

Article

A Critical Evaluation of Kolmogorov–Arnold Networks for Shear-Wave Slowness Prediction from Well Logs

Zhuoqing Li and Longqiang Han *

Faculty of Land Resource Engineering, Kunming University of Science and Technology, Kunming 650093, China

* Correspondence: lqhan2014@163.com

Abstract

Kolmogorov–Arnold networks (KANs) are increasingly applied to well-log and reservoir problems, but their evaluation protocols rarely control preprocessing leakage, replicate comparators across seeds, or test depth-shifted generalization. This study evaluates a KAN for predicting shear-wave slowness from seven conventional logs and two engineered proxies on a public, depth-indexed dataset of 1904 records, benchmarked against six model families under a leakage-free protocol with five-seed replication. On the random split, KAN is competitive ($R^2 \approx 0.99$) but not the best model: gradient-boosted ensembles and a multilayer perceptron score higher. Once adjacent-sample leakage is removed by blocked depth cross-validation, every family degrades sharply, and in a shallow-to-deep test with matched preprocessing, hyperparameters, and seeds, KAN loses most of its accuracy (an 8-fold RMSE increase) while compact symbolic regression retains the highest mean accuracy. Edge-level symbolic templates fit individual KAN edges well, but no template identity is stable across seeds, so edge fits serve as diagnostics rather than recovered physical laws. These results separate interpolation accuracy from depth generalization and indicate the evidence—multiwell blocked validation and calibrated uncertainty—required before KAN-based log reconstruction can be considered reliable.

Keywords: shear-wave slowness; sonic log reconstruction; well logging; Kolmogorov–Arnold network; critical evaluation; depth extrapolation; rock physics

1. Introduction

Shear-wave information supports elastic-modulus estimation, seismic interpretation, borehole-stability assessment, and geomechanical modeling [1,2]. In practice, a dipole shear-sonic log may be unavailable because of acquisition cost, tool limitations, poor borehole conditions, or incomplete legacy logging programs. Empirical relationships between compressional- and shear-wave velocities are therefore widely used to reconstruct missing shear information [3–5]. Their usefulness is well established, but their coefficients and validity ranges depend on lithology, porosity, clay content, saturation, pressure, and measurement conditions. A fixed relation can consequently become unreliable when it is transferred outside the calibration domain.

Machine-learning models provide a flexible alternative by learning multivariate relationships among conventional logs. Previous studies have used neural networks, ensemble methods, and other artificial-intelligence approaches to estimate shear-wave velocity or slowness from petrophysical logs [6,7]. Recent investigations have continued to improve accuracy and, in some cases, incorporate uncertainty or rock-physics validation [8,9]. However, a high score on a randomly sampled test set is not equivalent to reliable prediction



Academic Editors: Yong Yuan,
Cun Zhang and Bo Li

Received: 21 August 2026
Revised: 30 September 2026
Accepted: 6 October 2026
Published: 16 October 2026

Copyright: © 2026 by the authors.
Licensee MDPI, Basel, Switzerland.
This article is an open access article
distributed under the terms and
conditions of the [Creative Commons
Attribution \(CC BY\)](https://creativecommons.org/licenses/by/4.0/) license.

in a new depth interval or well. Adjacent well-log samples are strongly autocorrelated, and random partitions may place neighboring or geologically similar observations in both training and test subsets. Deployment performance must therefore be evaluated under an explicit distribution shift.

Interpretability is a second unresolved issue. Post hoc explanations can identify associations, but they do not automatically provide a stable or mechanistic model [10,11]. Kolmogorov–Arnold networks (KANs) replace fixed scalar activations with learnable univariate functions on network edges [12]. This representation has motivated the view that KANs may offer a useful compromise between predictive flexibility and inspectable functional structure; KAN 2.0 further emphasizes scientific applications and symbolic analysis [13]. Yet the existence of a fitted edge function does not make that function a physical law. The edge may belong to an intermediate hidden representation, may depend on standardization and network initialization, and may change while the complete network prediction remains similar. A credible KAN study must therefore separate three questions: predictive accuracy, extrapolative behavior, and stability of the claimed interpretation.

Applications have spread quickly through the geoscientific workflow. KANs and KAN hybrids have been used to predict reservoir parameters from well logs [14], to reconstruct resistivity logs for fluid characterization [15], and to forecast production on the public Volve dataset [16]. These studies report accuracy gains under within-dataset evaluation, which is the natural first question. The conditions studied here—leakage-free preprocessing, seed replication, and forward depth extrapolation—are the ones that decide whether a reconstructed log can be trusted beyond its calibration interval, and they have received far less attention.

Generalized additive models (GAMs) provide a relevant interpretable comparator because they combine univariate smooth functions additively [17]. Symbolic regression offers a more direct search for compact equations [18,19]. Tree boosting, support-vector regression, and multilayer perceptrons provide complementary high-performance baselines. Comparing these model families on the same split helps determine whether KAN offers an empirical advantage, an interpretability advantage, both, or neither.

The present study therefore evaluates KAN for shear-wave slowness reconstruction from conventional logs on an open, depth-indexed dataset, treating the response correctly as slowness (DTS, $\mu\text{s}/\text{ft}$) rather than velocity. Its contribution is threefold. It first establishes a leakage-free benchmark: the data partition precedes all fitting, every stochastic comparator is replicated across five seeds, and model selection uses validation data alone. It then characterizes generalization with two complementary designs—blocked depth cross-validation, which removes adjacent-sample optimism, and a shallow-to-deep test whose preprocessing, hyperparameters, and seeds match the benchmark, together with a wide-grid control that bounds the influence of spline clipping. Finally, it audits symbolic template fits to KAN edges across seeds and delimits what such fits can and cannot evidence. The aim is a reproducible assessment of where KAN is competitive, where it fails to generalize, and what its edge functions reveal.

2. Materials and Methods

2.1. Dataset and Response Definition

The workflow comprised data auditing, leakage-free preprocessing, random-split model comparison, blocked depth cross-validation, a harmonized depth-shift stress test, and an edge-level interpretability audit (Figure 1). The response was modeled in its recorded slowness unit.

The source was the public Geomechanical Properties of Rock dataset distributed through Kaggle [20]. The local comma-separated file contained 1904 depth-indexed rows

and eight measured variables after removal of an exported index column. No missing numeric values were present. The source package available for this project did not include a DOI, well identifier, lithology labels, geographic location, acquisition description, or a complete unit dictionary.

