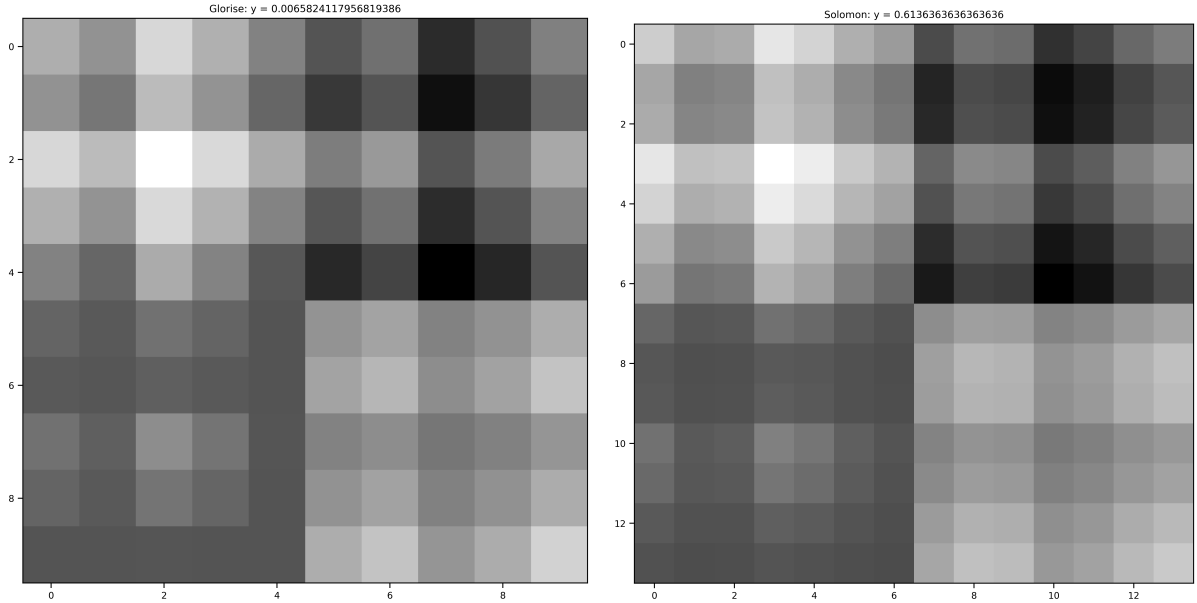
$$M_i(a_i, b_i, c_i) = \begin{bmatrix} a_i + a_i & a_i + b_i & a_i + c_i & a_i - a_i & a_i - b_i & a_i - c_i \\ b_i + a_i & b_i + b_i & b_i + c_i & b_i - a_i & b_i - b_i & b_i - c_i \\ c_i + a_i & c_i + b_i & c_i + c_i & c_i - a_i & c_i - b_i & c_i - c_i \\ a_i \times a_i & a_i \times b_i & a_i \times c_i & e^{-a_i} \times e^{-a_i} & e^{-a_i} \times e^{-b_i} & e^{-a_i} \times e^{-c_i} \\ b_i \times a_i & b_i \times b_i & b_i \times c_i & e^{-b_i} \times e^{-a_i} & e^{-b_i} \times e^{-b_i} & e^{-b_i} \times e^{-c_i} \\ c_i \times a_i & c_i \times b_i & c_i \times c_i & e^{-c_i} \times e^{-a_i} & e^{-c_i} \times e^{-b_i} & e^{-c_i} \times e^{-c_i} \end{bmatrix}$$

Figura 8 – Exemplificação da montagem da matriz - imagem



Fonte: Produzida pelo autor (2022).

Figura 9 – Matrizes geradas



(a) GloRiSe (b) Chuanzhong Uplift

Fonte: Produzida pelo autor (2022).

5 RESULTADOS E DISCUSSÕES

Neste capítulo serão discutidos os resultados obtidos da previsão do COT de cada abordagem apresentada anteriormente e também os resultados obtidos sobre as buscas de parâmetros obtidos a partir da evolução diferencial.

Os experimentos computacionais foram realizados utilizando implementações baseadas nas bibliotecas pygmo, pandas, framework scikit-learn e framework keras. Todos os experimento foram rodados em um computador com as seguintes especificações: Intel(R) Core(TM) i5-6500 (4 núcleos a 3.2 GHz), 16 GB RAM e sistema operacional Windows 10.

5.1 DESEMPENHO DA ABORDAGEM PROPOSTA

As tabelas 6 e 7 comparam as métricas para todas as técnicas de aprendizado de máquina com base em uma média de 100 execuções. Os melhores resultados aparecem em negrito e em parenteses temos o desvio padrão. Estas tabelas mostram que o Conv2D supera os outros algoritmos com relação as métricas R^2 e RMSE nas dua bases de dados estudadas. O processo de evolução diferencial foi utilizado para selecionar os parâmetros internos de cada método. A Tabela 5 exibe os parâmetros e o intervalo que os valores podem assumir.

