Deep Learning for CO_2 Leakage Detection and Risk Assessment

Rachel Mice

Department AI and Machine Learning, Norton University, Phnom Penh, Cambodia

Abstract

The viability of geological carbon dioxide (CO_2) storage as a climate mitigation strategy hinges on the reliable detection of subsurface leakage and the quantitative assessment of associated risks. Conventional monitoring techniques — seismic imaging, wellbore instrumentation, geochemical sampling, and remote sensing — produce voluminous, multi-modal datasets whose manual interpretation is increasingly impractical. Deep learning, leveraging hierarchical feature extraction through multi-layered neural architectures, has demonstrated significant promise in automating leakage detection, accelerating risk quantification, and enabling real-time decision support. This review critically examines the application of deep learning to CO_2 leakage monitoring and risk assessment within carbon capture and storage (CCS) projects.

Keywords: deep learning; carbon capture and storage; CO_2 leakage; seismic monitoring; risk assessment; physics-informed neural networks; anomaly detection; explainable AI; uncertainty quantification

Introduction

Geological carbon dioxide (CO_2) storage constitutes a key pillar of the global strategy to limit anthropogenic warming to 1.5 °C above pre-industrial levels. In this process, CO_2 captured from power generation, cement manufacture, steel production, and other industrial sources is compressed, transported, and injected into deep subsurface formations — principally saline aquifers and depleted hydrocarbon reservoirs — for permanent sequestration. The Global CCS Institute reports that as of late 2023, over 40 commercial-scale CCS facilities were operational worldwide, with a combined annual capture capacity approaching 50 million tonnes and more than 350 projects in various stages of development [Williams, M. O. (2023)]. Projections by the International Energy Agency (IEA) indicate that CCS must scale to approximately 6 gigatonnes of CO_2 per year by 2050 to meet net-zero commitments.

The environmental and regulatory credibility of every storage project rests on the guarantee that injected CO_2 remains permanently contained. Leakage — the unintended migration of CO_2 towards the surface or into potable aquifers — can occur through a range of pathways: compromised cement or corroded casing in legacy wellbores; caprock fractures induced by injection-related pressure build-up; reactivated pre-existing faults; and diffuse permeation through geochemically altered seal rock [International Energy Agency]. Even low-magnitude, persistent leaks can negate the climate benefit of storage and undermine public acceptance, making robust monitoring, verification, and accounting (MVA) systems indispensable.

Current MVA programmes draw on diverse geophysical, geochemical, and remote-sensing techniques. Time-lapse (4D) seismic surveys image plume evolution; downhole pressure and temperature gauges track reservoir response; soil-gas flux measurements and atmospheric eddy-covariance stations detect near-surface migration; and satellite interferometric synthetic-aperture radar (InSAR) identifies ground deformation. Each method, however, confronts inherent limitations — cost, acquisition frequency, spatial coverage, or susceptibility to natural background variability — that constrain its standalone detection capability [Benson, S.M. and Cole, D.R].

CO_2 Storage Integrity and the Detection Challenge

Four sequential trapping mechanisms underpin long-term storage security. Structural and stratigraphic trapping confines buoyant CO_2 beneath an impermeable caprock. Residual trapping immobilises disconnected CO_2 ganglia within pore throats via capillary forces. Solubility trapping dissolves CO_2 into the formation brine, increasing its density and promoting convective mixing. Mineral trapping — the slowest but most permanent mechanism — converts dissolved CO_2 into stable carbonate minerals through reactions with silicate lithologies [Williams, M. O. (2024)]. The effectiveness of these mechanisms depends on reservoir geology, injection strategy, and the geochemical composition of the formation fluids, all of which exhibit significant spatial variability and parametric uncertainty.

Leakage signals are diverse and pathway-dependent. A wellbore cement failure may produce an abrupt pressure anomaly localised around a single well, whereas diffuse caprock permeation generates a gradually broadening geochemical plume and faint, spatially distributed seismic amplitude changes. Fault-mediated leakage can manifest as microseismic activity, pressure perturbations propagating along the fault plane, and eventually surface deformation detectable by InSAR. A practical detection system must therefore be sensitive across a broad spectrum of signal magnitudes, spatial footprints, and temporal dynamics — precisely the multi-scale, multi-modal recognition task at which deep learning excels.

