NAS-Driven Quantitative Assessment of Overpressure Genesis in Sedimentary Basins

Haoyue Chen[1], Chaowei Li[1,*], Yang Xu[2], Song Deng[1], Huawei Zhang[3], Zhiqiang Hu[3] and Haoyu Pan[1]

[1] School of Petroleum and Natural Gas Engineering, Changzhou University, Changzhou, 213164, China

[2] Engineering Department, Sinopec International Petroleum Exploration and Production Corporation, Beijing, 100029, China

[3] Petroleum Engineering Technical Support Center, Sinopec Research Institute of Petroleum Engineering, Beijing, 102206, China

INFORMATION

Keywords:
Overpressure genesis mechanism
clustering algorithm
neural architecture search
quantitative assessment

DOI: 10.23967/j.rimni.2025.10.71602

Revista Internacional de Métodos Numéricos para Cálculo y Diseño en Ingeniería

NAS-Driven Quantitative Assessment of Overpressure Genesis in Sedimentary Basins

Haoyue Chen[1], Chaowei Li[1,*], Yang Xu[2], Song Deng[1], Huawei Zhang[3], Zhiqiang Hu[3] and Haoyu Pan[1]

[1]School of Petroleum and Natural Gas Engineering, Changzhou University, Changzhou, 213164, China

[2]Engineering Department, Sinopec International Petroleum Exploration and Production Corporation, Beijing, 100029, China

[3]Petroleum Engineering Technical Support Center, Sinopec Research Institute of Petroleum Engineering, Beijing, 102206, China

ABSTRACT

The study of overpressure genesis mechanism is the foundation of hydrocarbon reservoir formation and pressure prediction research, and a thorough understanding of the formation and distribution patterns of hydrocarbon resources is essential for practical hydrocarbon exploration. However, the prediction of anomalous high-pressure genesis currently encounters numerous challenges, including the complexity of overpressure genesis, the superimposed effect of multiple mechanisms, and the dependence of traditional models on manual analysis, resulting in inefficiency and quantification challenges. To this end, this paper proposes a method for identifying and quantitatively evaluating stratigraphic overpressure mechanisms based on neural architecture search, to enable rapid and accurate quantitative evaluation of overpressure mechanisms. The results show that the main anomalous high-pressure genesis mechanisms in the target work area include undercompaction, fluid expansion and tectonic compression, with contribution rates of approximately 73% for undercompaction, and 9% and 18% for fluid expansion and tectonic compression, respectively. The model's accuracy in the test set reaches 95.4%, significantly enhancing the identification accuracy of anomalous stratigraphic pressure genesis and its superposition relationships. The innovation of this paper lies in the combination of wave velocity-density rendezvous map with clustering algorithm and neural architecture search algorithm, offering an efficient approach to identify multiple overpressure genesis mechanisms and predict pore pressure through machine learning algorithms, which is of great theoretical significance and practical application value.

OPEN ACCESS

Received: 08/08/2025

Accepted: 14/10/2025

Published: 03/02/2026

DOI
10.23967/j.rimni.2025.10.71602

Keywords:
Overpressure genesis mechanism
clustering algorithm
neural architecture search
quantitative assessment

1 Introduction

Abnormally high pressure has long been a critical topic in oil and gas geology and exploration research. Accurate identification of the mechanisms underlying overpressure genesis is crucial for understanding overpressure and hydrocarbon formation in basins. However, the identification of the causes of overpressure and the accurate prediction of pressure remain challenging, primarily due to the

*Correspondence: Chaowei Li (chw25@cczu.edu.cn).

complex mechanisms underlying overpressure and the superimposition of multiple causative factors [1]. Different mechanisms of overpressure genesis lead to varying pressure prediction methods and distinct pressure evolution processes, which in turn result in different characteristics and distribution patterns of both overpressure and hydrocarbon accumulation. Therefore, identifying the specific overpressure mechanisms and assessing their relative contributions is a critical issue in the study of overpressure [2].

Significant progress has been made in empirical studies of overpressure genesis over the past two decades. Early-stage overpressure identification relied heavily on sonic logging data. Among these methods, combined log curve analysis has been widely recognized as a fundamental and reliable approach for overpressure genesis identification. Bowers [3] introduced a porosity-vertical effective stress relationship method (commonly referred to as the loading-unloading curve) for distinguishing different overpressure genesis types. This method has been extensively applied, particularly in the 21st century [4,5], demonstrating excellent performance. Hermanrud et al. [6] developed the porosity comparison method, identifying porosity anomalies in overpressure sections by comparing them with normally pressured sections at equivalent depths. Additionally, Bowers [7] established the acoustic-density crossplot method, effectively distinguishing three genesis mechanisms (undercompaction, fluid expansion, and tectonic extrusion) through data point positions relative to the ideal loading curve [8,9]. Some scholars [10,11] demonstrated that on density-velocity crossplots, both normally pressured and overpressured zones from undercompaction align with the loading curve, whereas other overpressure mechanisms deviate from this trend. While these direct and indirect methods have proven valuable for overpressure genesis identification, they remain limited to qualitative assessments rather than quantitative analyses.

Recent advances have enabled quantitative evaluation of overpressure generation mechanisms. The equivalent depth method [12,13] provides a reliable approach to calculate the contribution of undercompaction to overpressure development. Multi-basin simulation software has emerged as a powerful tool for quantitatively assessing overpressure mechanisms through geological process modeling [14–17]. The DCFM [18] offers another quantitative framework, where pre- and post-tectonic compression behaviors are analyzed to differentiate between undercompaction-induced overpressure and unloading overpressure. This approach facilitates quantitative analysis of how multiple overpressure sources influence hydrocarbon migration and seal integrity.

Moreover, Xu et al. [19] proposed a weight quantification model for overpressure mechanisms based on the LightGBM algorithm, further optimized via Bayesian optimization, offering a novel approach to machine learning applications in complex geological phenomena within analogous environments. However, as our preliminary research, this study exhibits limitations: the LightGBM-based hybrid model is inherently constrained by decision trees' weak capacity for characterizing high-order nonlinear interactions, while the fixed Bayesian optimization architecture is prone to local optima in parameter space when processing multimodal data such as well logs and seismic attributes. To address these issues, we developed a neural architecture search-based quantitative evaluation model for stratigraphic overpressure mechanisms. Our methodology involves: (1) cluster analysis of anomalous pressure samples using hierarchical clustering algorithms and acoustic velocity-density crossplots; (2) training samples with similar pressure mechanisms through neural architecture search; and (3) predicting overpressure mechanisms in the study area to quantify individual contributions to pressure evolution. The model's application in the Cameroon basin demonstrates its effectiveness in elucidating anomalous pressure genesis and multi-mechanism superposition relationships.

2 Algorithm Preference

2.1 Clustering Algorithm

Current research on clustering algorithms has achieved significant progress, with four primary categories emerging based on analytical approaches: K-means clustering, hierarchical clustering, DBSCAN, and model-based clustering (represented by GMM) [20,21]. Among these, hierarchical clustering, DBSCAN, and GMM demonstrate particular suitability for scenarios requiring a priori determination of cluster numbers. In contrast, partition-based algorithms like K-means prove less appropriate due to their mandatory cluster number specification requirement. For algorithm evaluation, hierarchical clustering, DBSCAN, and GMM clustering were implemented on stratigraphic velocity-density datasets. Quantitative assessment included inter-cluster and intra-cluster distance calculations, followed by performance evaluation of clustered anomalous pressure mechanism samples.

In terms of computational efficiency, both hierarchical clustering and K-means offer relatively fast processing speeds, with hierarchical clustering demonstrating favorable scalability for medium-scale formation data. In contrast, GMM and DBSCAN require longer training times due to model complexity and high-dimensional computational overhead. Regarding noise robustness, while DBSCAN inherently supports noise identification, it is highly sensitive to parameter settings; hierarchical clustering effectively controls noise through pruning and distance constraints, whereas GMM and K-means remain vulnerable to noise and outliers. In geological interpretability, hierarchical clustering excels: its dendrogram structure visually reflects multi-scale characteristics and genetic relationships within formation parameters, supporting the interpretation of geological processes such as sedimentary sequences and pressure evolution. The clustering results align closely with established geological classification systems like log facies and lithologic associations, providing a verifiable and traceable framework for geological analysis—overall outperforming DBSCAN, GMM, and K-means.

Effective clustering results should accurately capture the inherent structural characteristics of the original data. An optimal clustering algorithm should simultaneously minimize inter-class similarity while maximizing intra-class similarity, while also revealing implicit patterns and relationships within the dataset. Table 1 demonstrates that DBSCAN clustering exhibits relatively large intra-cluster distances, suggesting inadequate learning of the dataset's true structure. This results in poor cluster compactness and insufficient accuracy in classifying abnormal pressure causal mechanisms. Although GMM clustering shows comparable intra-cluster distances, its inter-cluster distances are smaller than hierarchical clustering due to the complexity of abnormal pressure mechanisms. Consequently, both methods underperform compared to hierarchical clustering, which achieves the smallest intra-cluster distance (5.19) among the three methods, indicating superior cluster compactness and regularity, along with the largest inter-cluster distance that facilitates further investigation.

Table 1: Cluster assessment of a sample of similar anomalous stress-causing mechanisms

Clustering methods	Number of categories	Intercluster distance	Intracluster distance
DBSCAN	3	2.49	47.77
GMM	7	9.37	12.38
Hierarchical clustering	9	19.16	5.19

In addition, the classification results of different clustering algorithms show obvious differences in the figure, as shown in Fig. 1. In Fig. 1a, DBSCAN clustering fails to effectively differentiate noise points from valid clusters, with cluster sizes deviating from the actual data distribution. Some clusters

become excessively large, incorporating data points with diverse characteristics. Fig. 1b reveals that the data does not conform to a mixture of Gaussian distributions, instead displaying highly non-convex, irregular shapes with complex multidimensional correlations that cannot be adequately characterized by Gaussian probability density functions. By contrast, Fig. 1c demonstrates hierarchical clustering's clear representation of cluster hierarchies with well-defined separation boundaries between clusters. The classification results align with the distribution patterns observed in wave velocity-density crossplots of abnormal pressure samples, offering high interpretability where cluster meanings can be directly inferred from the classification diagrams. Based on these findings, hierarchical clustering was selected for classifying abnormal pressure mechanisms.

