

Combinatorial feature engineering and deterministic rock typing for seismic facies prediction in the mixed clastic-carbonate Asmari formation, SW Iran

Sam Azarpour

sam.azarpour@gmail.com

Rockies Geoservices Ltd

Igor Morozov

University of Saskatchewan

Iman Zahmatkesh

Shahid Chamran University of Ahvaz

Mahmoud Poladzadeh

NIOC

Research Article

Keywords: Seismic facies classification, Reservoir rock typing, Feature engineering, Artificial neural networks, Mixed clastic-carbonate reservoirs, Asmari Formation

Posted Date: October 15th, 2026

DOI: <https://doi.org/10.21203/rs.3.rs-10735799/v1>

License: © ⓘ This work is licensed under a Creative Commons Attribution 4.0 International License.
[Read Full License](#)

Additional Declarations: No competing interests reported.

Combinatorial feature engineering and deterministic rock typing for seismic facies prediction in the mixed clastic-carbonate Asmari formation, SW Iran

Corresponding author: Sam Azarpour (sam.azarpour@gmail.com)

Sam Azarpour ^a, Igor Morozov ^b, Iman Zahmatkesh ^c, Mahmoud Poladzadeh ^d

^a *Rockies Geoservices, Calgary, Alberta, Canada*

^b *University of Saskatchewan, Department of Geological Science, Saskatoon, Canada*

^c *Department of Petroleum Geology and Sedimentary Basins, Faculty of Earth Sciences, Shahid Chamran University of Ahvaz, Ahvaz, Iran*

^d *NIOC, Department of Geophysics, Ahvaz, Iran*

Abstract

Accurate reservoir rock typing in highly heterogeneous formations remains a critical challenge in seismic reservoir characterization. Conventional unsupervised electrofacies classifications often lack petrophysical meaning, while raw seismic attributes often fail to resolve acoustic ambiguities in mixed clastic-carbonate systems. To address this, we propose an integrated workflow combining deterministic rock-physics principles with an ensemble-based feature engineering architecture. First, well-log data from 18 wells in the Asmari Formation were classified into three seismically consistent targets (Pay Sand, Porous Carbonate, and Tight Background) using strict petrophysical cut-offs. This step mitigates elastic non-uniqueness by consolidating acoustically similar non-productive lithologies into a unified background class.

The primary methodological novelty lies in tackling the redundancy and low discriminative power of raw seismic attributes. Rather than conventional attribute selection, we introduce a combinatorial multi-classifier scheme. By deploying four algorithms (Support Vector Machines, Random Forests, Neural Networks, and Logistic Regression) across all binary class splits, the initial 25 raw attributes are transformed into an enriched, discriminative 24-dimensional feature space. Following dimensionality reduction via Principal Component Analysis (PCA), this optimized feature set is fed into an artificial neural network, hereafter the Combinatorial-Feature-Engineering ANN (CFE-ANN). Evaluated on a blind well, the CFE-ANN workflow (76.2% accuracy) outperformed conventional ANN (50.8%) and Probabilistic Neural Network (PNN) (49.2%) baseline workflows, overcoming class-imbalance failures. The

resulting 2D confidence-filtered facies sections deliver a geologically coherent architecture of clastic-carbonate interbedding, demonstrating that transforming the seismic feature space, guided by deterministic petrophysics, enhances prediction reliability in complex reservoirs.

Keywords: Seismic facies classification; Reservoir rock typing; Feature engineering; Artificial neural networks; Mixed clastic-carbonate reservoirs; Asmari Formation

1. Introduction

Although machine learning (ML) algorithms are now widely applied in seismic reservoir characterization, a key problem remains the definition of a target variable that is both petrophysically meaningful and seismically resolvable. Conventional workflows frequently rely on unsupervised statistical clustering algorithms to generate electrofacies. However, these statistically derived classes often lack a direct physical meaning and may separate non-productive rocks that share the same lack of fluid-flow potential. Applying ML algorithms to predict such acoustically ambiguous targets inevitably leads to high uncertainty.

To establish a robust foundation, this study proposes a deterministic, rock-physics-driven framework for litho-reservoir rock typing prior to the application of seismic ML. Instead of utilizing black-box clustering, we define three specific reservoir classes (Pay Sand, Porous Carbonate, and Tight Background) governed by distinct petrophysical cut-offs. This strategy inherently integrates fluid-flow potential with acoustic impedance contrasts, mitigating elastic ambiguity by consolidating compositionally distinct but acoustically similar non-productive lithologies into a unified background class.

However, even with physically constrained target labels, extracting discriminative information from 3D seismic data remains a major challenge. Seismic attributes encapsulate critical lithological and structural information (Farzadi, 2006; Roy et al., 2013; Riedel et al., 2013a, b; Du et al., 2015; Zhao et al., 2015, 2018; Wrona et al., 2018; Zhang et al., 2019; Zahmatkesh and Kadkhodaie, 2025). Yet, integrating multiple attribute volumes produces high-dimensional datasets containing substantial redundancy and noise. Conventional attribute selection methods, such as linear correlation filters or stepwise regression often fail to resolve the complex, non-linear acoustic overlaps inherent to mixed clastic-carbonate systems (Kim et al., 2019; Qi et al., 2020; Babikir et al., 2024). Consequently, raw attribute spaces frequently lack the

discriminative power required to isolate productive sweet spots from heterogeneous tight backgrounds.

To address this limitation, modern data science increasingly relies on feature space transformations and meta-learning architectures (e.g., stacked generalization), though such concepts remain largely underexplored in seismic facies classification (Wolpert, 1992; Anifowose et al., 2017; Sagi and Rokach, 2018). Recognizing this gap, a central element of this study is a combinatorial multi-classifier feature-engineering scheme that underpins the Combinatorial-Feature-Engineering ANN (CFE-ANN). Instead of selecting a subset of the raw transformations, this approach generates an enriched feature space. By deploying diverse base classifiers (including Support Vector Machines, Random Forests, Neural Networks, and Logistic Regression) across all possible binary class splits, we engineer new discriminative features that explicitly capture the non-linear boundaries between specific rock types.

This study aims to predict and characterize the highly heterogeneous rock facies distribution in the Asmari Formation (Mansuri Oil Field) using the CFE-ANN model. The workflow integrates model-based acoustic impedance (AI) inversion, ML-based porosity estimation, and the aforementioned feature engineering followed by Principal Component Analysis (PCA). To evaluate this approach, its performance is compared with a conventional artificial neural network (ANN) and a commercial Probabilistic Neural Network (PNN). In addition, classification confidence is estimated and used to generate confidence-filtered facies volumes. This enables explicit assessment of prediction reliability, which is rarely examined in reservoir characterization. This paper presents an independently validated workflow that bridges the gap between deterministic petrophysics and ensemble feature engineering for complex mixed reservoir systems.

2. Geological Setting

The study area is the Mansuri Oil Field, located in southwest Iran in the northern Dezful Embayment of the Zagros fold belt (Figure 1). The target interval is the Asmari Formation, the principal oil-producing reservoir unit in Iran, consisting of interbedded carbonates, sandstones, and shales of Oligocene to Miocene age; within the Dezful Embayment, hydrocarbons are trapped in asymmetrical anticlines sealed by the evaporites of the overlying Gachsaran

Formation (Zahmatkesh et al., 2017). In the Mansuri field, the Asmari reservoir comprises approximately two-thirds limestones and dolomites and one-third clastic sediments, although production has been, and remains, predominantly from the sandstones. Based on core descriptions and wireline logs, the reservoir is subdivided into eight zones: carbonates dominate zones 1, 6, 7, and 8, and sandstones zones 2, 3, 4, and 5, so that apart from the carbonate-dominated uppermost zone, the upper half of the reservoir is predominantly clastic. This pronounced vertical and lateral interbedding of acoustically contrasting lithologies makes the Asmari of Mansuri a demanding test case for seismic facies classification.

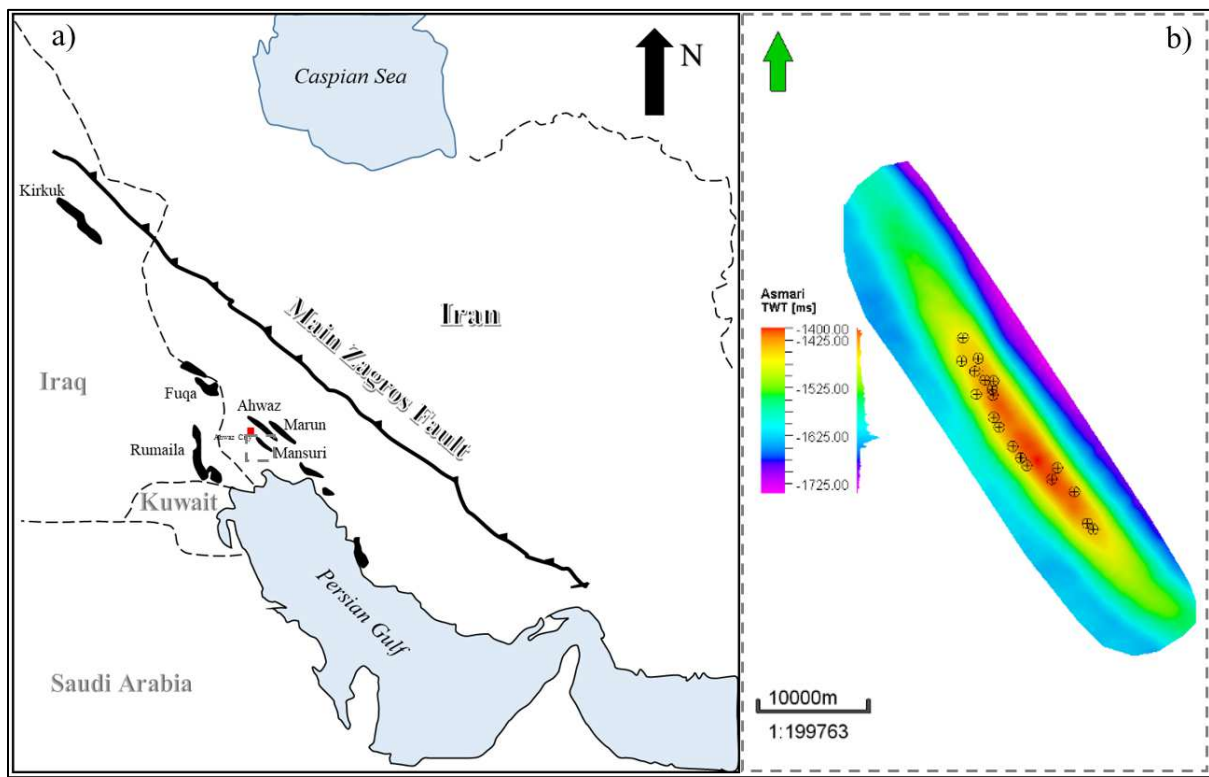


Figure 1: Regional map of the Mansuri oil field and time structure map of Asmari Formation. a) regional location; the Mansuri oil field is outlined by the dashed square. b) time structure map (TWT, ms) of the Asmari Formation (Oligo-Miocene) showing the study area and the location of wells used in this study. (Wells are shown as encircled plus symbols).

3. Methodology

3.1. Deterministic Litho-Reservoir Rock Typing

Prior to classification, a comprehensive quality control algorithm was applied to the well-log data to filter instrumental outliers and standardize volumetric scales. Subsequently, the volumetric fractions of key minerals, namely sand (V_{sand}), shale (V_{shale}), calcite (V_{calc}), dolomite (V_{dol}), and anhydrite (V_{anh}), were integrated with effective porosity (PHIE) to formulate the classification logic. To address the limitations of unsupervised electrofacies classification outlined in the Introduction, we developed a deterministic, rock-physics-driven framework for 1D well-log classification. Instead of utilizing black-box clustering, Reservoir Rock Types (RRTs) are defined based on explicit petrophysical cut-offs and computed lithological volumes.

Considering the pronounced heterogeneity of the Asmari Formation, the dataset was categorized into three distinct Litho-Reservoir classes. The specific cut-offs were designed to isolate reservoir sweet spots while consolidating non-productive intervals to reduce acoustic ambiguity. These three classes are described in Table 1.

Table 1: Definitions and petrophysical cut-offs of the three litho-reservoir rock types.

Class	Description
Class 1: Pay Sand (Reservoir Sand)	This class isolates the highly porous siliciclastic intervals with excellent reservoir quality. Due to the intrinsically lower density of the quartz matrix combined with high porosity, these intervals typically exhibit the lowest acoustic impedance within the formation. The deterministic criteria for this class mandate matrix dominance ($V_{\text{sand}} \geq 0.50$), an absence of pore-throat plugging by clay ($V_{\text{shale}} \leq 0.20$), and commercial fluid-flow potential ($\text{PHIE} \geq 0.12$).
Class 2: Porous Carbonate (Reservoir Carbonate)	This class captures the productive limestone and dolomitic intervals. Because carbonates inherently possess a stiffer rock framework and higher matrix density than unconsolidated sands, their acoustic impedance is generally moderate. Furthermore, due to the presence of secondary porosity systems (e.g., vugs and micro-fractures), a lower porosity threshold is sufficient for commercial production.

Class	Description
	The criteria for this class require a dominant carbonate matrix ($V_{calc} + V_{dol} \geq 0.50$) coupled with an effective porosity threshold of $PHIE \geq 0.08$.
Class 3: Tight / Non-Reservoir Background	All data intervals failing to satisfy the criteria for Class 1 or Class 2 were strictly categorized as the non-reservoir background (Class 3). This class is lithologically heterogeneous, encompassing highly cemented tight sands ($PHIE < 0.12$), tight carbonates ($PHIE < 0.08$), massive anhydrites, and dense shales. Despite their compositional differences, these rocks share a critical elastic signature: high bulk density and high compressional velocity, culminating in high acoustic impedance. Grouping these acoustically similar, non-productive rocks into a single background class prevents the subsequent machine learning model from forcing artificial distinctions between them, thereby maximizing the algorithm's precision in tracking the low-to-moderate impedance targets (Classes 1 and 2).

The deterministic classification in Table 1 is applied across the entire dataset comprising 18 wells, generating a discrete 1D target log for each well (sampled at 0.1524 m) as the foundational input for the subsequent seismic multi-attribute machine learning phase. To visually validate the vertical distribution, petrophysical accuracy, and geological consistency of the defined litho-reservoir rock types (RRTs), two representative wells (MI_004 and MI_015) are selected and presented as detailed case studies. The outputs for these selected wells were plotted using a standard multi-track petrophysical layout, which integrates the cumulative lithological fractions, gamma ray signatures, effective porosity, sand volume, stratigraphic zonation, and the resulting 3-class facies log (Figure 2 and Figure 3).

