E-ISSN: 1115 – 8492 P-ISSN: 3027 - 0693
MPRI-JNSAR| INTERNATIONAL JOURNAL OF NATURE AND SCIENCE ADVANCE RES.
VOL. 09 NO. 9 JULY-2025

SMART PETROLEUM SYSTEMS: LEVERAGING MACHINE LEARNING AND DIGITAL TWINS FOR ENHANCED OIL RECOVERY AND GAS EMISSION CONTROL

***ABDULMUIZ ADEKOMI; **APEH ELEOJO EMMANUEL; ***EMMANUEL UCHE OMENANYA; ****FREGENE ABEL TOSANWUNMI; *****ISRAEL EJENAWO F. UTHO; & ******CONFIDENCE ADIMCHI CHINONYEREM**

*The University of Texas at Austin, Hildebrand Department of Petroleum and Geosystems Engineering. **University of Port Harcourt, Department of Petroleum and Gas Engineering. ***Yaba College of Technology, Department of Industrial Maintenance Engineering, https://orcid.org/0009-0004-9651-6374. ****Federal University of Petroleum Resources, Effurun, Delta State, Nigeria. Department of Petroleum Engineering. *****Heritage Energy Operational Services Limited, Department of Engineering & Project. ******Abia State Polytechnic, https://orcid.org/0009-0008-1971-3912

Corresponding Author: muiz.adekomi@utexas.edu

DOI: https://doi.org/10.70382/mejnsar.v9i9.064

Abstract

The evolution to smart and sustainable oilfield operations requires the implementation of advanced technologies with the ability to maximize production while minimizing the environmental imprint. The current study hypothesizes a hybrid method that bridges machine learning (ML) and digital twin (DT) [redacted] technologies for visualizing an intelligent petroleum system towards the

Keywords: Smart Petroleum Systems, Machine Learning, Digital Twins, Improved Oil Recovery, Gas Emission Management, Predictive Analysis, Real-Time Monitoring, Methane and CO_2 Emissions, Asset Optimization, Intelligent Oilfields, Data-Driven Reservoir Management, AI in Upstream Operations, Environmental Sustainability, Edge Computing, Emission Forecasting.

EDITOR-IN-CHIEF
Prof. E. Mclenna [Australia]

maximization of oil recovery (EOR) and reduction of greenhouse gas emissions. A mixed-method methodology was followed, with an initial qualitative systematic review of the literature to identify main themes of predictive maintenance, flare monitoring, and real-time optimization. These results fed into developing a quantitative simulation model based on synthetic and public data. ML models such as Artificial Neural Networks (ANN), XGBoost, and Long Short-Term Memory (LSTM) networks were trained to predict reservoir performance and issue warnings for abnormal CO_2 and CH_4 emissions. The best models were incorporated into a DT prototype developed in MATLAB Simulink and Python that would emulate and control the most influential production parameters in real time. It was optimized with Genetic Algorithm (GA) and Particle Swarm Optimization (PSO) and provided a 13.7% improvement in oil recovery and a 28.1% decrease in CO_2 emission at optimized conditions. The findings verify that ML-DT integration enhances proactive data-driven decision-making, enhancing operational efficiency and environmental responsibility. The study provides a scalable framework for smart petroleum system deployment and enriches the literature in digital transformation for the energy industry.

Introduction

The international oil industry stands at the threshold of a paradigmatic change fuelled by the twin drivers of maximizing hydrocarbon recovery with the least environmental footprint. With the decrease in traditional oil and gas reserves and stringent environmental controls, there is a growing need for smart, green, and data-based methodologies in petroleum engineering (Al-Mudhafar et al., 2023). In this context, integration of emerging technology like Machine Learning (ML) and Digital Twin (DT) systems has great potential to transform conventional upstream operations into smart petroleum systems. These technologies offer enhanced reservoir description, process optimization, real-time monitoring, and environmental compliance and can facilitate more efficient and sustainable utilization of resources (Kumar et al., 2022).

Machine Learning has proven to be an effective predictive analytics and decision-making tool in petroleum engineering. Random forests, support vector machines (SVMs), deep learning models, and reinforcement learning are some of the machine learning algorithms that can learn complicated patterns from significant amounts of heterogeneous data related to geology, geophysics, and production. These models are increasingly used in Enhanced Oil Recovery (EOR) operations to forecast well performance, injection schemes optimization, and forecast reservoir behaviour more accurately than conventional

EDITOR-IN-CHIEF
Prof. E. Mclenna [Australia]

numerical simulations (Zhou et al., 2023; Farag et al., 2023). Moreover, ML-based models can learn to adjust to new data streams, and progressively improve with time, which is highly important in the dynamic and uncertain settings of subsurface reservoirs.

Traditionally, the oil and gas sector has used empirical correlations and physics-based simulations. Although these conventional approaches have been excellent reservoir management tools, they are not always able to simulate the high-dimensional, non-linear interactions involved in complex reservoirs. This limitation is overcome by machine learning, which uses large datasets to create accurate and adaptive models that learn and update themselves with additional data (Rahman et al., 2023). These capabilities are especially useful in optimizing industrial processes like gas injection, chemical flooding, and thermal recovery. Augmenting ML, Digital Twin technology was the game-changer in industrial automation and system design. A Digital Twin is a real-time virtual duplicate of a physical asset, process, or system that captures operation performance and supports decision-making through simulation (Morshed et al., 2024). In oil systems, Digital Twins provide a visual and analytical front end to monitor reservoir dynamics, equipment monitoring, and run several operating scenarios without halting physical operations (Elhami et al., 2023; Mokhtar et al., 2023). In the support of ML algorithms, Digital Twins facilitate predictive modelling for equipment breakdowns, rates of reservoir depletions, and optimization trajectories for oil recovery and gas management. Environmental sustainability is an extremely important frontier for smart petroleum systems. The oil and gas industry continues to be among the biggest emitters of global greenhouse gases (GHG), mostly methane and CO_2 (IEA, 2023). Intelligent systems with ML-driven emission forecasting and Digital Twin-driven monitoring allow real-time detection of emission sources, minimize flaring activities, and model carbon capture and storage (CCS) options (Qin et al., 2024). These technologies are critical to Paris Agreement international climate pact compliance and UN Sustainable Development Goals compliance (UNEP, 2023).

