Journal of Applied Geophysics 212 (2023) 104990

Contents lists available at ScienceDirect

Journal of Applied Geophysics

journal homepage: www.elsevier.com/locate/jappgeo

Towards automatic and accurate core-log processing

Artur Jordao [a,*], Joao Paulo da Ponte Souza [c], Michelle Chaves Kuroda [c], Marcelo Fagundes de Rezende [d], Helio Pedrini [a], Alexandre Campane Vidal [b]

[a] Institute of Computing, University of Campinas, Campinas, São Paulo, Brazil
[b] Institute of Geosciences, University of Campinas, Campinas, São Paulo, Brazil
[c] Centro de Estudos de Petroleo, University of Campinas, Campinas, São Paulo, Brazil
[d] Petróleo Brasileiro S.A. (Petrobras), Rio de Janeiro, Rio de Janeiro, Brazil

ARTICLE INFO

Keywords:
Core-log processing
Machine learning
Recurrence models
Transformer network

ABSTRACT

The analysis of rocks plays an important role in geological and petroleum-engineering problems. In these tasks, core (i.e., core-log) is a crucial element since it provides underlying information on the geophysical properties of the area. Thereby, works often leverage core data to assign the correct category from well-logs. Unfortunately, processing core is laborious and time-consuming; hence, its analysis takes long periods. For example, for a single well, a human could have to adjust thousands of core data. Besides, due to its nature, log analysis must handle noise and missing data. Given these issues, our goal is to propose an automatic (e.g., without any human intervention) and accurate core processing using the gamma-ray. To achieve this goal and additionally demonstrate the most promising pattern recognition strategies, we assess the effectiveness of several models to diagnostic the core-log. Such models include simple regressions, ensembles, gradient boosting, recurrence models and the relatively recent Transformer network. Our comprehensive evaluation is different from existing works on geoscience tasks, which evaluate a small number of models or variations of a single model. Particularly, we evaluate more than 200 unique models (i.e., models with different hyperparameters) and observe that deep learning techniques outperform other techniques by a large margin. Furthermore, we compare the compromise between predictive ability and computational cost of several models – a reliable information for real-time lithology, which is often overlooked by previous works. According to our results, we can effectively replace the manual (i.e., by a human expert – geologist) diagnostic of core by pattern recognition methods, mainly by the Transformer model, as it aligns the cores in accordance (i.e., in the same direction) with the geologist. More specifically, given the core and gamma-ray as input, the Transformer model outputs an adjusted core obtaining an R^2 of 93,63, which indicates a fine-grained adjustment. We empirically demonstrate that the success behind Transform is its effectiveness in expressing large sequences of core and well-log. On the other hand, other models such as RNN, LSTM and GRU meet collapse when processing large sequences of core-log. We further confirm that Transformers are among the top-performance models and often surpass recurrence models on several geoscience applications such as lithology and log-shape classification, and prediction of oil production. Regarding the first task, the models successfully classify different facies categories such as coarse sandstone, medium sandstone, fine sandstone, siltstone, dolomite, limestone and mudstone. In these tasks, Transformer outperforms the widely-employed LSTM by up to 14 percentage points. Our empirical observations encourage the exploration of multiple models and suggest Transformers as strong baselines for future research on geoscience tasks. In this direction, we release all data and trained models used throughout the work. To the best of our knowledge, we are the first study exploring Transformer models on geoscience applications.

1. Introduction

At the heart of modern society lies petroleum engineering and geoscience research, as we need to constantly discover and (green) explore energy and raw materials. Efforts towards improving these tasks become, therefore, fundamental requirements to modern society. In this direction, well-log processing and analysis play an important role, for example, to classify lithological boundaries in oil and gas reservoirs

* Corresponding author.
E-mail address: arturjordao@dcc.ufmg.br (A. Jordao).

https://doi.org/10.1016/j.jappgeo.2023.104990
Received 9 May 2022; Received in revised form 17 January 2023; Accepted 28 February 2023
Available online 4 March 2023

(Zhou et al., 2020; Karimi et al., 2021). Traditionally, a human expert – geologist – performs these tasks by carefully analyzing well-logs data composing the raw properties of the rocks. Such a process is prohibitively laborious and time-consuming and, to make things worse, log analysis must often handle noise and missing data (Wang and Chen, 2019; Datskiv et al., 2020; Singh et al., 2020). Importantly, the interpretation of logs can vary between experts; thus becoming an ambiguous process (Dubois et al., 2007; Song et al., 2020). Therefore, our goal is automatically and accurately adjust the core without any human intervention. For this purpose, we employ several pattern recognition methods including simple regressions, ensembles, gradient boosting, recurrence models and the relatively recent Transformer network. These methods take as input the core and the gamma-ray and output the adjusted core.

Previous efforts on automating well-log analysis have confirmed that pattern recognition methods can successfully replace or complement the work of a geologist. These works have obtained positive results in classifying lithological boundaries, identifying log shapes and predicting oil production. For example, Zhou et al. (2020) employed classifiers such as Support Vector Machine (SVM), random forest and gradient boosting to classify lithofacies. These techniques either use raw sensory data or design handcrafted features to extract patterns from well-logs (Karimi et al., 2021).

On the other end of the spectrum, studies have shown impressive results employing deep learning techniques (Lin et al., 2021; Santos et al., 2022). In this line of research, most works predominantly employ the Long Short-Term Memory (LSTM) model, as its positive results in natural language processing (NLP) and structured data consistently translate to well-log settings (Sagheer and Kotb, 2019; Lin et al., 2021; Santos et al., 2022). It is worth mentioning that in other cognitive tasks, researchers have replaced LSTM models by Transformers since the latter exhibit a better capacity to represent large sequences of data, even without any recurrence mechanism (Vaswani et al., 2017; Devlin et al., 2019; Rae et al., 2020). Surprisingly, in image classification tasks where convolutional networks are a paradigm of choice, Transformers are also on par with the state of the art (Carion et al., 2020; Dosovitskiy et al., 2021; Liu et al., 2021b). The role of the sequence size is the amount of information about the data that the models can represent. In the context of well-log processing and analysis, large sequence sizes enable the recurrent models to include more neighboring readings of the sensors, leading to better results (Wang and Chen, 2019; Song et al., 2020; Lin et al., 2021; Santos et al., 2022).

Regardless of the technique, studies on automatic well-log processing have confirmed the following aspects. First, gamma-ray benefits the predictive ability of the pattern recognition methods (Datskiv et al., 2020). Second, existing studies on deep learning often overlook evaluating LSTMs with models that share similar properties (Wang and Chen, 2019; Santos et al., 2022; Lin et al., 2021). Therefore, it is unclear if other related models (specifically, the recurrence ones) would achieve competitive results. In other words, there is weak evidence that LSTM is the more suitable model for processing well-log data. Third, existing works overlook the trade-off between predictive ability and computational cost of the models. In particular, based on our literature review, only the work by Sun et al. (2019) reports issues involving computational cost. The rationale behind discussing this quantitative metric is to demonstrate the suitableness of the models for real-time and resource-constraints applications as well as the computational budget for working on a specific task. Take the logging while drilling (LWD) equipment as an example: a classification model could perform real-time lithology as long as its computational cost satisfies the computational budget of the equipment. Additionally, since the equipment's logging speed is limited to its slowest device (service) (Serra, 1986), being aware of the computational cost of the model plays a role in real-time applications. At last but not least, many advances in automatic lithology employ private data. Hence, data are restricted[1] (not publicly available) and scarce, hindering reproducibility. To be more precise, only 33% of the papers we review provide the source data and/or the code. Among the papers satisfying such issues, none of them release the core from the well-log.

In this work, we investigate the performance of several pattern recognition models through the lens of core-log adjustment. Such a task consists of adjusting the core based on its relationship with gamma-ray, as illustrated in Fig. 1. We observe that while many studies explore gamma-ray and other well-log data (Imamverdiyev and Sukhostat, 2019; Song et al., 2020; Lin et al., 2021), the core from these logs often receives no attention. It turns out that the analysis involving core is costly and time-consuming; hence, its analysis can take long periods (Serra, 1986). On the other hand, it plays a fundamental role in precise lithology comprehension (Sun et al., 2019; Song et al., 2020; Asante-Okyere et al., 2020). For example, Song et al. (2020) assigned the correct shape (i.e., the label – ground-truth) from well-log based on a consensus of experts, which in turn define the shape by analyzing the morphological features from cores. Similarly, Imamverdiyev and Sukhostat (2019), and Sun et al. (2019) defined the true lithology category through cores. Overall, despite its drawbacks and limitations, many geoscience tasks depend on core-log integration.

According to the above observations, the core is indeed a key pillar of several applications, mainly during the labeling process of the data. Unfortunately, due to its nature, the core-log analysis must often handle noise and missing data, hindering interpretation and leading to ambiguous diagnoses of experts. Thus, core requires manual correction, which is laborious and takes long periods. To mitigate this problem, we conduct a comprehensive evaluation of models capable of adjusting the core and compare their effectiveness with the diagnostic by an expert geologist (i.e., the manual correction). The reason for such in-depth assessment is that there is weak evidence of the top-performance models, as previous works consider single models or perform comparisons with simple baselines. For example, Lin et al. (2021) assessed lithology classification by taking into account a single LSTM, for which the authors vary only the optimizer. Closely related, Santos et al. (2022) evaluated an LSTM architecture with standard pattern recognition methods; however, the authors overlook other recurrence models. Such observations also appear in works focusing on standard machine learning models since they avoid comparison with deep learning techniques. Our study bridge this gap since we perform a comprehensive evaluation of models often used by prior works ranging from standard models to modern deep learning techniques such as Transformers (Vaswani et al., 2017). Regarding the Transformer, studies on NLP have suggested its potential in replacing recurrence models (Vaswani et al., 2017; Devlin et al., 2019; Rae et al., 2020). Transformers have also demonstrated positive results in the computer vision community, obtaining competitive results to convolutional networks (Carion et al., 2020; Dosovitskiy et al., 2021; Liu et al., 2021b). Therefore, a natural question is whether the results from Transformers translate into geoscience applications. Not surprisingly, we find the answer is positive.