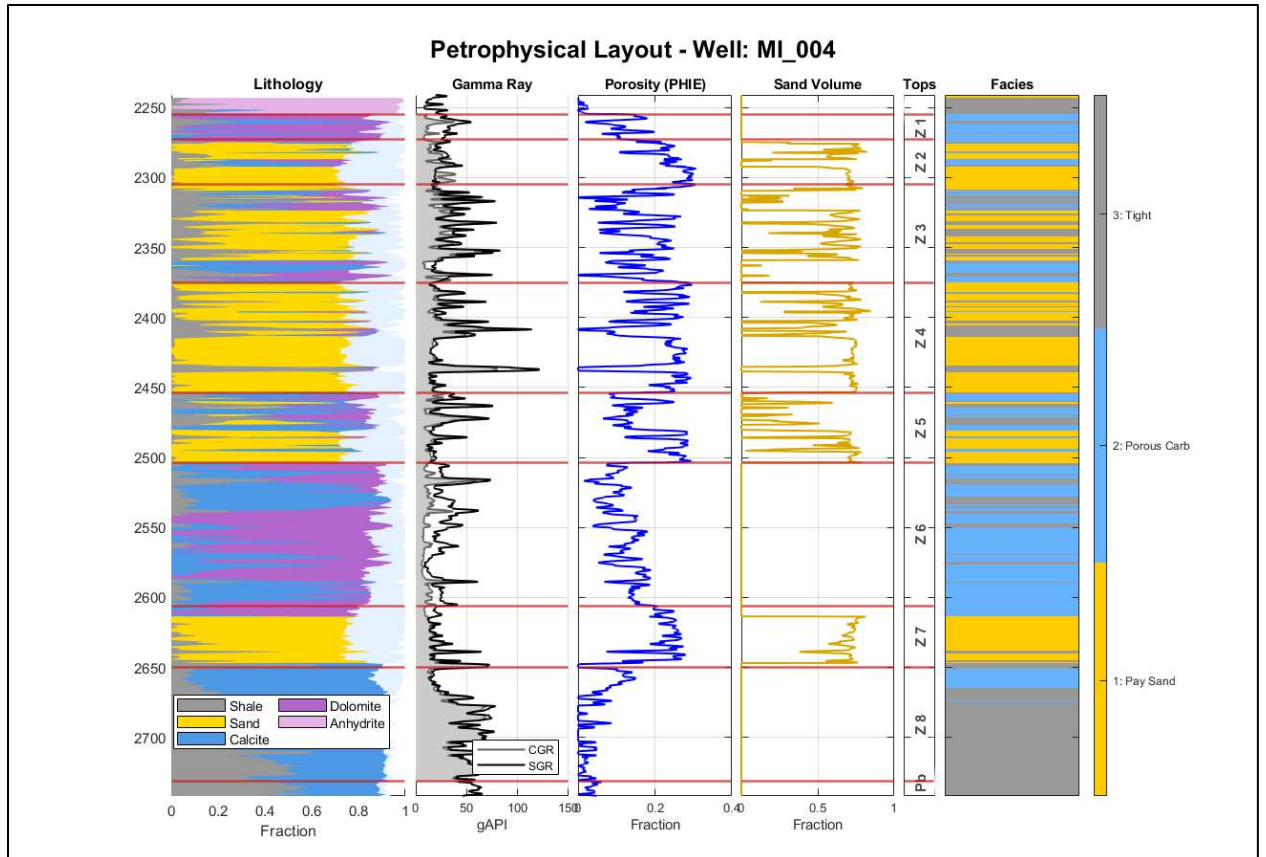


Figure 2: Petrophysical layout of well MI_004 with cumulative lithological fractions, Gamma Ray signatures, effective porosity, sand volume, stratigraphic zonation, and the resulting 3-class facies log.

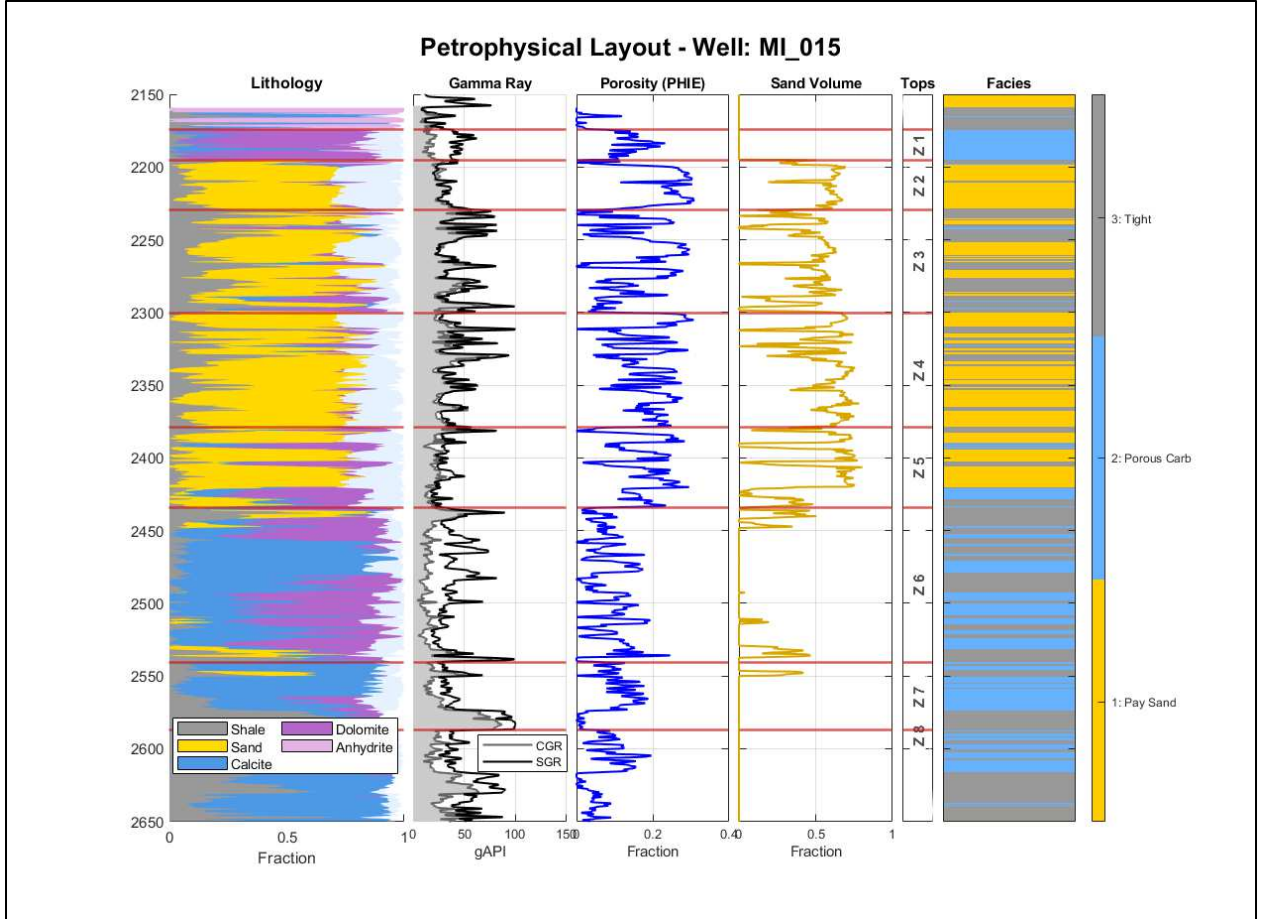


Figure 3: Petrophysical layout of well MI_015 with cumulative lithological fractions, Gamma Ray signatures, effective porosity, sand volume, stratigraphic zonation, and the resulting 3-class facies log.

Analysis of the classification results in well MI_004 reveals a distinct lithological contrast. The upper and middle stratigraphic zones (Z2 through Z5) are heavily dominated by siliciclastic deposits, exhibiting high sand volumes and substantial effective porosity. The implemented deterministic logic accurately captures these intervals, assigning them to Class 1 (Pay Sand, represented in yellow), which confirms their excellent fluid-flow potential. Conversely, the deeper intervals (Z6 to Z8) illustrate a transition into a carbonate-dominated environment characterized by elevated calcite and dolomite fractions. In these lower zones, the algorithm shifts to identify Class 2 (Porous Carbonate, represented in blue) where porosity exceeds the 8% commercial threshold, while correctly segregating the denser, low-porosity carbonate streaks into Class 3 (Tight, represented in gray). The robustness of this classification scheme is further demonstrated in well MI_015, which exhibits pronounced vertical heterogeneity. The upper sections (Z2 to Z5) display rapid vertical intercalations of Pay Sand

(Class 1) and non-reservoir backgrounds. The lower stratigraphic zones (Z6 to Z8) are predominantly carbonate sequences. The strict petrophysical cut-offs isolate the commercially viable porous carbonates (Class 2) from the heavily cemented tight carbonates and shales (Class 3) without being biased by natural radioactivity variations.

The visual validation across the representative wells confirms that Class 3 (tight background) effectively consolidates compositionally distinct but acoustically similar lithologies. For instance, the tight sands present in the upper zones and the tight carbonates in the lower zones are uniformly grouped into the gray background class. Unlike unsupervised statistical clustering algorithms that often artificially segregate these tight rocks due to variances in their nuclear log signatures, the proposed deterministic method groups them based on their shared lack of reservoir quality and characteristically high acoustic impedance. Consequently, applying this logic to all 18 wells provides a physically meaningful, seismically consistent target dataset devoid of circular reasoning, supporting reliable facies prediction from the seismic data.

To independently validate the deterministic classification and demonstrate the petrophysical distinctness of the defined classes across the entire field, statistical distributions and rock-physics crossplots were generated using the filtered dataset of all 18 wells (Figure 4, Figure 5 and Figure 6).

The statistical boxplots and overlaid density histograms of effective porosity (PHIE) (Figures 4 and 5) show a clear separation among the target classes. Class 1 (Pay Sand) exhibits a prominent high-porosity distribution (median ~25%), confirming its superior reservoir quality and unhindered flow potential. Class 2 (Porous Carbonate) displays an intermediate porosity range, which is commercially viable given the inherently rigid framework of carbonate rocks. In clear contrast, Class 3 (Tight Background) is strongly anchored near the zero-porosity baseline. Furthermore, the gamma-ray boxplot (Figure 4) illustrates the strong compositional heterogeneity within Class 3, which encompasses both clean tight carbonates (low radioactivity) and dense shales (high radioactivity). This highly variable nuclear signature highlights a practical limitation in conventional unsupervised algorithms (e.g., SOM): such data-driven methods would erroneously partition this non-reservoir class into multiple artificial clusters based on radioactivity variance, completely ignoring their shared lack of fluid-flow potential.

The elastic rock-physics crossplots (sonic transit time vs. bulk density) and the density overlap analysis provide strong justification for the seismic viability of this classification scheme (Figure 4 and Figure 5). In the elastic domain, Class 1 forms a distinct, easily separable cluster characterized by low bulk density and high transit time (i.e., minimal acoustic impedance), rendering it an ideal target for seismic anomaly detection. Conversely, the elastic analysis reveals a critical overlap between the non-reservoir rocks of Class 3 and parts of the carbonate framework. Despite their distinct mineralogical origins (e.g., highly cemented sandstones versus dense limestones), the non-productive rocks in Class 3 exhibit generally high structural stiffness and elevated bulk densities, translating to characteristically high acoustic impedance.

By deterministically consolidating these acoustically indistinguishable and petrophysically tight lithologies into a single unified background class, the proposed workflow reduces elastic ambiguity. The resulting target labels are aligned with the resolving limits of seismic data. Consequently, this physically constrained validation addresses the issue of circular reasoning, ensuring that the target variables fed into the subsequent supervised classification algorithms are both geologically meaningful and seismically predictable.

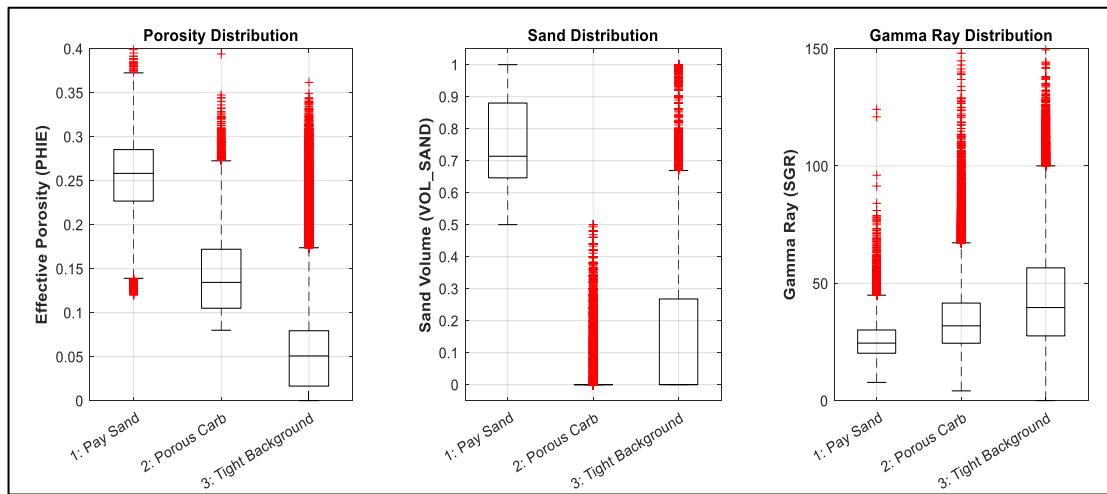


Figure 4: Statistical boxplots of effective porosity, sand volume, and gamma ray for pay sand, porous carbonate, and tight background intervals.

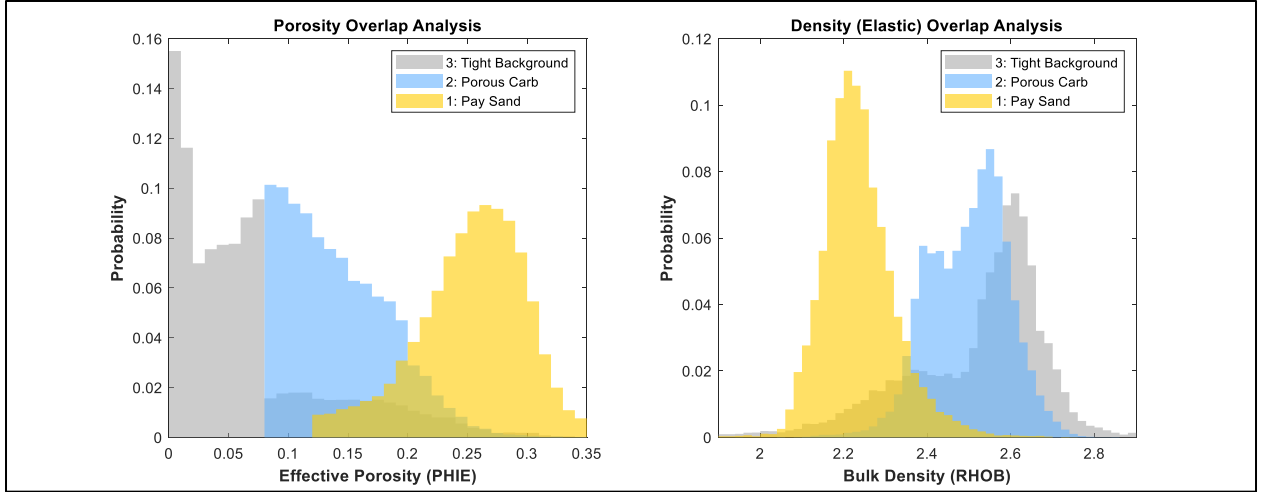


Figure 5: Overlap analysis of effective porosity (left) and bulk density (right) distributions for tight background, porous carbonate, and pay sand intervals.