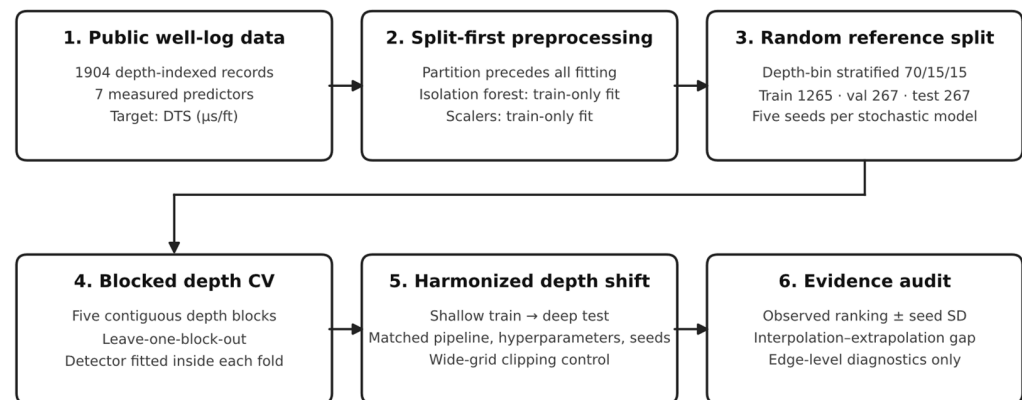


Figure 1. Evidence chain for the study.

The recorded response field is Shear Wave Travel Time. Its values range from 85.47 to 184.78 $\mu\text{s}/\text{ft}$ and are shear-wave transit time per unit length—commonly called shear-wave slowness or DTS. We denote it by Δt_s ; It does not refer to velocity. If conversion is required, shear-wave velocity in SI units is

$$V_s (\text{m s}^{-1}) = \frac{304,800}{\Delta t_s (\mu\text{s}/\text{ft})} \quad (1)$$

The analogous compressional-wave field is denoted as Δt_p (DTP). Modeling Δt_s directly avoids changing the response distribution and preserves the unit used by the source file. Depth was recorded in feet and is also reported in meters using $1 \text{ ft} = 0.3048 \text{ m}$. Table 1 show the measured variables in the source dataset before outlier filtering.

The labels RES, GR, and ρ_b follow common well-log terminology, but conventional units must not be assumed until they are checked against the original data producer’s documentation.

Table 1. Measured variables in the source dataset before outlier filtering.

Role	Source Field	Notation	Range	Unit Used in This Study
Predictor	Depth	D	6805.0–7756.5	ft (2074.2–2364.2 m)
Predictor	Resistivity	RES	17.413–7223.241	Not specified in source metadata
Predictor	Gamma Ray	GR	24.463–623.150	Not specified in source metadata
Predictor	Total Porosity	ϕ_t	0.00287–0.14208	Fraction
Predictor	Effective Porosity	ϕ_e	0.00212–0.11932	Fraction
Predictor	Bulk Density	ρ_b	2.42895–2.84976	Not specified in source metadata
Predictor	Compression Wave Travel Time	Δt_p	50.806–101.456	$\mu\text{s}/\text{ft}$
Response	Shear Wave Travel Time	Δt_s	85.474–184.779	$\mu\text{s}/\text{ft}$

2.2. Data Audit and Preprocessing

Two deterministic interaction features were added after cleaning:

$$x_{\rho\phi} = \rho_b \phi_t \quad (2)$$

and

$$x_{I^*} = 10^6 \frac{\rho_b}{\Delta t_p} \quad (3)$$

Equation (3) is called an acoustic-impedance proxy, not acoustic impedance, because the source metadata do not establish a fully consistent SI unit chain. The final input vector contained nine features: the seven measured predictors and the two engineered proxies.

The train–validation–test partition was drawn from the 1904 source rows before any fitting: ten quantile bins of depth served as stratification labels, with 70% of the observations assigned to the training subset and 15% each to the validation and test subsets. An Isolation Forest with contamination 0.05 and random seed 42 was then fitted to the seven measured predictors of the training subset and applied to the validation and test subsets [21]; 1265 training, 267 validation, and 267 test observations were retained. StandardScaler parameters were fitted on the retained training subset and applied to the validation and test subsets; the response was standardized in the same manner for optimization, and all reported predictions were inverse-transformed to $\mu\text{s}/\text{ft}$ before error computation. In the depth-shift and blocked protocols the detector is applied only to the development data, and the deliberately shifted test domains are evaluated unfiltered so that the extrapolation regime is retained. The influence of this difference in test filtering was measured directly: on the unfiltered random test set the order of the leading families (XGBoost, LightGBM, MLP, KAN) is unchanged and their RMSE values shift by at most $0.9 \mu\text{s}/\text{ft}$, while the most sensitive baselines are the additive GAM ($+5.1 \mu\text{s}/\text{ft}$) and the kernel SVR ($+2.2 \mu\text{s}/\text{ft}$). Test filtering therefore matters chiefly for the two least flexible baselines and does not alter the comparison of the leading families.

Figure 2 summarizes the unfiltered source table. Shear-wave slowness had a mean of $123.93 \mu\text{s}/\text{ft}$ and a standard deviation of $19.25 \mu\text{s}/\text{ft}$. The strongest pairwise relationship was with P-wave slowness ($r = 0.952$), but the depth profile and multimodal target distribution show that the data are not independent and identically distributed in the ordinary tabular sense.

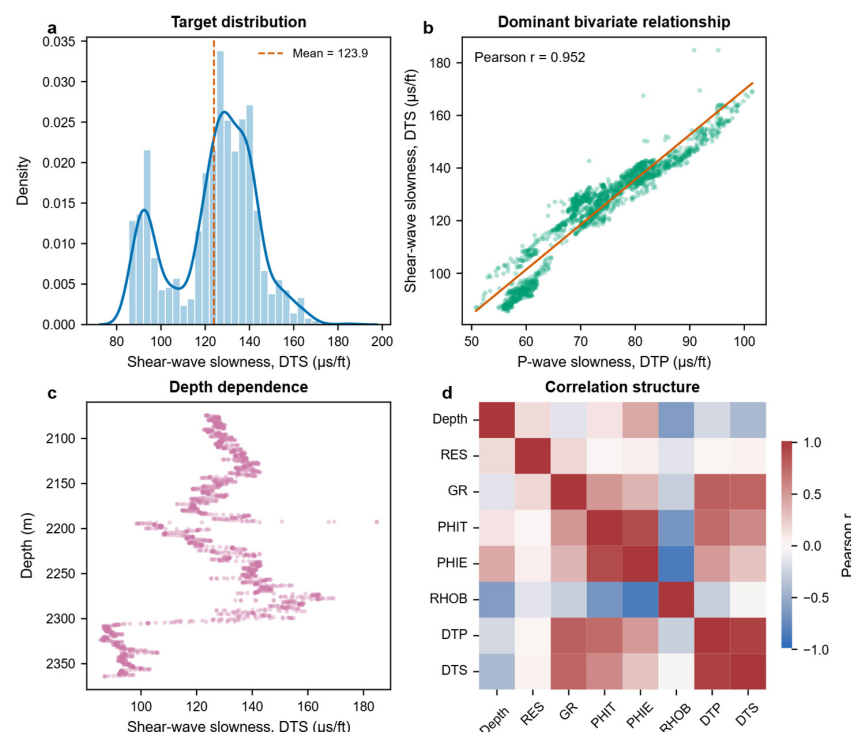


Figure 2. Source-data audit. (a) Distribution of shear-wave slowness. (b) Relationship between P-wave and shear-wave slowness; the line is an ordinary least-squares visual guide fitted to all source rows. (c) Shear-wave slowness versus depth. (d) Pearson correlation matrix for the measured predictors and response. RES: resistivity; GR: gamma ray; PHIT: total porosity; PHIE: effective porosity; RHOB: bulk density; DTP: P-wave slowness; DTS: shear-wave slowness.

