Petroleum Science 23 (2026) 4394–4413

Contents lists available at ScienceDirect

Petroleum Science

journal homepage: www.keaipublishing.com/en/journals/petroleum-science

Original Paper

Application of the AutoMix hybrid model for in-situ stress prediction in shale gas reservoirs

Hang Yuan [a,b,c,d], Yu-Ping Sun [c,d,*], Wei Xiong [a,b,c,d], Wen-Te Niu [e], Qian Cheng [f]

[a] University of Chinese Academy of Sciences, Beijing, 100049, China
[b] Institute of Porous Flow and Fluid Mechanics, Chinese Academy of Sciences, Langfang, 065007, Hebei, China
[c] Research Institute of Petroleum Exploration and Development, Beijing, 100089, China
[d] National Energy Shale Gas R&D (Experimental) Center, Langfang, 065007, Hebei, China
[e] School of Vehicle and Energy, Yanshan University, Qinhuangdao, 066004, Hebei, China
[f] Beijing Deep Green Energy Technology Company Limited, Beijing, 100071, China

A R T I C L E I N F O

Article history:
Received 18 May 2025
Received in revised form
28 October 2025
Accepted 11 March 2026
Available online 18 March 2026

Edited by Min Li

Keywords:
Unconventional reservoirs
In-situ stress prediction
Machine learning
Ensemble modeling
Logging data

A B S T R A C T

Accurate in-situ stress prediction is critical to optimizing hydraulic fracturing design and enhancing reservoir stimulation in deep shale gas development. However, traditional methods often struggle with accuracy and adaptability due to the geological complexity and high-dimensional, nonlinear nature of well-logging data. To address this challenge, we propose AutoMix, a hierarchical hybrid prediction model built on the AutoGluon framework. AutoMix integrates leading machine learning algorithms—LightGBM, CatBoost, ExtraTrees, RandomForest, and XGBoost—along with deep learning architectures such as NeuralNetTorch and NeuralNetFastAI, using multi-level stacking and weighted ensembling to enhance prediction performance for maximum and minimum horizontal stress (σ_H and σ_h). The model was trained and evaluated using 278,003 sets of real-world logging data from 14 horizontal shale gas wells in the Zigong area of the Sichuan Basin. Eleven algorithms were benchmarked under varying sampling ratios (1/1000, 1/100, 1/10, and full dataset). Results show that AutoMix consistently achieves the lowest errors across all scales (MSE = 0.0046, MRE = 0.454, R^2 = 0.9998), significantly outperforming any single model with strong generalization and robustness. Feature importance analysis identifies vertical stress, Poisson's ratio, feldspar content, and Young's modulus as key predictors, consistent with established geomechanical theory. Weight distribution analysis further reveals that deep learning models dominate in low-sample regimes, while tree-based models such as ExtraTrees and RandomForest take precedence as data volume increases—each excelling in σ_h and σ_H prediction, respectively. AutoMix demonstrates exceptional performance under data scarcity, nonlinearity, and multicollinearity, offering strong applicability and scalability for engineering-grade stress prediction in unconventional reservoirs.

1. Introduction

1.1. Necessity of reservoir stress prediction

The global energy sector is transitioning from conventional oil and gas to new clean energy. The depletion of traditional fossil fuels and the technical barriers to developing new clean energies, such as hydrogen and solar, have created an energy supply gap. This has led to the rise of unconventional oil and gas resources as strategic transitional resources (Wang et al., 2016). Shale gas, with its abundant reserves, high energy conversion efficiency, and low carbon emissions, plays an irreplaceable role in ensuring energy security and achieving carbon neutrality goals (Zou et al., 2023).

Shale reservoirs are typically of low porosity and permeability; efficient development relies on multi-stage fracturing of horizontal wells to create complex fracture networks (Lei et al., 2020). As fracturing technologies mature, precise reservoir stress characterization becomes essential. Accurate in-situ stress prediction

* Corresponding author.
E-mail address: sunyuping01@petrochina.com.cn (Y.-P. Sun).
Peer review under the responsibility of China University of Petroleum (Beijing).

https://doi.org/10.1016/j.petsci.2026.03.023
1995-8226/

therefore underpins unconventional-reservoir development, guides well-level engineering decisions, informs fracability evaluation, and helps mitigate environmental risks (Sun and Wang, 2024; Li et al., 2023).

1.2. Conventional reservoir stress calculation methods

Reservoir stress, defined as the stress state of reservoir rocks in their natural condition, results from various factors, including geological structural movements and temperature changes (Zhang et al., 2021). It comprises the vertical stress σ_V caused by the rock's self-weight and the maximum and minimum horizontal principal stresses σ_H and σ_h generated by tectonic activity. The magnitude and direction of reservoir stress are key parameters in designing drilling and fracturing programs, directly impacting casing steel grade selection and hydraulic fracture propagation. Conventional reservoir stress calculation methods, as summarized by (Lin et al., 2018), include empirical and statistical methods (Ljunggren et al., 2003), field testing (Haimson et al., 1991), geophysical experiments (Dinmohammadpour et al., 2022), and numerical simulation.

Empirical and statistical methods establish correlations based on regional geological data or historical records, making them suitable for preliminary estimation. They are applicable for predicting reservoir stress in shale reservoirs with relatively complete geological data and weak formation heterogeneity (Lee et al., 2018).

Field testing, a core technology for characterizing reservoir stress, directly measures rock stress parameters, including stress magnitude and direction, through on-site instrument responses (Robertson, 1986). This method avoids systematic errors from model assumptions in indirect approaches, offering high data reliability by reflecting true geological stress fields. It primarily includes hydraulic fracturing (Haimson and Fairhurst, 1969), stress recovery, core stress relief, and anelastic strain recovery (ASR) (Teufel, 1983). Field testing is best suited for brittle reservoirs with low fracture density and strong lithological homogeneity. To prevent instrument failure, it requires geological conditions with low reservoir temperature and pressure coefficients, typically in shallow to medium-depth formations. However, deep shale gas reservoirs present challenges due to their complex geological conditions, extremely high reservoir pressure (pressure coefficient >2 MPa/100 m), strong lithological heterogeneity, and complex natural fracture-weakness systems. In such environments, sensors for direct measurement are prone to zero drift and packaging failure due to high temperature and pressure. Moreover, the lack of detailed characterization of natural fractures and weak planes can lead to inaccurate stress measurement using hydraulic fracturing, as fracture fluid may propagate along weak planes or natural fractures. Thus, field testing in deep shale gas reservoirs faces challenges related to instrument tolerance, data interpretation, and engineering risks.

Geophysical experiments indirectly invert rock stress by establishing constitutive relationships between rock physical parameters and reservoir stress, based on continuum mechanics theory (Pitcher, 2015). This approach circumvents direct measurement failures caused by sensor intolerance in deep shale gas environments. They mainly include seismic wave inversion, acoustic emission, and geothermal methods (Pan et al., 2022; Seto et al., 1997; Kumari et al., 2017). The application of geophysical experiments requires relatively simple tectonic settings without strong fault activity and high lithological homogeneity. Additionally, high-resolution logging data (e.g., array acoustic, micro-resistivity imaging) and 3D seismic attribute volumes are needed to constrain the multi-solution problems in inversion processes. For example, Lai et al. (2019) integrated density and dipole sonic logs with borehole image logs to derive σ_H azimuth from DITFs/BOs and, via a modified Anderson model, continuously estimate SV σ_h, and σ_H along depth. Similarly, Sanei et al. (2023) built 1D/3D static geomechanical models from well logs—integrating density to compute SV, estimating σ_h via Blanton's relation, σ_H from poroelastic theory, and pore pressure by Eaton—and upscaling these properties to 3D with Kriging/SGS to support stress-field prediction.

The extreme temperature and pressure conditions and complex geological structures of deep shale gas reservoirs somewhat reduce the practicality of geophysical experimentation. Furthermore, theoretical models often oversimplify complex geological mechanical behaviors, making it difficult to distinguish anisotropic differences due to model assumptions. These shortcomings result in significant discrepancies in reservoir stress values obtained from different methods (e.g., seismic inversion and acoustic emission experiments). In segments with well-developed natural fractures, the diverse origins of anisotropic signals introduce greater uncertainty in stress azimuth interpretation. To reduce ambiguity in stress-azimuth interpretation from anisotropic signals, Cao et al. (2025) integrate 3D multi-azimuth seismic anisotropy as a lateral constraint and co-calibrate the inferred 3D SHmax field with FMI logs and earthquake focal-mechanism solutions.

Numerical simulation solves governing equations via FEM/DEM/BEM to reconstruct the stress state in complex geological bodies (Zhao and Qin, 2023; Singha and Chatterjee, 2015; Karatela et al., 2016; Li et al., 2009). It can depict heterogeneous reservoirs, multi-scale fractures, and coupled HPHT conditions, but depends heavily on high-precision geological models and multi-source data; cross-scale data integration and long runtimes limit efficiency and scalability in data-sparse, large-area settings.

These constraints motivate data-driven approaches, and recent advances in artificial intelligence provide new routes for big-data-based reservoir stress prediction.

1.3. Application of artificial intelligence in reservoir stress prediction

Artificial intelligence (AI) techniques have been increasingly applied to reservoir characterization and stress prediction. In practice, machine learning (ML) can be organized into supervised (e.g., KNN, SVM, decision trees, random forests, boosting), unsupervised (e.g., K-means, spectral clustering), and semi-supervised learning (e.g., label propagation). Compared with purely empirical formulas, these methods capture nonlinear interactions among logging features and stress responses, enabling rapid updates as new data accumulate.

Machine learning algorithms have deeply penetrated various fields of shale gas exploration and development and have shown significant advantages in reservoir stress prediction. Unlike traditional reservoir stress prediction methods that depend on downhole instrument measurements or complex geological modeling, machine learning offers a non-invasive, low-cost, and high-precision alternative. It can uncover the nonlinear relationships between rock mechanical parameters and reservoir stress fields through parameter mining, playing a significant role in guiding fracturing operations and fracture network inversion.

Lin et al. (2020) utilized 106 experimental data sets of sandstone and limestone, along with 23 field data sets, to develop a σ_H prediction method based on the Kriging machine learning model. This method estimates σ_H using drilling collapse geometric parameters and hole wall strength. In 2022, he compared the performance of 13 parametric and non-parametric machine learning models in predicting the minimum principal stress σ_h and found that the artificial neural network (ANN) model performed best on validation data, significantly improving upon traditional methods (Lin et al., 2022). Ibrahim et al. used data from two complex carbonate formation wells in the Middle East and applied random forests (RF), functional networks (FN), and adaptive neuro-fuzzy inference systems (ANFIS) to predict σ_h and σ_H (Ibrahim et al., 2021). Zhou et al. (2024) generated 2000 theoretical data sets based on triaxial test data and used a parameter identification method combining back propagation neural networks and exhaustive search algorithms to determine the constitutive parameters of geomaterial constitutive models. Rong et al., 2025 collected stress data from 66 measurement points in southwestern Guizhou and surrounding areas and proposed a hybrid model combining genetic algorithms (GA) and artificial neural networks (ANN) (GA-ANN) for predicting three-dimensional stress fields under complex geological conditions.

With algorithmic advances and increasing data availability, machine learning algorithms has improved the efficiency and accuracy of stress prediction in practice. That said, methods tailored to small datasets or narrow feature sets may scale less effectively when confronted with high dimensionality, strong multicollinearity, or wide inter-well variance—situations common in shale reservoirs.

In 2006, Hinton et al. (2006) introduced the deep belief network and overcame deep-model optimization barriers through layer-wise pre-training, catalyzing the deep-learning era. Traditional machine learning remains effective on small, structured datasets (typically $<10^4$) but is limited for nonlinear structure prediction when data scale to hundreds of thousands or millions and include unstructured components—conditions common in oilfield development. With powerful high-dimensional feature extraction, deep learning has thus gained wide attention in oil-and-gas parameter prediction.

For reservoir stress, representative studies highlight diverse deep architectures. Gao et al. (2020) built 877 FLAC3D cases and coupled 3D CNNs with capsule networks, adding feature enhancement and segmentation to improve 3D stress prediction. Wu et al. (2024) adopted GAN-based data generation to augment sample size/quality, followed by BPNN to learn the nonlinear mapping between wellbore-collapse geometry and σ_H. Lyu et al. (2024) generated data via model-driven high-precision pre-stack elastic-parameter inversion and proposed a CapsNet-BiLSTM hybrid that captures anisotropy (capsules) and temporal evolution (BiLSTM), with Mish activation improving nonlinear mapping and boosting 3D stress-field prediction in complex shale reservoirs. Collectively, these works show that deep learning is well suited to large-scale, unstructured data and multi-parameter coupling, and—through multi-source fusion and joint multi-stage prediction—can underpin more integrated geology-engineering workflows. Complementing these deep-learning advances, Zhang et al. (2025) demonstrate that a stacking ensemble can deliver continuous pore-pressure profiles from mud-log features alone, with Bayesian optimization of the Eaton index and end-to-end latency suitable for real-time or near–real-time analysis.

Despite the significant progress achieved in integrating deep learning with in-situ stress prediction, several methodological gaps remain to be addressed. First, many existing studies rely heavily on training datasets generated through numerical simulations (e.g., FLAC3D or finite element analysis), while real-world field data remain scarce. This limits the practical applicability and generalization capacity of the models. Second, the literature predominantly focuses on evaluating the performance of individual architectures (e.g., CNNs, RNNs, or their specific hybrids), lacking comprehensive comparative analyses across a broader range of deep learning frameworks, such as Transformers and graph neural networks. Third, there is currently no standardized or interpretable model selection criterion. Researchers often overlook inherent architectural limitations, such as the diminished directional sensitivity of CNNs or the vanishing gradient issue in RNNs. These unresolved challenges hinder the full-scale deployment of deep learning-based stress prediction in complex geological environments.

