symmetry

MDPI

Article

A Transformer and LSTM-Based Approach for Blind Well Lithology Prediction

Danyan Xie *, **Zeyang Liu**, **Fuhao Wang** and **Zhenyu Song** *

College of Information Engineering, Taizhou University, Taizhou 225300, China;
lewis0808zy@gmail.com (Z.L.); daowang1024@gmail.com (F.W.)
* Correspondence: xiedanyan@tzu.edu.cn (D.X.); songzhenyu@tzu.edu.cn (Z.S.)

Abstract: Petrographic prediction is crucial in identifying target areas and understanding reservoir lithology in oil and gas exploration. Traditional logging methods often rely on manual interpretation and experiential judgment, which can introduce subjectivity and constraints due to data quality and geological variability. To enhance the precision and efficacy of lithology prediction, this study employed a Savitzky–Golay filter with a symmetric window for anomaly data processing, coupled with a residual temporal convolutional network (ResTCN) model tasked with completing missing logging data segments. A comparative analysis against the support vector regression and random forest regression model revealed that the ResTCN achieves the smallest MAE, at 0.030, and the highest coefficient of determination, at 0.716, which are indicative of its proximity to the ground truth. These methodologies significantly enhance the quality of the training data. Subsequently, a Transformer–long short-term memory (T-LS) model was applied to identify and classify the lithology of unexplored wells. The input layer of the Transformer model follows an embedding-like principle for data preprocessing, while the encoding block encompasses multi-head attention, Add & Norm, and feedforward components, integrating the multi-head attention mechanism. The output layer interfaces with the LSTM layer through dropout. A performance evaluation of the T-LS model against established rocky prediction techniques such as logistic regression, k-nearest neighbor, and random forest demonstrated its superior identification and classification capabilities. Specifically, the T-LS model achieved a precision of 0.88 and a recall of 0.89 across nine distinct lithology features. A Shapley analysis of the T-LS model underscored the utility of amalgamating multiple logging data sources for lithology classification predictions. This advancement partially addresses the challenges associated with imprecise predictions and limited generalization abilities inherent in traditional machine learning and deep learning models applied to lithology identification, and it also helps to optimize oil and gas exploration and development strategies and improve the efficiency of resource extraction.

Keywords: transformer; LSTM; ResTCN; embedding; lithology prediction

check for updates

Citation: Xie, D.; Liu, Z.; Wang, F.; Song, Z. A Transformer and LSTM-Based Approach for Blind Well Lithology Prediction. *Symmetry* **2024**, *16*, 616. https://doi.org/10.3390/sym16050616

Academic Editor: Chong Wang

Received: 7 April 2024
Revised: 6 May 2024
Accepted: 14 May 2024
Published: 16 May 2024

1. Introduction

In oil and gas exploration, lithology prediction is crucial for determining potential investment [1,2]. Forecasting the rock type in the target area assists in constructing the underground trap structure and predicting oil and gas production. It is important to note that larger reservoirs can produce a higher return on investment. Traditional methods used to predict reservoir lithology in oil and gas exploration rely on geology and petrophysics theories, as well as field geological observations and well logging data analysis [3,4]. While these methods have achieved some success, they often require manual interpretation and rely on subjective evaluations, which are limited by factors such as data quality and geological conditions [5–7]. The development of data processing capabilities has resulted in the emergence of machine learning and artificial intelligence methods, such as deep learning models, that aim to improve the accuracy and efficiency of rockiness prediction. These methods can be integrated into traditional methods to create predictive models that combine data from multiple sources, which can better support decision making [8–10]. A "blind well" refers to a well where no prior drilling samples or geological information have been obtained. It is typically used to test the predictive capabilities of rock properties, stratigraphic features, or reservoir characteristics. The method for predicting lithology in blind wells uses deep learning analyses and models well logging data, which enable the prediction of rock types or lithology even in the absence of core data, enhancing the accuracy and efficiency of rock prediction. While mud logging provides valuable lithology information through magma analysis, there are situations where lithology must be inferred from logging records due to the absence of core samples, as core samples may not be available or may not have been obtained from some sections of the well due to various limitations. In these cases, logging data become the primary source for lithologic interpretation. Moreover, logging data offer continuous measurements across the entire wellbore, enabling the thorough analysis of lithologic variations. Common types of logging data include gamma ray, resistivity, sonic, density, neutron, porosity, and permeability data, which are typically presented as curves on logging charts corresponding to well depth [11]. Integrating these logging curves allows for a more comprehensive lithologic characterization, capturing subtle changes that may elude detection through rock chip analysis alone. The analysis and interpretation of these curves can yield detailed insights into formation and reservoir characteristics, aiding in well suitability assessments, the determination of production capacity, the evaluation of hydrocarbon reservoir potential, and making informed decisions regarding well completion or abandonment.

Traditional logging interpretation methods use traditional geology and physics for the interpretation and analysis of logging data. There are many traditional log interpretation methods, such as manually viewing and analyzing the morphology, trends, and interrelationships of logging curves to infer subsurface geologic features. For example, the resistivity, natural gamma radiation, and sonic velocity are used to understand the type of rock, hydrocarbon properties, and reservoir characteristics. Furthermore, the rock type, stratigraphic relationship, and reservoir characteristics can be deduced by observing the trend of change in different logging curves in the vertical direction, drawing the profiles of these logging curves, and combining the results with one's knowledge of stratigraphy. Traditional logging interpretation methods play an important role in the exploration and development stages, and with the development of machine learning and automation technology, these methods are gradually being combined with computer-aided interpretation to improve efficiency and accuracy. Lithology prediction methods based on logging data analysis have been an increasingly popular research topic in the oil and gas exploration field in recent years [12,13]. Researchers have aimed to improve the accuracy of predicting the rock type or lithology in blind wells without core data by combining artificial intelligence techniques with information from logging data. Machine learning methods applied to automated logging can reduce exploration costs and improve prediction accuracy compared to computationally intensive manual logging interpretations. Machine learning algorithms can automatically process large amounts of logging data, which can significantly reduce human resource and time costs compared to manual processing. Machine learning methods can extract valuable features from logging data and identify hidden patterns and correlations. This helps to speed up the exploration process and improve the prediction accuracy. Based on the prediction results of machine learning models, decision makers can make more informed decisions to reduce exploration risks and increase the success rate, reducing unnecessary trial and error and the waste of resources [4,14–16]. These methods focus on the identification of optimal features using unsupervised and supervised machine learning algorithms, as well as the application of automated logging data to achieve reliable lithology prediction and subsequent reservoir characterization. With the continuous updating of deep learning algorithms and the improvement of arithmetic power, more relevant methods are being used for reservoir lithology prediction. These methods include the CNN, recurrent neural network, and LSTM network. CNNs are used to extract complex features from logging data, and LSTM is used to extract vertical spatial relationships from its output characteristics. Finally, the mapping relationship between logging data and lithology type can be established. This model helps in the recognition of the lithology of complex formations [17]. A semi-supervised deep learning framework has been used with a closed-loop CNN and virtual logging labels. Closed-loop CNNs have predictive and generative sub-networks, and this model can be trained directly using seismic attribute data [18]. The spatial and temporal features of the logging data are extracted using a combination of a CNN and LSTM neural networks. A particle swarm optimization algorithm can also be used to determine the optimal hyperparameters for predicting log profiles [19]. The analysis must overcome several obstacles when employing conventional deep learning models. First, log data are often sparse, and the sample distribution is imbalanced. Second, log data can be influenced by noise, outliers, and other quality issues. Finally, the preprocessing and cleaning of data must be undertaken to enhance data quality and model robustness. Deep learning models usually function as black boxes, making it challenging to determine how they arrive at predictions and decisions. Model interpretability is crucial in well logging. Deep learning models may perform well on training sets due to the complexity of reservoir composition, but their ability to generalize to new data may be limited.

To address the above problems, we propose a Transformer- and LSTM-based hybrid approach to identify and classify lithology in blind wells. Using the multi-attention mechanism of the Transformer model and the ability of the LSTM network to capture the temporal spatial features of the lithology sequence, their combination can effectively learn the nonlinear relationship between logging curve data and their correlation in the depth dimension. After numerous experiments, we found that the Transformer–LSTM (T-LS) model outperformed several commonly used models in lithology prediction. To verify the model's generalization ability, we used the T-LS model and random forest (RF) model to predict the lithology of blind wells without core data in the stratum, which showed that the T-LS model had a better generalization ability. The main contributions of this study are as follows:

- A T-LS model is proposed to identify and classify lithology in blind wells. Combining the advantages of the Transformer model and LSTM network, the model effectively learns the nonlinear relationships between logging curve data and their correlation in the depth dimension;
- A nested ResTCN is deployed to address missing data, which can efficiently complete missing content, thereby ensuring the completeness and accuracy of training data;
- Comparative experiments demonstrate the advantages of the T-LS model in terms of several evaluation metrics. The results of neighboring blind well experiments further validate the model's generalization ability.

In this study, the T-LS hybrid model is proposed for the identification and classification of lithologies in blind wells. To address the issue of raw logging data quality, this study employed a Savitzky–Golay filter to remove anomalous data during data preprocessing and equalize data samples using a genetic algorithm-based sample interpolation method. This study also employed a nested residual convolutional network (ResTCN) to fill in the missing signals of some logging data, enabling more complete and accurate training data to be obtained. These methods effectively solved the data quality problem and improved the training effect of the model. Through model comparison experiments, it was found that the T-LS model outperformed other commonly used models in lithology prediction. In order to verify the generalization ability of the model, this study also predicted the lithology of blind wells with no core data in the formation and compared its predictions with those of the RF model. The results show that the T-LS model has a good generalization ability and can perform well on unknown data. Finally, a Shapley analysis was used for the T-LS model, and it was concluded that for the multi-sample lithology classification and prediction task, the more information contained in the logging data and the more eigenvalues they have, the better the accuracy of the classification and prediction. Methods for the fusion of multiple logging information can lead to a better understanding of the properties of subsurface rocks and fluids, the improved assessment of the production capacity and recoverable reserves in oil and gas reservoirs, an increased optimization of drilling and production decisions, and a reduction in exploration and development risks.

The rest of this paper is organized as follows. Section 2 focuses on related methods and techniques, including the ResTCN model, the proposed T-LS model, and Savitzky–Golay filtering. Section 3 describes the case studies, including data description and processing, and parameter settings, and discusses the experimental results. Section 4 presents our conclusions and discusses future work.

2. Methodology

2.1. ResTCN Model

To enhance the accuracy of model prediction, we used the ResTCN model to fill in the missing data from the raw logging data [20,21]. This improves the model's robustness and overfitting resistance while also alleviating the problem of gradient vanishing. The ResTCN combines the concepts of residual networks and temporal convolutional networks, allowing it to capture temporal relationships in time-series data. The ResTCN model is composed of multiple residual blocks, each containing a convolutional layer that extracts features from the time-series data. Additionally, each block has a residual connection that adds the original input to its output for information transfer. With the stacking of multiple residual blocks, the model can learn multiple levels of temporal features. The residual links enable it to learn residual representations, which are the differences between layer inputs and outputs. The ResTCN efficiently captures and models long-term dependencies in the input sequence by propagating residuals through the network.