2.3. Kolmogorov–Arnold Network

A KAN layer maps an input vector $\mathbf{x} \in \mathbb{R}^n$ to an output vector $\mathbf{z} \in \mathbb{R}^m$ by summing trainable univariate edge functions. In the present implementation,

$$z_j = \sum_{i=1}^n \left[w_{ji}^{(b)} \text{SiLU}(x_i) + s_{ji} \sum_{q=1}^Q c_{jiq} B_q(x_i) \right] \quad (4)$$

where B_q are B-spline basis functions, $w_{ji}^{(b)}$ is the residual-base weight, c_{jiq} are spline coefficients, and s_{ji} scales the spline component. The network architecture was [9, 5, 5, 1], with learnable edge functions in all three layers; the residual activation is $\text{SiLU}(x) = x \cdot \text{sigmoid}(x)$. Cubic splines (order 3) were defined on five grid intervals over the standardized range $[-2, 2]$, giving eight basis functions per edge, and inputs outside the grid were clipped to its endpoints for spline evaluation. To bound the influence of this clipping under extrapolation, a wide-grid variant of the same architecture spans the spline support over the full standardized training range (floor of the minimum minus 0.5 to ceiling of the maximum plus 0.5). The implementation follows the edge-function formulation of Liu et al. [12,13] and omits the grid-extension and symbolic-pruning utilities of the reference software.

KAN was optimized using mean-squared error only:

$$\mathcal{L}_{\text{data}} = \frac{1}{N} \sum_{k=1}^N (\hat{y}_k - y_k)^2 \quad (5)$$

AdamW was used with learning rate 10^{-3} , weight decay 10^{-4} , batch size 64, gradient-norm clipping at 1.0, a cosine learning-rate schedule, and a maximum of 500 epochs [22]. Training stopped if validation loss failed to improve for 40 epochs, and the checkpoint with minimum validation loss was restored. Five random seeds (42, 123, 456, 789, and 1024) were evaluated for the final KAN random-split result.

2.4. Comparator Models

Six comparators were evaluated on the same standardized random split:

1. Multilayer perceptron (MLP): hidden widths 64, 32, and 16; ReLU activation, batch normalization, and dropout 0.1, followed by an 8-unit ReLU head. The same optimizer, stopping rule, and data loss as KAN were used.
2. XGBoost: 300 trees, maximum depth 6, learning rate 0.05, row subsampling 0.8, and column subsampling 0.8 [23].
3. LightGBM: 300 trees, maximum depth 6, learning rate 0.05, row subsampling 0.8, and column subsampling 0.8 [24].
4. Support-vector regression (SVR): radial-basis-function kernel, $C = 1.0$, and $\epsilon = 0.1$ [25].
5. Generalized additive model (GAM): one univariate B-spline term per input feature. Candidate basis sizes of 6, 10, 15, and 20 splines were searched with smoothing parameter 0.1; the basis size was selected by validation-subset R^2 alone, and the selected model was then evaluated once on the test subset.
6. Symbolic regression: genetic programming with 5000 programs for 20 generations, a function set containing arithmetic, protected roots/logarithms, trigonometric functions, inversion, and negation, and a parsimony coefficient of 0.001 [19].

Stochastic models—KAN, MLP, XGBoost, LightGBM, and genetic-programming symbolic regression—were each repeated across five random seeds (42, 123, 456, 789, and 1024); SVR and pyGAM are deterministic algorithms under these settings and were evaluated once. All comparators use literature-default hyperparameters with the single validation-

based selection described above, so the benchmark characterizes default-configuration behavior under a uniform protocol rather than exhaustively tuned performance.

2.5. Blocked Depth Cross-Validation

Random partitions of depth-indexed logs place adjacent, strongly autocorrelated samples in both training and test subsets, so random-split scores can be optimistic even under depth-bin stratification. A blocked design removes this pathway. The 1904 rows were ordered by depth and divided into five contiguous blocks of 381 rows (the last containing 380). In fold k , block k was held out for testing and the remaining four blocks formed the development set, within which a depth-stratified 15% validation subset was drawn; because this validation subset remains interleaved with training depths, early-stopping estimates inherit a mild optimism that the held-out block itself does not. The isolation-forest detector and both scalers were fitted on the training subset of each fold only, and the held-out block was evaluated unfiltered, so every observation is tested exactly once under a partition that never mixes neighboring depths. All seven model families were evaluated with the hyperparameters of Section 2.4; stochastic models used three seeds per fold.

2.6. Depth-Extrapolation Stress Test

The 1904 source rows were ordered by depth and divided at the median, 7280.75 ft (2219.2 m): the shallower 952 rows formed the development domain and the deeper 952 rows formed the test domain. Within the shallow half, a depth-bin-stratified split assigned 85% of the rows to training and 15% to validation. Preprocessing, model configurations, hyperparameters, and seeds are identical to the random-split benchmark; the test domains differ in depth range and are evaluated unfiltered (Section 2.2), so performance differences between the protocols reflect the depth shift rather than tuning or architecture differences. An isolation forest fitted on the shallow training subset flags 66.4% of the deep rows as anomalous, and the mean absolute z-score of the deep half across the seven predictors is 1.9 relative to shallow-training statistics (depth 3.5σ ; gamma ray 2.7σ); both quantities summarize the covariate shift. All seven model families were evaluated, and KAN additionally in the wide-grid variant of Section 2.3.

2.7. Edge-Level Symbolic Surrogate Audit

The interpretability analysis inspected first-layer edges entering the first hidden node of trained KANs. For each input feature, 100 standardized values spanning the observed training range were passed through the selected edge while the remaining inputs were held at a single fixed training sample (the first row of the training set in standardized coordinates). Seven candidate templates—linear, quadratic, cubic, exponential, logarithmic, power, and sigmoid, covering the function families most commonly used in petrophysical trend modeling—were fitted by nonlinear least squares. For each seed, the five edges with the highest template-fit R^2 were retained; the retained feature sets are compared across seeds in Section 3.4.

