

Article

High-Precision Density Log Reconstruction Method Based on the RF-Transformer Algorithm

Junlei Su ¹, Xu Dong ^{2,*}, Yu Zeng ², Peidong Liu ², Xueying Shi ² and Wenqi Shi ²¹ Petroleum Exploration and Production Research Institute, Sinopec, Beijing 102206, China; sujl.syky@sinopec.com² State Key Laboratory of Continental Shale Oil, Northeast Petroleum University, Daqing 163318, China; zying886@163.com (Y.Z.); lpd_ldg@163.com (P.L.); sherylshi1211@gmail.com (X.S.); shiwenqi10@163.com (W.S.)

* Correspondence: dongxu985@nepu.edu.cn

Abstract

Under the backdrop of digital subsurface and intelligent field development, together with sustainable development planning, reliable and continuous well-log measurements are increasingly essential for reservoir evaluation and geological interpretation. Density (DEN) logging is critical for reservoir evaluation and geological interpretation, providing fundamental constraints for lithology/porosity-related assessment and integrated subsurface characterization. However, the DEN curve often contains missing intervals or distortions caused by borehole conditions and tool/environmental interference. This study proposes an RF-Transformer framework for DEN reconstruction that couples (i) Random-Forest-based feature screening to suppress redundant or low-contribution channels and (ii) a Transformer encoder with mask-aware self-attention to capture both local fluctuations and long-range depth dependencies. Experiments were conducted on logging data from nine vertical wells in the Lianggaoshan Formation (Sichuan Basin, China) with a unified sampling step of 0.125 m. Under a well-wise split protocol, RF-Transformer achieved RMSE = 0.0126 g/cm³, MAE = 0.0079 g/cm³, $R^2 = 0.9863$, and $r = 0.9932$, outperforming Random Forest, Decision Tree, KNN, LightGBM, LightGBM-NN, and a base Transformer. The pass rate reached 92.86% under an error tolerance of ± 0.02 g/cm³, demonstrating robust reconstruction in long missing sections and lithological transition zones. The proposed workflow provides an effective route for repairing density logs in complex reservoirs and for improving the continuity of multi-log interpretation.

Keywords: well-log curve reconstruction; Random Forest; Transformer model; feature selection; reservoir evaluation; geological interpretation

Academic Editor: Nikolaos Koukoulas

Received: 26 January 2026

Revised: 18 February 2026

Accepted: 19 February 2026

Published: 28 February 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.

1. Introduction

Well logs constitute the fundamental data source for petroleum geological interpretation, reservoir evaluation [1,2], and three-dimensional geological modeling; therefore, their continuity and accuracy directly determine the reliability of derived geological parameters [3,4]. In deep and unconventional reservoir settings, key logs such as density, acoustic, and resistivity curves are often affected by missing intervals or distortions, whereas re-logging is frequently impractical from both engineering and economic

perspectives. Consequently, it is imperative to achieve high-accuracy curve reconstruction by leveraging the information contained in available logging measurements [5–7].

To address this problem, the methodological trajectory has evolved from traditional interpolation/regression approaches to machine-learning techniques and, more recently, deep-learning models. Early studies primarily relied on spline interpolation, kriging, and linear regression [8,9], in which missing logs were treated as smooth and locally correlated random fields for estimation [10]. Although these approaches (e.g., kernel ridge regression, KRR) are straightforward to implement, they often fail to recover high-frequency fluctuations and abrupt variations in thin interbeds, strongly heterogeneous formations, or carbonate fractured–vuggy reservoirs [11]. Subsequently, machine-learning models were introduced to identify key inputs that are most informative for the target curve [12]. For instance, Random Forest, owing to its ensemble nature, is robust in handling noisy, multi-source, and partially missing data and can provide uncertainty analysis [13], whereas support vector machines and kernel-based methods exhibit certain advantages in local-neighborhood prediction [14]. With the rapid expansion of deep learning in geophysics, more complex nonlinear modeling has become feasible [15]. One-dimensional convolutional neural networks (CNNs) have been applied to extract local patterns within depth neighborhoods [16], bidirectional recurrent structures (BiGRU and BiLSTM) have been used to capture forward–backward dependencies [17–20], and temporal convolutional networks (TCNs) have also been demonstrated to be effective [21]. These methods establish nonlinear relationships among multiple log curves, improving the consistency of reconstructed results in both local details and overall trends. To overcome the limitations of recurrent architectures on long sequences, Zeng proposed an attention–BiGRU framework [22] that successfully captures vertical correlations in logging sequences. Furthermore, Liao introduced a multi-head self-attention mechanism [23] into density and sonic log reconstruction [24–26], enabling refined trend characterization in thin interbeds and complex reservoirs.

Although existing studies have substantially improved the accuracy of log-curve reconstruction, three common challenges remain. First, many deep-learning models simply fuse all available logging channels at the input stage, without physically motivated prior feature selection; consequently, ineffective or low-contribution features may be amplified simultaneously, thereby impairing model convergence and cross-well generalization. Second, although attention mechanisms can capture long-range dependencies, their modeling efficiency is still susceptible to interference from redundant channels when input-feature screening is absent. Third, due to differences in logging standards, tool replacement, and borehole-environment disturbances, multi-well datasets are prone to distribution shifts; this often leads to performance fluctuations in cross-well applications, and errors are more likely to be amplified in long missing intervals and at abrupt boundaries.

To address these limitations, this study proposes an RF-Transformer framework that integrates Random Forest and Transformer for high-accuracy reconstruction of the DEN curve. Specifically, Random Forest is employed to evaluate variable importance and to screen/rank an informative feature subset; the resulting feature sequence is then fed into the Transformer, where self-attention captures both near-depth details and cross-interval trends, enabling robust characterization of density-variation patterns. Overall, this work achieves high-precision DEN log reconstruction by unifying Random-Forest-based feature selection with Transformer-based sequence modeling, which also supports sustainable development in digital subsurface and intelligent field development by improving log continuity for more reliable reservoir evaluation and reducing avoidable re-logging or repeated operational interventions caused by poor data quality, thereby enhancing operational efficiency and minimizing unnecessary resource consumption.

2. Methods

2.1. Overall Architecture

For the DEN curve reconstruction task characterized by the coexistence of missing intervals and localized distortions, an interpretable workflow is established, consisting of preprocessing, RF-based feature screening, Transformer-based sequence modeling, and regression output. Multiple conventional logging curves together with a missingness mask are used as inputs; after unified resampling and normalization, the RF module is applied to select an informative feature subset and reduce the input dimensionality. The screened feature sequence is then fed into the Transformer encoder to learn multi-scale correlations along the depth direction, and finally, a regression head outputs continuous density values (Figure 1).

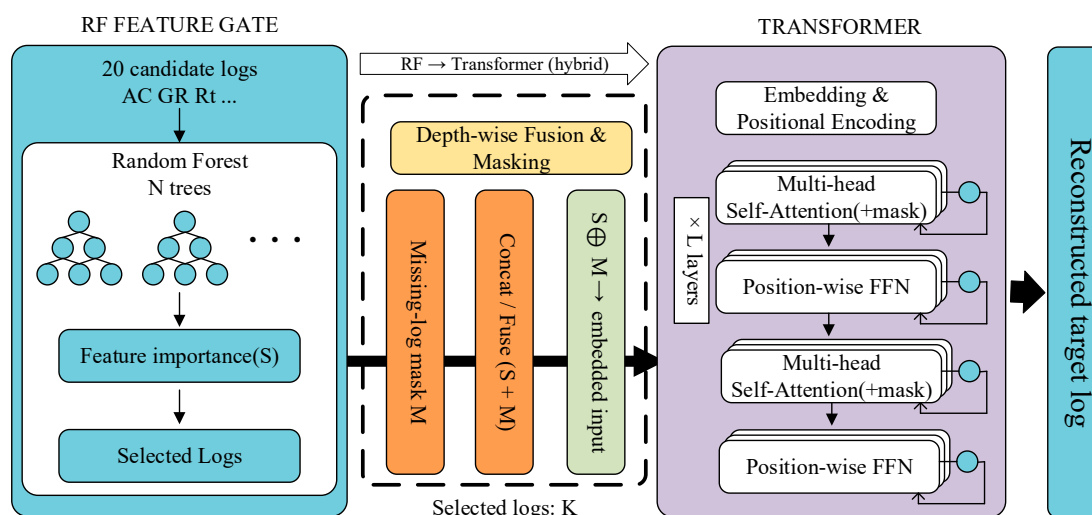


Figure 1. Overall workflow of RF-Transformer for density log reconstruction: preprocessing—RF screening—Transformer modeling—regression output. DEN denotes the density log (g/cm^3); depth is in m. RF denotes Random Forest.

After the basic preprocessing, the multi-source logging inputs are first fed into the Random-Forest-based feature-screening module to establish an integrated workflow including feature-importance evaluation, subset determination, and channel re-encoding. In this study, 20 logging channels are initially provided to the Random Forest for feature screening: CALX, DEVI, AZIM, AC, SP, CALY, CAL, LLD, LLS, GR, RFOC, RILM, RT, CILD, GRSL, K, KTH, TH, U, and CNL. Based on the RF importance ranking and the OOB-error stabilization criterion, the final retained subset is determined as $K = 6$, i.e., CNL, AC, CAL, AZIM, GR, and SP, which are used as the inputs to the Transformer encoder. In this way, without introducing any linearity assumptions, a feature subset with higher relevance to the density curve, richer mutual information, and lower redundancy can be selected. For channels containing complete-interval missingness or local anomalies, dedicated marker values and a mask vector are employed to record their positions; combined with the RF model's insensitivity to missing values, this strategy enables automatic discrimination between usable information and invalid information, thereby avoiding changes to the original sample distribution caused by imputation or deletion. Furthermore, once the RF subset is constructed, a mapping table from original channel indices to encoded channel indices is generated; subsequently, the Transformer interacts only with this mapping table, which markedly compresses the input dimensionality and constrains the input length and scale across different wells within a controllable range.