Como podemos observar na tabela 6 a RNC supera os outros algoritmos com relação as métricas R^2 e RMSE, porém fica atrás do algoritmo XGBoost nas métricas MAE, MARE e WI com relação a base de dados GloRise, mas com uma diferença muito baixa nos resultados, a diferença na métrica WI é de apenas 0.005, MAE de 0.003 e MARE de 0.043. Porém podemos observar na tabela 7 que a RNC supera todos os algoritmos quando utilizamos a base de dados Chuanzhong Uplift.

As figuras 10 e 11 mostram os resultados de previsão para o COT de acordo com os melhores modelos de cada método, encontrados a partir da evolução diferencial. Podemos observar que o método Conv2D e XGBoost produziram os melhores resultados. Apesar da abordagem XGBoost ter produzido um melhor modelo que supera o Conv2D quando olhamos a métrica RMSE, com relação a base GloRiSe, a rede convolucional ainda consegue superar o XGBoost pela sua consistência em produzir modelos com melhores RSME e R^2 em média.

Já com relação à base de dados Chuanzhong Uplift, a RNC supera as outras abordagens em média, e também produz o melhor modelo de previsão de COT.

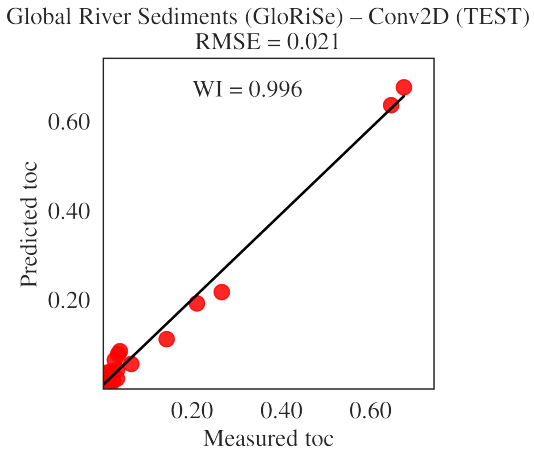
Tabela 6 – Média das métricas para cada abordagem de ML - GloRiSe

Estimator	R^2	WI	RMSE	MAE	MARE
Conv2D	**0.752** (0.272)	0.887 (0.191)	**0.075** (0.057)	0.037 (0.021)	0.454 (0.271)
ELM	0.336 (0.442)	0.668 (0.205)	0.135 (0.057)	0.080 (0.026)	0.899 (0.287)
EN	0.045 (0.561)	0.404 (0.283)	0.163 (0.062)	0.102 (0.030)	1.14 (0.400)
XGB	0.643 (0.494)	**0.892** (0.140)	0.083 (0.059)	**0.034** (0.017)	**0.411** (0.212)

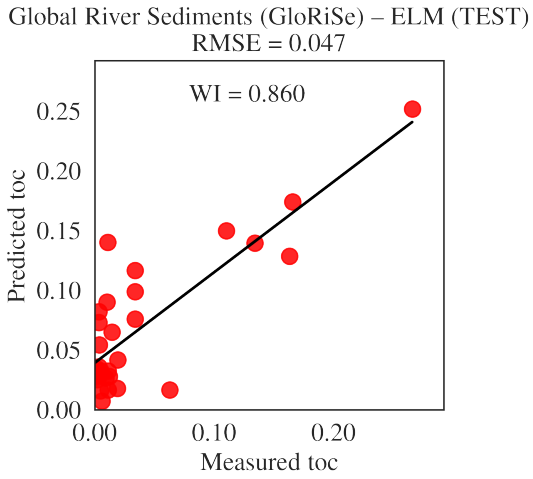
Tabela 7 – Média das métricas para cada abordagem de ML - Chuanzhong Uplift

Estimator	R^2	WI	RMSE	MAE	MARE
Conv2D	**0.794** (0.185)	**0.922** (0.148)	**0.096** (0.040)	**0.075** (0.031)	**0.216** (0.087)
ELM	0.498 (0.193)	0.768 (0.126)	0.159 (0.037)	0.123 (0.026)	0.344 (0.077)
EN	0.281 (0.314)	0.568 (0.278)	0.189 (0.048)	0.149 (0.040)	0.414 (0.109)
XGB	0.593 (0.162)	0.866 (0.054)	0.143 (0.030)	0.108 (0.022)	0.305 (0.067)

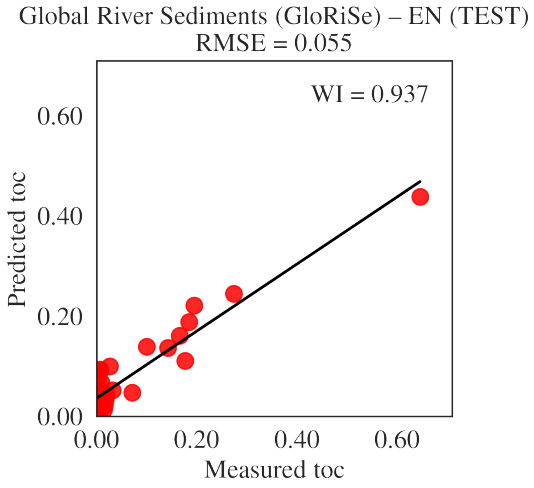
Figura 10 – Resultados dos melhores modelos - GloRiSe