In this study, we developed a hybrid ensemble prediction network, AutoMix, based on the AutoGluon automated machine learning framework for in-situ stress prediction (σ_H and σ_h) in shale reservoirs. AutoMix integrates several mainstream machine learning algorithms—LightGBM, CatBoost, ExtraTrees, RandomForest, and XGBoost—along with deep learning models such as NeuralNetTorch and NeuralNetFastAI. The final model is constructed through AutoGluon's automated algorithm selection and hyperparameter optimization, followed by performance-based weighted ensemble learning. By combining the interpretability and non-linear fitting capabilities of tree-based models with the feature representation power of deep neural networks, Automix achieves not only higher predictive accuracy but also superior generalization and robustness.

To comprehensively assess the performance of Automix, we constructed a benchmark comparison framework based on the principles of model diversity and complementary mechanisms. Ten representative models were selected, including ElasticNet, LightGBM, XGBoost, SVR, RandomForest, ExtraTrees, CatBoost, MLP, TabNet, and FTTransformer. These cover a wide spectrum of modeling paradigms, including linear models, ensemble methods, neural networks, and attention-based architectures. All models were trained and evaluated on 278,003 well-logging samples collected from 14 horizontal shale gas wells in the Zigong region of Sichuan Province, each comprising 33 geomechanical features and corresponding σ_H/σ_h labels. Results demonstrate that AutoMix significantly outperforms traditional models under conditions of multicollinearity and small sample size, exhibiting enhanced prediction stability and cross-sample adaptability.

2. Model

2.1. Data preprocessing and multi-scale sampling strategy

This study utilized complete logging datasets from 14 deep shale gas horizontal wells, all free of missing values, to ensure data integrity and representativeness. Data preprocessing strategies were tailored according to the sensitivity of different models to feature scaling. Specifically, for scale-sensitive models such as SVR and MLP, we applied standard normalization using the StandardScaler; for scale-insensitive models such as LightGBM, CatBoost, and ExtraTrees, the raw features were used directly without normalization. The same preprocessing logic was consistently applied to the ensemble models developed in this work. Consistent with this design choice, Deng emphasize that min–max normalization (scaling features to [0, 1]) improves training accuracy and convergence for multivariable pore-pressure models by eliminating magnitude imbalance among inputs such as porosity, shale content, and acoustic velocity (Deng et al., 2024).

To ensure consistency across models under different sampling conditions and to enhance reproducibility, we used a fixed random seed (random_state = 42) in all experiments. Four sampling scales were defined—1/1000, 1/100, 1/10, and full dataset—to simulate varying data volume scenarios and evaluate model robustness under different conditions. In each case, 80% of the sampled data were used for training and the remaining 20% for validation, with no data leakage between the two sets.

Table 1
Hyperparameters of Zeroshot-HPO model library.

Model type	Hyperparameters	Value range	Explanation
Neural network	Activation function	ReLu/ELU	This model enhances its ability to capture complex nonlinear relationships by adjusting the activation function, number of hidden layers, and number of units. Dropout is used to prevent overfitting, and batch normalization improves training stability.
	Number of hidden layers	1–5	
	Hidden layer size	32–250	
	Dropout rate	0.01–0.39	
	Learning rate	0.0004–0.018	
	Batch normalization	True/False	
	Weight decay	1e−12–1e−3	
LightGBM	Learning rate	0.005–0.05	A low learning rate and a large number of leaves are adopted to enhance model expressiveness. Feature subsampling and minimum leaf sample constraints help prevent overfitting, while enabling ExtraTrees increases model diversity.
	Number of leaves	16–252	
	Feature fraction	0.43–0.97	
	Min samples per leaf	3–56	
	Extra trees	True/False	
CatBoost	Tree depth	4–8	Categorical features are supported. A balance between generalization and expressiveness is achieved by controlling tree depth and regularization parameters. Multiple splitting strategies are supported to improve flexibility.
	Learning rate	0.017–0.085	
	L2 regularization	1.0–4.8	
	Grow policy	Symmetric/Depthwise	
	Max categorical features	2–10	
XGBoost	Max depth	4–10	High-order feature interactions in high-dimensional data are modeled through deep tree structures, low learning rates, and feature subsampling. Minimum child weight is set to prevent overfitting. The model can flexibly handle both numerical and categorical features.
	Learning rate	0.007–0.093	
	Min child weight	0.6–1.4	
	Colsample by tree	0.5–0.98	
	Enable categorical	True/False	
Fast AI neural network	Hidden layers	[400,200]/[800,400]	A wide and deep network with flexible Dropout settings and extended training epochs is suitable for high-dimensional continuous features. Large-batch training is effective when sufficient memory is available, and a low learning rate ensures stable convergence.
	Epochs	20–50	
	Batch size	128–2048	
	Dropout probability	0.02–0.67	
	Learning rate	0.0005–0.095	
RandomForest	Max features	0.75/1.0/log 2	A wide and deep network with flexible dropout settings and extended training epochs is suitable for high-dimensional continuous features. Large-batch training is effective when sufficient memory is available, and a low learning rate ensures stable convergence.
	Max leaf nodes	12,000–48,000	
	Min samples per leaf criterion	1–40	
		Gini/Entropy/ Squared Error	

For example, under the 1/1000 sampling condition (approximately 278 samples randomly drawn from 278,004 total records), 222 samples were used to train a LightGBM model and the remaining 56 samples were used for evaluation. This process was uniformly applied to all models and sampling scales.

To ensure efficiency and consistency across training procedures, a maximum time limit of 14,400 s (4 h) was imposed for each model training run.

2.2. Model selection, hyperparameter configuration, and multi-level stacking architecture

In the task of predicting maximum and minimum horizontal stresses (σ_H and σ_h), the data present high dimensionality, strong nonlinearity, and pronounced multicollinearity. These characteristics demand models with robust learning capabilities and strong generalization performance. To address this, we employed the AutoGluon framework for independent modeling and training of σ_H and σ_h, specifying the task type as regression to accommodate the continuous nature of the stress values. AutoGluon, a state-of-the-art AutoML platform built on PyTorch and MXNet, incorporates a diverse set of machine learning and deep learning algorithms. It offers automated model selection, hyperparameter optimization, and feature transformation, and supports ensemble strategies such as WeightedEnsemble and multi-layer Stacking to construct high-performance predictive systems.

Our proposed hybrid model leverages the Zeroshot-HPO (Zero-Shot Hyperparameter Optimization) mechanism for hyperparameter configuration. This method utilizes the TabRepo knowledge base to retrieve historically optimized hyperparameter sets tailored to

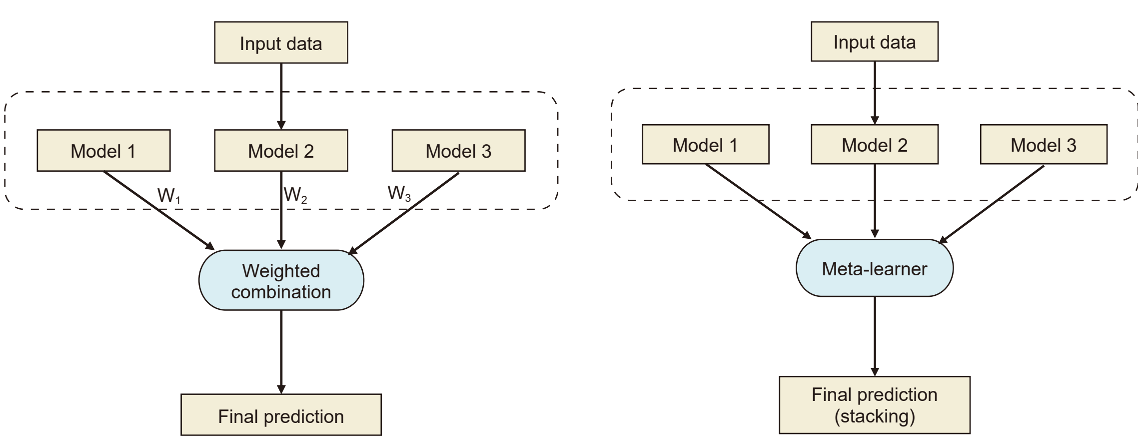

Fig. 1. Illustration of multi-stage stacking and weighted ensemble.

Table 2
Example of linear regression-based weighting using sub-model predictions.

Sample ID	Prediction by Model 1	Prediction by Model 2	Prediction by Model 3	True stress
Sample 1	A1	B1	C1	D1
Sample 2	A2	B2	C2	D2
Sample 3	A3	B3	C3	D3
Sample 4	A4	B4	C4	D4
Sample ….	A …	B …	C …	D …

Table 3
Feature range (min–max).

Feature	Unit	Range
Depth	m	2921.375–6799.75
Compressional slowness (Dt)	μs/ft	44.697–86.121
Brittleness index	%	30.126–71.774
Brittle minerals (quartz + feldspar)	%	0.764–82.704
Brittle minerals (quartz + feldspar + carbonates)	%	26.596–97.339
Shear slowness (St)	μs/ft	50.0–169.93
Stoneley slowness	μs/ft	180.0–317.308
Natural gamma ray	API	28.462–678.448
Permeability	nD	0.044–582.207
Poisson's ratio	/	0.001–0.33
Total gas content (pressure factor 2.0)	m^3/t	0.268–15.755
Total gas content (pressure factor 1.0)	m^3/t	0.192–10.48
Free gas (pressure factor 1.0)	m^3/t	0.001–7.478
Free gas (pressure factor 2.0)	m^3/t	0.001–12.805
Adsorbed gas (pressure factor 1.0)	m^3/t	0.035–3.883
Adsorbed gas (pressure factor 2.0)	m^3/t	0.035–4.549
Clay content	%	0.003–71.676
Vertical stress	MPa	71.299–111.066
Water saturation	%	2.409–100.0
Minimum principal stress	MPa	63.64–124.663
Maximum principal stress	MPa	72.961–139.835
Total organic carbon (TOC)	%	0.022–9.768
Kerogen content	%	0.036–13.084
Calcite content	%	0.093–76.288
Dolomite content	%	0.04–19.364
Feldspar content	%	0.076–9.728
Quartz content	%	0.688–74.434
Young's modulus	MPa	16344.346–81146.359
Sonic slowness	μs/ft	44.697–86.121
Average borehole diameter	in	8.062–14.839
Neutron porosity	p.u.	0.001–24.671
Density	g/cm^3	2.068–3.092
Potassium (K)	%	0.001–7.556
Photoelectric absorption cross section	b/e	1.854–29.083
Thorium (Th)	ppm	0.1–51.108
Uranium (U)	ppm	1.305–92.812

Table 4
Sub-model weights in AutoMix (sampling rate = 1/1000).

σ_h		σ_H	
Sub-model for σ_h prediction	Weight	Sub-model for σ_H prediction	Weight
NeuralNetFastAI_r103 _L1	0.6000	NeuralNetTorch_r12 _L1	0.4800
NeuralNetTorch_r89_L1	0.4000	NeuralNetFastAI_r18 _L1	0.5200

specific task types and data scales, thereby avoiding the computational overhead associated with traditional methods like grid search or Hyperband. Zeroshot ensures high efficiency and consistency across multiple training runs.

Under varying task and data size conditions, Zeroshot enables rapid extraction of the most appropriate configurations without the need for on-the-fly tuning. Finally, AutoGluon combines the predictions from multiple base models using multi-layer stacking and weighted ensembling, resulting in the AutoMix architecture that delivers optimal predictive performance (Table 1).

Fig. 1 illustrates the multi-stage stacking and weighted ensembling strategy adopted in the AutoMix framework. This hierarchical ensemble approach not only enhances predictive accuracy but also improves the model's ability to capture complex nonlinear relationships within geological data, thereby ensuring robust performance under high-dimensional feature conditions.

The model architecture is organized into three hierarchical levels:

Level 1 (base model layer): A diverse set of base models is trained, including gradient boosting algorithms (LightGBM, CatBoost), ensemble methods (RandomForest, ExtraTrees), and