After the RF feature subset is constructed, the selected logging-feature sequence, together with the depth-wise missingness mask, is jointly fed into the Transformer-based encoding module to achieve unified modeling of multi-scale correlations along the depth axis and abrupt interlayer transition patterns. First, the screened feature sequence is linearly projected into a unified high-dimensional representation space, and a learnable positional encoding is added to each depth location to explicitly represent the depth order. On this basis, the encoder employs a multi-head self-attention mechanism to simultaneously model short-range dependencies among adjacent layers and long-range constraints across intervals, thereby capturing both local abrupt variations in thin interbeds and the background trend in thick layers within the same network. Each encoder layer follows the standard architecture of multi-head attention–position-wise feed-forward network (Position-wise FFN)–residual connection–layer normalization, which enhances nonlinear representation capacity while maintaining training stability in deep networks. Moreover, to accommodate the objective existence of numerous masked positions in logging data, the correlations associated with masked locations are set to extremely small values before the attention softmax, preventing the model from allocating attention to data that were not measured. A lightweight regression head is then appended at the network end to output continuous values of the target curve. Transformer is adopted instead of BiLSTM or TCN-BiGRU because, in multi-well and multi-block joint training, the sequence length often increases multiplicatively, and the attention mechanism provides superior efficiency for long-sequence representation and cross-sample sharing [5]. Considering engineering deployability [27,28], the output side simultaneously computes MAE, RMSE, and structure-similarity-type metrics to distinguish between two error types—numerically close but morphologically distorted versus morphologically consistent but locally shifted—and these objectives are jointly optimized in a weighted manner during training. Overall, the unified data pathway of “RF screening followed by Transformer encoding” improves the stability of cross-well generalization and the controllability of field deployment by reducing ineffective dimensions and redundant noise.

2.2. Random-Forest Feature Screening and Subset Construction

To prevent attention dispersion and unstable generalization caused by high-dimensional redundant inputs, Random Forest is employed as an input-side denoising and compression module. A regression forest is trained to obtain a feature-importance ranking, and the subset size is determined by the convergence of the out-of-bag (OOB) error as a function of the number of retained features. The resulting subset is designed to simultaneously satisfy three criteria—importance, informational complementarity, and cross-well availability (Figure 2). For channels with missing measurements, fixed marker values are used in combination with a mask to record their locations, thereby ensuring that the original sample distribution is not altered.

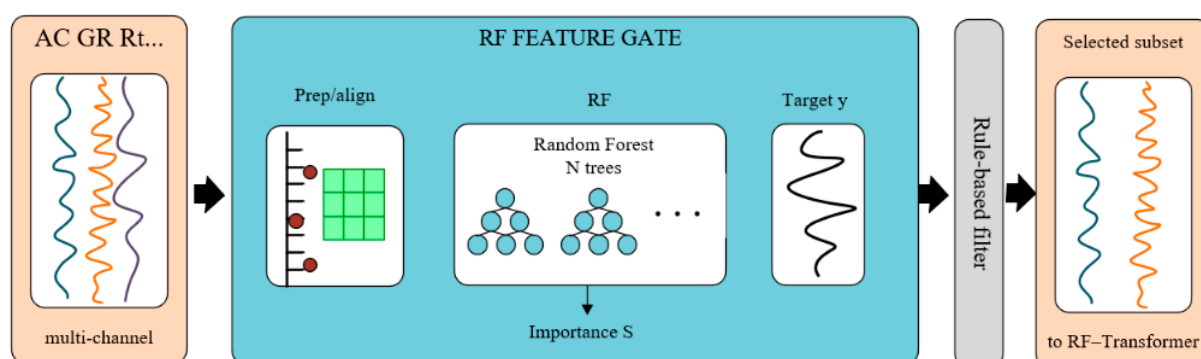


Figure 2. Workflow of RF feature-importance evaluation and feature-subset construction. The input consists of multiple conventional logging curves, and the output is a feature subset ranked by importance.

In terms of sample representation, the original input is denoted as

$$D = \{(x_i, y_i)\}_{i=1}^N \quad (1)$$

where N is the number of depth-sampling points; $x_i \in \mathbb{R}^M$ is the vector composed of M observed logging channels at the i -th depth point; and y_i is the scalar target curve value to be reconstructed at the same depth point.

For channels with complete-interval missingness, dedicated marker values are assigned, and their locations are recorded using a mask vector m_i , so that fields that do not participate in node splitting can be explicitly identified. During RF training, to make the feature-importance ranking better aligned with logging physics, a well-wise sampling and cross-well aggregation strategy is adopted: within each well, the depth order of samples is preserved to avoid disrupting stratigraphic continuity; across wells, samples are then randomly drawn to ensure that the training set covers diverse lithologies, different tool combinations, and multi-period logging operations [29]. With this design, the forest can exploit within-well local variations at split nodes while also learning, from inter-well differences, which curves behave as stable inputs in the study area. After training, RF provides each input log with an average split-gain (or Gini-importance) score. Based on these scores, a feature subset $S \subseteq \{1, \dots, M\}$ is selected in descending order of importance, and the corresponding screened observation at the i -th depth point is denoted as $x_i^{(S)}$. In practice, the subset is not constructed by simply taking the top- k features; instead, three principles are jointly considered: (i) importance must exceed a threshold; (ii) physically complementary information is prioritized while redundancy is suppressed; and (iii) cross-well availability is incorporated to prevent curves present only in a few wells from dominating the modeling, thereby ensuring usability for cross-well applications. This dual constraint of importance and availability was explicitly proposed in a 2025 GAN-based sequence-imputation benchmark [30] to mitigate the overfitting of high-order deep models to rare channels.

RF subset construction yields a mapping table from the original channel indices to the encoded channel indices, and the subsequent Transformer interacts only with this mapping table. Compared with directly feeding all channels into the attention network, this design markedly reduces the input dimensionality, allowing the attention computation to focus on features that are truly relevant to the target curve; meanwhile, it constrains the input length across different wells within a controllable range. Unlike kernel ridge regression, which performs kernel mapping only in the function space, the proposed strategy performs a physically meaningful denoising and compression step in the feature space in advance.

2.3. Transformer-Based Sequence Representation and Reconstruction Regression

The compact well-log feature sequence screened by RF, together with the depth-wise missingness mask, is jointly used as the input and fed into the Transformer encoder, so as to achieve unified modeling of multi-scale correlations along the depth direction and interlayer abrupt-transition patterns (Figure 3).

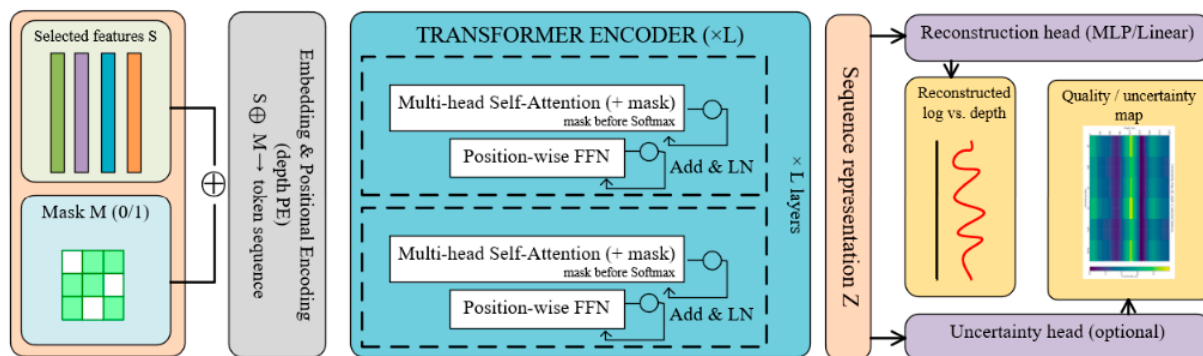


Figure 3. Multi-scale correlation modeling along the depth sequence and the density-regression output architecture of the Transformer encoder. The encoder is composed of multi-head self-attention and a feed-forward network to learn depth-direction correlations; the output is the reconstructed DEN result (g/cm^3).

The RF-screened feature sequence is first linearly projected into a unified d -dimensional vector space

$$z_i = W_e x_i^{(S)} + b_e \quad (2)$$

Here, W_e is the learnable weight matrix that maps the feature subset into the model hidden space; b_e is the corresponding bias vector; and z_i denotes the representation of the i -th depth point in the unified vector space.

To enable the model to recognize the depth order, a positional encoding is added to each depth point. Considering that logging depth is continuous and that the depth ranges vary substantially across wells, a learnable positional encoding is adopted instead of a fixed sinusoidal encoding, which is more suitable for automatically adapting to different depth intervals in cross-well training. The resulting input sequence is

$$h_i^{(0)} = z_i + p_i \quad (3)$$

Here, p_i is the learnable positional vector at the i -th depth location, and $h_i^{(0)}$ is the input representation of the encoder at layer 0 after incorporating positional information.

Each encoder layer adopts the standard multi-head self-attention mechanism [31]. For the first head at layer l , the attention output is computed as

$$\text{Attn}(Q, K, V) = \text{softmax}\left(\frac{QK^T}{\sqrt{d_k}}\right)V \quad (4)$$

Here, the query, key, and value matrices are obtained by linear projections of the previous-layer output: $Q = H^{(l-1)}W_Q$, $K = H^{(l-1)}W_K$, $V = H^{(l-1)}W_V$, $H^{(l-1)}$.