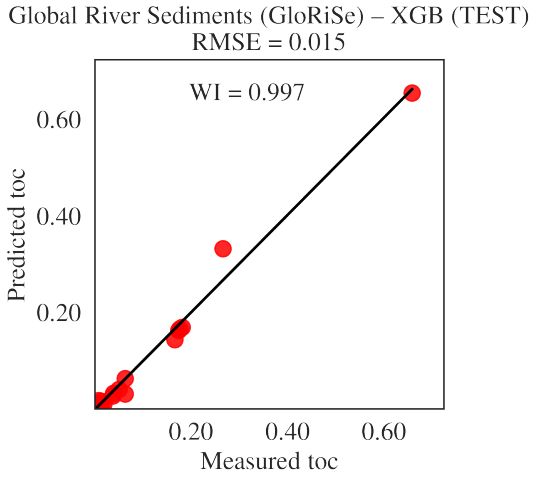
(a) Melhor modelo Conv2D



(b) Melhor modelo ELM



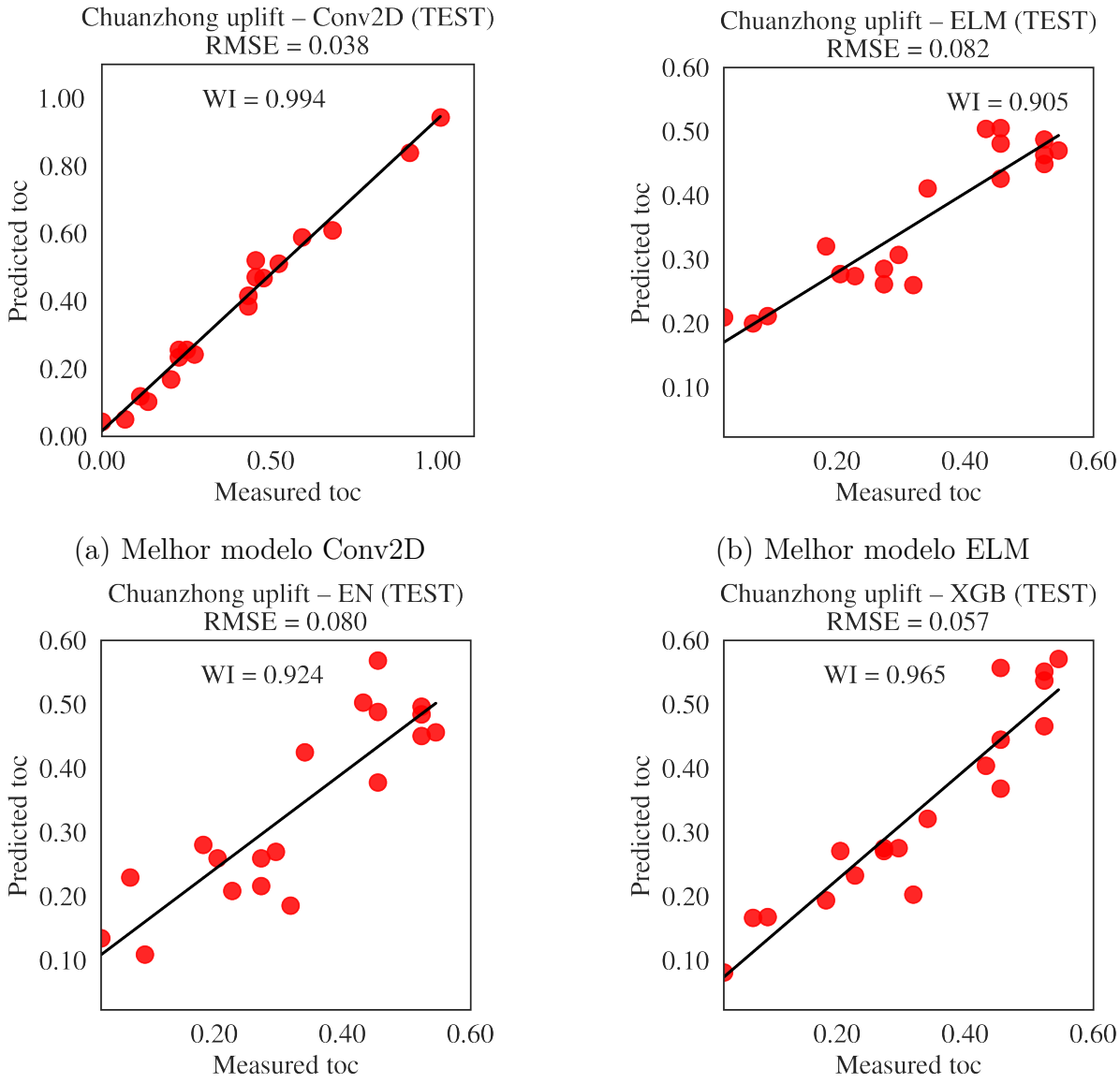
(c) Melhor modelo EN



(d) Melhor modelo XGB

Fonte: Produzida pelo autor (2022).