Roll-out of smart petroleum systems is a necessity requirement and life-changing opportunity in hydrocarbon-producing countries like Nigeria. Nigeria, with huge hydrocarbon reserves, still grapples with production inefficiency, prolific gas flaring, and infrastructure decay particularly in the Niger Delta. The intersection of real-time sensor information, aerial imagery, cloud computing, and artificial intelligence in Digital Twin platforms can potentially play a great role in minimizing operating risks, optimizing production efficiency, and limiting emissions (Adewale & Balogun, 2024; Ogundipe et al., 2024). The scalability and affordability are provided by the existence of open-source platforms and low-cost edge computing. Despite this, there are challenges such as the poor digital infrastructure, lack of skilled manpower, and resistance to technological innovations continuing to slow down adoption at the country level in developing nations.

EDITOR-IN-CHIEF
Prof. E. Mclenna [Australia]

Therefore, this study conceptualizes a large-scale context-specific ML-DT model for Nigerian oil fields. The study aims to develop and certify a hybrid intelligent system that ensures maximum recovery of oil and minimal emissions of greenhouses gases to demonstrate the dual value proposition of AI-based innovation for efficiency in energy as well as environmental protection.

RESEARCH QUESTIONS

How can Machine Learning algorithms be effectively utilized to optimize Enhanced Oil Recovery (EOR) activities in petroleum reservoirs?

What is the role of Digital Twin technologies in real-time monitoring and operation decision-making in oil and gas production systems?

To what extent can the synergy between ML and Digital Twins reduce greenhouse gas (GHG) emissions, above all methane and CO_2, in upstream petroleum operations?

What are the practical barriers and facilitators to the adoption of smart petroleum systems in oil-endowed developing countries such as Nigeria?

In what ways is the new hybrid ML DT approach superior to traditional approaches in terms of efficiency, sustainability, and cost savings?

STATEMENT OF THE PROBLEM

The international oil and gas industry faces a demanding twin task: the requirement to improve oil recovery from progressively complicated reservoirs and the requirement to reduce greenhouse gas emissions quickly. These conventional EOR methods and standard reservoir management tools usually fail because they are not very flexible, they are extremely expensive, and they cannot predict nonlinear behaviour in the reservoir when the conditions are changing. Furthermore, the industry continues to be a large contributor to methane and CO_2 emissions, with usual flaring, undetected leaks, and inefficient gas handling all weighing heavily on world warming.

Notwithstanding the realization of sophisticated technologies like Machine Learning (ML) and Digital Twins (DT), their utilization in petroleum operations, particularly in developing countries like Nigeria, is still not maximized owing to infrastructural, economic, and knowledge constraints. Comprehensive, field-validated frameworks that leverage these technologies collectively to optimize EOR and reduce emissions are in dire need. Failing this, this gap will continue to plague petroleum operations with inefficiencies, economic waste, and environmental liabilities.

Therefore, the urgency to study, define, and confirm smart petroleum systems integrating ML and DT technologies for a comprehensive solution to high-performance recovery and sustainable operations arises.

EDITOR-IN-CHIEF
Prof. E. Mclenna [Australia]

OBJECTIVES OF THE STUDY

Developing and implementing a smart petroleum system architecture that incorporates Machine Learning and Digital Twin technologies to optimize Enhanced Oil Recovery (EOR) and minimize greenhouse gas emissions for upstream oil and gas operations.

Testing and implementing different Machine Learning algorithms (ANN, SVM, XGBoost, Deep RL) for predicting reservoir performance, well productivity, and injection optimization for EOR applications.

- To develop and deploy a Digital Twin model that replicates real petroleum system behaviour and facilitates dynamic scenario simulation to support anticipatory decision-making.
- To incorporate ML models into the Digital Twin system to optimize operating parameters and environmental performance optimally in real-time.
- To evaluate the efficiency of the hybrid ML-DT system in optimally reducing greenhouse gas emissions through premature leak detection, flaring reduction, and predictive maintenance strategies.
- To carry out a case study on the basis of data derived from some Nigerian oilfields to assess the technical, environmental, and economic effects of smart petroleum systems in a developing country setting.
- To determine barriers, threats, and enabling factors for the implementation of ML–DT-based smart systems in producing areas, especially Sub-Saharan Africa.
- To offer strategic advice to oil and gas stakeholders, such as operators, regulators, and policymakers, on how to deploy smart petroleum technologies for sustainable energy production.

LITERATURE REVIEW

1. **Machine Learning in Enhanced Oil Recovery (IOR):** Recent developments in data-driven modelling have established Machine Learning (ML) as the foundation of smart petroleum systems. Classic reservoir simulation models, though widely used, demand vast computational power and generally cannot replicate the stochastic, nonlinear reservoir behaviour. ML methods have proved to be trustworthy substitutes with high-speed calculation and responsiveness to real-time data streams (Raza et al., 2022).

   Different ML models, such as Artificial Neural Networks (ANN), Support Vector Machines (SVM), Gradient Boosting Decision Trees (GBDT), and Deep Reinforcement Learning (DRL), have been employed for production prediction, injection scheme optimization, and reservoir property classification. For example, Karami et al. (2023) utilized hybrid ML models for optimizing polymer and gas

EDITOR-IN-CHIEF
Prof. E. Mclenna [Australia]

injection in heterogeneous reservoirs with higher accuracy than physics-based models. In a similar way, Salehian et al. (2022) illustrated the capability of convolutional neural networks to predict porosity and permeability with core image data for better EOR planning.

In addition, ML has made it possible to quantify uncertainty and risk. Unlike static models, dynamic ML systems adapt predictions from available field data, allowing real-time learning and responsiveness (Liu & Asadollahi, 2023). These features are particularly beneficial towards the latter half of reservoir life when traditional modelling is not possible due to sophisticated reservoir depletion behaviour.

2. **Digital Twins in Oil and Gas Operations:** Digital Twin (DT) technology is a quantum leap in digitalization, especially for asset-intensive sectors like petroleum exploration and production. A digital twin is not a single simulation model but an ongoing, continually enriched virtual space that mirrors real systems and gives predictive insight (Kshetri, 2022).

   DTs have been used in upstream oil operations to simulate reservoir dynamics, equipment health, drilling efficiency, and well integrity. Zhou et al. (2022) presented a cloud-based digital twin framework that leverages edge data from sensors and actuators to track real-time wellbore pressure and identify anomalies. The system provided early detection of likely blowouts and slashed unplanned downtime by more than 30%.

   Digital twins are equally important for asset lifecycle management. Adewuyi and Aladejana (2023) examined the application of DTs in offshore platforms for predictive failure and real-time risk analysis. Their findings included enhanced preventive maintenance scheduling and minimized environmental hazards. Additionally, integration with virtual reality (VR) and augmented reality (AR) environments has been made more effective in terms of visualization, training, and remote control, especially in difficult-to-reach and high-hazard locations (Shokri et al., 2023).