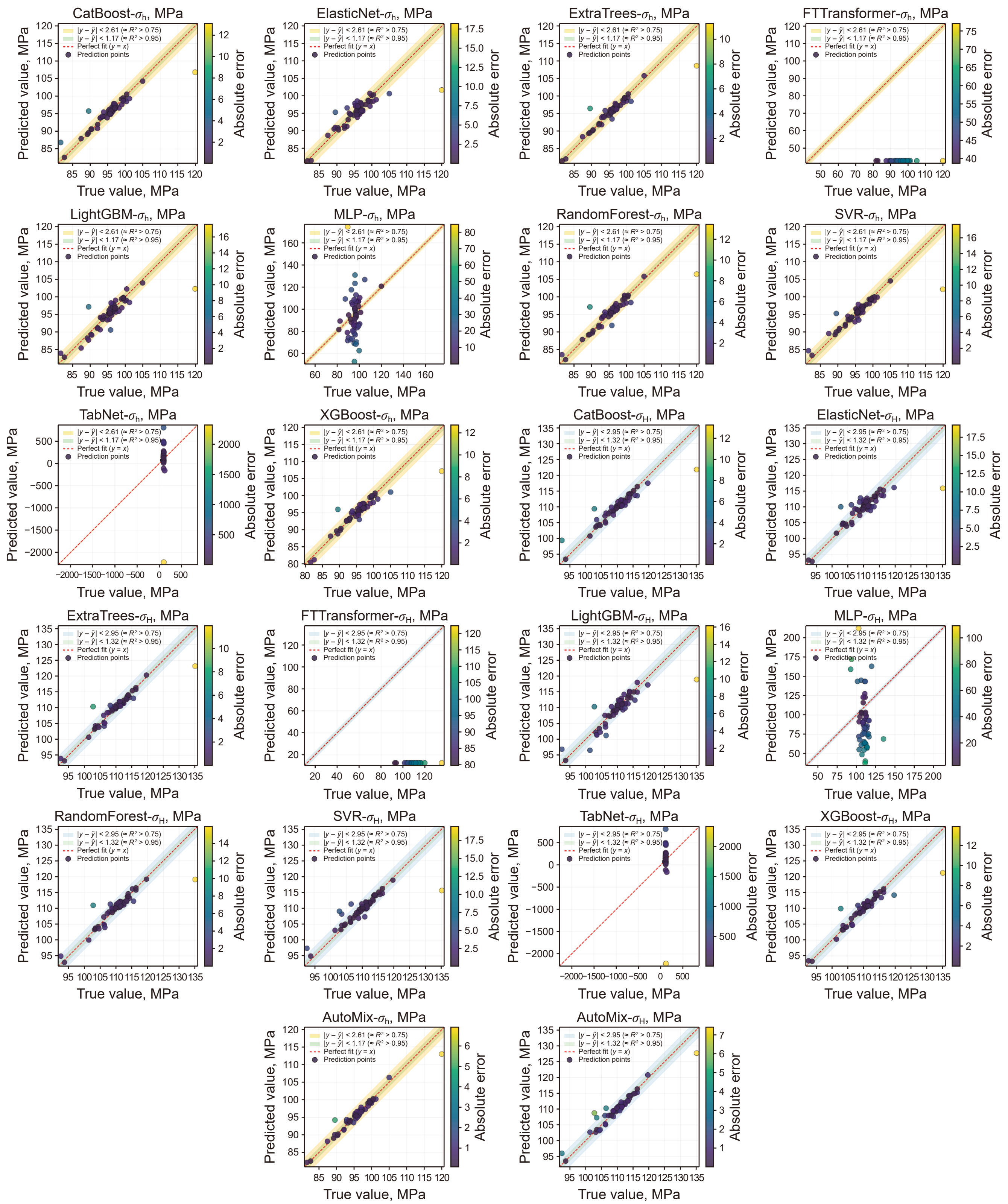

Fig. 2. Model prediction results under 1/1000 sampling ratio.

local learners (K-Nearest Neighbors). All models are configured using historical best hyperparameter sets retrieved from the Zeroshot library to ensure training efficiency and baseline performance.

Level 2 (intermediate stacking layer): High-performing Level 1 models are further combined through a first-stage stacking process. The outputs (predictions) of Level 1 models are used as meta-features and fed into Level 2 models. This nonlinear

Table 5
Performance metrics of all models under 1/1000 sampling ratio.

Model	σ_h			σ_H		
	MSE	MAE	R^2	MSE	MAE	R^2
CatBoost	4.79	1.01	0.82	5.73	1.19	0.84
ElasticNet	8.58	1.56	0.68	10.29	1.75	0.70
ExtraTrees	3.70	0.85	0.86	4.31	0.97	0.88
FTTransformer	2731.63	52.07	−134.19	9292.48	96.25	−317.60
LightGBM	8.46	1.49	0.69	8.69	1.73	0.75
MLP	355.92	14.24	−16.61	1577.69	33.17	−53.09
RandomForest	5.21	1.07	0.81	6.89	1.23	0.80
SVR	4.89	1.02	0.76	7.45	1.35	0.74
TabNet	122589.69	147.76	−4500.08	122095.92	143.37	−3506.50
XGBoost	4.69	1.09	0.83	5.68	1.12	0.84
AutoMix	1.61	0.70	0.94	2.86	1.00	0.92

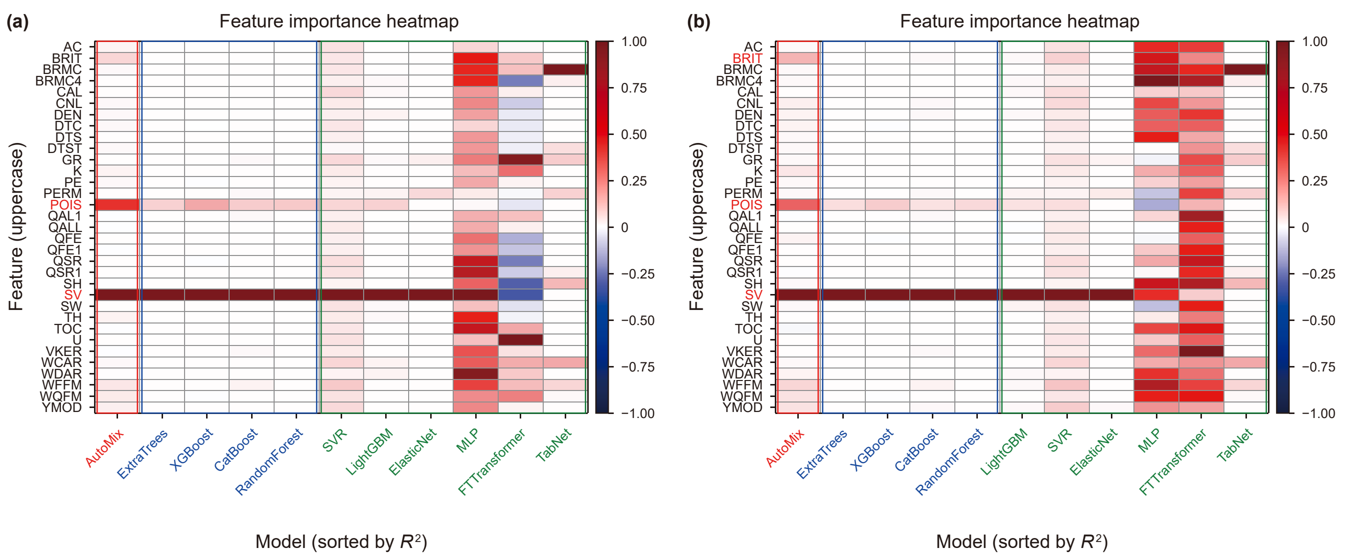

Fig. 3. Feature importance heatmaps under 1/1000 sampling ratio. **(a)** Influential features for minimum horizontal stress (σ_h); **(b)** Influential features for maximum horizontal stress (σ_H).

recombination helps suppress residual errors and improve predictive accuracy.

Level 3 (final ensemble layer): A selection of representative Level 2 models, along with a few high-performing Level 1 models, are aggregated through a linear weighted ensemble (WeightedEnsemble_L3). This final layer emphasizes complementary strengths across models and optimizes the overall prediction performance in terms of both accuracy and stability.

2.3. Benchmark model selection and ensemble strategy design

To establish a representative benchmarking framework, we selected ten diverse AI models spanning various learning paradigms and architectural types. The selection was based on the following key principles:

(1) Coverage of modeling paradigms: The chosen models encompass a wide spectrum of learning frameworks, including linear models, ensemble methods, deep neural networks, and attention-based architectures. This diversity ensures methodological complementarity and supports rigorous cross-paradigm comparisons.
(2) Scalability across sample sizes: Recognizing that well-logging datasets in engineering applications often exhibit significant variations in data volume, we evaluated all models under four sampling conditions (1/1000, 1/100, 1/10, and full dataset). Accordingly, the selected models are required to maintain robustness under small-sample conditions and fully leverage their learning capacity with large datasets—ensuring adaptability across varying computational budgets.
(3) High-dimensional and collinear feature handling: Given the high dimensionality (33 features) and notable multicollinearity (e.g., between depth and vertical stress) in our data, the models must possess adequate feature selection and regularization capabilities. This ensures effective identification of key variables and suppression of redundant or highly correlated features, ultimately enhancing model robustness and generalizability in complex feature spaces.

The final model set includes: ElasticNet (linear model), SVR, RandomForest, ExtraTrees (lightweight ensemble methods), LightGBM, XGBoost, CatBoost (boosting family), and three deep learning-based tabular models—MLP, TabNet, and FTTransformer. All baseline models were evaluated under default hyperparameter settings without additional tuning.

For the minimum principal stress (σ_h) prediction task, AutoGluon automatically trained and evaluated multiple algorithms, including tree models, ensemble models, and neural networks. It optimized

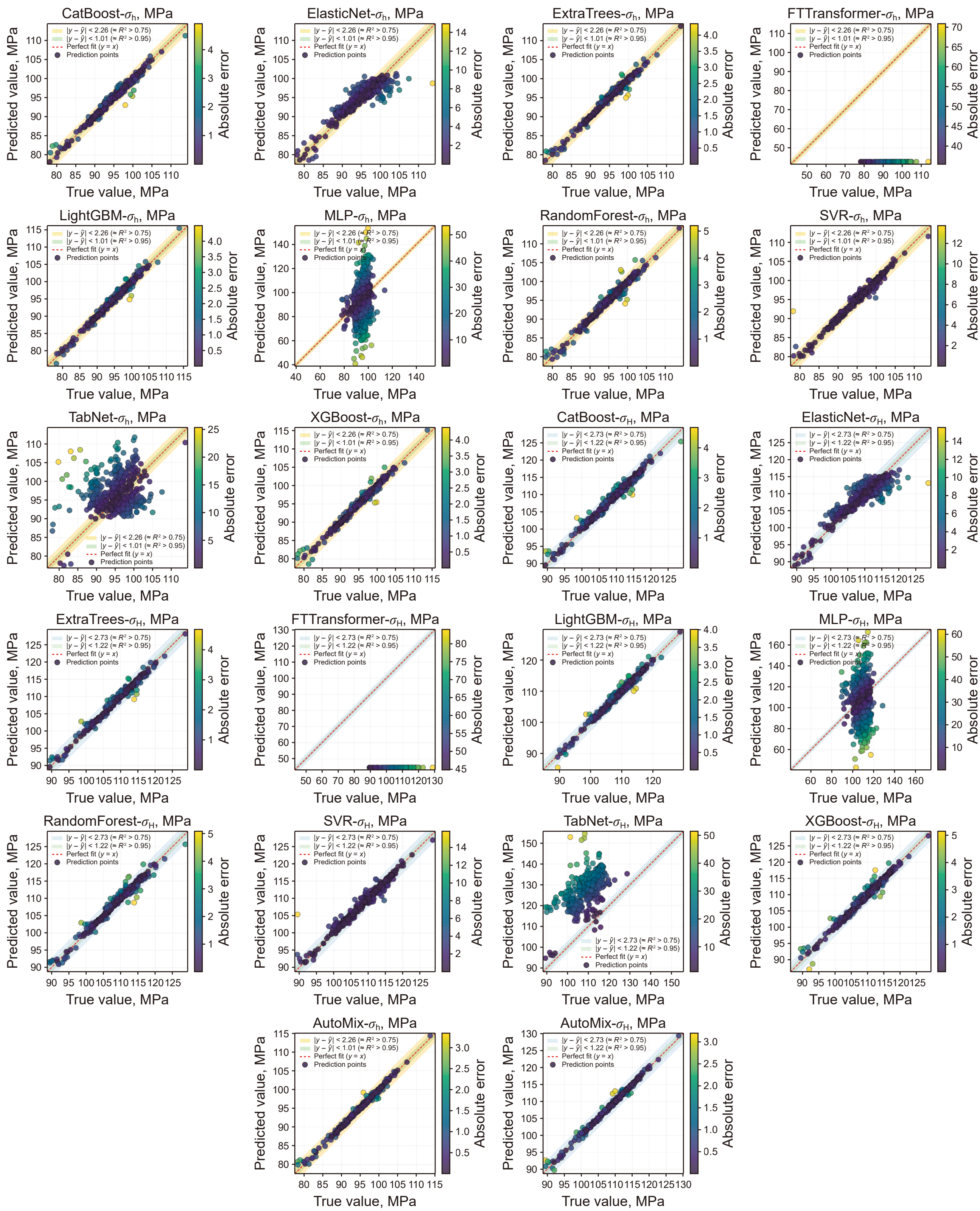

Fig. 4. Model prediction results under 1/100 sampling ratio.

the predictive model using a multi-layer stacked ensemble learning strategy and automatically selected and weighted the best hybrid model (LCER) via Weighted Ensemble. The final optimal model is a three-level weighted ensemble model, which includes LightGBM, CatBoost, ExtraTrees, and RandomForest.

The model hierarchy is as follows:

Level 1 (base level): Trained multiple categories of machine learning models, including gradient boosting (LightGBM, CatBoost), random forest (RandomForest), extremely randomized trees (ExtraTrees), and K-Nearest Neighbors (KNeighbors). For

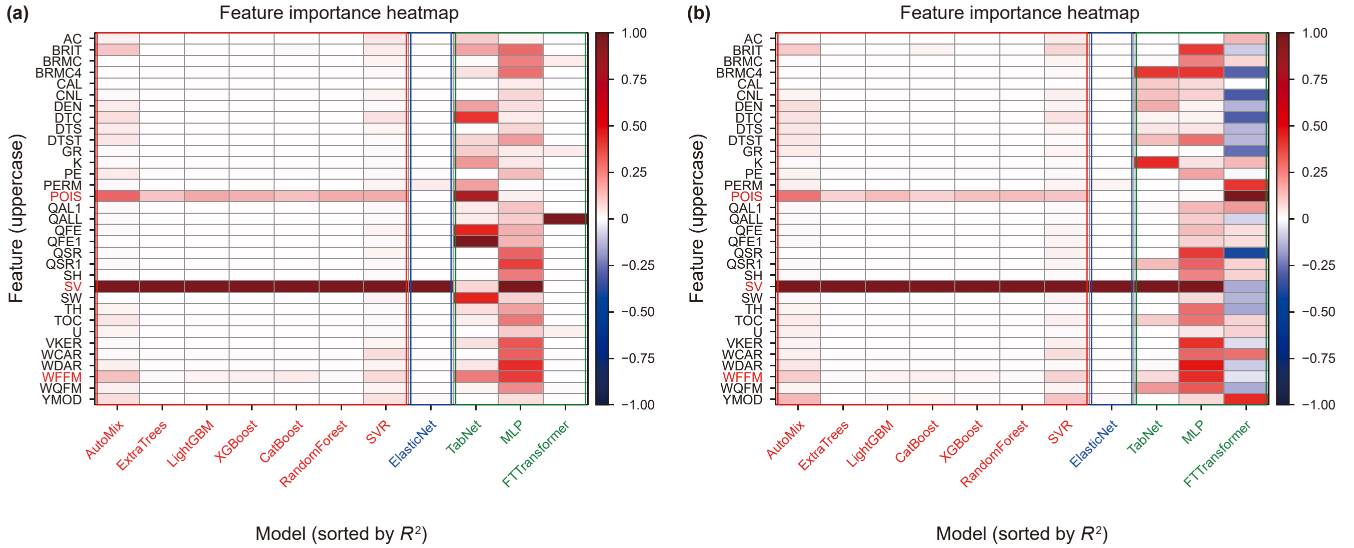

Fig. 5. Feature importance heatmaps under 1/100 sampling ratio. **(a)** Influential features for minimum horizontal stress (σ_h); **(b)** Influential features for maximum horizontal stress (σ_H).

each algorithm, multiple variants with different configurations were generated through hyperparameter search (e.g., RandomForest_r39, RandomForest_r195).