In assuming that the current output is $E(X)$ and the input of the previous layer is x, the output, after passing through the residual structure, is $F(x) = E(X) + x$. The residual is $E(x) = F(X) - x$. The output of the current layer is $F(x)$ when $x = 0$, which is the original neural network structure. This causes the output of the current neural network layer to be x, i.e., the cancellation of the current neural network layer when $F(x) = 0$, and the structure of the residual unit is shown in Figure 1.

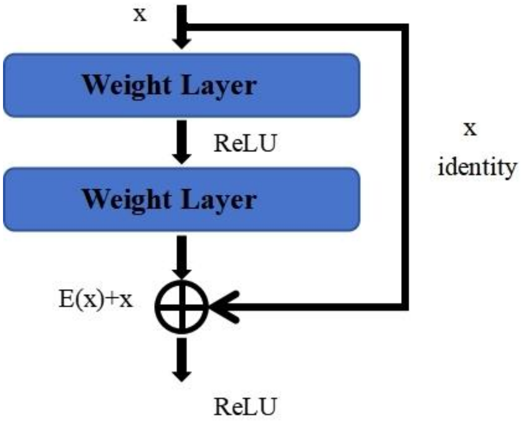
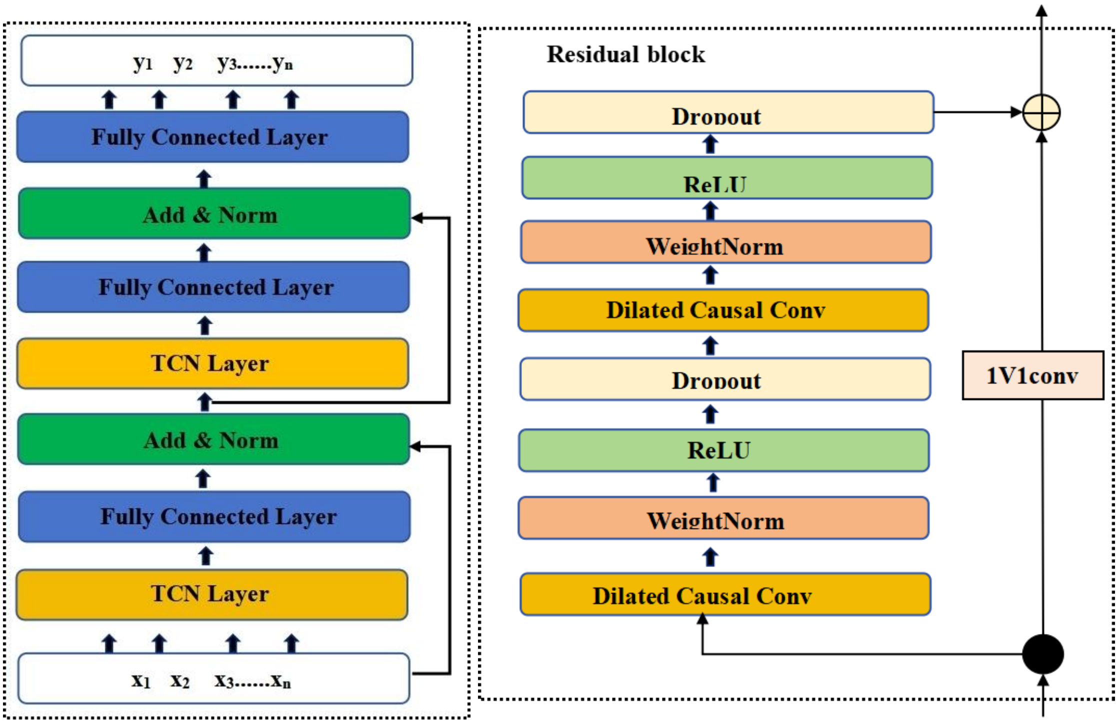


Figure 1. The structure of the residual unit.

Figure 2a shows the ResTCN model structure. Each of the two network blocks receives data that go through temporal convolutional layers, followed by a fully connected neural network layer and a residual structure. An input sequence X with length T is input into the ResTCN. The temporal convolutional layers consist of a set of convolutional kernels, W_1 and b_1. The input sequence X is convolved with convolutional kernel W_1, and bias b_1 is added. The first temporal convolutional layer generates output $H1$ using an activation function, such as ReLU, to introduce nonlinearity. Residual connections add connections to the output $H1$ of each time convolutional layer through the addition of the input sequence X to the output $H1$. This allows information to be jump-connected in the network, helping to solve the gradient vanishing problem.



(**a**) ResTCN model. (**b**) Residual block.

Figure 2. Architectural descriptions of the ResTCN model and residual block.

Figure 2b displays the structure of a residual block. Residual connectivity in the ResTCN aids the model in learning dependencies on different timescales, and guarantees training stability with increasing depth. The residual block's left side includes causal inflated convolution, dilated causal conv, a regularization layer weightnorm, tandem activation function ReLU, and dropout. The model's right side is concatenated with a residual link, and such modules are repeated twice. This forms a short-circuit connection, where the output of the current layer is passed to the next layer and added directly to the input. The use of residual blocks can improve the training and performance of deep neural networks by alleviating the problem of gradient vanishing through a direct flow of information from the input to output.

2.2. *Transformer–LSTM Model*

The proposed approach for identifying and classifying blind well lithology utilizes the T-LS model, which merges two distinct neural network architectures: a Transformer and LSTM. While traditional Transformer models excel in natural language processing tasks, they may encounter challenges when handling time-series data due to their limitations in sequence length and long-term dependency modeling. In contrast, a LSTM model, a classical recurrent neural network, adeptly processes time-series data and efficiently captures long-term dependencies. The T-LS model endeavors to comprehensively capture the temporal-spatial features of lithological sequences by leveraging the multi-head self-attention mechanism of the Transformer alongside the sequence modeling prowess of LSTM [22,23]. This combination facilitates the effective learning of nonlinear relationships between logging profile data and their correlations in the depth dimension. In amalgamating the strengths of both architectures, the T-LS model offers a robust framework for lithology classification. Geological exploration data and logging curve sequences are processed to delineate different lithology classes, leveraging learned geological features and patterns to predict lithology labels for each location based on the input sequences. The lithology classification prediction function is thereby realized. Figure 3 illustrates the structure of the T-LS model, which consists of five layers corresponding to specific operations.

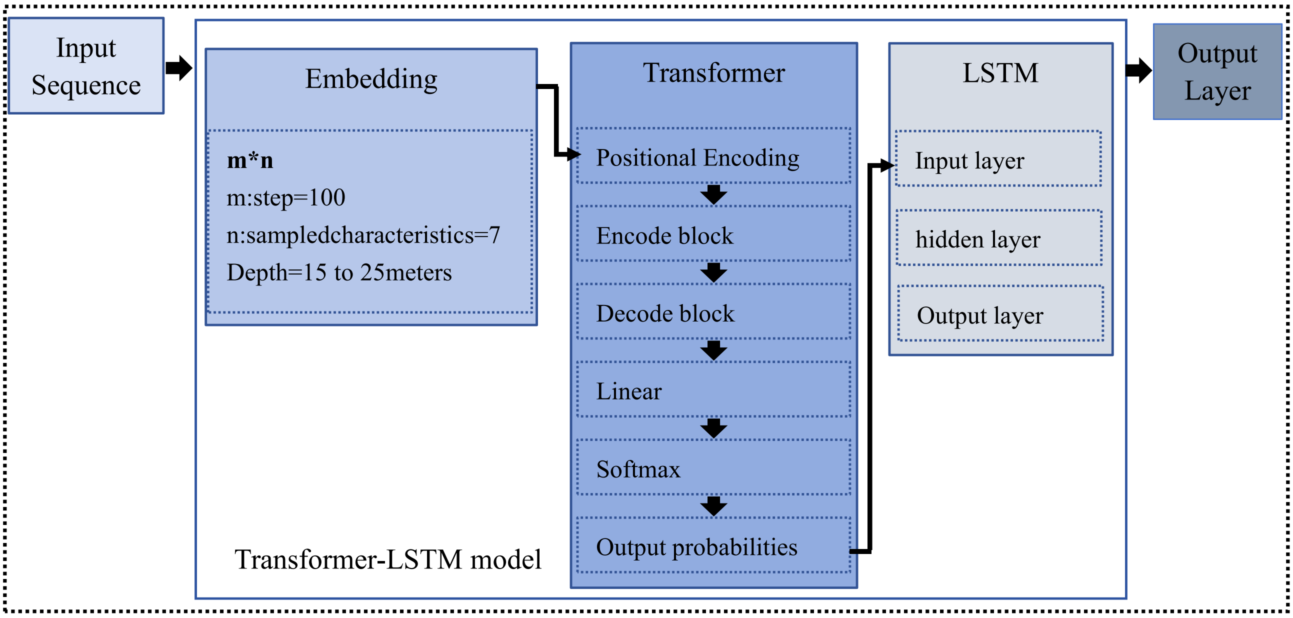


Figure 3. The proposed T-LS hybrid model.

The five layers are as follows:

- Input: the input layer is similar to the traditional Transformer. The input sequence is encoded through the embedding layer to obtain a vector sequence.
- Embedding layer: Traditional machine learning and deep learning models often struggle to accurately identify and generalize lithology due to their inability to properly link logging curves and depth changes in the data [24,25]. To address this issue, it is necessary to thoroughly investigate the correlations within the logging data. Scholars have studied logging curve data so as to obtain information on the correlation between logging curves and changes in depth. Correlation is evident when the depth interval is between 15 m and 25 m [26,27]. The realization principle of the Transformer model makes it is suitable for data with sequential relationships. In considering the differences between the logging dataset and the natural language processing dataset in terms of data types, quantities, and structures, an embedding-like principle was adopted to deal with the logging data to transform the time problem into a depth problem. With reference to the interrelated nature exhibited by the logging data in the depth interval of 15–25 m, the embedding step length was selected to be 100, and every 15.24 m was taken as a feature matrix, i.e., $m = 100$, in the embedding-like matrix. Serial number 0 to 99 is a piece, Serial number 1 to 100 is a piece, and so on. After the above data manipulation, several matrices composed of continuous depth logging features can be obtained. After processing through the embedding layer, the data can be transformed into an $m * n$ matrix, where m is the length of the input sequence, and n is the embedding dimension.
- Transformer Model: To enhance the accuracy of the deep learning model for lithology identification and classification in blind wells, we employed the Transformer model from natural language processing. This model integrates self-attention and attention mechanisms to discern data features within the current sequence, focusing on information from various locations within the input data. Comprising a positional encoding layer, encoder, and decoder, all interconnected through attention mechanisms, the Transformer model is adept at processing sequential data. The encoder and decoder modules are composed of multiple blocks, incorporating a multi-head self-attention mechanism layer and a point feedforward neural network layer, respectively. The input sequences undergo processing through the Transformer encoder, leveraging the self-attention mechanism to extract features and learn sequence element representations. This model effectively addresses the challenges associated with poor

classification conducted by conventional machine learning models that stem from the heterogeneous nature of subsurface reservoirs and the nonlinear relationships inherent in logging data. Notably, the model achieves efficient and accurate classification on blind well datasets. Figure 4 provides a visual representation of the Transformer model and illustrates its components, including inputs, the encoder, the decoder, and outputs.