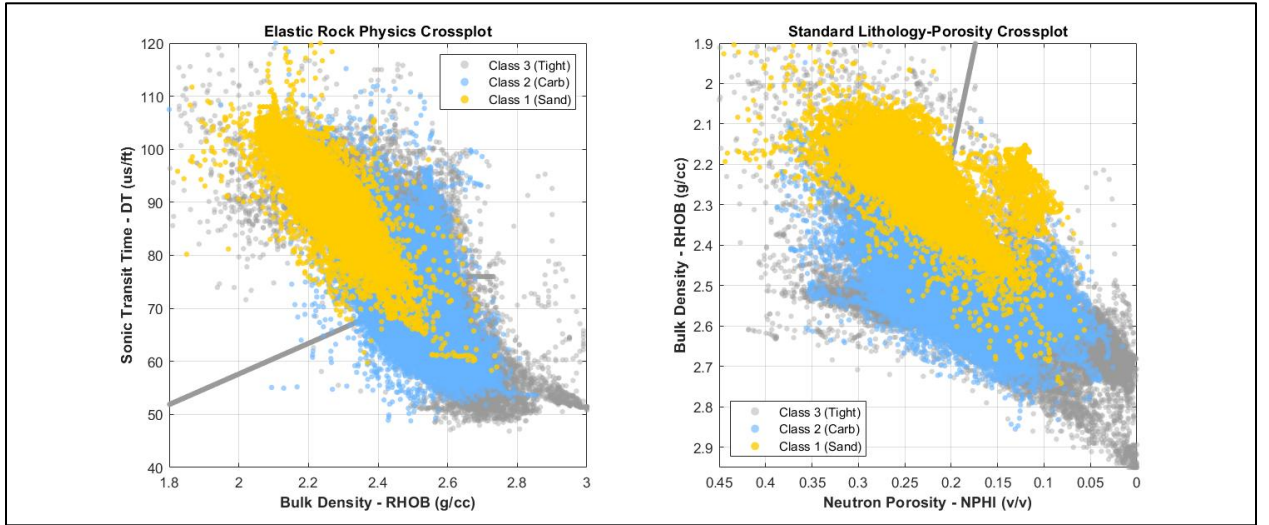


Figure 6: Elastic rock-physics crossplots (sonic transit time vs. bulk density) and (neutron porosity vs. bulk density).

3.2. Model-based estimation of acoustic impedance

To obtain geologically meaningful results, a constrained least-squares inversion method was used. In this approach, the solution for acoustic impedance depends not only on seismic data but also on the low-frequency background model, the seismic wavelet, and the weighting factor that controls how these components are combined. In addition, because seismic traces always contain noise, the resulting impedance model is also affected by this noise. During the inversion

analysis and testing, different low-frequency models and wavelets, together with different weighting values, were evaluated, and the optimal combination was selected based on the accuracy of the inverted acoustic impedance near the wells.

For this purpose, the agreement between two independently derived acoustic impedance estimates was quantified using the Pearson correlation coefficient computed at the well locations (Haykin, 2009). The first, available only at the well locations, was the well-log-derived acoustic impedance, calculated as the product of the density log and the acoustic (sonic) velocity log. The second, available throughout the seismic volume, was the acoustic impedance inverted from the migrated post-stack 3D seismic data.

3.3. Porosity estimation using Hampson-Russell software

The total or effective porosity is another key petrophysical material property of great interest in reservoir studies. Porosity is routinely measured in various ways: on core samples in the laboratory, from well-log data, from interval velocities of seismic traces, and from mathematical modelling. Well-log and core-sample observations yield high detail and spatial resolution of measurements, but they are only available at sparse locations. By contrast, seismic data provide continuous areal coverage between well locations. To obtain such coverage, one needs to find a combination of seismic attributes that can be calibrated at the wells and used as a proxy for porosity.

In this study, an ANN is used to predict porosity, incorporating AI together with additional seismic attributes. Several algorithms, including the Radial Basis Function Network, the Multi-Layer Feed-Forward Network, and the Generalized Regression Neural Network (GRNN) (Haykin, 2009), were evaluated using Hampson-Russell software, with the GRNN providing the most accurate porosity estimates.

3.4. Classification of seismic facies

Seismic facies are defined as generally 3-D units composed of certain groups of seismic reflections that differ from adjacent groups. Geologically, seismic facies represent sedimentary facies and/or certain structural and stratigraphic characteristics recognized within the patterns of seismic amplitudes, phases, or dips (Baaske et al., 2007). Broadly, these localized patterns in seismic data are represented by seismic attributes, which are secondary records extracted from

seismic data by various mathematical transformations. When wisely extracted using specifically targeted techniques, seismic attributes can highlight specific geological and petrophysical characteristics (Amendola et al., 2017; Hampson et al., 2001). Seismic attributes have been widely used in many studies to characterize and evaluate reservoir properties. (e.g., Ashraf et al., 2019; Du et al., 2015; Farzadi, 2006; Riedel et al., 2013a, 2013b; Roy et al., 2013; Wrona et al., 2018; Zhang et al., 2019; Zhao et al., 2015, 2018).

In modern practice, machine learning offers an automated solution for efficiently classifying reservoir facies from quantitative seismic attributes, as manual interpretation of continuous 3D data is time-consuming and impractical. Recently, there has been a tremendous increase in the number of studies using multivariate regression and ANN applications for automating the more challenging tasks in seismic interpretation and seismic pattern recognition. (e.g., Baaske et al., 2007; Bagheri and Riahi, 2015; Coléou et al., 2003; Feng et al., 2018a, 2018b; Linari et al., 2003; Liu et al., 2018; Marroquín, 2014; Roden et al., 2015; Saggaf et al., 2003; Sayago et al., 2012; Zahraa et al., 2017; Zahmatkesh and Kadkhodaie, 2025).

In this study, we compare the CFE-ANN workflow with several supervised pattern-recognition workflows to characterize lateral and vertical variations in the reservoir in terms of seismic facies. Three supervised classification methods were applied to the same dataset, and their performance, similarities, and differences were systematically evaluated. The results show that the proposed workflow, combining engineered attributes with an ANN classifier, provides improved classification performance compared to conventional approaches. In addition, incorporating statistical confidence measures leads to noticeable differences in the interpretation of the classification results.

3.4.1. Seismic attributes

A key step in this process is the careful selection of input attributes from the large number of transformations that can be derived from 3-D seismic data. Identifying an optimal set of attributes remains a major challenge in classification workflows.

In addition to acoustic impedance (AI) from model-based inversion and porosity estimated in the Hampson-Russell (EMERGE) software, 23 attributes were extracted from the

seismic reflection data, resulting in a total of 25 attributes (Table 2). Some of these attribute sections are shown in Figure 7.

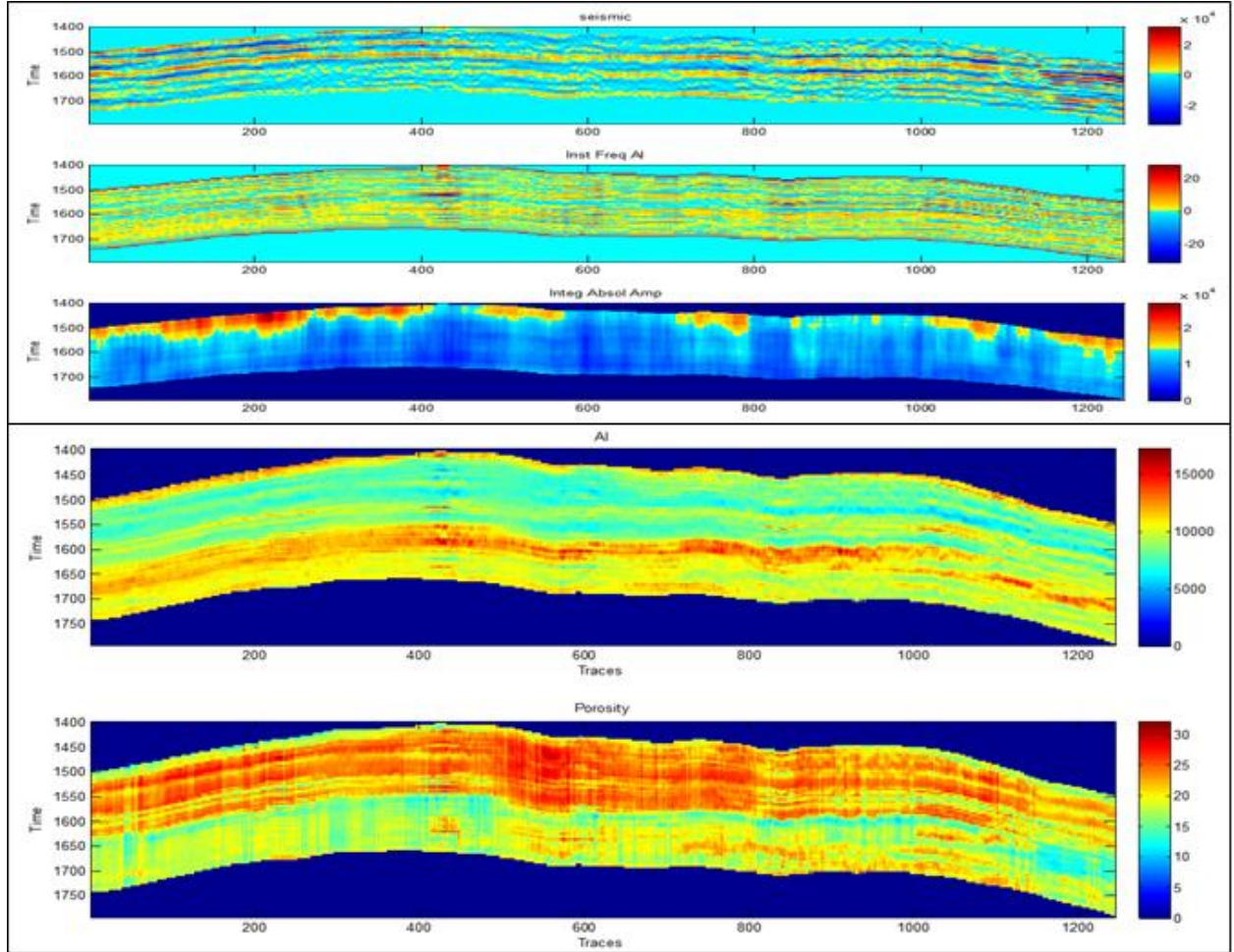


Figure 7: Some of the selected original attributes (from top to bottom: Seismic amplitude, Instantaneous frequency of acoustic impedance, Integrated absolute amplitude, acoustic impedance and porosity).

Table 2: List of the original 25 attributes.

<i>Number</i>	<i>Name</i>	<i>Description</i>	<i>Reservoir and geological implications</i>
1	FFT 20 Hz	Amplitude of Fourier spectrum at the given frequency calculated from a short time window	- Useful for sequence boundaries and thin bed tuning - Correlates with lithology and porosity
2	FFT 40 Hz		
3	FFT 60 Hz		
4	Similarity	Similarity of adjacent seismic traces is defined by the following equation: $\text{Similarity} = 1 - \frac{\sqrt{\sum_{i=1}^N (X_i - Y_i)^2}}{\sqrt{\sum_{i=1}^N X_i^2} + \sqrt{\sum_{i=1}^N Y_i^2}}$	- Indicates lateral continuity of lithofacies. - Visualize channels, reef edges, karst features, fracture zones
5	RMS Energy within floating time gate (-12ms, +12ms)	Root-mean square (sum of squared sample values within divided by the number of samples) of seismic record within the specified floating time gate	-Highlight hydrocarbons and thick sands - Identify stratigraphic features which are isolated from background features, like channels and bright spots
6	RMS Energy within floating time gate (-24ms, +24ms)		
7	Fingerprint Seismic & AI	Similarity between the traces of seismic signal and AI (or also NHI) taken at each sample position inside the seismic data volume.	-Useful to find similar good reservoirs as encountered in certain wells
8	Fingerprint Seismic & AI & NPHI		
9	Amplitude-Weighted Cosine Phase AI	Product of the amplitude envelope and the cosine of the instantaneous phase of the AI	
10	Amplitude-Weighted Frequency AI	Product of the amplitude envelope and the instantaneous temporal frequency of the AI	
11	Amplitude-Weighted Phase	Product of the amplitude envelope and the instantaneous phase (of the seismic signal or AI)	
12	Amplitude-Weighted Phase AI		
13	Average Frequency AI	Average of the amplitude spectrum of the AI over a small window centered on the time sample.	
14	Cosine Instantaneous Phase	Cosine of the instantaneous phase of seismic signal or AI	Highlight bedding very well and enhance the visibility of structural and stratigraphic features
15	Cosine Instantaneous Phase AI		
16	Derivative Instantaneous Amplitude AI	Derivative of the amplitude envelope of the AI	
17	Derivative Seismic	Derivative of the input trace, calculated as the difference between adjacent samples	
18	Dominant Frequency Seismic	Maximum of the amplitude spectrum over a small window around the time sample	
19	Filter 5-10-15-20 AI	AI or seismic record filtered within the specified frequency band	
20	Filter 5-10-15-20 Seismic		
21	Seismic	Post-stack seismic amplitude	
22	Instantaneous Frequency AI	Time derivative of the phase of the AI	Attenuation anomalies and thin bed tuning, variations of fluid content
23	Integrated Absolute Amplitude	Integration of the amplitude envelope of the input trace	Show bedding continuity, lithology, fluid content
24	AI	Acoustic Impedance derived from inversion of amplitude seismic data	Indicate lithological and porosity changes, discontinuities, unconformity surface, sequence boundaries and fluid content in the reservoir
25	Estimated Porosity	Porosity estimated using EMERGE	Primary petrophysical property of rock

3.4.2. Preconditioning and selection of optimal attributes

Feature selection, also known as dimensionality reduction in machine-learning applications, is used to optimize data inputs by identifying a subset of extracted features with the greatest predictive power. In this study, several conventional feature-selection methods were first investigated and found inadequate for the present dataset, motivating the feature-engineering approach described below.

A commonly used approach to feature reduction is based on pairwise correlations between attributes (Riedel et al., 2013a,b). The Pearson correlation matrix R is computed using the seismic attribute values (time series extracted near well locations), as shown in Figure 8. Highly correlated attribute pairs, indicated by strong positive or negative values (e.g., red or blue colors in Figure 8), are expected to provide similar information to the predictive model. Therefore, one attribute from each highly correlated pair can be considered redundant and removed from the classification process.

A standard feature-selection approach involves removing, from each pair of correlated attributes, the one with lower correlation to the target variable and lower relevance to the classification objective (Haykin, 2009). This approach was tested; however, it resulted in reduced model performance, likely due to the loss of potentially informative attributes.

Figure 9 illustrates the distributions of selected attributes and the corresponding Pearson correlation coefficients. Each off-diagonal subplot shows a scatterplot of a pair of variables with a least-squares regression line (pink), whose slope corresponds to the correlation coefficient. The diagonal plots display histograms of the individual attributes.