Level 2 (secondary ensemble): The high - performing Level 1 models were further integrated into a secondary model (WeightedEnsemble_L2). This layer automatically learned the contribution of different models through weighted averaging, significantly improving prediction robustness and reducing overfitting risk.

Level 3 (final ensemble): Based on Level 2, representative secondary models and a few high-performing models directly from Level 1 were selected to further build a tertiary ensemble (WeightedEnsemble_L3). This layer optimized overall prediction performance and achieved error complementarity among multiple models.

To guide model optimization, we adopt Mean Relative Error (MRE) as the core performance metric and the training loss because it normalizes errors across wells/formations with different stress magnitudes, aligns with percent-level engineering tolerances commonly used for σ_H/σ_h (e.g., whether deviations fall within safe operational windows), and avoids the high-magnitude bias that can affect MAE/RMSE when stress levels vary widely. The MRE is defined as:

$$\mathrm{MRE} = \frac{1}{n}\sum_{i=1}^{n}\left|\frac{y_i - \widehat{y}_i}{y_i}\right| \tag{1}$$

This formulation allows the model to focus on minimizing relative deviations, which is more relevant than absolute error in engineering contexts.

For deep learning models, we employed the ReLU activation function, defined as $\mathrm{ReLU}(x) = \max(0, x)$, due to its computational efficiency and ability to avoid vanishing gradients. Optimization was conducted using gradient descent algorithms such as Adam, with automatic learning rate and regularization tuning performed via the Zeroshot-HPO module.

2.4. Overfitting control and model contribution analysis

To enhance generalization, we applied at least five-fold cross-validation to all base models, ensuring stable performance under different training-validation splits. Independent cross-validation was conducted at each stacking level (L1, L2, L3) to maintain architectural independence and prevent information leakage across layers.

For overfitting control, we adopted a holdout strategy, strictly separating training and validation sets. This ensured that model evaluation was based solely on unseen data. In the final stacking layer (L3), the best-performing subset of models was automatically selected based on validation performance, thereby maximizing overall prediction accuracy.

To prevent data leakage, all feature normalization was performed exclusively on the training set. The validation set remained untouched by training statistics, ensuring an unbiased evaluation. Furthermore, no overlap occurred between training and validation samples at any point in the pipeline.

To quantify the contribution of each base model to the final ensemble, we introduced a linear regression-based weighting mechanism at the final integration stage. Specifically, the predictions of all candidate models were used as input features in a regression model, which was trained against the ground-truth stress values. The resulting regression coefficients were interpreted as the contribution weights of each sub-model in the AutoMix ensemble.

This approach not only facilitated effective output fusion but also provided a robust analytical framework for interpreting model importance across different sample sizes and experimental conditions (Table 2).

$$\widehat{D}_i = w_{Ai}A_i + w_{Bi}B_i + w_{Ci}C_i \tag{2}$$

where A, B, and C represent predictions from three different sub-models, while D denotes the true stress value for each sample. Linear regression is used to estimate the weight coefficients of each sub-model in the final ensemble. Here, w_i denotes the weight assigned to each sub-model within the ensemble. These weights are learned during training by minimizing the overall prediction error and can be interpreted as quantitative indicators of each model's contribution to the AutoMix ensemble.

3. Data analysis and geological interpretation

3.1. Geological background and reservoir characteristics of the study area

The Sichuan Basin is located on the northwestern flank of the Yangtze paraplatform and represents a secondary tectonic unit of

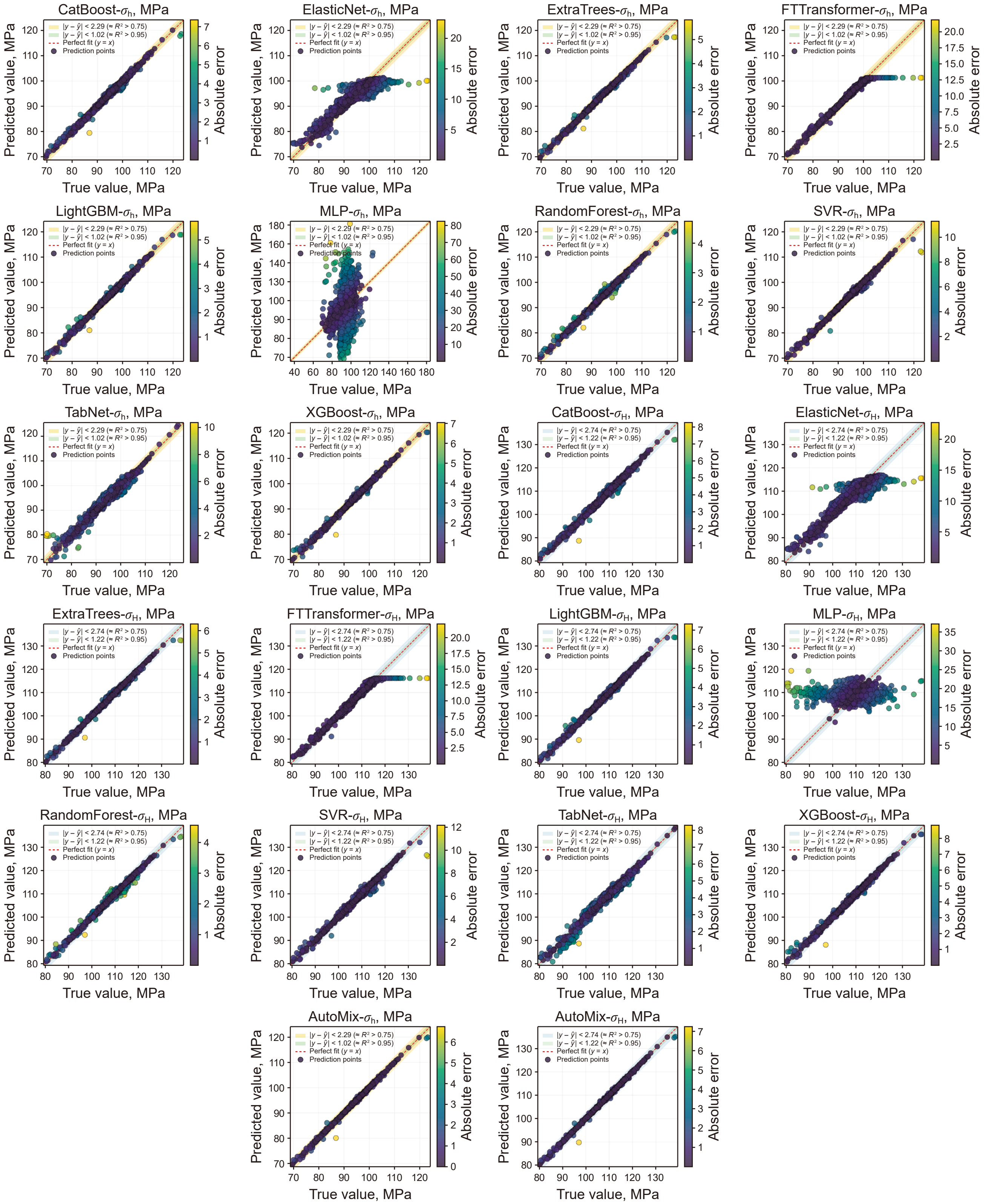

Fig. 6. Model prediction results under 1/10 sampling ratio.

this platform. The basin began to take shape during the Indosinian orogeny and subsequently developed its current structural features through intensive folding associated with the Himalayan movement. The tectonic framework of the region includes the Southeastern Sichuan fold-thrust belt, the Central Sichuan uplift belt, and the Northwestern Sichuan depression belt. The present study focuses on the Southwestern Sichuan low-amplitude fold belt, targeting the shale gas reservoir within the Longmaxi

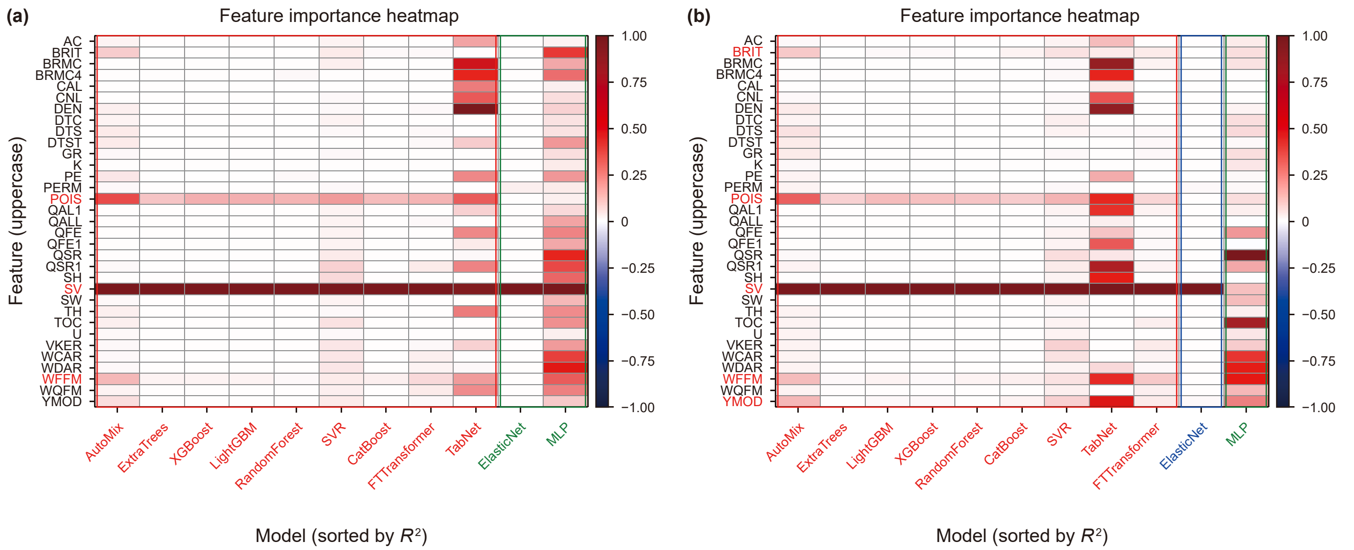

Fig. 7. Feature importance heatmaps under 1/10 sampling ratio. **(a)** Influential features for minimum horizontal stress (σ_h); **(b)** Influential features for maximum horizontal stress (σ_H).

Formation. The Longmaxi Formation is primarily composed of quartz and clay minerals, with minor amounts of carbonate minerals, plagioclase, and pyrite. Among the studied wells, Well Long_1^2 exhibits the highest quartz content, followed by Wells Long_1^3 and Long_1^1. In terms of clay mineral composition, illite is predominant, with subordinate illite/smectite mixed layers and minor amounts of chlorite and kaolinite. Reservoir lithofacies are categorized into six types based on core analysis, primarily comprising calcareous/silty shale, mudstone shale, and silty shale. In comparison, the Wufeng Formation displays four lithofacies types, mainly calcareous/silty shale, calcareous shale, and silty shale. The interval from Well Long_1^1 to the Wufeng Formation is characterized by a complex fracture network and dominant fracture connectivity, conducive to hydraulic fracturing. The middle-lower sublayers from Well Long_1^2 to Long_1^4 also exhibit complex fracture networks with high fracability. In contrast, the basal part of the Wufeng Formation and the overlying Baota Formation represent frac-barrier layers that typically generate only primary fractures and impede fracture propagation. In the study area, the maximum horizontal principal stress ranges from 80.51 to 117.23 MPa, while the minimum horizontal principal stress varies between 75.32 and 99.79 MPa. The differential horizontal stress ranges from 4.15 to 22.66 MPa, generally increasing with depth. The uranium-free gamma-ray (U-GR) values for the Longmaxi Formation range from 41 to 145 API and show a decreasing trend with depth. The total gamma-ray (GR) values vary from 78 to 250 API, remaining relatively stable with depth. Uranium content ranges between 2.8 and 23.0 ppm, also showing relative consistency with depth. Acoustic travel time (DT) ranges from 62 to 91 μs/ft and increases with depth. Neutron porosity decreases from 25.0% to 7.0%, while density varies from 2.38 to 2.75 g/cm^3 and remains relatively stable. Resistivity ranges from 3 to 95 Ω·m, showing an increasing trend with depth. In the Wufeng Formation, the U-GR ranges from 21 to 117 API, GR from 56 to 320 API, uranium content from 1.0 to 37.7 ppm, neutron porosity from 5.8% to 17.8%, density from 2.39 to 2.76 g/cm^3, and resistivity from 5 to 239 Ω·m.