This procedure approximates individual hidden-edge activations. It does not collapse the stacked layers into a complete mapping from the nine inputs to DTS, and its coefficients are expressed in standardized coordinates. The analysis therefore evaluates local shape and cross-seed stability, not a field-deployable equation or a new rock-physics law.

2.8. Evaluation and Statistical Reporting

Performance was measured using the coefficient of determination, root-mean-square error, and mean absolute error:

$$R^2 = 1 - \frac{\sum_{k=1}^N (y_k - \hat{y}_k)^2}{\sum_{k=1}^N (y_k - \bar{y})^2} \quad (6)$$

$$\text{RMSE} = \sqrt{\frac{1}{N} \sum_{k=1}^N (y_k - \hat{y}_k)^2} \quad (7)$$

and

$$\text{MAE} = \frac{1}{N} \sum_{k=1}^N |y_k - \hat{y}_k| \quad (8)$$

RMSE and MAE are reported in $\mu\text{s}/\text{ft}$ after inverse transformation. Stochastic models are summarized as mean $\pm$ population standard deviation (ddof = 0) across five seeds on the random and depth protocols and across three seeds per fold in blocked cross-validation; SVR and GAM are deterministic and carry no seed dispersion. Seed repetitions share one split and quantify algorithmic variability rather than geological sampling variability, so rankings are read as point estimates under the studied partition and differences smaller than the seed dispersion are not interpreted.

The analysis was implemented in Python 3.11 using PyTorch 2.6.0, pandas 2.2.2, NumPy 2.5.0, scikit-learn 1.9.0 [26], XGBoost 3.3.0, LightGBM 4.6.0, pyGAM 0.12.0, gplearn 0.4.3, Matplotlib 3.11.0, and Seaborn 0.13.2. Machine-readable split arrays, scalars, model summaries, seed-level metrics, and symbolic-fit parameters are retained in the project directory.

3. Results

3.1. Random-Split Model Comparison

Table 2 and Figure 3 compare the seven families on the random test split. XGBoost achieved the highest accuracy ($R^2 = 0.9956 \pm 0.0002$; $\text{RMSE} = 1.26 \pm 0.03 \mu\text{s}/\text{ft}$), followed by LightGBM and the MLP. KAN achieved a five-seed mean R^2 of 0.9891 with RMSE 1.98 $\mu\text{s}/\text{ft}$ and MAE 1.43 $\mu\text{s}/\text{ft}$ —essentially level with SVR (0.9890) and ahead of the GAM (0.9851) and symbolic regression (0.9666 ± 0.0114). On the unfiltered test set (286 rows) the leading families keep their order, with RMSE increases of 0.3–0.9 $\mu\text{s}/\text{ft}$ for XGBoost, LightGBM, the MLP, and KAN; the additive GAM and SVR are the most sensitive to the flagged rows (+5.1 and +2.2 $\mu\text{s}/\text{ft}$).

Table 2. Performance on the 267-record random test subset. Stochastic models: mean $\pm$ population SD across five seeds; SVR and GAM: single deterministic runs.

Model	Runs	R^2	RMSE ($\mu\text{s}/\text{ft}$)	MAE ($\mu\text{s}/\text{ft}$)
MLP	5	0.9912 ± 0.0014	1.77 ± 0.15	1.32 ± 0.09
XGBoost	5	0.9956 ± 0.0002	1.26 ± 0.03	0.89 ± 0.02
LightGBM	5	0.9939 ± 0.0003	1.48 ± 0.04	1.03 ± 0.02
KAN	5	0.9891 ± 0.0010	1.98 ± 0.09	1.43 ± 0.08
SVR	1 (det.)	0.9890	1.99	1.47
GAM	1 (det.)	0.9851	2.31	1.81
Symbolic regression	5	0.9666 ± 0.0114	3.42 ± 0.56	2.76 ± 0.50

Differences among the top three models are small in absolute terms. The lower symbolic-regression score indicates that a compact global expression drawn from the specified function set could not reproduce the multivariate relationship as accurately as the flexible nonlinear models.

Across seeds, KAN R^2 ranged from 0.9878 to 0.9908 and MAE from 1.31 to 1.56 $\mu\text{s}/\text{ft}$ —a dispersion that is small next to the random-to-depth performance gap discussed below.

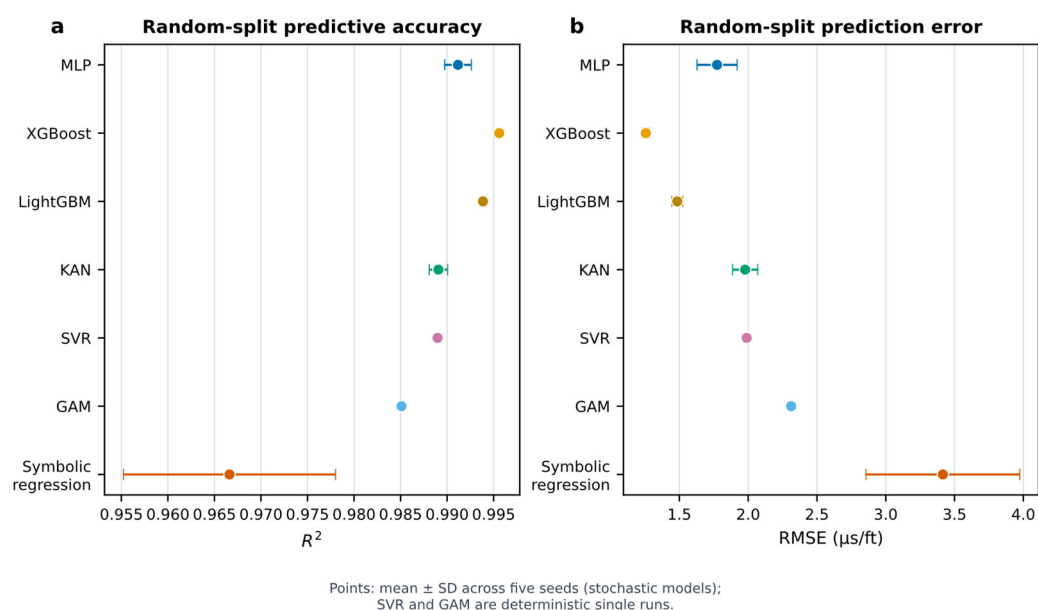


Figure 3. Random-split model comparison. (a) Coefficient of determination. (b) RMSE after inverse transformation to $\mu\text{s/ft}$. Points: mean $\pm$ SD across five seeds (stochastic models); SVR and GAM are single deterministic runs.