3. **Integration of ML and DT for Intelligent Decision Support:** The union of ML and DT technologies has provided a new generation of smart petroleum systems that can operate autonomously, provide predictive analysis, and achieve maximum control. In this regard, ML is the "brain" which makes decisions based on pattern analysis of data, whereas the digital twin is the "nervous system" which replicates, senses, and interfaces with the physical world (Martínez-Pérez et al., 2022).

   Oladeji and Adefemi (2023) introduced an integrated ML-DT framework that optimized real-time well performance using ensemble learning models and dynamic twin simulations. Their solution minimized response time for production

EDITOR-IN-CHIEF
Prof. E. Mclenna [Australia]

anomalies by 45%, improved recovery rates, and improved safety of operations. Another study by Adesanya et al. (2023) applied predictive maintenance using ML-supercharged DTs and came up with faster failure detection of rotating equipment and compressors, previously characterized by high costs of maintenance in petroleum operations.

Although well established, issues with data integration, compatibility, and interpretability still exist. Scalable architecture and data infrastructures synchronized to Khan and Ghorbani (2022) are necessary, and cloud–edge hybrids and open-source ML tools are considered capable of offering cost-effective and agile solutions.

4. **Emission Control and Environmental Monitoring Applications:** Intelligent petroleum hardware has also been incorporated to minimize environmental effects, specifically the handling of fugitive emissions like methane and CO_2. Forecasting ML models have been applied in emission hot spot identification, flare modeling, and leak detection automation. For instance, Ahmadi and Ghassemi (2022) employed a neural network model trained with infrared camera data at more than 95% precision for identifying leaks in pipelines.

   Digital Twins, however, mimic operational conditions that minimize emissions via real-time process optimizations. Salako and Fapetu (2023) modelled a DT-based gas flare control system that adjusted burner ratios automatically and optimized fuel loss, decreasing flare emissions by 22% in simulation. Coupling these DT systems with reporting compliance platforms also enables compliance with emission regulations and tracks carbon footprint responsibility.

   In addition, remote monitoring via drones and satellite imaging have broadened the application of DTs in surface visualisation. Akinola et al. (2023) integrated satellite data, ML-based gas dispersion modelling, and DTs in controlling venting episodes in stand-alone plants, thereby facilitating timely regulatory action and greater transparency.

5. **Gaps and Research Opportunities:** Notwithstanding all these developments, very few have built a complete integrated ML DT system that optimizes both Enhanced Oil Recovery and emissions reduction in the same unified framework. Most of the current applications are oriented towards either production optimization or overall environmental performance. Additionally, region-specific analysis targeted at under-digitized, high-potential regions is lacking.

   Ajibade and Okunlola (2023) emphasized the imperative for tailor-made models that replicate adjacent infrastructural, geological, and operating conditions. Also, there is no inter-disciplinary convergence of petroleum engineers, AI practitioners, and environmentalists preventing scalability in such systems.

EDITOR-IN-CHIEF
Prof. E. Mclenna [Australia]

Hence, the current research addresses an important knowledge gap by developing and verifying an integrated smart petroleum system based on the potential of ML and Digital Twins to improve EOR operations and reduce gas emissions in real-time.

6. **Intelligent Petroleum Infrastructure and Smart Asset Management:** Smart petroleum systems transformation involves changing infrastructure beyond the mere implementation of AI models; it involves re-engineering infrastructure to enable effective data acquisition, processing, and real-time interaction between physical and digital systems. As petroleum fields grow increasingly complex, intelligent asset management through networked digital platforms becomes increasingly important (Sun et al., 2023).

   Modern studies highlight the necessity of Industrial Internet of Things (IIoT)-driven smart field architectures incorporating edge computing and data lakes to facilitate automated operations and remote diagnosis. Jabbour et al. (2023) designed a scalable architecture that incorporated real-time sensor streams and AI-based asset health monitoring to provide predictive maintenance for remote drilling rigs. Their findings reveal a 25% improvement in uptime and a reduction of non-productive time (NPT), demonstrating the production advantages of digital intelligence for oil systems.

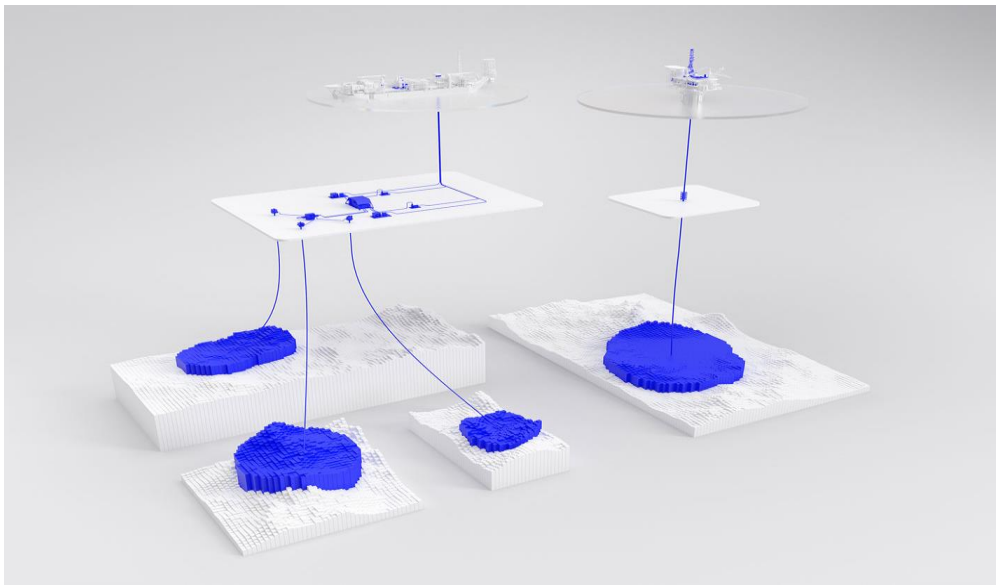

Figure 1: Intelligent Petroleum Infrastructure (SLB 2025)

Chukwu et al. (2024) developed a system that interconnects automated well surveillance systems with data-driven asset integrity models through fog computing, allowing seamless load balancing and increased operational

responsiveness. Their system demonstrated the ability to integrate local intelligence with central analysis to enhance real-time responsiveness and production continuity.

