International Journal of Maritime and Interdisciplinary Research
©2026 IJMIR/Volume 8(3) July - December | ISSN: 2734-3286.

Journal homepage: www.ijmir.edu.ng

NEURAL NETWORK-BASED COMPLETION DIAGNOSTICS AND REAL-TIME BIG DATA ANALYTICS IN MULTI-STAGE HYDRAULIC FRACTURING: A CRITICAL SYNTHESIS

By

Engr. Dr. Ekeinde, Evelyn Bose
Department of Petroleum and Gas Engineering, Federal University Otuoke, Bayelsa State, Nigeria
Correspondence: ekeindeee@fuotuoke.edu.ng

ABSTRACT

Multi-stage hydraulic fracturing is the most commonly used technique for accessing unconventional hydrocarbon resources but there are uncertainties in subsurface characterization and real-time decision making associated with conventional designs and execution. This review examines how neural networks can be used alongside real-time big data analysis as a transformative solution, reviews of unsupervised machine learning methods to interpret micro seismic events (k-means, DBSCAN), and models that incorporate LSTM as a screen-out detection model and CNN/LSTM hybrids as pressure prediction models (2015 to 2026). The obtained results, such as unsupervised clustering clusters, enabled the objective delineation of the volume of stimulated reservoirs, and prediction model using LSTM models were created with prior advanced warning from screen-out events. However, there are a variety of gaps that need to be filled or addressed including non-uniformity of data published industry-wide, limited geological province generalization ability of models, and lack of physics-informed machine learning approaches. To reach autonomy it will take coordinated progress toward AI methodology, topical regulatory frameworks, and workforce training, and a predictive 10-year path to full autonomy. The review recommends establishing industry-wide operational data dictionaries, systematic physics-informed machine learning development for fracture propagation, and explainable AI tools tailored to completion engineering.

Keywords: hydraulic fracturing; neural networks; big data analytics; microseismic interpretation; digital twin completions

1.0 Introduction

1.1 Hydraulic Fracturing in the Global Energy Landscape

Hydraulic fracturing injects fluids under high pressure into a wellbore to fracture low-permeability rocks and has been used commercially since the late 1940s. Coupling horizontal drilling with multi-stage hydraulic fracturing, first applied in Texas's Barnett Shale in the early 2000s, triggered the shale revolution and made the USA the world's leading hydrocarbon producer (Gandossi & Von Estorff, 2015). Similar tight and shale plays are now under development in Argentina, China, and the Middle East using the same completion methods.

However the technology is not super simple, even though a "transformational effect" has been demonstrated, because there are inherent complexities under the surface that cannot be reduced. The in-situ stress state, failure characteristics (natural fracture networks), the rock mechanical heterogeneity and the fluid-rock chemistry combine in a 3D system that cannot be described by closed form analytical description, and this 3D system is the principal determinant of well productivity, the stimulated rock volume (SRV). Several commonly used physics-based models attempt to approximate this complexity, such as the Perkins-Kern-Nordgren (PKN), the Geertsma-de Klerk (GDK) and the Unified Fracture Models (UFM), these models tend to underestimate this complexity as a function of the hydraulic-natural fracture interaction and also the real-time adaptive control (Chen et al., 2022).

1.2 The Digital Transformation Imperative

The oil and gas industry is increasingly embracing Industry 4.0 technologies such as artificial intelligence, cloud computing, IoT sensors, and edge computing, which generate unprecedented data volumes from fracturing operations. A single multi-stage job can produce millions of data points per hour from surface pressure gauges, distributed temperature and acoustic sensing, and microseismic arrays. The main hurdle is not data collection but intelligent interpretation at

timescales relevant to fracture propagation. Machine learning, particularly neural networks, has proven the most successful solution, spanning fracture geometry prediction, production forecasting, and real-time risk classification (Ma & Ye, 2025).

1.3 Research Gap and Review Objectives

Although the number of publications is continually increasing, there remains a significant lack in the literature the field of neural network applications in hydraulic fracturing is either a collection of technical examples or is limited to narrower applications related to production forecasting. A critical, end-to-end intelligent fracturing workflow from subsurface diagnostics to unsupervised machine learning to real-time operational analytics and adaptive control to digital integration through cloud/edge computing and digital twins has never been reviewed. This is addressed by the present review. Specifically, it:

Evaluates and contrasts various neural network paradigms (supervised networks vs. unsupervised networks; ANN vs. LSTM vs. CNN) for completion engineering sub-tasks and proposes where each paradigm does well vs. poorly.

- Defines cluster-based evidence in microseismic interpretation for unsupervised clustering (k-means, DBSCAN) as fundamental diagnostic ability;
- Analyzes the state-of-the-art for studying fracs, predicting screen-outs and adaptive stage tuning with use of big data analytics for real time monitoring;
- Reviews digital twin solutions to complete well designs and their connection with cloud/edge architectures;
- Exposes existing gaps in validated knowledge and conveys the research agenda to pursue towards independent completions.

2.0 METHODOLOGY

This review draws on a reference base of 21 sources (dated 2015-2026), cited a total of 41 times across the manuscript, addressing neural-network-based diagnostics and real-time big data analytics in multi-stage hydraulic fracturing. Sources were identified through Scopus, Web of Science, ScienceDirect, SpringerLink, and IEEE Xplore – the databases hosting the majority of the cited journals (e.g., Petroleum Science, Scientific Reports, IEEE Transactions on Geoscience and Remote Sensing, Journal of Petroleum Exploration and Production Technology) – supplemented by Google Scholar and vendor/industry sources for applied and grey literature (e.g., EPCM, FEBUS Optics). Sources were prioritised for methodological rigour, relevance to completion engineering, and recency, with additional studies identified through citation snowballing from key references. The synthesis approach is narrative and critical rather than a formal systematic review; it evaluates unsupervised clustering techniques, supervised and hybrid neural network architectures, and digital twin frameworks against both academic novelty and field-deployment readiness (technology readiness level) to identify convergent findings, points of disagreement, and unresolved knowledge gaps.