3.2. Parameter selection

This study utilizes logging data from 14 horizontal wells with complete logging suites. The dataset includes 33 logging parameters and approximately 270,000 samples. The available measurements cover compressional wave slowness (DTC), brittleness index (BRIT), brittle minerals (BRMC: quartz + feldspar), extended brittle minerals (BRMC4: quartz + feldspar + carbonate), shear wave slowness (DTS), Stoneley wave slowness (DTST), natural gamma-ray (GR), permeability (PERM), Poisson's ratio (POIS), total gas content (QAL1, pressure coefficients 2.0 and 1.0), free gas (pressure coefficients 2.0 and 1.0), adsorbed gas (pressure coefficients 2.0 and 1.0), clay content, vertical stress, water saturation, total organic carbon (TOC), kerogen, calcite, dolomite, feldspar, quartz, Young's modulus, sonic slowness, average borehole diameter, neutron porosity, density, potassium, photoelectric factor (PE), thorium, and uranium (Table 3).

The dataset comprehensively represents rock mechanical properties, reservoir lithology, and petrophysical characteristics necessary for quantitative evaluation of shale gas reservoirs.

3.3. Evaluation metrics and experimental setup

A total of 278,003 samples were obtained from 33 logging curves across 14 horizontal shale gas wells and used to train and evaluate 11 different algorithms, with the data partitioned into 80% training and 20% testing sets. To simulate various data-scale scenarios, the samples were subsampled at full, 1/10, 1/100, and 1/1000 levels. K-Fold cross-validation strategy was applied to expand the effective training-validation dataset to nearly one million samples, thereby improving model generalization and prediction accuracy. Model performance was assessed using mean squared error (MSE), mean relative error (MRE), and the coefficient of determination (R^2). Feature importance was further computed for each sampling condition and linearly normalized to the range of [−1,1] to facilitate cross-model comparison.

$$\mathrm{MSE} = \frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2 \tag{3}$$

$$\mathrm{MAE} = \frac{1}{n}\sum_{i=1}^{n}|y_i - \hat{y}_i| \tag{4}$$

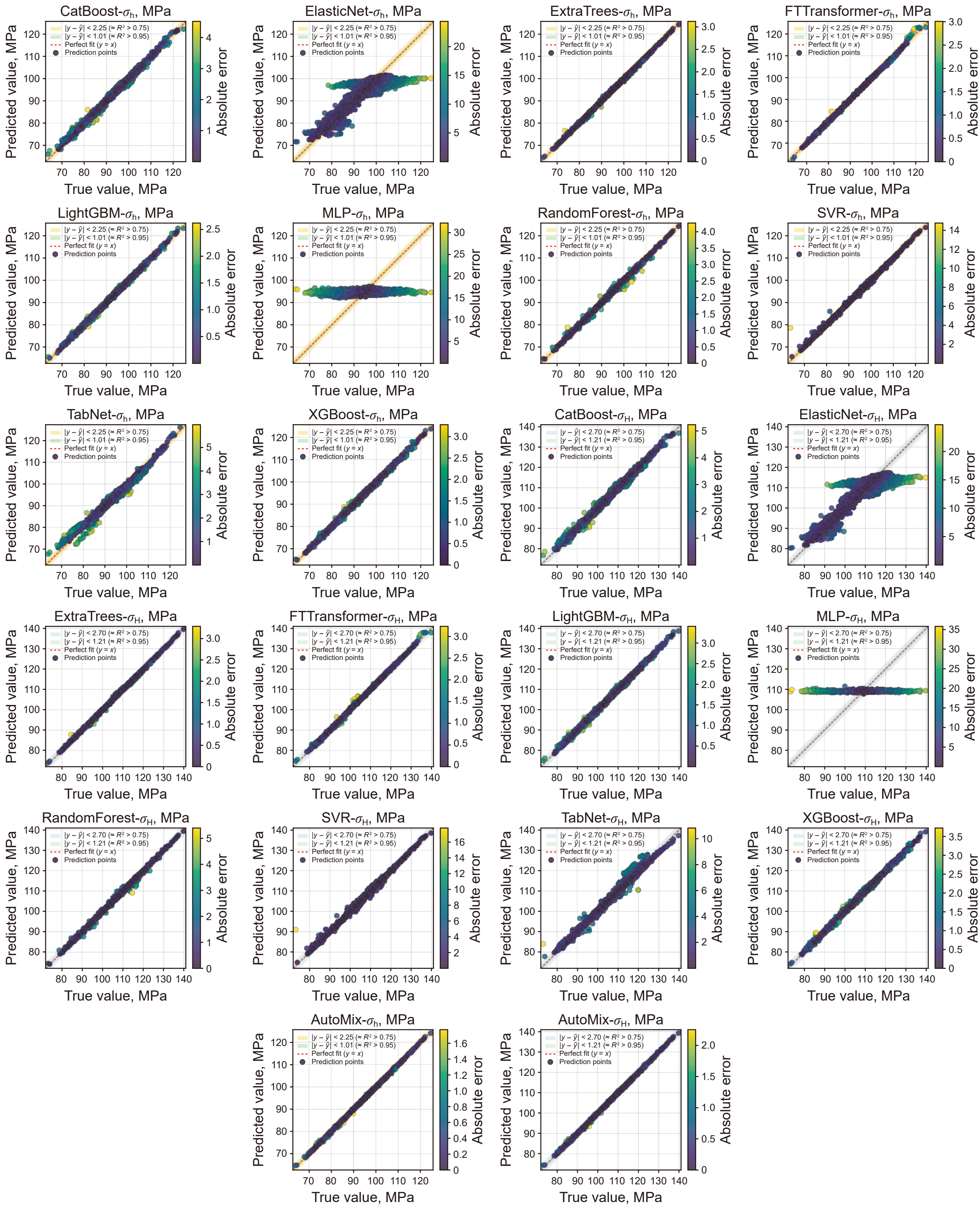

Fig. 8. Model prediction results under 1 sampling ratio.

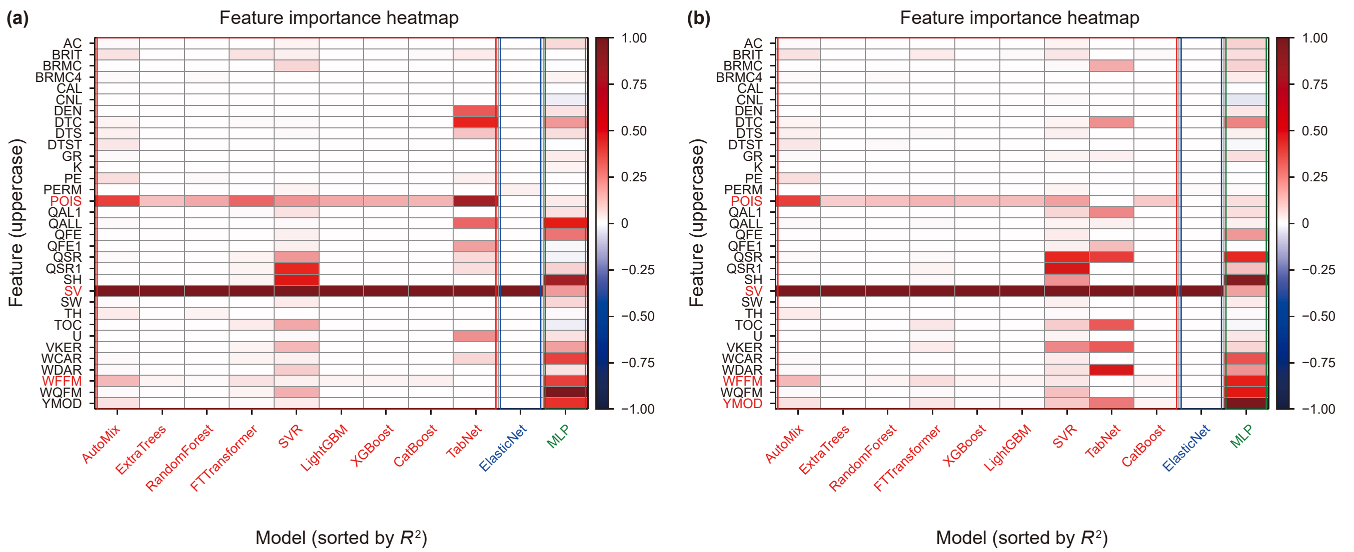

Fig. 9. Feature Importance heatmaps under 1 sampling ratio. **(a)** Influential features for minimum horizontal stress (σ_h); **(b)** Influential features for maximum horizontal stress (σ_H).

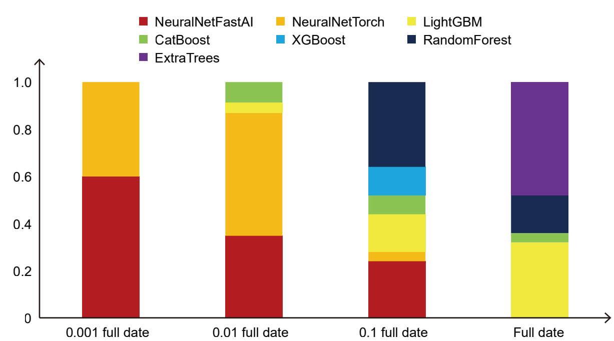

Fig. 10. AutoMix sub-model weight evolution for minimum stress (σ_h).

$$R^2 = 1 - \frac{\sum_{i=1}^{N} (y_i - \widehat{y}_i)^2}{(y_i - \overline{y})^2} \tag{5}$$

4. Results and discussion

To facilitate analysis of sub-models within the AutoMix ensemble, a standardized naming convention is adopted throughout this study. Each model name follows the format:

ModelType_HyperparameterID_StackLevel (NeuralNetFastAI_r103_L1).

The components are defined as follows:

ModelType (NeuralNetFastAI): Denotes the architecture of the model. AutoGluon supports a variety of base learners, such as LightGBM, CatBoost, NeuralNetTorch, and NeuralNetFastAI.

HyperparameterID (r103): Indicates the hyperparameter configuration used for this model. The identifier r103 refers to the 103rd configuration in AutoGluon's Zeroshot-HPO portfolio, derived from TabRepo's historical performance database.

If no r identifier is present (LightGBM_BAG_L1), the model uses default hyperparameters provided by AutoGluon.

StackLevel (L1, L2, L3): Refers to the model's position within the ensemble hierarchy. L1 indicates base models, L2 intermediate stacked models, and L3 the final weighted ensemble.

4.1. Model results and discussion with 1/1000 of full data (approx. 280 samples, testing set of about 50)

Under extremely low sample conditions, the ensemble primarily forms a two-level stacked structure. For both maximum and minimum principal stress predictions, NeuralNet-based models are dominant (Table 4).

As shown in Fig. 2, under low-sample conditions, individual models exhibit limited performance, with the highest R^2 reaching only 0.8763. At this scale, traditional machine learning models generally outperform deep learning models. Notably, the AutoMix ensemble maintains an R^2 above 0.9 even with minimal data, demonstrating robust predictive accuracy and stability (Table 5).

Fig. 3 presents the feature importance heatmaps for all models under the 1/1000 sampling scenario. Red frames indicate $R^2 \geq 0.9$, blue denotes $0.9 > R^2 \geq 0.8$, and green represents $R^2 < 0.8$. Among all models, only AutoMix reaches the red zone, indicating superior prediction performance. The heatmap results reveal that Poisson's ratio and vertical stress are the most influential features in predicting in-situ stress.

4.2. Model results and discussion with 1/100 of full data (approx. 2,800 samples, testing set of about 500)

Under medium sample size conditions (approximately 2800 samples), the prediction of minimum horizontal stress (σ_h) remains based on a two-level stacking structure, while the prediction of maximum horizontal stress (σ_H) evolves into a three-level ensemble. At this stage, machine learning and deep learning models contribute nearly equally within the AutoMix architecture, exhibiting a relatively balanced weight distribution (Table 6).

With the sample size increased to approximately 2,800, most machine learning models demonstrate strong predictive performance ($R^2 > 0.95$), making them practically applicable for engineering scenarios. In contrast, deep learning models generally perform less effectively at this stage. The single models including ExtraTrees, LightGBM, RandomForest, and CatBoost have already achieved high prediction accuracy in unimodal stress prediction (with R^2 approaching 0.98, indicating that only about 2% of the variance remains unexplained, and MAE below 0.43, indicating an

Table 6
Sub-model weights in AutoMix (sampling rate = 1/100).

σ_h		σ_H	
Sub-model for σ_h prediction	Weight	Sub-model for σ_H prediction	Weight
LightGBM _L1	0.0435	NeuralNetTorch_r22 _L1	0.2609
NeuralNetTorch_r22 _L1	0.3913	ExtraTreesMSE _L2	0.2609
CatBoost_r137 _L1	0.0870	NeuralNetFastAI _L2	0.0870
NeuralNetFastAI_r103 _L1	0.1304	LightGBMLarge_L2	0.1739
NeuralNetFastAI_r95 _L1	0.2174	LightGBM_r131 _L2	0.0870
NeuralNetTorch_r31 _L1	0.1304	NeuralNetTorch_r22 _L2	0.0435
		NeuralNetFastAI_r145 _L2	0.0870

Table 7
Performance metrics of all models under 1/100 sampling ratio.

Model	σ_h			σ_H		
	MSE	MAE	R^2	MSE	MAE	R^2
CatBoost	0.43	0.41	0.98	0.63	0.52	0.98
ElasticNet	3.91	1.40	0.81	5.79	1.80	0.81
ExtraTrees	0.33	0.32	0.98	0.46	0.40	0.98
FTTransformer	2740.64	52.16	−134.64	4178.88	64.42	−142.28
XGBoost	0.33	0.36	0.98	0.65	0.47	0.98
LightGBM	0.48	0.43	0.98	0.48	0.43	0.98
MLP	5.21	1.07	0.81	398.80	15.37	−12.67
RandomForest	0.70	0.50	0.98	0.70	0.50	0.98
SVR	1.17	0.55	0.96	1.17	0.55	0.96
TabNet	372.47	18.37	−11.53	372.47	18.37	−11.53
AutoMix	0.16	0.23	0.99	0.21	0.25	0.99

Table 8
Sub-model weights in AutoMix (sampling rate = 1/10).