Overall, our workflow operates as follows. Given a well, we present its core and the gamma-ray to a regression model, which, in turn, outputs an adjusted core. We perform this process for each well composing our database. This pipeline (summarized in Fig. 1) works efficiently, which indicates its advantage and feasibility in real-time scenarios. Since our study considers various pattern recognition methods (including the computationally extensive recurrence networks), we employ the minimum information (gamma-ray) possible to adjust the core. According to our experiments, we can replace the manual adjustment (i.e., the one made by a geologist) of core-gamma by pattern

[1] To avoid any conflict of interest, we do not mention the papers with restricted code/data.

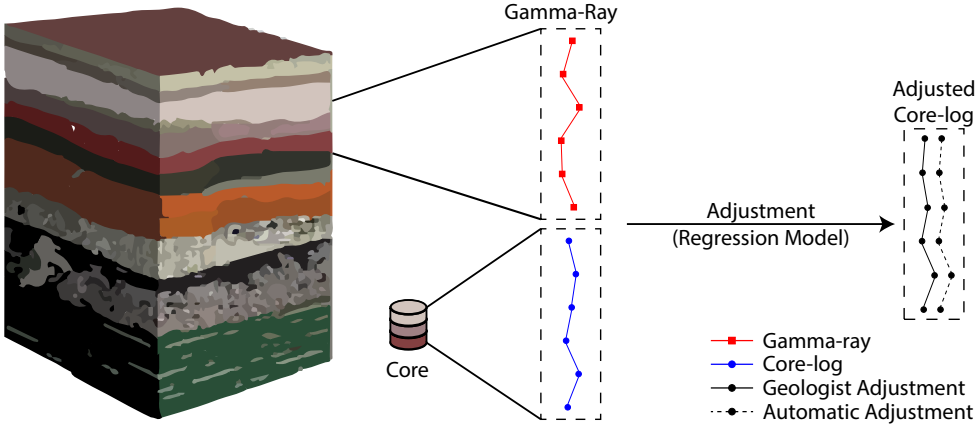

Fig. 1. Overall pipeline of the core-adjustment problem. Given a core data (blue curve) from a well-log (red curve), we apply a regression model to adjust the core (dashed-black curve). To assess the effectiveness of the adjustment done by a regression model, we compare it with the diagnostic provided by a human expert (geologist – solid-black curve). Our results demonstrate that the laborious and time-consuming core processing can be successfully replaced by pattern recognition methods, meaning an automatic adjustment without any human intervention. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

recognition techniques, which means a diagnosis without any human intervention – an important step towards automatic well-log analysis. To the best of our knowledge, this is the first study to explore Transformer models on geoscience applications. We believe that much more progress could be made by considering Transformers as standard baselines as well as carefully analyzing multiple models. We guess our findings would help future research to push the state of the art in petroleum engineering and geoscience tasks.

Contributions. Among our contributions, we highlight the following five folds:

- In the context of core-log adjustment, we show that deep learning models dominate several standard techniques such as gradient boosting and random forest, which are strong baselines in pattern recognition problems. We confirm this by a comprehensive evaluation of high-capacity models such as simple Recurrence Neural Network (RNN), Gated Recurrent Units (GRU), LSTM and Transformers.
- Regarding the Transformer model, we show that its effectiveness in expressing data with structural dependency translates into state-of-the-art performance in well-log data. We confirm that Transformers exhibit a higher capacity in representing long sequences of core data; thus achieving superior predictive ability than other recurrence models. From this perspective, we show that Transformers are less sensitive to sequence size change than recurrence models, meaning we can avoid careful tuning to guarantee competitive performance. We believe these findings encourage future efforts on Transformers.
- We compare the trade-off between predictive ability and computational cost of several deep learning models often employed in geoscience tasks. We observe that a common dilemma between these two metrics there also exists in geoscience tasks – high-capacity models incur a higher computational cost.
- While our main focus is on adjusting the core-log, we also demonstrate the performance of Transformers on several geoscience applications such as lithogology/facies classification, log-shape identification and prediction of the oil production. On these tasks, Transformers are always among the top-performance models. Thereby, Transformers emerge as potential baselines for future research. Surprisingly, we find that a simple RNN, often, surpasses LSTM. Such a contribution is not to discourage research on LSTM but, on the contrary, to offer motivation for more research on other models.
- In order to promote reproducibility, we release all data and trained models used in this work. The source code is avaiable at: https://github.com/arturjordao/TowardsAutomaticAccurateCore-logProcessing

2. Related Work

Due to the essence of the well-log data, past and ongoing research employ well-established elements of time-series analysis.

It is well known that different logs capture distinct characteristics from rocks. Despite this, previous studies confirmed that well-logs exhibit correlation with each other (Asante-Okyere et al., 2020; Karimi et al., 2021). In the task of water saturation prediction, Asante-Okyere et al. (2020) demonstrated that gamma-ray, sonic travel time, spontaneous potential, resistivity, and neutron porosity exhibit high correlation. Most importantly, the authors noted that this correlation negatively impacts the predictive ability of a model. To decorrelate the data, the authors proposed to employ the Principal Components Analysis (PCA) technique, which projects the original data onto subspaces (i.e., components) of maximum variance. Closely related to Asante-Okyere et al. (2020), Karimi et al. (2021) also employed PCA as preprocessing step, but in the context of identifying lithological boundaries. Both works demonstrated promising results by projecting (i.e., project onto PCA space) the data before classifying them. An underlying property of PCA is its unsupervised essence – it is unaware of the data category (label). In the supervised setting, Zhou et al. (2020) confirmed that some classifiers suffer from the imbalance distribution of types of rocks. For example, on the dataset by Sun et al. (2019), limestone and mudstone compose 3.41% and 27.74% of the data, respectively. To mitigate this issue, Zhou et al. (2020) proposed to use a synthetic sampling strategy for balancing the minority classes before the training phase. As evidenced by Zhou et al. (2020), the sampling strategy rarely affects the result of ensemble-like classifiers such as Adaboost due to their effectiveness in imbalanced data regimes. Besides imbalanced problems, some classifiers fail in modeling data when the data is non-linearly separable. To mitigate this issue, Al-Mudhafar (2017a) used the kernel version of SVM and obtained an accurate lithology prediction.

The works by Al-Mudhafar (2020); Abbas and Al-Mudhafar (2021); Al-Mudhafar et al. (2022); Al-Mudhafar and Wood (2022) confirmed the positive results of boosting and ensemble-based strategies in various applications of petroleum engineering and geoscience. In addition to this family of models, single models have also achieved outstanding results (Tang and White, 2008; Al-Mudhafar, 2017b; Ameur-Zaimeche et al., 2020). For example, Ameur-Zaimeche et al. (2020) combined a three-layer multilayer feedforward network (i.e., non-deep – shallow – neural networks) with clustering analysis to predict non-cored lithofacies in fluviatile reservoirs. Al-Mudhafar (2017b) employed Kernel Discriminant Analysis and generalized boosting models to predict lithofacies (discrete) distribution from data sequences, where the former can include missing intervals.

While the works above pay attention to standard machine learning models, another line of research devotes efforts to deep learning techniques. Studies on the latter category predominantly consider LSTMs

(Wang and Chen, 2019; Song et al., 2020; Lin et al., 2021; Santos et al., 2022). On the one hand, these works assess the effectiveness of LSTM on various (classification) geoscience tasks using different well-log data. For example, Lin et al. (2021) and Song et al. (2020) confirmed the success of LSTM to classify lithology and identify log shapes, respectively. On the other hand, by comparing the LSTM model between the works, we observe that it exhibits small changes, mainly on the number of neurons and the optimizer.

Overall, deep learning techniques have achieved unprecedented results and often outperform standard machine learning techniques in many applications. Despite this, Santos et al. (2022) observed that when classifying fine-sand facies Extreme Gradient Boosting surpasses LSTM. Such evidence reinforces the idea of assessing a vast range of classifiers instead of a single family of models (e.g., recurrence models). Differently from Santos et al. (2022), we observe that even with competitive outcomes, Extreme Gradient Boosting and other ensemble-like models underperform deep learning methods in the task of core adjustment.

The aforementioned works fall into (geoscience) classification tasks, but the success of LSTM models has also been confirmed into prediction tasks. Taking the work by Wang and Chen (2019) as an example, the authors successfully employed LSTM to predict pressure data from wells. Moreover, the authors showed that LSTM outperforms models that require prior knowledge of analytical models, thus reinforcing the potential of deep learning as an alternative to traditional techniques employed on geoscience and petroleum engineering applications. Sagheer and Kotb (2019) used LSTM to forecast the oil production based on past observations. Surprisingly, Sagheer and Kotb (2019) confirmed that stacking multiple layers lead to better results on oil prediction. Since previous efforts on related applications consider small (i.e., shallow) deep learning models (often a single-layer LSTM), the evidence by Sagheer and Kotb (2019) poses a question of whether deeper (more layers) models could achieve better results. Unfortunately, the answer to this question involves training and testing many possible architectures. Rather than exploring all possible architectures in an architecture-design space, Sagheer and Kotb (2019) applied a genetic algorithm to create high-performance LSTMs.

It is worth mentioning that in other pattern recognition tasks, most frequently image classification, many efforts are devoted to architecture design since it plays an important role in the predictive ability of deep networks (Tan and Le, 2019; Dong and Yang, 2020). Closely related to our work, Sagheer and Kotb (2019) also evaluated the predictive performance of other recurrence networks, but the authors overlook standard machine learning models, which are strong baselines. Besides, our work is the first to explore Transformer models.