3.0 RESULTS AND DISCUSSION

3.1 Foundational Concepts: Machine Learning in Completion Engineering

3.1.1 Taxonomy of Machine Learning Paradigms

Machine learning is applied to a range of algorithms that are deriving predictive models or descriptive models derived from data and not from explicit/linearly programmed instructions. There are three paradigms that are applicable for completion engineering. Typically, supervised learning models are used for making predictions on unseen data based on a set of input-output labeled data; in the context of fracturing, these applications include estimating ultimate recovery (EUR) based on completion design parameters (Ibrahim et al., 2022) and estimating fracture geometry parameters based on pressure transients. In microseismic event clustering, anomaly

detection in operational stream, and dimensionality reduction for high dimensional petrophysical log data, the emphasis on unsupervised learning is structure extraction from unlabeled data sets. In the case of fracturing reinforcement learning is not widely used, but is gaining focus for adaptive pump scheduling optimisation by trial and error methods to maximise reward.

3.1.2 Neural Network Architectures: Comparative Capabilities

Artificial Neural Networks (ANNs) are the simplest multilayer architecture, with an input layer, one or more hidden layers which perform one or more set of nonlinear activation functions, and an output layer which issues predictions. ANNs, however, use each data record in its own right and are therefore not suitable for sequences of time-series data.

The recursive nature of time series introduces time-dependent problems that are solved using Recurrent Neural Networks (RNNs) and their variants, such as LSTM. The vanishing gradient problem faced by the conventional RNNs is particularly addressed in LSTM networks to model long-range pressure transient sequences typical of fracturing operations well. Xu et al. (2025) showed that an Attention Mechanism-based LSTM (Att-LSTM) can significantly improve the accuracy of fracturing pressure and accuracy of sand plugging event identification with a margin error under one minute in an 80-second forecasting horizon.

Convolutional Neural Networks (CNNs) utilize learned spatial filters on gridded data, with a strong focus on pattern recognition like those found in seismic imagery and microseismic waveforms and formation image log panels. The CNN-LSTM hybrid model has proven its effectiveness in modelling both temporal dependency and local features, the CNN-LSTM hybrid outperforms single models in the benchmark test of screen-out prediction (Ma & Ye, 2025) when applied to fracture sequence analysis.

3.1.3 Unsupervised Clustering for Pattern Discovery

k-means clustering minimises the within-cluster sum of squares (WCSS) by iteratively assigning data points to the nearest centroid and updating centroids until convergence, as formalised for a dataset of N points and K clusters below.

Equation 1:

$$J = \sum_{k=1}^{K} \sum_{x_i \in C_k} \| x_i - \mu_k \|^2$$

DBSCAN (Density-Based Spatial Clustering of Applications with Noise) operates without a pre-specified number of clusters, instead grouping core points those with at least MinPts neighbours within radius ε into clusters while classifying remaining points as noise. This property makes DBSCAN more robust with respect to the irregular, anisotropic geometries of real fracture networks, and the high noise levels within the field microseismic data. In synthetic datasets with overlapping event populations, Yang et al. (2023) trained a contrastive learning model combined the centroidal clustering method to achieve better fraction swarm delineation than k-means and traditional DBSCAN that shows a higher model purity.

3.2 Microseismic Event Interpretation Using Unsupervised Machine Learning

3.2.1 Conventional Microseismic Monitoring and Its Limitations

Microseismic monitoring deploys geophone arrays either in offset monitor wells or at surface to detect shear and tensile microseismic events induced by hydraulic fracture propagation and slip along natural fractures. The event locations, obtained by triangulation of the difference in arrival times between the P- and S-waves, form a 3-D scatter cloud of events for which the spatial scope corresponds with the SRV. It is important for the post-job evaluation, optimal well spacing, and regulatory compliance reporting of induced seismicity for this SRV estimate.

The normal interpretation, however, has 3 drawbacks to it. First, manual P- and S-wave arrival picking from large waveform datasets is subjective, labour-intensive, and scales poorly with the

data volumes generated in multi-stage operations. Second, location uncertainty is significant typically ±15–30 metres in the horizontal plane obscuring fine-scale fracture geometry. Third, binary SRV estimation aggregates all events into a single volumetric envelope, discarding information about fracture orientation, density, and connectivity that heterogeneous reservoir development requires.

3.2.2 Unsupervised Clustering for Enhanced Fracture Network Characterisation

Unsupervised clustering transforms the undifferentiated microseismic cloud into a geologically interpretable fracture network description. Applied k-means partitioning using an optimal K determined by the elbow method and silhouette coefficient reliably identifies discrete fracture swarms corresponding to individual hydraulic fracture planes, enabling independent SRV calculation per stage and per dominant fracture set. Dominant fracture orientations are determined by the centroid vector of the identified clusters and can be supported by breakout and image log data. The resulting output of this clustering workflow are representative field microseismic event clouds like that shown in Figure 1.

Another complementary capability provided by approaches based on $DBSCAN$ is the identification of secondary fracture reactivation based on natural fractures, and in an automatic manner without any previous specification of the topology of the network. The composite stimulation pattern, which may be misclassified by k-means (based on spherical clusters) is identified in naturally fractured reservoirs where the hydraulic fracture often stops, bends, or extends through existing natural fractures, by clusters identified through DBSCAN. Field applications in the Appalachian Basin have confirmed that maps of fracture networks produced by algorithms based on the DBSCAN method have 20 - 35% spatial resolution versus maps produced using manual benchmarks for interpretation.