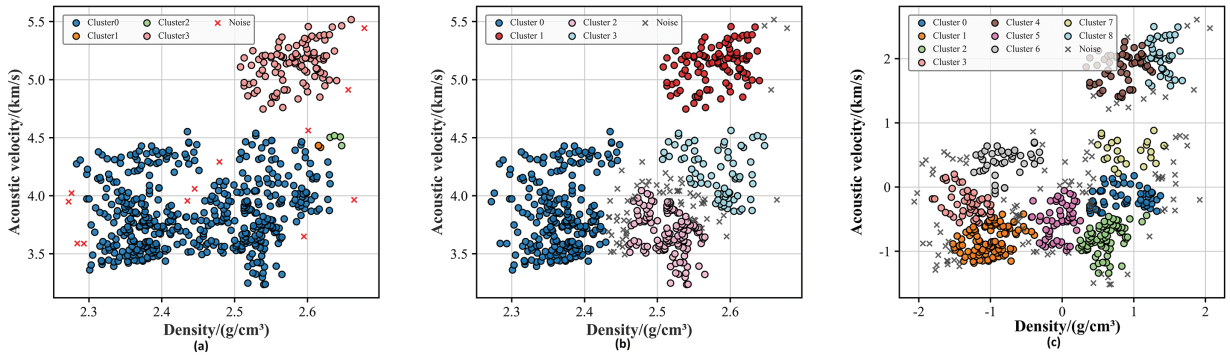


Figure 1: Visualization of classification results of different clustering algorithms. (a) DBSCAN clustering; (b) GMM clustering; (c) Hierarchical clustering

2.2 Supervised Learning Algorithm

Conventional methods for stratigraphic overpressure discrimination typically rely on empirically—based rules and linear models. It is difficult for them to capture the complex nonlinear relationships and high-dimensional data features in the formation mechanism of stratigraphic overpressure. With the continuous development of machine learning technology, data-driven prediction methods have gradually become important tools in fields such as geological exploration and oil and gas field development, data-driven models can automatically learn potential patterns and improve prediction accuracy [22]. When selecting machine learning algorithms, it is necessary to comprehensively consider the characteristics of the data and the specific requirements of the task. Therefore, the selection of intelligent algorithms needs to be comprehensively compared and weighed in the following aspects [23–25]:

1. Data Characteristics and Task Requirements: Traditional linear models and empirically-based rules struggle to capture complex geological feature associations, such as stratigraphic velocity anomalies and the nonlinear coupling mechanisms between rock physical property changes and overpressure. In essence, stratigraphic overpressure discrimination is a classification problem that requires high-precision prediction of the overpressure state, which poses higher requirements for the algorithm's nonlinear modeling ability and classification performance. Among classification algorithms, traditional classification algorithms intuitively express the logical relationship between stratigraphic velocity anomalies and pressure coefficients through feature splitting rules. Ensemble learning methods further optimize prediction accuracy through multi-model collaboration, and are suitable for handling the heterogeneity in geological data caused by differences in genetic mechanisms. Classification algorithms based on deep learning can automatically extract deep-seated features in high-dimensional data, through

multi-layer nonlinear transformations. Innovative architectures derived from NAS solve the problem of vague identification of the boundaries of pressure-sealing boxes by traditional methods through automatic optimization of network depth and connections. Considering classification ability, computational efficiency, and geological adaptability, this paper selects traditional classification algorithms, ensemble learning algorithms, deep learning algorithms, and classifiers derived from NAS for comparative analysis to explore the applicability of different algorithms in stratigraphic overpressure discrimination.

2. Robustness and Scalability of Algorithms: Datasets in geological exploration usually have the characteristics of large scale and high dimensionality. Therefore, the computational efficiency and scalability of algorithms are of great importance. When comparing a variety of common algorithms, the advantages and disadvantages of the following algorithms are mainly considered in Table 2:

Table 2: Supervised learning algorithm preferences

Algorithm type	Representative models	Applicability
Traditional	DT, RF	Basic: Small-scale, noise-sensitive
Ensemble	XGBoost, LightGBM	Improved: Stable but tuning-heavy
Deep Learning	DNN, CNN	Advanced: Powerful but data-hungry
Innovative	NAS	Cutting-edge: Auto-optimized

In conclusion, for the numerical classification problem of the given abnormal pressure genesis mechanism, the NAS algorithm is selected as the optimal solution. Its technical advantages are reflected in structural adaptability, multi-modal fusion capability, and transfer learning potential. Through a two-stage network architecture search, the NAS model can automatically eliminate redundant connections, significantly improving the discrimination efficiency of the stratigraphic overpressure mechanism. This selection fully takes into account the balance among algorithm robustness, computational efficiency, and interpretability under complex geological conditions, providing a new technical paradigm for the intelligent discrimination and quantitative evaluation of the overpressure mechanism.

3 Methodologies

To investigate stratigraphic anomalous high pressure resulting from multiple superimposed genesis mechanisms, a quantitative evaluation model was developed based on NAS. This model enables systematic analysis of individual overpressure mechanism contributions during regional pressure evolution. The complete model architecture is illustrated in Fig. 2.

3.1 Data Preprocessing

The identification of formation overpressure mechanisms depends on several logging parameters. well depth, wave velocity, density, natural gamma, resistivity, and acoustic time—difference are selected as key variables. These features are chosen based on their significance in formation characterization and pressure state determination. For instance, well depth reflects the variations in the subsurface geological structure and can offer spatial distribution information for identifying pressure mechanisms. Natural gamma, resistivity, and acoustic time-difference provide crucial information regarding fluid types, formation characteristics, and the elastic properties of rocks. Among these,

acoustic wave velocity and density are the most critical logging parameters. Wave velocity reflects the elastic properties of formations, while density serves as an indicator of the mass distribution and degree of compaction of formations [26].

To enhance the robustness and generalization ability of the model, the following steps are taken. First, the original data are randomly partitioned into training and test sets at a ratio of 7:3. This ensures that the training set is sufficiently rich and the test set has a reliable evaluation capacity. Second, to tackle the problem of class imbalance, the SMOTE algorithm is employed to oversample the minority class samples. Finally, the Max-Min Scaling method is used to standardize all input features. The feature values are scaled to the [0, 1] interval, eliminating the interference of magnitude differences on model training. After standardization, the value range of each index is consistent, avoiding training errors caused by scale differences and thus enhancing the stability and accuracy of the model during the training process [27]. By extracting, calculating, and pre-processing the data within the effective well depth range, the data quality is ensured to meet the requirements of subsequent analysis.

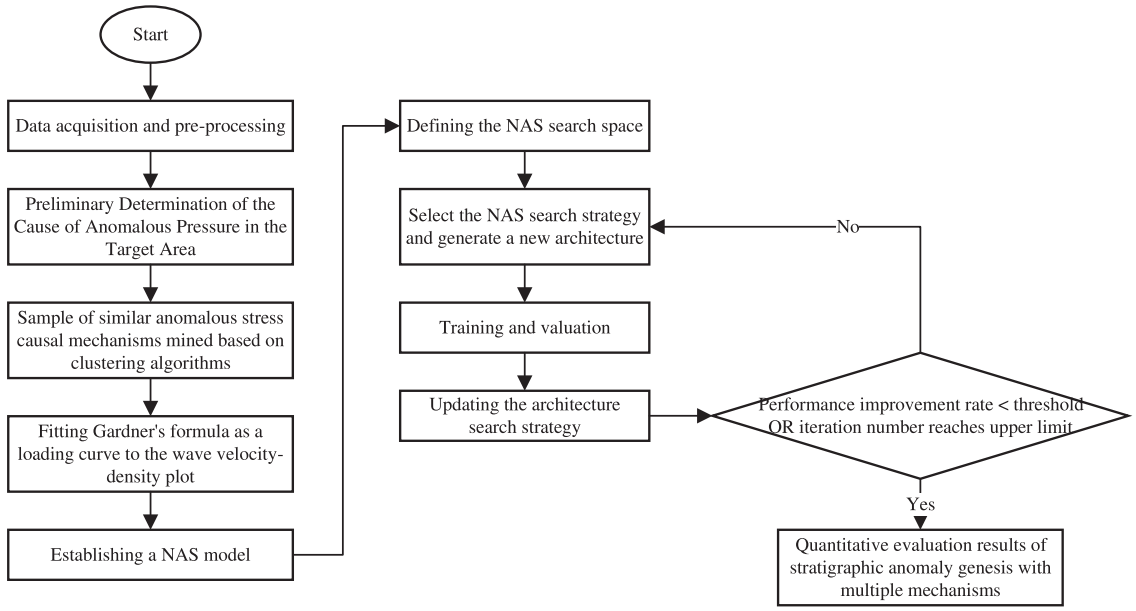


Figure 2: Overall flowchart of quantitative evaluation model of overpressure mechanism

3.2 Cluster Analysis

The application of hierarchical clustering algorithms in unsupervised learning enables automated classification of overpressure sample populations into meaningful clusters based on intrinsic similarity, without requiring predefined class identifiers. This data-driven approach effectively extracts latent yet valuable information from incomplete, noisy, and heterogeneous logging data, overcoming the limitations of conventional overpressure mechanism identification methods that heavily depend on empirical judgments. Particularly in complex formations with multiple overpressure origins, traditional methods often fail to discern mechanistic interactions, whereas clustering algorithms demonstrate superior accuracy and reliability in differential diagnosis.