Deep Learning Architectures Applied to Leakage Detection

Convolutional Neural Networks for Spatial Pattern Recognition

Convolutional neural networks (CNNs) underpin the majority of published leakage-detection studies targeting spatially structured monitoring data. In the typical workflow, a time-lapse seismic difference volume — generated by subtracting a pre-injection baseline survey from a subsequent monitor survey — is input to a CNN that classifies each sub-volume as “normal plume” or “anomaly.” Hierarchical convolutional filters learn local features (amplitude discontinuities, diffraction patterns) at early layers and aggregate them into global leakage signatures at deeper layers. Encoder–decoder architectures such as U-Net have been adapted for full-volume semantic segmentation, producing voxel-wise probability maps of CO_2 migration that delineate leakage plume geometry with accuracy rivalling expert manual interpretation [LeCun, Y. et. al.]. Attention mechanisms — channel attention (squeeze-and-excitation blocks) and spatial attention (convolutional block attention modules) — further enhance detection by dynamically weighting the most discriminative features, improving performance in low signal-to-noise environments [IPCC].

Recurrent and Temporal Architectures

Temporal monitoring data — continuous pressure records, distributed temperature profiles, time-sampled geochemical concentrations — encode leakage dynamics that spatial snapshots alone cannot capture. Long short-term memory (LSTM) networks learn to distinguish the characteristic temporal trajectories of normal reservoir response from those associated with wellbore failure, slow caprock permeation, or fault reactivation. Gated recurrent units (GRUs) offer a computationally lighter alternative with comparable performance for most CCS time-series tasks [Rutqvist, J]. Convolutional LSTM (ConvLSTM) hybrids extend this capability by simultaneously modelling spatial dependencies within each time step and temporal evolution across steps, making them well suited for processing sequences of seismic attribute maps or spatially distributed sensor arrays [Goodfellow, I., et. al.].

Physics-Informed Neural Networks

A persistent criticism of purely data-driven models is their potential to produce physically implausible predictions, particularly when extrapolating beyond training distributions. Physics-informed neural networks (PINNs) address this by incorporating

Table 1. Monitoring data modalities with associated deep learning architectures, strengths, and limitations.

Modality	*Typical DL Architecture*	*Detection Strength*	*Key Limitation*
4D seismic volumes	3D CNN, U-Net, ViT	High spatial resolution; plume geometry	Expensive; infrequent acquisition
Downhole P/T sensors	LSTM, GRU, CNN–1D	Real-time; rapid onset detection	Single-point spatial coverage
Distributed fibre optics (DAS/DTS)	ConvLSTM, 1D–2D CNN	Continuous spatial profile along wellbore	Large data rates; signal processing
InSAR / satellite imagery	2D CNN, ResNet	Wide-area surface deformation	Indirect; atmospheric artefacts
Soil-gas / geochemistry	Autoencoders, DNN	Sensitive to shallow migration	Highly variable baselines

the governing partial differential equations of multiphase porous-media flow — mass conservation, the Darcy–Muskat formulation, and energy balance — as soft constraints in the loss function [7]. The resulting models honour conservation laws by construction, generalise more reliably to unseen reservoir conditions, and require substantially fewer labelled training examples because the physics provides a strong inductive bias. PINNs have been demonstrated as rapid-turnaround surrogates for reservoir simulators, enabling probabilistic leakage scenario evaluation at computational costs orders of magnitude below full-physics simulation [Um, E.S. and Alumbaugh, D].

Generative and Self-Supervised Models for Anomaly Detection

The extreme rarity of confirmed leakage events at operational CCS sites creates a fundamental class-imbalance problem for supervised detection. Generative models circumvent this by learning the distribution of "normal" monitoring data and flagging deviations. Variational autoencoders (VAEs) learn a compressed latent representation from which reconstruction error serves as an anomaly score: elevated reconstruction loss indicates features absent from the normal training distribution. Generative adversarial networks (GANs) employ a discriminator trained on normal data; genuine leakage signals, being distributionally dissimilar, receive low discriminator confidence scores [Hu, J., Shen, L. and Sun, G.,]. Contrastive self-supervised learning — pre-training an encoder to distinguish positive (augmented) pairs from negative pairs without labels — has more recently been explored as a strategy for building powerful feature extractors from abundant unlabelled seismic volumes, which can then be fine-tuned with minimal labelled leakage examples [Um, E.S. and Alumbaugh, D].

Monitoring Data Modalities and Multi-Source Fusion

Risk Assessment and Uncertainty Quantification

Deep Learning Surrogate Models

Rigorous risk assessment for a CO_2 storage site requires exploring a high-dimensional uncertain parameter space — spanning permeability heterogeneity, fault transmissibility, relative permeability curves, caprock thickness, and injection scheduling — through thousands to millions of forward simulations. Full-physics reservoir simulators (e.g., TOUGH2, ECLIPSE, CMG-GEM) may require hours to days per run, rendering Monte Carlo analysis infeasible. Deep learning surrogates trained on strategically sampled simulation ensembles reproduce the simulator's input–output relationship in milliseconds, enabling rapid evaluation of leakage probability distributions, sensitivity rankings, and value-of-information analyses for monitoring network design [Hu, J., Shen, L. and Sun, G.,].