Altogether, existing works confirm the effectiveness of LSTM not only on classification-based geoscience tasks but also on prediction. The reason for selecting such a model is due to its success in time-series and NLP (Zeyer et al., 2017). After the work by Vaswani et al. (2017), studies put into question the dominance of recurrence-like models in expressing long-range dependence from data. Vaswani et al. (2017) proposed a deep learning architecture based on the self-attention mechanism: the Transformer model. The authors confirmed that Transformers express long sequences of data better than recurrence models. Moreover, Vaswani et al. (2017) demonstrated that Transformers are also more efficient in terms of computational cost on the constraints that the sequence length is smaller than the dimensionality of the self-attention projections. Since this seminal work, Transformers-like models have pushed the state-of-the-art in different tasks (Rae et al., 2020; Carion et al., 2020; Kreuzer et al., 2021; Xie et al., 2021). Surprisingly, these models perform on par or better than convolutional networks (Dosovitskiy et al., 2021; Liu et al., 2021b), which is the central paradigm in deep learning focused on computer vision tasks.

3. Methodology

3.1. Problem definition

In this section, we introduce the properties of our data and its constraints. Then, we define the problem of adjusting the core from gamma-ray logs.

The core-log data composing our dataset consists of continuous curves, meaning that the models only receive (and predict) one sequence from each well. On these data, we define some constraints for adjusting the core as automatically as possible. According to the nature of the core-log, a time series data with depth and gamma-ray information, we set its depth information as the starting point to select the region of the gamma-ray to be analyzed. We observe that the precision of the core-gamma depth is enough to conduct efficient shifts by the models.

Given the aforementioned issues, define $C \in \mathbb{R}^{1\times n}$ the points (i.e., the API readings) of a core from a well-log (gamma-ray). We use the terms core and core-log interchangeably. Define $G \in \mathbb{R}^{1\times n}$ the points of gamma-ray data. Let C_i and G_i be a single sample of C and G, respectively, at the depth i.

Given an arbitrary well, our problem is to adjust all the values of core $(C_1, C_2, \ldots, C_n)$ based on the gamma-ray. Intuitively, we can achieve this by predicting an adjusted core value through a regression model. Built upon this idea, consider a regression model $\mathscr{F}$ that takes as input a set of points and outputs an adjusted core-log value. In our modeling, $\mathscr{F}$ receives a set of points of core and gamma-ray and predicts the adjustment to a core value, at a single depth, $C_i^{\mathscr{F}}$. Formally, $\mathscr{F}(x) = C_i^{\mathscr{F}}$, in which $x = \{(C_i, G_i), (C_{i+1}, G_{i+1}), \ldots (C_{i+w}, G_{i+w})\}$. In this formulation, w is a parameter and indicates a sequence of data (pairs of core and gamma-ray). For example, $w = 3$ means that the models would receive three consecutive pairs of core and gamma-ray to adjust (i.e., predict) a single core point.

The regression model $\mathscr{F}$ can be a standard machine learning model (e.g., linear regression or random forest) or a deep learning model (e.g., recurrence- or Transformer-like networks). After applying $\mathscr{F}$ over all core points, we generate an adjusted core data $C^{\mathscr{F}}$ $(C_i^{\mathscr{F}}, \ldots, C_n^{\mathscr{F}})$.

In order to verify the effectiveness of $\mathscr{F}$ in adjusting the core, we compare the predicted values with a ground-truth (C^{geo}), which corresponds to the core adjusted (point-by-point) by an expert geologist. For this purpose, let $\mathscr{M}(.,.)$ be a performance metric that takes as input the ground-truth and the adjusted core by $\mathscr{F}$, and outputs a single value indicating how fine is the adjustment of $C^{\mathscr{F}}$ compared to C^{geo}. The better the value of $\mathscr{M}(C^{geo}, C^{\mathscr{F}})$, the more confident the model $\mathscr{F}$ is for the task of adjustment.

Following the aforementioned description, the depths of C, C^{geo} and $C^{\mathscr{F}}$ remain unchanged during all adjustment process, which means that $|C| = |C^{geo}| = |C^{\mathscr{F}}| = n$. We illustrate the overall pipeline of core adjustment in Figs. 1 and 2.

3.2. Transformer network

Given our problem definition, in this section, we describe the details behind the Transformer model.

The self-attention mechanism lies at the heart of Transform and works as follows. Let $Q \in \mathbb{R}^{d_k\times 1}$, $K \in \mathbb{R}^{d_k\times 1}$ and $V \in \mathbb{R}^{d_k\times 1}$ be a set of queries, keys and values. An attention operation consists of combining these matrices as

$$Attention(Q, K, V) = softmax\left(\frac{QK^T}{\sqrt{d_k}}\right)V, \tag{1}$$

where *softmax* (.) is a well-known activation function in deep learning defined in terms of

$$softmax(o) = \frac{e^{o_i}}{\sum_{i=1}^{|o|} e_i^o}, \tag{2}$$

with o representing the output of a given layer (i.e., the output of Eq.

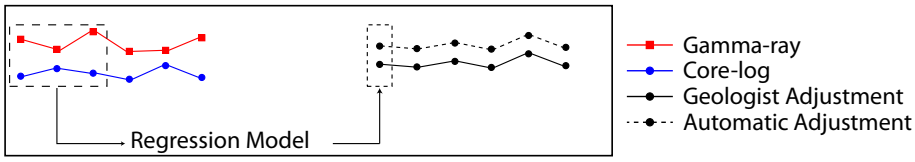

Fig. 2. The overall process to adjust a core curve. In this example, the model receives three consecutive pairs (sliding window) of core and gamma-ray to adjust (i.e., predict) a single core point. The sliding window slices over the curves (leftmost) to adjust all core data (rightmost). Observe that the model provides a curve with the same number of points as the ground-truth (geologist adjustment); therefore, the evaluation consists of a pair-wise (point-by-point) comparison.

(1)).

In the Transformer model, Q, K and V take as input the same input x (defined in Section 3.1); thus, such an operation receives the name of self-attention – the attention mechanism pays attention to x itself.

According to Vaswani et al. (2017), the model benefits from applying the self-attention mechanism multiple times for the same input such that the model learns different representations. For this purpose, the Transformer model employs multi-head attention by extending Eq. (1) as

$$Multihead(Q, K, V) = Concat(head_1, head_2, \ldots head_h)W^O, \quad (3)$$

where $head_i = Attention(QW_i^Q, KW_i^K, VW_i^V)$. Differently from Eq. (1), on the multi-head formulation, the attention mechanism first projects Q, K and V using learnable matrices (W_i^Q, W_i^K, W_i^V) randomly initialized. In Eq. 3, *Concat* indicates a simple concatenation operation. Importantly, after concatenating the output of each head, the model projects the output onto a matrix W^0.

Fig. 3 illustrates the overall architecture of a Transformer model. Note that the architecture stacks multiple layers of multi-head attention. In this structural setting, the output of a multi-head attention layer is the input to the next layer.

3.3. Evaluation Protocol

To assess the effectiveness of the models, we employ the holdout evaluation protocol (i.e., two-fold cross-validation). Such a protocol is a standard evaluation protocol employed by applications of machine learning in geoscience tasks (Zhang et al., 2017; Wang and Chen, 2019; Santos et al., 2022; Lin et al., 2021). The idea behind the holdout protocol is to split the data into two disjoint sets: training and testing sets. Fig. 4 illustrates the overall idea of the holdout protocol.

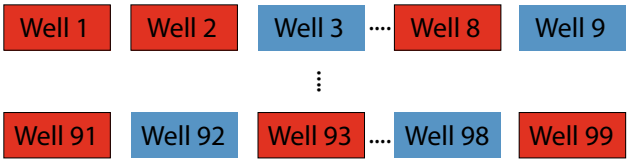

Fig. 4. Holdout evaluation protocol. Given a set of wells, the protocol splits them into training (red) and testing (blue) sets. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

In the holdout protocol, a common practice is to employ the 70–30 rule (Imamverdiyev and Sukhostat, 2019; Sagheer and Kotb, 2019; Wang and Chen, 2019), where 70% of data compose the training set and the remaining data (30%) constitute the testing set. We employ these values throughout the experiments.

4. Experiments and results

Dataset and Effectiveness Metrics. Our dataset consists of 99 wells, composed of three different facies (i.e., sandstone, sand–mudstone, and mudstone). The gamma-ray is sampled to each 0.05 m (i.e., the sampling rate is 0.05 m). Fig. 8 illustrates the data of an arbitrary well.

Throughout the experiments, we measure the effectiveness of the models (it says $\mathcal{M}(.,.)$) using the Root Mean Square Error (RMSE – leftmost in Eq. (4)) and Coefficient of Determination (R^2 – rightmost in Eq. (4)) metrics.