Building upon prior studies utilizing acoustic velocity, density, resistivity, and natural gamma as logging parameters, we specifically selected acoustic velocity and density as clustering algorithm inputs due to their diagnostic sensitivity to pressure mechanisms. As illustrated in Fig. 3, the hierarchical

clustering model successfully categorized samples into nine distinct groups, each representing a characteristic anomalous pressure genesis pattern.

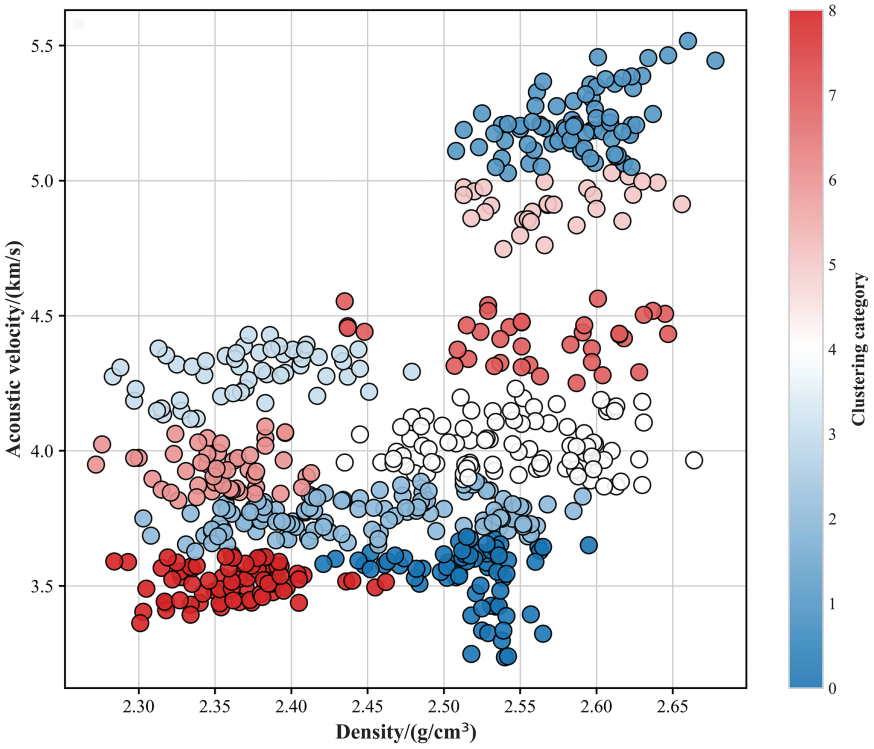


Figure 3: Clustering results for anomalous pressure samples

The hierarchical clustering algorithm can effectively identify sample clusters with similarities. However, during the analysis process, it is essential to integrate geological background knowledge and traditional research methods of overpressure mechanisms. Additionally, careful division of the clustering results is required to better interpret the model output. Thus, to further elucidate the clustering results of different anomalous pressure genesis mechanisms, in this study, based on Gardner's formula [28] which describes the wave velocity-density relationship, we fitted the data of the normal compaction section. We then established the normal trend line of the standard wave velocity-density cross-plot and applied it to the wave velocity-density cross-plot of the anomalously high-pressure section to supplement the additional loading curves.

Gardner's formula reveals the relationship between formation density and longitudinal wave velocity. It is an average transformed equation for the velocity-density relationship across different lithologies. This transformed equation provides the best fit for all lithologies. The only variables involved in the common Gardner formula are sonic time difference or sonic velocity. The equation is as follows (Eq. (1)):

$$\rho = A\left(\frac{1{,}000{,}000}{\Delta t}\right)^{B} \tag{1}$$

where ρ is the rock density, Δt is the acoustic time difference, A and B are constants.

However, the empirical formula may introduce errors when applied directly. To address this limitation, a more accurate Gardner correction formula was derived by fitting data from the normal compaction section, based on the conventional Gardner formula. The corrected Gardner normal trend line is shown in Fig. 4, which illustrates the typical relationship between wave velocity and density under normal compaction conditions.

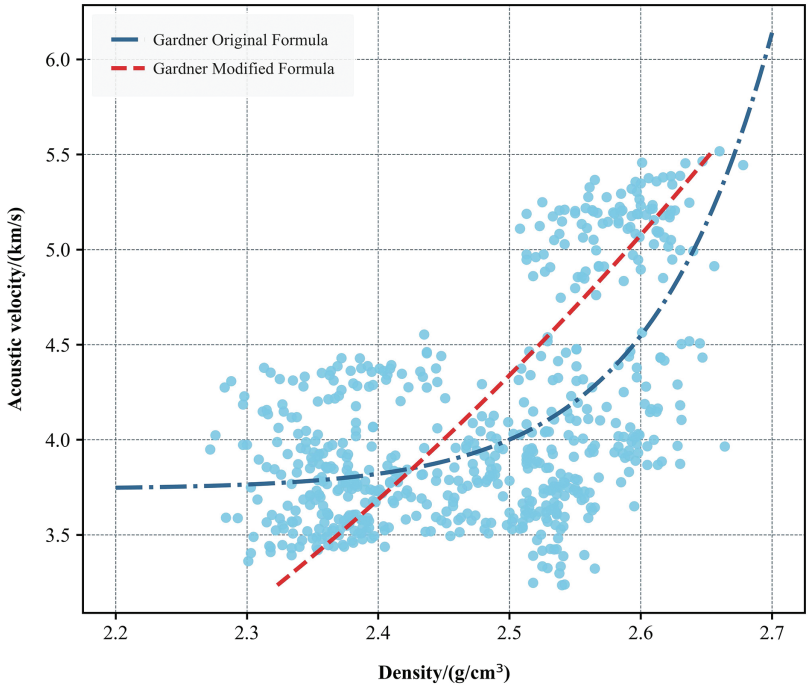


Figure 4: Gardner modified formula

In addition to tectonics, undercompaction and fluid expansion represent the two most prevalent mechanisms for the formation of anomalous high pressures. Consequently, this study focuses exclusively on these three causes of anomalous pressures: tectonics, undercompaction, and fluid expansion. Based on the discrimination theory of wave velocity-density rendezvous diagrams, the origin of anomalous pressure in formations can be identified by plotting the normal trend line in the rendezvous diagram and systematically analyzing the data groups within the target area.

The acoustic velocity-density rendezvous diagram method examines mechanical relationships during formation compaction, which can be categorized into two types: loading curve relationships and unloading curve relationships. In the loading curve scenario, effective vertical stress increases while porosity decreases. Conversely, in the unloading curve relationship, effective vertical stress diminishes with increasing porosity, resulting in elevated pore pressure or reduced overburden pressure. The specific classification is as follows [29]:

1. Undercompaction: Rapid deposition rates prevent normal compaction, leading to higher porosity and lower density relative to the normal trend. Longitudinal wave velocity and resistivity reflect porosity size, resulting in comparatively lower wave velocity and higher resistivity. Wave velocity-density scatter points deviate toward lower values relative to the normal compaction trend section but maintain distribution patterns consistent with the loading curve.

2. Fluid expansion: Following normal sedimentary compaction, fluid expansion props up the rock skeleton, causing a pronounced decrease in wave velocity and a slight reduction in density, accompanied by temperatures exceeding the normal geothermal gradient. Wave velocity-density scatter points are positioned directly below or slightly left of the normal trend line, clustering near the unloading curve within this depth range.
3. Tectonic compression: This process is essentially the inverse of fluid expansion, involving over-compaction beyond normal sedimentary compaction. Pores are compressed and fractures close, yielding higher acoustic wave velocity (with decreased resistivity and porosity) but only minimal density increase. Scatter points concentrate at higher wave velocity values above or slightly upper-right of the normal trend line.

Fig. 5 presents the acoustic velocity-density cross-plot with loading curve. According to the aforementioned theory, overpressure mechanisms in the stratum are identified as follows: Categories 1 and 5, located below the normal trend line, exhibit deviations toward lower wave velocity and density values while maintaining alignment with the loading curve, indicative of undercompaction. Categories 0, 4, and 7 cluster below the normal trend line near the unloading curve, suggesting fluid expansion origins. Categories 2, 3, 6, and 8, positioned above the normal trend line, demonstrate patterns consistent with tectonic compression, analogous to the mudstone bottom opening process.