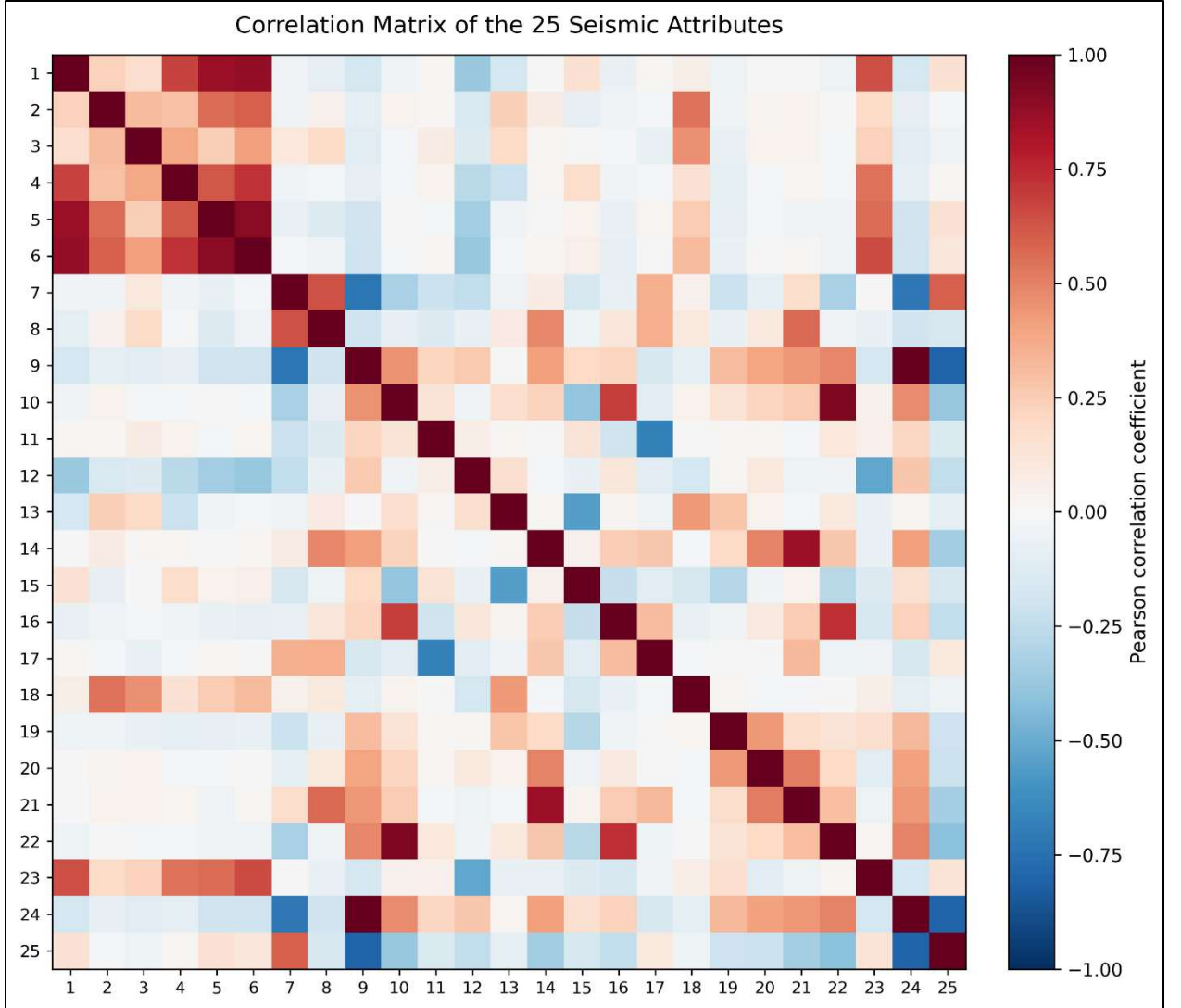


Figure 8: Correlation matrix of the 25 input seismic attribute volumes. Rows and columns of this matrix correspond to the sequence of seismic attributes in Table 2. A correlation of ± 0.7 or higher may indicate redundancy.

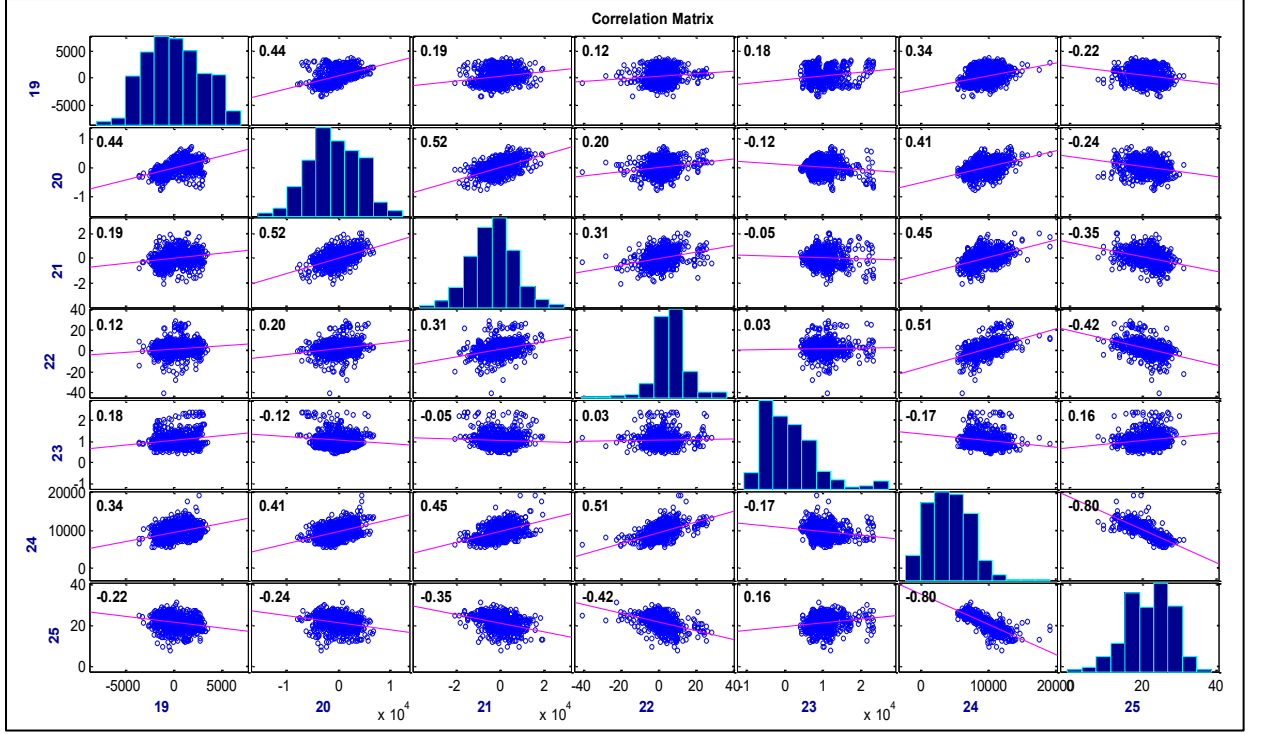


Figure 9: Cross-plots of pairs of the some raw attributes. Attribute numbers (as listed in Table 2) are shown along the bottom and left of the grid of panels. Histograms of the distributions of the attributes are shown on the diagonal of the grid. Pink lines indicate the correlations between pairs of attributes. Notice that AI (24) and porosity (25) have the highest correlation.

Bulb-shaped scatterplots with nearly horizontal regression lines (in pink) indicate weakly correlated attributes that are suitable for classification. However, several attributes show strong correlations and complex distribution patterns, which require careful selection and appropriate preconditioning.

Rather than selecting a subset of the raw attributes, we therefore adopt the combinatorial multi-classifier feature-engineering scheme that constitutes the central element of this study. The classification problem comprises three lithofacies classes. A three-class problem admits six binary decompositions: three one-versus-rest splits and three pairwise splits in which one class is excluded and the remaining two are compared directly. Using all six splits rather than an arbitrary subset, ensures that every possible binary grouping of the three classes is represented. This approach is conceptually related to error-correcting output codes (Dietterich and Bakiri, 1995) and stacked generalization (Wolpert, 1992), both of which employ auxiliary classifier outputs, rather than raw features alone, as inputs to a downstream model.

For each of the six binary splits, a separate binary classifier was trained on the original 25 attributes, and that classifier's continuous discriminative output was retained as a new engineered attribute. Four classifier types were employed: (1) regularized logistic regression, (2) a support vector machine (SVM), (3) a shallow feed-forward neural network, and (4) a random forest classifier. Each was applied to all six splits, yielding six engineered attributes per method. The four types were chosen deliberately to span a range of functional forms, from linear to fully nonlinear and ensemble-based:

- **Regularized logistic regression** (Haykin, 2009): The retained attribute is the raw linear score $X\theta$ (prior to the sigmoid transform), where θ is the fitted coefficient vector for that split. This score is unbounded and directly inspectable. The magnitude and sign of each entry in θ indicate how strongly, and in which direction, each of the 25 original attributes contributes to that split's decision.
- **SVM with a radial basis function (RBF) kernel** (Cortes and Vapnik, 1995): The retained attribute is the SVM decision-function value: the signed distance of each sample from the separating hyperplane in the nonlinear feature space induced by the RBF kernel. Its sign indicates the predicted side of the binary split, and its magnitude reflects the strength of that assignment. Like the logistic-regression score, this quantity is unbounded, but it derives from a nonlinear kernel projection of the original attributes rather than a directly inspectable linear combination.
- **Shallow feed-forward neural network**: The retained attribute is the softmax output of a small network trained on the same split, bounded on $[0, 1]$ and interpretable as a class-probability estimate.
- **Random forest** (Breiman, 2001): The retained attribute is the out-of-bag (OOB) vote fraction from a 100-tree ensemble, also bounded on $[0, 1]$. Using OOB rather than in-sample prediction ensures that each sample's derived attribute value comes only from trees that did not train on that sample, avoiding a data-leakage bias that would otherwise artificially inflate downstream classification performance.

Applying all four classifier types to all six splits yields 24 engineered attributes, intended to augment (not replace) the original 25-attribute suite as input to the final multiclass lithofacies classifier.

Because all four derived-attribute types are, by construction, functions of the same 25 original attributes, (differing in functional form and interpretability, not in information source) their added value was tested empirically. Five candidate attribute sets were constructed: the original 25 attributes alone (hereafter set 0, the baseline); its incremental augmentation with the logistic regression (31 attributes), SVM (37 attributes), and neural network (43 attributes) attribute blocks; and all 24 derived attributes combined with the original suite (49 attributes, hereafter set 5). Each set was evaluated with a 100-tree random forest classifier using out-of-bag (OOB) accuracy (an unbiased generalization estimate requiring no separate held-out split, since each tree's OOB samples are excluded from its training) To account for training stochasticity, each set was evaluated over ten runs with independent random seeds, and per-class OOB accuracies were recorded as mean $\pm$ standard deviation (Figure 10).

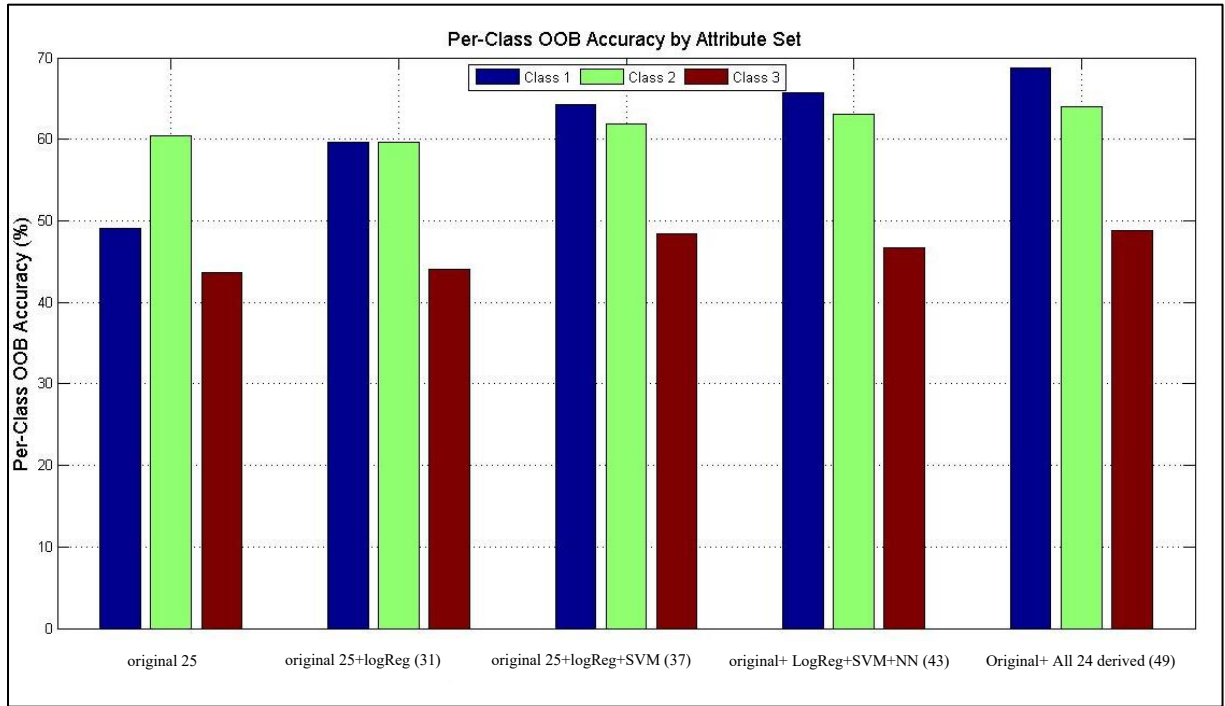


Figure 10: Per-class out-of-bag (OOB) classification accuracy across the five candidate attribute sets, using a 100-tree random forest classifier.

Three candidate feature sets were then evaluated for dimensionality reduction: set 0 (25 raw attributes), set 5 (all 49 attributes), and set 6 (the 24 derived attributes only). set 6 was constructed to test whether the engineered attributes render the originals redundant. Each set was subsequently reduced by principal component analysis (PCA) to the number of components required to reach 95% cumulative explained variance (14, 18, and 9 components, respectively), generating three additional sets. Attributes were z-score standardized prior to PCA, and components were computed by eigenvalue decomposition of the covariance matrix and retained in order of descending eigenvalue. All six candidates were evaluated with a 100-tree random forest classifier using OOB accuracy. Each evaluation was repeated over ten runs with independent random seeds, and overall and per-class accuracies were recorded as mean $\pm$ standard deviation. This is the same validation protocol used in the preceding feature-engineering evaluation.

Figure 11 presents the per-class results, which mirror the overall accuracies class by class. Set 6 PCA achieves the highest per-class OOB accuracy for all three classes: approximately 69% (class 1), 66–67% (class 2), and 55% (class 3). The class 3 result is the most consequential. It exceeds the next-best candidate (set 6 raw, ~50%) by roughly five percentage points on the class that has proven hardest to resolve throughout this study. Relative to the baseline (set 0), the advantage of the engineered sets is concentrated in class 1 (~64–69% versus 49%) and class 3 (~45–55% versus 42–44%), while class 2 varies only modestly (~60–67%). PCA affected the two engineered sets in opposite directions: set 5 PCA declined relative to its raw form on all three classes, whereas set 6 PCA improved on all three, indicating that the benefit of PCA is specific to the compact, derived-only attribute space rather than a general property of dimensionality reduction on this dataset.