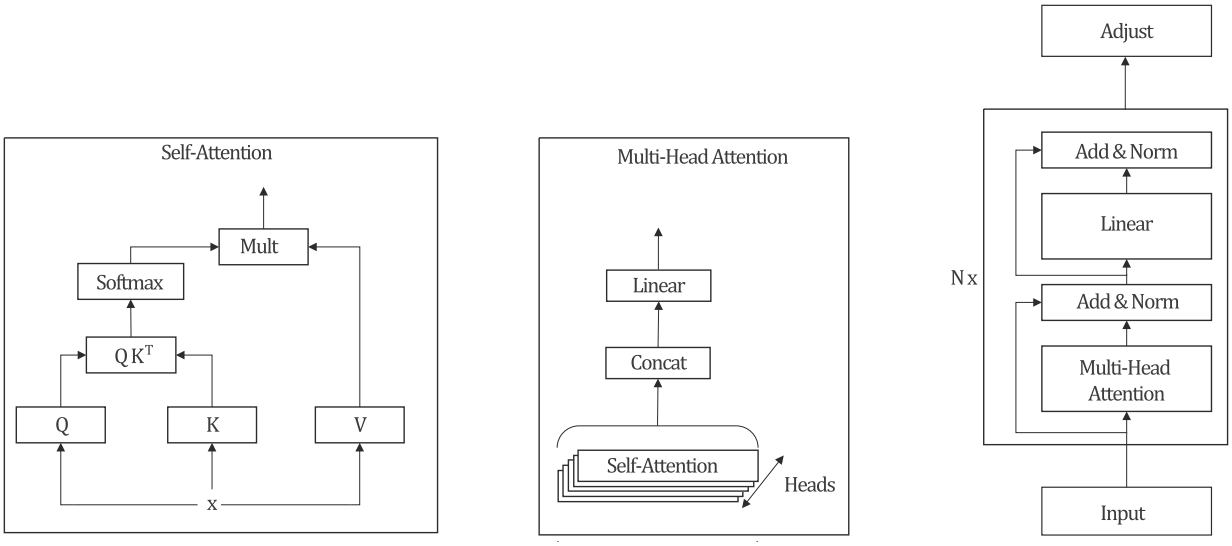

Fig. 3. The overall Transformer architecture. Left: Self-attention mechanisms. Middle: Multi-head Attention. Right: The final model architecture is obtained by stacking multiple layers of multi-head attention and other well-known layers. The layers *Add & Norm* indicate, respectively, a point-wise addition operation and batch normalization (Ioffe and Szegedy, 2015; Santurkar et al., 2018), whereas the linear layer means a linear projection onto a learnable matrix.

$$\text{RMSE} = \sqrt{\frac{1}{n}\sum_{i=1}^{N}\left(C_i^{Geo} - C_i^{\mathscr{F}}\right)^2}, \quad \text{R}^2 = 1 - \frac{\sum_{i=1}^{n}\left(C_i^{Geo} - C_i^{\mathscr{F}}\right)^2}{\sum_{i=1}^{n}\left(C_i^{Geo} - \sigma^2\right)^2}. \tag{4}$$

In Eq. (4), the symbol σ^2 stands for the variance of data. Finally, since each well will have an RMSE and R2, we report the average performance considering all wells.

Experimental Setup. To select the best hyperparameters for each model, we employ a grid-search scheme and evaluate the performance of each hyperparameter combination on a validation set, which consists of 10% of the training set. For fair comparison, we ensure that all models see the same training, validation and test samples. Regarding the deep learning models, we train the models for 500 epochs using the Adaptive Moment Estimation (Adam) optimizer. While there exist other potential optimizers such as Stochastic Gradient Descent (SGD), previous works have argued that Adam often provides competitive performance (Schneider et al., 2019; Schmidt et al., 2021). Particularly, Zeyer et al. (2017) suggested that recurrence models prefer Adam optimizer while Liu et al. (2020) and Davis et al. (2021) observed that Transformers are sensitive to SGD. On lithology identification, Lin et al. (2021) experimentally confirmed that Adam achieves better performance than other optimizers. We leverage these remarks and set the optimizer as Adam, which enables us to explore other hyperparameters more efficiently.

Comparison among Competing Models. In this experiment, we compare the performance of several models in the task of core-log adjustment. Table 1 summarizes the results. According to this table, Least Squares and Bayesian Ridge obtain similar performance. Interestingly, ensemble-based techniques such as Random Forest, Extreme Gradient Boosting and Light Gradient Boosting attain better performance than single models. Indeed, we observe that using a large number of estimators to compose the ensembles benefits their predictive capacity. It is worth mentioning that these findings are not unexpected, as previous works have been demonstrated remarkable results in ensembles not only in terms of predictive ability but also in the reliability of machine learning (i.e., robustness to adversarial samples and model uncertainty) (Lee and Chung, 2020; Wen et al., 2021; Zaidi et al., 2021). We believe that these results might encourage future research to investigate other ensemble strategies and integrate existing models into such strategies.

Despite the promising results, ensemble-based models underperform deep learning techniques. From Table 1, it is possible to observe that, in terms of RMSE, even the simple RNN outperforms the best ensemble model (Extreme Gradient Boosting) by an absolute difference of 0.51. This difference increases when we take into account the GRU and LSTM models. More precisely, by comparing the best recurrence model with Extreme Gradient Boosting, the absolute difference in RMSE is 0.59. The same trend holds on the R^2 metric. Such empirical evidence demonstrates that deep learning techniques are better candidates for an automatic and accurate core adjustment. In this direction of research, as we discussed before, most works on geoscience applications predominantly employ LSTM. Unfortunately, there is no drastic difference between LSTM and other recurrence models, as shown in Table 1. Thereby, a natural question that arises is whether LSTMs are, in fact, the best models for geoscience and petroleum engineering tasks. Taking GRU as an example, the different in RMSE and R^2 is only 0.02 and 0.0004, respectively. It is worth pointing out that LSTM and GRU share similar aspects in their architecture – GRU simplifies the LSTM cell –, and we believe that this simple modification slightly impacts the core adjustment. We shall see later that, on other geoscience tasks, the performance of LSTM compared with other recurrence models is even small and sometimes a simple RNN attains superior performance. Hence, a reasonable answer to the above question is that other recurrence models are as competitive as LSTMs.

Table 1
Comparison among competing models. We divide the models into two categories: standard machine learing models, which include ensemble-based strategies (Random Forest, Extreme Gradient Boosting and Light Gradient Boosting), and deep learning models. The arrows indicate which direction is better for each qualitative metric (RMSE and R^2).

Model	RMSE (↓)	R^2 (↑)
Support Vector Regression	6.01	0.8988
Least Squares	5.27	0.8987
Bayesian Ridge	5.27	0.8988
Light Gradient Boosting	5.12	0.9102
Random Forest	5.07	0.9097
Extreme Gradient Boosting	5.04	0.9092
RNN	4.53	0.9134
GRU	4.47	0.9160
LSTM	4.45	0.9164
Transformer	3.74	0.9363

Finally, the Transformer model surpasses all models we evaluate by a large margin. Specifically, compared to the best ensemble model, Transformer achieves an absolute difference of 1.29 and 0.0271 in RMSE and R^2, respectively. In terms of RMSE, Transformer outperforms RNN, GRU and LSTM by a difference of 0.79, 0.73 and 0.71, respectively. Note that this difference is higher than the one when we compare recurrence models with ensemble models. Most importantly, Transformer was the unique model capable of achieving an RMSE below 4 and an R^2 higher than 0.93. Regarding the latter metric, none of the models (including recurrence models) achieved an R^2 higher than 0.92. One particular aspect behind the results of Transformer is that it can express long sequences of data, which, in turn, translates directly into a finer core adjustment (we elaborate on this fact in the next experiment – see Fig. 5). The recurrence models, on the other hand, demonstrate sensitivity to process sequences longer than 2.

Although the discussion in this experiment considered only the core-adjustment problem, we confirm that many of the observations above hold to geoscience-related tasks. For example, we observe that the Transformers perform comparably (or best) to well-employed deep learning techniques in other geoscience tasks. Surprisingly, RNNs also demonstrate a strong baseline, achieving competitive results. Prior works, however, predominantly evaluate LSTM-like models and do not perform an in-depth comparison with competitive models, which poses a blind spot of the best performance models. We believe that our results would encourage research on other types of deep learning techniques.

Effectiveness in Expressing Long Sequences of Well Log. Most works on geoscience applications leverage the ability of recurrence models to express structured and long sequences of data (Sagheer and Kotb, 2019; Song et al., 2020; Lin et al., 2021). Such an ability becomes, therefore, a desirable and relevant property when designing/selecting a model.

To demonstrate the effectiveness of the models in expressing long sequences of data, in this experiment, we assess their predictive ability as a function of the length of the data sequence. It is important to remember that the sequence length, w, indicates the number of pairs of points (core and gamma-ray) a model receives. For this purpose, we vary the length of w using the values {2,3,4,10} and, for each value, we report a quality metric (RMSE and R^2) on the validation set. Since the sequence length is an additional hyperparameter to be set, we report the average performance of all hyperparameter configurations (i.e., combinations) that include a specific sequence length.

Fig. 5 illustrates the RMSE and R^2 metrics (y-axis) as a function of the length of the data sequence (x-axis). From this figure, all recurrence models obtained the best results using the smallest sequence length, which is 2. On the one hand, this means that recurrence models effectively adjust a value of core using only the current and the subsequent core and gamma-ray points (e.g., $\{(C_i, G_i), (C_{i+1}, G_{i+1})\}$). On the other hand, these models are sensitive in expressing long sequences of data. Specifically, with a sequence of length 4, recurrence models are no better than a simple regression model. Moreover, according to Fig. 5, LSTM seems to be more sensitive than GRU and RNN to process long sequences, since its performance degrades faster as w increases.