3.2. Performance Under Blocked Depth Cross-Validation

Blocked cross-validation removes the adjacent-sample optimism of the random split, and the effect is severe (Table 3, Figure 4). Mean per-fold RMSE rises to 6.8 $\mu\text{s/ft}$ for KAN and 7.8 $\mu\text{s/ft}$ for the MLP (random-split means: 2.0 and 1.8 $\mu\text{s/ft}$), and on the extrapolative end folds every family produces negative R^2 . Two mechanisms produce the extreme values. On fold 4, where the GAM and symbolic models reach R^2 of -38.1 and -6.2 , the failure is divergence: 31% of GAM and 34% of symbolic predictions fall outside the range of their training response. On fold 5, the mechanisms combine: the held-out block's response SD is only 4.1 $\mu\text{s/ft}$ (fold 1: 4.7), so a given absolute error depresses R^2 far more than in the random split, and the MLP and GAM place 52% and 27% of predictions outside the training range. Because R^2 divides the squared error by the within-fold variance, it magnifies both failure modes; RMSE is therefore used as the cross-fold summary and per-fold R^2 is reported for transparency.

Table 3. Blocked depth cross-validation (leave-one-block-out over five contiguous depth intervals; test folds unfiltered). Cells: per-fold R^2 , mean $\pm$ population SD across three seeds for stochastic models (KAN, MLP, symbolic regression); single values for deterministic SVR and GAM. Mean RMSE is the average across folds.

Model	Fold 1	Fold 2	Fold 3	Fold 4	Fold 5	Mean RMSE ($\mu\text{s/ft}$)
MLP	-0.18 ± 0.42	0.59 ± 0.05	0.84 ± 0.04	0.49 ± 0.18	-12.53 ± 7.43	7.76
XGBoost	-0.25 ± 0.00	0.18 ± 0.00	0.88 ± 0.00	0.54 ± 0.00	-0.42 ± 0.00	6.07
LightGBM	-0.18 ± 0.00	-0.06 ± 0.00	0.84 ± 0.00	0.42 ± 0.00	-2.59 ± 0.00	7.13
KAN	-0.37 ± 0.53	0.58 ± 0.06	0.83 ± 0.01	-0.02 ± 0.17	-0.49 ± 0.55	6.78
SVR	-0.15	0.68	0.78	-1.28	-23.20	11.15
GAM	-1.35	0.29	0.80	-38.12	-0.10	21.44
Symbolic regression	-1.51 ± 1.42	0.71 ± 0.06	0.79 ± 0.10	-6.16 ± 4.88	-8.98 ± 1.51	12.20

Within this design, the interior folds separate the families cleanly: on fold 3, where the held-out block is bracketed by training depths, most families reach R^2 of 0.78–0.88, with KAN and the MLP the strongest of the neural models. The end folds collapse for every family. No model trained on this sequence retains usable accuracy under genuine

depth extrapolation, which is the deployment-relevant characterization of this dataset; the random-split scores of Table 2 should be read against this reference.

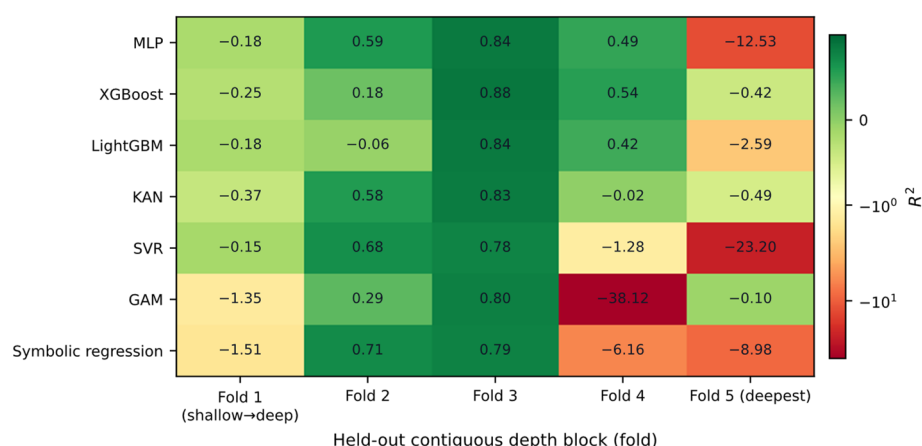


Figure 4. Blocked depth cross-validation: per-fold mean R^2 by held-out contiguous depth block. Fold 1 (shallowest) and Fold 5 (deepest) require extrapolation; middle folds interpolate across a removed depth interval.

3.3. Shallow-to-Deep Generalization

All seven families lose most of their accuracy when the development domain is restricted to the shallow half (Table 4, Figure 5). KAN falls to $R^2 = 0.5624 \pm 0.1660$ (RMSE 16.48 ± 2.99 $\mu\text{s}/\text{ft}$, an 8.3-fold increase over its random-split mean) and the MLP to $R^2 = 0.4258 \pm 0.0227$ (a 10.8-fold increase); the kernel SVR collapses almost completely ($R^2 = 0.1361$). The boosted ensembles retain more of their accuracy (LightGBM 0.7438 ± 0.0031 ; XGBoost 0.6776 ± 0.0269), the additive GAM falls to 0.5118, and the symbolic models retain the highest mean accuracy ($R^2 = 0.8513 \pm 0.0925$) though with the widest seed-to-seed spread of any family. Standardized deep-half inputs fall outside the default $[-2, 2]$ spline grid for 42.3% of sample-feature entries (up to 94.2% for depth). The wide-grid KAN variant, whose spline support covers the training-standardized range, improves deep RMSE only to 15.21 ± 0.69 $\mu\text{s}/\text{ft}$, establishing an upper bound of roughly 1.3 $\mu\text{s}/\text{ft}$ on the contribution of boundary clipping; the remaining degradation reflects the difficulty of forward depth extrapolation for this architecture and dataset.

Table 4. Shallow-to-deep stress test with preprocessing, hyperparameters, and seeds matched to the random-split benchmark. RMSE in $\mu\text{s}/\text{ft}$; ratio is deep RMSE divided by the random-split mean. Stochastic models: mean $\pm$ SD across five seeds; SVR and GAM: deterministic single runs.