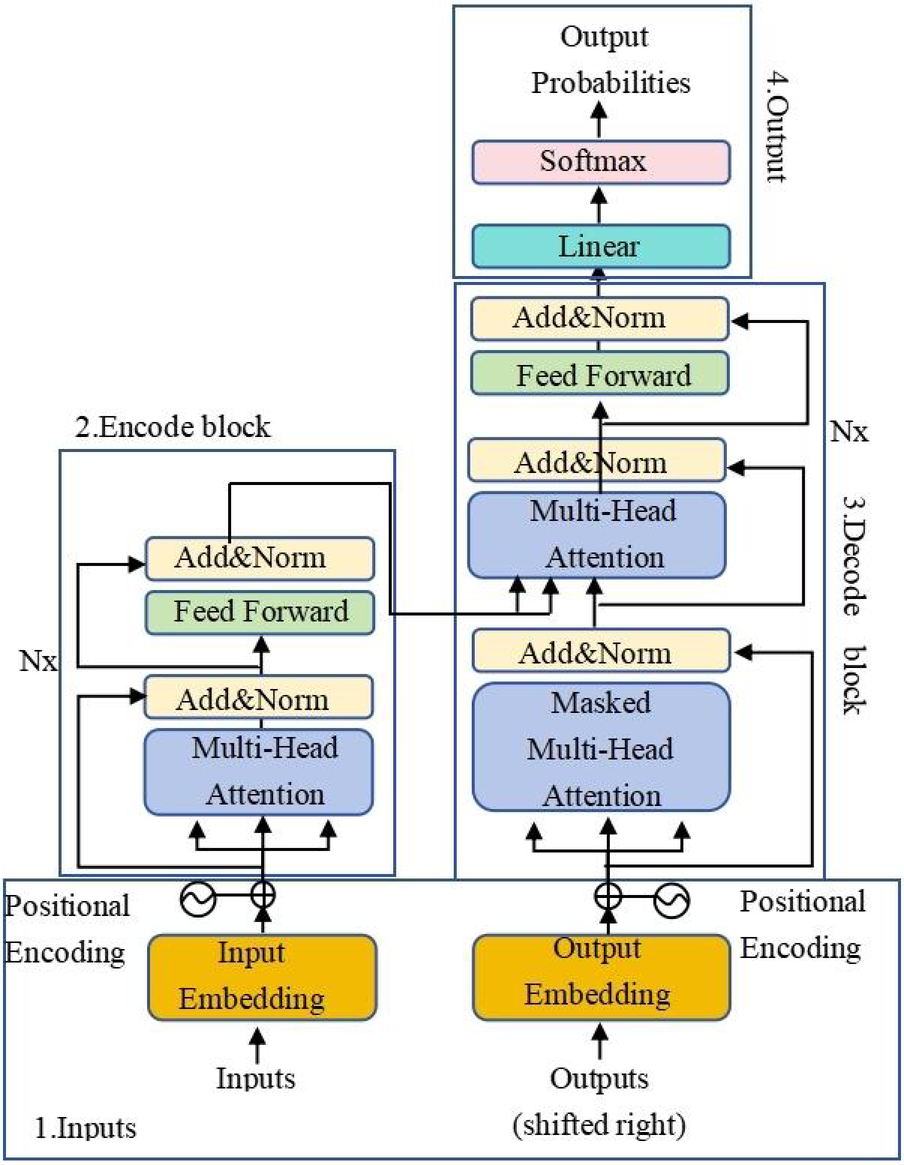


Figure 4. The Transformer model.

- LSTM layer: The LSTM layer processes the sequence data and captures long-term dependencies, using the output of the transformer encoder as its input. It maintains an internal memory called the cell state, and has forget, input, and output gates to control the updating and use of the cell state. This enables it to selectively remember and forget the information of input sequences, and to better handle long-term dependencies. The structure of LSTM is shown in Figure 5.

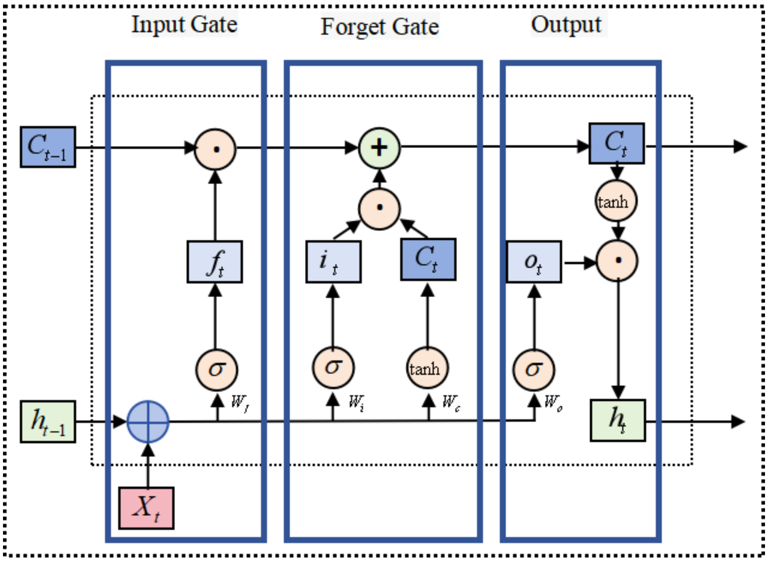


Figure 5. The structure of LSTM.

- Output layer: For classification and regression tasks, a softmax layer is added after the LSTM layer. The T-LS model combines the parallel computation and attention mechanism of the Transformer with the sequence and long-term dependency modeling capabilities of LSTM.

2.3. Savitzky–Golay

During data acquisition, certain anomalies are generated in the training and testing datasets, which can have a significant impact on the accuracy of the neural network model if they are not processed before being entered. It is important to apply a technique to the datasets to ensure optimal classification accuracy [28,29]. In traditional log data processing, a smoothing filter is used to remove noise, reduce fluctuations or mutations, and extract trends and characteristics of the formation. Smoothing filters include the moving average, median, weighted moving average, Savitzky–Golay, and Gaussian smoothing filter. We comprehensively compared several filtering methods, and selected Savitzky–Golay filtering for anomalous data processing [30–32]. The Savitzky–Golay method, developed by Abraham Savitzky and Marcel J. E. Golay in the 1960s [33], is widely used for data smoothing and derivation in signal processing and data analysis. The central idea of the Savitzky–Golay filter is to estimate the smoothed value by fitting a polynomial over the local neighborhood of the signal. The use of a symmetry window can help to maintain the symmetry or local symmetry of the signal and reduce the bias introduced during the fitting process. It helps to extract useful information from noisy data. The method fits a polynomial function to a small window of neighboring data points, and uses the coefficients of the polynomial to estimate the smoothed or derived values. This approach is based on polynomial fitting:

$$y[n] = \sum_{i=-M}^{M} C_i x[n+i], \tag{1}$$

where $y[n]$ is the filtered output value, $x[n]$ is the current data point of the input signal, and M is the size of the filtering window. The Savitzky–Golay filter preserves the overall trend and characteristics of the data and is more effective in smoothing continuous signals.

2.4. Normalization

Logging data may contain curves of varying lengths. To enable a comprehensive analysis of the multiple measurement curves in the selected dataset, normalization can be utilized to convert the measurements of different curves into a standard normal distribution with the same scale, allowing comparisons. Logging data collected using various instruments often differ significantly in magnitude. If the data are used directly for model training, indicators with higher values will have a more significant impact on the analysis. An effective data standardization method can significantly reduce the range of raw data, ensuring that data values have the same order of magnitude, which balances their roles in model training [34]. Commonly used normalization methods include $Min - Max$ and the Z-score [35,36]. We applied Z-score data normalization to the original logging data. The set of sequences for X was $X = \{x_1, x_2, x_3, \ldots, x_n\}$, from which a new set of sequences was generated after normalization, $Y = \{y_1, y_2, y_3, \ldots, y_n\}$, with mean 0 and variance 1. The calculation is as follows:

$$y_i = \frac{x_i - \overline{x}}{s}, \tag{2}$$

where $\overline{x}$ and s are defined as

$$\overline{x} = \frac{1}{n}\sum_{i=1}^{n} x_i \ \ and \ \ s = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(x_i - \overline{x})^2}. \tag{3}$$

In Equation (3), $\overline{x}$ and s denote the respective mean and standard deviation of the original data. The data processed using Z-score normalization conform to a standard normal distribution, which accelerates gradient descent and training.

2.5. Simulated Genetic-Based Interpolation of Sample Data

Imbalanced categories in training samples can cause a model to misclassify and favor larger categories, leading to a reduced recognition accuracy [37]. To address this, we standardized the number of logging data points for each petrophysical phase in each well using sample data interpolation. This created a standard sample layer tensor for network model training and testing. To generate new samples from small classes, a simulated genetic method was employed, which involved calculating the actual distance between each pair of objects (x_i, x_j) in the small-category sample:

$$d(x_i, x_j) = \sqrt{(x_i - x_j)^T s^{-1}(x_i - x_j)}, \tag{4}$$

where s^{-1} is the covariance matrix, in which the concrete objects were ranked and divided into two sets, and b_1 and b_2 were used as biparent nodes to generate new objects. From each set, we selected one concrete object, x_m from b_1 and x_n from b_2, and calculated their average to create a new object, x_{new}, for the minority class. The process was repeated until all instances in both groups were included. If the number of specific objects in the small-category samples remained significantly less than in the large-category samples, x_{new} was utilized as a biparental node in the next generation. This continued until the numbers of objects in each category were approximately equal.

3. Case Studies

3.1. Data Description

The dataset comprised 10 wells from the Cornwall Grove Gas Field, a natural gas reservoir situated between Brown County and Nemaha County in eastern Kansas, USA. The basin is an ancient sedimentary formation primarily composed of Lower Cambrian to Ordovician rocks. Of the 10 wells, two are missing PE values, and the rest have complete data. There are 4150 data samples from the 10 wells, including 3232 complete data samples from 8 of the wells and 918 data samples with missing PE values from 2 of the wells. In the ResTCN complementation task, the training set and validation set were divided according to a 7:1 ratio. In the T-LS model lithology prediction classification task, the training set and validation set were divided according to an 8:2 ratio. Two additional neighboring blind wells without lithology labels were used as the test set. The reservoir is primarily composed of sandstone and shale formations. Shale gas refers to the natural gas confined within the minute pores and fractures of shale formations. Alongside their high gas content, shale formations exhibit a notable gas storage capacity. The rock's low permeability impedes gas flow, leading to its retention within the shale matrix. The abundance of micro-fractures and nanopores further enhances the storage capacity by offering supplementary storage sites for gas molecules. The logging curve data include natural gamma (GR), depth (ILD log10), rock density (PE), neutron and density porosity (DeltaPHI), mean neutron and density porosity (PHIND), and two geologically constrained variables: the non-marine and marine metrics (NM_M) and relative position information (PELPOS). Table 1 shows example sample data and shows only a very small portion of the dataset.

The logging curve dataset comprises two main lithological classes: sandstone (SS), including coarse sandstone (CSiS), fine sandstone (FSiS), and marine sand shale (SiSh); and carbonate rock (MS), including Wacker limestone (WS), dolomite (D), mudstone (MS), muddy grey grained limestone (PS), and foliated algal tuff (BS). These classes are not discrete, but gradually merge. Mislabeling may occur among adjacent lithofacies, so it is important to use these codes accurately. Table 2 provides lithological name abbreviations and numeric codes. Table 3 shows some of the raw logging data for the two wells, ALEXANDER D and KIMZEY A, that have missing PE values.