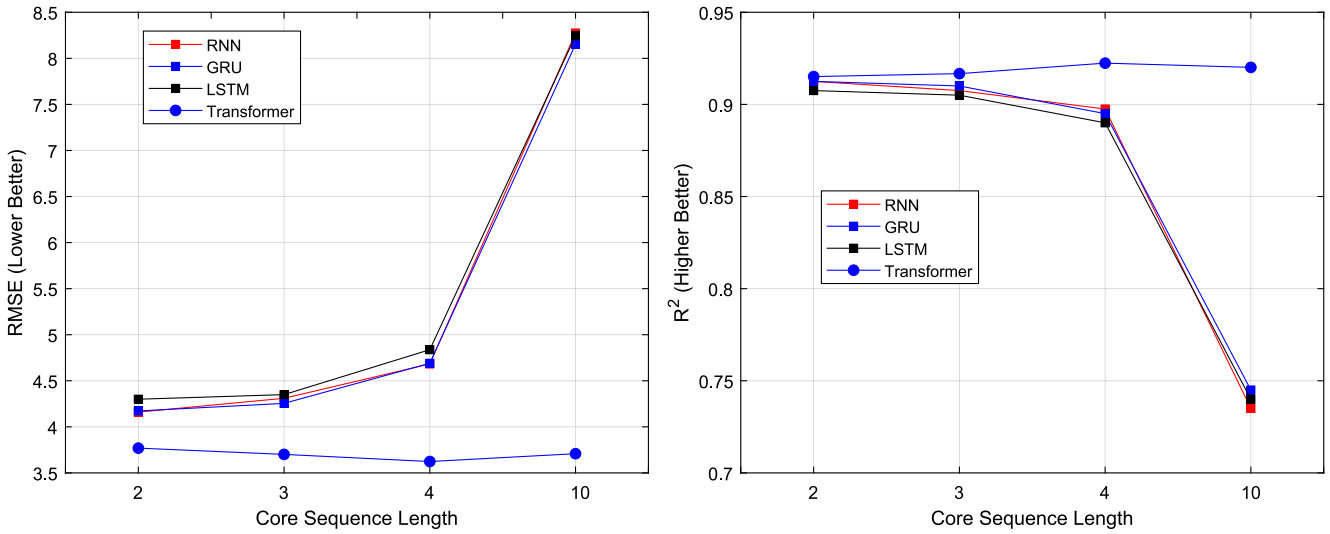

Fig. 5. Effectiveness of recurrence models as a function of data sequence length (w – more details in Section 3.1. It is evident RNN, LSTM and GRU are sensitive in processing large sequences of data since both RMSE and R^2 meet collapse when the sequence length is above three. Transformer, on the other hand, exhibits consistent results regardless of the sequence length.

Unlike recurrence models, Transformer exhibits consistent results on long sequences of data. On both RMSE and R^2, its predictive ability has a negligible variation. Taking the RMSE metric as an example, the best and worst result of Transformer is 3.62 and 3.76, respectively. Such a finding demonstrates that Transformer is not sensitive to the sequence length, as opposed to recurrence models. Notably, on a very long sequence regime (i.e., $w > 3$), all recurrence models meet collapse while Transformer successfully approximates the diagnosis done by a human expert, as indicates its (low) RMSE and (high) R^2. In particular, the predictive ability of Transformer when $w = 10$ outperforms recurrence models by an absolute difference of 4.56 in terms of RMSE.

Overall, the empirical evidence above exposes the effectiveness of the recurrence and Transformer models in expressing long sequences data, for which the latter exhibits notably better results. This behavior is expected, as previous works have demonstrated the success of Transformer in processing sequential and structured data (Vaswani et al., 2017; Devlin et al., 2019; Rae et al., 2020), even though it does not contain any recurrence mechanisms.

Relationship between Computational Cost and Predictive Ability. So far, we demonstrate that deep learning techniques outperform standard machine learning models to adjust core data. However, a practical concern when exploring deep learning models is the computation cost in both the training and testing phase, as high-capacity models tend to be computationally prohibited and hardware-costly (Tan and Le, 2019; Han et al., 2020; Tan and Le, 2021). According to our literature review, prior works on automatic well-log processing rarely report the computational performance of the methods. Specifically, only the study by Sun et al. (2019) discusses the computational cost of the models evaluated. Unfortunately, it does not take into account deep learning techniques.

To bridge this gap, in this experiment, we discuss the computational cost of the deep learning techniques. Besides, we analyze the relationship of this metric with the predictive ability of the models. For this purpose, we employ the following metrics: the number of parameters and the total of floating point operations (FLOPS).[2] Despite parameters and FLOPs being useful to indicate the computational cost, they are not always the better metrics for measuring the computational performance of a model, as evidenced by previous works (Ding et al., 2019; Liu et al., 2021a). Thereby, we also consider the average inference time (average of 30 runs), which is the time for a model to predict the regression value of a point (i.e., the adjustment of one core data). We highlight that the computational cost refers to the architecture employed in Table 1, which in turn was found by a dense hyperparameter tuning on a validation set. Table 2 shows the computational cost of the models.

According to Table 2, GRU is the costly model[3] in terms of FLOPs and the number of parameters. Similarly, Transformer exhibits a high number of parameters and FLOPs. Additionally, its inference time is the slowest compared to recurrence models. Roundly speaking, its inference takes twice of time compared to the slowest recurrence model (GRU). It turns out that we employ the Transformer architecture based on the work by Vaswani et al. (2017), which contains more layers than other models. Besides, the Transformer architecture contains complex structures such as skip connections (the output of a layer is concatenated with the result of preceding layers). Altogether, these factors incur higher inference time.

By taking into account all metrics in Table 2, RNN is the more efficient model. Since its RMSE and R^2 is only 0.79 and 0.02 inferior with respect to the best model (Transformer – see Table 1), we believe RNN emerges as a promising candidate for real-time and resource-constrained geoscience applications. We highlight that the difference in RMSE and R^2 decreases when we compare it with GRU and LSTM, further reinforcing its efficiency.

Table 2
Computational cost of the deep learning models in terms of the number of parameters, floating point operations (FLOPs) and average time (in seconds) for adjusting a set of core points. On these metrics, lower values indicate better computational performance.

Model	Parameters	FLOPs	Inference Time
RNN	321	707	0.0665
GRU	13121	26115	0.0737
LSTM	4513	8997	0.0673
Transformer	9689	9729	0.1308

[2] In this work, the term FLOPs is **unrelated** to floating-point operations **per second**.

[3] On the same number of neurons, GRU is around 20% more efficient than LSTM and this percentage is even higher when the number of neurons surpasses 64.

From the aforementioned discussion, it is evidence that low-capacity models exhibit better computational performance and vice-versa. Fig. 6 poses such a dilemma. From this figure, it is evident that the RNN, LSTM, and Transformer models are non-dominated solutions, meaning that no model (i.e., solution) dominates the others in both metrics. We highlight that GRU was dominated by LSTM, in which the latter provides superior predictive ability and computational performance. Finally, Fig. 6 shows that the accurate core adjustment by Transformers comes at the price of additional computational overhead. This issue does not necessarily mean a limitation since there are successful mechanisms for improving the computational performance of Transformers without degrading their expressive capacity, for example, pruning strategies (Michel et al., 2019; Fan et al., 2020; Han et al., 2021; Rao et al., 2021).

Qualitative Analysis of Core Adjustment. In the previous experiments, we analyze and discuss the results according to quantitative metrics. Here, we show qualitative results of the models in adjusting the core from gamma-ray. Due to the large number of models we evaluate, we show only the top(2)-performance models of each category (standard machine learning and deep learning). Besides, since a well consists of many core points (on average 501.36), we split it into two arbitrary and non-overlapping segments for better visualization.

Fig. 7 illustrates the adjustment done by different models (dashed curves) and the one made by a human expert (geologist – solid line). From this figure, both standard machine learning and deep learning techniques successfully adjust the core, as the resulting adjustment is close to the one done by the geologist. It is worth mentioning that the Transformer model aligns the cores in accordance (i.e., in the same direction) with the geologist. The high value of R^2 of Transformer further supports this observation. Fig. 8 illustrates a complete adjustment for an arbitrary well, where it is possible to observe the similarity between the human and the model (Transformer) adjustments.

In summary, the laborious and time-consuming process of adjusting core could be effectively replaced by pattern recognition methods, meaning a core diagnostic without any human intervention.

Performance on Geoscience-Related Tasks. In our last experiment, we demonstrate the performance of the models on other geoscience applications. We group these applications into two groups: classification and regression. In the first group, we consider the following tasks: (i) Lithological facies classification and (ii) well-log shape classification (log-shape for short). In the second group, we consider the prediction of oil production. In these tasks, we employ datasets publicly available by previous works. Table 3 summarizes the details of each dataset.

For each task, we follow the quality metric indicated by the authors in their original work. Regarding the training and testing sets, we employ the same as in the original work when available; otherwise, we split the data into 70% training and 30% testing. Finally, to reduce the number of experiments and hence the computational cost, we consider only deep learning techniques.

Table 4 summarizes the results. According to this table, Transform is always among the top-performance methods. Specifically, it is among the first and second-best results in 3 out of 4 applications. Surprisingly, the RNN model exhibits notable results obtaining competitive performance on all tasks. On the one hand, all models achieve similar results in lithological facies classification and prediction of oil production. On the other hand, on the log-shape classification, RNN outperforms the second best method by 14.29 percentage points. We believe that the reason for this behavior is due to the small number of samples provided by this dataset (see Table 3 last column), which favors low-capacity models such as RNN. The work by Song et al. (2020) corroborates this evidence, where the authors demonstrated a large improvement by RNN but compared to simple classifiers such as Support Vector Machines. As an additional remark, we observe that LSTM obtained the second-best results in 2 out of 4 tasks, however, the best performance remains between RNN and Transformer. Fig. 9 shows qualitative results in the lithological classification task for the datasets by Dubois et al. (2007) (left) and Sun et al. (2019) (right). This figure illustrates that the models exhibited a similar classification behavior, which clarifies their similar performance in the lithology task.

From the previous discussion, we highlight the following observations. First, our results pose a blind spot: LSTMs are not the best performing models in all geoscience tasks. This evidence reinforces the idea of evaluating more techniques rather than considering small variations of a single model. Second, RNN and Transformer emerge as strong baselines for future research.

5. Conclusions

In this work, we take a promising step towards automatic and reliable core-log processing. Our goal focused on automatically and accurately adjusting the core without any human intervention. For this purpose, we present core and the gamma-ray from a well to a regression model that outputs an adjusted core. This workflow allows employing several regression models such as ensemble-based strategies (Random Forest, Extreme Gradient Boosting and Light Gradient Boosting) and deep learning techniques (recurrence and Transformer network). We demonstrate that pattern recognition methods provide an effective and fine-grained core adjustment; thus replacing (or complementing) the laborious and time-consuming diagnostic by a human expert. Our best model, the Transformer network, obtained an R^2 of 93.63 compared to the manual adjustment (the one performed by a geologist), which indicates that pattern recognition models can achieve a fine-grained adjustment.