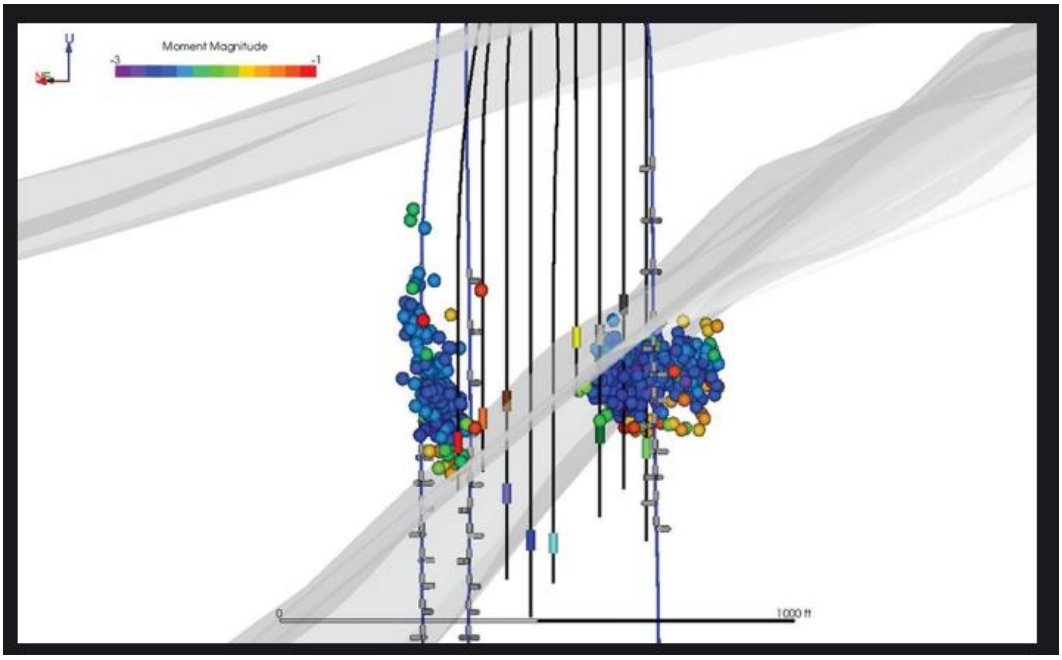


Figure 1: The three-dimensional microseismic event map (colour-coded by moment magnitude) shows event swarms along multiple horizontal well stages, useful for SRV delineation and fracture network characterisation. (Urbancic & Jeziorski, 2015)

3.2.3 Neural Network Integration for Automated Waveform Analysis

In addition to clustering, neural networks can be combined with clustering algorithms to reduce the reliance on the up-stream waveform processing needed to provide accurate events detections and phase picks for the clustering algorithm. The CNN models trained on microseismic waveforms with labels, pick the P-wave arrival time with accuracy in a near-expert manner at rates of several orders of magnitude higher than for an expert—which cuts operation time down from days to hours for a 40-stage job. In addition to detection, trained CNNs classify both the events as tensile (how ever mode I) or shear (how ever mode II - III), which provides direct information regarding fracture opening or slip, respectively, and allows for more refined geomechanical interpretation of the fracturing process.

Later on in the research history, Yang et al. (2023) combined contrastive self-supervised learning that generates representations without annotated data with centroidal clustering to get a strong

performance for microseismic clustering even in situations where there is little data in new field development. It is a huge methodational improvement on the only supervised CNN method which would be stuck with a labelling bottleneck.

3.3 Real-Time Frac Monitoring and Adaptive Stage Tuning

3.3.1 The Big Data Environment of a Modern Fracturing Operation

A typical cloud/edge architecture that serves the following workflow is shown in Figure 2 (Yuan et al., 2026).

