Journal of Environmental & Earth Sciences
https://journals.bilpubgroup.com/index.php/jees

REVIEW

Machine-Learning-Enhanced Transient Electromagnetic Signal Processing for Advanced Mineral Exploration

Changchun Wang [1,2]***, Maohong You*** [2*]***, Fei Qin*** [2]

[1] *School of Geophysics and Geomatics, China University of Geosciences (Wuhan), Wuhan 430074, China*
[2] *Shanxi Dibao Energy Co., Ltd., Taiyuan 030000, China*

ABSTRACT

This review synthesizes recent progress in machine-learning-enhanced transient electromagnetic signal processing for advanced mineral exploration, with emphasis on methods that improve reliability under realistic field conditions. Transient Electromagnetic (TEM) data provide strong sensitivity to subsurface conductivity, yet practical interpretation is often limited by non-stationary interference, platform- and instrument-dependent artifacts, wide dynamic range decay, and ill-posed inversion. We organize the literature using a pipeline perspective spanning automated quality control, denoising and interference suppression, system-response correction and normalization, representation learning, Machine learning (ML) assisted inversion (including surrogate forward models and hybrid physics ML inference), and target detection and ranking. Particular attention is given to the exploration-specific constraints that shape evidence quality, including label scarcity and bias, the synthetic-to-field gap, spatial leakage in evaluation splits, and the need for cost-aware metrics tied to drill decisions rather than pointwise regression error. Across reported studies, the strongest and most transferable benefits are observed in quality control (QC) automation and interference-aware denoising that improve repeatability and stabilize downstream inversion. More ambitious end-to-end inversion and target-ranking models remain promising but are highly sensitive to domain shift across waveforms, gate schedules, noise regimes, and geology, making calibrated uncertainty estimation and out-of-distribution detection central requirements for deployment. We conclude by outlining reproducibility and reporting practices suitable for SCI-standard evidence and by identifying priority research directions, including realistic

*CORRESPONDING AUTHOR:
Maohong You, Shanxi Dibao Energy Co., Ltd., Taiyuan 030000, China; Email: link8197@foxmail.com

ARTICLE INFO
Received: 1 January 2026 | Revised: 30 January 2026 | Accepted: 3 March 2026 | Published Online: 17 April 2026
DOI: https://doi.org/10.30564/jees.v8i4.13043

CITATION
Wang, C., You, M., Qin, F., 2026. Machine-Learning-Enhanced Transient Electromagnetic Signal Processing for Advanced Mineral Exploration. Journal of Environmental & Earth Sciences. 8(4): 157–183. DOI: https://doi.org/10.30564/jees.v8i4.13043

synthetic data generation, hybrid inversion with error control, and benchmark tasks aligned with exploration value.
Keywords: Transient Electromagnetics; Mineral Exploration; Machine Learning; Denoising; Inversion

1. Introduction

The transient electromagnetic (TEM) technique has been extensively applied to explore changes in electrical conductivity of the ground surface, which offers significant data relating to mineralization, groundwater distribution, and geology [1]. Conductive signatures associated with massive sulfides, graphite, and certain alteration or fluid pathways can often be detected even when targets are buried beneath transported cover, obscured by complex structure, or only weakly expressed in magnetic or radiometric data. As exploration gradually moves to deeper and less obvious deposits, and discovery steadily moves to those locations where outcrop is limited and difficult access is constrained, TEM will continue to hold a special place: within reasonable acquisition geometry and processing faithfulness, it will have the capability to offer rapid regional screening and, at the same time, be able to contribute to deposit-scale delineation [2].

Close-surface variations in conductivity, system noise and electromagnetic coupling effects can largely dominate early- and mid-time gates of TEM responses and will complicate the interpretation of deeper geological structures [3,4]. A TEM measurement tries to measure a response that decays rapidly over a large dynamic range with diagnostic content spread between early and late times. Early-time gates and other gates can be dominated by the response of the system, or by the coupling between transmitters and receivers, saturation mechanisms, or near-surface effects, whereas late-time gates, which are frequently the most important to use in deriving information regarding deeper conductors, can be dominated by noise, just when the signal amplitude is at its minimum. This signal-processing problem is inherently non-stationary and multi-source: numerous noise processes, which have various temporal and spectral characteristics, may act on the true geological response. Since later interpretation may always need to be inverted via an imperfect forward model, any bias present at the time of preprocessing can be multiplied and reinforced, and finally affect costly and difficult-to-undo decision-making, including drill targeting.

Developed mineral exploration conditions provide a unique combination of interference sources and acquisition artifacts. Platform motion and time-varying variable geometry cause time-varying coupling and noise in airborne TEM (AEM), which may vary quickly in line direction and inter-flight line. Fluctuations in altitude, attitude, and the stability of transmitter currents interplay with the system transfer functions as well as gating strategies and render the successful data quality spatially different even within one survey. In spite of its reduced mobilization dependence, Ground TEM may also be compromised by imperfect loop geometry, variable coupling to conductive layers near the surface, and practical factors of the difficulty of laying instruments in a rough or vegetated environment. The transient being measured in either of the two modalities is determined by the transmitter waveform, front-end bandwidth, instrument impulse response, and the stacking, or gating, of the transistor to trade off between time resolution and noise reduction [5]. The products of the resulting data may vary considerably across contractors, systems, and vintages and thus make cross-survey comparisons and model transfer difficult.