In contrast to existing works on geological applications, which

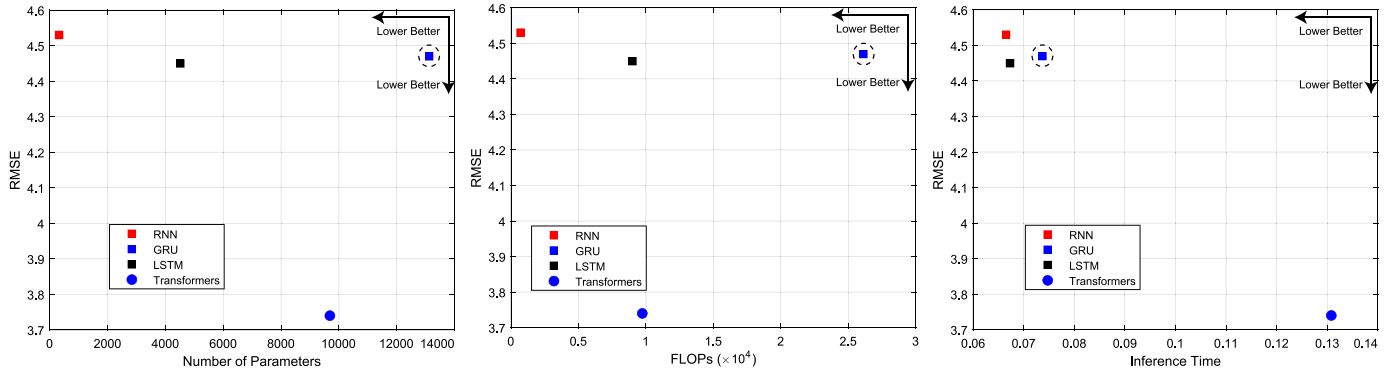

Fig. 6. Relationship between predictive ability (*y*-axis) and computational cost (*x*-axis). We measure the computational cost using well-established metrics: number of parameters (leftmost), floating point operations (FLOPs – middle) and inference time (rightmost). This figure poses a dilemma between predictive ability and computational cost: high-capacity models incur higher computationally cost. Furthermore, it is possible to observe that there is no model (i.e., solution) dominating the others on both metrics (except GRU that is dominated by LSTM). The dashed circle indicates dominated models.

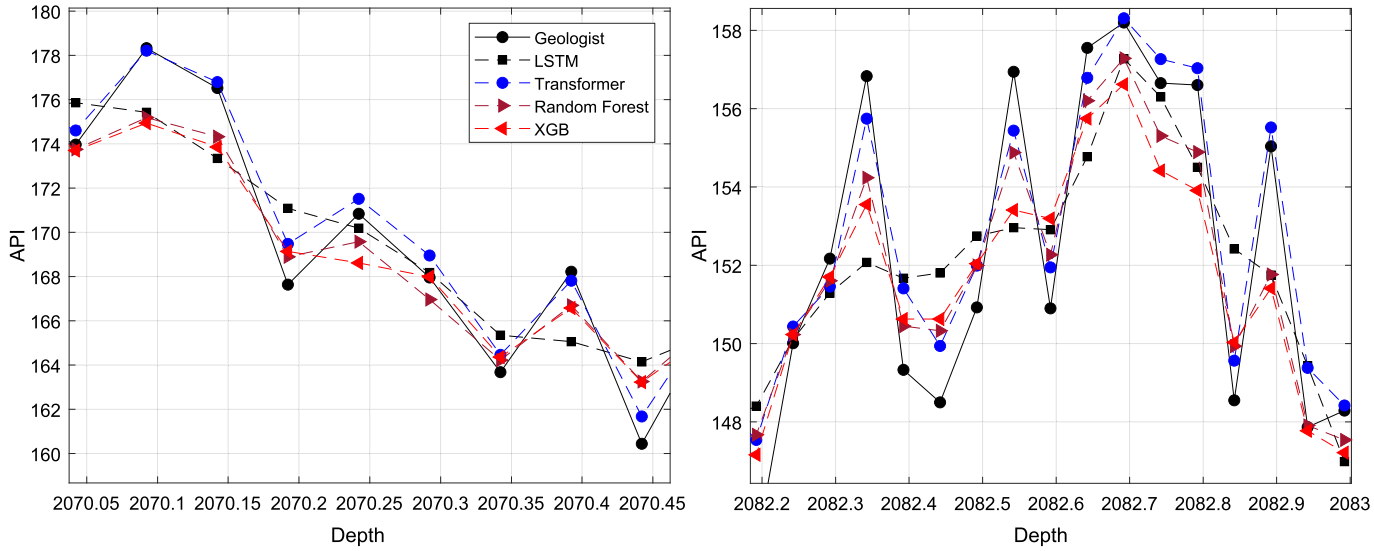

Fig. 7. Qualitative results of the best-performance models of the two categories we consider: standard machine learning and deep learning. Dashed and solid curves indicate, respectively, the adjustment made by pattern recognition models and an expert geologist. From this figure, we can see that the models successfully adjust the core. Interestingly, the adjustments made by Transformer (blue curve) proceed in the same direction as the geologist's diagnostics. In the figures, the API value (*y*-axis) is different since they come from different segments of a well. XGB stands for Extreme Gradient Boosting. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

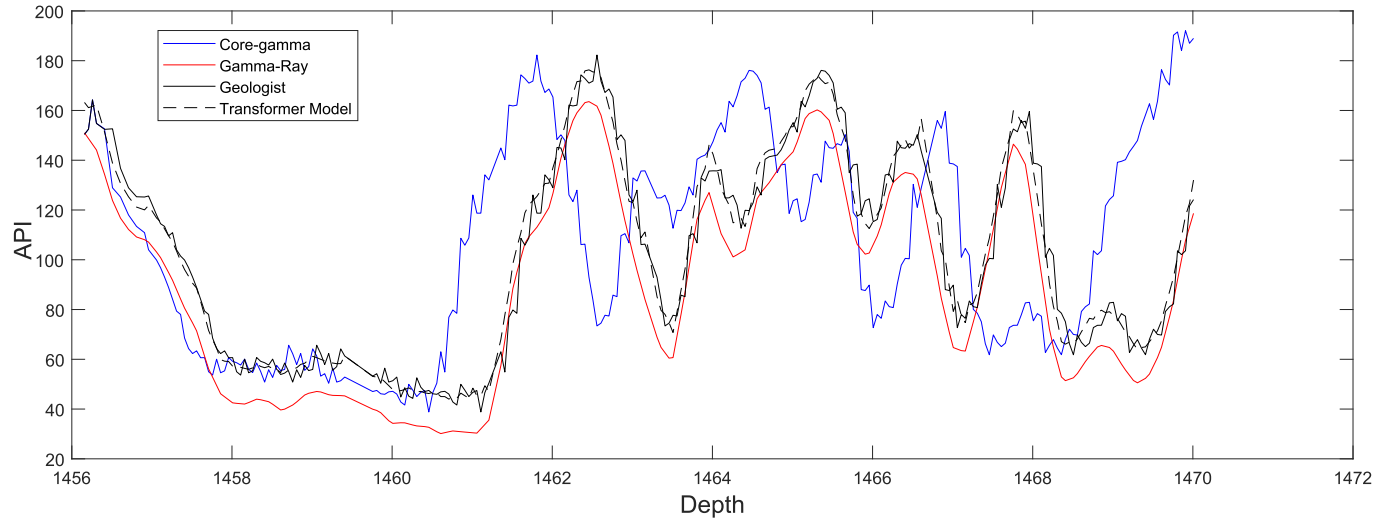

Fig. 8. Adjustment of an entire well. For better visualization, we remove the dots from the curves.

evaluate a small number of models or vary a single model, we conduct a comprehensive evaluation of several models including simple regression models, ensemble-based and deep learning techniques. In particular, we evaluate more than 200 unique models. Regarding the deep learning techniques, we explore recurrence models and the relatively recent Transformers. Through extensive analysis, we demonstrate that deep learning techniques outperform other techniques by a large margin. More specifically, a simple RNN model obtained superior results than the best ensemble-based model and this result is even better when the Transformer model is taken into account. We empirically demonstrate that the reason for such results relies on the capacity of Transformers in expressing long sequences of data, which translates directly into a better core adjustment. Importantly, in the long-sequence regime, other deep learning techniques led to inferior results when attempting to learn patterns. From a qualitative perspective, we show that the adjustment made by Transformers follows the same direction as the diagnostic of a human expert (geologist), while other models exhibit a certain disagreement. To the best of our knowledge, this is the first study to investigate Transformer models on geoscience applications.

We further analyze the performance of both Transformers and recurrence models on related applications such as lithology and log-shape classification, and prediction of oil production. In the lithology classification, the models successfully classify different facies categories such as coarse sandstone, medium sandstone, fine sandstone, siltstone, dolomite, limestone and mudstone. Regardless the task, Transformers are always among the best-performance models; thus emerging as a strong baseline for future research. Furthermore, we observe that LSTM exhibits no drastic difference from other recurrence models and, sometimes, even a simple RNN obtains superior results. Therefore, as opposed to existing works, we believe that much more progress could be done by evaluating more deep learning techniques with careful hyperparameter tuning.

Throughout our work, we also discuss the computational cost of several models and we evidence that there exists a dilemma between

Table 3
Overall characteristics of datasets of geological tasks we evaluate the models. The first and second columns indicate the work in which we take the dataset and its corresponding geological task, in this order. The third column shows the categories of each task. The last column reports the number of samples (training and testing) available in each dataset.