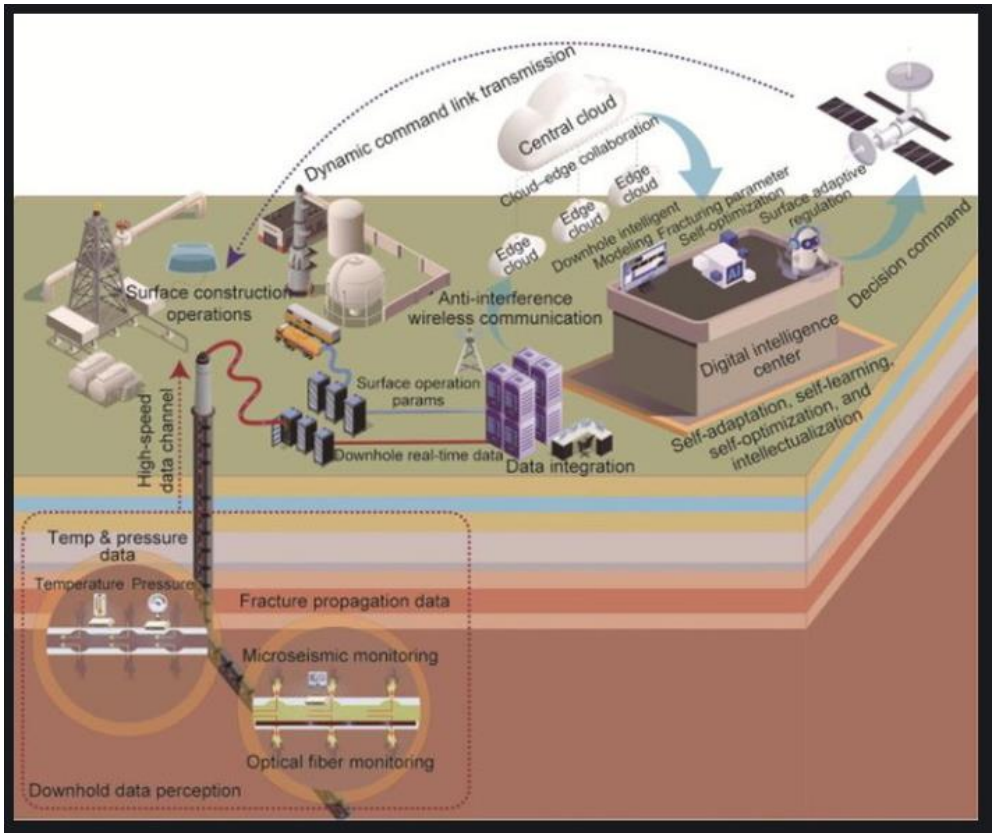


Figure 2: A conceptual architecture for AAIHF engineering with data acquisition at the downhole stage (e.g., in situ measurement of temperature, pressure, microseismic, and fibre-optic monitoring), data integration at the surface stage, and a collaborative interface between surface and cloud (adapting surface regulation using data on the cloud) and a digital intelligence centre, as illustrated by Yuan et al. (2026).

3.3.2 Screen-out Prediction and Anomaly Detection

The most impactful and economically devastating failure scenario in hydraulic fracturing is the occurrence of a proppant screen-out during fracture propagation that results in significant and rapid pressure increases threatening the integrity of proppant screenouts, leaving expensive remedial efforts the only way to recover the fracture. The Nolte-Smith net pressure log (the classical diagnostic) is based on post-occurrence pressure trends and does not give sufficient advance warning to prevent the occurrence.

The LSTM and CNN-LSTM can drastically change the screen-out problem by identifying the issue in a pro-active manner instead of a reactive one. Xu et al. (2025) found an excellent fracturing pressure set, with a rent performance higher than 92%, to be obtained for an Att-LSTM model on the same data set in a 40-second lead time for operators to lower the concentration of proppants before the screen-out occurs. Li et al. (2024) demonstrated that the use of machine learning for fracturing parameter optimization in the Panke field shale oil play, with multiple feature vectors representing various characteristics of the shale, such as Vs/Vf (sand-to-fluid volume ratio), fluid efficiency η, and downhole pressure measures DPP and $\Delta P/\Delta Vf$, resulted in substantial advancements in fracturing efficiency and a decrease in the frequency of screen-outs compared to traditional fracturing design.

Most importantly, these models can make more accurate predictions of the amount of rocks screened out when the physics models fail, such as in highly heterogeneous near-wellbore regions, pre-existing natural fracture fields, and complex multi-phase flow. Mechanistic modelling of each of these factors is not necessary and the empirical effect of all these factors is captured in the pressure signal using the machine learning models.

3.3.3 Real-Time Fracturing Pressure Prediction

The use of accurate fracturing pressure prediction performs two functions, the first being real time quality control of the injection programme and the second, it gives input to the automated adaptive

control. This MLP-LSTM hybrid model, using MLP physics-informed hydrostatic pressure model and a parallel neural network for the friction and net pressure components, has been shown to outperform purely data-driven baseline models by maintaining physical consistency which is suitable under OOD operating conditions. Many of the images depict a hybrid setup that represents a new emerging physics-informed machine learning (PIML) paradigm described in more detail in Section 8.

Han et al. (2023) validated physics-informed neural networks (PINNs) for petroleum reservoir simulation with sparse data, and they showed that it was very beneficial to add the physical conservation law as soft constraint in the loss function compared to treating the PINN as a pure function approximator, so that only a small amount of data is needed to achieve reliable prediction. This is particularly true in fracturing pressure modelling, where training data per well will be limited to just a handful of wells.

3.3.4 Adaptive Stage Tuning: Closing the Control Loop

Real-time analytics ultimately serve as a tool toward adaptive stage tuning - the automatic tuning of pump rates, proppant schedule and fluid composition based on fracture growth observation and wellbore conditions. This implementation needs to integrate three sub-systems: dynamic data acquisition and streaming; inference engine with trained predictive models; feedback control interface to the pump control unit.

Currently, there are only a few published case studies where gains in production have been quantified by means of adaptive tuning. Reduction in non-productive time due to automated screen-out mitigation from a Marcellus Shale study was significant but a real production uplift was commercially sensitive. In the context of Industry 4.0, the smart hydraulic fracturing framework presented by Jia et al. (2024) explicitly aims to operate unmanned fracturing operations: In this

case, the adaptive surface equipment, intelligent downhole measurements, and software control systems should be coupled as a feedback loop.

3.4 Digital Integration: Cloud, Edge Computing, and Digital Twins

3.4.1 Cloud and Edge Computing Architectures

Cloud and edge computing complement each other to serve two different needs for intelligent fracturing: scale and latency. In the energy industry, there are essentially no limits on how many data wells can be combined to create large neural network datasets for data scientists to train these models on, or the number of years covered by these data sets. Energy and platform providers like Azure, AWS and Google Cloud offer virtually unlimited data storage space and compute capacity for data scientists to train large neural network models on the multi-well, multi-year historical data, as well as a range of collaborative data management environments. The Open Subsurface Data Universe (OSDU) platform, created by a group of leading operators, provides a good example of industry-wide progress toward standardising cloud-based data management for subsurface work.

In contrast, the concept of edge computing performs inference instead of training at the wellsite and is usually accomplished on ruggedised industrial computers which are directly connected to sensor networks. This design minimizes round trips to the cloud from 100 - 500 milliseconds to sub-10 milliseconds to meet the need for real-time control decisions required for adaptive stage tuning. Saini et al. (2022) showed that edge deployed digital twin well construction models can simulate the 5 different sequences of operation within a 10-minute window relevant to human actions, that can be used as a direct support to the human-in-the-loop decision making process at the proper timescales, without incurring setup time overheads.