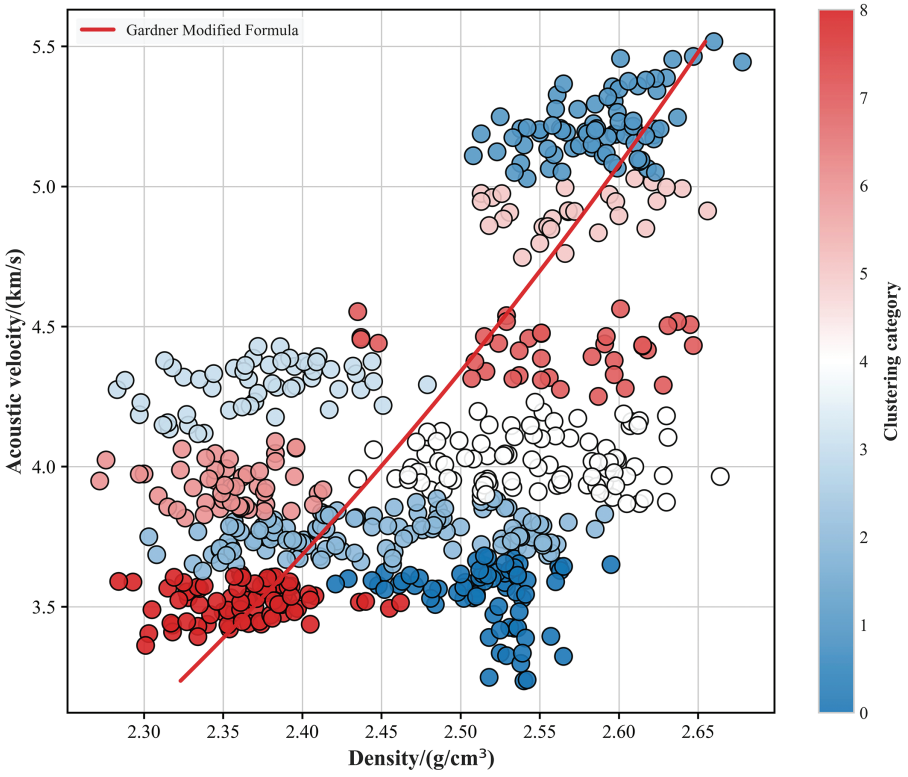


Figure 5: Acoustic velocity-density cross-plot with loading curve

The clustering analysis of anomalous pressure homologous samples employs a hierarchical clustering algorithm to achieve preliminary classification of causal mechanisms. Subsequently, the deviation between acoustic-density rendezvous relationships and theoretical compaction trend lines is

utilized to validate these mechanisms. This process culminates in the development of an identification method for anomalous high-pressure causes in the study area, integrating data-driven and physically-constrained approaches for preliminary mechanism assessment. The proposed multidimensional fusion analysis method effectively resolves interpretation ambiguities inherent to single geophysical methods, thereby establishing a novel technical framework for pressure genesis analysis in deep and complex strata.

3.3 Neural Architecture Search

The numerical classification of anomalous pressure genesis mechanisms under complex geological conditions presents a significant challenge. As a key technology in automated machine learning, NAS addresses the critical challenge of deep learning model tuning, representing an interdisciplinary approach combining optimization and machine learning principles. Prior to the advent of deep learning, traditional machine learning modeling faced similar model tuning challenges. However, the relatively simple architecture of shallow models allowed most studies to incorporate model structure as hyperparameters for optimization. These hyperparameters were typically optimized using black-box methods, though model expansion introduced additional hyperparameters, thereby complicating the optimization process. Traditional modeling approaches prove both labor-intensive and time-consuming, with excessive coding space often preventing algorithm convergence. Furthermore, the extended training durations of deep learning models diminish the computational efficiency of black-box optimization methods. The constraints of manual design paradigms may additionally restrict the exploration of novel architectures. As illustrated in Fig. 6, NAS comprises three core components: search space, search strategy, and evaluation strategy. The search space defines permissible architecture configurations, while the search strategy determines the exploration methodology. The evaluation strategy assesses the performance of identified network architectures. Through iterative application of the search strategy within predefined bounds and continuous feedback from evaluation results, the process continues until satisfactory outcomes are achieved [30].

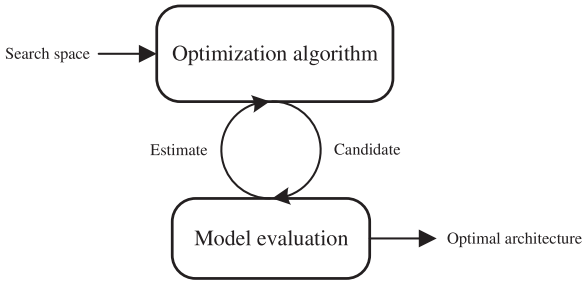


Figure 6: NAS framework

3.3.1 Search Space

Neural Architecture Search (NAS) for petroleum well-log data incorporates domain-specific adaptations in its search space operations: Convolutional operations extract local features and multi-scale anomalies from log curves, such as identifying acoustic interval time anomalies or resistivity step changes. Pooling operations enhance adaptability to vertical resolution and thickness variations through downsampling, highlighting critical anomalies or smoothing noise. Activation functions introduce nonlinearity to model complex coupling relationships between parameters and output causation probabilities. These operations are flexibly combined via directed acyclic graphs (DAGs),

enabling NAS to automatically construct geologically consistent and highly discriminative network architectures. As depicted in Fig. 7, each node in the directed acyclic graph represents an operation, and the edges signify the direction of information transmission. The core of this algorithm lies in shortening the training process through parameter sharing among architectures and achieving fast search.

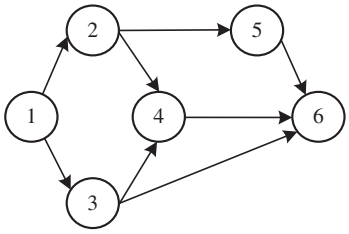


Figure 7: Directed acyelic graph

While retaining basic operations such as convolution and pooling, this study explicitly constrains the search space using geophysical prior knowledge to eliminate structural designs that may violate geological principles (e.g., contradictions between acoustic and resistivity responses). In response to the high sampling rate of well-log data, hyperparameters were adaptively optimized: kernel sizes were set to 3, 5, 7, 9 based on a sampling interval of 0.125 m, corresponding to physical scales of 0.375–1.125 m to capture both thin and thick formations; network depth was limited to 4–8 layers according to the distribution of well-section lengths, balancing long-sequence feature extraction and overfitting risks. These adaptations significantly enhance the geological rationality and interpretability of the generated architectures.

3.3.2 Optimization Algorithm

The neural architecture serves as the optimization target in NAS, wherein the optimization function constitutes a global black-box problem. This formulation does not necessitate an explicit mathematical expression but requires only the optimization objectives and constraints. NAS aims to identify the optimal architecture α that maximizes the objective function on the validation set Dvalid, as formally defined in Eq. (2):

$$\boldsymbol{a}^{*} = \arg\min_{\boldsymbol{a}\in\mathcal{A}} \mathcal{O}\left(\Lambda(\boldsymbol{a}, d_{\text{train}}), d_{\text{valid}}\right) = \arg\max_{\boldsymbol{a}\in\mathcal{A}} f(\boldsymbol{a}) \tag{2}$$

where d_{train}, d_{valid} are the test set and validation set, respectively, the first parameter of $\mathcal{O}$ is defined by Eq. (3):

$$\Lambda(\boldsymbol{a}, d) = \arg\min_{\boldsymbol{\eta}_{\boldsymbol{\theta}}\in\mathcal{M}_{\mathcal{E}}} \left[\mathcal{L}(\boldsymbol{\eta}_{\boldsymbol{\theta}}, \boldsymbol{\theta}, d_{\text{train}}) + \mathcal{R}(\boldsymbol{\theta})\right] \tag{3}$$

where $\mathcal{L}(m_{\boldsymbol{\alpha}}, \boldsymbol{\theta}, d_{\text{train}})$ is the loss function and $\mathcal{R}(\boldsymbol{\theta})$ is the regular term.

In this study, the architecture search did not employ a fully open graph structure but explicitly incorporated physical constraints derived from formation pressure mechanisms. In terms of methodology, NAS primarily relies on two strategies: reinforcement learning (RL), which treats architectures as actions and rewards them based on model evaluation feedback, and evolutionary algorithms (EA), which optimize architectures through population-based selection, crossover, and mutation. Although both RL and EA enable effective exploration of the search space, their discrete nature limits gradient-based optimization efficiency. Instead, this study adopts the gradient-based DARTS method, which transforms the discrete search space into a continuous domain, enabling efficient gradient descent

optimization. Specifically, operation weights are computed via softmax functions to select optimal configurations (Fig. 8), improving search efficiency while introducing geophysically interpretable custom operations (e.g., time-series feature extraction based on compaction trends). Additionally, connection paths that violate geological principles—such as the requirement for synergistic anomalies between acoustic interval time and resistivity in overpressure zones—were prohibited based on prior knowledge. A geological consistency penalty term was also incorporated to prioritize subnetworks consistent with overpressure mechanisms, significantly enhancing the interpretability and generalization capability of the architectures [31].

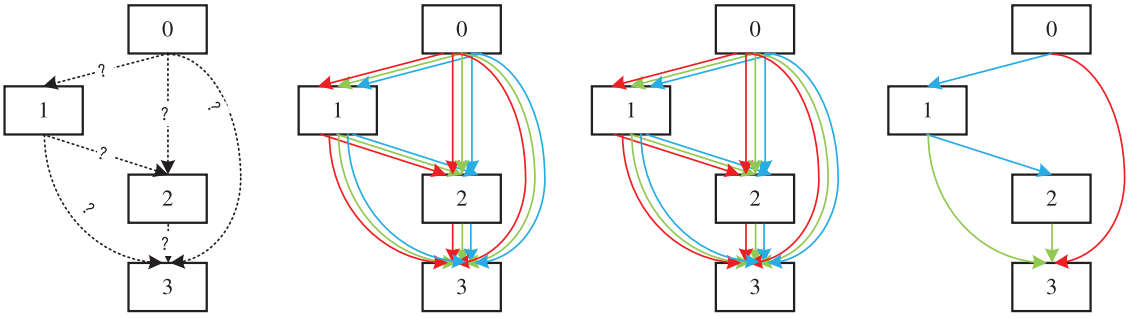


Figure 8: Overall flowchart of DARTS

3.3.3 Model Evaluation

Model evaluation dominates NAS time consumption, prompting several optimization methods. The first approach uses lower fidelity: shortened training time, dataset subsets, or low-pixel data. However, architecture ranking discrepancies tend to widen with data disparities. Another method employs surrogate models to estimate individuals, reducing evaluation costs. Bayesian optimization's model evaluation represents one surrogate type, applicable to RL or EA. A key surrogate model challenge involves management trade-offs: more accurate models become time-consuming, while inaccurate ones poorly estimate offspring quality. Notably, optimal surrogates needn't be maximally accurate—only capable of tracking original model trends to correctly rank and select individuals.