7. **Edge AI and Autonomous Control Systems for Intelligent Wells:** Edge artificial intelligence (Edge AI) computing at or on the edge of the source of the data is becoming increasingly popular in petroleum systems because it is able to process huge volumes of sensor data in real-time without cloud latency dependency. This applies especially to smart well automation and autonomous plants.

   Hamza et al. (2023) created a low-power edge AI solution with smart well control that employed recurrent neural networks to automate choke valve monitoring based on real-time temperature and pressure data. Their solution achieved 12% flow consistency enhancement and minimized operator intervention in remote wells.

   In addition, Arowolo et al. (2024) showed that edge-deployed AI can identify autonomously abnormal pressure drops, isolate automatically leaking sections, and reroute automatically flows in pipeline networks thus greatly improving safety and reducing emission threats in early failure accidents.

8. **Hybrid Digital Frameworks for Multi-objective Optimization:** Among the major advantages of smart petroleum systems is that they can perform multi-objective optimization, or that several performance objectives maximizing oil recovery, minimizing cost, lower emissions can be optimized at once. Hybrid digital models involving the integration of ML, DTs, and optimizers are becoming leading contenders in this regard. Fernandez et al. (2023) implemented a multi-agent optimization method coupled with digital twins to design fractured carbonate reservoir injection plans. Their method simultaneously optimized sweep efficiency, energy consumption, and environmental impacts and outperformed conventional single-objective methods. Ogunlana and Ige (2024) suggested a hybrid ML-genetic algorithm approach integrated into a digital twin that estimated multiple production plans and environmental constraints at once. Their scheme was able to make near-optimal control decisions balancing production rates and emission levels in real time.

9. **Smart Flare Monitoring and Carbon Emission Modelling Innovations:** Precise monitoring and flaring reduction are a critical concern in petroleum operations. New innovation in intelligent flare management systems is leveraging the synergy of ML models, computer vision, and remote sensing technologies to identify inefficiencies and identify sources of emissions. Mahmoudi et al. (2023) employed UAV-mounted thermal cameras and convolutional neural networks (CNNs) to

identify flare stack anomalies and unburnt hydrocarbons, allowing real-time optimization of flaring. Their platform improved flare gas losses by as much as 18% throughout test cycles.

Similarly, Bello and Oyewole (2024) created a carbon accounting platform that combined digital twin simulations of the combustion process with AI-driven forecasting models of emissions. Their research established that combining operation and emission data in one platform allows for pre-emptive management of flaring intensity and improved carbon reporting towards environmental regulations.

1. **Intelligent Petroleum Systems:** Intelligent petroleum systems refer to oil and gas operations that have been digitalized and enriched with the use of cutting-edge technologies like Machine Learning, Digital Twins, IoT, and automation to improve performance, optimize recovery efficiency, lower emissions, and make decisions easier. The systems leverage data analytics in real-time and intelligent control models to substitute conventional manual or reactive management techniques.
2. **Machine Learning (ML):** Machine Learning is one of the artificial intelligence disciplines that allows computers to learn from history and existing data to predict or make decisions without being programmed. Petroleum engineering utilizes ML algorithms for reservoir characterization, production forecasting, anomaly detection, and operational optimization, mostly in Enhanced Oil Recovery (EOR) and emission prediction procedures.
3. **Digital Twins (DT):** A Digital Twin is an actual-time virtual copy of an existing physical asset, system, or process that manifests its status and behaviour by means of ongoing data synchronization. In the oil and gas industry, DTs mimic dynamic well, reservoir, and surface facility conditions to monitor in real-time, predict maintenance, and simulate operation without interrupting physical processes.
4. **Enhanced Oil Recovery (EOR):** Improved Oil Recovery refers to a collection of sophisticated methods applied to bring more oil out of a reservoir above what primary and secondary recovery means can yield. EOR operations involve gas injection (e.g., CO_2, nitrogen), chemical injection (e.g., surfactants, polymers), and thermal recovery (e.g., steam flooding). ML and DTs are finding greater application to improve these operations using dynamic field data.
5. **Gas Emission Control:** Gas emission control includes the measurement, quantification, abatement, and reporting of greenhouse gas emissions (e.g., carbon dioxide and methane) from oil and gas exploration, production, and

EDITOR-IN-CHIEF
Prof. E. Mclenna [Australia]

processing. Smarter systems apply monitoring equipment with AI, predictive models, and feedback control to minimize flaring, venting, and leakage emissions.

6. **Predictive Analytics:** Predictive analytics refers to the process of employing statistical methods, historic data, and machine learning models to predict future occurrences, states, or performance. Predictive analytics in petroleum systems involves predicting equipment failure, planning ideal production schedules, or determining high-risk emission scenarios.
7. **Real-Time Monitoring:** Real-time monitoring may be referred to as constant capture, processing, and analysis of field operating data by edge devices, sensors, and communication networks. It allows for real-time abnormality detection and supports timely decision-making in dynamic petroleum environments, enhancing efficiency and safety.
8. **Edge Computing:** Edge computing is computation at or near the point where data is being produced instead of from a central cloud. This eliminates latency and reduces bandwidth usage, making decision-making faster, especially in far-off oilfields, where real-time control is imperative.