Table 1. Selected datasets from the Cornwall Grove natural gas reservoir in Kansas.

Well Name	Depth	GR	ILD _log10	Delta PHI	PHIND	PE	NM_M	RELPOS
STUART	2808	66.276	0.63	3.3	10.65	3.591	1	1
STUART	2808.5	77.252	0.585	6.5	11.95	3.341	1	0.978
STUART	2809	82.899	0.566	9.4	13.6	3.064	1	0.956
STUART	2809.5	80.671	0.593	9.5	13.25	2.977	1	0.933
STUART	2810	75.971	0.638	8.7	12.35	3.02	1	0.911
STUART	2810.5	73.955	0.667	6.9	12.25	3.086	1	0.889

Table 2. Comparison of abbreviations and numerical codes for lithological names.

Lithology Abridge	SS	CSiS	FSiS	SiSh	MS	WS	D	PS	BS
Numeric Code	1	2	3	4	5	6	7	8	9

Table 3. Selected logging datasets with missing PE values.

Facies	Fm	Well Name	Depth	GR	ILD _log10	Delta PHI	PHIND	PE	NM_M	RELPOS
3	A1 SH	ALE	2887.5	88.71	0.612	6.7	10.605	/	1	1
3	A1 SH	ALE	2888	92.71	0.583	11	12.515	/	1	0.974
3	A1 SH	ALE	2888.5	94.54	0.579	12	13.41	/	1	0.949
3	A1 SH	ALE	2889	95.31	0.579	11.5	13.75	/	1	0.923
3	A1 SH	ALE	2889.5	93.79	0.572	10.3	13.405	/	1	0.897

3.2. Data Preprocessing

The collected logging data have varying curve lengths and anomalies that require data cleaning. The number of rock samples in the dataset varies, and when the difference is significant, this can affect the results of the lithology prediction. Therefore, data preprocessing is necessary. The logging curve segment was processed using the Savitzky–Golay method introduced in Section 2.3, which involves using a sliding window, selecting a set of data points within the window, and estimating the smoothed value at the center of the window using a least squares polynomial fit. The window is then slid to the next position, and the fitting process is repeated until the entire data series is covered. Here, the outliers were handled by adjusting the size of the window; the larger the window, the better the outliers are handled. However, too large a window can also affect the other data, so it is crucial to use an appropriate window value. The results are shown in Figure 6, which displays the difference graph of the effect of one of the logging curve segments in the dataset before and after the filtering process. The logging curve segment before filtering is blue, while the orange curve shows the shape of the curve after processing. The method effectively removes sharp points on the original curve, resulting in a much smoother curve.

Data preprocessing involves analyzing the sample balance, which is crucial to avoiding bias in training results. Unbalanced samples can cause the model to learn better for majority categories, resulting in lower recall and higher misclassification rates for minority categories. Additionally, common evaluation metrics such as accuracy may be misleading due to evaluation bias. The model's tendency to predict the majority category due to its large sample size may result in high accuracy but poor performance for the minority category. In addition, the model may learn features associated with the majority category and ignore those associated with the minority category, leading to a decrease in its ability to judge the minority category. Overfitting occurs when the model is trained on a limited number of samples from the minority category, causing it to fail in generalizing to new samples [38,39].

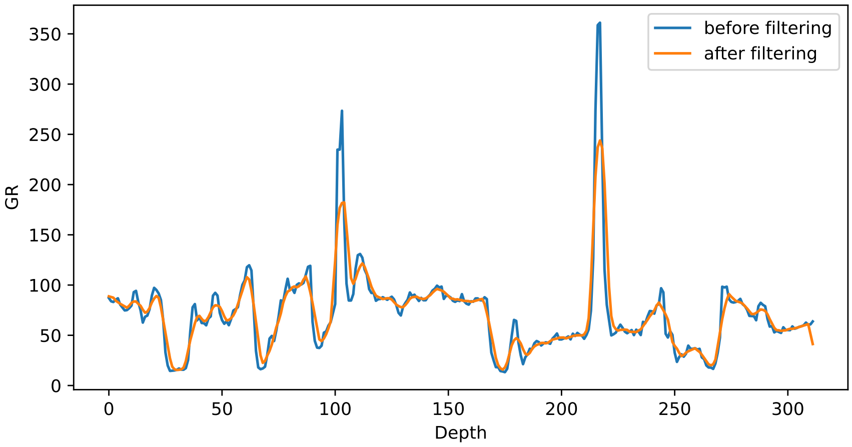
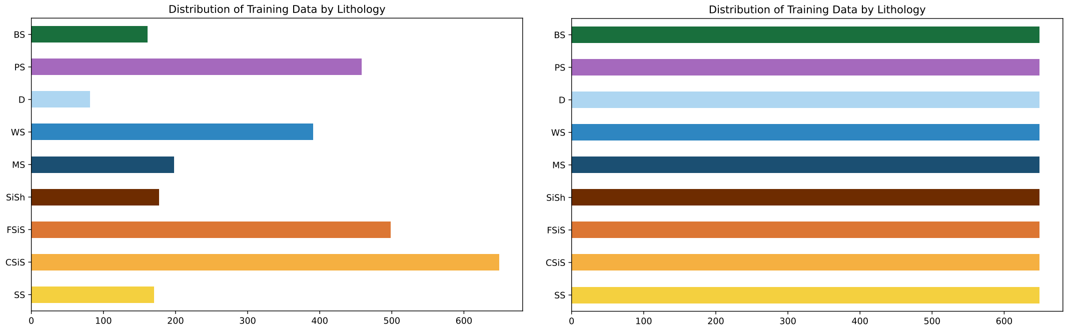


Figure 6. Comparison of logging curves before and after filtering.

To mitigate the aforementioned issues, we generated a histogram illustrating the distribution of training samples, depicted in Figure 7a. This revealed an imbalance in sample counts across different lithology categories, with PS, WS, FSiS, and CSiS having higher sample counts, and D having significantly fewer samples. To address this imbalance, we employed a sample interpolation method based on genetic algorithms, as outlined in Section 2.4. This method involved augmenting the existing data samples until a more balanced distribution was achieved, thereby increasing the sample counts for certain lithology categories. Following the equalization process, the nine lithologies had approximately 5841 samples. Figure 7b illustrates the histogram, showcasing the distribution of training samples after equalization.



(**a**) Original dataset.

(**b**) After data equalization.

Figure 7. Histograms of the number of training samples.

Interrelationships between sample features are crucial in data analysis and machine learning, as they can provide valuable information about the structure, relevance, and importance of the data, and can significantly impact the performance and explanatory power of the model. The interrelationships between the sample features were analyzed, as shown in Figure 8. There were discrete relationships between different samples, such as, GR and ILD _log10, Delta PHI, PHIND, and PE. No obvious signs of multiple covariance between the features were found, which proves that each of the samples is independent. If two features covary strongly, covariance likely exists, and so only one feature should be selected.

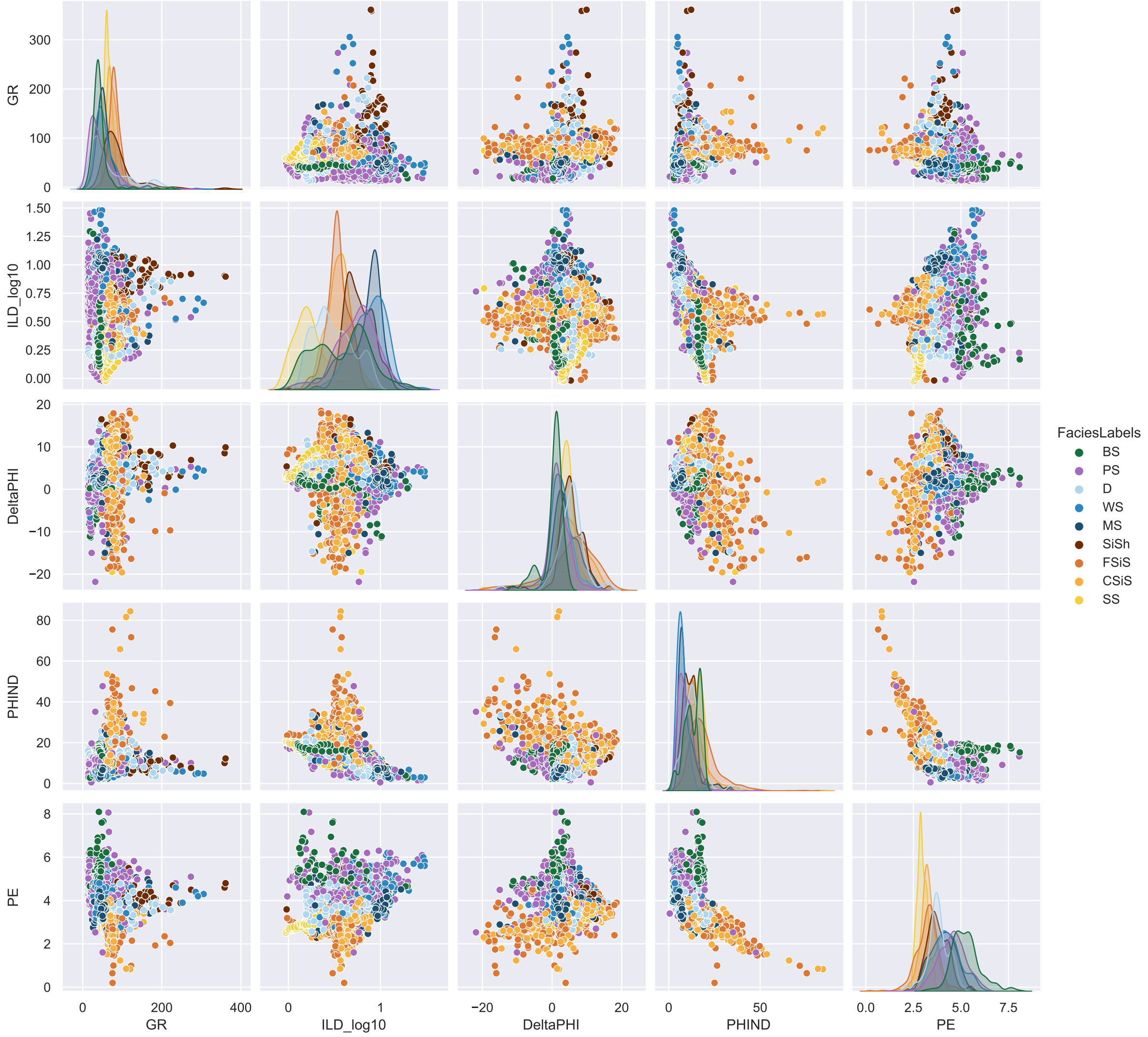


Figure 8. Plots of correlations between sample features.

3.3. Parameter Settings

The following models were involved in this study. The ResTCN accomplished the task of missing data completion. The T-LS, LR, KNN, DT, RF, GB, LSVM, MSVM, and BNB models accomplished the task of lithology classification prediction. The hyperparameters were selected as shown in Table 4.

Table 4. Model hyperparameter selection.