To tackle noise and class imbalance in well-log data, DARTS optimization utilized weighted cross-entropy loss (dynamically adjusted by class sample size) and adversarial regularization for noise robustness. Given the computational burden of model evaluation in NAS, surrogate models built via Bayesian optimization were introduced to approximate performance rankings efficiently. These surrogates prioritized capturing original model ranking trends over absolute accuracy, significantly reducing evaluation costs. The assessment also incorporated geological consistency validation—such as deviation penalties based on velocity-density crossplot comparisons–to ensure geologically plausible architectures.

3.4 Quantitative Evaluation

Quantitative evaluation of anomalous stratigraphic pressure genesis is achieved by integrating machine learning model predictions with statistical weighting analysis. The statistical weighting analysis core involves establishing a vertical frequency weighting model, as expressed in Eq. (4):

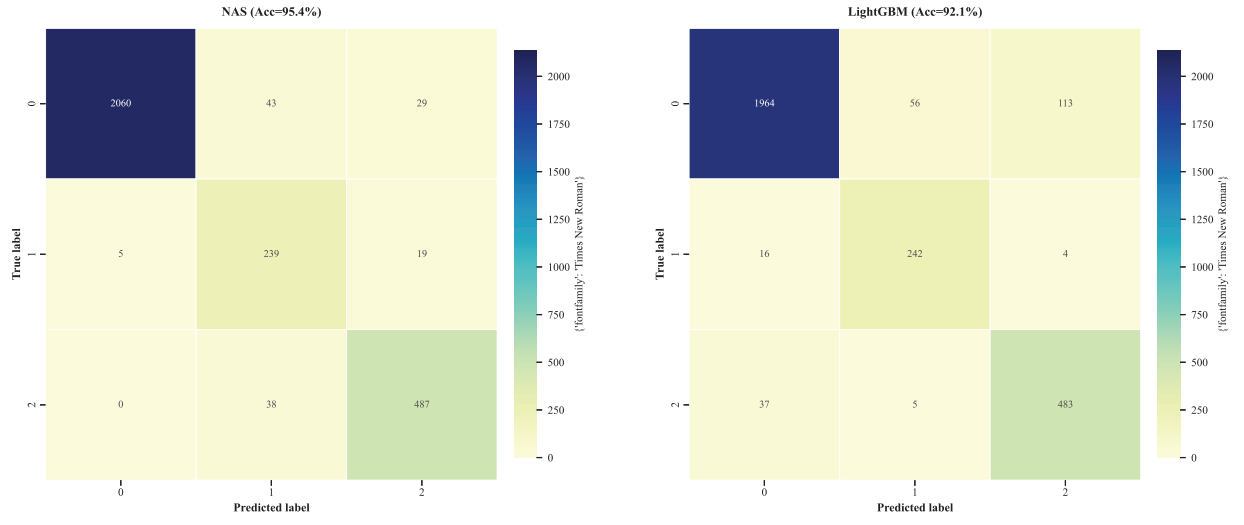
$$W_i = \frac{f_i}{\sum_j f_j} \times \frac{\Delta h_i}{H_{\text{total}}} \tag{4}$$

where f is the frequency of occurrence of the cause at a given depth, Δh_i is the vertical thickness of the cause-dominated section and H_{total} is the total thickness of the studied section.

The target wells undergo quantitative vertical stratification analysis based on calculated weight results. This methodology integrates machine learning characterization with dynamic statistical weight correction, elucidating both the spatial distribution patterns of various genesis categories and their relative contributions within the study area. Notably, high-frequency causal categories predominantly govern regional overpressure mechanisms, while low-frequency categories exert minimal localized pressure influence. Such differentiation enables precise identification of dominant orogenic factors controlling pressure distribution patterns vs. secondary contributors. These findings provide critical insights into geopressure formation mechanisms while establishing a robust quantitative framework for subsequent geohazard risk assessment and prevention strategies.

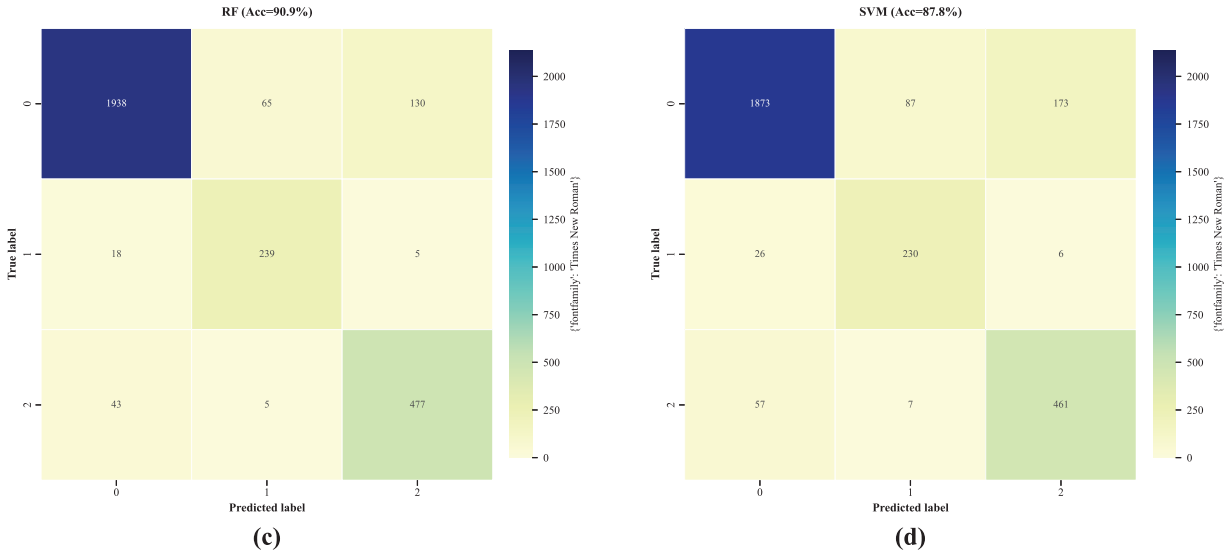
3.5 Model Validation

In this study, SVM, RF, and LightGBM machine learning models were built to verify the accuracy of the reliability of NAS. Data records gathered from well W1 and well W2 were unified into a single dataset. Specifically, 70% of the data records were used for algorithm training, 15% of the data points were used for algorithm testing, and the remaining 15% were used for validation. To ensure all training and validation records were included in the evaluation, 8-fold cross-validation was performed on the combined training and validation subsets. Fig. 9 compares four algorithms using confusion matrices, showing NAS achieves the highest accuracy (95.4%), outperforming LightGBM (92.1%), RF (90.9%), and SVM (87.8%). NAS demonstrates excellent pore pressure genetic classification with 99.8% precision and 96.6% recall for undercompaction by capturing velocity-stress linearity, 90.9% recall for hydrothermal pressurization despite feature overlaps, and 92.7% recall with 91.9% F1-score for mud diapirism. These results confirm NAS's superior pattern recognition and generalization in complex tasks.



(a) (b)

Figure 9: (Continued)



(c) (d)

Figure 9: Confusion matrix analysis for SVM, RF, LightGBM, and NAS models on the total dataset (n = 2921 from wells W1 & W2). (a) NAS; (b) LightGBM; (c) RF; (d) SVM

Fig. 10 compares the classification performance of four machine learning algorithms on the total dataset (n = 2921) using AUC-ROC curves. The results show that the NAS model (AUC = 0.82), with its auto-optimized neural network architecture, achieved the best overall performance. Its curve is closest to the top-left corner, indicating its ability to maintain a high True Positive Rate (TPR) while controlling the False Positive Rate (FPR) at a low level. LightGBM (AUC = 0.80) performed closely behind, with a minimal performance gap, and demonstrated strong competitiveness—especially in the low-FPR (high-specificity) region—making it a highly efficient and robust benchmark. In contrast, Random Forest (RF) and Support Vector Machine (SVM) showed relatively poorer performance. SVM, in particular, with an AUC of 0.61, performed near random guessing levels, indicating limited discriminative ability on this dataset.

To address concerns regarding the high accuracy achieved with limited data, we conducted data volume sensitivity analyses to validate the reliability of our results. Specifically, we gradually reduced the size of the training dataset from 100% to 30% of the original data. The results showed that when the data volume was reduced to 30%, the accuracy of the NAS model decreased to 88.5%, while those of LightGBM, RF, and SVM dropped to 84.2%, 80.7%, and 75.1%, respectively. Although all models exhibited a reasonable performance decline with reduced data, NAS maintained a relative advantage in stability.

Furthermore, feature ablation experiments were performed to assess the model's dependence on critical features. When the velocity feature was excluded, the accuracy of NAS decreased by 4.3 percentage points, whereas that of SVM dropped by 8.7 percentage points. This indicates that while NAS exhibits a certain degree of reliance on key features, the magnitude of its performance fluctuation remains within an acceptable range. The comprehensive analysis demonstrates that NAS can maintain relatively stable performance even with limited data, though its practical application requires reliable access to key features.