Where $H^{(l-1)}$ denotes the encoder output of the $(l-1)$ -th layer. Multi-head attention concatenates multiple heads computed in parallel and then restores the representation to the original dimensionality through a linear layer. This design allows the model to simultaneously focus on lithological transition points, mud-filtrate-contamination-sensitive intervals, and locations that are physically highly correlated with the target curve, thereby recovering both the macroscopic trend and fine-scale fluctuations within a single reconstructed curve. Compared with RNN-based structures, this attention-only encoding does not suffer from sequential dependence [32], enabling efficient processing of long-interval data and parallel training on samples from multiple wells within the same mini-batch; moreover, within the previously reported large-model framework for well-log reconstruction, it has been shown to achieve improved cross-well generalization without increasing the parameter count [5].

After each self-attention sublayer, a position-wise feed-forward network (Position-wise FFN) is appended, formulated as

$$\text{FFN}(x) = \max(0, xW_1 + b_1)W_2 + b_2 \quad (5)$$

Here, x is the input vector at each position; W_1 and b_1 are the parameters of the first linear layer; W_2 and b_2 are the parameters of the second linear layer; and d_{ff} denotes the intermediate dimensionality of the feed-forward sublayer.

This structure acts independently at each depth location to enhance nonlinear representation capacity. Residual connections and layer normalization are placed after the attention and FFN sublayers, respectively, to stabilize deep training as network depth increases and to mitigate gradient vanishing on long sequences. To accommodate the objective presence of numerous masked positions in well-log data, the correlations associated with masked locations are set to extremely small values before the attention softmax [33], preventing the model from allocating attention to unmeasured data; this is consistent with the masking strategy adopted by Han in log super-resolution [34] to address erroneous alignment in low-resolution intervals.

Transformer encoding yields a high-dimensional representation of the entire well interval; therefore, a regression head is required to project it back into the scalar space of the target curve. In practice, the regression head can be implemented using a one- or two-layer fully connected network

$$\hat{y} = W_o h_i^{(L)} + b_o \quad (6)$$

Here, h_i^L denotes the output representation of the encoder at the L -th layer for the i -th depth position; W_r and b_r are the weight matrix and bias term of the linear regression head, respectively; and $\hat{y}_i$ is the predicted scalar value of the target curve at this depth position.

During training, the objective function is dominated by MAE and is supplemented with MSE or the Huber loss to enhance robustness against outliers

$$\mathcal{L} = \lambda_1 \frac{1}{N} \sum_{i=1}^N |\hat{y}_i - y_i| + \lambda_2 \frac{1}{N} \sum_{i=1}^N (\hat{y}_i - y_i)^2 \quad (7)$$

Here, $\hat{y}_i$ denotes the model prediction at the i -th depth position, y_i is the corresponding ground-truth target value, and $\lambda_1, \lambda_2 \geq 0$ are the weighting coefficients for the two loss terms.

If integration with a GAN- or CGAN-based imputation framework is required, the Transformer output can be treated as a prior input to the generator. In this way, the physical interpretability of RF-Transformer is retained, while the distribution-fitting capability of generative models under small-sample conditions can also be leveraged. This joint strategy is consistent with the latest sequence-based GAN well-log imputation studies [35]; however, in the present framework, sequence-feature learning is moved upstream to the Transformer, and the generator is mainly responsible for fine-detail correction and uncertainty representation.

3. Experiments

3.1. Experimental Setup

Model training and validation were conducted on a high-performance workstation oriented to petroleum data processing (Table 1), equipped with a multi-core CPU, large-capacity memory, and an NVIDIA GPU with large VRAM, so as to support parallel slicing of long multi-well sequences and the associated self-attention computation. The algorithms were implemented uniformly in Python 3.10: deep-learning components were developed in PyTorch 2.3, whereas ensemble baselines such as Random Forest and

LightGBM were implemented using scikit-learn; NumPy and Pandas were used for data ingestion, depth registration, and missing-interval marking; and Matplotlib 3.9 was used for visualization. A fixed random seed was adopted to ensure reproducibility, and logs and model weights were written to a local SSD to reduce I/O-induced variability. Prior to modeling, all input curves were normalized using either min–max (0–1) scaling or Z-score standardization to stabilize gradients and eliminate dimensional inconsistency.

Table 1. Experimental environment configuration.

Module	Specification/Version
CPU	16-core x86_64 (≥ 2.5 GHz)
Memory	64 GB DDR4
GPU	NVIDIA RTX 4090, 24 GB VRAM
Storage	2 TB NVMe SSD
Operating system	Ubuntu 22.04 LTS
Python	3.10
PyTorch	2.3
CUDA/cuDNN	12.1/9
Major libraries	NumPy 1.26; SciPy 1.11; Pandas 2.2; scikit-learn 1.4; Matplotlib 3.9

3.2. Dataset

The experimental data were collected from development well-logging records of the Lianggaoshan Formation, China. This block is characterized by a high logging completeness, diverse curve types, and frequent vertical facies/lithological transitions, which can fully expose typical issues such as unstable correlations among multi-source logs, complete-curve absence over local intervals, and inconsistency in standards between old and new wells—consistent with practical experience reported in multi-curve joint digitization and completion studies [36–38]. The dataset comprises 9 vertical wells with a unified sampling step of 0.125 m, and the compiled dataset contains 3012 depth samples spanning 3680.250–4056.625 m (total span: 376.375 m). The interpretation label includes six lithology-based categories (IDs 1–6) with proportions of 17.37%, 15.78%, 16.81%, 10.59%, 12.09%, and 27.37%, respectively. Along depth, 49 category transitions are observed (i.e., 50 contiguous lithology segments), indicating frequent lithological alternations within the studied interval. Depth registration was performed using the main interval depth as the reference, and quality flags were assigned to tool-change sections, cased-hole sections, and intervals with severe borehole enlargement; the flagged points were excluded from loss computation. The full raw candidate set contains 20 logging channels. After RF screening, only the RF-retained subset ($K = 6$; CNL, AC, CAL, AZIM, GR, and SP) was used as the input to the Transformer encoder, and the density log DEN was selected as the reconstruction target because it is sensitive to reservoir evaluation and is prone to discontinuities under the influence of borehole conditions, invasion effects, and instrument drift.

This study adopts a well-wise split protocol to evaluate cross-well generalization. Specifically, 6 wells are used for training, 1 well is used for validation (hyperparameter selection and early stopping), and the remaining 2 wells are used for testing. Under the unified sampling step of 0.125 m, the compiled dataset contains 3012 depth samples in total, with approximately 2008/335/669 samples in the training/validation/testing splits, respectively.

3.3. Evaluation Metrics

Because the present task is inherently a regression-type reconstruction along a depth sequence, the focus is not on classification accuracy but on whether the correct curve morphology can be recovered at the correct depth positions with sufficient dimensional accuracy. Therefore, the evaluation system is primarily based on continuous metrics, complemented by an engineering-oriented pass-rate metric that is more intuitive for field use. First, the root mean square error (RMSE) and mean absolute error (MAE) are used to quantify the overall dimensional deviation between the reconstructed and ground-truth density curves

$$\text{RMSE} = \sqrt{\frac{1}{N} \sum_{t=1}^N (y_t - \hat{y}_t)^2}, \text{MAE} = \frac{1}{N} \sum_{t=1}^N |y_t - \hat{y}_t| \quad (8)$$

Here, N is the number of sampling points; y_t is the ground-truth density value at depth index t ; and $\hat{y}_t$ is the reconstructed value at the same depth.

Next, the coefficient of determination R^2 and the Pearson correlation coefficient r are introduced to assess whether the model truly learns the depth-dependent trend of the curve rather than merely producing a smoothed interpolation. When $r > 0.9$ while RMSE remains relatively large, it often indicates that the model still exhibits insufficient dimensional tracking in intervals with strong abrupt variations; this phenomenon is most evident in oil–water transition zones, deep high-pressure mudstone sections, and intervals with rapid borehole-diameter changes. When metrics are computed as

$$R^2 = 1 - \frac{\sum_{t=1}^N (y_t - \hat{y}_t)^2}{\sum_{t=1}^N (y_t - \bar{y})^2}, r = \frac{\sum_{t=1}^N (y_t - \bar{y})(\hat{y}_t - \bar{\hat{y}})}{\sqrt{\sum_{t=1}^N (y_t - \bar{y})^2} \sqrt{\sum_{t=1}^N (\hat{y}_t - \bar{\hat{y}})^2}} \quad (9)$$

where $\bar{y}$ is the full-well mean of the ground-truth density curve and $\bar{\hat{y}}$ is the mean of the reconstructed curve.

Finally, an engineering-threshold pass rate is reported to reflect the proportion of depth points that can be directly used for interpretation under a prescribed error tolerance. In summary, a combined system of continuous error and consistency metrics is adopted—RMSE and MAE for dimensional deviation, and R^2 and r for trend consistency—together with a threshold-based pass rate to quantify the effective depth coverage for practical use.

4. Results and Discussion

4.1. Overall Accuracy and Trend Consistency

Under a unified data split, preprocessing procedure, and evaluation protocol, the reconstruction performance of RF-Transformer is systematically compared with multiple algorithms, including Random Forest, Decision-Tree, KNN, LightGBM-NN, Base-Transformer, and LightGBM (Table 2).

Table 2. Performance comparison of density-curve reconstruction across different models (RMSE, MAE, R^2 , Pearson r , and pass rates under error thresholds). The underlined values indicate the best results.

