GeoTS: A TSC Framework for Estimating Geological Formation to Model Carbon Storage Reservoirs

Shwetha Salimath
Laboratoire Interdisciplinaire des Sciences du Numérique,
Paris-Saclay University, CNRS, SLB
Montpellier, France
shwetha.salimath@universite-paris-saclay.fr

Francesca Bugiotti
Laboratoire Interdisciplinaire des Sciences du Numérique,
CentraleSupélec, CNRS
Paris, France
francesca.bugiotti@centralesupelec.fr

Sylvain Wlodarczyk
SLB
Montpellier, France
swlodarczyk@slb.com

Abstract

In geoscience, it is necessary to study the lithography of the Earth's subsurface, which consists of different stratified layers called geological formations. This study performs well correlation task to model and characterize reservoirs. This operation links the beginning of specific geological formations called tops using measurements from drilled wells. Although data are abundant, the traditional algorithms used for well correlation are semi-automated, requiring significant time and high computational cost. This paper introduces GeoTS, a Python library to apply cutting-edge time series classification models to perform well correlation in a completely automated setting. As input, take the drilling trajectory depth and gamma-ray well logs, which measure the natural radioactivity across the well depth trajectory. The top depths of the formations are predicted as an output. The gamma-ray signatures are extracted around the top depths assigned by geologists. Preprocessing is performed to clean and cluster these signatures using the Dynamic Time Wrapping (DTW) distance and HDBSCAN. Implementation of existing deep learning architectures (FCN, InceptionTime, XceptionTime, XCM, LSTM-FCN) and new architecture (LSTM-2dCNN, LSTM-XCM) are performed. Our experiments demonstrate faster computation with an increase in accuracy. GradCAM has also been implemented for model explainability. Experiments were performed using Colorado oil fields and deployed on Wyoming oil fields. The deployment has provided us with critical insights regarding the improvements needed.

Keywords

Deep Learning, Time series classification, Carbon storage, Reservoirs, Lithography

ACM Reference Format:
Shwetha Salimath, Francesca Bugiotti, and Sylvain Wlodarczyk. 2025. GeoTS: A TSC Framework for Estimating Geological Formation to Model Carbon Storage Reservoirs. In *Proceedings of the 31st ACM SIGKDD Conference on Knowledge Discovery and Data Mining V.2 (KDD '25), August 3–7, 2025, Toronto, ON, Canada.* ACM, New York, NY, USA, 11 pages. https://doi.org/10.1145/3711896.3737228

KDD '25, August 3–7, 2025, Toronto, ON, Canada

ACM ISBN 979-8-4007-1454-2/2025/08
https://doi.org/10.1145/3711896.3737228

1 Introduction

Energy is crucial for industrial and economic growth [6, 45], but this has led to rising fossil fuel consumption, increasing carbon dioxide (CO_2) emissions, and contributing to global warming [23]. From the 1960s to 2023, CO_2 emissions from fossil fuels surged from about 11 billion tons to approximately 36.6 billion tons [29]. Research is focused on reducing carbon emissions through renewable energy sources and carbon capture and storage (CCS) [13, 30]. CCS is particularly important for power and industrial plants, capturing CO_2 through various methods, such as post-combustion [12]. The captured CO_2 is then transported to storage sites, often deep geological formations, where it is injected as a high-pressurized liquid, creating a closed-loop system [19, 39].

The CCS storage sites must be monitored to prevent leakage and potential long-term stewardship [9, 34]. The first step is to understand the geological characteristics of the storage reservoir [50]. One of the most widely used methods for studying geological formations is drilling a well and obtaining logging data. Well logging is a method for collecting and analyzing measurement data regarding the rocks' electrical, radioactive, and acoustic properties [15]. The well logs identify the formations depending on the rock's nature in a depth interval. As a result, it becomes possible to estimate the structure and composition of the subsurface and model the storage reservoir by correlating the formations of relatively closely located wells. This process is called well correlation [5, 12].

Problem The main problem is that most of the information related to geological formation is currently estimated by geologists using mud logs and the rocks extracted during borehole drilling to study their characteristics. This process is tedious and time-consuming [44]. A marker top refers to the beginning of the intersection between the well and the formation. According to domain experts, a pattern is observed at the change of the formations, known as marker signature. The marker signatures are expected to be similar across the wells, at least within a region [33]. Machine learning (ML) and DTW algorithms [35] have been used to semi-automate the process of identifying these formations using wireline logs [10, 28]. Wireline logs are records of formation properties across the well depth measured using a variety of sensors. The nearest neighboring wells are selected through clustering. Observing these well logs, experts select the marker signature manually.

DTW is used to search for the best match in the new wells. However, this method still has a high time complexity and is subject to human bias [2]. Using deep learning to extract information from wireline logs efficiently would save time and resources.

Proposed Solution To completely automate the process, we first cluster the wells according to their geological location. We then extract a sequence of formation that follows an order of occurrence. The marker signature are clustering using the DTW distance to be able to group and identify a distinct marker pattern in the region. We approach the problem as pattern identification using time series classification(TSC) [40, 47].

The paper documents a comprehensive framework encompassing everything from cleaning well log data to estimating the marker depth for geological formations. GeoTS allows to compare state of the art time series classification models such as FCN [31], LSTM [18], Inception Time [22], LSTM-FCN [24], and new models for well-correlation tasks. We have used the DTW-based algorithm as our baseline. We have implemented GradCAM [43] for time series, which allows us to understand the feature importance and selection for the classification task. Our framework has achieved a higher score than the industrial baseline process.

The contributions of this paper are summarized as follows:

- Provides a new streamlined framework called GeoTS for reservoir modeling, allowing its utilization by geologists to understand the lithography of the earth's subsurface.
- Introduces new hybrid TSC algorithms and provides model explainability with GradCAM implementation for time series.
- Deployment on real-world dataset to compare the results of GeoTS framework.

Section 2, describes the dataset used and defines evaluation metrics, and in Section 3, discusses the state of the art in the domain. Section 5 illustrates the clustering algorithms and the data cleaning process. Section 6 explains the methodology and compares our models with the baseline model. We discuss the results of the experiments conducted and the model explainability in Section 7. The deployment of the framework is discussed in Section 8 and end with a conclusion in Section 9.

2 Dataset and evaluation metrics

This section introduces the well log dataset used in the experiments and definitions of evaluation metrics used to evaluate and compare model performances.

2.1 Dataset

Well logging has two main types: (i) geological well logs, made from mud logs and rock samples, and (ii) geophysical well logs, which record physical measurements collected by instruments lowered into the drilled hole. These geophysical well logs are analyzed to identify lithologies, to differentiate between porous and non-porous rock, and to get reservoir characteristics from subsurface formations [3, 26]. This paper focuses on geophysical well logs, which we will henceforth refer to simply as well logs.

Well logs record the magnitude of a specific formation property, such as resistivity, gamma radiation, density, neutron porosity,

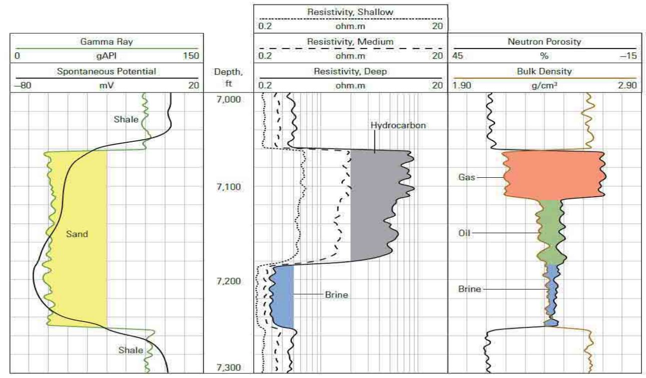

Figure 1: Example of gamma ray, resistivity, and neutron porosity well logs [3]

and sonic properties, measured by the tool as it traverses an interval defined by depth. Well logs present a concise, detailed plot of formation parameters vs depth, as shown in Figure 1. Our study uses gamma rays (GR) because they are present across many wells. The GR device measures naturally occurring radioactivity from the formation due to the presence of radioactive elements, primarily potassium, uranium, and thorium [8, 17].