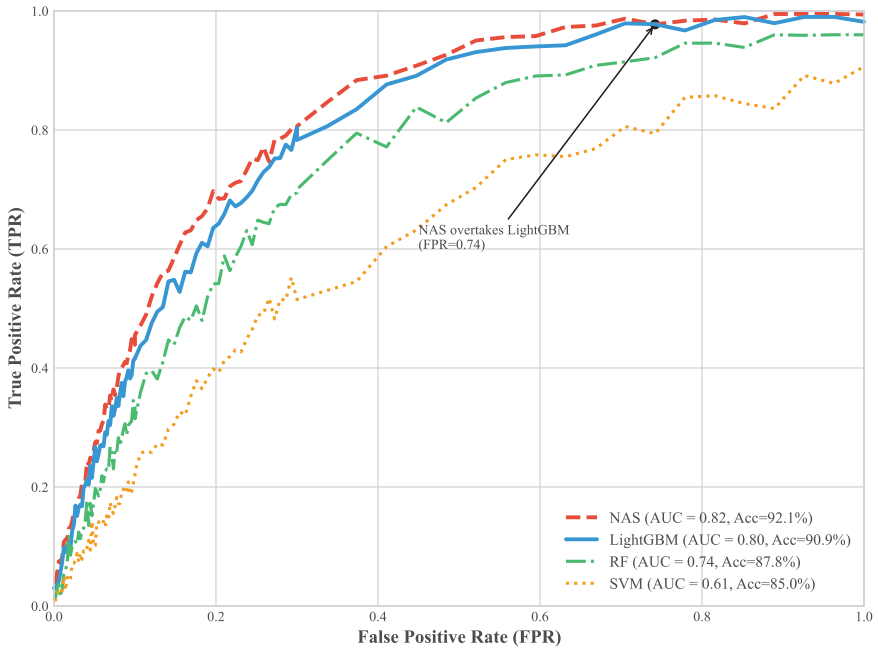


Figure 10: AUC-ROC validation curves for different models on the total dataset (n = 2921 from wells W1 & W2)

4 Case Analysis

This study investigates the Rio-Dere Basin in Cameroon as a representative case for applying quantitative evaluation models of stratigraphic overpressure mechanisms, with the main stratigraphic sequences illustrated in Fig. 11. Situated at the northeastern margin of the Niger Delta Basin—a significant hydrocarbon province—the study area features deep turbidite sandstone reservoirs (2500–3500 m depth) exhibiting extreme pressure-temperature conditions: maximum pressure coefficient of 2.10, wellbore temperatures reaching 161°C, and geothermal gradients up to 4.6°C/100 m. These Neogene formations, influenced by fracture blocks and mudstone underplating, demonstrate complex overpressure genesis through three primary mechanisms: stratigraphic undercompaction driven by rapid sedimentation, high-temperature fluid expansion associated with geothermal anomalies, and mudstone underplating coupled with fracture block development.

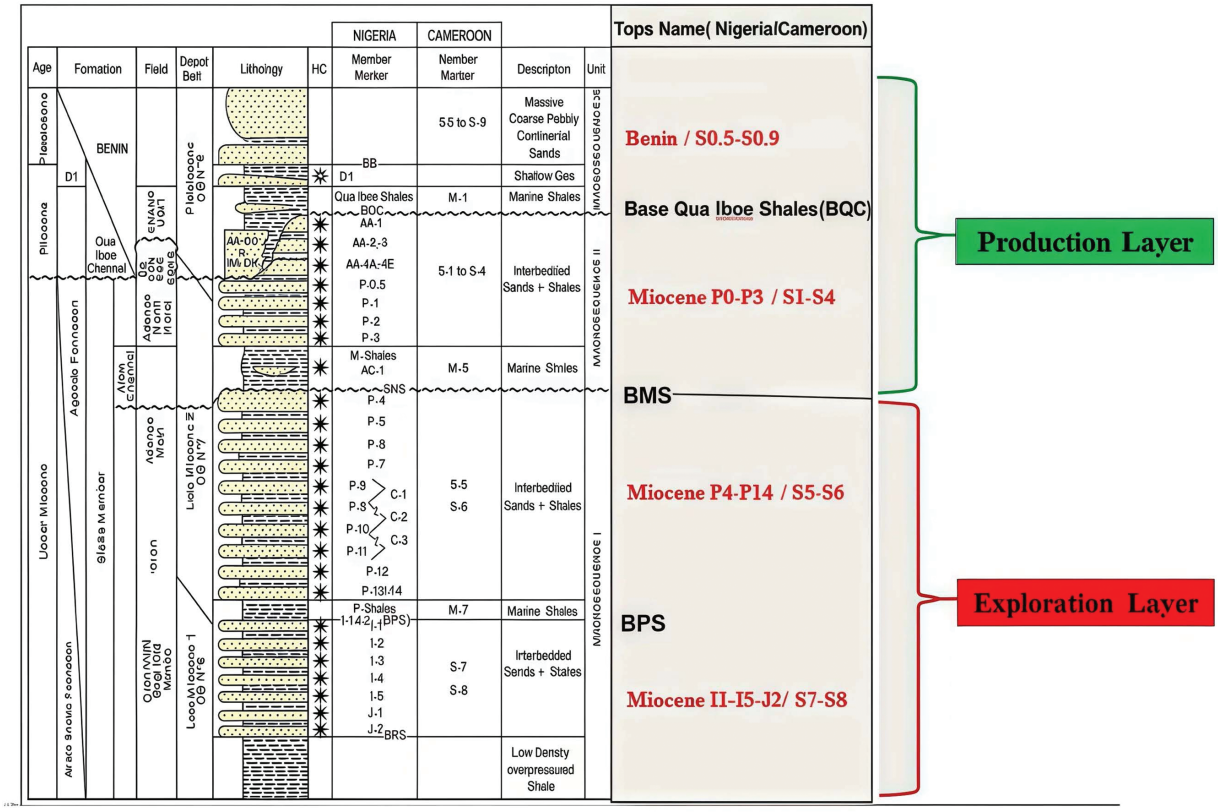


Figure 11: Stratigraphic sequence map of the target work area

Three representative wells (W2, W3, and W4) exhibiting overpressure characteristics were selected for detailed analysis. Comprehensive datasets including effective well depth, sonic velocity, bulk density, resistivity, and natural gamma ray measurements were systematically extracted and processed to establish a robust foundation for investigating formation pressure genesis mechanisms. Initial analysis focused on the sonic velocity-depth relationship, which identified a well-defined normal compaction trend within the upper 1300 m interval, as illustrated in Fig. 12.

The logging data from the normal compaction section contributed to the derivation of Gardner's modified equation. This made it possible to obtain the normal trend line of wave velocity-density, which was used as the analytical baseline. The corrected Gardner equation, established with the data from this section, is presented in Eq. (5):

$$\rho = 1.7321 \left(\frac{1{,}000{,}000}{\Delta t} \right)^{0.25} \tag{5}$$

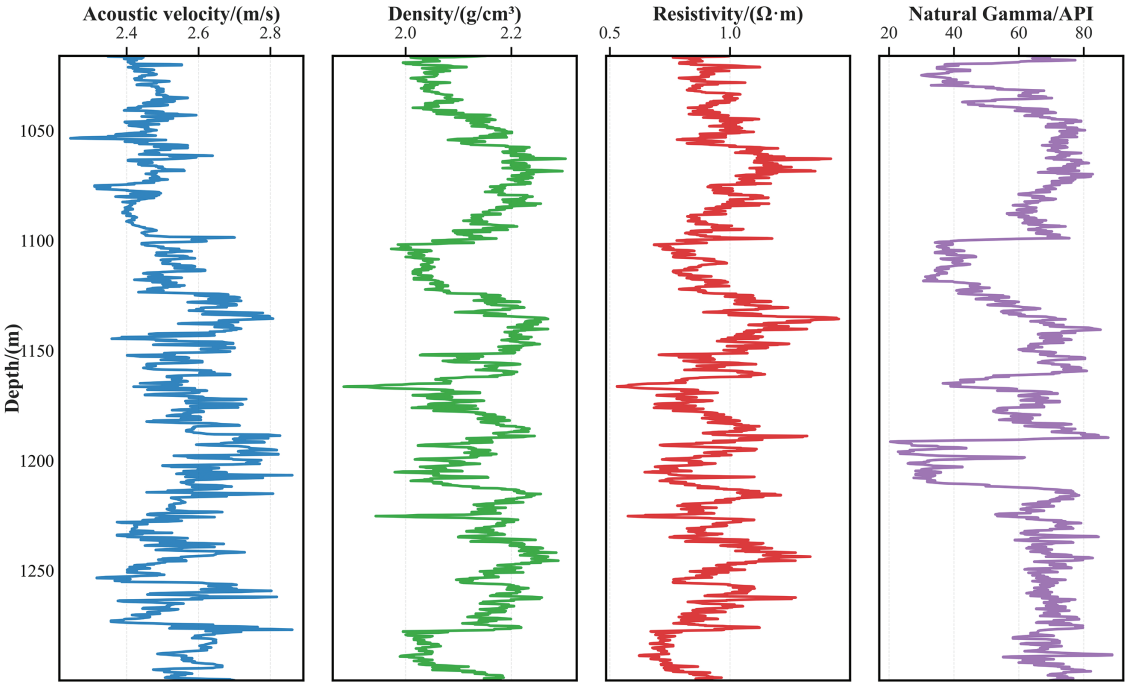


Figure 12: Logging data related to normal compaction sections

To investigate the genesis of abnormal formation pressure, longitudinal wave velocity and density data were analyzed using a velocity-density crossplot approach. This method, combined with established overpressure mechanism identification techniques, enabled the extraction of sample data clusters exhibiting similar characteristics. Fig. 13 displays the standard velocity-density normal trend line derived from normal compaction zone data, serving as a reference curve. Comparative analysis between the velocity-density crossplot and the geological context of the study area revealed distinct patterns in abnormal pressure origins:

1. Categories 1, 4, and 5, positioned below or laterally adjacent to the normal trend line, indicate undercompaction-related mechanisms.
2. Category 3, located below the normal trend line, suggests hydrothermal pressurization as the primary cause.
3. Categories 0 and 2, clustered at high-velocity values directly above or slightly right of the normal trend line, correlate with anomalies induced by mudstone bottoming effects.