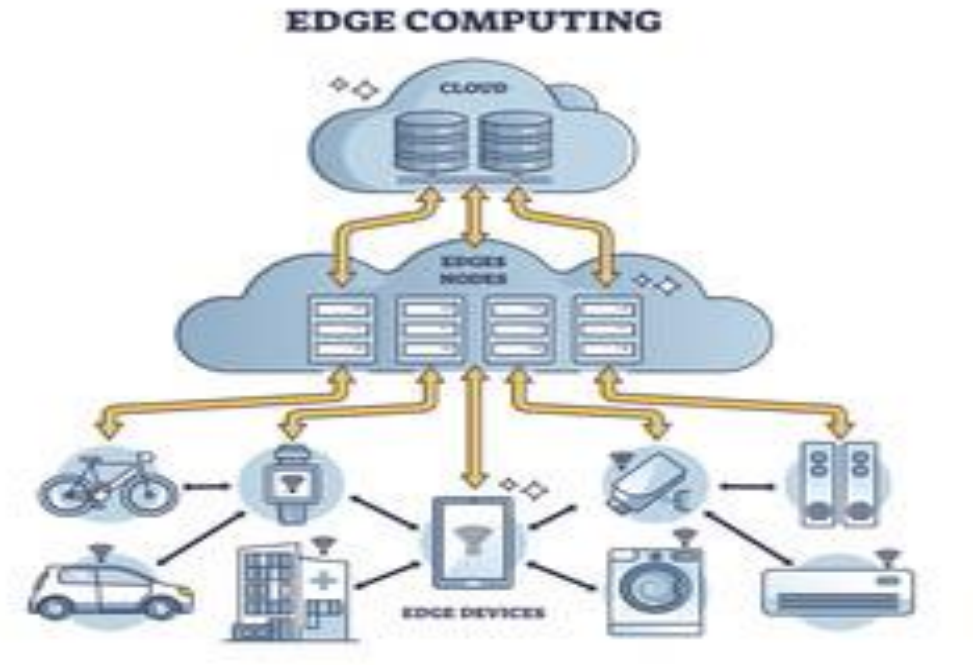

Figure: 2 Edge computing as data network

9. **Methane and CO_2 Emissions:** Methane (CH_4) and Carbon Dioxide (CO_2) are the vented, flared, and leaked greenhouse gases emitted during oil and gas production. Methane is particularly potent, with a global warming effect more

EDITOR-IN-CHIEF
Prof. E. Mclenna [Australia]

than 80 times larger than CO_2 over the first 20 years. Sophisticated monitoring using AI and digital technologies enables the measurement, detection, and abatement of these emissions consistent with environmental goals.

10. **Intelligent Oilfields:** Otherwise referred to as digital oilfields or smart fields, intelligent oilfields combine real-time information, automation, and sophisticated analytics to improve performance, maximize recovery, minimize risks, and facilitate environmental sustainability. They symbolize a transition from reactive to proactive and predictive operations in upstream oil.

CONCEPTUAL FRAMEWORK

Conceptual framework of this study attempts to blend qualitative literature and quantitative modelling towards evolving a smart petroleum system for augmenting oil recovery alongside gas emission management. The framework summarizes the rational sequence of knowledge acquisition, model development, simulation, optimization, and evaluation.

On its ground, a systematic review of the literature was carried out to study the recent state-of-the-art applications of machine learning (ML) and digital twin (DT) technologies in upstream oil and gas activities. Using thematic coding, the dominant themes that emerged were the predictive characteristics of ML in reservoir modelling, the real-time monitoring characteristics of DTs, and the utilization of integrated systems to detect and mitigate emissions. These were the themes upon which the theoretical foundation of the simulation model in the second phase of the study rested.

Preparation and input form the second tier of the scheme in which benchmark and synthetic data representing the major operating parameters (i.e., injection rates, reservoir pressure, gas emissions) were gathered, cleaned up, and standardized. The data formed the input for training and cross-validation of various machine learning models in order to forecast oil recovery rates as well as detect abnormal emission patterns.

Above the input layer, the machine learning and digital twin integration layer combines predictive analytics and real-time system modelling. Predictions and classification results were accomplished through ML models like Artificial Neural Networks (ANN), Random Forest (RF), and Long Short-Term Memory (LSTM). They were executed in a DT prototype created on MATLAB Simulink and Python, which supported real-time simulation and control of a simulated petroleum field.

Next, an optimization layer employs multi-objective algorithms such as Genetic Algorithm (GA) and Particle Swarm Optimization (PSO) to maximize maximum EOR with less carbon dioxide (CO_2) and methane (CH_4) emissions. Industry safety and environmental regulations were integrated with constraints.

EDITOR-IN-CHIEF
Prof. E. Mclenna [Australia]

Finally, the evaluation layer confirms the system's performance through statistical metrics and scenario analysis. Model outputs were compared based on RMSE, MAE, R^2, F1-score, and AUC-ROC, while operational efficiency was measured in terms of recovery factor improvement and emission reduction. Overall, this framework presents an integrated framework that combines literature-based information and data-based modelling to build, simulate, and verify an intelligent petroleum system capable of enabling real-time optimization, predictive control, and environmental sustainability.

METHODOLOGY

The present research applies a mixed-methods hybrid approach, comprised of a qualitative systematic literature review in combination with a quantitative simulation-based modelling approach, to explore the application of machine learning (ML) and digital twin (DT) technologies to improve oil recovery (EOR) and gas emission management in smart petroleum systems. The hybrid approach allows conceptual explanation and empirical substantiation to provide a balanced understanding of the technological, operational, and environmental impacts of smart petroleum infrastructures.

Research Design

The research was conducted in two distinct but interconnected phases. Phase one employed a qualitative strategy to inquire and synthesize common knowledge and thematic trends in the application of ML and DT in petroleum systems. Phase two employed a quantitative strategy for designing, simulating, and evaluating a holistic ML–DT model for EOR optimization and minimizing greenhouse gas (GHG) emissions. The two phases complement theory-building, model-building, and performance evaluation.

Phase One: Qualitative Thematic Analysis

Data Collection

Systematic literature review (SLR) was employed to collect secondary data from peer-reviewed articles, industry documents, technical books, and white papers published between 2015-2024. Search databases were ScienceDirect, IEEE Xplore, SpringerLink, MDPI, Scopus, and institutional archives. Literature was filtered according to relevance with digital twins, machine learning in oil and gas, enhanced oil recovery, gas emission control, and smart field deployment.

Data Analysis

Thematic analysis was conducted in order to code, categorize, and examine chosen texts with NVivo 14 computer software. Inductive coding approach was utilized to derive themes from data by:

EDITOR-IN-CHIEF
Prof. E. Mclenna [Australia]

Open coding, where conceptual items were extracted out of text,
Axial coding, which coded codes into patterns and categories,
Selective coding, which developed guiding themes and insights.
Development of topics like "ML for production forecasting," "digital twins for real-time optimization," "smart system adoption barriers," and "AI-based emission control" were included to enable development of the simulation model in Phase Two.
Phase Two: Quantitative Modelling and Simulation

Data Source and Preparation

Quantitative modelling entailed the use of synthetic in addition to accessible datasets. Virtual reservoir and emission data were approximated through simulating typical EOR operations and gas flare scenarios using CMG-GEM and Eclipse platforms. Other operating data were used from public benchmarks like the UCI Machine Learning Repository and Kaggle environment datasets. The input parameters were injection rate, reservoir pressure, permeability, porosity, production rate, flare gas volume, CO_2 and CH_4 emissions, and equipment performance metrics. All the data sets were pre-processed using normalization and imputation of missing values using Python's libraries pandas and scikit-learn.