Model	Parameters
ResTCN	$Discard\ probability = 0.3$, $Epoch = 50$, $Batch\ size = 128$, $Learning\ rate = 0.00001$
T-LS	$Hidden\ size = 32$, $Layer = 3$, $Class = 9$, $Batch\ size = 128$, $Epoch = 50$, $Learning\ rate = 0.0001$
LR	$C = 1.0$, $Iteration = 1000$, $Penalty = L2$
KNN	$Leaf\ size = 30$, $Neighbors = 10$
DT	$Min\_leaf = 1$, $Min\_split = 2$
RF	$Min\_leaf = 1$, $Min\_split = 2$, $Estimator = 100$
GB	$Learning\ rate = 0.1$, $Loss$: $deviance$, $Estimator = 100$, $Random\ state = 42$
LSVM	$Cost(c) = 1.0$, $Kernel$: $Linear$
MSVM	$Cost(c) = 1.0$, $Kernel$: $Radial\ basis\ function$
BNB	$\alpha = 1.0$, $Binarize = 0.0$

3.3.1. ResTCN Model Hyperparameter Selection

We gathered seismic, logging, and core data from blind wells to create a dataset for model training and validation. During oil exploration, logging instruments may not accurately detect all logging curve data due to the complexity of the underground reservoir structure and signal transmission. Using machine learning to complete missing logging curve data can improve the accuracy of identifying and classifying the lithology of blind wells. The ResTCN model with selected hyperparameters, including 16 convolution kernels, a discard probability of 0.3, 64 neurons, and a ReLU activation function with 12 regularization, was used to complete the missing data. The model was trained for 50 epochs using the Adam optimizer with a batch size of 128. The learning rate decay mechanism was adopted with an initial learning rate of 0.001. The learning rate was multiplied by 0.5 every 10 consecutive epochs if the mean absolute error (MAE) did not decrease, and the minimum learning rate was set to 0.00001. Table 4 shows the hyperparameter selection.

3.3.2. Transformer–LSTM Model Hyperparameter Selection

The Transformer model requires the logging curve data to be sliced in the depth direction. To achieve this, the input data were processed using the embedding-like logging curve data in Figure 4. The final shape of the input data had three parts: the number of data blocks, the slicing step size m, and the dimension of the input data features. The sampling step size hyperparameter was set to 100, which corresponds to 100 sampling points. The feature matrix had a size of 15.24 m. The Transformer model was constructed using an encoder and padding_mask, and the LSTM network was connected through a dropout function. The model's hyperparameters were set as follows: the LSTM layer had a hidden state of 32 dimensions, and the three-head auto-attention network layer had Quer_dim and Value_dim of 32 dimensions. The discard probability was set to 0.2 using sdrop. This was combined with a residual structure that used the ReLU activation function and introduced l2 regularization and a fully connected layer with a tanh activation function. The fully connected network layer was wrapped by a TimeDistributed wrapper with a softmax activation function. The training process involved randomly selecting 128 groups of training data from the dataset, with a batch size of 128, and repeating this process for 50 training epochs. A cross-entropy loss function was employed, with an initial learning rate of 0.001. A learning rate decay mechanism was added, and its learning rate was multiplied by 0.5 when the test set's loss function value did not decrease for 10 consecutive times. The minimum learning rate was set to 0.0001.

3.4. Evaluation Metrics

The ResTCN model employed two metrics, the MAE and coefficient of determination (R^2), to measure its capacity to complete missing log curve data. The MAE is the average absolute difference between the predicted and true values, as shown in Equation (5), where m is the number of data points, y_i is the predicted value, and $h(x_i)$ is the true value. All data points were given equal weight in the calculation of the MAE. A smaller MAE value indicates a reduced average discrepancy between the predicted and true values of the model. R^2 shows the proportion of changes in the dependent variable that can be explained by the independent variable, and is calculated as shown in Equation (6), where m is the number of data points, y_i is the true value, $\overline{y}$ is the average of the true value, and p_i is the predicted value. A smaller MAE indicates a smaller average difference between the predicted and true values of the model. R^2 expresses the proportion of changes in the dependent variable that can be explained by the independent variable, and ranges from 0 to 1. In Equation (6), y_i is the true value, $\overline{y}$ is the average of the true value, and p_i is the predicted value. A value closer to 1 indicates a better explanation of the observed data variability, while a value closer to 0 indicates a poorer explanation.

$$MAE(X,h) = \frac{1}{m}\sum_{i=1}^{m}|h(x_i) - y_i|, \tag{5}$$

$$R^2 \equiv 1 - \frac{\sum_{i=1}^{m}(y_i - \overline{y})^2}{\sum_{i=1}^{m}(y_i - p_i)^2}. \tag{6}$$

Lithological prediction involves assessing the predictive ability of machine learning classification algorithms or models using precision, recall, and $F1$-score. Precision is the proportion of samples correctly predicted to be in the positive class. Precision measures the accuracy of a classifier's positive predictions. High precision indicates that a model makes fewer errors in predicting negative instances as positive instances, meaning that the model is more accurate in predicting positive instances. Precision is calculated using Equation (7). In Table 5, true positive (TP) and false positive (FP) represent the respective numbers of positive instances that are correctly and incorrectly predicted as positive, and true negative (TN) and false negative (FN), respectively, represent the number of positive instances that are correctly and incorrectly predicted as negative.

$$Precision = TP/(TP + FP), \tag{7}$$

$$Recall = TP/(TP + FN). \tag{8}$$

Table 5. Confusion matrix representation.

	Predicted Positive	**Predicted Negative**
Actual Positive	TP	FN
Actual Negative	FP	TN

Recall is the proportion of true positives that the classifier correctly identifies. It measures the rate at which the classifier correctly identifies positive-class samples. A high recall indicates that the model is better at identifying positive examples. Recall is calculated using Equation (8). The model's ability to identify lithology is evaluated based on its accuracy and completeness, which quantitatively reflect its generalization ability. The $F1$-score, which combines precision and recall, is used to balance the trade-off between the two. A higher $F1$-score indicates higher precision and recall. It can comprehensively assess a model's performance, and is calculated as shown in Equation (9).

$$F1 = 2 * (Precision * Recall)/(Precision + Recall). \tag{9}$$

3.5. Experimental Results and Discussion

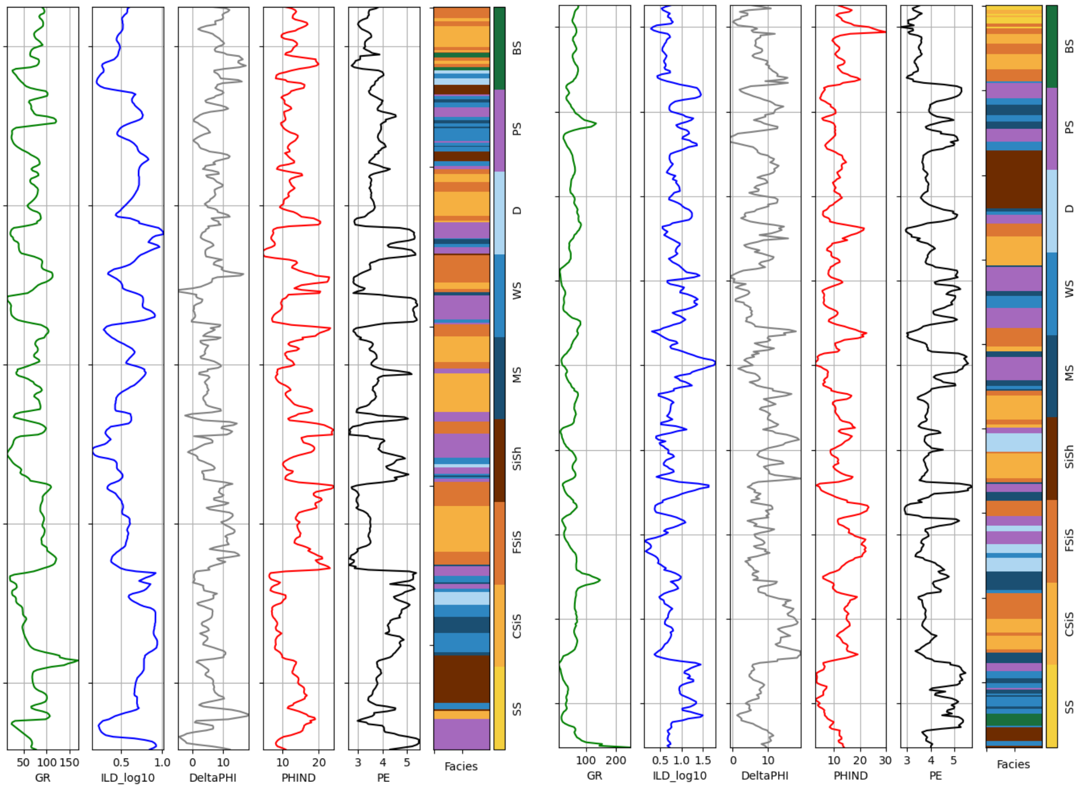
3.5.1. Discussion of Using the ResTCN Model to Complete Missing Data

The ResTCN model (Section 2.1) was used to complete the logging curves of rock density PE for wells with missing logging data, i.e., ALEXANDER D and KIMZEY A. The model was trained, validated, and tested using six of the remaining eight wells for training, one for validation, and one for testing. After testing, we complemented the density logging curves of the rocks in the two missing wells. To demonstrate the accuracy of the ResTCN model in completing missing logging curves, we conducted comparative experiments using support vector regression (SVR) and random forest regression (RFR) models, whose results are presented in Table 6.

Table 6. Experimental results of density comparison of missing rocks.

Model Name	MAE	R^2
SVR	0.0480	0.473
RFR	0.0347	0.667
ResTCN	0.0300	0.716

Among the three models, the ResTCN had the smallest MAE, indicating that its results were closest to the true values, and the largest R^2 value, meaning that it best explained the variability of the observed data. The ResTCN has been shown to be more effective than other deep learning models in completing missing logging curve data. After testing the model, the logging curves for rock density PE were completed for wells with missing data (ALEXANDER D and KIMZEY A), as shown in Figure 9, where black curves represent the completed data. The use of the ResTCN model in this study provided several advantages due to the nested residual structure and fully connected layers of the TCN model. This allows for the improved extraction of nonlinear relationships between input features, and the acquisition of correlations between the same features in the depth dimension. Therefore, the ResTCN model is more effective in completing missing logging curve data.



(**a**) ALEXANDER. (**b**) KIMZEY.

Figure 9. Logging curves for ALEXANDER D and KIMZEY complementary PE values.

3.5.2. Comparison of Predictive Modeling of Lithology in Blind Wells

We utilized the logging dataset from the Hugoton and Panoma Fields in Kansas, USA, following compensation for missing data, as discussed previously. The dataset was partitioned into training, validation, and test sets. The training and validation sets comprised 10 exploration wells from this dataset. The missing density curves in these wells were filled using the ResTCN model. Among these, eight wells were allocated to the training set, and two constituted the validation set. The test set comprised two neighboring blind wells from this dataset, i.e., the STUART and CRAWFORD wells.

We employed the T-LS model (Section 2.2) for blind well lithology classification. To evaluate our model, we applied a number of machine learning and deep learning models to the same dataset. These eight models included logistic regression (LR), k-nearest neighbors (KNN), RF, decision tree (DT), gradient boosting (GB), linear support vector machine (LSVM), multiclass support vector machine (MSVM), and Bernoulli naive Bayes (BNB) models. The hyperparameter selection of all models is summarized in Table 4. Precision, recall, and *F*1-score were employed to evaluate the lithology prediction capabilities of the models, with results as shown in Table 7. Based on these results, the RF and T-LS models demonstrate superior performances in the precision, recall, and *F*1-score. The T-LS model outperformed RF across most samples, except for a few rocky samples such as CSiS and FSiS. The overall evaluation value of the T-LS model is approximately 0.88, which surpasses that of the RF model, which is at 0.74.