σ_h		σ_H	
Sub-model for σ_h prediction	Weight	Sub-model for σ_H prediction	Weight
NeuralNetFastAI_r191 _L1	0.0800	NeuralNetFastAI_r191 _L1	0.0500
LightGBMXT _L2	0.0400	LightGBMXT_ L2	0.0500
LightGBM _L2	0.1200	LightGBM_ L2	0.2000
RandomForestMSE _L2	0.3600	RandomForestMSE _L2	0.2000
NeuralNetFastAI_r95 _L1	0.1200	CatBoost _L2	0.0500
NeuralNetTorch_r31 _L1	0.0400	ExtraTreesMSE _L2	0.1500
NeuralNetFastAI _L2	0.0400	NeuralNetFastAI_r95_L2	0.1000
XGBoost _L2	0.1200	NeuralNetFastAI_r2_L2	0.0500
CatBoost_r177 _L2	0.0800	CatBoost_r177 _L2	0.1500

Table 9
Performance metrics of all models under 1/10 sampling ratio.

Model	σ_h			σ_H		
	MSE	MAE	R^2	MSE	MAE	R^2
XGBoost	0.084	0.162	0.996	0.133	0.208	0.996
TabNet	0.895	0.643	0.957	0.789	0.652	0.974
SVR	0.166	0.162	0.992	0.304	0.253	0.990
RandomForest	0.106	0.153	0.995	0.143	0.183	0.995
MLP	230.147	11.626	−10.39	31.549	4.261	−0.082
LightGBM	0.088	0.167	0.996	0.130	0.211	0.996
FTTransformer	0.695	0.355	0.966	0.859	0.455	0.971
ExtraTrees	0.070	0.123	0.997	0.092	0.147	0.997
ElasticNet	4.214	1.418	0.799	5.909	1.783	0.803
CatBoost	0.196	0.287	0.991	0.300	0.364	0.990
AutoMix	0.032	0.082	0.998	0.038	0.093	0.999

average error of less than 0.43). These models can also serve as a reference for constructing in-situ stress fields (Table 7, Fig. 4).

The feature importance heatmaps reveal that seven models achieve an R^2 greater than 0.9 under the 1/100 sampling condition. Among these, more than half assign significant weight to the following features: vertical stress (SV), Poisson's ratio, and feldspar content (WFFM) (Fig. 5).

Table 10
Sub-model weights in AutoMix (sampling rate = 1).

σ_h		σ_H	
Sub-model for σ_h prediction	Weight	Sub-model for σ_H prediction	Weight
LightGBMXT_L2	0.0400	LightGBMXT_L2	0.05000
LightGBM _L2	0.2800	LightGBM_L2	0.35000
RandomForestMSE_L2	0.1600	RandomForestMSE_L2	0.55000
CatBoost_L2	0.0400	CatBoost_L2	0.05000
ExtraTreesMSE_L2	0.4800		

Table 11
Performance metrics of all models under 1 sampling ratio.

Model	σ_h			σ_H		
	MSE	MAE	R^2	MSE	MAE	R^2
XGBoost	0.0356	0.1229	0.9982	0.0536	0.1535	0.9982
TabNet	0.9944	0.8512	0.9508	0.2055	0.2759	0.9930
SVR	0.0199	0.0812	0.9990	0.0585	0.1312	0.9980
RandomForest	0.0155	0.0538	0.9992	0.0161	0.0578	0.9994
MLP	18.8820	3.2118	0.0655	28.8872	4.0943	0.0096
LightGBM	0.0352	0.1276	0.9983	0.0567	0.1638	0.9981
FTTransformer	0.0158	0.0891	0.9992	0.0180	0.0950	0.9994
ExtraTrees	0.0060	0.0420	0.9997	0.0069	0.0450	0.9998
ElasticNet	4.0259	1.4034	0.8008	5.8175	1.7891	0.8005
CatBoost	0.1339	0.2516	0.9934	0.2179	0.3216	0.9925
AutoMix	0.0046	0.0454	0.9998	0.0054	0.0483	0.9998

4.3. Model results and discussion with 1/10 of full data (approx. 28,000 samples, testing set of about 5,000)

Under medium-to-high sample size conditions (approximately 28,000 samples), both minimum and maximum stress prediction models evolve into three-level stacked ensembles. At this stage, machine learning models begin to dominate the AutoMix architecture (Table 8).

With 20,000 samples, most machine learning algorithms have achieved exceptionally high prediction accuracy, which is generally sufficient for in-situ stress field construction and hydraulic fracturing simulations. Deep learning models, such as TabNet and FTTransformer, have also shown improvements in predictive performance at this stage. However, the deep model FTTransformer performs relatively poorly in high-stress prediction (Table 9, Fig. 6).

In the feature importance heatmap for minimum horizontal stress prediction, Young's modulus (YMOD) emerges as a significant feature for the first time. This aligns with conventional geomechanical theory, suggesting that with increased sample size, the model is capable of automatically identifying physically meaningful and relevant parameters (Fig. 7).

4.4. Model results and discussion with full data (approx. 280,000 samples, testing set of about 50,000)

Under full-sample training conditions (approximately 280,000 data points), the number of sub-models in the AutoMix ensemble is significantly reduced, and all retained models are based on machine learning algorithms (Table 10).

With the full training dataset of approximately 200,000 samples, nearly all models achieve exceptionally high prediction accuracy, with R^2 values generally exceeding 0.99. This level of performance meets—and often exceeds—the precision requirements for engineering-grade stress prediction, highlighting the substantial benefits of increased data volume on model learning capacity and stability (Table 11, Fig. 8) .

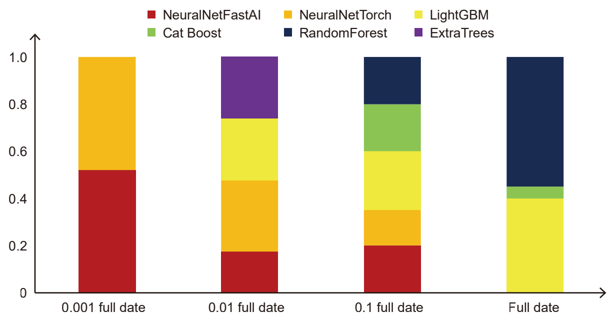

Fig. 11. AutoMix sub-model weight evolution for maximum stress (σ_H).

Under full-sample conditions, the feature importance heatmap indicates that Poisson's ratio, vertical stress (SV), feldspar content (WFFM), and Young's modulus (YMOD) are the most influential features for in-situ stress prediction. This outcome further confirms the AutoMix model's strong physical consistency and robust feature recognition capabilities in large-scale data environments (Fig. 9).

4.5. Summary of results

As the sample size increases, the weights assigned to deep learning models such as NeuralNetFastAI and NeuralNetTorch in the AutoMix ensemble gradually decline, eventually reaching zero under full-sample training. In contrast, the weights of machine learning models increase steadily, becoming dominant in large-data scenarios. Specifically, ExtraTrees emerges as the leading contributor to minimum horizontal stress (σ_h) prediction, while RandomForest plays a major role in predicting maximum horizontal stress (σ_H) (Figs. 10 and 11).

The AutoMix ensemble consistently achieves the lowest prediction error and highest stability across all sampling scales. Even with extremely limited data (e.g., 1/1000 sampling), it maintains high predictive accuracy. In comparison experiments, several standalone models—such as ExtraTrees, RandomForest, and XGBoost—also demonstrate excellent performance when the sample size exceeds 2000. These models reliably identify key features strongly correlated with maximum and minimum horizontal stress (σ_H and σ_h), including vertical stress, Poisson's ratio, feldspar content, and Young's modulus.

In geomechanics, interlayer principal stresses are commonly estimated using the Eaton empirical formula, expressed as (Iqbal et al., 2018):

$$\sigma_H = \frac{\nu}{1-\nu}(\sigma_v - \alpha P_p) + \alpha P_p + \frac{E\nu\varepsilon}{1-\nu^2} \tag{6}$$

$$\sigma_h = \frac{\nu}{1-\nu}(\sigma_v - \beta P_p) + \beta P_p + \frac{E\varepsilon}{1-\nu^2} \tag{7}$$

Here, ν denotes Poisson's ratio, σ_v represents vertical stress, P_p is formation pore pressure, and α and β are structural coefficients that describe the regional stress transformation, ε is strain factor.

The feature importance analysis consistently identifies the following well-logging parameters as highly influential across all sampling levels, underscoring their robust geomechanical relevance:

Vertical stress (σ_v): A fundamental input in stress field modeling, directly involved in stress computation, and strongly correlated with depth and rock density.

Poisson's ratio (ν) and Young's modulus (E): Key elastic parameters that govern rock deformation behavior. In the Eaton model, Poisson's ratio regulates the transformation of vertical stress into horizontal stress.

Feldspar content (WFFM): Reflects lithological characteristics and may indirectly influence the structural coefficients α and β, thereby affecting in-situ stress estimation.

It is worth noting that while Eqs. (6) and (7) are commonly used in field practice to reflect the relationship between Poisson's ratio, vertical stress, and horizontal stress, other models—such as the combined spring model—are also widely applied in shale formations. The selection of stress calculation method often depends on the geological conditions and engineering practices of specific blocks.

4.6. Potential limitations of the model and discussion on the reliability of in-situ stress interpretation

The samples used in this study are primarily derived from the lower section of the Wufeng–Longmaxi Formation in the Sichuan Basin. This stratigraphic interval represents a complete sedimentary cycle, characterized by deep-water shelf facies composed of organic-rich black shale, with relatively uniform geological background. However, the generalization capability of the constructed model may degrade significantly under the following scenarios:

Regions with significantly different lithology: If the training data do not include such lithological types, the model's generalization ability diminishes, leading to unstable prediction accuracy.

Structurally complex areas: In regions with dense faulting or well-developed folds, local stress fields and fracture orientations are highly variable. Due to drastic changes in logging responses and insufficient valid data in such zones, the model may fail to learn the underlying patterns, resulting in misjudgments or extrapolation errors.

Insufficient representativeness of training samples: With only 14 wells, the current dataset may not fully capture the geological variability of the entire study area, thereby limiting the extrapolation capacity of the model.

Taking the Qiongzhusi Formation as an example, this shale interval exhibits a classic thick, multilayered vertical structure. Layers 1, 3, 5, and 7 consist of organic-rich black shale, while layers 2, 4, 6, and 8 are composed of organic-lean silty shale. These alternating lithologies are vertically stacked and deposited in a distal marine fine-grained sedimentary environment, with frequent lithological variations. Due to data limitations, this study only conducted predictions on a few horizontal wells within the Qiongzhusi Formation and analyzed and corrected cases where model performance was unsatisfactory. The data preprocessing (including removal of outliers and blanks) followed the same approach as previously described. The prediction results using the Wufeng–Longmaxi-trained model are shown in Fig. 12.

When the model trained on the Wufeng–Longmaxi Formation was applied to the target well, noticeable deviations were observed between predicted and actual values. Possible reasons include geological complexity and differences in logging acquisition methods. To improve model performance, a regional retraining strategy was applied, using partial data from the target well to retrain the model. The improved results are shown in Figs. 13 and 14.

The retrained model achieved significantly better performance, with minimum principal stress prediction yielding MSE = 0.002503, MAE = 0.03465, R^2 = 0.9997, and maximum principal stress prediction with MSE = 0.00245, MAE = 0.03465, R^2 = 0.9997. These results validate the effectiveness of the regional retraining strategy.

Therefore, for basin datasets with significantly different lithology, structural features, or logging acquisition conditions, applying a regional retraining approach can significantly enhance the predictive capability of the AutoMix model. In addition, the following optimization strategies are also recommended:

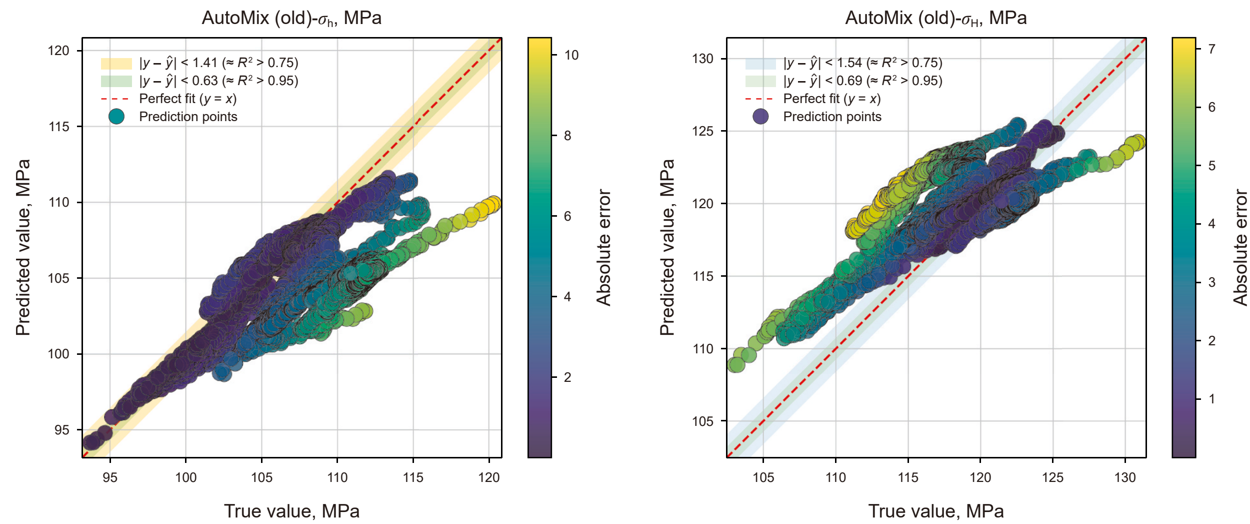

Fig. 12. Prediction performance on Qiongzhusi Formation using the Wufeng–Longmaxi model.