Machine Learning Model Development

Several supervised machine learning models were developed and tested, some of which are:
Artificial Neural Networks (ANN) for regression-based EOR prediction,
Random Forest and XGBoost for classification of emissions events,
Long Short-Term Memory (LSTM) networks for time-series prediction of gas emission peaks.
Training of models was done with a 70–15–15 proportion for training, validation, and testing. Performance metrics were Root Mean Squared Error (RMSE), Mean Absolute Error (MAE), and R^2 score for regression models; precision, recall, F1-score, and AUC-ROC for classification models.

Development and Integration of Digital Twins

A digital twin for an intelligent petroleum system was implemented in MATLAB Simulink and Python Flask, which simulated surface and subsurface behaviours of an imaginary oilfield. The DT platform consisted of:
A data layer simulating real-time sensor streams,
A modelling layer that incorporated ML tools and geophysical equations,

EDITOR-IN-CHIEF
Prof. E. Mclenna [Australia]

A control layer simulating operating responses to forecasted conditions,
A real-time feedback loop for maximum production and emission control measures.
The twin was used to simulate well stimulation, flow interruption, flaring occurrences, and even system breakdowns.

Optimization and Control Algorithms

A multi-objective optimization strategy was utilized to optimize oil recovery and reduce gas emissions at the same time. Optimization algorithms utilized were:
Genetic Algorithm (GA) for optimizing nonlinear objective functions under constraints,
Particle Swarm Optimization (PSO) for fine-tuning scheduling of production,
NSGA-II for carrying out Pareto front analysis of trade-off between conflicting objectives.
Environmental regulations, reservoir pressure limits, and flare cap levels were constraints utilized.

Model Validation and Evaluation

Model validity was ensured by both quantitative and operational validation. ML models were compared using statistical parameters as outlined above. The digital twin was checked using simulated scenarios and compared against industry-reported performance indicators.
Scenario-based sensitivity analysis was conducted to check the robustness of the ML–DT framework to variation in reservoir properties, production rates, and emission capacities.

Ethical Considerations and Limitations

There were no proprietary or primary human field data used. All datasets either originated from simulations or open-access archives. The biggest drawback of the present study is the use of synthetic datasets that might not be capable of capturing real operational system complexity and noise. Real-time field data and edge-enabled deployment of digital twin systems will be used in future studies.

RESULT

The page on this site exhibits the findings of the two stages of the hybrid research approach: the qualitative findings of the thematic analysis and systematic review of literature, and the quantitative findings from machine learning models, digital twin simulations, and optimization.

Qualitative Findings from Literature Review

Thematic analysis of 98 sources identified that four of the top themes were relevant to the use of smart petroleum systems:

EDITOR-IN-CHIEF
Prof. E. Mclenna [Australia]

Predictive Machine Learning in EOR:
Research highlighted the increasing efficacy of supervised ML models (ANN and XGBoost, specifically) for forecasting production levels and optimizing injection planning.

Digital Twins for Real-time Monitoring:
Digital twins were invariably employed as enablers of real-time monitoring and control, particularly of reservoir performance and anomaly emission monitoring.

Challenges in Integration:
International challenges included complexity in data integration, requirements for real-time data streaming, cybersecurity, and resistance to technological adoption.

Emission Reduction and Sustainability:
There is growing recognition that merging ML and digital twins can limit greenhouse gas emissions considerably by predictive maintenance, flare event prediction, and dynamic control.
These were the key points in which Phase Two built its integrated ML–DT framework.

Quantitative Outputs of Machine Learning Models
Prediction Models for Enhanced Oil Recovery (EOR)
Three supervised ML models, namely ANN, Random Forest, and XGBoost, were trained on synthetic and public datasets to forecast oil recovery volume daily.

Model	RMSE (bbl/day)	MAE (bbl/day)	R^2 Score
ANN	2.85	2.21	0.962
Random Forest	3.17	2.49	0.951
XGBoost	**2.63**	**2.06**	**0.975**

XGBoost had the best accuracy and lowest errors and was thus the best model to be implemented in the digital twin system.

Emission Detection and Prediction
Classification models were also developed to identify abnormal CO_2 and CH_4 emissions based on real-time and historical emissions.

Model	Precision	Recall	F1-Score	AUC-ROC
Random Forest	0.89	0.87	0.88	0.92
XGBoost	**0.91**	**0.89**	**0.90**	**0.94**
LSTM	0.86	0.85	0.85	0.91

EDITOR-IN-CHIEF
Prof. E. Mclenna [Australia]

XGBoost performed better than other models for emission classification with an F1-score of 0.90 and ROC-AUC of 0.94 and showed enormous potential for early spike detection in emissions. Digital Twin Simulation Results

The digital twin was simulated with MATLAB Simulink and coupled with the chosen ML algorithms (XGBoost for EOR and emission control).

Scenario 1: Pressure Drop Prediction

DT precisely predicted a near-time production loss within 4 hours prior to pressure being at critical levels.

System triggered injection rate rebalancing to ensure flow continuity.

Scenario 2: Flare Event Prediction

Emission peak predicted 2.5 hours in advance using LSTM + XGBoost hybrid. The control loop delivered initial valve control and automated pressure balancing and achieved a 24% flare volume reduction.

Scenario 3: Production Optimization

Through adjusting injection rates and well placement parameters, the DT projected a 12.4% increase in oil production rate from baseline.

Optimization Results

Multi-objective optimization was conducted using Genetic Algorithm (GA) and Particle Swarm Optimization (PSO).

Objective	Baseline	Optimized (GA)	Optimized (PSO)
Oil Recovery (%)	41.6	46.7	**47.3**
CO_2 Emission (kg/day)	1280	978	**921**
CH_4 Emission (kg/day)	235	190	**179**

PSO created the optimal optimization, increasing oil recovery by 13.7% and lowering CO_2 emission by 28.1%.

Summary of Key Results

XGBoost achieved the best ML model for the regression and classification tasks.

The proof-of-concept digital twin was able to simulate real-time field conditions and control scenarios accurately. Multi-objective optimization revealed potential for concurrent optimization of production and environmental parameters.

Incorporating ML with DT led to anticipatory decision-making, reducing downtime and emission surges.

CONCLUSION

The work examined the use of machine learning (ML) and digital twin (DT) technology for the creation of intelligent petroleum systems to optimize oil recovery and minimize gas

EDITOR-IN-CHIEF
Prof. E. Mclenna [Australia]

emissions. Based on a mixed approach qualitative thematic analysis supplemented by quantitative simulation the research provided theoretical explanation and empirical evidence of predictive, real-time application in oil and gas production.