Figura 11 – Resultados dos melhores modelos - Chuanzhong Uplift



(a) Melhor modelo Conv2D

(b) Melhor modelo ELM

(c) Melhor modelo EN

(d) Melhor modelo XGB

Fonte: Produzida pelo autor (2022).

5.2 ANÁLISE DA DISTRIBUIÇÃO DE PARÂMETROS

Nesta seção serão discutidos os parâmetros encontrados nas construções do modelo RNC para cada base de dados.

Um estudo com a distribuição dos parâmetros determinados pelo processo de evolução diferencial e com validação cruzada foi realizado para avaliar a variação dos parâmetros dos modelos finais encontrados em todas as 100 avaliações. As figuras 12 e 13 descrevem a distribuição dos parâmetros internos para o método de aprendizado de máquina Conv2D em 100 execuções.

A partir da análise das imagens, percebemos que para ambas as bases de dados o

parâmetro *pool size*= 2 com 68 de 100 execuções para a base GloRise (12d) e 84 de 100 execuções para a base Chuanzhong Uplift (13d). Isso significa que a camada de pooling sempre reduzirá o tamanho de cada mapa de ativação por um fator de 2, por exemplo. Cada dimensão é reduzida pela metade, reduzindo o número de pixels ou valores em cada mapa de ativação para um quarto do tamanho.

Para a primeira camada de convolução, podemos observar que a quantidade de 16 filtros prevalece na base de Chuanzhong (12a). Os filtros da primeira camada com relação a base Glorise (12a) possuem uma distribuição mais uniforme na escolha, mas permanecendo o valor 16 o mais escolhido, cerca de 20% das execuções escolheram este valor. Como a entrada dos dados é apenas uma matriz que representa uma imagem em escala de cinza, não se faz necessário o uso de muitos filtros para a convolução. Isso também se repete para a segunda camada de convolução para a base Uplift (13b). Já com a base GloRise (12b) temos três valores que se destacam, $f = 64$, $f = 16$ e $f = 4$.

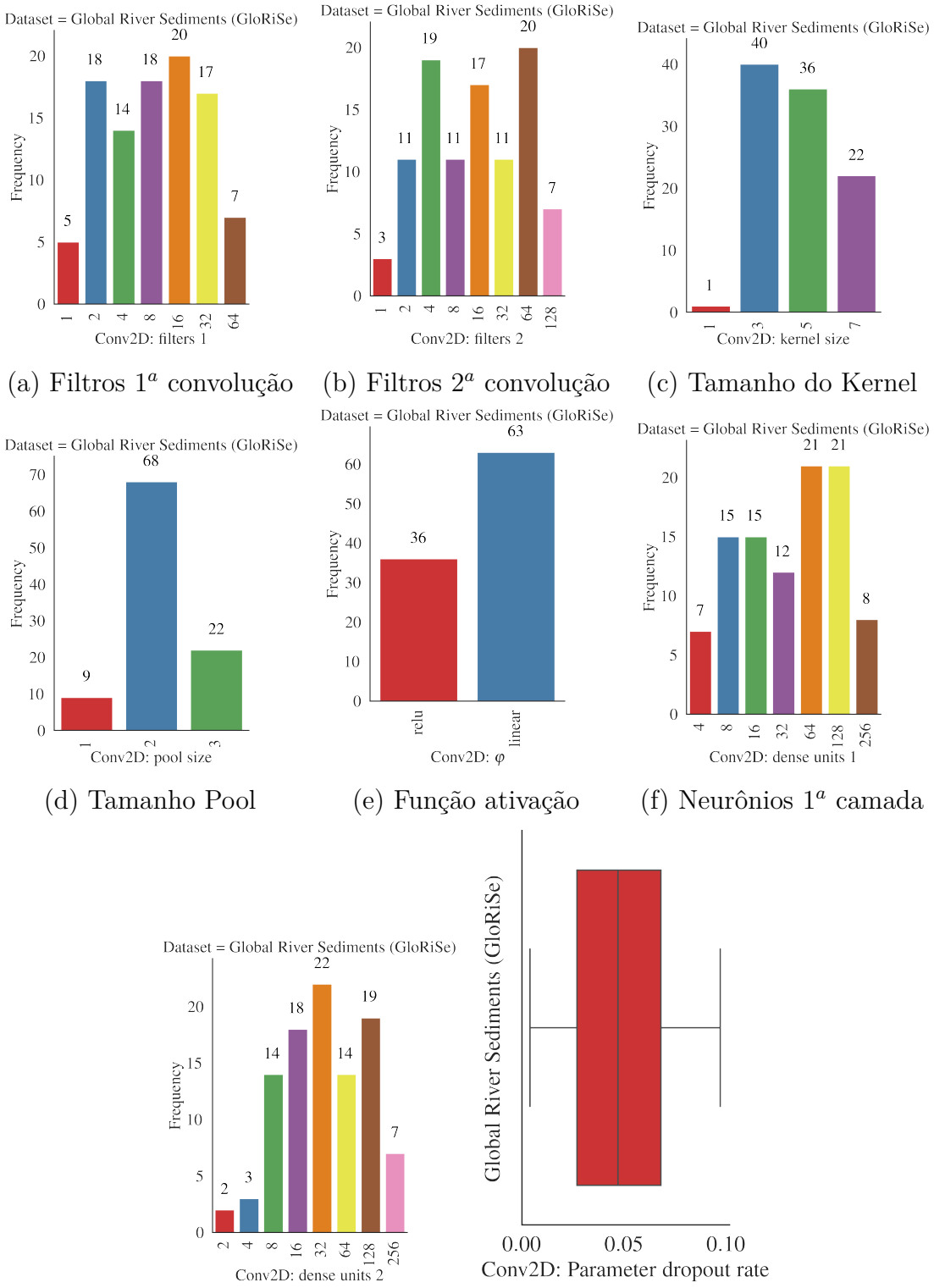
O tamanho do kernel, que gera os mapas de ativação, teve valores $k = 3$ em 53 de 100 execuções para a base Uplift (13c) e em 40 de 100 execuções para a base GloRiSe (12c). Como a dimensão de entrada dos dados é pequena, pelo fato da quantidade de variáveis de ambas as bases de dados serem pequenas, um tamanho pequeno de kernel foi o mais escolhido.