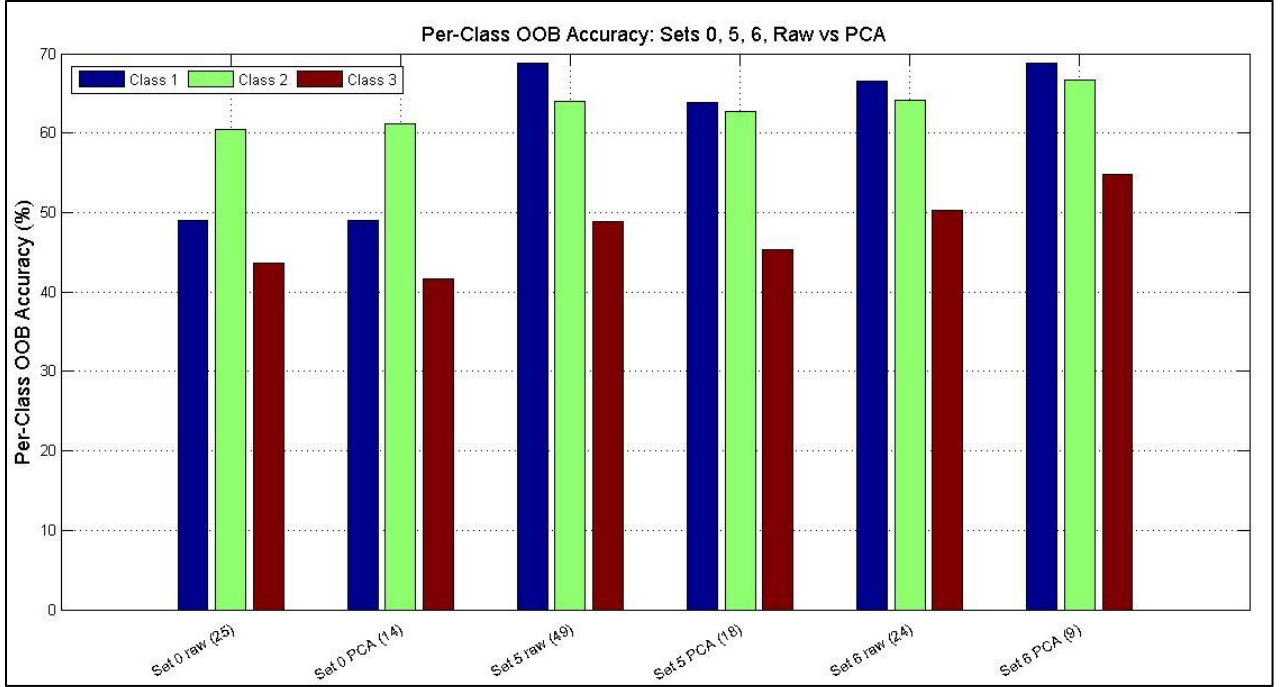


Figure 11: Per-class OOB accuracy for sets 0, 5, and 6, evaluated raw and after PCA reduction.

Because OOB accuracy is computed from samples drawn from the training wells and is specific to one classifier family, all six candidate feature sets were additionally evaluated with a feed-forward neural network on a blind-well basis. Well 096 was withheld entirely from training, normalization, dimensionality reduction, and hyperparameter tuning. Normalization statistics and PCA projections were fitted on the development data only and applied unchanged to the blind well, preventing information leakage. Hyperparameters (regularization strength and hidden-layer size) were tuned independently per set via grid search on a depth-blocked inner tuning set, and the winning configuration was retrained over three random initializations. The network was a single-hidden-layer architecture trained by conjugate-gradient minimization of a regularized cross-entropy cost with softmax outputs.

Figure 12 compares training and blind-well accuracy. All four sets containing the original attributes and their PCA overfitted, most severely set 5 raw, with a training-to-blind gap of ~42 percentage points and blind-well accuracy of only ~57%. The derived-only sets generalized

markedly better: set 6 PCA achieved the highest blind-well accuracy (~76%), exceeding set 6 raw by ~6 percentage points, consistent with its advantage under OOB evaluation.

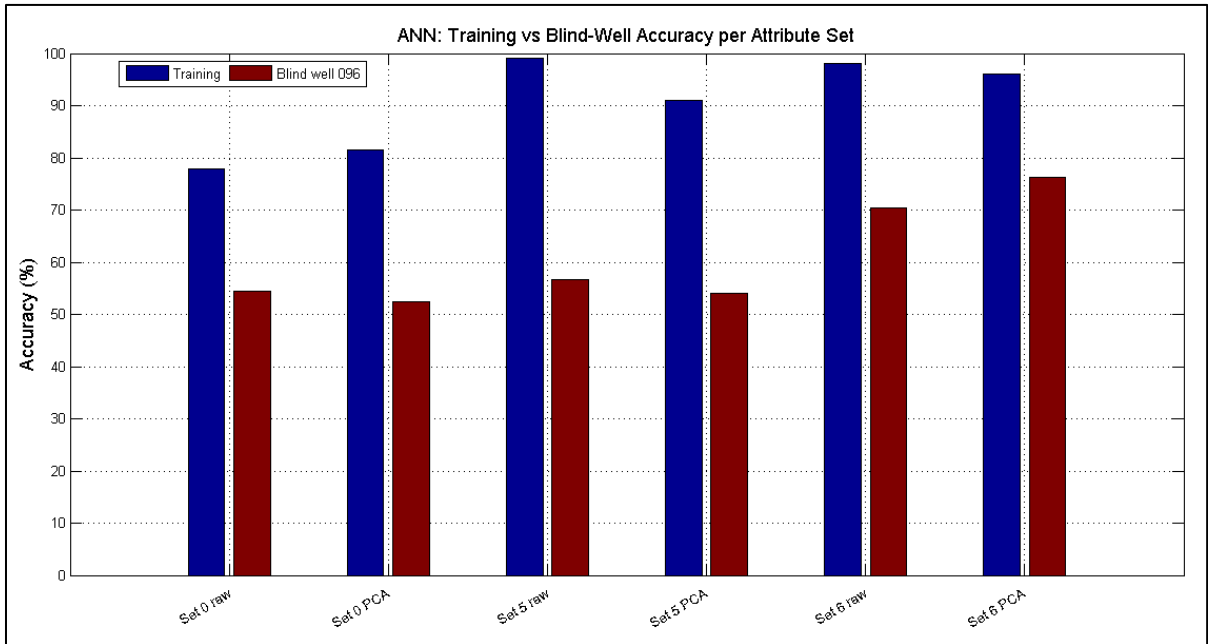


Figure 12: Neural network training accuracy versus blind-well (well 096) accuracy for all six candidate feature sets, each with independently tuned regularization and hidden-layer size.

Multiple lines of evidence, across different models and validation methods, all point to the same conclusion. First, the engineered attributes are unambiguously valuable: they outperform the original 25-attribute suite by ~9–13 percentage points under random forest OOB evaluation and by ~22 percentage points under blind-well neural network evaluation, confirming that the combinatorial multi-classifier attribute generation strategy encodes discriminative structure the raw attributes do not readily express. Second, the value of PCA proved set-dependent. It was costly for set 5 and inconsequential for set 0, but decisively positive for the derived-only space. For set 6, projection onto nine principal components improved overall and per-class OOB accuracy, delivered the best blind-well generalization, and yielded by far the most favorable sample-to-dimension ratio. Third, set 6 PCA is therefore recommended as the final training set, achieving the highest overall OOB accuracy (~63.5%), the highest per-class accuracy for all three classes (including ~55% on Class 3, the study's most difficult class) and the

highest blind-well accuracy ($\sim 76\%$). By contrast, set 5 raw degrades to $\sim 57\%$ on the unseen well despite near-perfect training accuracy ($\sim 99.5\%$), indicating that it overfits the training data. This recommendation revises the preliminary preference for the full 49-attribute set: the blind-well evidence shows that retaining the raw attributes substantially reduces predictive performance on wells not used in training.

3.4.3 Confidence of predicted classes

The primary objective of most applications in this study is to obtain a discrete classification of the dataset, that is, to assign an integer class label to each data point. To classify the data into N_c classes, the training dataset must be split into N_c subsets and N_c predictors $f_j(\mathbf{x})$ must be constructed. In most applications, the values of $f_k(\mathbf{x})$ will lie within the interval $[0,1]$. When training the algorithm or constructing another model for the k^{th} predictor, the target output for the i^{th} data point is specified as $f_k(x_i) = 1$ if the point belongs to the k^{th} class and $f_k(x_i) = 0$ otherwise. After training or inversion, the model will predict larger values $f_j(\mathbf{x}_i)$ for feature vectors $\mathbf{x}_i$ belonging to class j . Using these outputs, the predictors can be normalized as:

$$p_k(x) = \frac{f_k(x)}{\sum_{i=1}^K f_i(x)} \quad (1)$$

These normalized predictors sum to one, and consequently $p_k(\mathbf{x})$ can be interpreted as the probability of finding class k at point $\mathbf{x}$. Note that this is the same normalization as used, for example, in the PNN method.

Based on the normalized predictors, the classification decision appears straightforward: for a feature vector $\mathbf{x}$, the class k with the largest $p_k(\mathbf{x})$ should be selected (Haykin, 2009). Let us denote the corresponding highest p_k by $P_{\max}$. However, this selection would clearly have a low degree of statistical confidence if other values of $p_j(\mathbf{x})$ are close to the selected one. Therefore, the classification procedure outputs not only the most likely class label k but also the estimated probability $p_k(\mathbf{x})$. This approach generates images of the estimated probability for each selected class across subsurface samples, as well as images depicting classification confidence.

The advantage of using ANN predictors for classification lies in their nonlinearity, universality, and flexibility. Numerous ANN architectures can be used, with different numbers of layers, neurons, and shapes of the activation functions. With larger numbers of ANN weights and

bias terms, training can be made more accurate. At the same time, the universality of the ANN model can be a limitation. The predictors are obtained by purely empirical fitting of the binary training data consisting of zeros and ones, and the resulting values cannot be rigorously viewed as probabilities of the classes. With large numbers of ANN weights, the ANN can overfit the training data and allow noise to influence the model.

When logistic regression is used to obtain class predictors, its advantage lies in the clear physical meaning of the output quantity as the probability of finding the specified class k at point (feature set) $\mathbf{x}$ in the data. This probability is defined within the expected range $[0,1]$, although, unfortunately, the different classes do not exclude each other. A disadvantage of the standard logistic regression approach is the assumption of linearity between variables and consequently limited accuracy of data fitting. In one pass of logistic regression, the probability of only one class is estimated. For multiclass classification, one class k , can be selected while all other classes are considered as non-occurrences of k , and the probability of class k is estimated. This method is known as One-versus-All classification. After repeating this estimation for each class k , their predictors $f_k(\mathbf{x})$ can be combined by the equation (1). Alternatively, multinomial regression can be used (Haykin, 2009).

Thus, a large number of choices for selecting prediction and/or classification algorithms are available, and the appropriate algorithm should be selected by trials and quality control. However, the selection of the criteria for this model quality control may become a nontrivial issue in geophysical applications.

Using probability normalization in equation (1), outputs of any ANN or similar prediction algorithms can be normalized into the range $[0,1]$, which allows them to be interpreted as estimated probabilities (likelihoods) of finding these classes at a given data point $\mathbf{x}$ (Haykin, 2009).

In many applications, it is useful to output not only the resulting class label k but also some measure of statistical confidence of the classification. Unfortunately, this issue of confidence is usually not considered in conventional approaches to AI-based classification, but it is important for reservoir characterization.

Two useful measures of classification confidence can be proposed using the normalized estimators $p_k(\mathbf{x})$ defined in equation (1). The first is $P_{\max}$ itself: it indicates the inferred probability of the class assigned to data point $\mathbf{x}$. Values of $P_{\max}$ close to one correspond to high confidence in this assignment, and values close to $1/N_c$ indicate low confidence.

A second measure, denoted $C(\mathbf{x})$, is obtained as the ratio of $P_{\max}$ to the second-largest of the $p_k(\mathbf{x})$:

$$C = \frac{P_{\max}}{P_{\text{next max}}} \quad (2)$$

This value is always greater than or equal to one, and larger values indicate lower ambiguity of the classification, i.e., situations in which class k is notably more likely than the next competing alternative.

Using the prediction probabilities or either of the two confidence measures defined above, the classified seismic sections were filtered to display only zones of sufficiently high confidence.

3.4.4 Selection of algorithm parameters

Following the selection of set 6 PCA as the final feature set, the definitive classification model was constructed. A grid search was performed over regularization strength $\lambda \in \{0, 0.5, 1, 2, 4, 8\}$ and hidden-layer size $\in \{10, 20, 30, 40, 50\}$ and the tuning results were plotted across the grid, identifying $\lambda = 1$ and a hidden-layer size of 40 as the optimal configuration. With these parameters, the final single-hidden-layer network was trained on all development samples by conjugate-gradient minimization of the regularized cross-entropy cost (1,000 iterations, with a fixed random seed for reproducibility). The resulting weight matrices were saved together with the complete preprocessing chain, the normalization statistics (μ , σ) and the PCA projection matrix fitted on the training data, since any unseen data must be normalized and projected with these same training-derived parameters before the model can be applied. This packaged model constitutes the deterministic classifier subsequently used to predict lithofacies on new, unseen seismic data. Application to unseen data therefore follows an identical transformation sequence:

the original attributes are first converted to the 24 engineered attributes using the saved split-classifier parameters, normalized with the stored μ and σ , projected onto the saved principal components, and finally classified by forward propagation through the saved network weights.

4. Results

After feature selection and ANN parameter tuning, three classifications of the studied reservoir were derived (Table 3): 1) a regular ANN (based on conventional feature selection), 2) the CFE-ANN (an identical network trained on the engineered feature set), and 3) a probabilistic neural network (PNN) implemented in Hampson–Russell software. The training and validation accuracies in Table 3 were computed using a random split of the development dataset, with a subset of samples held out for validation.

Table 3: Comparison of three classification workflows on training and validation sets.

	CFE-ANN	Regular ANN	PNN
Number of features	9 principal components	7 principal components	9 original features
Type of preconditioning and feature selection	Class-discriminative feature engineering via combinatorial multi-classifier binary splits (24 derived attributes), z-score standardization, principal component analysis (95% cumulative variance)	Z-score standardization and orthogonalization via singular value decomposition (principal component analysis)	Stepwise regression
Training Accuracy (%)	96	98	85
Validation accuracy (%)	77	75	57

4.1 Blind test evaluation

Figure 13 shows the results of applying the three methods described above to the classification of test data from blind well 096. The results indicate that misclassifications mainly

occur over short time intervals, likely due to high-frequency noise in the seismic attributes. The best match is between the labeled classes (far left) and the CFE-ANN (far right).