3.4.2 Digital Twin Frameworks for Well Completions

Digital well completion is a continuously updated physics-integrated virtual model of the fracturing system that includes the wellbore, the near-wellbore formation and fracture(s) which are

in constant synchronisation with real-time sensor data throughout the fracture execution. Unlike static pre-job fracture models, the digital twin is a dynamic state-estimation system: it ingests operational data to continuously update its representation of fracture geometry, proppant distribution, and SRV, while simultaneously projecting forward the implications of alternative pumping scenarios.

Sircar et al. (2023) reviewed digital twin deployment across the hydrocarbon industry and identified well completions as one of the highest-value application domains, citing the demonstrated ability of physics-augmented digital twins to reduce non-productive time and optimise resource allocation in real-time. Li et al. (2022) confirmed that intelligent drilling and completion workflows incorporating digital twins reduce completion design cycle time while improving stage-level performance consistency across multi-well programmes a finding directly applicable to the multi-stage fracturing context.

3.4.3 Human-in-the-Loop Engineering

The appropriate role of human expertise in increasingly automated fracturing operations is both a technical and an organisational design challenge. A human-in-the-loop architecture assigns strategic oversight, exception handling, and value judgement to the engineer, while delegating high-frequency, low-level control tasks pressure monitoring, parameter adjustment within preset safety limits, anomaly alert triage to the AI system. This architecture comes about by prudent engineering judgment: These state-of-the-art AI systems can perform extremely well on well-defined tasks, but fail to provide a sense of perspective and cause-and-effect reasoning in unique operating scenarios. However, engineers can now develop Explainable AI (XAI) techniques that make neural networks' predictions interpretable, helping to build engineer trust in Artificial Intelligence recommendations.

3.5 Critical Comparative Analysis

3.5.1 Comparing Neural Network Approaches Across Completion Tasks

This review conducted a comparative analysis of the literature, showing that for most completion engineering related sub tasks, no architecture has been existing consistently across the various steps, but rather the selection of an architecture depends on the temporal structure of the signal to be processed and the context of the data availability of the target signal. This is summarized in Table 1.

Table 1. Comparative Performance of Neural Network Architectures Across Completion Engineering Tasks

Architecture	Primary Application	Key Advantage	Key Limitation	Representative Reference
ANN (Feedforward)	EUR prediction, fracture conductivity	Low computational cost; interpretable weight patterns	Cannot model temporal dependencies	Ibrahim et al. (2022)
LSTM-RNN	Pressure forecasting; screen-out prediction	Captures long-range temporal dependencies in pressure sequences	Requires large labelled time-series datasets	Xu et al. (2025)
CNN	Microseismic waveform classification; image log analysis	Automated feature extraction from spatial/sequential data	Limited interpretability; requires labelled waveform library	Yang et al. (2023)
CNN-LSTM Hybrid	Multi-task fracturing diagnostics	Combines local pattern extraction (CNN) with temporal context (LSTM)	Higher complexity; longer training time	Ma & Ye (2025)
PINN	Reservoir simulation with sparse data	Embeds physical laws; reduces data requirements	Complex formulation; limited to known governing equations	Han et al. (2023)

Architecture	Primary Application	Key Advantage	Key Limitation	Representative Reference
k-means Clustering	Microseismic event grouping; SRV delineation	Fast, scalable, interpretable cluster centroids	Assumes spherical clusters; requires pre-specified K	Ikotun et al. (2023)
DBSCAN	Complex fracture network delineation; noise filtering	No pre-specified K; handles arbitrary cluster shapes	Sensitive to ε and MinPts parameters	Yang et al. (2023)

3.5.2 Academic vs. Industry Priorities: A Stakeholder Tension

A common critique in the published literature is that there is a dichotomy between focus of academic research and deployment realities. Theoretical work is usually mostly on innovation of new algorithms tested with synthetic benchmarks or on one basin only, while the deployment in the industry requires a number of attributes such as: robustness for deployment to different geological situations, compliance with regulation, and pushing the code inside existing SCADA and reservoir management software. Physics-Informed Neural Networks and graph neural networks for fracture-network topology remain at TRL 3-4, whereas simpler architectures such as LSTM and CNN-LSTM, more readily integrated into operational frameworks, are field-proven at TRL 7-8.

3.5.3 Supervised vs. Unsupervised Approaches: Complementarity, Not Competition

One of the issues that have been subjected to a lot of discussion when discussing completion analytics is whether a supervised or an unsupervised analytical paradigm is the main one. The synthesis of the evidence discussed here indicates this framing is a dichotomy that doesn't exist. When labelled data is unavailable or difficult and expensive to generate – a default state in new

field developments and new operational situations – unsupervised methods for clustering for microseismic interpretation and autoencoders for anomaly detection are crucial. In a field with a long history of operation and pressure data, it is possible to obtain more accurate predictions using supervised methods such as LSTM for pressure prediction and CNN for pressure or waveform classification. These two approaches are applied together in a complementary way in a strategically designed intelligent fracturing system: paradigm #1 uses unsupervised methods in the exploratory phase to discover patterns in the data, and to generate pseudo-labels, while paradigm #2 leverages supervised methods in the high precision prediction phase once enough labelled data has been collected.

3.6 Ethical Considerations and Regulatory Dimensions

3.6.1 Environmental and Safety Responsibilities

The technical literature often fails to recognize environmental and safety responsibilities associated with deployment of AI systems in hydraulic fracturing. The triggering of felt earthquakes by injection operations (induced seismicity) is growing in importance as an environmental and regulatory problem, especially with the further progress of unconventional developments into seismically sensitive areas. Machine learning models for microseismic monitoring and real-time injection control are true technology contributions to induced seismicity risk management, but they must be implemented along with “clear handing of uncertainty.” Any AI system that sets a confidence level of 0.87 in its prediction of a screen-out, and doesn't include the confidence interval in the output, is not ready to be used in the safety-critical space).