Colorado dataset Data were sourced from the Colorado state government website [1], where we obtained two files for each well: one with well log measurements at various depths (including GR, density, and resistivity) and another is well report containing geological formations with their marker depths. We have used a subset of three markers in 800 wells, with log depths ranging from approximately 300 ft to 7200 ft. The markers of interest are Niobrara, Fort Hays, and Codell. Of the 800 wells, 150 are testing wells with manually verified marker depths by geophysicists due to the inherent noise in log data. The noise can be from the measuring instruments or an error in picking the top marker depth. Figure 2 displays the GR logs for select wells, showing overall variability and highlighting a section near the bottom that contains the formations to identify. The marker signature of these formations is also extracted and displayed. The green and red dotted lines in the middle subplot of Figure 2 indicate the correct depths of the marker for the marker originally assigned incorrectly. The black dotted line indicates the midpoint of the formation marker signature, considered as the marker depth.

2.2 Evaluation Metrics

Marker propagation is to be able to predict the marker depth using the GR well log data and the well location. We use accuracy, mean absolute error (MAE) and recall to compare model performance. All these terms are explained in this subsection.

Accuracy is computed as the proportion of labels that are correctly predicted over all of the labels for the classification model.

MAE The mean absolute error (MAE) is the absolute difference between predicted and actual depth for each marker top. MAE is then averaged over all the wells.

Recall The recall is calculated after applying a threshold T. If the absolute difference between the predicted depth $\hat{d}$ and actual

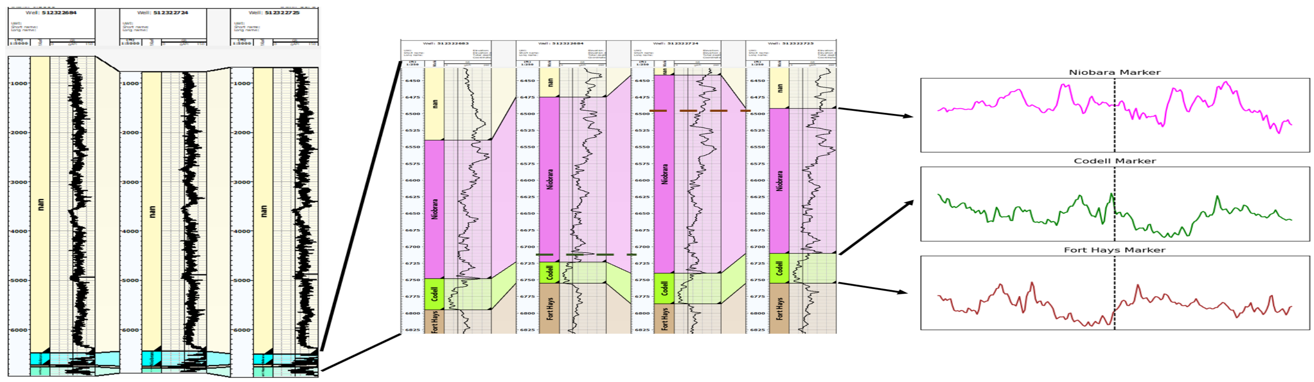

Figure 2: Gamma ray logs for different wells within a region

depth d is less than the threshold T, we consider it as a correct prediction and assign it as a true positive else false negative. The recall is then calculated per class and macro averaged. It measures how many of the labels for a class are correctly predicted as:

$$IF(|\hat{d} - d| < T) = TP$$

3 Related Work

The DTW algorithm computes the optimal correlation between two logs by minimizing the cumulative distance along a correlation path [35]. It is used for well correlation as it accommodates the time variations and has better non-linear behavior than other correlation techniques [16]. In a case study conducted on 22 wells of the Karatube Oilfield in Kazakhstan, a minimum spanning tree algorithm was used to find well-pairs, and later DTW was performed to see the correlation between them [52].

Active research is being conducted to use machine learning to automate well correlation. One hundred wells from Kansas state oil and gas fields, USA, were used to train and test a 1D CNN Autoencoder [?]. Nearest wells are computed based on proximity. During training, the autoencoder tries to recognize a series of matching points from the nearest wells to match geological well tops. These methods strongly depend on the reference wells, with respect to which the other wells are correlated.

3.1 Time series classification

Intensive research has been done on time series classification for industrial applications [42, 49]. The traditional methodology uses a nearest-neighbor classifier coupled with distance functions. DTW is the most popular distance function and is considered as a baseline for many applications [21]. We explore various deep learning models for classification used in the study.

LSTM network [18] - Effective for tasks like time forecasting and anomaly detection, it can learn long-term dependencies through its memory structure [36]. The Bi-LSTM processes sequences in both forward and backward directions.

FCN network [7, 31] - Utilizes 1D CNN blocks for time series, featuring layers that upsample and downsample data. Average pooling reduces output dimensions before feeding into a linear layer to obtain class probabilities as output.

Resnet Network [14] - Resnet blocks are used to build the network. The main concept introduced is residual connections in the block, which helps in convergence and stabilization of the network. These blocks are used to up-sample the input data and then apply average adaptive pooling followed by a linear layer.

InceptionTime [22, 51] - It is an ensemble of six Inception modules [48] with different randomly initialized weight values. The inception module is composed of CNN layers with varying filter lengths. This allows the neural network to extract relevant features from long and short-time series.

XceptionTime [41] - Its architecture integrates depthwise separable CNN with different filter sizes, adaptive average pooling, and a non-linear normalization technique. They are concatenated in series to capture temporal and spatial information without manual feature extraction.

XCM [11] - Designed for multivariate time series, it merges 2D and 1D CNNs in a parallel architecture with upsampling and downsampling layers. Their output is concatenated in the first dimension and passed through an upsampling CNN layer. A linear layer at the end follows a pooling layer to get the output.

LSTM-FCN [24, 38] - Combines LSTM and FCN in parallel, allowing for concatenation of results and the potential addition of an attention mechanism to weigh the importance of sequence components. The ALSTM-FCN adds an attention layer to the LSTM layer.

4 GoeTS framework

This section presents a detailed explanation of the GeoTS framework through a flowchart. Oil fields often extend hundreds of kilometers, resulting in a wide spatial distribution of wells used for extraction. Due to the potential variation in subsurface stratigraphy over such large distances, it becomes essential to group wells based on their geographic locations. From these clusters, a smaller subset is selected for focused analysis. Within this subset, only a few wells typically contain detailed, labeled information about the geological formations they penetrate. These labeled wells serve as the foundation for training a model that can infer similar information for the remaining unlabeled wells in the cluster.

To ensure meaningful propagation, the framework only considers formation tops, referred to as markers, that appear in at least

10% of tops. of the regional wells. This threshold helps to guarantee that the formations are likely to occur in the unlabeled wells. For each marker, a specific segment of the well log data surrounding its depth is extracted. This segment, known as the marker signature, captures the characteristic pattern of the formation. These signatures are then clustered using the DTW distance to identify and remove outliers. Clusters that are too small or inconsistent are excluded from the training process, as the goal is to retain only those patterns that are reliably observed across neighboring wells.