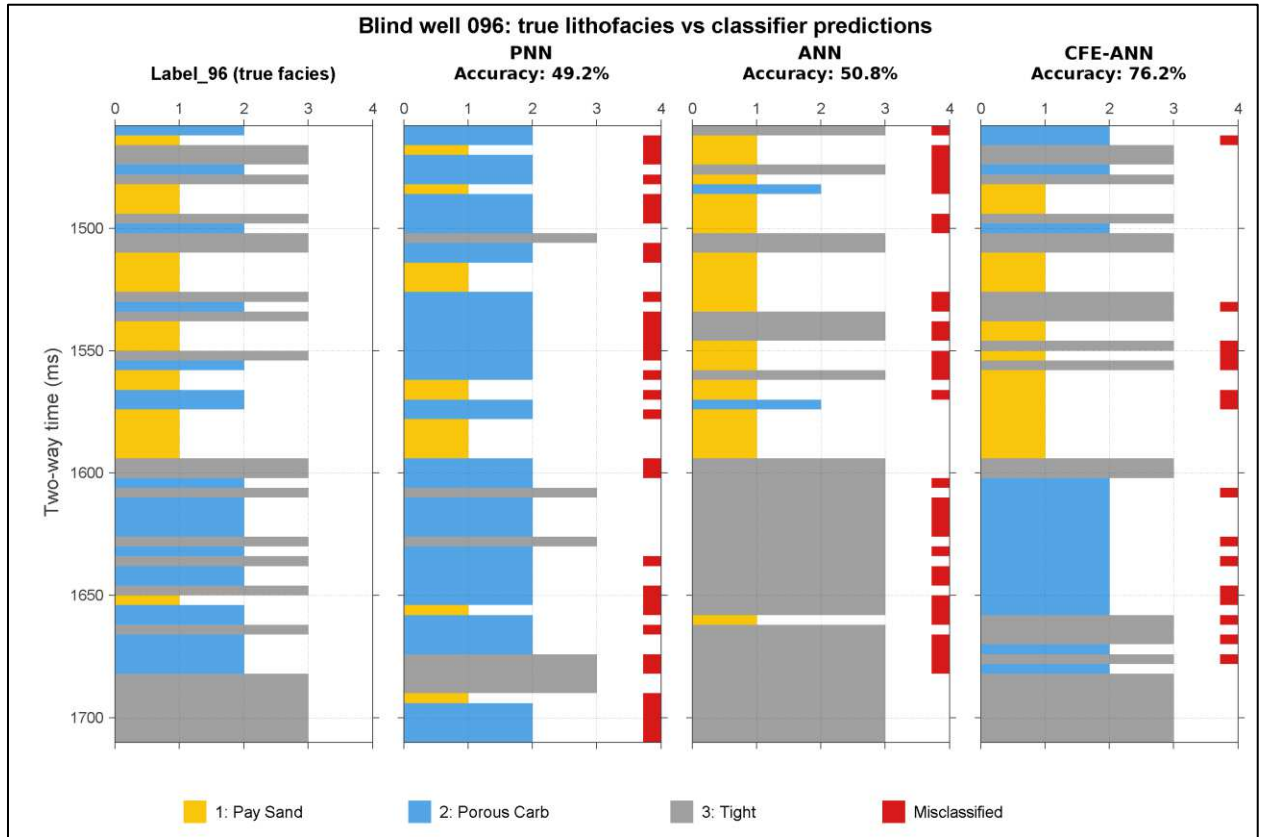


Figure 13: Results of classification of blind well 096. The panels are, from left to right: 1) known classification in the well, 2) results of the PNN, 3) the ANN, and 4) the CFE-ANN classification workflows. Yellow, blue, and grey colors represent classes 1, 2 and 3 respectively. Panel headings indicate the prediction accuracy.

Figure 14 shows the estimated confidence levels P_k for the class k identified at each sample in the preceding figure. The average level of confidence is highest for the CFE-ANN and lowest for the PNN. The confidence levels are moderate when the ANN is applied to the blind well. Overall, the CFE-ANN workflow appears superior to the ANN and PNN workflows, based on validation accuracy and classification confidence. The confidence level patterns obtained from the CFE-ANN and ANN show similar depth-dependent trends.

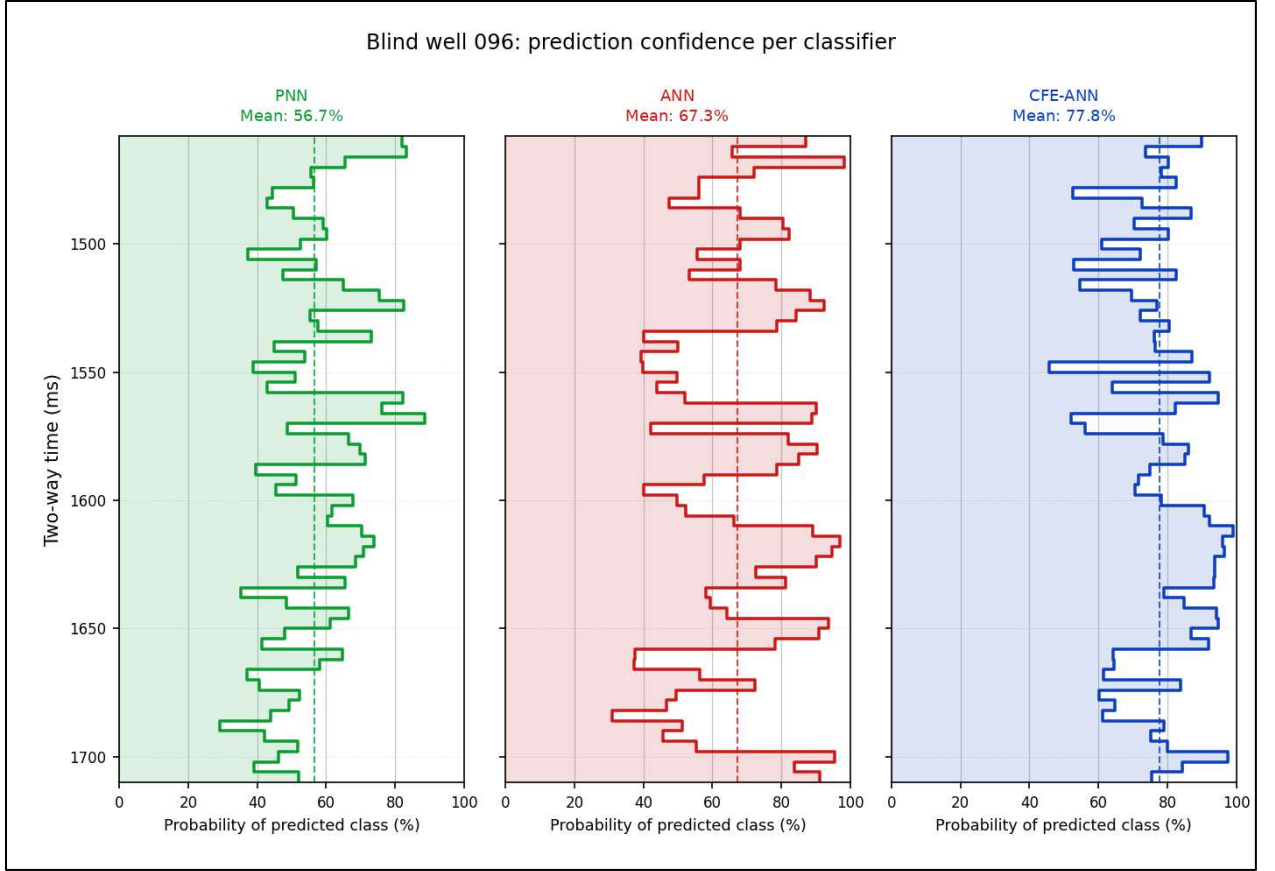


Figure 14: Probability of finding the class identified at each point of the subsurface for the three classification workflows on test data (well 096). The left panel shows the results of the PNN, the middle panel the ANN, and the right panel the CFE-ANN. Notice that the overall probability for the CFE-ANN is higher than for the ANN and PNN.

For a more detailed assessment of the accuracy of class predictions, confusion matrices are often used (Haykin, 2009). The confusion matrix visualizes the performance of the classification model by comparing the classes known from well interpretations and those predicted from seismic records. The k^{th} element of the j^{th} row of the confusion matrix represents the percentage of cases in which a data point known as class k is classified by the algorithm as class j . Thus, the diagonal of the confusion matrix represents the percentages of correctly classified samples, and the off-diagonal values are the percentages of various misclassifications. Ideally, a perfect classifier would result in a confusion matrix equal to the identity matrix, i.e., with values of one only on the diagonal.

Various metrics can be derived from confusion matrices to quantify the performance of machine learning algorithms, such as accuracy, the F1 score and Cohen's kappa statistic. The overall accuracy is the simplest performance metric used in classification problems. This metric is defined as the ratio of the number of correctly predicted examples to the total number of examples. In the present case, Cohen's kappa is a convenient measure that can handle both the multi-class and imbalanced class aspects of the dataset. The kappa score shows how well the model performs compared to a completely random classification. This score is defined as:

$$\kappa = \frac{P_0 - P_e}{1 - P_e} \quad (3)$$

where P_0 is the overall accuracy and P_e is the agreement expected by chance, computed from the class proportions. For a good class prediction ($P_0 \rightarrow 1$), kappa is also close to one, and for a near-random model, it is close to zero.

Figure 15 presents the confusion matrices for the three classification workflows applied to the test set (blind well 096); overall accuracies and Cohen's kappa are summarized in Table 4. Columns represent actual classes and rows predicted classes: diagonal cells contain correct classifications, the gray column and row report per-class precision and recall, respectively, and the blue cell gives overall accuracy.

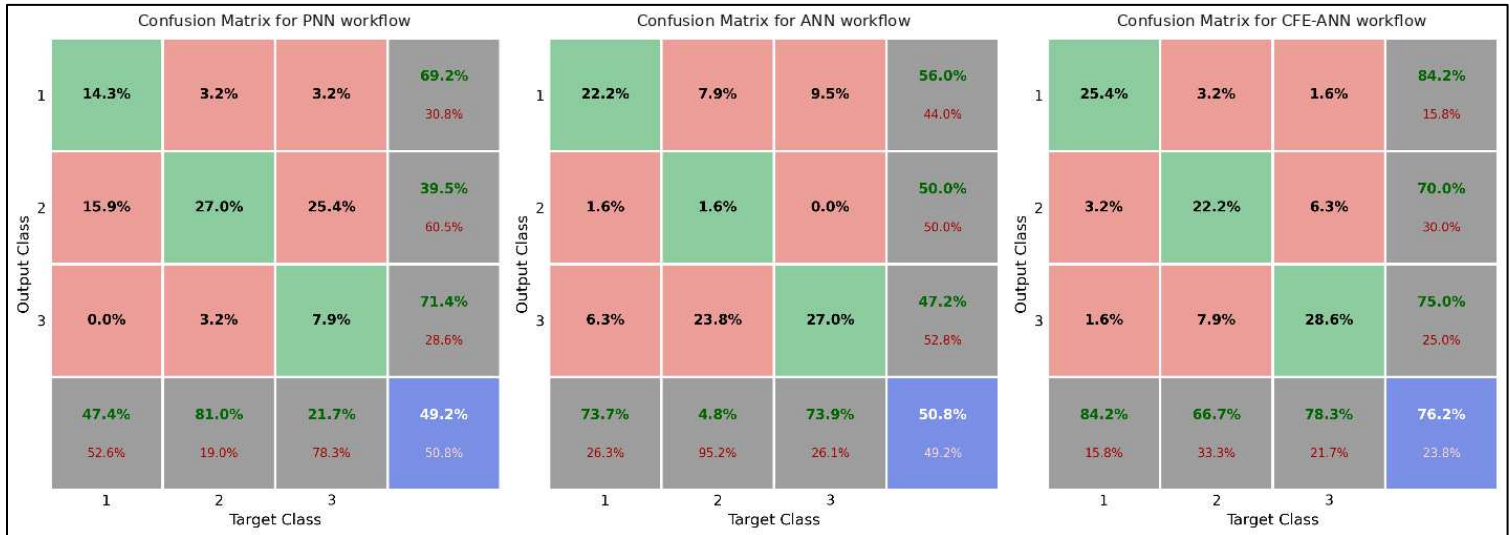


Figure 1: Confusion matrices for the PNN, ANN and CFE-ANN classification workflows. Columns represent the known class labels from the wells, and rows are the class labels predicted by the algorithm.

The PNN achieves 49.2% overall accuracy and is biased toward class 2, retrieving 81.0% of its samples but with only 39.5% precision and a class-3 recall of just 21.7%. The regular ANN (50.8%) shows the opposite pathology, an almost complete collapse on class 2 (recall 4.8%). The CFE-ANN attains 76.2% accuracy and is the only classification workflow with balanced performance across all three classes (recall 84.2%, 66.7%, and 78.3%), indicating a genuine gain in discriminative capability of the engineered feature space rather than preferential treatment of an easily separable class.

The predictive performance of the three classification workflows (CFE-ANN, ANN, and PNN) was evaluated on the blind well (well 096) using metrics derived from the confusion matrix. Precision is the fraction of the number of samples predicted as a class that truly belong to it. Recall is the fraction of the number of samples of a class that are correctly predicted.

The F1 score is the harmonic mean of precision and recall. It penalizes any imbalance between the two and vanishes when either quantity collapses. The macro average is the unweighted mean of a per-class metric over the three classes, treating each lithofacies as equally important regardless of its sample count. Balanced accuracy is the macro-averaged recall. Two chance-corrected measures are also reported. Cohen's kappa (κ) rescales the observed accuracy by the agreement expected from the class proportions alone, with κ near one for perfect prediction and near zero for a near-random model. The Matthews correlation coefficient (MCC) is a multiclass correlation between predicted and true labels and is robust to class imbalance. Table 4 summarizes all metrics for the obtained classification.

Table 4: Per-class and aggregate classification metrics for the CFE-ANN, ANN, and PNN classification workflows on blind well 096

Metric		CFE-ANN	ANN	PNN
Class 1	Precision (%)	84.2	56.0	69.2
	Recall (%)	84.2	73.7	47.4
	F1 score (%)	84.2	63.6	56.2
Class 2	Precision (%)	70.0	50.0	39.5
	Recall (%)	66.7	4.8	81.0
	F1 score (%)	68.3	8.7	53.1
Class 3	Precision (%)	75.0	47.2	71.4

Metric		CFE-ANN	ANN	PNN
	Recall (%)	78.3	73.9	21.7
	F1 score (%)	76.6	57.6	33.3
Macro average	Precision (%)	76.4	51.1	60.1
	Recall (%)	76.4	50.8	50.0
	F1 score (%)	76.4	43.3	47.6
Overall	Accuracy (%)	76.2	50.8	49.2
	Cohen's kappa	0.64	0.26	0.24
	MCC	0.64	0.29	0.29
	Balanced accuracy (%)	76.4	50.8	50.0

For this application, overall accuracy alone is an unreliable indicator of performance because it can remain moderate while a classifier fails entirely on one lithofacies. This is evident for the ANN, whose 50.8% accuracy conceals a near-total collapse on Class 2 (recall of 4.8%, F1 of 8.7%): the network rarely predicts this class at all. The PNN exhibits the complementary pathology, over-predicting Class 2 (recall of 81.0% at only 39.5% precision) while detecting barely one fifth of Class 3 samples. The most reliable metrics for this study are therefore those which expose these failure modes: the F1 score averaged across all classes and the coefficients corrected for chance agreement (κ and MCC), all of which weight every lithofacies equally or account for the class distribution. By these measures, the contrast between the methods is considerably sharper than the accuracy values suggest: the CFE-ANN attains a macro F1 of 76.4% against 43.3% (ANN) and 47.6% (PNN), and its κ of 0.64 falls within the substantial-agreement range (Landis and Koch, 1977), whereas $\kappa \approx 0.25$ for the two baseline networks indicates only fair, near-chance performance. The close agreement between κ and MCC for all three classification workflows further confirms the robustness of this ranking. The CFE-ANN is thus the only method that maintains a balanced precision–recall trade-off across all three facies, supporting its selection as the preferred classification workflow.

In summary, the confusion matrices and derived statistics show that the CFE-ANN workflow substantially outperformed both the regular ANN and the PNN on the blind well, achieving the highest accuracy, the only balanced per-class performance, and chance-corrected agreement in the substantial range. The CFE-ANN supervised classification workflow thus

provides a convenient and robust method for classifying the reservoir into the known lithofacies classes, with performance numerically superior to the two other benchmark workflows.