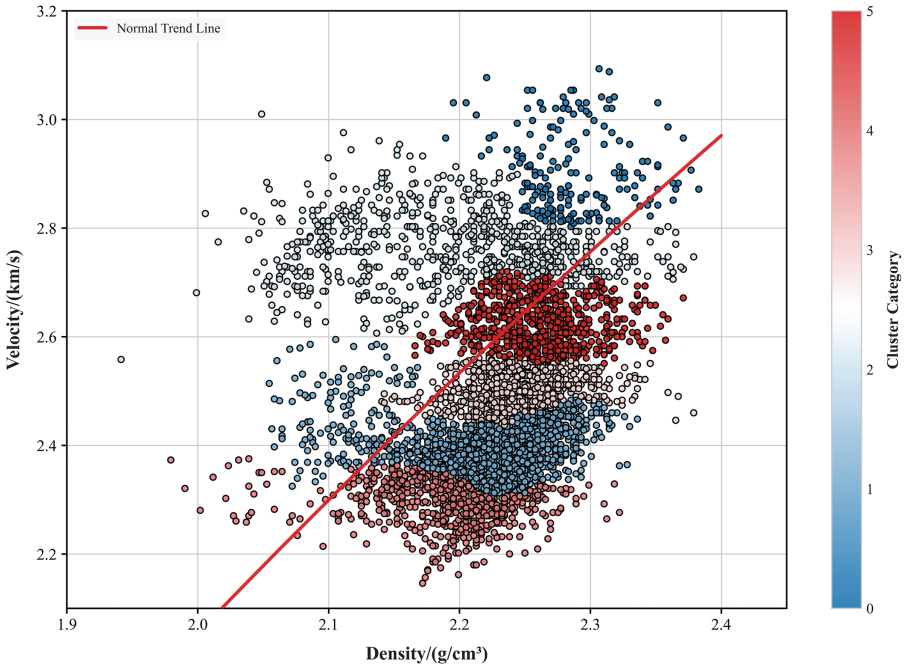


Figure 13: Acoustic velocity-density cross-plot of selected wells

The quantitative evaluation model of overpressure mechanism is applied to assess the overpressure mechanisms in the prospective well W5. For distinct stratigraphic layers, the model yields varying superposition relationships among multiple genetic mechanisms, including undercompaction, fluid expansion, and mudstone bottom opening. Consequently, when computing overpressure genesis weights in this region, the superposition relationship within specific depth intervals can be determined through depth-based segmentation. Fig. 14 presents the quantitative evaluation results of genesis mechanisms for high-pressure sections at varying depths in the proposed well W5 within Cameroon's Rio-Delay Basin. The target wells are vertically stratified and analyzed based on both the calculated weight results and the geological context of the study area:

1. Rapid Depositional Layers: Rapid depositional layers in well W5 predominantly occur in shallow—middle layers, particularly in Sections S2–M3 (1522–1611 m). The development of overpressure is chiefly attributed to the rapid accumulation of thick-bedded shale and lenticular sandstone in the former deltaic facies. This accumulation results in restricted drainage of pore water and undercompaction. The confined sandstone lenses function as localized fluid storage units, with pore pressure gradients reaching up to 1.8 MPa/100 m. Lateral compression at the bottom of the mudstone's penetration boundary further restricts the release of vertical pressure, enabling pore fluids to bear 40%–60% of the overburden load. Logging data exhibit significant departures from the normal compaction trend. These departures include shifts in acoustic time-difference, reduced rock chip density, and elevated clay mineral content, all of which are characteristics of undercompaction. This undercompaction mechanism contributes to 73% of the total overpressure genesis in the study area and represents the dominant high-pressure confinement mechanism in the shallow-medium horizons.
2. Fault—Influenced Zone: This stratigraphic interval ranges from Section M3 (1546.5–1611 m) to Section S5 (1908.5–2036.5 m) and is vertically distributed along the dorsal fault. Tectonic

extrusion concentrates stress, and the horizontal stress increases to 85%, especially at the root of the reverse fault (1900–2036.5 m). Here, the modulus of elasticity of the faulted mud increases non—linearly, and the rupture pressure coefficient drops sharply to 0.33. Initial faulting activity initiates tectonic extrusion and pressure enhancement, and subsequent relaxation of the fault zone leads to periodic pressure fluctuations. Enhanced vertical stress gradients at depth promote the development of high-angle rupture surfaces. The combination of mudstone bottoming and faulting amplifies the deep high-pressure closure, accounting for 9% of the regional overpressure.

3. Diagenetic Evolution Zone: The overpressure in Section S4 (1664–1736.5 m) and the deep hydrocarbon window (1600–2000 m) is mainly caused by fluid expansion. The transformation of montmorillonite to illite releases interlayer water, leading to an expansion of pore fluid volume. The thermal evolution of organic matter generates hydrocarbons, which, in combination with stratigraphic temperature gradients, drives hydrothermal pressurization. Clay minerals reorganize to form a dual-pore system. Lateral extrusion of openings at the base of the mudstone impedes vertical escape channels in the lenticular sandstone, creating a deep high-pressure storage system that accounts for 18% of the overpressure in the study area.

Based on the quantitative evaluation results, geological interpretation and exploration strategies in this area can be significantly improved through a compartmentalized analysis that accounts for the dominant overpressure mechanisms at different depths. In the shallow to middle intervals (1522–1611 m), it is essential to identify undercompaction compartments caused by rapid deposition, using acoustic time difference anomalies and sandstone lenses as key indicators for predicting high-pressure sweet spots. In the fault-affected zone (1546–2036 m), the characteristics of the tectonic stress field should be integrated to assess the impact of periodic pressure fluctuations near the root of reverse faults on wellbore stability. In the deep diagenetic evolution interval (1600–2000 m), particular attention should be paid to the significant contribution of fluid expansion to overpressure generation. This mechanism not only reflects hydrocarbon generation from active source rocks but also indicates effective fluid retention conditions in deep strata, making such intervals favorable targets for natural gas and light oil exploration. By clarifying the dominant overpressure mechanisms and their weighting in different intervals, the design of drilling fluid density and the accuracy of pressure prediction can be optimized, thereby effectively reducing exploration risks.

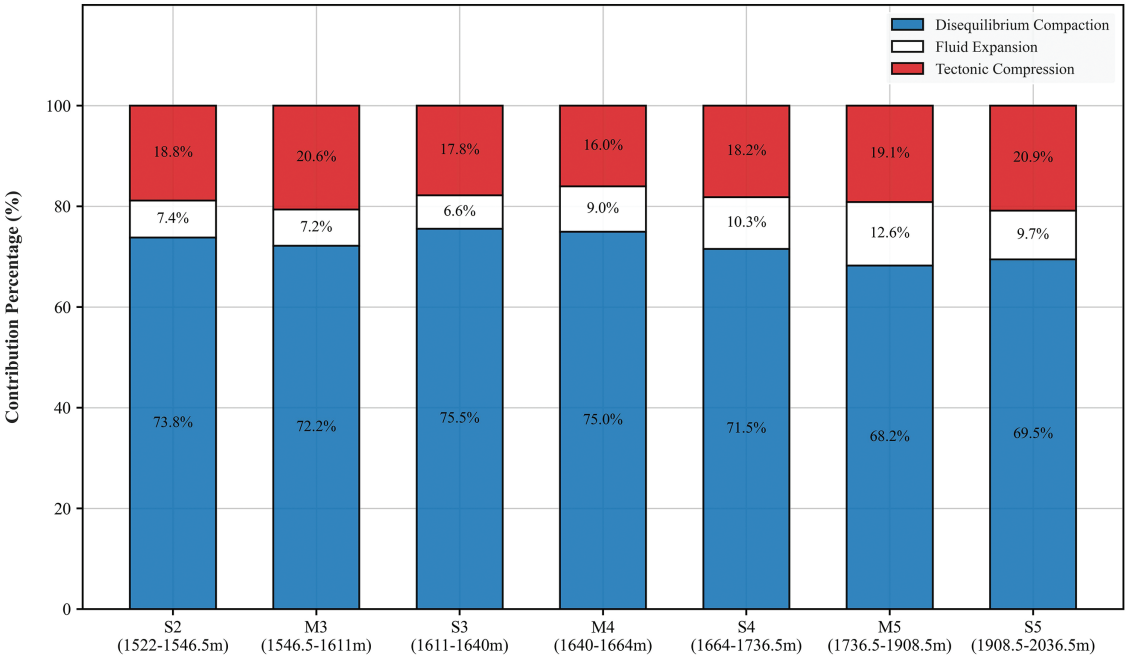


Figure 14: Quantitative evaluation results of stratified overpressure genesis for the new well W5

5 Conclusions

Based on the neural structure search algorithm, a quantitative evaluation model of the stratigraphic overpressure mechanism was established. The research findings can be principally summarized as follows:

1. A model for identifying the formation overpressure mechanism based on the clustering algorithm was proposed. Initially, this model mines similar samples of the anomalous pressure genesis mechanism via hierarchical clustering analysis. It utilizes two crucial geophysical parameters, namely wave velocity and density, for a preliminary determination of the anomalous high-pressure genesis mechanism in the study area. Subsequently, the data clusters are further partitioned. This is achieved by supplementing Gardner's correction equation as a normal trend line on the wave velocity-density scatter plot of the anomalous high-pressure section.
2. A quantitative evaluation model of the stratigraphic overpressure mechanism was established. The core of this model is to construct an efficient and precise quantitative evaluation model of the stratigraphic overpressure mechanism. This construction is based on the identification results of overpressure causes obtained from the clustering analysis of anomalous pressure homologous samples, in combination with the neural architecture search algorithm. The model is designed to address the numerical categorization problem of the anomalous pressure genesis mechanisms. It accurately identifies different anomalous pressure genesis mechanisms and quantifies the contribution of each mechanism by employing logging data such as wave velocity and density as inputs.
3. The experimental results indicate that the anomalous formation pressure genesis identification model proposed in this study exhibits excellent performance in the application of well W1,

with an overall classification accuracy of 95.4%. Through high-dimensional feature fusion and discriminant boundary optimization, the model accomplishes specific identification of three main control mechanisms: undercompaction (with an accuracy of 99.8%), hydrothermal pressurization (with a recall of 90.9%), and mudstone bottom-penetration (with an F1 score of 91.9%). Notably, its ability to capture the acoustic velocity-effective stress linear response features of the undercompaction mechanism is particularly remarkable. A small number of misclassifications reveal the coupling among the genesis mechanisms.