Model	Random R^2	Deep R^2	Deep RMSE ($\mu\text{s}/\text{ft}$)	RMSE Ratio
MLP	0.9912 ± 0.0014	0.4258 ± 0.0227	19.18 ± 0.38	$10.8\times$
XGBoost	0.9956 ± 0.0002	0.6776 ± 0.0269	14.36 ± 0.60	$11.4\times$
LightGBM	0.9939 ± 0.0003	0.7438 ± 0.0031	12.81 ± 0.08	$8.6\times$
KAN	0.9891 ± 0.0010	0.5624 ± 0.1660	16.48 ± 2.99	$8.3\times$
KAN (wide grid)	0.9891 ± 0.0010	0.6381 ± 0.0336	15.21 ± 0.69	$7.7\times$
SVR	0.9890	0.1361	23.53	$11.8\times$
GAM	0.9851	0.5118	17.69	$7.7\times$
Symbolic regression	0.9666 ± 0.0114	0.8513 ± 0.0925	9.39 ± 2.67	$2.7\times$

With preprocessing, hyperparameters, and seeds matched across protocols, the deterioration in Table 4 reflects the depth shift itself. The test changes depth and correlated log distributions simultaneously and cannot isolate a single causal factor; the blocked

cross-validation of Section 3.2, which varies the train–test distance more gradually, shows the same ordering of failure modes.

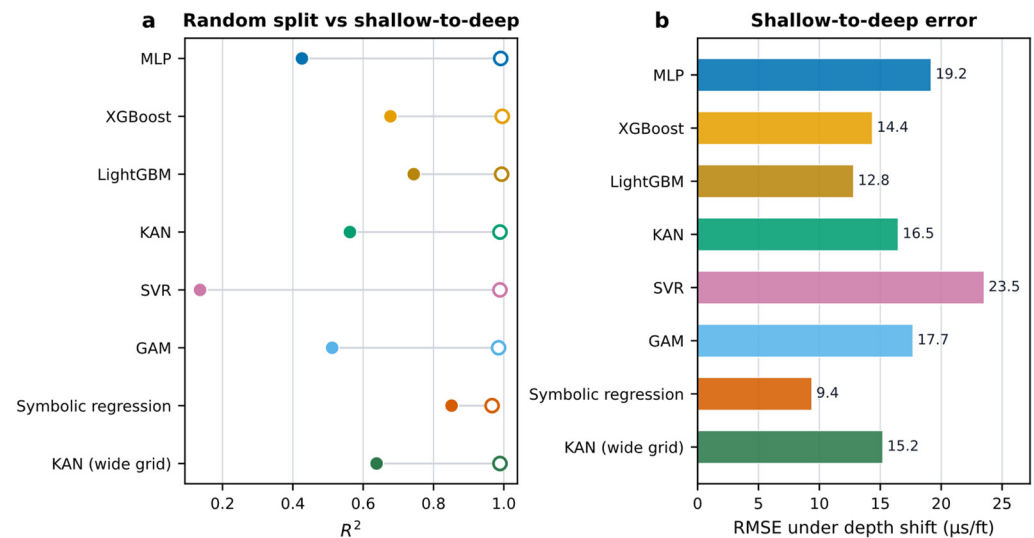


Figure 5. Shallow-to-deep extrapolation. (a) R^2 on the random split (open) versus the shallow-to-deep partition (filled) with matched preprocessing and hyperparameters. (b) RMSE under depth shift in the original DTS unit; the wide-grid KAN bounds the influence of spline boundary clipping.

3.4. Edge-Level Symbolic Surrogates

Many first-layer edges remain compressible into simple templates with high local R^2 (Figure 6). The DTP edge was fitted by a sigmoid template for seed 42 ($R^2 = 0.988$); a cubic template for seed 123 ($R^2 = 0.966$); a cubic template for seed 456 ($R^2 = 0.976$). DTP and the acoustic-impedance proxy appear among the five best-fitted edges in all three seeds, but the remaining slots differ, and no feature–template pair recurs in all three seeds.

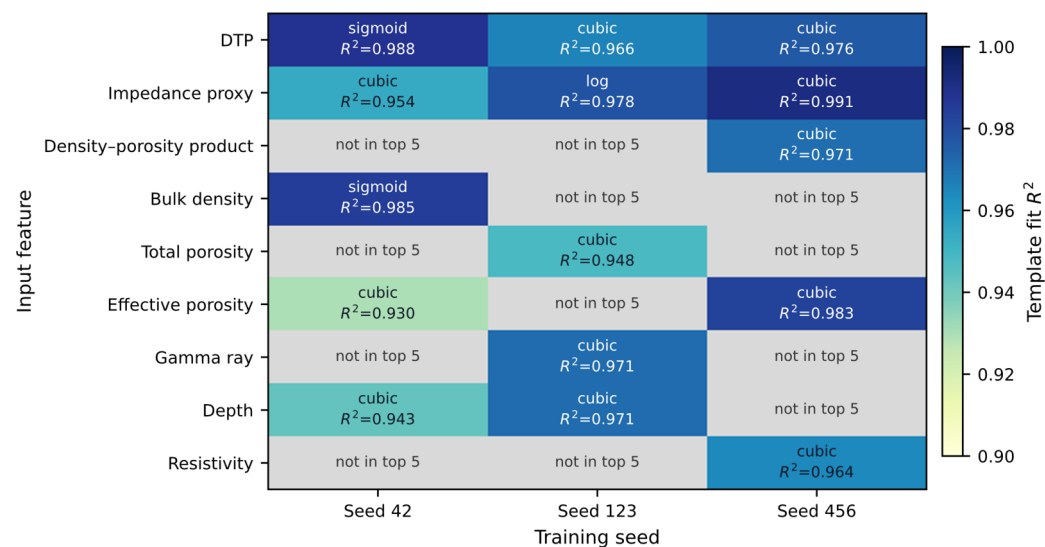


Figure 6. Best-fitting templates for selected first-layer edges entering the first KAN hidden node across three seeds. Cell text gives the fitted template family and its local R^2 ; grey cells mark features not among the five best-fitted edges for that seed. Values describe local edge activations in standardized coordinates, not a complete input-to-output mapping.

The high within-edge fit scores show that some learned spline shapes are compressible into familiar mathematical forms; the changing template identity shows that this compressibility is not a stable physical relation. Edge plots are therefore best used as

hypothesis-generating diagnostics, with coefficients, coordinate transformations, edge location, and cross-seed stability reported alongside.

4. Discussion

On the random split, KAN sits in the middle of a tightly packed field: the boosted ensembles and the MLP score higher, SVR essentially matches it (R^2 0.9890 versus 0.9891), and the additive GAM follows closely. The narrow margin over the GAM is the most instructive comparison, because both families learn univariate spline components—the GAM sums them additively while KAN composes them across layers. On this dataset the additional composition depth bought only a small error reduction, suggesting that once the dominant DTP relationship is accounted for, the residual multivariate structure of these logs is close to additive. An RMSE near 2 $\mu\text{s}/\text{ft}$ is small relative to the 19.25 $\mu\text{s}/\text{ft}$ response standard deviation, so the interpolation performance of all competitive families is practically equivalent; the choice among them should rest on secondary criteria—training cost, inspectability, and extrapolation behavior—rather than on random-split accuracy.