Table 7. Classification and identification evaluation table of nine models for nine lithology types.

Model	Appraise Value	SS	CSiS	FSiS	SiSh	MS	WS	D	PS	BS	Total
T-LS	Precision	**0.93**	0.70	**0.77**	**0.97**	**0.94**	**0.9**	**0.95**	**0.86**	**0.95**	**0.88**
	Recall	**0.99**	0.73	0.69	**0.99**	**0.96**	**0.85**	**0.98**	**0.81**	**0.97**	**0.89**
	*F*1-score	**0.96**	0.72	0.73	**0.98**	**0.95**	**0.87**	**0.96**	**0.83**	**0.96**	**0.88**
LR	Precision	0.64	0.59	0.69	0.54	0.11	0.46	0.85	0.54	0.82	0.57
	Recall	0.62	0.7	0.57	0.54	0.02	0.55	0.55	0.69	0.56	0.58
	*F*1-score	0.63	0.64	0.62	0.54	0.04	0.5	0.67	0.61	0.67	0.57
KNN	Precision	0.65	0.72	0.74	0.72	0.63	0.59	0.93	0.64	0.81	0.69
	Recall	0.65	0.76	0.75	0.78	0.63	0.59	0.7	0.61	0.81	0.69
	*F*1-score	0.65	0.74	0.74	0.75	0.63	0.59	0.8	0.63	0.81	0.69
DT	Precision	0.59	0.69	0.66	0.65	0.49	0.63	0.64	0.67	1	0.66
	Recall	0.62	0.59	0.78	0.59	0.47	0.63	0.8	0.68	0.84	0.66
	*F*1-score	0.6	0.63	0.72	0.62	0.48	0.63	0.71	0.67	0.92	0.66
RF	Precision	0.84	**0.75**	0.74	0.75	0.6	0.65	0.94	0.7	0.93	0.74
	Recall	0.73	**0.8**	**0.76**	0.73	0.56	0.67	0.8	0.69	0.88	0.74
	*F*1-score	0.78	**0.78**	**0.75**	0.74	0.58	0.66	0.86	0.69	0.9	0.74
GB	Precision	0.67	0.68	0.72	0.67	0.62	0.58	0.81	0.64	0.94	0.68
	Recall	0.65	0.73	0.7	0.65	0.47	0.63	0.65	0.65	0.91	0.68
	*F*1-score	0.66	0.7	0.71	0.66	0.53	0.6	0.72	0.64	0.92	0.67
LSVM	Precision	0.65	0.61	0.67	0.57	0.00	0.44	1	0.54	0.84	0.57
	Recall	0.67	0.71	0.55	0.62	0.00	0.61	0.45	0.61	0.66	0.58
	*F*1-score	0.66	0.65	0.61	0.6	0.00	0.51	0.62	0.57	0.74	0.57
MSVM	Precision	0.7	0.65	0.66	0.68	0.33	0.5	1	0.59	1	0.63
	Recall	0.6	0.72	0.65	0.62	0.05	0.64	0.65	0.75	0.59	0.63
	*F*1-score	0.65	0.68	0.65	0.65	0.08	0.56	0.79	0.66	0.75	0.62
BNB	Precision	0.5	0.57	0.52	0.39	0.00	0.4	0.47	0.5	0.7	0.48
	Recall	0.35	0.45	0.74	0.38	0.00	0.58	0.45	0.54	0.44	0.49
	*F*1-score	0.41	0.5	0.61	0.38	0.00	0.47	0.46	0.52	0.54	0.47

To visually represent the overall performance and imbalance of the models, we analyzed them using a categorical confusion matrix heatmap, as depicted in Figure 10, and through its analysis, it was observed that the models exhibited varying degrees of effectiveness in the lithology classification of blind wells. LR, LSVM, MSVM, and BNB demonstrated lower degrees of effectiveness, as evidenced by low the precision, recall, and *F*1-scores, particularly for LSVM and BNB, which failed to identify MS lithology altogether. T-LS and RF showed superior overall identifications, well distinguishing the nine lithology types, with T-LS performing marginally better than RF. The T-LS model was compared with traditional machine learning methods and the RF model, with the former proving more effective. The lithology classification results on ALEXANDER D wells using the T-LS and RF models are illustrated in Figure 11. Both models demonstrate commendable classification predictions for the SS, CSIS, FSIS, SISH, and BS lithology types. Notably, for the thinner SISH layer, RF performed slightly better than T-LS. However, in the classification of MS and PS lithology types, T-LS outperformed RF. While both models effectively differentiated sandstones and silicates, there remains room for improvement in the internal differentiation of these lithologies.

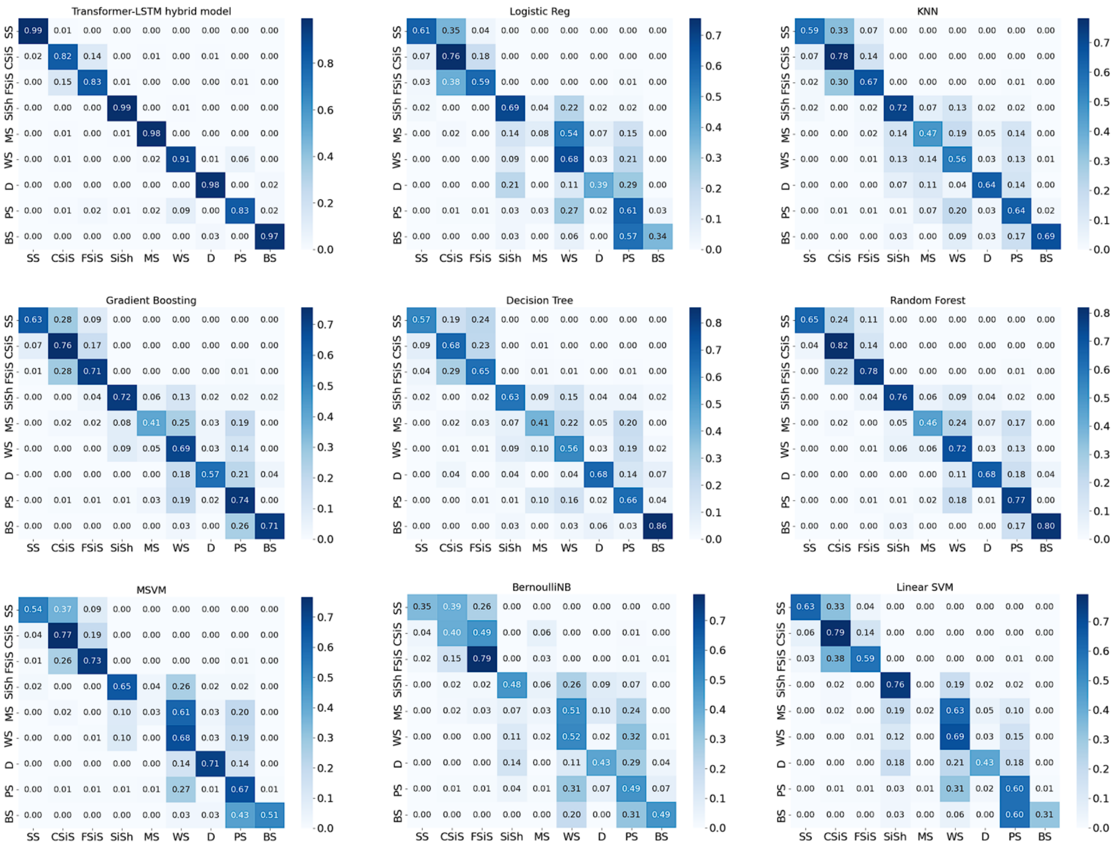
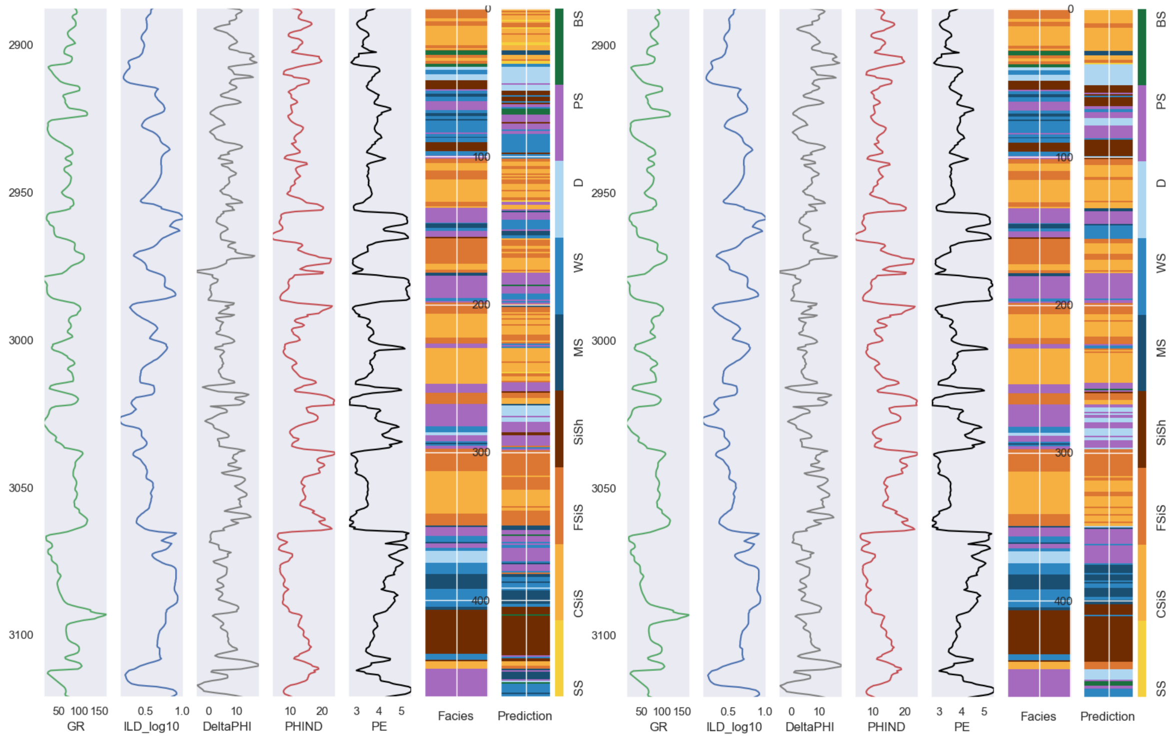


Figure 10. Heat map of confusion matrices for nine models.

In summary, traditional methods often overlook the long-term spatial correlation between logging curve data and lithology labels, thus missing a crucial foundation for enhancing lithology classification accuracy. The T-LS model replaces the recurrent neural network structure with a multi-head self-attention mechanism, enabling a more effective extraction of the long-term spatial correlation between logging curve data and lithology labels, which significantly improves its classification effectiveness. To assess the generalization ability of the model, two adjacent blind wells, STUART and CRAWFORD, were selected for evaluation. The results of the blind well lithology prediction using the T-LS and RF models are depicted in Figure 12. In the comparative experiments detailed in this paper, the lithology prediction results of the T-LS and RF models for adjacent unmarked blind wells exhibit close proximity, particularly in their ability to accurately classify sandstone. However, for locations where a mixture of several lithology types may be present, the identification effectiveness of both models was average. Overall, the T-LS model demonstrated a certain degree of generalization ability, showcasing its potential for application across diverse geological contexts.