A ativação mais escolhida para as camadas de convolução para a base GloRiSe foi a *linear* com 63 de 100 execuções (12e), para a base Chuanzhong tivemos 51% de escolha para a função de ativação linear e 49% para a ReLu (Rectified Linear activation Function), percebemos então que a função de ativação não teve muita variação em ambos os casos.

Com relação as duas camadas ocultas lineares, obtivemos valores de 64 e 128 para a quantidade de neurônios na primeira camada e 32 para a segunda camada para a base GloRiSe Para a base Uplift obtivemos 128 neurônios na primeira camada e 64 e 128 para a segunda.

Por fim, o nível de distribuição do parâmetro *dropout* para as duas bases de dados ficou com média perto de 0.05%

Figura 12 – Resultados parâmetros GloRiSe



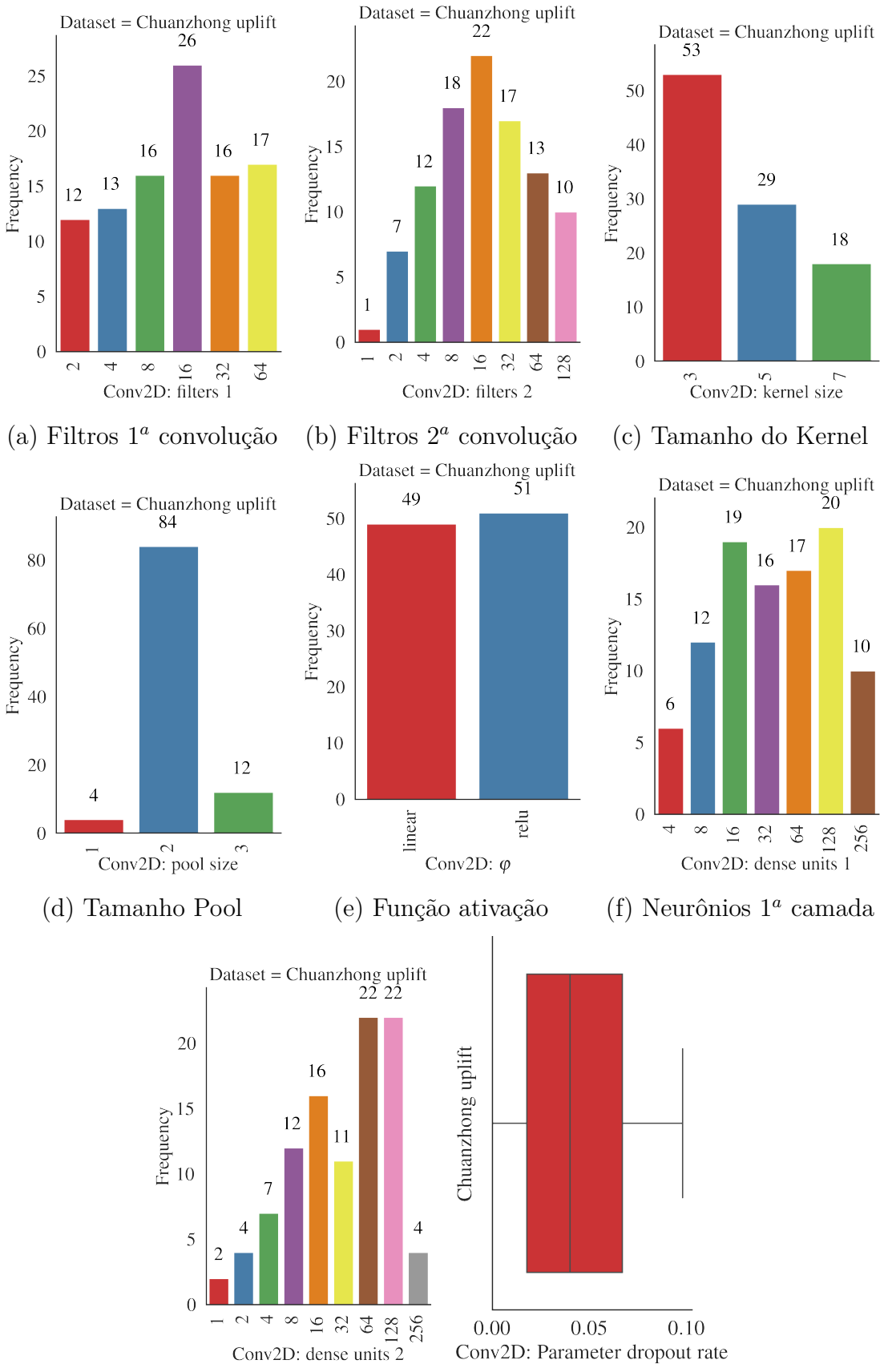
(a) Filtros 1^a convolução (b) Filtros 2^a convolução (c) Tamanho do Kernel