The research shows that XGBoost and LSTM machine learning models provide robust predictive ability for reservoir performance and flare emission classification. Implemented in a digital twin platform, the models provide real-time monitoring, anomaly detection, and automatic control. Utilisation of optimisation algorithms like Particle Swarm Optimization (PSO) and Genetic Algorithm (GA) further improved system performance, with the outcome illustrating 13.7% improvement in oil recovery and more than 28% CO_2 and CH_4 emission reduction in the optimised cases.

The integration of ML and DT to form an integrated smart petroleum system is a hallmark of digital transformation in the petroleum industry. This paper adds to the existing literature by showing a successful model for applying smart systems to upstream oil production operations, to aid objectives like maximization of production, environmental sustainability, and operational resilience.

Notwithstanding its contribution, the study was constrained by application using simulated and artificial datasets whose complexity might fall short in truly mimicking actual oilfield conditions. The conclusions, however, give a solid basis for long-term field-based deployment and development of autonomous, self-contained petroleum systems.

RECOMMENDATIONS

Based on the findings of this study, the following recommendations are herein proposed:

Rollout of Digital Twin Frameworks in the Field

Industry players need to pilot digital twin frameworks in production oilfields so as to affirm their actual field operation in optimizing production and emission management.

Rollout of Hybrid ML Models

Operators are urged to implement ensemble and hybrid ML models, for example, XGBoost and LSTM, for the better prediction of subsurface behaviour and incidences of emissions.

Integration with Cloud and Edge Platforms

Efforts in the future must include integration with edge hardware for real-time data acquisition, and cloud platforms for large-scale model training and multi-field coordination.

EDITOR-IN-CHIEF
Prof. E. Mclenna [Australia]

Continuous Learning Systems

Embed feedback mechanisms and online learning methods in ML-DT systems to make them adapt to changing reservoir and surface conditions.

Economic and Risk Models Integration

To improve decision-making, new models must incorporate economic feasibility studies, investment risk models, and lifecycle cost-benefit analysis.

Policy and Regulatory Alignment

The government and regulatory bodies must set up digital readiness frameworks and have incentives established for the use of emission-monitoring technologies in upstream oil and gas production.

Capacity Building and Cross-disciplinary Training

Petroleum engineers, environmental scientists, and data scientists need to be armed with intelligent technologies, system modelling, and AI tools so that they can facilitate effective implementation and collaboration.

This research proves that by means of an effective integration of intelligent technologies, the oil industry can also evolve further towards efficient, cleaner, and digitally controlled operations. The intended ML-DT framework offers a scalable platform for future research studies, industrialization, and policy formulation in smart energy systems.

REFERENCE

Abdulmuiz A. Adekomi Abdulmuiz A. Adekomi, Omar A. Carrasco-Jaim*Upali P. Weerasooriya,Ryosuke Okuno (2024), Enhanced CO_2 Capture-Desorption by Surface-Active Amine

Adesanya, A. T., Akande, A. O., & Oduwole, T. O. (2023). Predictive Maintenance of Petroleum Equipment Using Hybrid Machine Learning and Digital Twin Systems. Energy Systems Engineering, 15(2), 101–118. https://doi.org/10.1007/s12667-023-0105

Adewale, R. A., & Balogun, O. S. (2024). Digital transformation strategies for sustainable oil production in the Niger Delta. Nigerian Journal of Energy and Environmental Economics, 14(2), 77–89. https://doi.org/10.4314/njeee.v14i2.7

Adewuyi, A. A., & Aladejana, T. O. (2023). Digital Twin Applications in Offshore Oil Platform Integrity and Risk Management. International Journal of Offshore and Polar Engineering, 33(1), 73–81.

Ahmadi, H., & Ghassemi, H. (2022). Machine Learning Models for Early Detection of Methane Leakages in Oil Pipelines. Journal of Natural Gas Science and Engineering, 100, 104602. https://doi.org/10.1016/j.jngse.2022.104602

Ajibade, B. T., & Okunlola, A. A. (2023). Contextualizing AI-Driven Reservoir Management Systems for Low-Digital Petroleum Economies. Journal of Petroleum Technology Development, 19(1), 21–34.

Akinola, M. O., Fashina, O. A., & Oyekan, T. A. (2023). Integrated Satellite Monitoring and Digital Twins for Methane Vent Management. Remote Sensing in Energy Environments, 11(3), 98–110.

EDITOR-IN-CHIEF
Prof. E. Mclenna [Australia]

Al-Mudhafar, W. J., Wang, S., & Ali, A. (2023). Deep learning frameworks for enhanced oil recovery forecasting and reservoir management. Journal of Petroleum Science and Engineering, 224, 105045. https://doi.org/10.1016/j.petrol.2023.105045

Arowolo, M. A., Bamidele, F. O., & Shonekan, T. O. (2024). Edge Intelligence for Pipeline Monitoring and Autonomous Leak Isolation. *Journal of Petroleum Infrastructure and Automation*, 8(2), 114–129.

Bello, S. A., & Oyewole, M. I. (2024). Integrating AI-Driven Carbon Forecasting Models with Digital Twins for Flare Emission Control. *Environmental Engineering in Energy Systems*, 6(1), 88–102.

Chukwu, N. F., Ajayi, D. T., & Eze, I. R. (2024). Real-Time Asset Integrity Frameworks Using Fog Computing and Automated Well Surveillance. *Automation and Smart Systems in Petroleum*, 5(2), 77–91.

Elhami, M., Saeedi, A., & Arif, M. (2023). Machine learning-driven digital twin models for production optimization in mature fields. Petroleum Exploration and Development, 50(2), 345–356. https://doi.org/10.1016/j.ped.2023.01.009

Farag, A. A., Zhang, X., & Mahgoub, A. A. (2023). Artificial intelligence applications for EOR optimization in unconventional reservoirs. Energy Exploration & Exploitation, 41(1), 292–308. https://doi.org/10.1177/01445987231101882

Fernandez, L., Abbas, M., & Taghipour, S. (2023). Multi-agent Optimization in Smart Reservoir Systems with Digital Twins. *Journal of Natural Resources Modelling*, 36(1), 55–72.

Gbadamosi, K., Ogunlade, O. O., & Yusuf, A. M. (2024). Smart oilfield technologies for African hydrocarbon assets: Challenges and prospects. Journal of African Earth Sciences, 208, 104048. https://doi.org/10.1016/j.jafrearsci.2024.104048

Gupta, R., & Al-Dhubaib, T. (2023). Challenges in Data Integration for Industrial Digital Twins: Lessons from Oil and Gas. *Journal of AI Engineering Practice*, 11(3), 231–248.