Bayesian Uncertainty Quantification

Deterministic point predictions are insufficient for safety-critical decisions; regulators and operators require calibrated uncertainty bounds. Bayesian deep learning approaches — Monte Carlo dropout, deep ensembles, and stochastic variational inference — produce predictive distributions that distinguish high-confidence detections from ambiguous signals. In a leakage detection context, epistemic uncertainty maps highlight subsurface regions where model confidence is low, directly informing decisions about where to deploy additional monitoring assets for maximum risk reduction.

Validation: Field-Scale and Benchmark Applications

The Sleipner CO_2 storage project (Norwegian North Sea), operational since 1996, has served as the primary open-access testbed for deep learning monitoring research. CNN-based models trained on Sleipner's time-lapse seismic dataset have demonstrated automated tracking of nine distinct CO_2 layers within the Utsira Sand, achieving plume delineation accuracy within the confidence intervals of expert interpretation while reducing processing time from weeks to minutes.

Open Challenges

Training Data Scarcity and Domain Shift

The near-total absence of confirmed leakage from operational CCS sites means that models are overwhelmingly trained on simulator-generated data. Distributional differences between synthetic and field data — arising from idealised physics, simplified noise models, and incomplete geological characterisation — can degrade field performance. Domain adaptation techniques (adversarial domain alignment, optimal transport-based distribution matching) and few-shot learning paradigms are being investigated to mitigate this simulator-to-field gap, but their effectiveness for CCS-specific data regimes remains to be comprehensively demonstrated.

Scalability and Computational Efficiency

Processing full 3D seismic volumes or high-rate distributed acoustic sensing (DAS) streams with deep neural networks demands substantial compute infrastructure. Efficient inference at remote, bandwidth-limited CCS sites requires model compression (pruning, quantisation, knowledge distillation) or edge-AI hardware deployment. Furthermore, models must remain performant over multi-decadal operational horizons as sensor characteristics drift and subsurface conditions evolve, necessitating automated retraining and continual learning pipelines.

False Alarm Management

In operational settings, each false leakage alarm can trigger costly well interventions, injection shutdowns, and regulatory investigations. The base rate of genuine leakage is extremely low, so even a modest false-positive rate translates into a poor positive predictive value. Multi-tier detection architectures — coupling a sensitive screening model with a high-specificity confirmatory model, or requiring concordance across multiple independent data streams before issuing an alert — are necessary to achieve acceptable operational false-alarm rates.

Conclusions

Deep learning is rapidly advancing the capability, speed, and reliability of CO_2 leakage detection and risk assessment for geological carbon storage. Convolutional networks extract spatial leakage signatures from seismic and satellite imagery; recurrent and convolutional-recurrent hybrids model temporal dynamics in sensor time series; physics-informed architectures enforce physical consistency and reduce data requirements; and generative and self-supervised models enable anomaly detection in the absence of labelled leakage examples. Multi-modal fusion architectures, Bayesian uncertainty quantification, and explainable AI techniques collectively bring these capabilities closer to the standards of transparency and reliability demanded by regulatory authorities.

References

1. International Energy Agency. (2021). Net zero by 2050: A roadmap for the global energy sector. IEA Publications.
2. Benson, S. M., & Cole, D. R. (2008). CO_2 sequestration in deep sedimentary formations. Elements, 4(5), 325–331.
3. Williams, M. O. (2024). Development of reactive heat exchangers for enhanced geothermal energy recovery. Power System Protection and Control, 52(2), 18–37.
4. LeCun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. Nature, 521(7553), 436–444. https://doi.org/10.1038/nature14539
5. Intergovernmental Panel on Climate Change. (2005). Special report on carbon dioxide capture and storage. Cambridge University Press.
6. Rutqvist, J. (2012). The geomechanics of CO_2 storage in

deep sedimentary formations. Geotechnical and Geological Engineering, 30(3), 525–551.
7. Goodfellow, I., Bengio, Y., & Courville, A. (2016). Deep learning. MIT Press.
8. Um, E. S., & Alumbaugh, D. (2022). Deep learning for CO_2 leakage detection at geologic sequestration sites. Geophysics, 87(3), M111–M124.
9. Hu, J., Shen, L., & Sun, G. (2018). Squeeze-and-excitation networks. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (pp. 7132–7141).