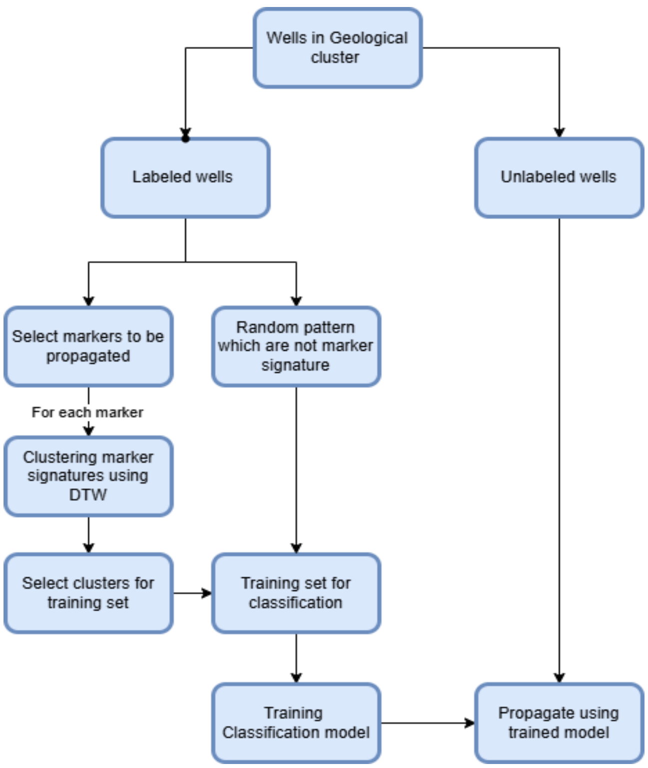

Figure 3: Flowchart of GeoTS framework

Once the marker signatures are cleaned and clustered, they form the training dataset for a classification model. The model is then applied to the unlabeled wells using a sliding window approach, scanning through the well logs to identify the depth at which each marker is most likely to occur. Since a large portion of the well log data does not correspond to any known marker, a special 'None' class is introduced. This class is trained using data sampled from random depths across various wells, enabling the model to distinguish between marker and non-marker regions effectively. Through this process, the GeoTS framework enables robust and scalable interpretation of geological formations across large and complex oil fields.

5 Data cleaning

The well log data set is naturally noisy due to fluid turbulence, subsurface conditions, and technical errors such as sensor noise or hardware failure. This section outlines the data cleaning and preprocessing pipeline in the GeoTS framework. Initially, missing values are addressed by imputation with averages or nearest neighbors, ensuring uniform time stamps and units.

After the initial cleaning, the GR reading around the marker depth is extracted as the marker signature with a specific window size. Geologists assign marker depths, leading to subjective bias in interpretation. In cases where the readings are wrong, the marker depth assigned need to be corrected. Therefore, further cleaning is necessary to identify wells with clear marker signatures. Clustering with DTW distance is employed to filter wells for training the deep learning model.

DTW [35] distance is used for measuring similarity between two temporal sequences, which may vary in speed. The DTW distance is computed after realigning (warping) two time series of equal lengths. The marker signatures with a given window size are extracted and DTW distance matrix is calculated. The DTW distance matrix contains the DTW distances between all pairs of extracted signatures. The matrix size is $[N_m, N_m]$, where N_m is the number of signatures for particular markers in the dataset. This matrix is used to perform clustering.

5.1 Clustering marker signatures

The Hierarchical Density-Based Spatial Clustering of Applications with Noise (HDBSCAN) [32] was used for clustering the signatures [27]. It clusters based on the density distribution of data points, identifying high-density regions as part of the cluster and others as noise, thus creating a hierarchical tree of clusters for different levels of granularity.There is no requirement to predefined the number of clusters to be created and are robust to noise. The DTW distance matrix is used for clustering. The outcome is regarded as favorable when the patterns of the signature in the clusters are similar, and are of substantial size.

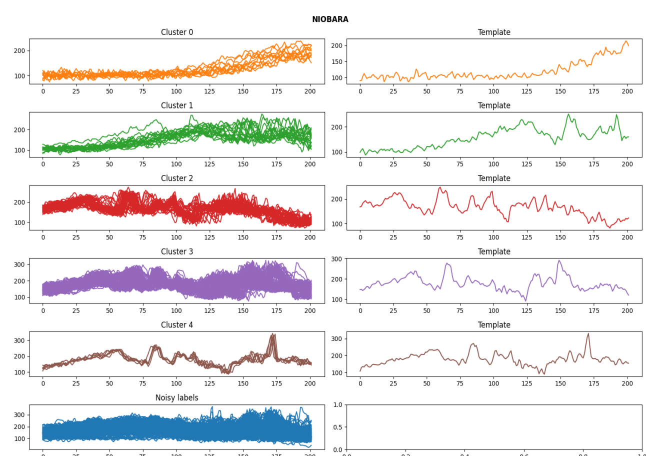

Figure 4: Results of the clustering algorithms on Niobrara signatures

For HDBSCAN, the most impactful parameters are maximum and minimum cluster size, and minimum samples. We set a small minimum sample of 5, a larger minimum cluster size of 15, and the maximum cluster size to one-fourth of the total sample size to capture various patterns and maintain large clusters. The results of

HDBSCAN on Niobrara marker signatures are shown in Figure 4. The template of each cluster is the barycenter of the samples in the cluster.

6 Classification operation

This section outlines the DTW baselin, the new hybrid DL architectures and the classification section of the framework. It discusses data preparation, DL model training mechanisms for optimal convergence in GeoTS, and post-processing to obtain predicted marker depths. Our objective is to determine the depth at which the formation starts. Marker signatures are used to identify this depth as discussed in Section 2.1.

6.1 Baseline

A marker signature template needs to be selected for the DTW baseline approach. Then, compute the DTW distance between this signature template and all the GR signals extracted from the well using a sliding window approach. The template of the largest cluster obtained after clustering 5 is selected as the marker signature template. The smallest DTW distance refers to the depth of closest similarity between the signature template and the observed GR well log. This depth is returned as the predicted marker depth. The time required for the Industrial baseline for multiple time sequences has a complexity of $O(n \cdot m \cdot s \cdot W)$ where n is the number of wells, m is the number of markers,s is the window size, and W is the average well length. Thus, we can propagate only one marker signature at a time. Figure 5 shows the DTW process where we see the signature template on the left side and the query sequence in which we need to find the best match on top. The similarity matrix indicates the region of minimum distance. Figure 6 shows the result of propagating all three marker templates on Well A. The dark blue indicates a low DTW distance, and the yellow indicates a high one. The disadvantage of DTW is that the method is based on pure statistics. It is heavily dependent on the signature template, making it rigid. This issue is overcome by using the DL models, which allows us to include all variations of marker signatures to learn the template pattern.

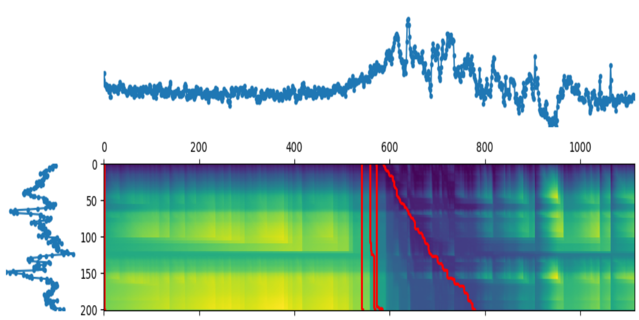

Figure 5: DTW match process for NIOBRARA

6.2 Designed Hybrid models

It appears that hybrid models such as LSTM-FCN and XCM, which incorporate parallel networks of CNNs and RNNs, tend to exhibit improved performance compared to these models used individually or structured within a series network architecture. Taking inspiration from them, three new architectures were built using the LSTM, FCN, and parts of the XCM model as building blocks. These models are explained in this section.