The generalization results are the paper's central finding, and the two designs agree. Blocked cross-validation, which merely prevents neighboring depths from sharing the training and test sides, raises mean per-fold RMSE to 6.8 $\mu\text{s}/\text{ft}$ for KAN and 7.8 $\mu\text{s}/\text{ft}$ for the MLP, and the shallow-to-deep test raises KAN's RMSE by a factor of 8 and the MLP's by 11. Model families differ systematically in how gracefully they fail: the compact global symbolic expressions retain the highest mean shifted-domain accuracy, the boosted ensembles are intermediate, and the dense and stacked-spline networks together with the kernel SVR fare worst. A model selected on random-split accuracy alone would therefore appear far more deployment-ready than either shifted evaluation supports, which motivates leave-one-well-out, rolling-origin, or lithology-blocked validation before operational use.

The symbolic audit adds a caution about interpretability. Many first-layer edges compress into simple templates with high local R^2 , but the selected template identity changes across seeds: DTP and the acoustic-impedance proxy are consistently the most compressible inputs, yet no feature–template pair recurs in all three seeds and the DTP edge switches template family between runs. This is local functional compressibility, not discovery of a physical law. The analyzed functions are edges entering a hidden node rather than the complete mapping from inputs to DTS; standardization alters apparent shapes and coefficients; and multiple parameterizations can represent similar network functions. Stable shapes observed across seeds, wells, and preprocessing choices could still motivate geological hypotheses, but such hypotheses require validation in original units, comparison with established rock-physics relationships, and testing on independent wells. Figure 6 is only the first step of that process.

A related caution concerns physics-informed regularization. Physics-based penalties are valuable when formulated in the correct physical space [27,28], but they fail in characteristic ways when the constraint and the representation mismatch [29]. In this project, exploratory penalties enforcing positivity, an empirical DTP–DTS relation, and an elastic ratio were evaluated on standardized predictions and inputs, where zero represents the training mean rather than zero physical slowness, and a ratio of standardized quantities has no invariant rock-physics meaning; those experiments were therefore set aside. A valid formulation must inverse-transform predictions and inputs before evaluating any dimensional constraint, following the practice of physics-guided well-log and petrophysical inversion studies [30–32], and must be tested on independent geological domains. The lesson generalizes beyond KAN: physical meaning must survive preprocessing.

This study delivers point predictions only. An exploratory five-model ensemble produced 46.3% empirical coverage for a nominal 95% interval, so no calibrated uncertainty

is offered here; deployment of reconstructed sonic logs requires intervals that respond to measurement noise and distribution shift, calibrated on independent data and evaluated under blocked validation. Before operational use, a reconstructed DTS log should be screened for out-of-domain inputs, plotted against depth, compared with rock-physics constraints in original units, and flagged with a domain-of-applicability indicator.

Three limitations bound these conclusions. First, the public table appears to represent a single depth-indexed sequence without well, field, lithology, or acquisition metadata, so neither the blocked nor the depth-shift design can establish cross-well or cross-formation generalization. Second, native units are missing for several predictors, which prevents verification of dimensional physical constraints and qualifies the engineered acoustic-impedance proxy. Third, the symbolic audit covers one first-layer node and three seeds rather than the full network. Within these bounds the study is an auditable proof of concept rather than a claim of field validity.

5. Conclusions

This study evaluated a KAN for shear-wave slowness prediction on a public well-log table under a leakage-free protocol with seed replication, blocked depth cross-validation, and a matched shallow-to-deep test. Within the table, KAN interpolates accurately and stably, but boosted ensembles, the MLP, and SVR match or exceed it. Under both shifted evaluations every family degrades severely—KAN's depth-shift RMSE rises by a factor of 8—and the extrapolative end folds collapse for all families, while compact symbolic models retain the highest mean shifted-domain accuracy. Edge-level template fits compress individual KAN edges well but are not stable across seeds, and no complete physical equation is recovered.

The defensible conclusion is, therefore, conditional: KAN is a competitive interpolator within this table and a hypothesis-generating source of edge-level diagnostics, but this dataset does not establish predictive superiority, stable law discovery, or deployment readiness. The paper's central empirical statement is the separation between interpolation accuracy and depth-generalization ability, which persists under leakage-free preprocessing, seed replication, and controlled partitioning. A deployment-grade continuation requires multiwell blocked validation, exact unit metadata, full-network interpretability analysis, and calibrated uncertainty under distribution shift.

Author Contributions: Conceptualization, Z.L. and L.H.; methodology, Z.L.; software, Z.L.; validation, Z.L. and L.H.; formal analysis, Z.L.; investigation, Z.L.; resources, Z.L. and L.H.; data curation, Z.L.; writing—original draft preparation, Z.L.; writing—review and editing, Z.L. and L.H.; visualization, Z.L.; supervision, L.H.; project administration, L.H.; funding acquisition, L.H. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The Geomechanical Properties of Rock dataset analyzed in this study is publicly available on Kaggle at <https://www.kaggle.com/datasets/minhazengg/geomechanical-properties-of-rock> (accessed on 18 August 2026). The split arrays, trained-model summaries, and seed-level metrics generated during the current study are available from the corresponding author upon reasonable request.

Acknowledgments: During the preparation of this manuscript/study, the author(s) used Kimi-k3 for the purposes of text translation and text polishing. The authors have reviewed and edited the output and take full responsibility for the content of this publication.