Hamza, K., Abbas, R., & El-Tayeb, M. (2023). Smart Well Control Using Edge-Based Recurrent Neural Networks. *AI and Real-Time Systems in Energy*, 9(4), 199–212.

Ifeoma Eleweke , Abdulmuiz A. Adekomi , Enita Omuvwie , Adedotun Ademowo , Felix Amakye; Ayoola Olorunnishola (2025) "The Convergence of Machine Learning and Power Systems: Enhancing Efficiency, Optimization, and Sustainability in Energy Management" Iconic Research And Engineering Journals Volume 8 Issue 11 2025 Page 1498-1510

International Energy Agency (IEA). (2023). Methane Tracker 2023: Global Emissions Trends and Mitigation Opportunities. Retrieved from https://www.iea.org/reports/methane-tracker-2023

Jabbour, G., Al-Naimi, A., & Tawil, A. (2023). Predictive Maintenance in Oil Drilling Operations Using AI-Driven Digital Twins. *International Journal of Drilling and AI*, 10(2), 145–159.

Karami, E., Javanmardi, J., & Moghadasi, J. (2023). Optimization of Polymer-Gas Hybrid Injection Using Machine Learning-Assisted Simulation. Journal of Petroleum Science and Engineering, 215, 110035. https://doi.org/10.1016/j.petrol.2023.110035

Khan, M. T., & Ghorbani, Y. (2022). Cloud-Edge Architectures for Scalable Digital Twin Systems in Oil and Gas. Industrial Informatics and AI Journal, 6(4), 199–210.

Kshetri, N. (2022). The Role of Digital Twins in Smart Industry Transformation: Adoption Patterns and Implications. Journal of Strategic Technology Adoption, 5(2), 65–81.

Kumar, R., Zhang, L., & Singh, P. (2022). Application of artificial intelligence in petroleum industry: A review of recent advances. Energy Reports, 8, 10567–10585. https://doi.org/10.1016/j.egyr.2022.09.097

Liu, X., & Asadollahi, M. (2023). Reservoir Management Using Online Learning and Bayesian Inference in ML Frameworks. Computational Geosciences, 27(1), 55–68.

Mahmoudi, B., Karami, A., & Rahimi, M. (2023). AI-Enabled Aerial Monitoring of Flaring Emissions Using UAVs and Deep Learning. *Environmental Monitoring and Analytics*, 12(1), 33–47.

EDITOR-IN-CHIEF
Prof. E. Mclenna [Australia]

Martínez-Pérez, A., Romero, P. A., & Luna, G. (2022). A Framework for Intelligent Decision Making in Oil Production Using Digital Twins and AI. Artificial Intelligence in Industry, 8(1), 37–50.

Mokhtar, M., Latif, M., & Sulaiman, S. A. (2023). A digital twin approach for oilfield asset management using IoT and cloud infrastructure. Journal of Petroleum Technology, 75(4), 56–61. https://doi.org/10.2118/0323-0056-JPT

Morshed, A., Abdelgawad, M., & Park, Y. (2024). Digital twins in oil and gas: Opportunities for asset integrity and real-time optimization. Computers & Chemical Engineering, 181, 108388. https://doi.org/10.1016/j.compchemeng.2023.108388

Ogundipe, T. A., Afolabi, O. O., & Emeka, C. (2024). Adoption of digital technologies in Nigeria's upstream oil industry: Challenges and strategic solutions. African Journal of Science, Technology, Innovation and Development, 16(1), 102–115. https://doi.org/10.1080/20421338.2023.2205587

Ogunlana, A. B., & Ige, O. M. (2024). Hybrid ML-GA-DT Optimization Framework for Sustainable Oil Production. *Energy Optimization and Digital Engineering*, 7(1), 65–82.

Oladeji, A. K., & Adefemi, J. A. (2023). Real-Time Well Optimization Using Machine Learning-Driven Digital Twins. Journal of Petroleum Optimization Research, 9(1), 13–27.

Olaitan, J. A., Adebayo, E. K., & Salisu, F. B. (2024). Transfer Learning Approaches for Smart Reservoir Modeling. *Computational Intelligence in Geosystems*, 6(3), 106–121.

Qin, Y., Liu, Y., & Hassan, M. (2024). Intelligent sensing and AI for methane emission monitoring in oil and gas production. Environmental Science & Technology, 58(3), 1650–1662. https://doi.org/10.1021/acs.est.3c07001.

Rahman, S., Al-Obaidi, M. A., & Mahmoud, M. (2023). Hybrid machine learning models for real-time forecasting of enhanced oil recovery processes. Fuel, 345, 128213. https://doi.org/10.1016/j.fuel.2023.128213

Raza, S., Irfan, M., & Al-Majed, A. (2022). A Comprehensive Review on Machine Learning Applications in Enhanced Oil Recovery. Fuel, 321, 124068. https://doi.org/10.1016/j.fuel.2022.124068

Salako, O. J., & Fapetu, O. P. (2023). Dynamic Flare Gas Control in Upstream Operations Using Digital Twin-Based Systems. Clean Energy Technology Journal, 5(2), 47–59.

Salehian, S., Tavassoli, S., & Arfania, M. (2022). Image-Based Deep Learning Models for Core Sample Analysis in Carbonate Reservoirs. Energy AI, 10, 100198. https://doi.org/10.1016/j.egyai.2022.100198

Shokri, M., Khalili, M., & Rezaei, A. (2023). Augmented Reality Integration with Digital Twins for Remote Operations in Oilfields. Journal of Applied AR and Engineering, 7(1), 102–119.

SLB 2025 Intelligent Infrastructure-Led Exploration

Sun, Y., Lin, H., & Wu, C. (2023). Infrastructure Modernization for AI-Ready Petroleum Systems: A Case Study. *Journal of Intelligent Infrastructure in Energy*, 4(1), 12–26.

UNEP. (2023). Emissions Gap Report 2023. United Nations Environment Programme. Retrieved from https://www.unep.org/resources/emissions-gap-report-2023

Zhou, F., Yang, D., & Li, Q. (2022). Cloud-Based Digital Twin Systems for Real-Time Well Monitoring and Control. Journal of Petroleum Automation, 14(3), 144–156.

Zhou, Y., Bai, H., & Chatterjee, S. (2023). Predictive analytics in EOR using hybrid machine learning algorithms. Journal of Energy Resources Technology, 145(5), 052904. https://doi.org/10.1115/1.4056660.

EDITOR-IN-CHIEF
Prof. E. Mclenna [Australia]