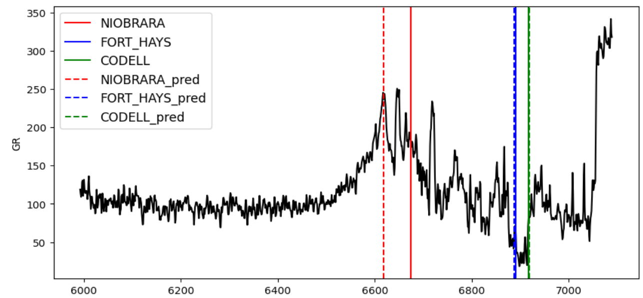

Figure 6: DTW result for propagation

LSTM-XCM is the augmentation of LSTM with XCM submodules. We combine an LSTM parallel network with the existing 1D and 2D parallel networks of the XCM. The results are concatenated and the second part follows the XCM architecture.

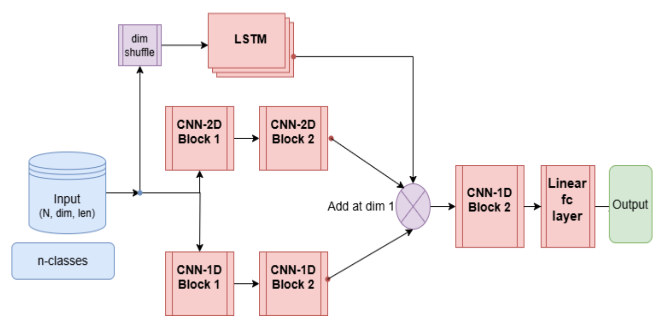

Figure 7: LSTM-XCN model architecture

LSTM-FCN-2dCNN consists of a parallel network of LSTM, FCN, and 2DCNN layers. The output of the 2DCNN is reshaped to allow for the concatenation. In this case, after the concatenation of the outputs of the parallel network, the result is passed through the dropout and linear layer.

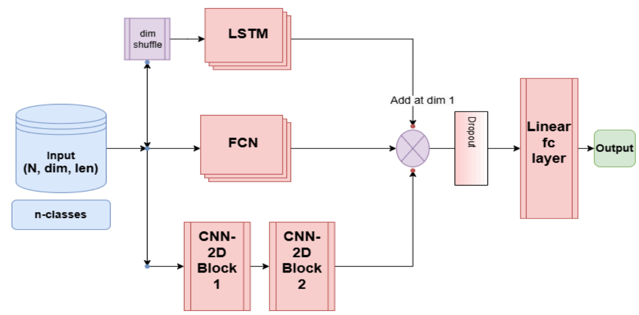

Figure 8: LSTM-FCN-2DCNN model architecture

LSTM-2dCNN does not contain the FCN part of the LSTM-FCN-2DCNN model. This experiment was performed to understand the importance of the FCN part.

6.3 Data preparation for DL model

The results of the clustering operation helps create a set of wells with valid marker signatures. We consider the largest cluster to be the one that represents, at best, the marker signature. Cluster having similar template to the largest one are also considered as part of the learning dataset. From a dataset of 650 wells, 400 wells were filtered and used for training the DL model. The well-log marker signatures with their associated depths, latitude, and longitude are extracted. This is the input to the classification model. The classification model is trained to categorize sequences into four classes: 'None', 'Niobrara', 'Codell', and 'Fort Hays'. 'None' indicates depths without markers. As seen in the well logs in Section 2.1, most logs belong to the 'None' class, creating a class imbalance problem. To address the problem, a subset of total samples is selected to create the final 'None' category.

6.4 DL model training

Hyperparameter tuning enables DL models to achieve better convergence. Selecting a correct learning rate (Lr) influences whether the model's optimization will reach global minima and the time required for training the model. The Lr range test [46] is used to select Lr. Two training policies have been introduced to minimize fluctuations during regularization. The Lr is reduced if the validation loss increases over three epochs, enhancing the stability of the convergence. The second one stops the training if there is no significant improvement in validation loss for eight epochs. These policies reduce training time and help avoid overfitting or underfitting. This method brings more stability for the validation curve. The hyperparameter tuning is performed using the Tsai [37] and fastai [20] libraries. Cross-Entropy loss is applied for backpropagation using an Adam optimizer [25].

6.5 Marker Prediction

The trained model is applied on the complete well-log data using sliding window to generate probability curves to predict marker depths. A moving average filter is applied to smooth the probability curve. The predicted marker top depth is the one with the highest probability. Figure 9 illustrates the probability curve for well A and the extracted marker depth.

The time complexity of using the classification models after training is $O(n \cdot s \cdot W)$, where n is the number of wells, s is the sliding window length, and W is the average well length. This complexity is independent of the number of markers, as probability curves for all markers are generated simultaneously. Additionally, the approach does not depend on a single template like in Industrial baseline, enhancing its robustness.

7 Experiments and Results

This section presents the results of various classification models in GeoTS and discusses the experiments that identify key factors influencing these results. It also addresses the impact of different window sizes. The evaluation metrics from Section 2.2 were applied

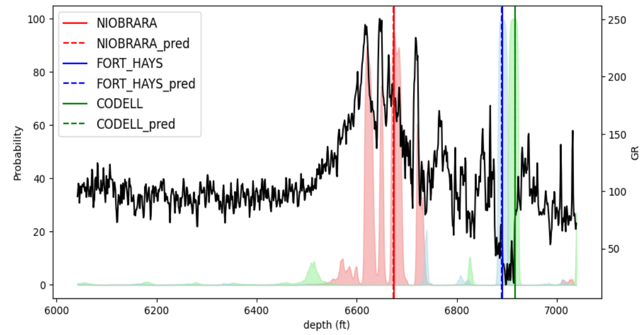

Figure 9: Propagation results for LSTM-XCM

for model comparisons. The experiments were carried out over ten runs to account for randomness in the data selection and model weights initialization. The training dataset consists of 3,500 signal samples across four classes. When testing for prediction, the sample size increases significantly; for a well log of 7,000 ft, there are 7,000 test samples, leading to about one million samples across 150 wells.

When applying the classification models to predict the marker depth over the whole well, the recall varies over different models. Figure 10 shows the recall with a threshold of 10ft for the DL models and DTW baseline. Models like LSTM, FCN, and Resnet have a low recall, with FCN performing the worst, followed by Resnet. LSTM-2dCNN and LSTM-XCM emerged as the best performers, exhibiting a high accuracy of around 95% and low variation, indicating robustness. The DTW baseline averages around 80% similar to that of the recall of inception model.

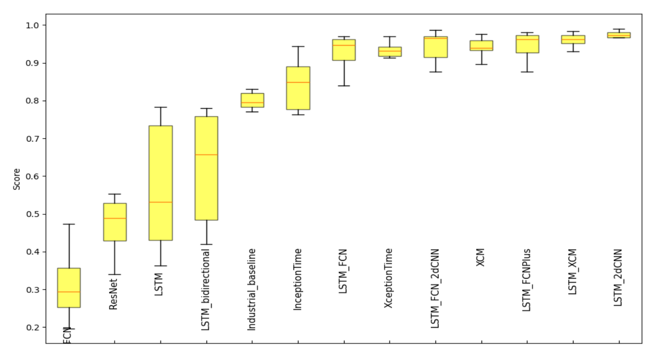

Figure 10: Recall score

The test has been performed to check for model convergence by calculating the accuracy for a balanced testing dataset as discussed in Section 6.3. We get an accuracy greater than 90% for all the deep learning models in such a scenario. We conclude that since we decimate the 'None' class, some models have difficulty classifying the 'None'.