DataSet	Task	Categories	Number of Samples
Sun et al. (2019)	Lithological Facies	coarse sandstone, medium sandstone, fine sandstone, siltstone, dolomite, limestone, mudstone	3711
Dubois et al. (2007)	Lithological Facies	nonmarine sandstone, coarse siltstone, fine siltstone, marine siltstone and shale, mudstone, wackestone, dolomite, packstone-grainstone, phylloid-algal bafflestone	3232
Song et al. (2020)	Log-Shape	Bell shape, Funnel shape, Egg Shape, Cylinder shape	85
Sagheer and Kotb (2019)	Oil Prediction	–	43

Table 4
Performance of deep learning techniques on several geoscience applications. We group these applications into two categories: classification (left side) and regression (right side). These results are the first step towards a benchmark in geoscience tasks. Bold and underline indicate the top-1 and top-2 best results, in this order. The arrows indicate which direction is better for each task.

	Classification (Accuracy ↑)			Regression (RMSE ↓)
	Lithological Facies (Sun et al., 2019)	Lithological Facies (Dubois et al., 2007)	Log-Shape (Song et al., 2020)	Oil Prediction (Sagheer and Kotb, 2019)
RNN	86.34	**90.07**	**92.86**	**0.030**
GRU	84.70	89.14	71.43	0.038
LSTM	86.07	89.91	64.29	0.037
Transformer	**88.80**	87.51	78.57	**0.030**

predictive ability and computational performance: higher-capacity models incur higher computational costs. This is a useful information for resource-constrained and real-time applications, but, unfortunately, it is rarely reported by previous research. Our work, on the other hand, bridges this gap.

6. Open problems

Despite the positive results in core processing and geoscience-related applications, there is much room for improvements. We highlight the following. (i) *Ensemble.* According to our results, models based on ensemble attain competitive performance underperforming deep learning techniques only. Thus, integrating ensemble mechanisms to strong classifiers emerges as a potential line of research. For example, we would combine RNN, GRU, LSTM and Transformer to compose an ensemble. (ii) *Architecture Design.* Due to resource constraints, we only vary the architecture proposed by Vaswani et al. (2017) during the Transformer hyperparameter tuning. However, previous works have demonstrated that small variations in architecture design lead to more stable training and better predictive ability (Liu et al., 2020;, Parisotto et al., 2020; Davis et al., 2021). Thus, we believe that Transformers could attain even better performance taking into account other basis architectures. In the direction of architecture exploration, we could combine recurrence and convolutional architectures with Transformers, as previous efforts have confirmed positive results in this sense (Bello et al., 2019; Yang et al., 2021). (iii) *Transfer Learning.* In the computer vision community, a common practice is to learn a model using a large-scale dataset and, then, fine-tune (i.e., transfer) its parameters on a small (proxy) dataset. From a practical perspective, transfer learning often leads to better results than training a model from scratch (Hendrycks et al., 2019; Shafahi et al., 2020; Kolesnikov et al., 2020), especially when using Transformer-like models (Liu et al., 2021b). For example, the visual Transformer by Dosovitskiy et al. (2021) obtains state-of-the-art performance on small datasets when transferring its parameters from a large-scale dataset. More aligned with the scope of our work, Yang et al. (2021) showed positive results in transferring acoustic models for time series applications. Despite these observations, to the best of our knowledge, this topic remains unexplored in geoscience applications. Therefore, transfer learning is an encouraging line of research.

Code availability section

Contact: arturjordao@dcc.ufmg.br.

Hardware requirements: We test all code on an Intel i5-1135G7 2.40GHz CPU with 8GB RAM.

Program language: Python.

Packages required: TensorFlow 2.4.1.

License: MIT License.

Program size: 2.14 MB (2.254.841 bytes).

Code is available at the link: https://github.com/arturjordao/TowardsAutomaticAccurateCore-logProcessing

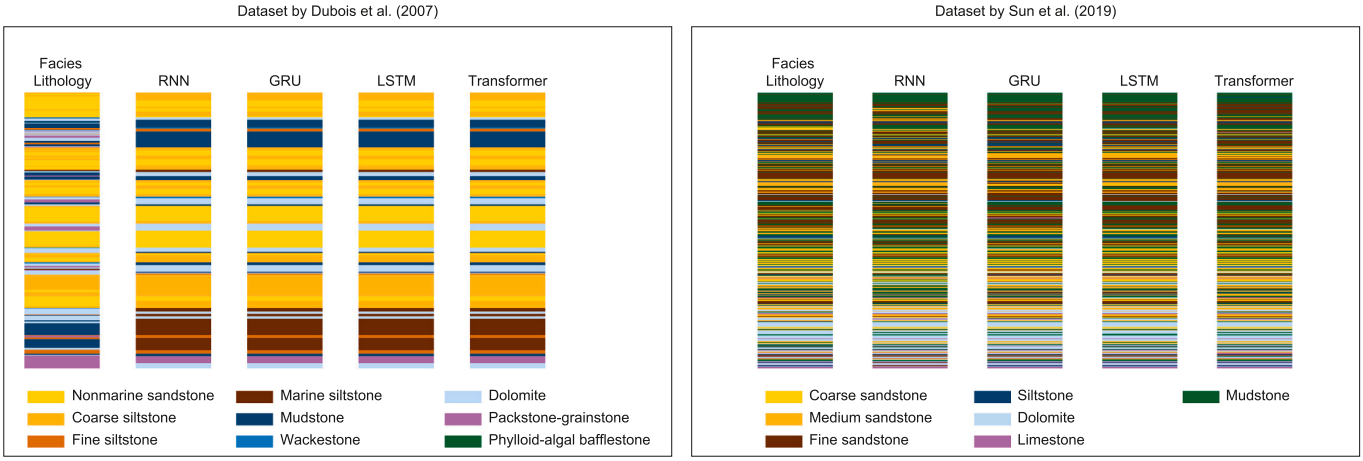

Fig. 9. Qualitative results from the lithology classification. Left: Dataset by Dubois et al. (2007). Right: Dataset by Sun et al. (2019).

CRediT authorship contribution statement

Artur Jordao: Conceptualization, Methodology, Software, Formal analysis, Investigation, Writing – original draft. **Joao Paulo Ponte Souza:** Methodology, Investigation, Writing – review & editing. **Michelle Chaves Kuroda Avansi:** Methodology, Investigation, Writing – review & editing. **Marcelo Fagundes de Rezende:** Formal analysis, Investigation, Resources, Data curation, Writing – review & editing. **Helio Pedrini:** Conceptualization, Formal analysis, Writing – review & editing, Project administration. **Alexandre Campane Vidal:** Conceptualization, Formal analysis, Writing – review & editing, Project administration, Funding acquisition.

Declaration of Competing Interest

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

Data availability

Data will be made available on request.

Acknowledgments

The authors would like to thank the Petróleo Brasileiro S.A. (Petrobras) e Agência Nacional do Petróleo, Gás Natural e Biocombustíveis (Grant 2018/00607-6).