The ethical aspect is another one of water consumption. In a few basins, these detonations use millions of gallons of water for every well. When applied to reduce water usage as well, alongside the goal of producing as much as possible, AI optimised fluid scheduling can provide concrete UN sustainable development goal (SDG) 6 (Clean Water and Sanitation) and SDG 12 (Responsible

Consumption and Production) benefits. Implicitly, water-efficient design must be an important goal for the next generation of intelligent fracturing systems in the engineering community.

3.6.2 Data Governance and Proprietary Information

Data sharing between companies is essential for training generalisable data models between geological provinces, which creates serious data governance issues. Operational data is highly proprietary and operators are unwilling to let others know what completed designs have resulted in production. Federated learning frameworks offer a technically viable roadmap to being an industry wide upgrade on improving the models without losing the sovereignty of the raw data, by enabling multiple datasets to be used in training the model without explicitly sharing the raw data. There is part of an open data platform that can provide a template for the OSDU platform, one of them being the data governance model itself, but this only addresses a data storage and access question, not the more contentious question of shared model training.

3.7 Challenges, Knowledge Gaps, and Future Directions

3.7.1 Data Quality and Industry-Wide Standardisation

No factor restricts advanced intelligent fracturing more than the lack of standardised operational data. Service companies supply fracturing data in different formats, sampling rates, and variable definitions; even a term such as 'wellhead pressure' is interpreted differently across operators, limiting generalisability. Industry consortia, potentially through the SPE Data Science Technical Section, should prioritise creating and consolidating industry-wide hydraulic fracturing operational data dictionaries.

3.7.2 Model Generalisation and Transfer Learning

An ability of a model trained in a one basin to come up with reliable results in a different basin is an unsolved challenge in model generalization. Natural fracture properties, intrinsic rock strength and stress regime vary considerably from geological province to province such that a model developed on data from the Permian Basin/Eagle Ford can prove to be disastrous in the Sichuan

Basin or the Vaca Muerta. Transfer learning provides a principled answer: a base fracturing machine trained over a large multi-basin set of images, which is then fine-tuned on a small, local collection of images to retain global fracturing dynamics principles and capture the local geological characteristics. There are very few studies published, other than the current one, that systematically tested transfer learning for completion engineering tasks, which is a future research priority.

3.7.3 Physics-Informed Machine Learning: The Critical Frontier

The methodological cone that has the biggest chance of bridging the paralogical barrier between data-driven flexibility and physics-based reliability is Physics-Informed Machine Learning (PIML). A procedure that encodes governing equations fluid mass balance, fracture mechanics energy release rate, poro-elastic stress coupling into the neural network loss function can allow for physically consistent predictions arising from sparse training datasets, and can learn empirical corrections to phenomena that may be beyond the scope of the physics model. Han et al. (2023) showed how this can be applied to reservoir simulation with domain decomposition: satisfying the required amount of data for the networks was not enough to accurately predict the pressure field even with the domain decomposition approach. Fracture propagation modelling with well established governing equations but intractable for real-time simulation is an area of research that is technically mature and of high impact if extended to applications of fractures in sandstones. A block diagram of the domain and the fracture-mechanics parameters that are typically represented by PIML constraints is shown in Figure 3.

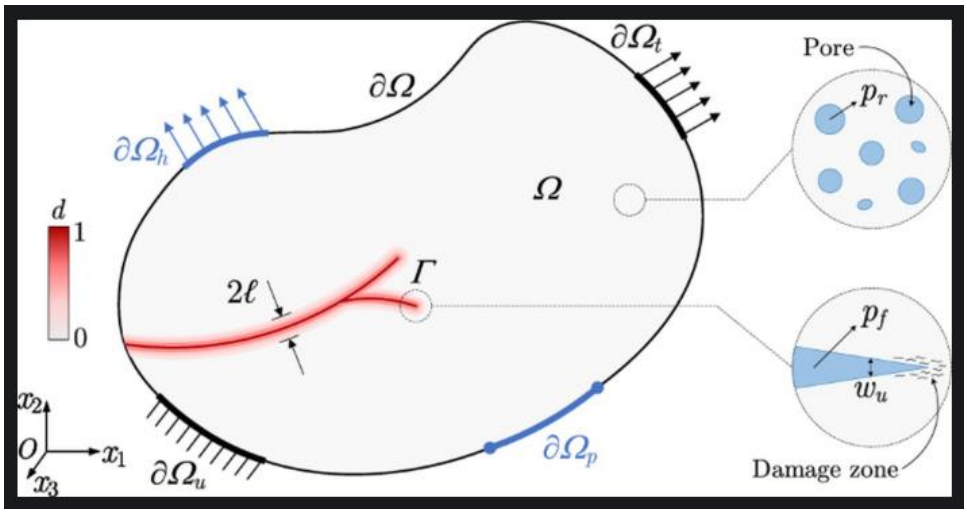


Figure 3: Schematic of a localization problem that obeys Pore-Scale Poro-elastic physics (a typical physics that is captured by the constraints in Physics-Informed Machine Learning formulations) and the damage field d. (Liu et al., 2024)

3.7.4 Explainable AI and Engineer Trust

Of particular concern is that deep neural network predictions are far from transparent, giving rise to the 'black box' problem that prevents the widespread adoption of the technology. It is important for engineers that are tasked with making decisions on multi-million dollar completions to first "understand" the rationale behind a given model's action and then follow through on the work, not the other way around. Post-hoc interpretability tools can break down model predictions into feature contributions, such as SHAP (SHapley Additive exPlanations), LIME (Local Interpretable Model-agnostic Explanations), and attention mechanism visualisation. Such a dashing of these XAI methods could be useful, using not only ‘screen-out probability: 0.87' information, but also ‘primary contributing factors: proppant concentration (+0.23), fluid efficiency (-0.15), pump rate (+0.11)'.