Model	RMSE	MAE	R^2	Pearson r	Within ± 0.02	Within ± 0.05
RF-Transformer	<u>0.0126</u>	<u>0.0079</u>	<u>0.9863</u>	<u>0.9932</u>	<u>92.86</u>	<u>99.34</u>
Random Forest [12]	0.0154	0.0103	0.9796	0.9899	87.42	98.84
Decision-Tree [30]	0.0219	0.0149	0.9585	0.9791	74.97	96.38
KNN [7]	0.0252	0.0179	0.9453	0.9724	67.56	95.29
LightGBM-NN [39]	0.0264	0.0199	0.9399	0.9709	61.39	93.92
Base-Transformer [26]	0.0305	0.0219	0.9196	0.9590	57.90	92.23
LightGBM [40]	0.0315	0.0233	0.9142	0.9671	53.85	90.67

From the numerical metrics, RF-Transformer achieves the best or near-best performance across all evaluation criteria. Its RMSE is 0.0126, and MAE is 0.0079 (Figure 4), with $R^2 = 0.9863$ and Pearson $r = 0.9932$; the proportions of depth points with errors controlled within ± 0.02 g/cm³ [41] and ± 0.05 g/cm³ reach 92.862% and 99.336%, respectively. Compared with the best-performing conventional baseline, RF-Transformer reduces RMSE from 0.0154 to 0.0126 and MAE from 0.0103 to 0.0079, while R^2 and r increase from 0.9796 and 0.9899 to 0.9863 and 0.9932, respectively. These results indicate that, while maintaining robustness, introducing long-sequence modeling can further compress the overall error and improve the consistency of depth-dependent trends. Taking the commonly used engineering threshold of ± 0.02 g/cm³ as an example, RF-Transformer increases the coverage by approximately 5.4% relative to Random Forest and by nearly 39% relative to LightGBM. In terms of deployability, typical baseline models can complete single-well inference in ~ 30 s, while the proposed RF-Transformer takes ~ 1 min under the same protocol. Although slightly slower, this latency is still acceptable for practical single-well density reconstruction.

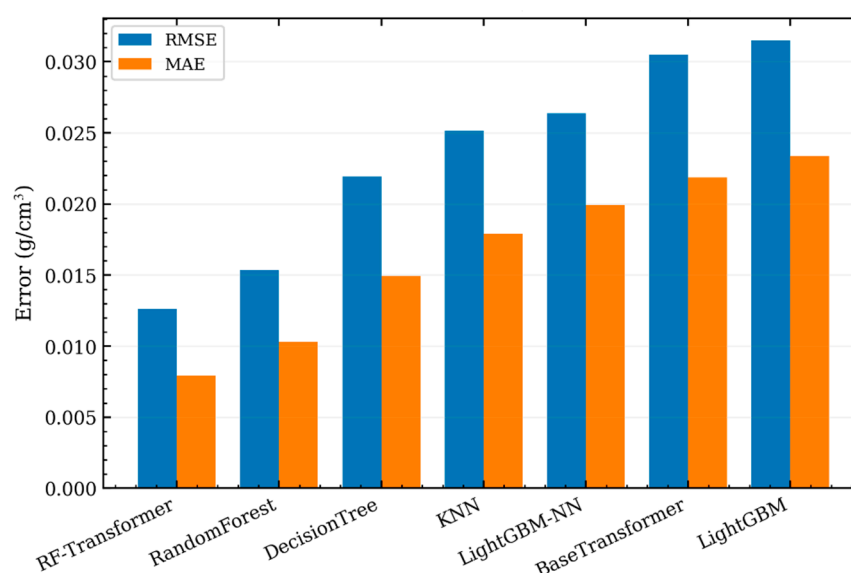


Figure 4. Comparison of DEN reconstruction errors (MAE and RMSE) between RF-Transformer and baseline models. MAE and RMSE are in g/cm³; smaller values indicate lower reconstruction error.

From the perspective of overall trends, all models are able to follow the dominant variation direction of the density curve (Figure 5). However, their differences become pronounced in intervals with frequent oscillations induced by thin interbeds: Random-Forest-

based models tend to exhibit step-like behavior or overly smoothed responses near local extrema, whereas sequence models place greater emphasis on continuity along the depth direction. In comparison, RF-Transformer reproduces peak–trough locations, turning-point slopes, and weak high-frequency perturbations more consistently with the measured curve; meanwhile, it does not introduce excessive oscillations in intervals with strong local noise.

Although the RF-Transformer demonstrates strong performance on the Liang-gaoshan Formation dataset, its accuracy may degrade when transferred to other basins or formations due to domain shift in both geology and acquisition conditions. Typical failure cases include (i) different lithofacies assemblages and diagenetic trends that alter the relationship between density and auxiliary logs (e.g., shale content, pore-fluid type, and compaction regime), (ii) changes in borehole environment and tool calibration (mud system, borehole enlargement, standoff, and logging standards) that distort the measured responses, and (iii) trajectory-related and structural complexities that introduce non-stationary depth dependencies. Importantly, the RF screening step is domain-dependent: the retained feature subset is not guaranteed to remain unchanged across fields. While our final subset is fixed as $K = 6$ (CNL, AC, CAL, AZIM, GR, and SP) for the studied dataset, different geological contexts may yield a different importance ranking and thus a different optimal subset. To assess and improve transferability, we recommend (1) conducting leave-one-field/leave-one-area validation when multi-field data are available, (2) re-running RF screening on the target domain to regenerate the selected subset and its mapping, (3) performing lightweight fine-tuning of the Transformer using a small labeled interval from the target area, and (4) aligning normalization and quality-control rules with the target acquisition standard.

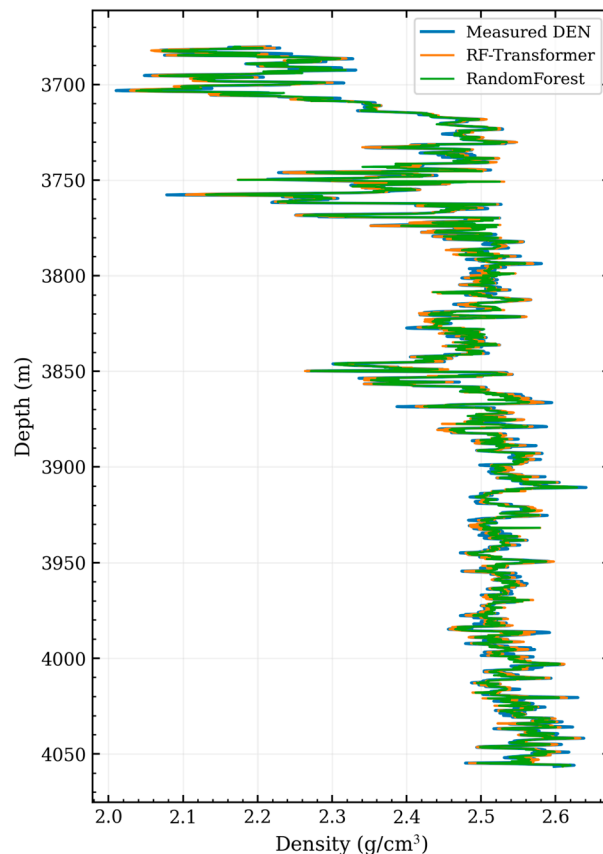


Figure 5. Overlay comparison between the measured density curve and the reconstructed results over the full well interval, illustrating overall trend consistency and deviation levels; used to compare trend consistency and bias across different models over the entire depth range.

4.2. Input-Feature Correlation Analysis Supporting RF Screening

To further explain—at the input level—why the proposed model can stably reconstruct DEN and to provide a verifiable basis for the subsequent RF-based feature screening, the correlation structure among the input features is analyzed. This structure reflects the informational complementarity among different logs, and it also reveals the degree of redundancy as well as potential pathways through which noise may propagate. Note that the correlation analysis is performed over the raw candidate channels to illustrate redundancy/complementarity, whereas the Transformer receives only the RF-retained subset as inputs.

As indicated by the correlation matrix (Figure 6), DEN exhibits stable and relatively strong correlations with several inputs, while a pronounced multicollinearity structure is also present. Specifically, DEN shows strong negative correlations with CNL, AC, and CAL, with the absolute correlation coefficients approaching or exceeding 0.85; moreover, DEN varies consistently with borehole-size-related variables such as CAL, CALX, and CALY. This suggests that porosity responses, acoustic responses, and borehole-environment variations jointly control the dominant sources of density-measurement fluctuations. In contrast, DEN shows moderate positive correlations with trajectory/attitude variables (e.g., AZIM and DEVI) and natural gamma indicators (e.g., GR and GRSL), implying that wellbore trajectory and lithological information provide auxiliary constraints on density variations. If all channels are directly fed into the model without screening, the network will simultaneously receive multiple near-duplicate information groups, which not only increases dimensionality and training instability but may also amplify the influence of borehole or tool effects on attention allocation, thereby weakening cross-well generalization.