References

Abbas, M.A., Al-Mudhafar, W.J., 2021. Lithofacies classification of carbonate reservoirs using advanced machine learning: A case study from a southern iraqi oil field. In: OTC Offshore Technology Conference.
Al-Mudhafar, W.J., 2017a. Integrating kernel support vector machines for efficient rock facies classification in the main pay of zubair formation in south rumaila oil field, Iraq. Model. Earth Syst. Environ. 3, 2363–6211.
Al-Mudhafar, W.J., 2017b. Integrating well log interpretations for lithofacies classification and permeability modeling through advanced machine learning algorithms. J. Pet. Explor. Prod. Technol. 7, 1023–1033.
Al-Mudhafar, W.J., 2020. Integrating machine learning and data analytics for geostatistical characterization of clastic reservoirs. J. Pet. Sci. Eng. 195, 107837.
Al-Mudhafar, W.J., Wood, D.A., 2022. Tree-based ensemble algorithms for lithofacies classification and permeability prediction in heterogeneous carbonate reservoirs. In: OTC Offshore Technology Conference.
Al-Mudhafar, W.J., Abbas, M.A., Wood, D.A., 2022. Performance evaluation of boosting machine learning algorithms for lithofacies classification in heterogeneous carbonate reservoirs. Mar. Pet. Geol. 145, 105886.
Ameur-Zaimeche, O., Zeddouri, A., Heddam, S., Kechiched, R., 2020. Lithofacies prediction in non-cored wells from the sif fatima oil field (berkine basin, southern Algeria): a comparative study of multilayer perceptron neural network and cluster analysis-based approaches. J. Afr. Earth Sci. 166, 103826.
Asante-Okyere, S., Shen, C., Ziggah, Y.Y., Rulegeya, M.M., Zhu, X., 2020. Principal component analysis (pca) based hybrid models for the accurate estimation of reservoir water saturation. Comput. Geosci. 145, 1–11.
Bello, I., Zoph, B., Le, Q., Vaswani, A., Shlens, J., 2019. Attention augmented convolutional networks. In: International Conference on Computer Vision (ICCV), pp. 3285–3294.
Carion, N., Massa, F., Synnaeve, G., Usunier, N., Kirillov, A., Zagoruyko, S., 2020. End-to-end object detection with transformers. In: European Conference on Computer Vision (ECCV), pp. 213–229.
Datskiv, O., Struk, A., Maksymenko, M., Veselovska, M., Bondarenko, O., Karpiv, V., 2020. Framework for automatic globally optimal well log correlation. In: Neural Information Processing Systems (NeurIPS) Workshop on AI for Earth Sciences, pp. 1–5.
Davis, J.Q., Gu, A., Choromanski, K., Dao, T., Ré, C., Finn, C., Liang, P., 2021. Catformer: Designing stable transformers via sensitivity analysis. In: International Conference on Machine Learning (ICML), pp. 2489–2499.
Devlin, J., Chang, M., Lee, K., Toutanova, K., 2019. BERT: pre-training of deep bidirectional transformers for language understanding. In: Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, pp. 4171–4186 (NAACL-HLT).
Ding, X., Ding, G., Guo, Y., Han, J., Yan, C., 2019. Approximated oracle filter pruning for destructive CNN width optimization. In: International Conference on Machine Learning (ICML), pp. 1607–1616.
Dong, X., Yang, Y., 2020. Nas-bench-201: Extending the scope of reproducibl neural architecture search. In: International Conference on Learning Representations (ICLR), pp. 1–16.
Dosovitskiy, A., Beyer, L., Kolesnikov, A., Weissenborn, D., Zhai, X., Unterthiner, T., Dehghani, M., Minderer, M., Heigold, G., Gelly, S., Uszkoreit, J., Houlsby, N., 2021. An image is worth 16x16 words: Transformers for image recognition at scale. In: International Conference on Learning Representations (ICLR), pp. 1–21.
Dubois, M.K., Bohling, G.C., Chakrabarti, S., 2007. Comparison of four approaches to a rock facies classification problem. Comput. Geosci. 33, 599–617.
Fan, A., Grave, E., Joulin, A., 2020. Reducing transformer depth on demand with structured dropout. In: International Conference on Learning Representations (ICLR), pp. 1–16.
Han, K., Wang, Y., Zhang, Q., Zhang, W., Xu, C., Zhang, T., 2020. Model rubik's cube: Twisting resolution, depth and width for tinynets. In: Neural Information Processing Systems (NeurIPS), pp. 19353–19364.
Han, Y., Huang, G., Song, S., Yang, L., Wang, H., Wang, Y., 2021. Dynamic neural networks: a survey. IEEE Trans. Pattern Anal. Mach. Intell. 1–20.
Hendrycks, D., Lee, K., Mazeika, M., 2019. Using pre-training can improve model robustness and uncertainty. In: International Conference on Machine Learning (ICML), pp. 2712–2721.
Imamverdiyev, Y., Sukhostat, L., 2019. Lithological facies classification using deep convolutional neural network. J. Pet. Sci. Eng. 174, 216–228.
Ioffe, S., Szegedy, C., 2015. Batch normalization: Accelerating deep network training by reducing internal covariate shift. In: International Conference on Machine Learning (ICML), pp. 448–456.
Karimi, A.M., Sadeghnejad, S., Rezghi, M., 2021. Well-to-well correlation and identifying lithological boundaries by principal component analysis of well-logs. Comput. Geosci. 157, 1–13.
Kolesnikov, A., Beyer, L., Zhai, X., Puigcerver, J., Yung, J., Gelly, S., Houlsby, N., 2020. Big transfer (bit): General visual representation learning. In: European Conference on Computer Vision (ECCV), pp. 491–507.
Kreuzer, D., Beaini, D., Hamilton, W.L., Létourneau, V., Tossou, P., 2021. Rethinking graph transformers with spectral attention. In: Neural Information Processing Systems (NeurIPS), pp. 1–18.
Lee, J., Chung, S., 2020. Robust training with ensemble consensus. In: International Conference on Learning Representations (ICLR), pp. 1–15.
Lin, J., Li, H., Liu, N., Gao, J., Li, Z., 2021. Automatic lithology identification by applying LSTM to logging data: a case study in X tight rock reservoirs. IEEE Geosci. Remote Sens. Lett. 18, 1361–1365.
Liu, L., Liu, X., Gao, J., Chen, W., Han, J., 2020. Understanding the difficulty of training transformers. In: Conference on Empirical Methods in Natural Language Processing (EMNLP), pp. 5747–5763.
Liu, L., Zhang, S., Kuang, Z., Zhou, A., Xue, J., Wang, X., Chen, Y., Yang, W., Liao, Q., Zhang, W., 2021a. Group fisher pruning for practical network compression. In: International Conference on Machine Learning (ICML), pp. 7021–7032.
Liu, Y., Sangineto, E., Bi, W., Sebe, N., Lepri, B., Nadai, M.D., 2021b. Efficient training of visual transformers with small datasets. Neural Inform. Process. Syst. (NeurIPS) 1–13.
Michel, P., Levy, O., Neubig, G., 2019. Are sixteen heads really better than one?. In: Neural Information Processing Systems (NeurIPS), pp. 14014–14024.
Parisotto, E., Song, H.F., Rae, J.W., Pascanu, R., Gülçehre, Ç., Jayakumar, S.M., Jaderberg, M., Kaufman, R.L., Clark, A., Noury, S., Botvinick, M., Heess, N., Hadsell, R., 2020. Stabilizing transformers for reinforcement learning. In: International Conference on Machine Learning (ICML), pp. 7487–7498.
Rae, J.W., Potapenko, A., Jayakumar, S.M., Hillier, C., Lillicrap, T.P., 2020. Compressive transformers for long-range sequence modelling. In: International Conference on Learning Representations (ICLR), pp. 1–19.
Rao, Y., Zhao, W., Liu, B., Lu, J., Zhou, J., Hsieh, C., 2021. Dynamicvit: Efficient vision transformers with dynamic token sparsification. In: Neural Information Processing Systems (NeurIPS), pp. 1–13.
Sagheer, A.E., Kotb, M., 2019. Time series forecasting of petroleum production using deep LSTM recurrent networks. Neurocomputing 323, 203–213.
Santos, D.T.D., Roisenberg, M., Nascimento, M.D.S., 2022. Deep recurrent neural networks approach to sedimentary facies classification using well logs. IEEE Geosci. Remote Sens. Lett. 19, 1–5.
Santurkar, S., Tsipras, D., Ilyas, A., Madry, A., 2018. How does batch normalization help optimization?. In: Neural Information Processing Systems (NeurIPS), pp. 2488–2498.
Schmidt, R.M., Schneider, F., Hennig, P., 2021. Descending through a crowded valley - benchmarking deep learning optimizers. In: International Conference on Machine Learning (ICML), pp. 9367–9376.
Schneider, F., Balles, L., Hennig, P., 2019. Deepobs: a deep learning optimizer benchmark suite. In: International Conference on Learning Representations (ICLR), pp. 1–14.
Serra, O., 1986. Fundamentals of well-log interpretation: the interpretation of logging data. Elsevier, Amsterdam, pp. 1–679.
Shafahi, A., Saadatpanah, P., Zhu, C., Ghiasi, A., Studer, C., Jacobs, D.W., Goldstein, T., 2020. Adversarially robust transfer learning. In: International Conference on Learning Representations (ICLR), pp. 1–14.
Singh, H., Seol, Y., Myshakin, E.M., 2020. Automated well-log processing and lithology classification by identifying optimal features through unsupervised and supervised machine-learning algorithms. SPE J. 25, 2778–2800.
Song, S., Hou, J., Dou, L., Song, Z., Sun, S., 2020. Geologist-level wireline log shape identification with recurrent neural networks. Comput. Geosci. 134, 104313.
Sun, J., Li, Q., Chen, M., Ren, L., Huang, G., Li, C., Zhang, Z., 2019. Optimization of models for a rapid identification of lithology while drilling - a win-win strategy based on machine learning. J. Pet. Sci. Eng. 176, 321–341.

Tan, M., Le, Q.V., 2019. Efficientnet: Rethinking model scaling for convolutional neural networks. In: International Conference on Machine Learning (ICML), pp. 6105–6114.
Tan, M., Le, Q.V., 2021. Efficientnetv2: Smaller models and faster training. In: International Conference on Machine Learning (ICML), pp. 10096–10106.
Tang, H., White, C.D., 2008. Multivariate statistical log log-facies classification on a shallow marine reservoir. J. Pet. Sci. Eng. 61, 88–93.
Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A.N., Kaiser, L., Polosukhin, I., 2017. Attention is all you need. In: Neural Information Processing Systems (NeurIPS), pp. 5998–6008.
Wang, S., Chen, S., 2019. Application of the long short-term memory networks for well-testing data interpretation in tight reservoirs. J. Pet. Sci. Eng. 183, 106391.
Wen, Y., Jerfel, G., Muller, R., Dusenberry, M.W., Snoek, J., Lakshminarayanan, B., Tran, D., 2021. Combining ensembles and data augmentation can harm your calibration. In: International Conference on Learnin Representations (ICLR), pp. 1–21.
Xie, E., Wang, W., Yu, Z., Anandkumar, A., Alvarez, J.M., Luo, P., 2021. Segformer: simple and efficient design for semantic segmentation with transformers. In: Neural Information Processing Systems (NeurIPS), pp. 1–14.
Yang, C.H., Tsai, Y., Chen, P., 2021. Voice2series: Reprogramming acoustic models for time series classification. In: International Conference on Machine Learning (ICML), pp. 11808–11819.
Zaidi, S., Zela, A., Elsken, T., Holmes, C.C., Hutter, F., Teh, Y.W., 2021. Neural ensemble search for uncertainty estimation and dataset shift. In: Neural Information Processing Systems (NeurIPS), pp. 1–32.
Zeyer, A., Doetsch, P., Voigtlaender, P., Schlüter, R., Ney, H., 2017. A comprehensive study of deep bidirectional LSTM RNNS for acoustic modeling in speech recognition. In: International Conference on Acoustics, Speech and Signal Processing (ICASSP), pp. 2462–2466.
Zhang, J., Liu, S., Li, J., Liu, L., Liu, H., Sun, Z., 2017. Identification of sedimentary facies with well logs: an indirect approach with multinomial logistic regression and artificial neural network. Arab. J. Geosci. 10, 247.
Zhou, K., Zhang, J., Ren, Y., Huang, Z., Zhao, L., 2020. A gradient boosting decision tree algorithm combining synthetic minority oversampling technique for lithology identification. Geophysics 85. WA147–WA158.