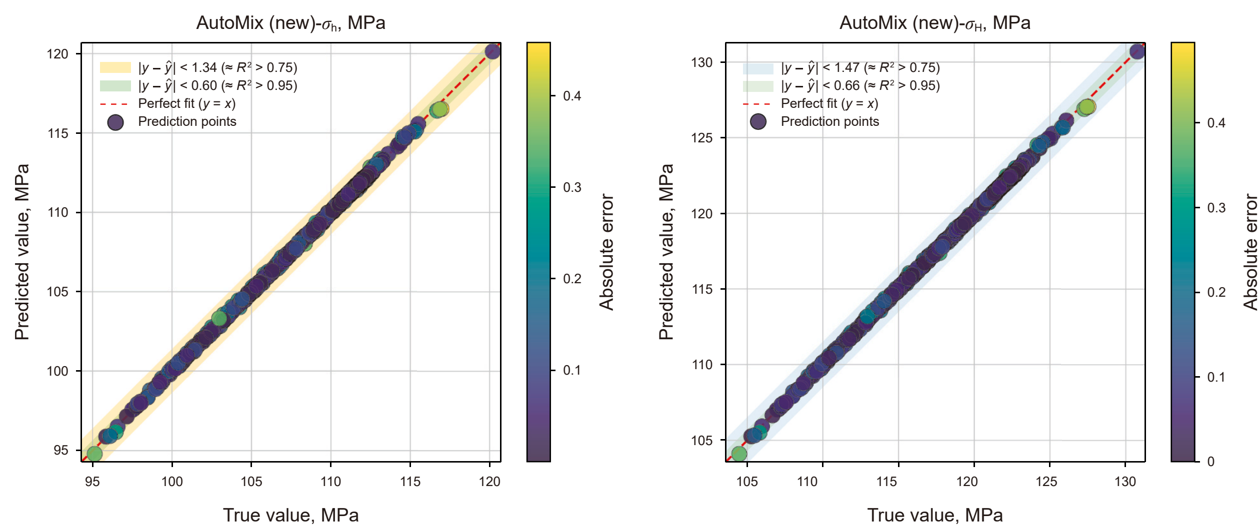

Fig. 13. Prediction performance after retraining on the target block.

Introduce categorical variables such as “geological block ID” and “lithology type” into the training data to improve the model's ability to recognize different structural units;

Expand the training sample coverage to include other tectonic blocks, and integrate structural interpretation results and lithofacies maps, enabling geological zoning-based logic regression or expert rule embedding;

Employ quantile models or Bayesian uncertainty modeling from AutoGluon to generate confidence interval outputs, thereby providing users with insight into the reliability of model predictions.

Furthermore, several wells with available DFIT (Diagnostic Fracture Injection Test) data were selected to compare the minimum horizontal stress derived from well log interpretation with that estimated from shut-in pressure decline analysis. Since each DFIT test typically corresponds to a single fracturing stage, we used the average minimum horizontal stress from log interpretation for the corresponding interval as a reference. The comparison results show that the log-interpreted minimum horizontal stress is typically about 7 MPa lower than the value obtained from DFIT analysis. This deviation suggests that even when model predictions align with the log-derived stress, the actual minimum in-situ stress may still be underestimated by approximately 7 MPa (Fig. 15, Table 12) .

4.7. Engineering implications for hydraulic fracturing

The fast and accurate predictions of σ_{H} and σ_{h} delivered in this study can be used as first-order inputs to fracturing design—informing stage placement and cluster spacing, fluid system and rate windows, as well as net-pressure targets—and enabling a pre-job assessment of fracability. In particular, the differential stress $\Delta\sigma = \sigma_{\mathrm{H}} - \sigma_{\mathrm{h}}$ directly constrains inter-stage/inter-layer stress contrast and fracture-deflection risk, and it governs cluster competition, shut-in timing, and feasible viscosity–rate combinations. Accordingly, we propose a practical “stress-to-design” mapping:

Large $\Delta\sigma$: shorten stage length and increase fluid viscosity/proppant strength to suppress deflection and stabilize primary-fracture conductivity;

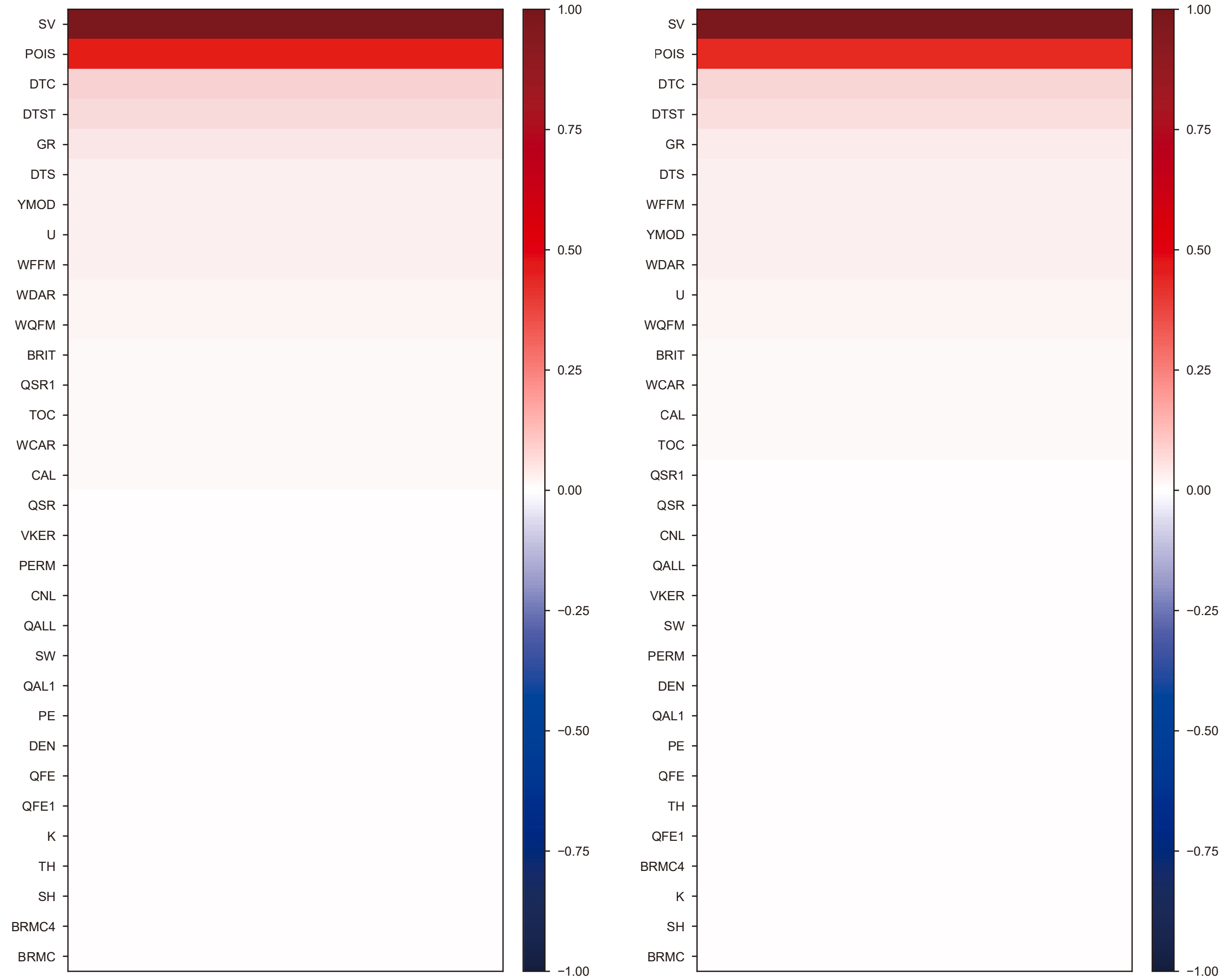

Fig. 14. Feature importance of the retrained model for the target block. **(a)** Influential features for minimum horizontal stress (σ_h); **(b)** Influential features for maximum horizontal stress (σ_H).

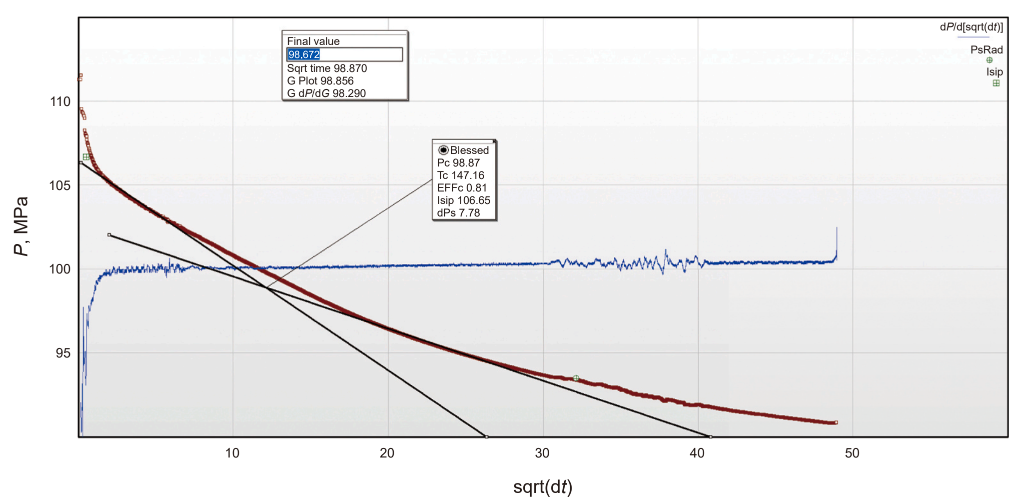

Fig. 15. Minimum principal stress estimated from shut-in pressure decline curve.

Small $\Delta\sigma$: moderately tighten cluster spacing to enhance fracture complexity and improve flow-capacity balance while managing competition.

Compared with numerical simulation and downhole tests, AutoMix—trained on 278,003 field logging samples with a stacked/weighted ensemble—achieves MSE $\approx$ 0.0046, MRE 0.454,

Table 12
In-situ stress estimates from different interpretation methods for selected wells.

Well	Instantaneous Shut-in Pressure (ISIP)	Depth, m	Average minimum horizontal stress ($\sqrt{t}$ Plot, G-function Plot, G-function Derivative), MPa	Average minimum horizontal stress from logging interpretation, MPa
H2	106.7	4009	98.7	91
H3	101.7	3973	95.1	88.95
H4	113.6	4116	104.1	97.36

$R^2 \approx 0.9998$ on the full dataset and maintains high accuracy even under 1/1000-sampling scarcity. This enables large-area, rapid, rolling generation of stress profiles/fields, substantially reducing pre-design cost and providing a stable foundation for downstream multi-disciplinary optimization (e.g., stage/cluster design and fluid-rate strategies).

In summary, continuous predictions of σ_H, σ_h, and $\Delta\sigma$ from AutoMix support an integrated geology-to-engineering workflow: use the stress field to bound stage/cluster layout, then iterate within a viscosity–rate–net-pressure window to balance robustness and operational economy.

Traditional in-situ stress logging interpretation primarily relies on formula-based calculations, where geological coefficients, such as the Biot coefficient, are difficult to determine when the number of wells is limited. However, it is well established that these coefficients are related to geological conditions. Our algorithm integrates well log geological parameters with mechanical parameters, bypassing the need for determining the Biot coefficient. This allows for relatively accurate in-situ stress interpretation across multiple wells, even when the number of wells and logging data is sparse. In the future, iterative analysis of in-situ stress data from multiple wells can be used to construct a 3D in-situ stress field, which will provide valuable guidance for subsequent fracturing design and optimization.

5. Conclusion

In response to the challenges posed by the complex geological conditions of deep shale reservoirs, the high-dimensional nonlinear characteristics of well logging data, and the limited adaptability of traditional methods, this study develops a multi-level hybrid prediction model named AutoMix, based on the AutoGluon automated ensemble learning framework. The model integrates mainstream algorithms including LightGBM, CatBoost, ExtraTrees, and RandomForest, and employs a multi-stage stacking and weighted ensemble strategy to significantly enhance the accuracy and stability of in-situ stress (σ_H and σ_h) prediction.

Using 278,003 sets of measured logging data from 14 horizontal shale gas wells in the Zigong area of the Sichuan Basin, a systematic comparison was conducted among 11 AI models under different sampling rates (1/1000, 1/100, 1/10, and full data). The results show that the AutoMix model consistently achieved the lowest error across all sample sizes (MSE = 0.0046, MRE = 0.454, $R^2 = 0.9998$), demonstrating excellent generalization ability and robustness to data variations.

Feature importance analysis revealed that vertical stress, Poisson's ratio, feldspar content, and Young's modulus have strong correlations with in-situ stress. These highly important features are not only prioritized by the statistical learning process of the model but also theoretically consistent with geomechanical stress calculation models, reflecting a strong alignment between model output and rock physics mechanisms.

In low-sample scenarios, deep learning algorithms accounted for a larger proportion of the AutoMix model's weight, while in high-sample cases, machine learning algorithms played a more dominant role. ExtraTrees emerged as the leading model in minimum principal stress prediction, whereas RandomForest was dominant in maximum principal stress prediction.

Moreover, the study found that traditional machine learning models such as ExtraTrees, RandomForest, XGBoost, and CatBoost maintained stable performance under medium to high sample sizes (>2000). This provides a practical reference for model selection in future engineering applications where computational resources are limited or data distribution is imbalanced.