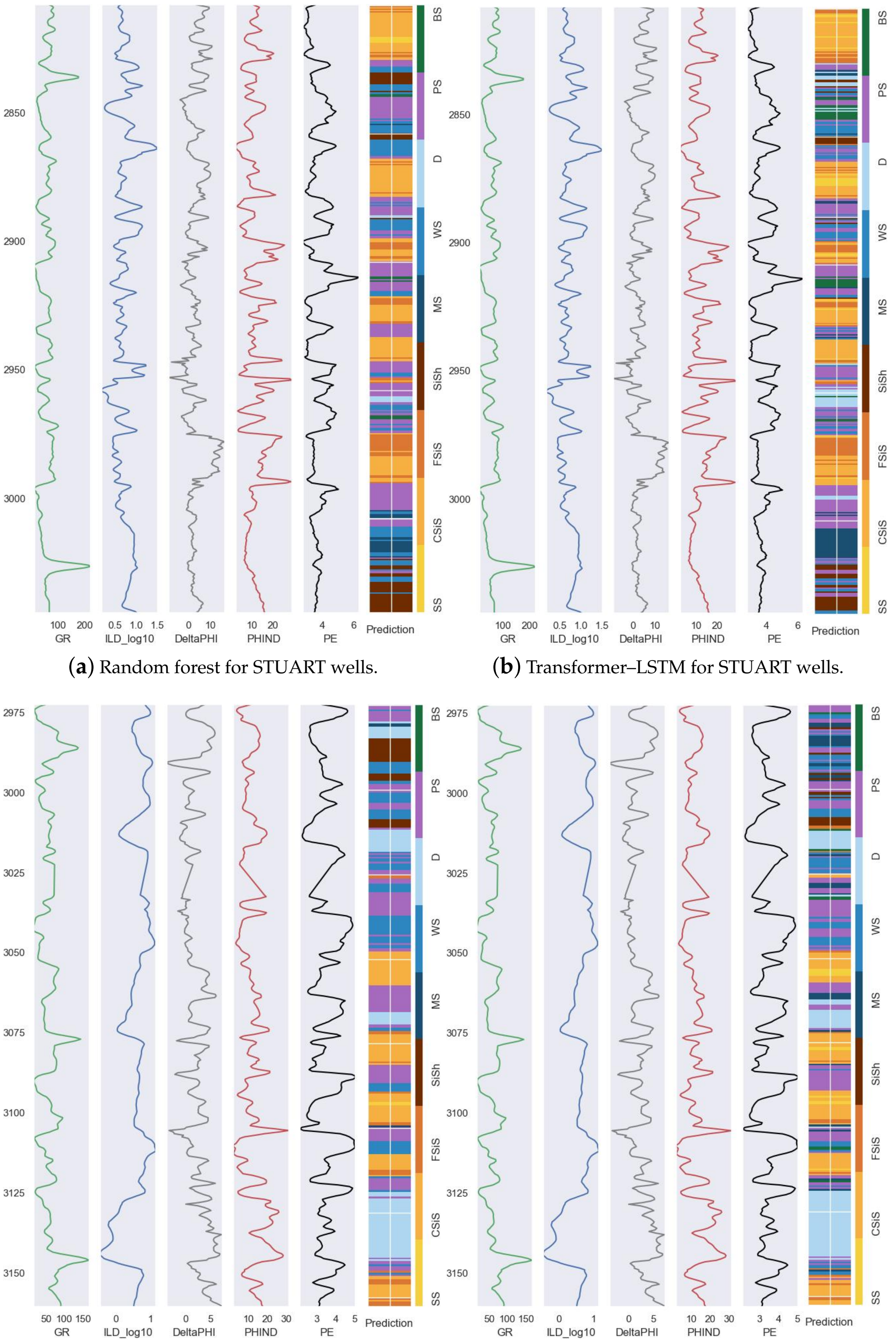


(**a**) Transformer–LSTM model. (**b**) Random forest model.

Figure 11. Plots of lithological classification results for ALEXANDER D wells using the T-LS model and random forest models.

3.5.3. Shapley Analysis of the T-LS Model

The use of Shapley analysis scatter plots can help researchers understand the extent to which a model contributes to each logging data feature. Figure 13 shows the Shapley analysis scatter plot for the identification of nine lithology types using the T-LS model, where the horizontal coordinates indicate the Shapley values, the vertical coordinates indicate the names of the features, and each point represents a sample instance. The six features selected were GR, ILD_log10, DeltaPHI, PHIND, PE, and NM_M, and the names of the nine lithology types are shown in Table 2. A positive Shapley value means that the feature contributes more positively to the prediction results, while a negative Shapley value means that it contributes more negatively. The darker the color, the larger the value of the feature. The lighter the color, the smaller the value of the feature. By observing the change in color coding, the degree of influence of different feature values on the prediction results can be understood. Among all the features, NM_M, i.e., terrestrial or marine stratigraphy, has the greatest influence on lithology prediction, while other features have different influences on lithology prediction for different categories. For example, for SS formations, PHIND and DeltaPHI have a greater impact, while for CSIS formations, GR and PHIND have a greater impact, and DeltaPHI has a smaller impact. In the multi-sample lithology classification prediction task undertaken as part of the comprehensive analysis, the more information and feature values contained in the logging data, the better the classification prediction accuracy. Multi-information fusion logging data analysis can provide a better understanding of the nature of subsurface rocks and fluids, assess the capacity and recoverable reserves of oil and gas reservoirs, optimize drilling and production decisions, and reduce exploration and development risks by combining data from multiple logging tools. In practical applications, factors such as logging tool characteristics, data quality, calibration, and alignment must be considered to ensure the accuracy and reliability of fusion analysis.



(**a**) Random forest for STUART wells.

(**b**) Transformer–LSTM for STUART wells.

(**c**) Random forest for CRAWFORD wells.

(**d**) Transformer–LSTM for CRAWFORD wells.

Figure 12. Lithology classification predictions of neighboring unlabeled blind wells using the T-LS and RF models.

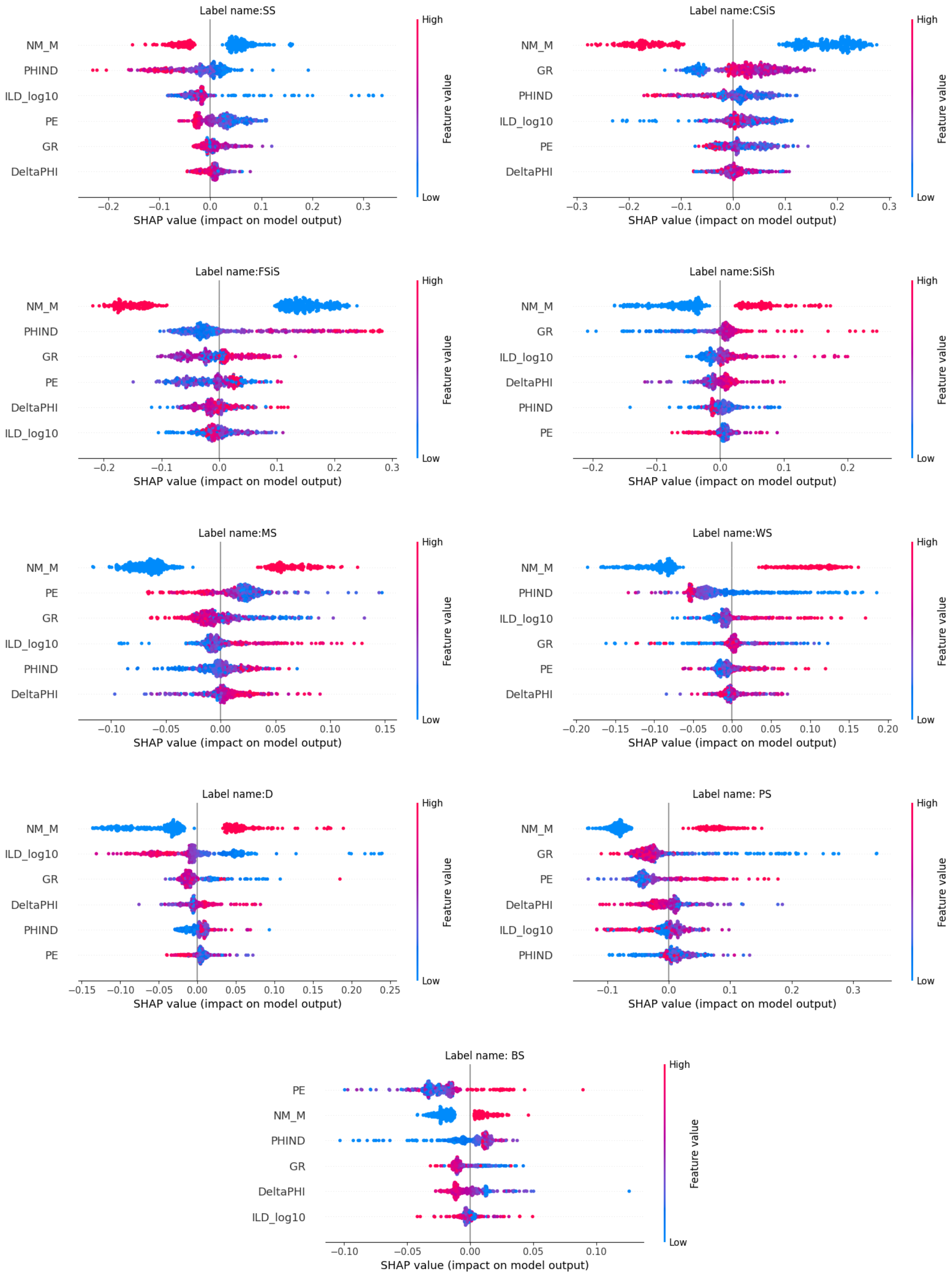


Figure 13. Shapley analysis scatter plot for the identification of nine lithology types via the T-LS model.

4. Conclusions

The prediction of lithology in oil and gas exploration target areas holds paramount importance for the development of oil and gas resources. Addressing the inherent subjectivity and limitations of traditional logging methods, we proposed a deep learning-based approach to realize lithology prediction in blind wells. Raw logging data often present practical challenges such as outliers influenced by noise, variations in raw data from different logging instruments, uneven distribution of logging data samples, and partially missing data. Initially, the Savitzky–Golay filtering method was employed to mitigate outliers in logging curves. A comparative analysis with the commonly used SVRs and RFR methods revealed the superior evaluation performance of the ResTCN in terms of MAE and R^2, demonstrating its efficacy in completing missing data. Eight commonly used models for lithology classification prediction, including LR, KNN, DT, and RF, were selected for comparative experiments. Comprehensive evaluation metrics such as the precision, recall, and $F1$-score were employed to assess the model performance. Among the nine models evaluated, T-LS and RF exhibited superior evaluation results. Notably, in the prediction task encompassing nine rock samples, T-LS outperformed RF for seven samples, and achieved slightly lower ratings for two samples. The overall rating of the T-LS model reached 0.88, surpassing the RF model, which reached 0.74.