(d) Tamanho Pool (e) Função ativação (f) Neurônios 1^a camada

(g) Neurônios 2^a camada (h) Dropout(%)

Fonte: Produzida pelo autor (2022).

Figura 13 – Resultados parâmetros Chuanzhong Uplift



(a) Filtros 1^a convolução (b) Filtros 2^a convolução (c) Tamanho do Kernel

(d) Tamanho Pool (e) Função ativação (f) Neurônios 1^a camada

(g) Neurônios 2^a camada (h) Dropout(%)

Fonte: Produzida pelo autor (2022).

6 CONCLUSÃO

Este trabalho apresentou a comparação entre quatro métodos de aprendizado de máquina distintos (Conv2D, ELM, EN, XGBoost) usados para previsão do conteúdo de carbono orgânico total em duas bases de dados diferentes, usando uma estratégia computacional de ajuste de hiperparâmetros pela Evolução Diferencial. A estrutura computacional fornecida aqui se mostrou benéfica para selecionar o modelo de aprendizado de máquina e ajustar seus hiper-parâmetros. A rede convolucional desenvolvida neste estudo é o modelo que se destaca, demonstrando uma ótima performance na previsão do COT, seguida do XGBoost.

Dois dos quatro modelos estudados mostraram boa performance considerando o conjunto de métricas utilizados, com destaque ao modelo convolucional, que conseguiu superar os outros modelos ao comparar as métricas R^2 e RMSE em ambas as bases de dados. A abordagem proposta foi suficiente para poder definir um modelo RNC que produz predições com uma boa correlação com os dados obtidos ($R^2_{GloRise} = 0.752$ e $R^2_{Uplift} = 0.794$) e baixos erros associados ($\text{RMSE}_{GloRise} = 0.075$ e $\text{RMSE}_{Uplift} = 0.096$ / $\text{MAE}_{GloRise} = 0.037$ e $\text{MAE}_{Uplift} = 0.075$) em média para as duas bases de dados aplicadas.

Análises dos melhores modelos convolucionais mostraram a qualidade de predição para as diferentes bases de dados, considerando as métricas RMSE ($\text{RMSE}_{GloRise} = 0.021$ e $\text{RMSE}_{Uplift} = 0.038$) e WI ($\text{WI}_{GloRise} = 0.996$ e $\text{WI}_{Uplift} = 0.994$).

Isto demonstra que há uma vantagem significativa na transformação dos dados em matrizes para serem utilizados em uma rede convolucional de duas dimensões (15), sendo interpretadas como imagens pela rede neural convolucional.

Podemos afirmar que o aprendizado de máquina e abordagens evolutivas são ótimas ferramentas de trabalho para poder auxiliar geólogos e outros profissionais da área na previsão do conteúdo de carbono orgânico total. A combinação de modelos robustos de aprendizado de máquina com um algoritmo de otimização eficiente obteve bons resultados. Essa abordagem pode minimizar o trabalho de laboratório, otimizar potencialmente o planejamento experimental e reduzir o tempo de produção de amostras e a carga das atividades associadas.

REFERÊNCIAS

1 FERNANDES, L. *Petróleo e gás natural*. [S.l.]: Departamento Nacional de Produção Mineral, 2009.

2 CEVOLANI, J. et al. Visualização e classificação automática de petrofácies sedimentares. In: . [S.l.: s.n.], 2011.

3 PASSEY S. CREANEY, J. Q. R. A practical model for organic richness from porosity and resistivity logs. *AAPG Bulletin*, American Association of Petroleum Geologists AAPG/Datapages, v. 74, 1990. Disponível em: <https://doi.org/10.1306%2F0c9b25c9-1710-11d7-8645000102c1865d>.