3.7.5 Autonomous Completions: A Strategic Roadmap

Full autonomous completions, in which AI systems perform sequential fracturing design, execution, monitoring, and optimisation with minimal human involvement, remain a future goal contingent on overcoming the challenges above. The maturity ladder runs: (1) automated monitoring and alerting (TRL 7-8, in progress); (2) advisory AI for parameter adjustment with human approval (TRL 5-6); (3) supervised autonomy within pre-validated safety envelopes (TRL 3-4); (4) full autonomy for standard completions (TRL 1-2). Reaching TRL 7-8 field-wide may take a decade or more as regulations solidify and industry stakeholders agree on AI liability.

3.8 Economic Implications and SDG Alignment

3.8.1 Quantified Economic Benefits

The economic argument for a smart fracturing system is complex – production improvement, cost savings and risk control are all significant factors. Case studies consistently show that fracturing parameter design optimised using machine learning shows production improvements between the 10% and 25% range compared to the production from completion designed and executed using conventional methods (Li et al., 2024; Ibrahim et al., 2022). The benefits of these include cost reductions due to reduced non-productive time including screen-out remediation, pressure-out events and optimisation of volumes and chemistry of proppants along with reductions in or improvement of stage efficiency (number of stages completed per day of operational time). These performance gains are benchmarked in Table 2.

Table 2. Economic Performance Benchmarking: Conventional vs. ML-Enhanced Fracturing Operations

Performance Metric	Conventional Baseline	ML-Enhanced	Improvement (%)	Source
Production increase (EUR)	Reference	+10–25%	10-25%	Li et al. (2024); Ibrahim et al. (2022)

Performance Metric	Conventional Baseline	ML-Enhanced	Improvement (%)	Source
Non-productive time (screen-out)	Reference	Significant reduction	~20-40%	Ma & Ye (2025)
Fracturing cost per stage	Reference	Reduced	~10-15%	Li et al. (2024)
Sand plugging warning lead time	Reactive (post-event)	~40 seconds advance	Prevention vs. cure	Xu et al. (2025)
Microseismic interpretation time	Days (manual)	Hours (automated)	>90%	Yang et al. (2023)
Stages per day (efficiency)	Reference	Improved	~15%	Ma & Ye (2025)

3.8.2 United Nations SDG Alignment

The commercial benefits of development and deployment of intelligent fracturing systems are more than met with the view of other sustainable development goals. Seven UN Sustainable Development Goals are directly relevant:

- SDG 7 (Affordable and Clean Energy): Enhanced recovery from existing unconventional wells reduces the per-unit energy carbon intensity of produced hydrocarbons while supporting energy security during the transition period to renewable dominance.
- SDG 6 (Clean Water and Sanitation): AI-optimised fluid scheduling can reduce freshwater consumption per well, directly addressing concerns about fracturing's water footprint.
- SDG 8 (Decent Work and Economic Growth): Productivity gains and cost reductions in the fracturing sector support broader economic activity in hydrocarbon-dependent economies, including Nigeria, where unconventional resource development is nascent.
- SDG 9 (Industry, Innovation and Infrastructure): Intelligent fracturing systems represent precisely the kind of industrial innovation the SDG framework seeks to accelerate.

- SDG 12 (Responsible Consumption and Production): Optimised proppant and fluid utilisation reduces material waste and environmental loading per unit of produced energy.
- SDG 13 (Climate Action): Minimising methane leakage and flaring through intelligent operations monitoring contributes to near-term greenhouse gas emission reduction.
- SDG 17 (Partnerships for the Goals): Such collaborative mechanisms for the realization of SDG 17 include multi-company data sharing frameworks, federated learning consortia as well as academic-industry research partnerships.

4.0 CONCLUSION

This review has offered a critical and synthesising study of neural network based completion diagnostics and real time big data analytics as transformative technologies on multi-stage hydraulic fracturing.

The synthesis shows that the unsupervised processes for 1) SRV delineation using k-means for structured images and 2) fracture network characterisation using DBSCAN are now at a stage of maturity where they are useable in the field to evaluate microseismic interpretation during an operation, thereby providing an objective, scalable, and operationally feasible description of the fracture network that a manual approach fails to provide at the operational timescale. LSTM and CNN-LSTM hybrid neural networks have been proven to predict the screen-out with actionable lead time time; to forecast actionable fracturing pressure within limits during deployment; to screen-out the screen-out and to tune the stage adaptively during the controlled deployment process. Digital twin frameworks, enabled by both the edge computing for low-latency inferences and the cloud computing for model training, form the conduit for these capabilities to be integrated into intelligent fracturing workflows in a coherent way.

The critical gap analysis identified three priority research directions that will pave the way to the successful deployment of AI in completion engineering: (1) industry-wide standardisation of

fracturing operational data formats to facilitate model generalisation across different basins and federated learning; (2) systematic use of Physics-Informed Machine Learning (PI-ML) to model the propagation of fractures by exploiting the well-known flow-duration equations of hydraulic fracturing to guide the learning of the neural network in the absence of significant quantity of data; and (3) development and evaluation of explainable AI (XAI) tools adapted to completion engineering contexts, to gain engineer trust in the AI recommendations as a prerequisite to autonomous use.

Achieving autonomous completions is a 10-year journey that involves concurrent developments in the field of AI methodology, legal structures, liability and responsibility recognition, and restructuring of the workforce. Up to this task must go the engineering community, work in a similarly technical, ethical, and collaborative manner as the best elements of the shale revolution.