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Mavko, G.; Mukerji, T.; Dvorkin, J. *The Rock Physics Handbook: Tools for Seismic Analysis of Porous Media*, 2nd ed.; Cambridge University Press: Cambridge, UK, 2009. [CrossRef]
2. Fjær, E.; Holt, R.M.; Horsrud, P.; Raaen, A.M.; Risnes, R. *Petroleum Related Rock Mechanics*, 2nd ed.; Elsevier: Amsterdam, The Netherlands, 2008.
3. Castagna, J.P.; Batzle, M.L.; Eastwood, R.L. Relationships between compressional-wave and shear-wave velocities in clastic silicate rocks. *Geophysics* **1985**, *50*, 571–581. [CrossRef]
4. Greenberg, M.L.; Castagna, J.P. Shear-wave velocity estimation in porous rocks: Theoretical formulation, preliminary verification and applications. *Geophys. Prospect.* **1992**, *40*, 195–209. [CrossRef]
5. Han, D.-H.; Nur, A.; Morgan, D. Effects of porosity and clay content on wave velocities in sandstones. *Geophysics* **1986**, *51*, 2093–2107. [CrossRef]
6. Anemangely, M.; Ramezanzadeh, A.; Amiri, H.; Hoseinpour, S.-A. Machine learning technique for the prediction of shear wave velocity using petrophysical logs. *J. Pet. Sci. Eng.* **2019**, *174*, 306–327. [CrossRef]
7. Maleki, S.; Moradzadeh, A.; Riabi, R.G.; Gholami, R.; Sadeghzadeh, F. Prediction of shear wave velocity using empirical correlations and artificial intelligence methods. *NRIAG J. Astron. Geophys.* **2014**, *3*, 70–81. [CrossRef]
8. Ali, A.; Qian, R. Machine-learning prediction of shear-wave velocity from conventional well logs with uncertainty quantification and rock-physics validation. *J. Appl. Geophys.* **2026**, *250*, 106251. [CrossRef]
9. Wibowo, R.C.; Arby, F.M.; Mulyatno, B.S.; Dewanto, O.; Kumalasari, I.N.; Sarkowi, M. Applicability of machine learning for shear-wave velocity prediction from conventional well logs: The LSBoost approach. *Turk. J. Earth Sci.* **2026**, *35*, 236–249. [CrossRef]
10. Rudin, C. Stop explaining black box machine learning models for high stakes decisions and use interpretable models instead. *Nat. Mach. Intell.* **2019**, *1*, 206–215. [CrossRef] [PubMed]
11. Molnar, C. *Interpretable Machine Learning: A Guide for Making Black Box Models Explainable*, 2nd ed.; Lulu.com: Morrisville, NC, USA, 2022. Available online: <https://christophm.github.io/interpretable-ml-book/> (accessed on 18 August 2026).
12. Liu, Z.; Wang, Y.; Vaidya, S.; Ruehle, F.; Halverson, J.; Soljačić, M.; Hou, T.Y.; Tegmark, M. KAN: Kolmogorov–Arnold networks. In Proceedings of the Thirteenth International Conference on Learning Representations, Singapore, 24–28 April 2025.
13. Liu, Z.; Ma, P.; Wang, Y.; Matusik, W.; Tegmark, M. KAN 2.0: Kolmogorov–Arnold networks meet science. *arXiv* **2024**, arXiv:2408.10205. [CrossRef]
14. Yang, H.; Zhang, C.; Xiong, L.; Xiong, W.; Lin, G.; Huang, K.; Zhang, W. Research on prediction method of well logging reservoir parameters based on Multi-TransFKAN model. *Sci. Rep.* **2025**, *15*, 18057. [CrossRef] [PubMed]
15. Gong, A.; Zhang, L.; Sun, Y.; Zhang, Y. Transformer–Kolmogorov–Arnold network for high-fidelity resistivity reconstruction toward fluid characterization. *Phys. Fluids* **2025**, *37*, 126608. [CrossRef]
16. Lu, X.; Cao, J.; Zou, J. Well production forecasting in Volve field using Kolmogorov–Arnold networks. *Energies* **2025**, *18*, 3584. [CrossRef]
17. Hastie, T.; Tibshirani, R. *Generalized Additive Models*; reprint of the 1990 original; Routledge: New York, NY, USA, 2017.
18. Schmidt, M.; Lipson, H. Distilling free-form natural laws from experimental data. *Science* **2009**, *324*, 81–85. [CrossRef] [PubMed]
19. Udrescu, S.-M.; Tegmark, M. AI Feynman: A physics-inspired method for symbolic regression. *Sci. Adv.* **2020**, *6*, eaay2631. [CrossRef] [PubMed]
20. Minhazeng. Geomechanical Properties of Rock. Kaggle. Available online: <https://www.kaggle.com/datasets/minhazeng/geomechanical-properties-of-rock> (accessed on 18 August 2026).
21. Liu, F.T.; Ting, K.M.; Zhou, Z.-H. Isolation Forest. In Proceedings of the 2008 IEEE International Conference on Data Mining, Pisa, Italy, 15–19 December 2008; pp. 413–422. [CrossRef]
22. Loshchilov, I.; Hutter, F. Decoupled weight decay regularization. In Proceedings of the International Conference on Learning Representations, New Orleans, LA, USA, 6–9 May 2019.
23. Chen, T.; Guestrin, C. XGBoost: A scalable tree boosting system. In Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, San Francisco, CA, USA, 13–17 August 2016; pp. 785–794. [CrossRef]
24. Ke, G.; Meng, Q.; Finley, T.; Wang, T.; Chen, W.; Ma, W.; Ye, Q.; Liu, T.-Y. LightGBM: A highly efficient gradient boosting decision tree. In *Advances in Neural Information Processing Systems 30*; Curran Associates: Red Hook, NY, USA, 2017.
25. Cortes, C.; Vapnik, V. Support-vector networks. *Mach. Learn.* **1995**, *20*, 273–297. [CrossRef]
26. Pedregosa, F.; Varoquaux, G.; Gramfort, A.; Michel, V.; Thirion, B.; Grisel, O.; Blondel, M.; Prettenhofer, P.; Weiss, R.; Dubourg, V.; et al. Scikit-learn: Machine learning in Python. *J. Mach. Learn. Res.* **2011**, *12*, 2825–2830.
27. Raissi, M.; Perdikaris, P.; Karniadakis, G.E. Physics-informed neural networks: A deep learning framework for solving forward and inverse problems involving nonlinear partial differential equations. *J. Comput. Phys.* **2019**, *378*, 686–707. [CrossRef]

28. Karniadakis, G.E.; Kevrekidis, I.G.; Lu, L.; Perdikaris, P.; Wang, S.; Yang, L. Physics-informed machine learning. *Nat. Rev. Phys.* **2021**, *3*, 422–440. [[CrossRef](#)]
29. Krishnapriyan, A.S.; Gholami, A.; Zhe, S.; Kirby, R.M.; Mahoney, M.W. Characterizing possible failure modes in physics-informed neural networks. In *Advances in Neural Information Processing Systems 34*; Curran Associates: Red Hook, NY, USA, 2021.
30. Vashisth, D.; Mukerji, T. Direct estimation of porosity from seismic data using rock- and wave-physics-informed neural networks. *Lead. Edge* **2022**, *41*, 840–846. [[CrossRef](#)]
31. Li, P.; Liu, M.; Alfarraj, M.; Tahmasebi, P.; Grana, D. Probabilistic physics-informed neural network for seismic petrophysical inversion. *Geophysics* **2024**, *89*, M17–M32. [[CrossRef](#)]
32. Li, P.; Grana, D.; Liu, M. Bayesian neural network and Bayesian physics-informed neural network via variational inference for seismic petrophysical inversion. *Geophysics* **2024**, *89*, M185–M196. [[CrossRef](#)]

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.