When applied to regions with significantly different geological conditions, the model exhibits limited generalization ability. For example, applying the model trained on the Wufeng–Longmaxi Formation to the Qiongzhusi Formation resulted in a noticeable decline in prediction accuracy. Regional retraining effectively restored performance (MSE < 0.0025, $R^2 > 0.9997$), highlighting the importance of incorporating geological diversity in the training dataset.

The AutoMix model is therefore best suited for geologically homogeneous formations. For structurally complex areas, regional retraining is recommended, along with the inclusion of categorical features such as “geological block ID” and “lithology type” to improve adaptability. Incorporating quantile regression or Bayesian methods for uncertainty modeling can further enhance prediction reliability in practical applications.

In conclusion, the proposed AutoMix hybrid model not only overcomes the adaptability limitations of traditional stress prediction methods in deep heterogeneous reservoirs but also delivers excellent performance under multiple constraints such as data scarcity, strong feature collinearity, and high precision requirements. It demonstrates promising potential for practical applications.

CRediT authorship contribution statement

Hang Yuan: Writing – review & editing, Writing – original draft, Visualization, Validation, Supervision, Software, Resources, Project administration, Methodology, Investigation, Funding acquisition, Formal analysis, Data curation, Conceptualization. **Yu-Ping Sun:** Funding acquisition. **Xiong Wei:** Conceptualization. **Wen-Te Niu:** Writing – review & editing. **Qian Cheng:** Writing – review & editing.

Funding

This research was funded by the China National Petroleum Corporation (CNPC)'s Basic and Foresight-Oriented Science and Technology Program, project “Mechanisms of Shale Oil and Gas Development and Volumetric Development Technology” (Grant No. 2023ZZ08).

Declaration of interests

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

The authors gratefully acknowledge the support of the National Energy Shale Gas R&D Center for this work.

Parameter description

AC	Sonic transit time, μs/ft
BRIT	Britteness index, %
BRMC	Briterness mineral (quartz + feldspar), %
BRMC4	Briterness mineral (quartz + feldspar + carbonate), %
CAL	Caliper (average borehole diameter), in
CNL	Neutron porosity, p.u.
DEN	Density, g/cm^3
DEPTH	Depth, m
DTC	Compressional sonic travel time, μs/ft
DTS	Shear sonic travel time, μs/ft
DTST	Stoneley wave transit time, μs/ft
GR	Gamma ray, API
K	Potassium, %
PE	Photoelectric absorption index, b/e
PERM	Permeability, mD
POIS	Poisson's ratio
QAL1	Total gas content (pressure coefficient 2.0), m^3/t
QALL	Total gas content (pressure coefficient 1.0), m^3/t
QFE	Free gas (pressure coefficient 1.0), m^3/t
QFE1	Free gas (pressure coefficient 2.0), m^3/t
QSR	Adsorbed gas (pressure coefficient 1.0), m^3/t
QSR1	Adsorbed gas (pressure coefficient 2.0), m^3/t
SH	Shale content, %
SV	Vertical stress, MPa
SW	Water saturation, %
TH	Thorium, ppm
TOC	Total organic carbon, %
U	Uranium, ppm
VKER	Kerogen content, %
WCAR	Calcite content, %
WDAR	Dolomite content, %
WFFM	Feldspar content, %
WQFM	Quartz content, %
YMOD	Young's modulus, MPa

References

Cao, H., Zhao, Y., Chen, H.C., Zhang, L.L., Xian, C.G., Yang, J.D., Liu, L., 2025. Integrating well logs, 3D seismic, and earthquake data for comprehensive prediction of 3D in-situ stress orientations: a case study from the Weiyuan area in the Sichuan Basin, China. Pet. Sci. 22 (1), 210–221. https://doi.org/10.1016/j.petsci.2024.07.015.

Deng, S., Pan, H.Y., Wang, H.G., Xu, S.K., Yan, X.P., Li, C.W., Zhao, F., 2024. A hybrid machine learning optimization algorithm for multivariable pore pressure prediction. Pet. Sci. 21 (1), 535–550. https://doi.org/10.1016/j.petsci.2023.09.001.

Dinmohammadpour, M., Nikkhah, M., Goshtasbi, K., Ahangari, K., 2022. Application of the Kaiser Effect in in-situ stress measurement in rocks – an overview. Rudarsko-Geolosko-Naftni Zb. 37 (4), 1–16. https://doi.org/10.17794/rgn.2022.4.1.

Gao, W., Lu, X., Peng, Y., Wu, L., 2020. A deep learning approach replacing the finite difference method for in situ stress prediction. IEEE Access 8, 44063–44074. https://doi.org/10.1109/access.2020.2977880.

Haimson, B., Fairhurst, C., 1969. In-situ stress determination at great depth by means of hydraulic fracturing. In: ARMA US Rock Mechanics/Geomechanics Symposium. ARMA. ARMA-69.

Haimson, B.C., Lee, M., Herrick, C., 1991. Recent advances in in-situ stress measurements by hydraulic fracturing and borehole breakouts. In: Proc. 7th Int. Congress of Rock Mechanics. Aachen, Germany.

Hinton, G.E., Osindero, S., Teh, Y.W., 2006. A fast learning algorithm for deep belief nets. Neural Comput. 18 (7), 1527–1554. https://doi.org/10.1162/neco.2006.18.7.1527.

Ibrahim, A.F., Gowida, A., Ali, A., Elkatatny, S., 2021. Machine learning application to predict in-situ stresses from logging data. Sci. Rep. 11 (1), 23445. https://doi.org/10.1038/s41598-021-02959-9.

Iqbal, O., Ahmad, M., Abd Kadir, A., 2018. Effective evaluation of shale gas reservoirs by means of an integrated approach to petrophysics and geomechanics for the optimization of hydraulic fracturing: a case study of the Permian Roseneath and Murteree Shale Gas reservoirs, Cooper Basin, Australia. J. Nat. Gas Sci. Eng. 58, 34–58. https://doi.org/10.1016/j.jngse.2018.07.017.

Karatela, E., Taheri, A., Xu, C., Stevenson, G., 2016. Study on effect of in-situ stress ratio and discontinuities orientation on borehole stability in heavily fractured rocks using discrete element method. J. Pet. Sci. Eng. 139, 94–103. https://doi.org/10.1016/j.petrol.2015.12.016.

Kumari, W.G.P., Ranjith, P.G., Perera, M.S.A., Shao, S., Chen, B.K., Lashin, A., Rathnaweera, T.D., 2017. Mechanical behaviour of Australian Strathbogie granite under in-situ stress and temperature conditions: an application to geothermal energy extraction. Geothermics 65, 44–59. https://doi.org/10.1016/j.geothermics.2016.07.002.

Lai, J., Li, D., Wang, G., Xiao, C., Hao, X., Luo, Q., Qin, Z., 2019. Earth stress and reservoir quality evaluation in high and steep structure: the Lower Cretaceous in the Kuqa Depression, Tarim Basin, China. Mar. Petrol. Geol. 101, 43–54. https://doi.org/10.1016/j.marpetgeo.2018.11.036.

Lee, H., Ong, S.H., 2018. Estimation of in situ stresses with hydro-fracturing tests and a statistical method. Rock Mech. Rock Eng. 51, 779–799. https://doi.org/10.1007/s00603-017-1349-1.

Lei, Q., Xu, Y., Cai, B., Gong, S., Xin, W., Bi, G., Tao, L., 2020. Progress and prospects of horizontal well fracturing technology for shale oil and gas reservoirs. Pet. Explor. Dev. 49 (1), 191–199. https://doi.org/10.1016/s1876-3804(22)60015-6.

Li, G., Mizuta, Y., Ishida, T., Li, H., Nakama, S., Sato, T., 2009. Stress field determination from local stress measurements by numerical modelling. Int. J. Rock Mech. Min. Sci. 46 (1), 138–147. https://doi.org/10.1016/j.ijrmms.2008.07.009.

Li, P., Cai, M., Gorjian, M., Ren, F., Xi, X., Wang, P., 2023. Interaction between in-situ stress states and tectonic faults: a comment. Int. J. Miner., Metall. Mater. 30 (7), 1227–1243. https://doi.org/10.1007/s12613-023-2607-8.

Lin, H., Oh, J., Masoumi, H., Canbulat, I., Zhang, C., Dou, L., 2018. A Review of in Situ Stress Measurement Techniques.

Lin, H., Kang, W.H., Oh, J., Canbulat, I., 2020. Estimation of in-situ maximum horizontal principal stress magnitudes from borehole breakout data using machine learning. Int. J. Rock Mech. Min. Sci. 126, 104199. https://doi.org/10.1016/j.ijrmms.2019.104199.

Lin, H., Singh, S.K., Xiang, Z., Kang, W.H., Raval, S., Oh, J., Canbulat, I., 2022. An investigation of machine learning techniques to estimate minimum horizontal stress magnitude from borehole breakout. Int. J. Mining Sci. Technol. 32 (5), 1021–1029. https://doi.org/10.1016/j.ijmst.2022.06.005.

Ljunggren, C., Chang, Y., Janson, T., Christiansson, R., 2003. An overview of rock stress measurement methods. Int. J. Rock Mech. Min. Sci. 40 (7-8), 975–989. https://doi.org/10.1016/j.ijrmms.2003.07.003.

Lyu, F., Liu, J., Chen, L., Tao, B., Liu, X., 2024. 3D in-situ stress prediction for shale reservoirs based on the CapsNet-BiLSTM hybrid model. Int. J. Rock Mech. Min. Sci. 183, 105937. https://doi.org/10.1016/j.ijrmms.2024.105937.

Pan, X., Liu, Z., Wang, P., Zheng, Y., Li, L., Wang, X., Liu, J., 2022. Estimation of in situ stresses from PP-wave azimuthal seismic data in fracture-induced anisotropic media. Geophysics 87 (6), C139–C154. https://doi.org/10.1190/geo2022-0175.1.

Pitcher, T.M., 2015. Joint Geophysical and Geomechanical Analysis of in-situ Stress, Wattenberg Field, Colorado, USA. Colorado School of Mines. https://doi.org/10.1190/geo2022-0175.1. M.Sc. thesis.

Robertson, P.K., 1986. In situ testing and its application to foundation engineering. Can. Geotech. J. 23 (4), 573–594. https://doi.org/10.1139/t86-086.

Rong, P., Zuo, Y., Lin, J., Zheng, L., Pan, C., Sun, W., Chen, B., 2025. In situ stress prediction model in complex geology: a hybrid GA-ANN with nonlinear boundary condition. J. Rock Mech. Geotech. Eng. 17 (7), 4349–4366. https://doi.org/10.1016/j.jrmge.2024.11.003.

Sanei, M., Ramezanzadeh, A., Delavar, M.R., 2023. Applied machine learning-based models for predicting the geomechanical parameters using logging data. J. Pet. Explor. Prod. Technol. 13 (12), 2363–2385. https://doi.org/10.1007/s13202-023-01687-2.

Seto, M., Utagawa, M., Katsuyama, K., Nag, D.K., Vutukuri, V.S., 1997. In situ stress determination by acoustic emission technique. Int. J. Rock Mech. Min. Sci. 34 (3–4), 281.e1-281.e16281, https:doi.org/10.1016/S1365-1609(97)00156-1.

Singha, D.K., Chatterjee, R., 2015. Geomechanical modeling using finite element method for prediction of in-situ stress in Krishna–Godavari basin, India. Int. J. Rock Mech. Min. Sci. 73, 15–27. https://doi.org/10.1016/j.ijrmms.2014.10.003.

Sun, R., Wang, J., 2024. Effects of in situ stress and multiborehole cluster on hydraulic fracturing of shale gas reservoir from multiscale perspective. J. Energy Eng. 150 (2), 04024002. https://doi.org/10.1061/jleed9.eyeng-5226.

Teufel, L.W., 1983. Determination of in-situ stress from anelastic strain recovery measurements of oriented core. In: SPE Rocky Mountain Petroleum Tech. Conf. (SPE-11649). https://doi.org/10.2523/11649-ms.

Wang, H., Ma, F., Tong, X., Li, Z., Wu, Z., Li, D., Yang, L., 2016. Assessment of global unconventional oil and gas resources. Pet. Explor. Dev. 43 (6), 925–940. https://doi.org/10.1016/b978-0-12-802238-2.00002-x.

Wu, B., Zhang, H., Wu, S., Nie, Y., Zhang, X., Jeffrey, R.G., 2024. Prediction of the maximum horizontal principal stress from breakout data using generative adversarial networks and backpropagation neural network. SPE J. 29 (2), 606–619. https://doi.org/10.2118/217970-pa.

Zhang, Y., Yin, S., Zhang, J., 2021. In situ stress prediction in subsurface rocks: an overview and a new method. Geofluids 2021, 6639793. https://doi.org/10.1155/2021/6639793.
Zhang, D.Y., Ma, T.S., Liu, Y., Zhang, D.C., 2025. A stacking ensemble approach for pore pressure prediction in real-time during drilling based on mud log data. Pet. Sci. 22 (12), 5047-5067. https://doi.org/10.1016/j.petsci.2025.09.020.
Zhao, T., Qin, Q., 2023. Characterization methods for current in-situ stress in oil and gas reservoirs: a mini review. Front. Earth Sci. 11, 1276807. https://doi.org/10.3389/feart.2023.1276807.
Zhou, C., Gao, B., Yan, B., Zhu, W., Ye, G., 2024. A combined machine learning/search algorithm-based method for the identification of constitutive parameters from laboratory tests and in-situ tests. Comput. Geotech. 170, 106268. https://doi.org/10.1016/j.compgeo.2024.106268Getrightsandcontent.
Zou, C., Wu, S., Yang, Z., Pan, S., Wang, G., Guan, M., Shen, Y., 2023. Progress, challenge and significance of building a carbon industry system in the context of carbon neutrality strategy. Pet. Explor. Dev. 50 (1), 210–228. https://doi.org/10.1016/s1876-3804(22)60382-3.