References

Chen, B., Barboza, B. R., Sun, Y., Bai, J., Thomas, H. R., Dutko, M., Cottrell, M., & Li, C. (2022). A review of hydraulic fracturing simulation. Archives of Computational Methods in Engineering, 29(1), 1–58. https://doi.org/10.1007/s11831-021-09653-z

Gandossi, L., & Von Estorff, U. (2015). An overview of hydraulic fracturing and other formation stimulation technologies for shale gas production. Publications Office of the European Union. https://doi.org/10.2790/379646

Han, J.-X., Xue, L., Wei, Y.-S., Qi, Y.-D., Wang, J.-L., Liu, Y.-T., & Zhang, Y.-Q. (2023). Physics-informed neural network-based petroleum reservoir simulation with sparse data using domain decomposition. Petroleum Science, 20(6), 3450–3460. https://doi.org/10.1016/j.petsci.2023.10.019

Ibrahim, A. F., Alarifi, S. A., & Elkatatny, S. (2022). Application of machine learning to predict estimated ultimate recovery for multistage hydraulically fractured wells in Niobrara shale formation. Computational Intelligence and Neuroscience, 2022, Article 7084514. https://doi.org/10.1155/2022/7084514

Li, G., Song, X., Tian, S., & Zhu, Z. (2022). Intelligent drilling and completion: A review. Engineering, 18, 33–48. https://doi.org/10.1016/j.eng.2022.07.014

Li, W., Zhang, T., Liu, X., Dong, Z., Dong, G., Qian, S., Yang, Z., Zou, L., Lin, K., & Zhang, T. (2024). Machine learning-based fracturing parameter optimization for horizontal wells in Panke field shale oil. Scientific Reports, 14, Article 6046. https://doi.org/10.1038/s41598-024-56660-8

Ma, Y., & Ye, M. (2025). Application of machine learning in hydraulic fracturing: A review. ACS Omega, 10(11), 10769–10785. https://doi.org/10.1021/acsomega.4c11342

Saini, G. S., Fallah, A., Ashok, P., & van Oort, E. (2022). Digital twins for real-time scenario analysis during well construction operations. Energies, 15(18), Article 6584. https://doi.org/10.3390/en15186584

Sircar, A., Nair, A., Bist, N., & Yadav, K. (2023). Digital twin in hydrocarbon industry. Petroleum Research, 8(3), 270–278. https://doi.org/10.1016/j.ptlrs.2022.04.001

Xu, Y., Guo, B., Zhang, W., Shen, J., Yuan, B., Zhang, W., Zhao, M., Xiong, H., & Jin, A. (2025). Real-time warning method for sand plugging in offshore fracturing wells. Scientific Reports, 15, Article 6062. https://doi.org/10.1038/s41598-025-90768-9

Yang, Z., Li, H., Tuo, X., Li, L., & Wen, J. (2023). Unsupervised clustering of microseismic signals using a contrastive learning model. IEEE Transactions on Geoscience and Remote Sensing, 61, Article 5903212. https://doi.org/10.1109/TGRS.2023.3240728

Jia, J., Fan, Q., Jing, J., Lei, K., & Wang, L. (2024). Intelligent hydraulic fracturing under Industry 4.0: A survey and future directions. Journal of Petroleum Exploration and Production Technology, 14(12), 3161–3181. https://doi.org/10.1007/s13202-024-01877-6

Yuan, B., Zhao, M., Zhang, W., Meng, S., Jin, A., & Dindoruk, B. (2026). Artificial intelligence-driven subsurface hydraulic fracturing engineering: Connotation and practices. Engineering, 58(3), 144–156. https://doi.org/10.1016/j.eng.2025.12.024

Ikotun, A. M., Ezugwu, A. E., Abualigah, L., Abuhaija, B., & Heming, J. (2023). K-means clustering algorithms: A comprehensive review, variants analysis, and advances in the era of big data. Information Sciences, 622, 178–210. https://doi.org/10.1016/j.ins.2022.11.139

EPCM. (2016). Hydraulic fracturing overview. Retrieved August 16, 2026, from https://epcmholdings.com/hydraulic-fracturing-overview/

Simple ML. (2025). Most people don't really understand K-Means here's the visual way that finally clicks. Artificial Intelligence in Plain English. Retrieved August 1, 2026, from https://ai.plainenglish.io/k-means-clustering-explained-from-intuition-to-code-and-visualization-133775129f95

Al-Shuhail, A., Kaka, S. I., Zerguine, A., Iqbal, N., Latif, A. L. A., Bacha, S. S., Yaqub, M. U. M., Putra, R. A., McClellan, J., Liu, E., & Zhu, L. (2016). Microseismic data processing | Research. Center for Energy & Geo Processing. Retrieved August 1, 2026, from https://cegp.ece.gatech.edu/research/projects/microseismic/

Urbancic, T., & Jeziorski, K. (2015). Microseismic monitoring applications in heavy oil reservoirs. CSEG Recorder, 40(1), 38–42. Retrieved August 1, 2026, from https://csegrecorder.com/articles/view/microseismic-monitoring-applications-in-heavy-oil-reservoirs

FEBUS Optics. (2024). Unconventional well monitoring. Retrieved August 1, 2026, from https://www.febus-optics.com/en/page/unconventionalwellmonitoring

Liu, S., Wang, W., Jia, Y., Bian, H., & Shen, W. (2024). Modeling of hydro-mechanical coupled fracture propagation in quasi-brittle rocks using a variational phase-field method. Rock Mechanics and Rock Engineering, 57(9), 7079–7101. https://doi.org/10.1007/s00603-024-03896-5