Figure 11 compares training and prediction times, demonstrating the advantages of the proposed GeoTS framework, as it reduces the propagation time by more than 20 compared to the DTW approach when used to detect three markers. The process is also faster due to the use of GPU's for the deep learning models. The experiments

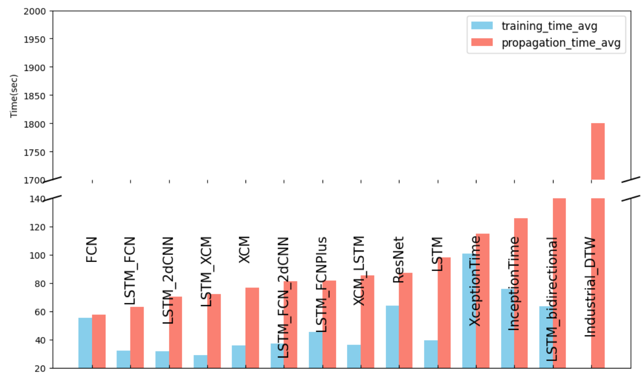

Figure 11: Time consumed in training and propagation

run on NVIDIA RTX 2000 Ada GPU. The time gap will increase as the number of markers to be detected increases.

7.1 Detailed model performance analysis

To understand the prediction behavior of the models, we pick three models, one with low recall, one with average, and one with high recall: LSTM-bidirectional, LSTM-FCN, LSTM-XCM. The results for LSTM-bidirectional, LSTM-FCN, and LSTM-XCM on well B are presented in Figure 12. The probability curves computed are steep and less dispersed for models with higher recall. The difference in accuracy and recall helps us understand that the results for a DL model on the real-world dataset can be quite different than when checked over a testing subset. This is particularly true when the data are noisy and can have minor variations over time or space. The LSTM-bidirectional model mispredicts the depths of the Niobrara, Codell, and Fort Hays markers, while the LSTM-FCN only mispredicts Codell. This does not imply entirely incorrect predictions, as actual depths also show high probability values. However, the result of the LSTM-XCM model show high probability only close to actual depths, leading to correct predictions.

The MAE is analyzed to compare model performance. LSTM-2dCNN, LSTM-XCM, and Xception have a low MAE as shown in Figure 13.

7.2 Sliding Window size

The window size used to extract the signature should effectively represent the geological event of the marker; thus, it is an important parameter. A large window size can lead to clipped results at the well-log's start and end, which is especially problematic when markers are at these locations. Padding can mitigate this issue. Conversely a short window size may not capture the geological event.

Table 1 illustrates the impact of various window sizes on the recall. The green cells highlight the optimal size. A window of 201ft has yielded good recall in most cases, and we used this size for all experiments. Smaller or larger window sizes tend to decrease final recall. For the deployment it is set as a changeable variable, but for the Wyoming results shown, we have used the same window size.

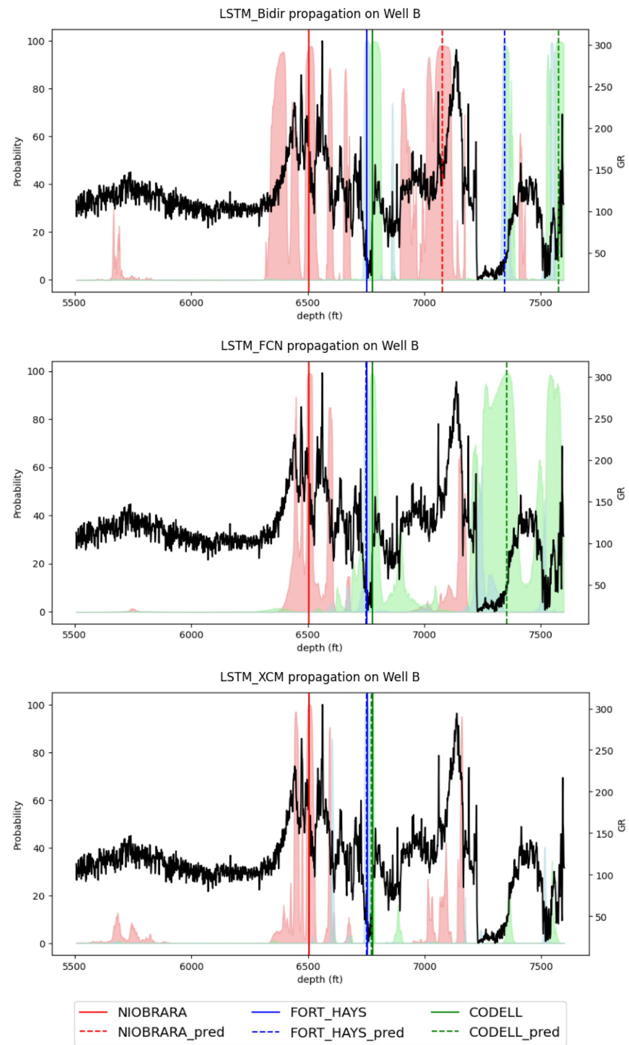

Figure 12: Propagation result on Well B

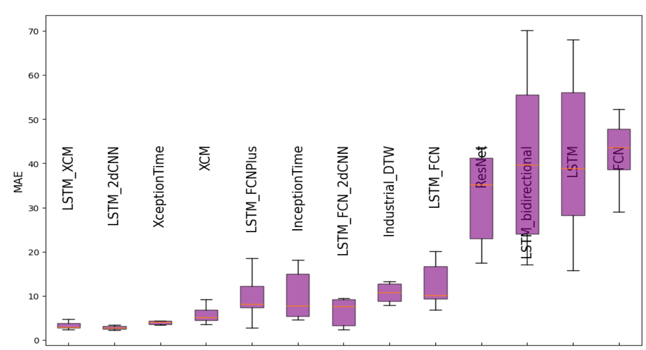

Figure 13: Mean Absolute Error

Table 1: Results for different methods and window sizes (WS)

Methods	WS_201	WS_301	WS_101	WS_51
LSTM_FCN	0.97	0.92	0.84	0.84
LSTM_XCM	0.97	0.95	0.73	0.65
LSTM_2dCNN	0.97	0.98	0.93	0.66
LSTM_FCNPlus	0.96	0.93	0.65	0.66
InceptionTime	0.95	0.84	0.91	0.66
XCM	0.93	0.79	0.95	0.66
LSTM_FCN_2dCNN	0.94	0.96	0.76	0.62
XceptionTime	0.90	0.92	0.95	0.64
LSTM_bidirectional	0.76	0.83	0.72	0.59
LSTM	0.56	0.68	0.68	0.64
ResNet	0.53	0.66	0.39	0.76
FCN	0.27	0.41	0.79	0.85

7.3 Model Interpretability

We implement GradCAM [43], which can be applied to all models in the benchmark. GradCAM helps us with model interpretability and understand the model decision making process at a deeper level. It is crucial in complex architecture using a combination of different models, to gain insight into the success and failure of each part. Unreasonable predictions could have a reasonable explanation or vice versa. It allows us to tune models to make them more robust. It enables us to identify the distribution of feature importance in relation to the different classes. We can determine if the model has been able to pick the correct features to distinguish the classes.

We have adapted the gradCAM function for multivariate time series and propose a novel visualization. We implemented three variations, one for each RNN, 1dCNN, and 2dCNN layer, as they are standard building blocks for time series classification models. This would allow us to implement gradCAM on most time series models. To explain the functionality and effectiveness of the gradCAM function, we have performed a detailed analysis of the LSTM-2dCNN model.

Figure 14a represents the results of grad-CAM for the 2dCNN part of the network. Figure 14a has two subplots, which are heatmap, with a darker shade of red indicating important regions for decision making. The left signal represents a marker class, with higher importance on GR values, while the right signal represents a 'None' class, with more weight on depth, latitude, and longitude values. Figure 14b represents the results of grad-CAM for the LSTM part of the network. The input to the LSTM is a multivariate time series with a single time step. Thus, we aggregated the results on the input variables of GR, depth, latitude, and longitude. The first subplot illustrates the signal categories and the second the gradcam results. The purple represents the 'None' class and blue, green and yellow represent 'Niobrara', 'Codell', and 'Fort Hays' respectively. It can be observed that the GR has a lower dependency for the 'None' class than for the marker classes. The marker 'Codell' is the least represented and does not occur at all locations, thus it having a higher dependence on latitude and longitude means it takes the neighboring wells in the training set as a reference to propagate this formation top.