The proposed method is developed based on basin settings with well-developed mudstone diapir structures, where overpressure mechanisms and their interactions exhibit characteristic coupling patterns. Consequently, its applicability may be limited in basins lacking such tectonic activity, as the dominant overpressure mechanisms and their coupling behaviors may differ fundamentally. Additionally, the method is sensitive to variations in data quality and geological heterogeneity. Future research will focus on incorporating multi-scale geological constraints—such as seismic attributes and sedimentary facies models—along with data augmentation and quality control techniques to enhance the method's generalizability across diverse basin types and complex heterogeneous conditions.

Acknowledgement: Not applicable.

Funding Statement: Not applicable.

Author Contributions: Conceptualization: Chaowei Li and Haoyu Pan; Methodology: Haoyue Chen and Yang Xu; Formal analysis and investigation: Huawei Zhang and Zhiqiang Hu; Language polishing: Song Deng. All authors reviewed the results and approved the final version of the manuscript.

Availability of Data and Materials: The data cannot be provided due to confidentiality agreements. The code can be provided on request.

Ethics Approval: Not applicable.

Conflicts of Interest: The authors declare no conflicts of interest to report regarding the present study.

References

1. Sun S, Shi Z, Dong D, Bai W, Wei L, Yin J, et al. Formation and evolution of shale overpressure in deep wufeng-longmaxi formation in southern sichuan basin and its influence on reservoir pore characteristics. Front Earth Sci. 2024;12:1375241. doi:10.3389/feart.2024.1375241.
2. Li C, Luo X, Zhang L, Fan C, Xu C, Liu A, et al. New understanding of overpressure responses and pore pressure prediction: insights from the effect of clay mineral transformations on mudstone compaction. Eng Geol. 2022;297(5):106493. doi:10.1016/j.enggeo.2021.106493.
3. Bowers GL. Pore pressure estimation from velocity data: accounting for overpressure mechanisms besides undercompaction. SPE Drill Complet. 1995;10(02):89–95. doi:10.2118/27488-pa.
4. Dasgupta S, Chatterjee R, Mohanty SP. Magnitude, mechanisms, and prediction of abnormal pore pressure using well data in the krishna–godavari basin, east coast of India. Am Assoc Pet Geol Bull. 2016;100(12):1833–55. doi:10.1306/05131615170.
5. Kaeng GC, Sausan S, Simatupang Z. Overpressure mechanisms in compressional tectonic of borneo deepwater fold-thrust belt. In: The AAPG/datapages combined publications database. Jakarta, Indonesia: Indonesian Petroleum Association; 2016.

6. Hermanrud C, Wensaas L, Teige G, Bolås HN, Hansen S, Vik E. Shale porosities from well logs on haltenbanken (offshore mid-norway) show no influence of overpressuring. In: Abnormal pressures in hydrocarbon environments. Vol. 70. Tulsa, OK, USA: AAPG Memoir; 1998. p. 65–85.
7. Bowers GL. Determining an appropriate pore-pressure estimation strategy. In: Offshore Technology Conference; 2001 Apr 30–May 3; Houston, TX, USA. doi:10.4043/13042-ms.
8. Satti IA, Ghosh D, Yusoff WI, Hoesni MJ. Origin of overpressure in a field in the southwestern malay basin. SPE Drill Complet. 2015;30(3):198–211.
9. Wang Q-C, Chen D-X, Gao X-Z, Li M-J, Shi X-B, Wang F-W, et al. Overpressure origins and evolution in deep-buried strata: a case study of the jurassic formation, central Junggar Basin, western China. Pet Sci. 2023;20(3):1429–45.
10. O'connor S, Swarbrick R, Lahann R. Geologically-driven pore fluid pressure models and their implications for petroleum exploration. Introduction to thematic set. Geofluids. 2011;11(4):343–8. doi:10.1111/j.1468-8123.2011.00354.x.
11. Tingay MR, Morley CK, Laird A, Limpornpipat O, Krisadasima K, Pabchanda S, et al. Evidence for overpressure generation by kerogento-gas maturation in the northern malay basin. Am Assoc Pet Geol Bull. 2013;97(4):639–72. doi:10.1306/09041212032.
12. Liu C, Yang T, Li P, Feng D, Huo H. Origin mechanism of overpressure in saline lacustrine formation of the paleogene and neogene in the western Qaidam Basin, NW China. Carbonates Evaporites. 2024;39(1):2. doi:10.1007/s13146-023-00912-7.
13. Shi L, Jin Z, Zhu X, Lin M, Guan B. Generation mechanism of overpressures caused by disequilibrium compaction in the northwestern Bozhong Subbasin, China. J Pet Explor Prod Technol. 2024;14(7):1843–57. doi:10.1007/s13202-024-01811-w.
14. Kong L, Chen H, Ping H, Zhai P, Liu Y, Zhu J. Formation pressure modeling in the baiyun sag, northern south China sea: implications for petroleum exploration in deep-water areas. Mar Pet Geol. 2018;97(11):154–68. doi:10.1016/j.marpetgeo.2018.07.004.
15. Hakimi MH, Alaug AS, Mohialdeen IM, Kahal AY, Abdulelah H, Hadad YT, et al. Late jurassic arwa member in south-eastern Al-jawf Sub-basin, NW sabatayn basin of Yemen: geochemistry and basin modeling reveal shale-gas potential. J Nat Gas Sci Eng. 2019;64:133–51. doi:10.1016/j.jngse.2019.01.022.
16. Li C, Zhang L, Luo X, Lei Y, Yu L, Cheng M, et al. Overpressure generation by disequilibrium compaction or hydrocarbon generation in the paleocene shahejie formation in the chezhen depression: insights from logging responses and basin modeling. Mar Pet Geol. 2021;133:105258. doi:10.1016/j.marpetgeo.2021.105258.
17. Zhang L, Li C, Luo X, Zhang Z, Zeng Z, Ren X, et al. Vertically transferred overpressures along faults in mesozoic reservoirs in the central Junggar Basin, northwestern China: implications for hydrocarbon accumulation and preservation. Mar Pet Geol. 2023;150:106152. doi:10.1016/j.marpetgeo.2023.106152.
18. Zhang W, Wu C, Liu S, Liu X, Wu X, Lu X. Impact of disequilibrium compaction and unloading on overpressure in the southern Junggar Foreland Basin, NW China. Mar Pet Geol. 2024;164:106819. doi:10.1016/j.marpetgeo.2024.106819.
19. Xu Y, Yang J, Hu Z, Zhao Q, Li L, Yin Q. Causes of multi-mechanism abnormal formation pressure in offshore oil and gas wells. Appl Sci. 2024;14(22):10149. doi:10.3390/app142210149.
20. Han JW, Kamber M. Data mining: concepts and techniques. San Francisco, CA, USA: Morgan Kaufmann Publishers; 2006.
21. Wang W-C. Setting up evaluate indicators for slope control engineering based on spatial clustering analysis. Nat Hazards. 2018;93(2):921–39. doi:10.1007/s11069-018-3332-x.
22. Yu H, Chen G, Gu H. A machine learning methodology for multivariate porepressure prediction. Comput Geosci. 2020;143:104548. doi:10.1016/j.cageo.2020.104548.
23. Radwan AE, Wood DA, Radwan AA. Machine learning and data-driven prediction of pore pressure from geophysical logs: a case study for the Mangahewa gas field, New Zealand. J Rock Mech Geotech Eng. 2022;14(6):1799–809. doi:10.1016/j.jrmge.2022.01.012.

24. Zhang G, Davoodi S, Band SS, Ghorbani H, Mosavi A, Moslehpour M. A robust approach to pore pressure prediction applying petrophysical log data aided by machine learning techniques. Energy Rep. 2022;8(3):2233–47. doi:10.1016/j.egyr.2022.01.012.
25. Deng S, Pan H-Y, Wang H-G, Xu S-K, Yan X-P, Li C-W, et al. A hybrid machine learning optimization algorithm for multivariable pore pressure prediction. Pet Sci. 2024;21(1):535–50. doi:10.1016/j.petsci.2023.09.001.
26. Deng S, Pei C, Yan X, Hao H, Cui M, Zhao F, et al. Lost circulation prediction method based on an improved fruit fly algorithm for support vector machine optimization. ACS Omega. 2023;8(36):32838–47. doi:10.1021/acsomega.3c03919.
27. Deng S, Pan H, Li C, Yan X, Wang J, Shi L, et al. A real-time lithological identification method based on smote-tomek and icsa optimization. Acta Geol Sin. 2024;98(2):518–30. doi:10.1111/1755-6724.15144.
28. Pan W, Ting L, Feng T. Rock density optimization algorithm based on gardner formula. Mud Logg Eng. 2016;27(04):12–5.
29. Ye Z, Fan H, Cai J, Ji R, Li C, Liu G. Investigation and application of a discrimination method for abnormal high formation pressure forming mechanism. J China Unive Petrol. 2012;36(3):102–7.
30. Jin C, Huang J. Inner loop-based modified differentiable architecture search. IEEE Access. 2024;12(31):41918–33. doi:10.1109/access.2024.3377888.
31. Jing K, Xu J, Zhang Z. A neural architecture generator for efficient search space. Neurocomputing. 2022;486:189–99. doi:10.1016/j.neucom.2021.10.118.