The comparative experimental results underscore the superior classification prediction of the proposed model compared to traditional methods. The lithology of adjacent unlabeled blind wells was predicted using the T-LS and RF models, revealing superior performance of the new model in terms of its generalization ability, thereby partially mitigating the challenges of inaccurate prediction and weak generalization abilities encountered by traditional machine learning and deep learning models in lithology recognition. Our future work will focus on improving the T-LS model by incorporating additional features and refining its architecture, so as to further enhance the lithology prediction accuracy. Additionally, exploring the integration of advanced data augmentation techniques and domain-specific knowledge could enhance the model's robustness and generalization ability across diverse geological formations.

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

Funding: This study was partially supported by the Qinglan Project of Jiangsu Universities, the Talent Development Project of Taizhou University (No. TZXY2018QDJJ006), and the Young Science and Technology Talent Support Project of Taizhou.

Data Availability Statement: All data are contained within the article.

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

Abbreviations

The following abbreviations are used in this manuscript:

T-LS	Transformer–Long Short-Term Memory
LSTM	Long Short-Term Memory
ResTCN	Residual Temporal Convolutional Networks
KNN	K-Nearest Neighbor
LR	Logistic Regression;
CNN	Convolutional Neural Network
SVR	Support Vector Regression
MAE	Mean Absolute Error
R^2	Coefficient of Determination
TP	True Positive

FN	False Negative
FP	False Positive
TN	True Negative
LR	Logistic Regression
DT	Decision Tree
RBF	Random Forest Regression
RF	Random Forest
GB	Gradient Boosting
LSVM	Linear SVM
MSVM	Multiclass Support Vector Machine
BNB	BernoulliNB

References

1. Yao, G.; Wu, X.; Sun, Z.; Yu, C.; Ge, Y.; Yang, X. Status and prospects of exploration and exploitation key technologies of the deep petroleum resources in onshore Chinan. *J. Nat. Gas Geosci.* **2018**, *31*, 125–135.
2. Guo, Q.; Ren, H.; Yu, J.; Wang, J.; Liu, J.; Chen, N. A method of predicting oil and gas resource spatial distribution based on Bayesian network and its application. *J. Pet. Sci. Eng.* **2022**, *208*, 109267. [CrossRef]
3. Qian, K.R.; He, Z.L.; Liu, X.W.; Chen, Y.Q. Intelligent prediction and integral analysis of shale oil and gas sweet spots. *Pet. Sci.* **2018**, *15*, 744–755. [CrossRef]
4. Mishra, A.; Sharma, A.; Patidar, A.K. Evaluation and development of a predictive model for geophysical well log data analysis and reservoir characterization: Machine learning applications to lithology prediction. *Nat. Resour. Res.* **2022**, *31*, 3195–3222. [CrossRef]
5. Logging, C. Reservoir characteristics of oil sands and logging evaluation methods: A case study from Ganchaigou area, Qaidam Basin. *Lithol. Reserv.* **2015**, *27*, 119–124.
6. Min, X.; Pengbo, Q.; Fengwei, Z. Research and application of logging lithology identification for igneous reservoirs based on deep learning. *J. Appl. Geophys.* **2020**, *173*, 103929. [CrossRef]
7. Saporetti, C.M.; da Fonseca, L.G. A lithology identification approach based on machine learning with evolutionary parameter tuning. *IEEE Geosci. Remote. Sens. Lett.* **2019**, *16*, 1819–1823. [CrossRef]
8. Park, S.Y.; Son, B.K.; Choi, J. Application of machine learning to quantification of mineral composition on gas hydrate-bearing sediments, Ulleung Basin, Korea. *J. Pet. Sci. Eng.* **2022**, *209*, 109840. [CrossRef]
9. Handhal, A.M.; Ettensohn, F.R. Spatial assessment of gross vertical reservoir heterogeneity using geostatistics and GIS-based machine-learning classifiers: A case study from the Zubair Formation, Rumaila oil field, southern Iraq. *J. Pet. Sci. Eng.* **2022**, *208*, 109482. [CrossRef]
10. Antariksa, G.; Muammar, R.; Lee, J. Performance evaluation of machine learning-based classification with rock-physics analysis of geological lithofacies in Tarakan Basin, Indonesia. *J. Pet. Sci. Eng.* **2022**, *208*, 109250. [CrossRef]
11. Mukhopadhyay, P. Advances in Well Logging Techniques for Shale Reservoirs Exploration. In *Unconventional Shale Gas Exploration and Exploitation: Current Trends in Shale Gas Exploitation*; Springer International Publishing: Cham, Switzerland, 2024; pp. 31–47.
12. Li, Z.; Deng, S.; Hong, Y.; Wei, Z.; Cai, L. A novel hybrid CNN–SVM method for lithology identification in shale reservoirs based on logging measurements. *J. Appl. Geophys.* **2024**, *223*, 105346. [CrossRef]
13. Lee, H.; Lee, H.P. Formation lithology predictions based on measurement while drilling (MWD) using gradient boosting algorithms. *Geoenergy Sci. Eng.* **2023**, *227*, 211917. [CrossRef]
14. Agrawal, R.; Malik, A.; Samuel, R. Real-time prediction of Litho-facies from drilling data using an Artificial Neural Network: A comparative field data study with optimizing algorithms. *J. Energy Resour. Technol.* **2022**, *144*, 043003. [CrossRef]
15. Singh, H.; Seol, Y.; Myshakin, E.M. Automated well-log processing and lithology classification by identifying optimal features through unsupervised and supervised machine-learning algorithms. *SPE J.* **2020**, *25*, 2778–2800. [CrossRef]
16. Joshi, D.; Patidar, A.K.; Mishra, A.; Mishra, A.; Agarwal, S.; Pandey, A. Prediction of sonic log and correlation of lithology by comparing geophysical well log data using machine learning principles. *GeoJournal* **2021**, *88*, 47–68. [CrossRef]
17. Wang, J.; Cao, J. A Lithology Identification Approach Using Well Logs Data and Convolutional Long Short-Term Memory networks. *IEEE Geosci. Remote Sens. Lett.* **2023**, *20*, 7506405. [CrossRef]
18. Song, C.; Lu, W.; Wang, Y. Reservoir prediction based on closed-loop CNN and virtual well-logging labels. *IEEE Trans. Geosci. Remote Sens.* **2022**, *60*, 5919912. [CrossRef]
19. Wu, L.; Dong, Z.; Li, W. Well-logging prediction based on hybrid neural network model. *IEEE Trans. Geosci. Remote Sens.* **2021**, *14*, 8583. [CrossRef]
20. Yang, W.; Xia, K.; Fan, S. Oil logging reservoir recognition based on TCN and SA-BiLSTM deep learning method. *Eng. Appl. Artif. Intell.* **2023**, *121*, 105950. [CrossRef]
21. Smith, R.; Bakulin, A.; Golikov, P. Predicting sonic and density logs from drilling parameters using temporal convolutional networks. *Lead. Edge* **2022**, *41*, 617–627. [CrossRef]
22. Vaswani, A.; Shazeer, N.; Parmar, N. Attention is all you need. *Adv. Neural Inf. Process. Syst.* **2017**, *30*.
23. Hochreiter, S.; Schmidhuber, J. Long short-term memory. *Neural Comput.* **1997**, *9*, 1735–1780. [CrossRef] [PubMed]

24. Jacinto, M.V.G.; Silva, M.A.; de Oliveira, L.H.L.; Medeiros, D.R.; de Medeiros, G.C.; Rodrigues, T.C.; de Almeida, R.V. Lithostratigraphy Modeling with Transformer-Based Deep Learning and Natural Language Processing Techniques. In Proceedings of the Abu Dhabi International Petroleum Exhibition and Conference, Abu Dhabi, United Arab Emirates, 2–5 October 2023; SPE, D031S110R003.
25. Viggen, E.M.; Merciu, I.A.; Løvstakken, L.; Måsøy, S.E. Automatic interpretation of cement evaluation logs from cased boreholes using supervised deep neural networks. *J. Pet. Sci. Eng.* **2020**, *195*, 3107539. [CrossRef]
26. Zhang, D.; Yuntian, C.; Jin, M. Synthetic Well Logs Generation via Recurrent Neural Networks. *Pet. Explor. Dev.* **2018**, *45*, 629–639. [CrossRef]
27. Wang, J.; Cao, J.; You, J. A method for well log data generation based on a spatio-temporal neural network. *J. Geophys. Eng.* **2021**, *18*, 700–711. [CrossRef]
28. Tewari, S.; Dwivedi, U.D. A comparative study of heterogeneous ensemble methods for the identification of geological lithofacies.Journal of Petroleum. *Explor. Prod. Technol.* **2020**, *10*, 1849–1868. [CrossRef]
29. Ismail, A.; Ewida, H.F. Identification of gas zones and chimneys using seismic attributes analysis at the Scarab field, offshore, Nile Delta, Egypt. *Pet. Res.* **2020**, *5*, 59–69. [CrossRef]
30. Liu, Y.; Dang, B.; Li, Y. Applications of savitzky-golay filter for seismic random noise reduction. *Acta Geophys.* **2016**, *64*, 101–124. [CrossRef]
31. Roy, I.G. An optimal Savitzky–Golay derivative filter with geophysical applications: An example of self-potential data. *Geophys. Prospect.* **2020**, *68*, 1041–1056. [CrossRef]
32. Sabah, M.; Talebkeikhah, M.; Wood, D.A. A machine learning approach to predict drilling rate using petrophysical and mud logging data. *Earth Sci. Inform.* **2019**, *12*, 319–339. [CrossRef]
33. Savitzky, A.; Golay, M.J.E. Smoothing and differentiation of data by simplified least squares procedures. *Anal. Chem.* **1964**, *36*, 1627–1639. [CrossRef]
34. Zhou, K.; Zhang, J.; Ren, Y. A gradient boosting decision tree algorithm combining synthetic minority oversampling technique for lithology identification. *Geophysics* **2020**, *85*, 147–158. [CrossRef]
35. Bacal, M.C.J.O.; Hwang, S.; Guevarra-Segura, I. Predictive lithologic mapping of South Korea from geochemical data using decision trees. *J. Geochem. Explor.* **2019**, *205*, 106326. [CrossRef]
36. Shier, D.E. Well log normalization: Methods and guidelines. *Petrophys.-SPWLA J. Form. Eval. Reserv. Descr.* **2004**, *45*, SPWLA-2004-v45n3a4.
37. Sun, J.; Li, Q.; Chen, M. Optimization of models for a rapid identification of lithology while drilling-A win-win strategy based on machine learning. *J. Pet. Sci. Eng.* **2019**, *176*, 321–341. [CrossRef]
38. Jiang, H.; Nachum, O. Identifying and correcting label bias in machine learning. In Proceedings of the International Conference on Artificial Intelligence and Statistics, Online, 26–28 August 2020; PLMR; pp. 702–712.
39. Elmgerbi, A.; Chuykov, E.; Thonhauser, G.; Nascimento, A. Machine learning techniques application for real-time drilling hydraulic optimization. In Proceedings of the International Petroleum Technology Conference, Dhahran, Saudi Arabia, 21–23 February 2022; IPTC; D011S018R002.

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.