The grad-CAM method helps us understand each parallel network's feature importance in such complicated hybrid models. It can be concluded that the parallel sub-model specializes in different input features with respect to the output classes, allowing the hybrid model to outperform the sub-models trained individually.

8 Deployment

We have conducted deployment of GeoTS on the Wyoming oil fields consisting of 5600 wells. Clustering according to the well location is performed using OPTICS [4]. For each geo-cluster, we select the

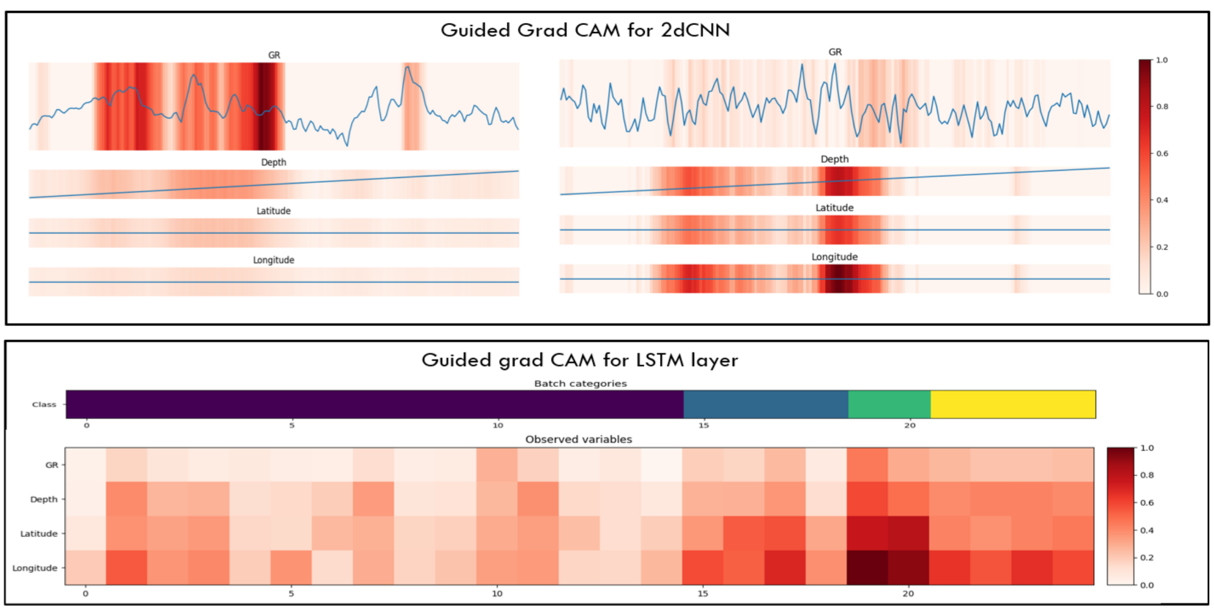

Figure 14: Grad Cam implementation for the LSTM-2dCNN model

most represented formations to be used for marker propagation using GeoTS. The process is conducted using a pre-existing Kubeflow pipeline. The GeoTS pipeline, newly integrated into Kubeflow, is performed for wells within a region. The steps followed are as stated in the flowchart shown in Figure 3. As long as we have a set of wells within a region with labeled and unlabeled formation data, we can apply the GeoTS pipeline to find marker depths for the wells and use it to create reservoir models.

To test the propagation results, we have selected a few sequences for which geologists performed manual verification of the data. Table 2 presents the recall for the propagation of four formation sequences from the four different geo-clusters previously obtained. The sequences are Bg-W(Big George and Werner formations), Bgc-Bc-Fc(Big George coal, Badger coal and Felix coal formations), Fh-La-Al (Fox Hill, Lance and Almond formations), and Ls-Lm-Ll (Lance shale, Lance middle and Lance lower formations). We used a 200 ft window size for the processing.

Table 2: Deployment results for Wyoming (Recall)

Sequence	DTW	LSTM-FCN	LSTM-XCM	LSTM-2dCNN
Bg-W	0.21	0.71	0.73	0.74
Bgc-Bc-Fc	0.31	0.85	0.84	0.85
Fh-La-Al	0.25	0.45	0.55	0.48
LS-LM-LL	0.33	0.46	0.48	0.47

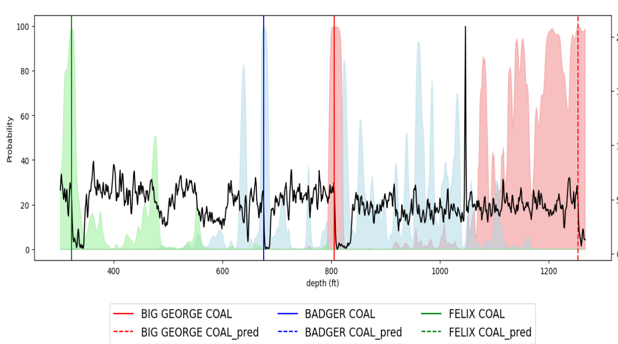

Figure 15: Wyoming propagation results: Bgc-Bc-Fc

We can clearly see that LSTM-XCM and LSTM-2dCNN perform better than the DTW algorithm. However, it can also be observed that the recall for the Fh-La-Al and Ls-Lm-Ll sequences is lower overall and closer to the recall of DTW. We have realized that the signatures of these markers do not follow a distinctive pattern in the GR well log. With no pattern being followed, even manually, we are unable to position the marker correctly using just the GR well log. This indicates the need for additional logs such as density, resistivity, and sonics, but these logs were not available in most of the use cases. Some of the formations may have also been identified using seismic data. Another observation is that these markers have a larger depth distribution of occurrence of three to five thousand feet, which makes the depth information redundant. As for the formation sequences of Bg-W and Bgc-Bc-Fc, we have been able to increase recall by more than 50%. The incorrect predictions were cases in which the GR log at the depth being similar to the marker signature. Figure 15 represents the case in which Big George coal has been inaccurately picked due to a similar pattern present at incorrect depth. With this study, we can conclude that we need to consider not only the local features of the marker signature but also the global features of the well log between the markers.

9 Conclusion and future work

In this paper, we analyze well log data and we demonstrate the deep learning model's efficiency, ease, and accuracy in pattern identification task. We present GeoTS a framework which implements the workflow for the complete process: the preprocessing data, the model implementation, and the post processing. The implemented grad-CAM can be applied to many deep-learning models. It will help understand the logic behind the decision-making of neural networks better for time series. The benchmark evaluation section clearly shows the grad-CAM's ability to recognize the input feature's importance.

GeoTS framework can be applied across diverse well log datasets obtained from different companies without following a standard uniformity. Data cleaning ensures uniformity across well logs for the study. The primary feature in the paper is set to GR as it is the most available well-log. Additional well-log information, such as resistivity, density, and compressional slowness, if available, will be automatically integrated during the data preparation process. We expect the model performance to improve with additional logs.

The main advantage of using GeoTS is that it is a completely autonomous process. We achieve an increase in recall and faster computational time when compared with the DTW baseline. Most of the existing frameworks used for marker propagation in wells are dependent on the neighboring wells log. As the model is trained on hundreds of wells, it is not dependent on a particular well, thus allowing the results to be resilient to the noise in the well logs. Our new models the LSTM-2dCNN and the LSTM-XCM outperform the state of art and baseline models.