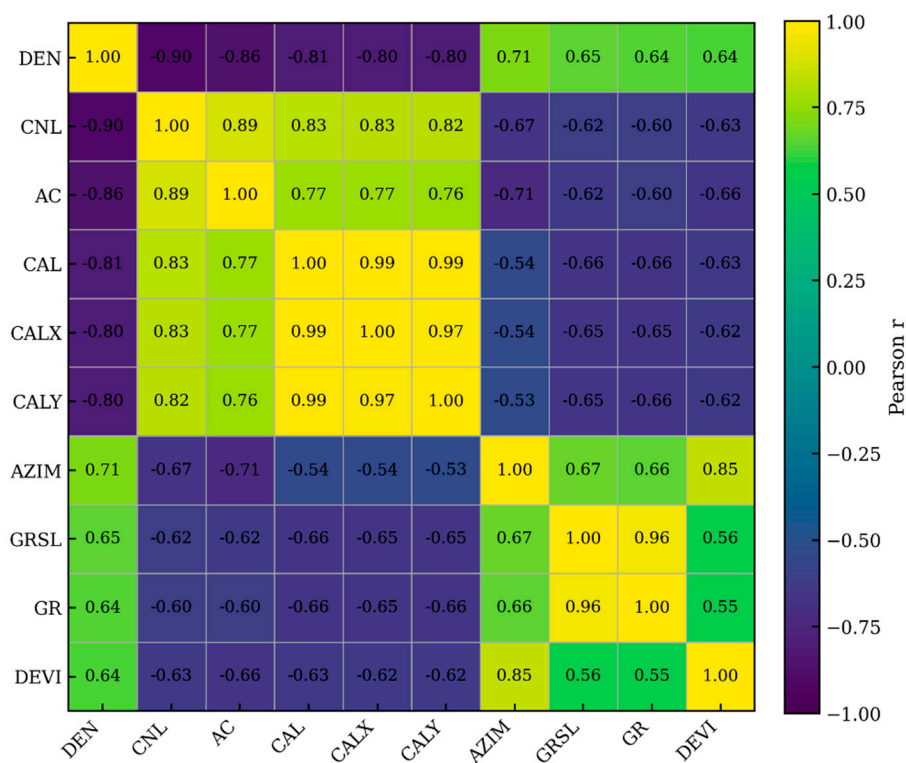


Figure 6. Pearson- r correlation heatmap among multiple well-log curves.

Univariate scatter fitting further corroborates the above findings (Figure 7) and reveals pronounced differences in the explanatory strength of individual inputs with respect to DEN. Specifically, the goodness-of-fit between DEN and CNL, AC, and CAL

reaches $R^2 = 0.7932$, 0.7021 , and 0.6147 , respectively, and all exhibit clear negative trends, indicating that porosity response and acoustic slowness provide the most direct constraints on density, while borehole-size variations can introduce systematic biases through environmental effects. In contrast, GR shows only a moderate association with DEN ($R^2 = 0.4180$), whereas the linear explanatory power of LLS and LLD is weak ($R^2 \approx 0.0847$ and 0.0732 , respectively). This reflects the fact that resistivity is jointly controlled by multiple coupled factors—such as fluid properties, invasion, and clay content—so a monotonic linear constraint on DEN is not stable. These discrepancies imply that DEN reconstruction should not rely on a single strongly correlated channel; instead, high-contribution features should be retained while collinearity redundancy and weakly constraining channels should be suppressed to avoid noise-driven model behavior.

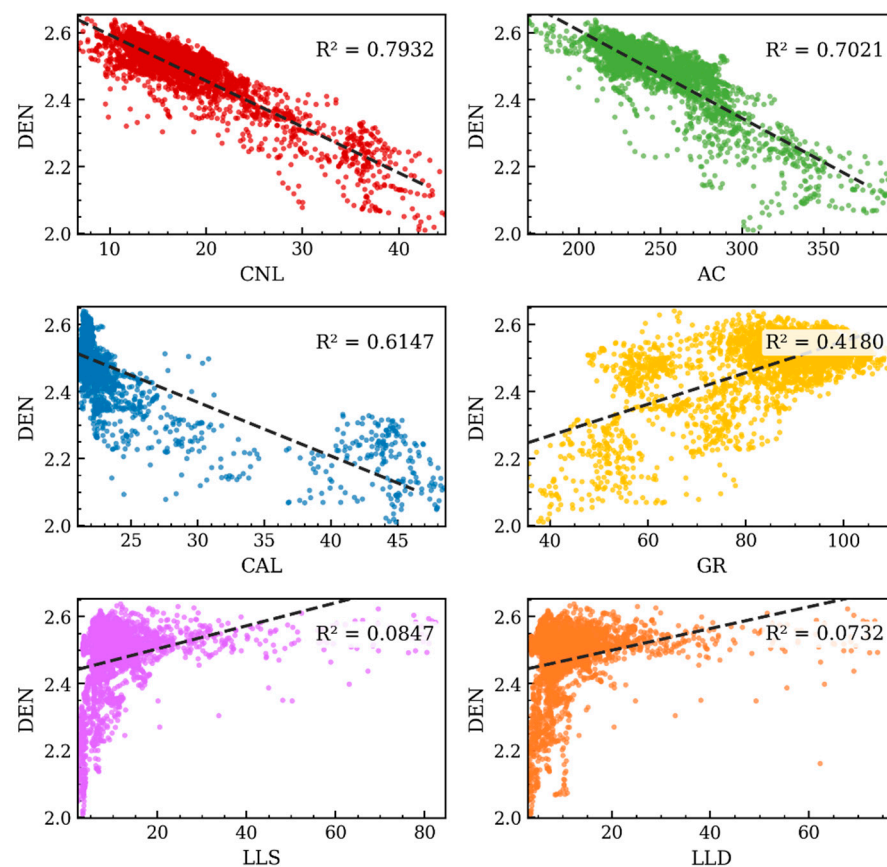


Figure 7. Scatter relationships and linear fits between DEN and key input curves, reporting the correlation strength and the trend direction.

4.3. Local-Interval Detail Fidelity and Abrupt-Boundary Tracking

After the overall metrics and the input-correlation structure have been validated, the model's detail fidelity at fine-scale stratigraphic sequences must be further examined. This is because the engineering usability of a density curve often depends on whether high-frequency fluctuations in thin interbeds, sharp spike-and-fall behaviors, and the phase alignment and amplitude recovery at abrupt boundaries can be accurately reproduced, rather than merely achieving a low full-well average error.

In the local-interval comparison (Figure 8), all models are able to follow the dominant variation direction of the density curve; however, their differences become pronounced in intervals with frequent oscillations induced by thin interbeds. Near local extrema, Random-Forest-based models are more prone to step-like or overly smoothed responses: the peak–trough amplitudes at certain depth points are compressed, and the slopes at turning

points tend to become blunted. In contrast, sequence models emphasize continuity along the depth direction and are more sensitive to phase consistency of high-frequency perturbations; therefore, they show advantages in reproducing peak–trough positions and small-amplitude fluctuations. Across the column-wise comparison, RF-Transformer matches the measured curve more closely in terms of peak–trough locations, turning-point slopes, and weak high-frequency perturbations; meanwhile, it is less likely to introduce excessive oscillations in noisy intervals, thereby achieving more stable detail fidelity in thin-interbedded sections.

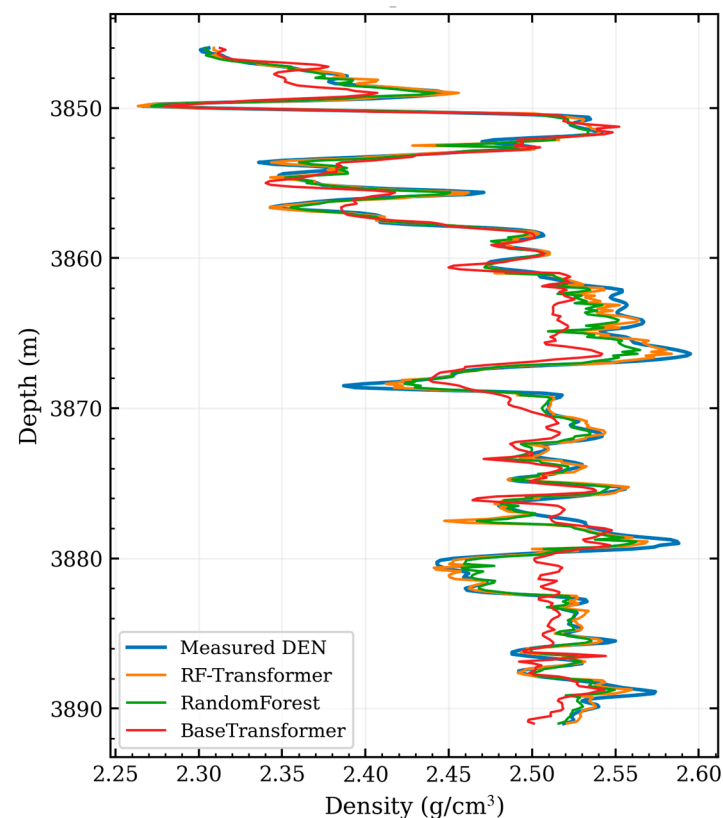


Figure 8. Overlay comparison of density curves in the enlarged interval (3846–3892 m), illustrating the tracking performance at abrupt boundaries and peak–trough locations.

Further multi-model overlays over the interval shown in Figure 9 enable a more direct assessment of detail fidelity and abrupt-boundary tracking among different models. Overall, all models can reproduce the dominant DEN trend; however, marked discrepancies are observed at high-frequency oscillations induced by thin interbeds and at local abrupt transitions. Random Forest, Decision Tree, and LightGBM exhibit, at certain peak–trough positions, either amplitude compression or small-scale jitters caused by local over-fitting, leading to visible deviations from the measured curve around turning points. Base Transformer places greater emphasis on sequence smoothness and thus produces relatively stable overall trends, but they tend to flatten sharp peaks, show insufficient rebound at troughs, or exhibit slight phase lag at steep boundaries, thereby reducing the resolution of fine-scale layer interfaces. By contrast, RF-Transformer achieves a higher overlap with the measured curve at most depth points: it preserves boundary alignment while retaining thin-bed oscillation amplitudes without introducing additional oscillations, demonstrating a more stable advantage in tracking local details and abrupt boundaries.