4 NIKRAVESH, M.; AMINZADEH, F. *Soft Computing for Reservoir Characterization and Modeling*. [S.l.]: Springer Berlin, 2004.

5 TAN, M.; LIU, Q.; ZHANG, S. A dynamic adaptive radial basis function approach for total organic carbon content prediction in organic shale. *GEOPHYSICS*, Society of Exploration Geophysicists, v. 78, n. 6, p. D445–D459, nov 2013. Disponível em: <https://doi.org/10.1190%2Fgeo2013-0154.1>.

6 LEE, T. R.; WOOD, W. T.; PHRAMPUS, B. J. A machine learning (kNN) approach to predicting global seafloor total organic carbon. *Global Biogeochemical Cycles*, American Geophysical Union (AGU), v. 33, n. 1, p. 37–46, jan 2019. Disponível em: <https://doi.org/10.1029%2F2018gb005992>.

7 YU, H. et al. A new method for TOC estimation in tight shale gas reservoirs. *International Journal of Coal Geology*, Elsevier BV, v. 179, p. 269–277, jun 2017. Disponível em: <https://doi.org/10.1016%2Fj.coal.2017.06.011>.

8 SHI, X. et al. Application of extreme learning machine and neural networks in total organic carbon content prediction in organic shale with wire line logs. *Journal of Natural Gas ...*, Elsevier, 2016. Disponível em: <https://www.sciencedirect.com/science/article/pii/S1875510016303742>.

9 HANDHAL, A. M. et al. Prediction of total organic carbon at rumaila oil field, southern iraq using conventional well logs and machine learning algorithms. *Marine and Petroleum Geology*, Elsevier BV, v. 116, p. 104347, jun 2020. Disponível em: <https://doi.org/10.1016%2Fj.marpetgeo.2020.104347>.

10 ELKATATNY, S. A self-adaptive artificial neural network technique to predict total organic carbon (TOC) based on well logs. *Arabian Journal for Science and Engineering*, Springer Science and Business Media LLC, v. 44, n. 6, p. 6127–6137, dec 2018. Disponível em: <https://doi.org/10.1007%2Fs13369-018-3672-6>.

11 MAHMOUD, A.; ELKATATNY, S.; AL-ABDULJABBAR, A. Application of machine learning models for real-time prediction of the formation lithology and tops from the drilling parameters. *Journal of Petroleum ...*, Elsevier, 2021. Disponível em: <https://www.sciencedirect.com/science/article/pii/S0920410521002345>.

12 SIDDIG, O.; IBRAHIM, A. F.; ELKATATNY, S. Application of various machine learning techniques in predicting total organic carbon from well logs. *Computational*

Intelligence and Neuroscience, Hindawi Limited, v. 2021, p. 1–9, aug 2021. Disponível em: <https://doi.org/10.1155%2F2021%2F7390055>.

13 WANG, P.; PENG, S.; HE, T. A novel approach to total organic carbon content prediction in shale gas reservoirs with well logs data, tonghua basin, china. *Journal of Natural Gas Science and Engineering*, v. 55, p. 1–15, 2018. ISSN 1875-5100. Disponível em: <https://www.sciencedirect.com/science/article/pii/S1875510018301471>.

14 SAFAEI-FAROUJI, M.; KADKHODAIE, A. Application of ensemble machine learning methods for kerogen type estimation from petrophysical well logs. *Journal of Petroleum Science and ...*, Elsevier, 2022. Disponível em: <https://www.sciencedirect.com/science/article/pii/S0920410521010986>.

15 ASANTE-OKYERE, S.; ZIGGAH, Y. Y.; MARFO, S. A. Improved total organic carbon convolutional neural network model based on mineralogy and geophysical well log data. *Unconventional Resources*, v. 1, p. 1–8, 2021. ISSN 2666-5190. Disponível em: <https://www.sciencedirect.com/science/article/pii/S2666519021000017>.

16 MüLLER, G.; , Middelburg; , Sluijs. *Global River Sediments (GloRiSe)*. Zenodo, 2021. Disponível em: <https://zenodo.org/record/4447435>.

17 LECUN, Y.; BENGIO, Y.; HINTON, G. Deep learning. *Nature*, Springer Science and Business Media LLC, v. 521, n. 7553, p. 436–444, may 2015. Disponível em: <https://doi.org/10.1038%2Fnature14539>.

18 KUO, C.-C. J. Understanding convolutional neural networks with a mathematical model. *Journal of Visual Communication and Image Representation*, Elsevier BV, v. 41, p. 406–413, nov 2016. Disponível em: <https://doi.org/10.1016%2Fj.jvcir.2016.11.003>.

19 AJIT, A.; ACHARYA, K.; SAMANTA, A. A review of convolutional neural networks. In: *2020 International Conference on Emerging Trends in Information Technology and Engineering (ic-ETITE)*. [S.l.: s.n.], 2020. p. 1–5.