The approach has been validated with the Colorado and Wyoming datasets, and additional tests for new datasets are scheduled by the company. To solve the issues we encountered during the deployment for Wyoming for the future, we will be trying sequence to sequence transformer models. The whole well-log can be given as an input, allowing the model to understand the relative depth dependencies between the markers.

Acknowledgments

We thank the Colorado Energy and Carbon Management Commission and WOGCC for publicly providing well log datasets for informational and research purposes.

References

[1] [n. d.]. Colorado Energy and Carbon Management Comission. https://ecmc.state.co.us/data.html#/cogis.
[2] Mustafa A. Al Ibrahim. 2022. Uncertainty in Automated Well-Log Correlation Using Stochastic Dynamic Time Warping. *Petrophysics - The SPWLA Journal* 63, 06 (12 2022), 748–761. https://doi.org/10.30632/PJV63N6-2022a11 arXiv:https://onepetro.org/petrophysics/article-pdf/63/06/748/3056402/spwla-2022-v63n6a11.pdf
[3] Mark A. Andersen. [n. d.]. Defining log interpretation. https://www.slb.com/resource-library/oilfield-review/defining-series/defining-log-interpretation

[4] Mihael Ankerst, Markus M. Breunig, Hans-Peter Kriegel, and Jörg Sander. 1999. OPTICS: ordering points to identify the clustering structure *(SIGMOD '99)*. Association for Computing Machinery, New York, NY, USA, 49–60. https://doi.org/10.1145/304182.304187

[5] Mohammadreza Ebadi Asieh Zare. Majid Bagheri. 2020. Reservoir facies and porosity modeling using seismic data and well logs by geostatistical simulation in an oil field. *Carbonates and Evaporites* 35, 65 (Oct. 2020), 336–359. https://doi.org/10.1007/s13146-020-00605-5

[6] Robert U. Ayres, Jeroen C.J.M. van den Bergh, Dietmar Lindenberger, and Benjamin Warr. 2013. The underestimated contribution of energy to economic growth. *Structural Change and Economic Dynamics* 27 (2013), 79–88. https://doi.org/10.1016/j.strueco.2013.07.004

[7] Xiao Bian, Ser Nam Lim, and Ning Zhou. 2016. Multiscale fully convolutional network with application to industrial inspection. In *2016 IEEE Winter Conference on Applications of Computer Vision (WACV)*. 1–8. https://doi.org/10.1109/WACV.2016.7477595

[8] Richard L Caldwell, Willett F Baldwin, James D Bargainer, James E Berry, George N Salaita, and Raymond W Sloan. 1963. Gamma-ray spectroscopy in well logging. *Geophysics* 28, 4 (1963), 617–632.

[9] Apoorva Chawla, Yessica Arellano, Martin Viktor Johansson, Hossein Darvishi, Khadija Shaneen, Matteo Vitali, Francesco Finotti, and Pierluigi Salvo Rossi. 2022. IoT-Based Monitoring in Carbon Capture and Storage Systems. *IEEE Internet of Things Magazine* 5, 4 (2022), 106–111. https://doi.org/10.1109/IOTM.001.2200175

[10] Chinedu Pascal Ezenkwu, John Guntoro, Andrew Starkey, Vahid Vaziri, and Maurillio Addario. 2023. Automated Well-Log Pattern Alignment and Depth-Matching Techniques: An Empirical Review and Recommendations. *Petrophysics - The SPWLA Journal* 64, 01 (02 2023), 115–129. https://doi.org/10.30632/PJV64N1-2023a9 arXiv:https://onepetro.org/petrophysics/article-pdf/64/01/115/3065797/spwla-2023-v64n1a9.pdf

[11] Kevin Fauvel, Tao Lin, Véronique Masson, Élisa Fromont, and Alexandre Termier. 2021. XCM: An Explainable Convolutional Neural Network for Multivariate Time Series Classification. *Mathematics* 9, 23 (Dec. 2021), 3137. https://doi.org/10.3390/math9233137

[12] British geological survey. [n. d.]. Understanding carbon capture and storage. https://www.bgs.ac.uk/discovering-geology/climate-change/carbon-capture-and-storage/

[13] Jon Gibbins and Hannah Chalmers. 2008. Carbon capture and storage. *Energy Policy* 36, 12 (2008), 4317–4322. https://doi.org/10.1016/j.enpol.2008.09.058 Foresight Sustainable Energy Management and the Built Environment Project.

[14] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. 2015. Deep Residual Learning for Image Recognition. *CoRR* abs/1512.03385 (2015). arXiv:1512.03385 http://arxiv.org/abs/1512.03385

[15] J.R. Hearst and P.H. Nelson. 2025. *Well logging for physical properties*. Mc Graw-Hill Book Co.

[16] Roberto Henry Herrera and Mirko van der Baan. 2012. Automated seismic-to-well ties? arXiv:1209.0201 [physics.geo-ph]

[17] RC Hertzog and RE Plasek. 1979. Neutron-excited gamma-ray spectrometry for well logging. *IEEE Transactions on Nuclear Science* 26, 1 (1979), 1558–1567.

[18] Sepp Hochreiter and Jürgen Schmidhuber. 1997. Long Short-Term Memory. *Neural Computation* 9, 8 (1997), 1735–1780. https://doi.org/10.1162/neco.1997.9.8.1735

[19] Wan Yun Hong. 2022. A techno-economic review on carbon capture, utilisation and storage systems for achieving a net-zero CO2 emissions future. *Carbon Capture Science & Technology* 3 (2022), 100044. https://doi.org/10.1016/j.ccst.2022.100044

[20] Jeremy Howard et al. 2018. fastai. https://github.com/fastai/fastai.

[21] Hassan Ismail Fawaz, Germain Forestier, Jonathan Weber, Lhassane Idoumghar, and Pierre-Alain Muller. 2019. Deep learning for time series classification: a review. *Data Mining and Knowledge Discovery* 33, 4 (March 2019), 917–963. https://doi.org/10.1007/s10618-019-00619-1

[22] Hassan Ismail Fawaz, Benjamin Lucas, Germain Forestier, Charlotte Pelletier, Daniel F. Schmidt, Jonathan Weber, Geoffrey I. Webb, Lhassane Idoumghar, Pierre-Alain Muller, and François Petitjean. 2020. InceptionTime: Finding AlexNet for time series classification. *Data Mining and Knowledge Discovery* 34, 6 (Sept. 2020), 1936–1962. https://doi.org/10.1007/s10618-020-00710-y

[23] Saidi Kais and Hammami Sami. 2016. An econometric study of the impact of economic growth and energy use on carbon emissions: Panel data evidence from fifty eight countries. *Renewable and Sustainable Energy Reviews* 59 (2016), 1101–1110. https://doi.org/10.1016/j.rser.2016.01.054

[24] Fazle Karim, Somshubra Majumdar, Houshang Darabi, and Shun Chen. 2018. LSTM Fully Convolutional Networks for Time Series Classification. *IEEE Access* 6 (2018), 1662–1669. https://doi.org/10.1109/access.2017.2779939

[25] Diederik P. Kingma and Jimmy Ba. 2017. Adam: A Method for Stochastic Optimization. arXiv:1412.6980 [cs.LG] https://arxiv.org/abs/1412.6980

[26] Jin Lai, Yang Su, Lu Xiao, Fei Zhao, Tianyu Bai, Yuhang Li, Hongbin Li, Yuyue Huang, Guiwen Wang, and Ziqiang Qin. 2024. Application of geophysical well logs in solving geologic issues: Past, present and future prospect. *Geoscience Frontiers* 15, 3 (2024), 101779. https://doi.org/10.1016/j.gsf.2024.101779

[27] T Warren Liao. 2005. Clustering of time series data—a survey. *Pattern recognition* 38, 11 (2005), 1857–1874.