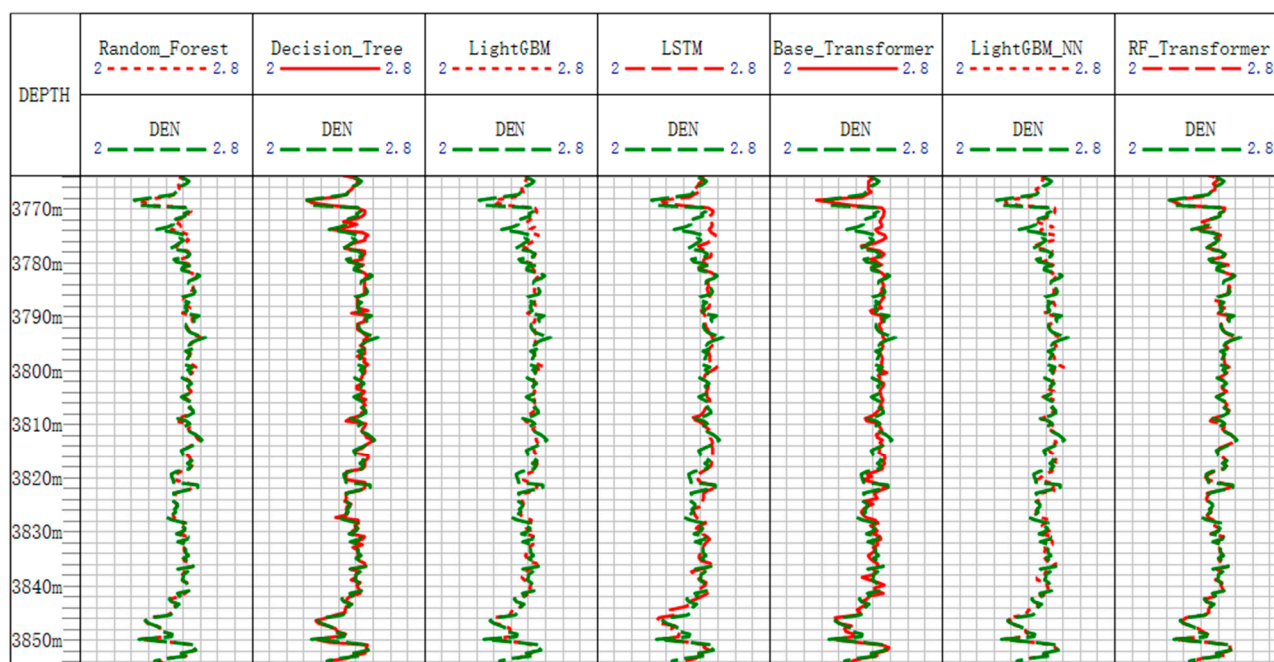


Figure 9. Overlay comparison of the measured DEN curve and reconstructions from multiple models over 3750–3850 m.

4.4. Prediction Consistency and Statistical Distribution of Errors

A single average error is insufficient to characterize model reliability across different density ranges, stratigraphic assemblages, and abnormal borehole conditions. Therefore, prediction consistency, the shape of the error distribution, threshold-coverage efficiency, and stratified residual structures are further analyzed to ensure that the conclusions are both statistically robust and engineering-interpretable.

From the perspective of prediction consistency, the cross-plot relationship between predicted and measured values can directly verify whether the dimensional mapping remains stable over the full range. The scatter cloud of RF-Transformer generally forms a relatively compact band along the 1:1 reference line (Figure 10), indicating that the model does not merely reproduce local trends through smoothed interpolation but instead maintains consistent amplitude characterization across different density levels. The limited deviations are mainly concentrated in extreme ranges or in sample segments with the most drastic variations. Such dispersion is typically associated with abrupt boundaries, borehole-environment perturbations, and imbalanced sample proportions; nevertheless, the overall convergent pattern still supports its cross-interval prediction consistency.

Error-distribution statistics further characterize the robustness of this consistency at an engineering scale. RF-Transformer exhibits a lower median absolute error, a narrower interquartile range, and a more strongly suppressed long tail (Figure 11), indicating that errors are more concentrated across most depth points and that the proportion of outliers is lower. Correspondingly, its cumulative distribution function rises faster in the small-error region (Figure 12), implying that more sample points can be covered under the same error threshold. Consistent with the numerical metrics, under the unified evaluation protocol across the nine wells, RF-Transformer achieves $RMSE = 0.0126 \text{ g/cm}^3$ and $MAE = 0.0079 \text{ g/cm}^3$; moreover, the pass rates under $|\hat{y} - y| \leq 0.02 \text{ g/cm}^3$ and $|\hat{y} - y| \leq 0.05 \text{ g/cm}^3$ reach 92.862% and 99.336%, respectively, demonstrating that the overall convergence of the error distribution and the advantage in threshold coverage hold simultaneously.

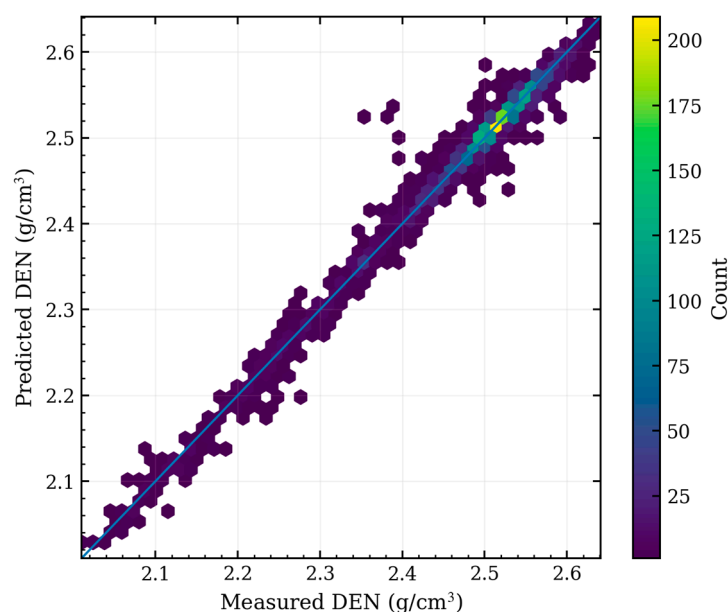


Figure 10. Cross-plot of RF-Transformer predictions versus measured values.

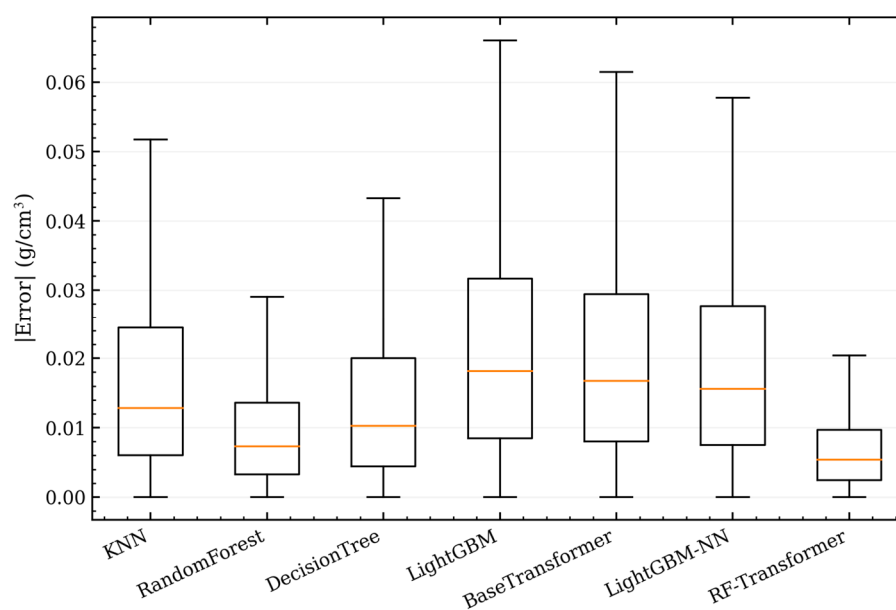


Figure 11. Comparison of the statistical distributions of absolute density-reconstruction errors among different models. The absolute error is $|\text{predicted} - \text{measured}|$ (g/cm³); the box indicates the interquartile range, and the center line indicates the median.

In addition to the overall distribution, the residual structure across different interpretation categories can further reveal whether the sources of error exhibit stratigraphic dependence. The residual dispersion increases noticeably in certain categories (Figure 13), indicating that, under specific lithological combinations, the input logs provide weaker or more unstable constraints on DEN; consequently, variance is more readily amplified, forming the major contributing intervals to long-tail risk. This observation is consistent with the previously identified category-dependent differences in input correlations, and it further suggests that category priors or stratified modeling strategies could be introduced in subsequent work to further compress the residual dispersion for specific categories without sacrificing overall consistency, thereby improving transferability across stratigraphic scenarios and enhancing risk controllability.

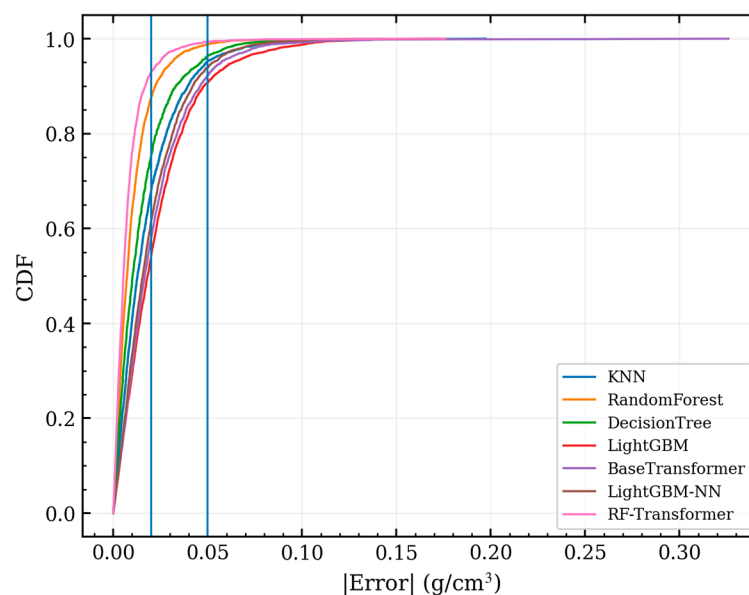


Figure 12. Comparison of the cumulative distribution functions of absolute errors among different models. The x-axis is $|\text{predicted} - \text{measured}| (\text{g}/\text{cm}^3)$, and the y-axis is the cumulative proportion; vertical lines denote the preset error-threshold references.