20 ZHOU, D.-X. Theory of deep convolutional neural networks: Downsampling. *Neural Networks*, v. 124, p. 319–327, 2020. ISSN 0893-6080. Disponível em: <https://www.sciencedirect.com/science/article/pii/S0893608020300204>.

21 MISHRA, M. *Convolutional Neural Networks, explained*. Towards Data Science, Sep 2020. Disponível em: <https://towardsdatascience.com/convolutional-neural-networks-explained-9cc5188c4939#:~:text=A%20Convolutional%20Neural%20Network%2C%20also,binary%20representation%20of%20visual%20data.>

22 HUANG, G.-B.; ZHU, Q.-Y.; SIEW, C.-K. Extreme learning machine: Theory and applications. *Neurocomputing*, v. 70, n. 1, p. 489–501, 2006. ISSN 0925-2312. Neural Networks. Disponível em: <https://www.sciencedirect.com/science/article/pii/S0925231206000385>.

23 SORIA-OLIVAS, E. et al. Belm: Bayesian extreme learning machine. *IEEE Transactions on Neural Networks*, v. 22, n. 3, p. 505–509, 2011.

24 HUANG, G.-B. et al. Extreme learning machine for regression and multiclass classification. *IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics)*, v. 42, n. 2, p. 513–529, 2012.

25 ZOU, H.; HASTIE, T. Regularization and variable selection via the elastic net. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, Wiley, v. 67, n. 2, p. 301–320, apr 2005. Disponível em: <https://doi.org/10.1111%2Fj.1467-9868.2005.00503.x>.

26 HOERL, A. E.; KENNARD, R. W. Ridge regression: Biased estimation for nonorthogonal problems. *Technometrics*, Informa UK Limited, v. 12, n. 1, p. 55–67, feb 1970. Disponível em: <https://doi.org/10.1080%2F00401706.1970.10488634>.

27 MOL, C. D.; VITO, E. D.; ROSASCO, L. Elastic-net regularization in learning theory. *Journal of Complexity*, Elsevier BV, v. 25, n. 2, p. 201–230, apr 2009. Disponível em: <https://doi.org/10.1016%2Fj.jco.2009.01.002>.

28 ARAVEEPORN, A. The higher-order of adaptive lasso and elastic net methods for classification on high dimensional data. *Mathematics*, MDPI AG, v. 9, n. 10, p. 1091, may 2021. Disponível em: <https://doi.org/10.3390%2Fmath9101091>.

29 ZOU, H.; HASTIE, T. Regression shrinkage and selection via the elastic net, with applications to microarrays. *JR Stat Soc Ser B*, researchgate.net, 2003. Disponível em: <https://www.researchgate.net/profile/Trevor-Hastie/publication/228781252_Regression_shrinkage_and_selection_via_the_elastic_net_with_applications_to_microarrays/links/0c960521b946ea7e32000000/Regression-shrinkage-and-selection-via-the-elastic-net-with-applications-to-microarrays.pdf>.

30 TIBSHIRANI, R. Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society: Series B (Methodological)*, Wiley, v. 58, n. 1, p. 267–288, jan 1996. Disponível em: <https://doi.org/10.1111%2Fj.2517-6161.1996.tb02080.x>.

31 BASILIO, S. A.; GOLIATT, L. Gradient boosting hybridized with exponential natural evolution strategies for estimating the strength of geopolymer self-compacting concrete. *Kbes*, Knowledge-based Engineering and Sciences, v. 3, n. 1, p. 1–16, abr. 2022.

32 CHEN, T.; GUESTRIN, C. XGBoost. In: *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. ACM, 2016. Disponível em: <https://doi.org/10.1145%2F2939672.2939785>.

33 BENTéJAC, C.; CSöRGő, A.; MARTíNEZ-MUñOZ, G. A comparative analysis of xgboost. 11 2019.

34 WILLMOTT, C. J. On the validation of models. *Physical Geography*, Taylor Francis, v. 2, n. 2, p. 184–194, 1981. Disponível em: <https://doi.org/10.1080/02723646.1981.10642213>.

35 GEORGIOUDAKIS, M.; PLEVRIS, V. A comparative study of differential evolution variants in constrained structural optimization. *Frontiers in Built Environment*, Frontiers Media SA, v. 6, jul 2020. Disponível em: <https://doi.org/10.3389%2Ffbuil.2020.00102>.

36 STORN, R. On the usage of differential evolution for function optimization. In: *Proceedings of North American Fuzzy Information Processing*. [S.l.: s.n.], 1996. p. 519–523.

37 STORN, R.; PRICE, K. *Journal of Global Optimization*, Springer Science and Business Media LLC, v. 11, n. 4, p. 341–359, 1997. Disponível em: <https://doi.org/10.1023%2Fa%3A1008202821328>.

38 BILAL et al. Differential evolution: A review of more than two decades of research. *Engineering Applications of Artificial Intelligence*, Elsevier BV, v. 90, p. 103479, apr 2020. Disponível em: <https://doi.org/10.1016%2Fj.engappai.2020.103479>.