4.2 Spatial predictions and classification confidence

Figure 16 compares the lithofacies sections predicted by the three classification workflows along the studied seismic line, with the well locations overlaid for reference. The CFE-ANN produces a markedly more coherent geological image than either benchmark: predicted facies bodies form laterally continuous, subparallel layers that conform to the structural dip of the horizon, with pay-sand intervals concentrated in the upper part of the section and porous carbonate and tight facies organized into persistent bands beneath. This stratified architecture is consistent with the known depositional character of the study reservoir (a mixed system of interbedded sandstones and carbonates) in which lithofacies are expected to exhibit strong lateral persistence and vertical alternation rather than random spatial variation. By contrast, the regular ANN over-predicts pay sand, amalgamating the upper section into a single quasi-continuous sand body and thereby suppressing the interbedded carbonate intervals, while the PNN yields a fragmented, speckled distribution with poor lateral continuity, reflecting its class-2 bias identified in the confusion-matrix analysis. The geological plausibility of the CFE-ANN section (laterally traceable layering, preserved sand–carbonate interbedding, and facies transitions that honor the well control) therefore provides an independent, qualitative line of validation complementing the quantitative blind-well statistics, and indicates that the proposed feature-engineering strategy improves not only classification accuracy but the spatial fidelity of the predicted facies architecture.

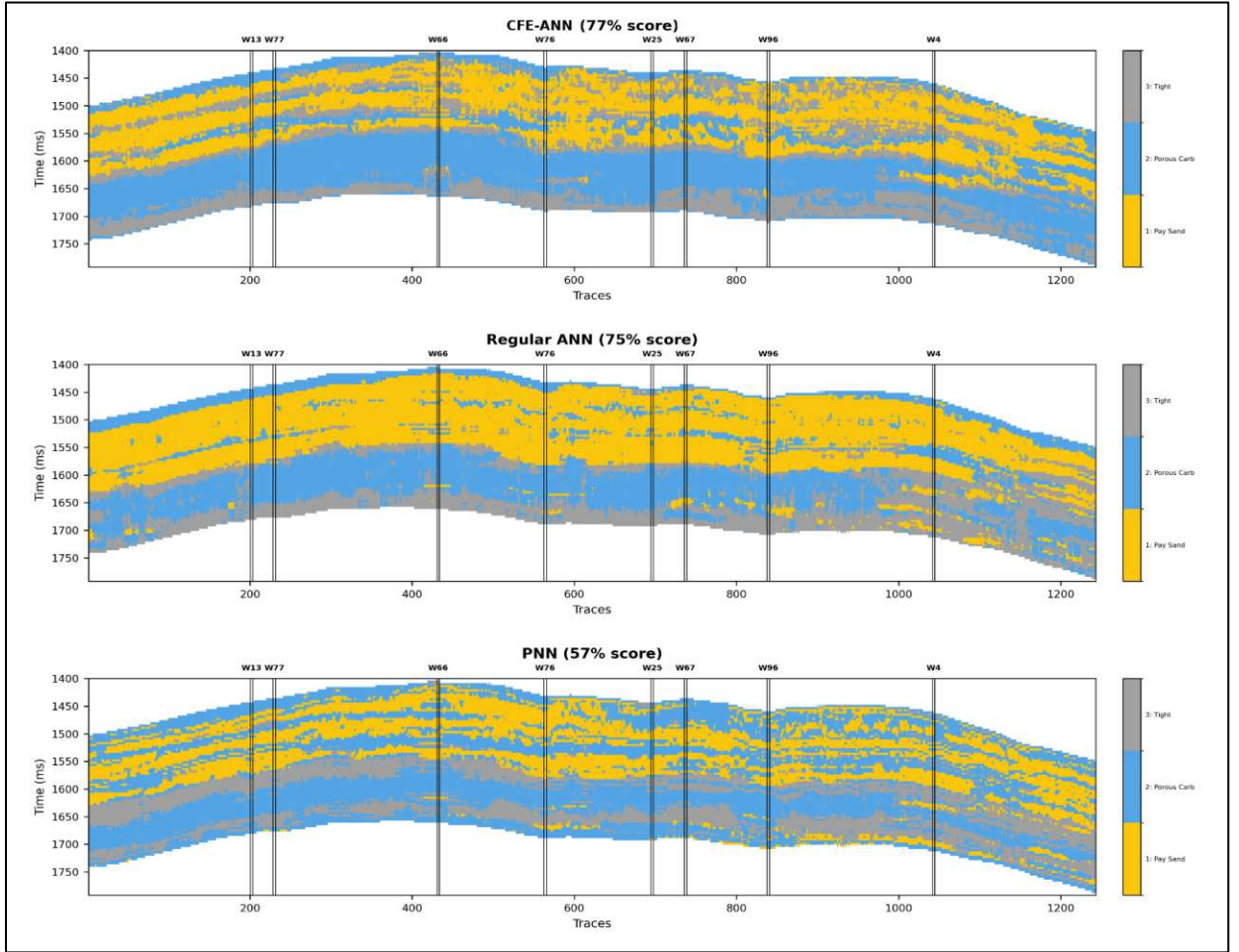


Figure 16: Classification cross-sections along the selected line using the CFE-ANN, Regular ANN, and PNN classification workflows (panel headings; percentages indicate validation scores). The color bar indicates the lithofacies classes.

The final classifications in Figure 16 consist of integer class labels determined from the estimated probabilities P_k of each of the classes and confidence levels C of the prediction. These probabilities contain important additional information that is, unfortunately, rarely utilized for characterizing hydrocarbon reservoirs. Nevertheless, this information may have a critical impact on interpretation. Figure 17 shows the confidence levels of the three classifications, also plotted in the form of continuous cross-sections.

As shown in these figures, the confidence values show clearer and more consistent layering within the reservoir than do the images of interpreted classes. The confidence levels for

the CFE-ANN exceed those of the PNN in particular, while the regular ANN also exhibits generally high confidence values. However, the meaning of the predictors P_k in the PNN algorithm is different fundamentally from outputs of an ANN, and therefore the confidence thresholds should likely be interpreted differently for these methods.

The estimated confidence levels can also be used to create enhanced and easier to interpret images of the classification. To highlight subsurface areas with reliable classification, a filter (mask) was applied to display data only where confidence $C \geq 70\%$ for ANN-based methods or $C \geq 50\%$ for the PNN. The filtered displays of classifications are shown in Figure 18. These images emphasize the layered character of the subsurface. Compared to Figure 16, the filtered section allows several significant observations. In the filtered CFE-ANN section, all three lithofacies are retained at high confidence, and the layered sand–carbonate architecture remains laterally traceable. In the regular ANN section, the upper pay-sand interval and the underlying porous carbonate band likewise persist above the 70% threshold, although the masked (white) areas are more extensive along the flanks of the section. The filtered PNN section is markedly more fragmented: substantial portions of the section fall below even the 50% threshold, and the tight facies (class 3) near the base of the section becomes intermittent due to low confidence.

However, the PNN predictors P_k are not directly comparable to the ANN softmax outputs: the PNN estimates class probabilities via Gaussian kernel density estimation, which yields inherently more diffuse posteriors than the near-decisive softmax values of the ANN-based methods. A numerically lower threshold therefore represents a comparable level of confidence, and thresholds of 70% (ANN and CFE-ANN) and 50% (PNN) were adopted for the filtered displays.

Notably, in the filtered PNN image, class 2 (porous carbonate) near the base of the section is retained at high confidence within intervals in which ANN-based methods (and the expected stratigraphy) indicate class 3 (tight facies). This misclassification shows that despite a high estimated statistical confidence, the actual classification may still be erroneous. This observation reinforces the class-2 over-prediction bias of the PNN identified in the confusion-matrix analysis.

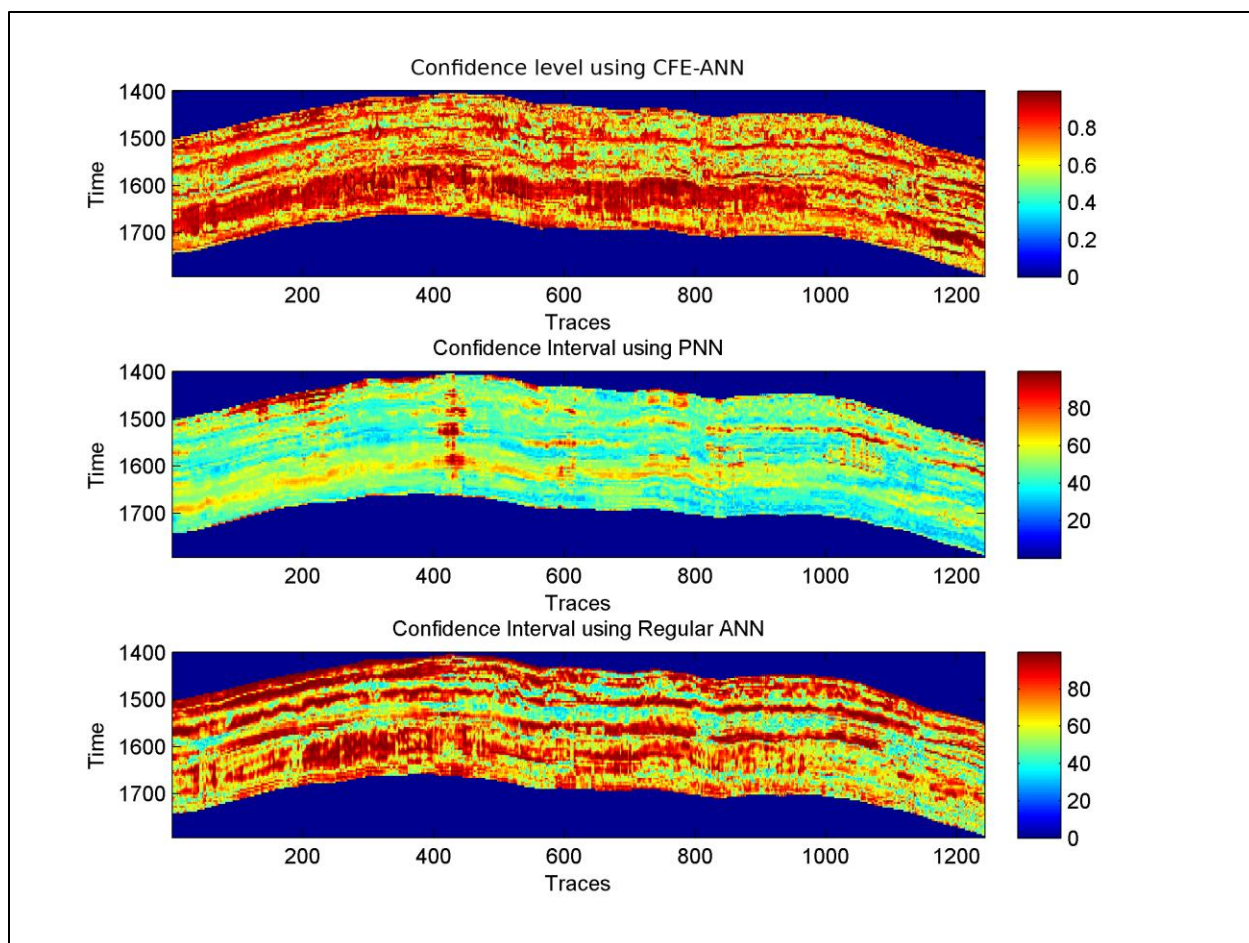


Figure 17: Confidence levels for the three classification workflows, CFE-ANN (top), PNN (middle), and Regular ANN (bottom).

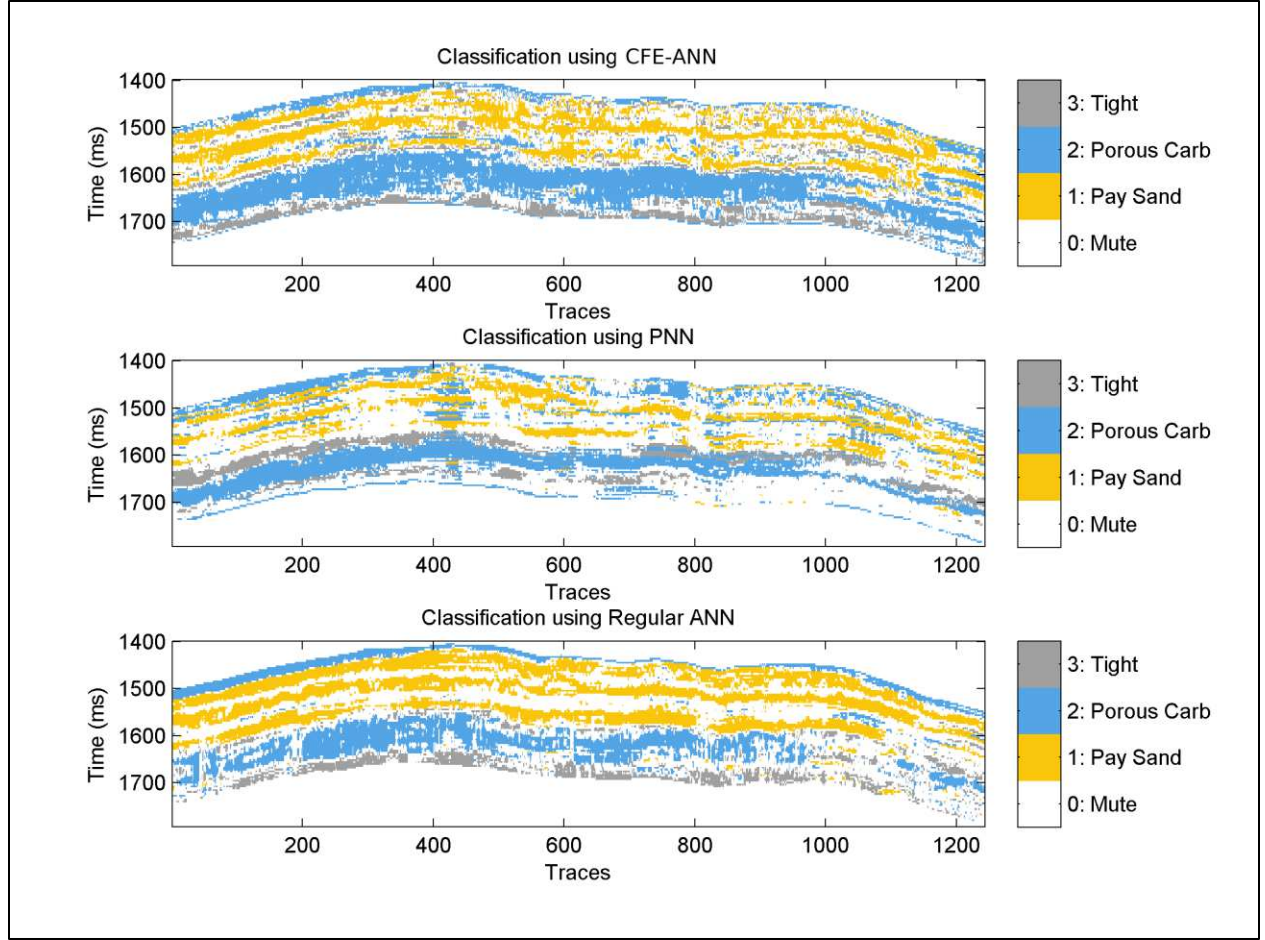


Figure 18: Filtered seismic facies classification using the confidence of the predicted class: CFE-ANN (top), PNN (middle), and Regular ANN (bottom). The threshold is 70% for the ANN and CFE-ANN, and 50% for the PNN. White areas (label 0, "mute") indicate samples below the confidence threshold.