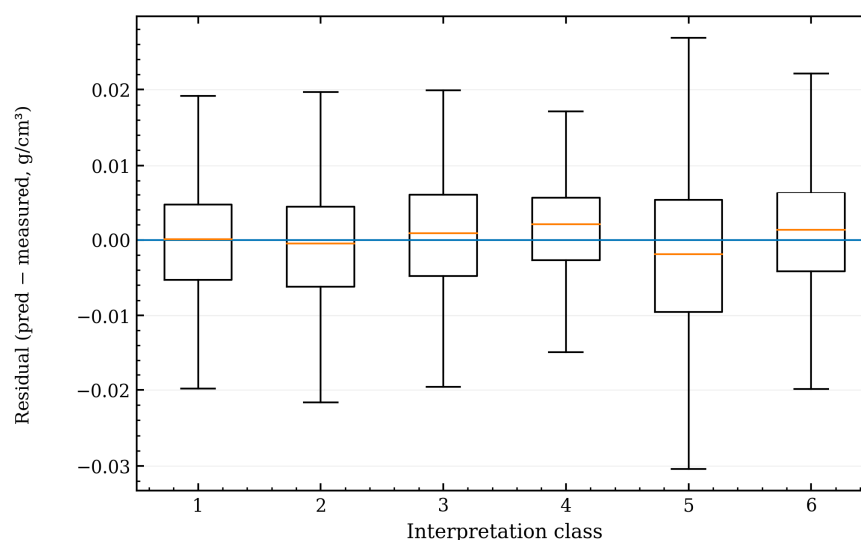


Figure 13. Residual distribution of density reconstruction across different interpretation categories. The residual is defined as $\hat{y} - y (\text{g}/\text{cm}^3)$; each category summarizes the dispersion characteristics of residuals to evaluate category-dependent reconstruction stability.

Recent studies have demonstrated that sequence models enhanced by attention mechanisms can substantially improve the recovery of depth-wise dependencies, especially when the reconstruction needs to preserve both local fluctuations and longer-range correlations. For example, attention-augmented architectures combining temporal convolution and bidirectional recurrence have been further strengthened by introducing lithology indicators as physical constraints, which help stabilize density reconstruction under complex stratigraphic settings and borehole enlargement effects (e.g., TCN-BiGRU with multi-head self-attention and lithology constraints) [23]. In contrast, our results indicate that a complementary improvement can be achieved by addressing the “input-side” issue before sequence learning: Random-Forest screening suppresses redundant/low-contribution channels and fixes an informative subset, while the mask-aware Transformer focuses its attention budget on measured positions and depth-consistent patterns. This design

rationale aligns with the general observation that feature selection and redundancy control can improve robustness and reduce unnecessary variance amplification in well-log prediction tasks [13], and it provides a practical route to stabilize attention allocation when multi-log inputs exhibit strong multicollinearity and acquisition-dependent noise.

At the same time, the literature also highlights that generalization remains sensitive to domain shift and that the “best” input set may vary across fields. Therefore, the RF-selected subset reported in this study should be interpreted as dataset-specific rather than universally fixed; in a new basin or formation, re-running feature screening and recalibrating the mapping is recommended prior to deployment. Moreover, recent generative imputation frameworks (e.g., sequence-based GANs) suggest a viable pathway to represent distributional uncertainty and recover plausible variability under limited observations [30]. Consistent with uncertainty-aware RF imputation practices that use prediction intervals to quantify confidence along depth [13], a key future direction is to augment RF–Transformer with uncertainty quantification (e.g., prediction intervals or calibrated uncertainty scores) and domain adaptation/fine-tuning using small labeled intervals from the target area, so that long-tail errors and category-dependent dispersion can be further compressed while improving transferability across operational conditions.

5. Conclusions

This study addresses density (DEN) log discontinuities induced by borehole conditions and tool/environmental interference by proposing an RF–Transformer workflow that combines Random-Forest-based feature screening with a mask-aware Transformer encoder for depth-sequence reconstruction. Using nine vertical wells (0.125 m sampling) under a well-wise split protocol, the method delivers accurate and trend-consistent DEN reconstruction, improving density-log continuity for multi-log interpretation in complex reservoirs.

Across all compared models, RF–Transformer achieves the best overall performance with RMSE = 0.0126 g/cm³, MAE = 0.0079 g/cm³, R² = 0.9863, and r = 0.9932, and a pass rate of 92.862% within ±0.02 g/cm³. The reconstructed curve better preserves phase and amplitude in intervals with thin interbeds and abrupt spike–fallback patterns, benefiting from RF de-redundancy and mask-constrained attention that strengthens long-range structure while avoiding attention to unmeasured positions.

Several limitations remain. The RF-selected input subset is domain-dependent and may change across basins, lithofacies, or acquisition standards. Residual dispersion is category-dependent (Figure 13), implying higher variance in more heterogeneous lithologies and frequent transition zones, and transferability may degrade under geological and borehole-condition shifts.

Future work will focus on improving transferability and compressing long-tail errors via lithology-aware (stratified) modeling, category priors, and domain adaptation, together with stricter cross-block validation and uncertainty quantification to better control prediction risk under varying operational conditions.

Author Contributions: Conceptualization, J.S. and X.D.; methodology, J.S.; validation, X.D., Y.Z. and P.L.; formal analysis, J.S.; investigation, Y.Z. and P.L.; data curation, X.S. and W.S.; writing—original draft preparation, J.S.; writing—review and editing, X.D. and Y.Z.; supervision, X.D.; project administration, X.D.; funding acquisition, X.D. All authors have read and agreed to the published version of the manuscript.

Funding: This study was supported by the National Natural Science Foundation of China (Young Scientists Fund, No. 42204131), the Heilongjiang Provincial Excellent Young Scientists Fund (No. YQ2023D004), and the National Science and Technology Major Project for New Oil and Gas

Exploration and Development, “Enrichment patterns of coalbed methane and evaluation of geological and engineering sweet spots” (No. 2025ZD1404202).

Data Availability Statement: The raw data supporting the conclusions of this article, together with the source code and scripts used for data processing, feature screening, model training, and evaluation, are publicly available in the authors’ GitHub repository: <https://github.com/CHTjw/> (accessed on 2 February 2026).

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

Abbreviations

DEN	Density log (bulk density)
CNL	Compensated neutron log
AC	Acoustic/Sonic slowness log
CAL	Caliper log
CALX	Caliper log (X-direction component)
CALY	Caliper log (Y-direction component)
GR	Gamma ray log
GRSL	Gamma ray (auxiliary)
SP	Spontaneous potential log
LLD	Deep laterolog resistivity (deep resistivity channel)
LLS	Shallow laterolog resistivity (shallow resistivity channel)
RT	True formation resistivity
RILM	Resistivity channel
RFOC	Resistivity-related channel
CILD	Resistivity-related channel
AZIM	Azimuth (wellbore/tool azimuth)
DEVI	Deviation (well deviation angle)
K	Potassium (spectral gamma-related channel)
TH	Thorium (spectral gamma-related channel)
U	Uranium (spectral gamma-related channel)
KTH	Spectral gamma-related channel

References

1. Asquith, G.B.; Krygowski, D. *Basic Well Log Analysis*, 2nd ed.; American Association of Petroleum Geologists (AAPG): Tulsa, OK, USA, 2004.
2. Schlumberger. *Log Interpretation: Principles/Applications*; Schlumberger Educational Services: Houston, TX, USA, 1989.
3. Gama, P.H.T.; Faria, J.; Sena, J.; Neves, F.; Riffel, V.R.; Perez, L.; Korenchender, A.; Sobreira, M.C.A.; Machado, A.M.C. Imputation in well log data: A benchmark for machine learning methods. *Comput. Geosci.* **2025**, *196*, 105789. <https://doi.org/10.1016/j.cageo.2024.105789>.
4. Garini, S.A.; Shiddiqi, A.M.; Utama, W.; Insani, A.N.F. Filling-well: An effective technique to handle incomplete well-log data for lithology classification using machine learning algorithms. *MethodsX* **2024**, *14*, 103127. <https://doi.org/10.1016/j.mex.2024.103127>.
5. Chen, J.-R.; Yang, R.-Z.; Li, T.-T.; Xu, Y.-D.; Sun, Z.-P. Reconstruction of well-logging data using unsupervised machine learning-based outlier detection techniques (UML-ODTs) under adverse drilling conditions. *Appl. Geophys.* **2025**, *22*, 1–17. <https://doi.org/10.1007/s11770-025-1178-z>.
6. Jiang, C.; Zhang, D.; Chen, S. Handling missing data in well-log curves with a gated graph neural network. *Geophysics* **2023**, *88*, D13–D30. <https://doi.org/10.1190/geo2022-0028.1>.
7. Kim, M.J.; Cho, Y. Imputation of missing values in well log data using k-nearest neighbor collaborative filtering. *Comput. Geosci.* **2024**, *193*, 105712. <https://doi.org/10.1016/j.cageo.2024.105712>.
8. Fan, P.; Deng, R.; Qiu, J.; Zhao, Z.; Wu, S. Well logging curve reconstruction based on kernel ridge regression. *Arab. J. Geosci.* **2021**, *14*, 1559. <https://doi.org/10.1007/s12517-021-07792-y>.
9. Bader, S.; Wu, X.; Fomel, S. Missing log data interpolation and semiautomatic seismic well ties using data matching techniques. *Interpretation* **2019**, *7*, T347–T361. <https://doi.org/10.1190/INT-2018-0044.1>.