[28] Peng Lin, Ji-gen Xia, Qiu-yuan Hou, Yong-li Ji, and Chen Li. 2024. An Intelligent Depth Correction Method for Logging Curves Based on Pearson Correlation Coefficient and DTW. In *Proceedings of the International Field Exploration and Development Conference 2023*, Jia'en Lin (Ed.). Springer Nature Singapore, Singapore, 102–114.

[29] Rebecca Lindsey. [n. d.]. Climate Change: Atomospheric carbon dioxide. https://www.climate.gov/news-features/understanding-climate/climate-change-atmospheric-carbon-dioxide

[30] Jiang Liu, Hua Wu, Xingguang Xu, Bo Wang, Yunkun Wu, and Jun Gu. 2025. Fabrication of CO2-Saturated Nanofluids with Optimized Interfacial Properties for Carbon Capture, Utilization, and Storage. *SPE Journal* (01 2025), 1–17. https://doi.org/10.2118/224429-PA arXiv:https://onepetro.org/SJ/article-pdf/doi/10.2118/224429-PA/4444114/spe-224429-pa.pdf

[31] Jonathan Long, Evan Shelhamer, and Trevor Darrell. 2015. Fully convolutional networks for semantic segmentation. In *2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. 3431–3440. https://doi.org/10.1109/CVPR.2015.7298965

[32] Claudia Malzer and Marcus Baum. 2020. A Hybrid Approach To Hierarchical Density-based Cluster Selection. In *2020 IEEE International Conference on Multisensor Fusion and Integration for Intelligent Systems (MFI)*. IEEE. https://doi.org/10.1109/mfi49285.2020.9235263

[33] J.L. Mari, P. Gaudiani, and J. Delay. 2011. Characterization of geological formations by physical parameters obtained through full waveform acoustic logging. *Physics and Chemistry of the Earth, Parts A/B/C* 36, 17 (2011), 1438–1449. https://doi.org/10.1016/j.pce.2011.07.011 Clays in Natural & Engineered Barriers for Radioactive Waste Confinement.

[34] Ajan Meenakshisundaram, Olusegun Stanley Tomomewo, Laalam Aimen, and Shree Om Bade. 2024. A comprehensive analysis of repurposing abandoned oil wells for different energy uses: Exploration, applications, and repurposing challenges. *Cleaner Engineering and Technology* 22 (2024), 100797. https://doi.org/10.1016/j.clet.2024.100797

[35] Meinard Müller. 2007. *Dynamic Time Warping*. Springer Berlin Heidelberg, Berlin, Heidelberg, 69–84. https://doi.org/10.1007/978-3-540-74048-3_4

[36] H.D. Nguyen, K.P. Tran, S. Thomassey, and M. Hamad. 2021. Forecasting and Anomaly Detection approaches using LSTM and LSTM Autoencoder techniques with the applications in supply chain management. *International Journal of Information Management* 57 (2021), 102282. https://doi.org/10.1016/j.ijinfomgt.2020.102282

[37] Ignacio Oguiza. 2023. tsai - A state-of-the-art deep learning library for time series and sequential data. Github. https://github.com/timeseriesAI/tsai

[38] Patxi Ortego, Alberto Diez-Olivan, Javier Del Ser, Fernando Veiga, Mariluz Penalva, and Basilio Sierra. 2020. Evolutionary LSTM-FCN networks for pattern classification in industrial processes. *Swarm and Evolutionary Computation* 54 (2020), 100650.

[39] Peter.D.Warwick. [n. d.]. Using Petroleum Reservoirs to Store Carbon. https://www.usgs.gov/, accessed2024-06-05.

[40] Shriram Tallam Puranam Raghu, Dawn T. MacIsaac, and Erik J. Scheme. 2025. Self-supervised learning via VICReg enables training of EMG pattern recognition using continuous data with unclear labels. *Computers in Biology and Medicine* 185 (2025), 109479. https://doi.org/10.1016/j.compbiomed.2024.109479

[41] Elahe Rahimian, Soheil Zabihi, Seyed Farokh Atashzar, Amir Asif, and Arash Mohammadi. 2020. XceptionTime: Independent Time-Window Xceptiontime Architecture for Hand Gesture Classification. In *ICASSP 2020 - 2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. 1304–1308. https://doi.org/10.1109/ICASSP40776.2020.9054586

[42] Hansheng Ren, Bixiong Xu, Yujing Wang, Chao Yi, Congrui Huang, Xiaoyu Kou, Tony Xing, Mao Yang, Jie Tong, and Qi Zhang. 2019. Time-series anomaly detection service at microsoft. In *Proceedings of the 25th ACM SIGKDD international conference on knowledge discovery & data mining*. 3009–3017.

[43] Ramprasaath R. Selvaraju, Michael Cogswell, Abhishek Das, Ramakrishna Vedantam, Devi Parikh, and Dhruv Batra. 2019. Grad-CAM: Visual Explanations from Deep Networks via Gradient-Based Localization. *International Journal of Computer Vision* 128, 2 (Oct. 2019), 336–359. https://doi.org/10.1007/s11263-019-01228-7

[44] A. Shakirov, A. Molchanov, L. Ismailova, and M. Mezghani. 2023. Quantitative assessment of rock lithology from gamma-ray and mud logging data. *Geoenergy Science and Engineering* 225 (2023), 211664. https://doi.org/10.1016/j.geoen.2023.211664

[45] Susan Sunila Sharma. 2010. The relationship between energy and economic growth: Empirical evidence from 66 countries. *Applied Energy* 87, 11 (2010), 3565–3574. https://doi.org/10.1016/j.apenergy.2010.06.015

[46] Leslie N. Smith. 2018. A disciplined approach to neural network hyper-parameters: Part 1 – learning rate, batch size, momentum, and weight decay. arXiv:1803.09820 [cs.LG] https://arxiv.org/abs/1803.09820

[47] Stephan Spiegel, Julia Gaebler, Andreas Lommatzsch, Ernesto De Luca, and Sahin Albayrak. 2011. Pattern recognition and classification for multivariate time series.

In *Proceedings of the Fifth International Workshop on Knowledge Discovery from Sensor Data* (San Diego, California) *(SensorKDD '11)*. Association for Computing Machinery, New York, NY, USA, 34–42. https://doi.org/10.1145/2003653.2003657

[48] Christian Szegedy, Wei Liu, Yangqing Jia, Pierre Sermanet, Scott Reed, Dragomir Anguelov, Dumitru Erhan, Vincent Vanhoucke, and Andrew Rabinovich. 2015. Going deeper with convolutions. In *2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. 1–9. https://doi.org/10.1109/CVPR.2015.7298594

[49] Qingsong Wen, Linxiao Yang, Tian Zhou, and Liang Sun. 2022. Robust time series analysis and applications: An industrial perspective. In *Proceedings of the 28th ACM SIGKDD Conference on Knowledge Discovery and Data Mining*. 4836–4837.

[50] Tian Zhang, Wanchang Zhang, Ruizhao Yang, Yimo Liu, and Masoud Jafari. 2021. CO2 capture and storage monitoring based on remote sensing techniques: A review. *Journal of Cleaner Production* 281 (2021), 124409. https://doi.org/10.1016/j.jclepro.2020.124409

[51] Ye Zhang, Yi Hou, Kewei OuYang, and Shilin Zhou. 2022. Multi-scale signed recurrence plot based time series classification using inception architectural networks. *Pattern Recognition* 123 (2022), 108385.

[52] Aman Zhumekeshov, Mukhtar Kunyrbayev, Ilyas Turgazinov, Daniyar Nassipov, Shukhrat Mametov, and Kassymzhomart Ulasbek. 2023. Automating Conventional Well Correlation Using Machine Learning Techniques. https://doi.org/10.2139/ssrn.4588834