Generally, the CFE-ANN image shows more laterally continuous layering compared to previous study results (Zahmatkesh et al., 2018, 2021). Based on prior stratigraphic knowledge, the study reservoir consists of interbedded sandstones and carbonates, and this pattern appears in the CFE-ANN results in the top half of the reservoir.

The conventional ANN and the CFE-ANN pipelines, including the deterministic litho-reservoir rock typing, the combinatorial attribute engineering, PCA compression, network training, and confidence estimation, were implemented in MATLAB and Python using custom code developed for this study. The PNN classification and the GRNN-based inversion and the

porosity estimation were performed in the Hampson–Russell software suite (Strata and EMERGE modules).

5. Conclusions

A case study of the highly heterogeneous Asmari Formation in the Mansuri oil field, SW Iran, was conducted to predict litho-reservoir rock types from seismic data. To address the limitations of conventional unsupervised clustering and raw seismic attributes, this study introduced a framework combining a deterministic rock-physics-driven target definition with a combinatorial multi-classifier feature-generation strategy. The principal conclusions are as follows:

First, classifying the well-log data into three petrophysically governed classes (Pay Sand, Porous Carbonate, and Tight Background) produced seismically consistent target labels. Consolidating compositionally distinct but acoustically similar non-reservoir lithologies into a unified background class mitigated the elastic ambiguity that typically causes unsupervised electrofacies schemes to define targets beyond seismic resolution.

Second, the proposed feature-engineering strategy (utilizing six binary class splits processed by four distinct classifier types to yield 24 engineered attributes) proved highly effective. Combined with PCA projection (retaining 95% cumulative variance), this ensemble-based feature space transformation substantially reduced the overfitting gap observed when using raw attribute sets and achieved superior blind-well generalization.

Third, evaluated on a fully blind well, the integrated CFE-ANN workflow (engineered features) clearly outperformed the conventional ANN (using raw features) and the commercial PNN baselines. The proposed framework attained a chance-corrected agreement (Cohen's kappa) in the substantial range and maintained a balanced precision–recall trade-off across all three lithofacies, effectively overcoming the severe class-imbalance failures of the benchmark models.

Fourth, applied along the studied seismic line, the CFE-ANN produced laterally continuous, geologically plausible facies architectures consistent with the known interbedded

sandstone–carbonate character of the reservoir. Furthermore, the introduction of confidence-filtered sections provided an explicit, critically needed measure of spatial prediction reliability.

Finally, while the proposed workflow is robust and portable, several limitations warrant consideration for future research. The use of strict Boolean cut-offs for rock typing imposes crisp boundaries on continuous geological variations, potentially introducing artificial class separations near threshold margins. Additionally, treating inverted acoustic impedance and ML-derived porosity as error-free inputs may propagate and compound their uncertainties. Future studies could replace the fixed cut-offs with fuzzy membership functions, allowing samples near a threshold partial membership in adjacent classes, or with probabilistic thresholds that propagate cut-off uncertainty into class probabilities at the wells. Such gradational labels would integrate naturally with the confidence-based framework proposed here. The scalability of the multi-classifier scheme should be assessed on full 3D seismic volumes.

Funding

The authors declare that no funds, grants, or other support were received during the preparation of this manuscript.

Competing interests

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.

Author contributions

Sam Azarpour: Conceptualization, Methodology, Software (MATLAB, Python, HRS), Formal analysis, Investigation (seismic interpretation and analysis), Writing (Original Draft) Writing (Review & Editing), Project administration. Performed the CFE-ANN, ANN, and PNN classifications and generated the final results; finalized and submitted the manuscript.

Igor Morozov: Methodology (feature engineering), Supervision. Supervised S. Azarpour's Master's thesis at the University of Saskatchewan, from which this study originated.

Iman Zahmatkesh: Investigation (petrophysical analysis), Writing (Original Draft), Writing (Review & Editing).

Mahmoud Poladzadeh: Validation (cross-checked findings against recent reservoir studies of the Mansuri field at NIOC), Writing (Review & Editing). Reviewed the full manuscript, including for compliance with applicable Iranian national regulations.

All authors reviewed and approved the final manuscript.

Data availability

The seismic and well-log data analysed in this study are proprietary to the National Iranian Oil Company (NIOC) and are not publicly available due to confidentiality restrictions. Access is governed by NIOC data-release policy, and the authors are not permitted to distribute the data. All information necessary to evaluate the conclusions of this study is presented in the paper.

References

- Amendola, A., Gabbriellini, G., Dell'Aversana, P., Marini, A.J., 2017. Seismic facies analysis through musical attributes. *Geophys. Prospect.* 65 (1), 49–58. <https://doi.org/10.1111/1365-2478.12504>
- Anifowose, F.A., Labadin, J., Abdulraheem, A., 2017. Ensemble machine learning: An untapped modeling paradigm for petroleum reservoir characterization. *J. Pet. Sci. Eng.* 151, 480–487. <https://doi.org/10.1016/j.petrol.2017.01.024>
- Ashraf, U., Zhu, P., Yasin, Q., Anees, A., Imraz, M., Mangi, H.N., Shakeel, S., 2019. Classification of reservoir facies using well log and 3D seismic attributes for prospect evaluation and field development: A case study of Sawan gas field, Pakistan. *J. Pet. Sci. Eng.* 175, 338–351. <https://doi.org/10.1016/j.petrol.2018.12.060>
- Baaske, U.P., Mutti, M., Baioni, F., Bertozzi, G., Naini, M.A., 2007. Using multi-attribute neural networks classification for seismic carbonate facies mapping: A workflow example from mid-Cretaceous Persian Gulf deposits. In: Davies, R.J., Posamentier, H.W., Wood, L.J., Cartwright, J.A. (Eds.), *Seismic Geomorphology: Applications to Hydrocarbon Exploration and Production*. Geol. Soc. Lond. Spec. Publ. 277, 105–120. <https://doi.org/10.1144/GSL.SP.2007.277.01.06>
- Babikir, I., Abdul Latiff, A.H., Elsaadany, M., Pratama, H., Sajid, M., Mad Sahad, S., Anwar Ishak, M., Laudon, C., 2024. Enhancing machine learning-based seismic facies classification through attribute selection: application to 3D seismic data from the Malay and Sabah Basins, offshore Malaysia. *Geomech. Geophys. Geo-Energy Geo-Resour.* 10, 138. <https://doi.org/10.1007/s40948-024-00846-x>
- Bagheri, M., Riahi, M.A., 2015. Seismic facies analysis from well logs based on supervised classification scheme with different machine learning techniques. *Arab. J. Geosci.* 8 (9), 7153–7161. <https://doi.org/10.1007/s12517-014-1691-5>
- Breiman, L., 2001. Random forests. *Mach. Learn.* 45 (1), 5–32. <https://doi.org/10.1023/A:1010933404324>
- Coléou, T., Poupon, M., Azbel, K., 2003. Unsupervised seismic facies classification: A review and comparison of techniques and implementation. *Lead. Edge* 22 (10), 942–953. <https://doi.org/10.1190/1.1623635>
- Cortes, C., Vapnik, V., 1995. Support-vector networks. *Mach. Learn.* 20 (3), 273–297. <https://doi.org/10.1007/BF00994018>

- Dietterich, T.G., Bakiri, G., 1995. Solving multiclass learning problems via error-correcting output codes. *J. Artif. Intell. Res.* 2, 263–286. <https://doi.org/10.1613/jair.105>
- Du, H., Cao, J., Xue, Y., Wang, X., 2015. Seismic facies analysis based on self-organizing map and empirical mode decomposition. *J. Appl. Geophys.* 112, 52–61. <https://doi.org/10.1016/j.jappgeo.2014.11.007>
- Farzadi, P., 2006. Seismic facies analysis based on 3D multi-attribute volume classification, Dariyan Formation, SE Persian Gulf. *J. Pet. Geol.* 29 (2), 159–173.
- Feng, R., Luthi, S.M., Gisolf, D., Angerer, E., 2018a. Reservoir lithology classification based on seismic inversion results by hidden Markov models: Applying prior geological information. *Mar. Pet. Geol.* 93, 218–229. <https://doi.org/10.1016/j.marpetgeo.2018.03.004>
- Feng, R., Luthi, S.M., Gisolf, D., Angerer, E., 2018b. Reservoir lithology determination by hidden Markov random fields based on a Gaussian mixture model. *IEEE Trans. Geosci. Remote Sens.* 56 (11), 6663–6673. <https://doi.org/10.1109/TGRS.2018.2841059>
- Hampson, D.P., Schuelke, J.S., Quirein, J.A., 2001. Use of multiattribute transforms to predict log properties from seismic data. *Geophysics* 66 (1), 220–236. <https://doi.org/10.1190/1.1444899>
- Haykin, S., 2009. *Neural Networks and Learning Machines*, third ed. Pearson Prentice Hall, Upper Saddle River, NJ.
- Kim, Y., Hardisty, R., Marfurt, K.J., 2019. Attribute selection in seismic facies classification: Application to a Gulf of Mexico 3D seismic survey and the Barnett Shale. *Interpretation* 7 (3), SE281–SE297. <https://doi.org/10.1190/INT-2018-0246.1>
- Landis, J.R., Koch, G.G., 1977. The measurement of observer agreement for categorical data. *Biometrics* 33 (1), 159–174. <https://doi.org/10.2307/2529310>
- Linari, V., Santiago, M., Pastore, C., Azbel, K., Poupon, M., 2003. Seismic facies analysis based on 3D multiattribute volume classification, La Palma Field, Maracaibo, Venezuela. *Lead. Edge* 22 (1), 32–36. <https://doi.org/10.1190/1.1542753>
- Liu, J., Dai, X., Gan, L., Liu, L., Lu, W., 2018. Supervised seismic facies analysis based on image segmentation. *Geophysics* 83 (2), O25–O30. <https://doi.org/10.1190/geo2017-0223.1>
- Marroquín, I.D., 2014. A knowledge-integration framework for interpreting seismic facies. *Interpretation* 2 (1), SA1–SA9. <https://doi.org/10.1190/INT-2013-0072.1>
- Qi, J., Zhang, B., Lyu, B., Marfurt, K., 2020. Seismic attribute selection for machine-learning-based facies analysis. *Geophysics* 85 (2), O17–O35. <https://doi.org/10.1190/geo2019-0223.1>
- Riedel, M., Bahk, J.-J., Kim, H.-S., Scholz, N.A., Yoo, D.-G., Kim, W.-S., Ryu, B.-J., Lee, S.R., 2013a. Seismic facies analyses as aid in regional gas hydrate assessments. Part-I: Classification analyses. *Mar. Pet. Geol.* 47, 248–268. <https://doi.org/10.1016/j.marpetgeo.2013.04.011>
- Riedel, M., Bahk, J.-J., Scholz, N.A., Ryu, B.-J., Yoo, D.-G., Kim, W.-S., Kim, G.Y., 2013b. Seismic facies analyses as aid in regional gas hydrate assessments. Part-II: Prediction of reservoir properties, gas hydrate petroleum system analysis, and Monte Carlo simulation. *Mar. Pet. Geol.* 47, 269–290. <https://doi.org/10.1016/j.marpetgeo.2013.04.012>
- Roden, R., Smith, T., Sacrey, D., 2015. Geologic pattern recognition from seismic attributes: Principal component analysis and self-organizing maps. *Interpretation* 3 (4), SAE59–SAE83. <https://doi.org/10.1190/INT-2015-0037.1>
- Roy, A., Dowdell, B.L., Marfurt, K.J., 2013. Characterizing a Mississippian tripolitic chert reservoir using 3D unsupervised and supervised multiattribute seismic facies analysis: An example from Osage County, Oklahoma. *Interpretation* 1 (2), SB109–SB124. <https://doi.org/10.1190/INT-2013-0077.1>
- Saggaf, M.M., Toksöz, M.N., Marhoon, M.I., 2003. Seismic facies classification and identification by competitive neural networks. *Geophysics* 68 (6), 1984–1999. <https://doi.org/10.1190/1.1635052>
- Sagi, O., Rokach, L., 2018. Ensemble learning: A survey. *WIREs Data Min. Knowl. Discov.* 8 (4), e1249. <https://doi.org/10.1002/widm.1249>
- Sayago, J., Di Lucia, M., Mutti, M., Cotti, A., Sitta, A., Broberg, K., Przybylo, A., Buonaguro, R., Zimina, O., 2012. Characterization of a deeply buried paleokarst terrain in the Loppa High using core data and multiattribute seismic facies classification. *AAPG Bull.* 96 (10), 1843–1866. <https://doi.org/10.1306/02271211137>
- Wolpert, D.H., 1992. Stacked generalization. *Neural Netw.* 5 (2), 241–259. [https://doi.org/10.1016/S0893-6080\(05\)80023-1](https://doi.org/10.1016/S0893-6080(05)80023-1)
- Wrona, T., Pan, I., Gawthorpe, R.L., Fossen, H., 2018. Seismic facies analysis using machine learning. *Geophysics* 83 (5), O83–O95. <https://doi.org/10.1190/geo2017-0595.1>

- Zahmatkesh, I., Soleimani, B., Kadkhodaie, A., Golalzadeh, A., Abdollahi, A.A.M., 2017. Estimation of DSI log parameters from conventional well log data using a hybrid particle swarm optimization–adaptive neuro-fuzzy inference system. *J. Petrol. Sci. Eng.* 157, 842–859. <https://doi.org/10.1016/j.petrol.2017.08.002>
- Zahmatkesh, I., Kadkhodaie, A., Soleimani, B., Golalzadeh, A., Azarpour, M., 2018. Estimating V_{sand} and reservoir properties from seismic attributes and acoustic impedance inversion: A case study from the Mansuri oilfield, SW Iran. *J. Pet. Sci. Eng.* 161, 259–274. <https://doi.org/10.1016/j.petrol.2017.11.060>
- Zahmatkesh, I., Kadkhodaie, A., Soleimani, B., Azarpour, M., 2021. Integration of well log-derived facies and 3D seismic attributes for seismic facies mapping: A case study from Mansuri oil field, SW Iran. *J. Pet. Sci. Eng.* 202, 108563. <https://doi.org/10.1016/j.petrol.2021.108563>
- Zahmatkesh, I., Kadkhodaie, A., 2025. Addressing the heterogeneity of mixed clastic-carbonate Asmari formation through seismic-facies aided electrofacies modeling in a Southern Iran oilfield. *Carbonates Evaporites* 40 (4), 154. <https://doi.org/10.1007/s13146-025-01191-0>
- Zahraa, A., Zailani, A., Ghosh, D.P., 2017. Characterizing geological facies using seismic waveform classification in Sarawak Basin. *IOP Conf. Ser.: Earth Environ. Sci.* 88, 012001. <https://doi.org/10.1088/1755-1315/88/1/012001>
- Zhang, S.-H., Xu, Y., Abu-Ali, M., Teng, M.K., 2019. Seismic facies recognition and stratigraphic trap characterization based on neural networks. In: *International Petroleum Technology Conference*. International Petroleum Technology Conference.
- Zhao, T., Jayaram, V., Roy, A., Marfurt, K.J., 2015. A comparison of classification techniques for seismic facies recognition. *Interpretation* 3 (4), SAE29–SAE58. <https://doi.org/10.1190/INT-2015-0044.1>
- Zhao, T., Li, F., Marfurt, K.J., 2018. Seismic attribute selection for unsupervised seismic facies analysis using user-guided data-adaptive weights. *Geophysics* 83 (2), O31–O44. <https://doi.org/10.1190/geo2017-0192.1>