10. Mirhashemi, M.; Khojasteh, E.R.; Manaman, N.S.; Makarian, E. Efficient sonic log estimations by geostatistics, empirical petrophysical relations, and their combination: Two case studies from Iranian hydrocarbon reservoirs. *J. Pet. Sci. Eng.* **2022**, *213*, 110384. <https://doi.org/10.1016/j.petrol.2022.110384>.
11. Hallam, A.; Mukherjee, D.; Chassagne, R. Multivariate imputation via chained equations for elastic well log imputation and prediction. *Appl. Comput. Geosci.* **2022**, *14*, 100083. <https://doi.org/10.1016/j.acags.2022.100083>.
12. Mukherjee, B.; Sain, K.W. Missing log prediction using machine learning perspectives: A case study from upper Assam basin. *Earth Sci. Inform.* **2024**, *17*, 3071–3093. <https://doi.org/10.1007/s12145-024-01323-5>.
13. Feng, R.; Grana, D.B. Imputation of missing well log data by random forest and its uncertainty analysis. *Comput. Geosci.* **2021**, *152*, 104763. <https://doi.org/10.1016/j.cageo.2021.104763>.
14. Qiao, L.; Cui, Y.; Jia, Z.; Xiao, K.; Su, H. Missing Well Logs Prediction Based on Hybrid Kernel Extreme Learning Machine Optimized by Bayesian Optimization. *Appl. Sci.* **2022**, *12*, 7838. <https://doi.org/10.3390/app12157838>.
15. Akmal, L.; Pyrcz, M.J. Physics-Based DiscreXucy Modeling for Well Log Imputation. *Math. Geosci.* **2025**, *57*, 1235–1264. <https://doi.org/10.1007/s11004-025-10203-7>.
16. Haritha, D.S. Generation of missing well log data with deep learning: CNN-Bi-LSTM approach. *J. Appl. Geophys.* **2025**, *233*, 105628. <https://doi.org/10.1016/j.jappgeo.2025.105628>.
17. Zhang, D.; Chen, Y.M. Synthetic well logs generation via recurrent neural networks. *Pet. Explor. Dev.* **2018**, *45*, 629–639. [https://doi.org/10.1016/S1876-3804\(18\)30068-5](https://doi.org/10.1016/S1876-3804(18)30068-5).
18. Li, J.G. Digital construction of geophysical well logging curves using the LSTM deep-learning network. *Front. Earth Sci.* **2023**, *10*, 1041807. <https://doi.org/10.3389/feart.2022.1041807>.
19. Zhou, W.; Zhao, H.; Li, X.; Qi, Z.; Lai, F.; Yi, J. Missing well logs reconstruction based on cascaded bidirectional long short-term memory network. *Expert Syst. Appl.* **2025**, *259*, 125270. <https://doi.org/10.1016/j.eswa.2024.125270>.
20. Pham, N.; Wu, X.; Naeini, E.Z. Missing well log prediction using convolutional long short-term memory network. *Geophysics* **2020**, *85*, WA159–WA171. <https://doi.org/10.1190/geo2019-0282.1>.
21. Zhang, L.G. A deep learning multi-module fusion method for well logging curve reconstruction. *Phys. Fluids* **2025**, *37*, 076627. <https://doi.org/10.1063/5.0276398>.
22. Zeng, L.; Ren, W.S. Attention-based bidirectional gated recurrent unit neural networks for well logs prediction and lithology identification. *Neurocomputing* **2020**, *414*, 153–171. <https://doi.org/10.1016/j.neucom.2020.07.026>.
23. Liao, W.; Gao, C.; Fang, J.; Zhao, B.; Zhang, Z. A TCN-BiGRU Density Logging Curve Reconstruction Method Based on Multi-Head Self-Attention Mechanism. *Processes* **2024**, *12*, 1589. <https://doi.org/10.3390/pr12081589>.
24. Fan, X.; Meng, F.; Deng, J.; Semnani, A.; Zhao, P.; Zhang, Q. Transformative reconstruction of missing acoustic well logs using multi-head self-attention BiRNNs. *Geoenergy Sci. Eng.* **2025**, *245*, 213513. <https://doi.org/10.1016/j.geoen.2024.213513>.
25. Wang, J.; Cao, J.; Fu, J.; Xu, H. Missing well logs prediction using deep learning integrated neural network with the self-attention mechanism. *Energy* **2022**, *261*, 125270. <https://doi.org/10.1016/j.energy.2022.125270>.
26. Lin, L.; Wei, H.; Wu, T.; Zhang, P.; Zhong, Z.; Li, C. Missing well-log reconstruction using a sequence self-attention deep-learning framework. *Geophysics* **2023**, *88*, D391–D410. <https://doi.org/10.1190/geo2022-0757.1>.
27. Liu, Q.; Kong, F.L. A method for training while drilling to predict electromagnetic wave logging curves based on long short-term memory neural networks. *Earth Sci. Inform.* **2025**, *18*, 431. <https://doi.org/10.1007/s12145-025-01936-4>.
28. Mulashani, A.K.; Shen, C.; Nkurlu, B.M.; Mkono, C.N.; Kawamala, M. Enhanced group method of data handling (GMDH) for permeability prediction based on the modified Levenberg Marquardt technique from well log data. *Energy* **2022**, *239*, 121915. <https://doi.org/10.1016/j.energy.2021.121915>.
29. Xu, B.; Feng, Z.; Zhou, J.; Shao, R.; Wu, H.; Liu, P.; Tian, H.; Li, W.; Xiao, L. Transfer learning for well logging formation evaluation using similarity weights. *Artif. Intell. Geosci.* **2024**, *5*, 100091. <https://doi.org/10.1016/j.aiig.2024.100091>.
30. Al-Fakih, A.; Koeshidayatullah, A.; Mukerji, T.; Al-Azani, S.; Kaka, S.I. Well log data generation and imputation using sequence based generative adversarial networks. *Sci. Rep.* **2025**, *15*, 11000. <https://doi.org/10.1038/s41598-025-95709-0>.
31. Vaswani, A.; Shazeer, N.; Parmar, N.; Uszkoreit, J.; Jones, L.; Gomez, A.N.; Polosukhin, I. Attention Is All You Need. *Adv. Neural Inf. Process. Syst.* **2017**, *30*, 5998–6008. <https://doi.org/10.48550/arXiv.1706.03762>.
32. Wang, Z.; Wang, Y. Integrating unsupervised learning and transformer for missing logging data prediction. *Geophysics* **2025**, *90*, D85–D100. <https://doi.org/10.1190/geo2024-0558.1>.
33. A.Y.; Koç, E.; Koç, A. Multivariate Time Series Imputation with Transformers. *IEEE Signal Process. Lett.* **2022**, *29*, 2517–2521. <https://ieeexplore.ieee.org/document/9964035>.

34. Han, J.; Deng, Y.; Zheng, B.; Cao, Z. Well logging super-resolution based on fractal interpolation enhanced by BiLSTM-AMPSO. *Geomech. Geophys. Geo-Energy Geo-Resour.* **2025**, *11*, 54. <https://doi.org/10.1007/s40948-025-00969-9>.
35. Qu, F.; Liao, H.; Liu, J.; Wu, T.; Shi, F.; Xu, Y. A novel well log data imputation methods with CGAN and swarm intelligence optimization. *Energy* **2024**, *293*, 130694. <https://doi.org/10.1016/j.energy.2024.130694>.
36. Mukherjee, B.; Sain, K.; Kar, S.; Srivardhan, V. Deep learning-aided simultaneous missing well log prediction in multiple stratigraphic units: A case study from the Bhogpara oil field, Upper Assam, Northeast India. *Earth Sci. Inform.* **2024**, *17*, 4901–4928. <https://doi.org/10.1007/s12145-024-01425-0>.
37. Ali, M. A novel machine learning approach for detecting outliers, rebuilding well logs, and enhancing reservoir characterization. *Nat. Resour. Res.* **2023**, *32*, 1047–1066. <https://doi.org/10.1007/s11053-023-10184-6>.
38. Onalo, D. Dynamic data driven sonic well log model for formation evaluation. *J. Pet. Sci. Eng.* **2019**, *175*, 1049–1062. <https://doi.org/10.1016/j.petrol.2019.01.042>.
39. Garini, S.A.; Shiddiqi, A.M.; Utama, W.; Abduh, M.U.N. Reconstructing Missing Well-Log Data with LightGBM and BiRNN in Oil Fields. In Proceedings of the 2025 International Conference on Smart Computing, IoT and Machine Learning (SIML), Surakarta, Indonesia, 3–4 June 2025. <https://doi.org/10.1109/SIML65326.2025.11081019>.
40. Kumar, K.I.; Tripathi, B.K.; Singh, A. Synthetic well log modeling with light gradient boosting machine for Assam-Arakan Basin, India. *J. Pet. Sci. Eng.* **2022**, *203*, 104679. <https://doi.org/10.1016/j.jappgeo.2022.104697>.
41. Gjerdingen, T.; Hilton, J.; Bounoua, N. Sourceless LWD Porosity Determination: A Fit for Purpose Formation Evaluation with Significant HS&E Benefits. In Proceedings of the SPE Annual Technical Conference and Exhibition, San Antonio, TX, USA, 8–10 October 2012. <https://doi.org/10.2118/159522-MS>.

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.