There is also structural diversification in interference in TEM data. Power infrastructure harmonics may place narrowband components and harmonics in places that contain useful parts of the transient and impulsive events, like sferics, and may inject broadband spikes into several gates. Other cultural electromagnetic emissions, which are not always continuous but are spatially localized, may cause line-dependent artifacts that are hard to locate with global thresholds. Although the level of environmental noise is moderate, conditions of the subsurface can make interpretation challenging: conductive overburden, saline groundwater, graphitic metasediments, and other non-economic conductors may have strong responses that either obscure signal indications of ore or give false positives. Stereotyped inversion methods are based on overlaying simplification of assumptions on distributions of subterranean conductivities, as well as on noise properties, both of which are often contravened in geologically complicated settings [6–8].

The classical TEM processing and inversion processes present fundamental foundations of addressing such problems, yet are fragile to assumptions that are broken. Stacking and robust statistics may be useful in the cancellation of random noise, but cannot effectively counteract structured interference correlated in time or direction of line [9]. Notch filtering can minimize powerline contamination, although it may not work when harmonics shift, change in amplitude, or the interference occupies the same time-frequency representation as the actual signal. This can work in cases of apparent failures like saturation or dead channels; however, in less obvious cases of distortion, rule-based quality control tends to fail, and it can also tend to reject informative late-time data in the hardest environments. Inversion techniques, such as 1D layered-earth techniques to more recently developed laterally constrained inversion and full 3D methods, are ill-posed and need regularization decisions that compromise resolution versus stability. When inputs are biased or incompletely corrected, inversion can produce smooth, plausible-looking conductivity models that mask uncertainty and can misrepresent conductor geometry, depth, or continuity. In operational exploration, where the cost of false positives and false negatives is asymmetric and context-dependent, reliability under non-ideal conditions is often more important than peak performance on clean case studies.

Machine learning has become attractive in this context because it offers tools for learning complex, context-dependent mappings from data, for extracting representations that can remain stable under non-stationary noise, and for automating decisions across large survey volumes. Several trends have converged to make (Machine learning) ML more viable for TEM than in earlier cycles of interest. TEM data volumes, especially for AEM campaigns, have grown to the point where manual QC and bespoke processing become limiting factors, and where consistent handling across surveys is difficult to maintain. Meanwhile, deep learning has advanced substantially for sequential data, providing architectures such as temporal convolutional networks and transformer-based encoders that match the structure of gated transients, where informative patterns are distributed across time channels and may depend on long-range temporal context. At the same time, the broader geoscience community has increasingly adopted hybrid strategies that combine physical modeling with learning-based components, creating a path to integrate ML without abandoning known physics or interpretability requirements [10,11].

The promise of ML in TEM signal processing, however, is inseparable from new risks, especially when models are deployed under domain shift [12]. Overly aggressive denoising can oversmooth late-time responses and suppress weak but genuine conductors, increasing false negatives in precisely the scenarios where TEM is relied upon most. Conversely, models trained to “repair” corrupted gates may hallucinate structure, creating false positives that propagate into inversion and target ranking. End-to-end approaches that map directly from waveforms to conductivity models or target labels can be fast and appealing, but may generalize poorly across differences in transmitter waveforms, gate schedules, system responses, flight dynamics, and geology. Label scarcity compounds the challenge: reliable ground truth is limited because drilling is sparse and biased toward anomalies, and because many exploration datasets are proprietary. As a result, many studies depend on proxy labels derived from legacy inversions or manual interpretations, which can encode prior biases, or on synthetic training data, which can create misleading performance if the synthetic world is too similar to the testing scenario. Avoiding “inverse crime” conditions and establishing credible evidence, therefore, becomes as important as proposing new architectures.

A pipeline viewpoint will help to understand where the ML can play the most useful role and how it is to be assessed. The TEM process may be considered a series of transductions between raw or less process transients and outcomes, such as data ingestion and metadata harmonization; automated quality control and artifact recognition; noisy and suppressing processes; response system entrapment and normalization; feature extraction or representing learning; inversion or surrogate modeling; and ultimately target detection, prioritization, and prospectivity mapping, frequently in combination with other geoscience input listed layers of magnetics, gravity, geology, geochemistry, or drill data. The introduction of the ML method may occur at any stage, and the correct learning paradigm, as well as the profile of operational risks, varies by stage. When teaching early processing, ML may act as an adaptive filter that uses contextual information to distinguish signal and various types of interference, and self-supervised and weakly supervised approaches are especially applicable since clean targets are scarce. ML can be used in the system correction and normalization process and help in transferring calibration and compensating for drift, thereby minimizing the variability across surveys. In representation learning, ML can generate decay behavior and spatial context line-based embeddings to help in clustering, ranking anomalies, and interpreting downstream. In this review, we will use a pipeline perspective, where TEM acquisition and preprocessing are linked to inversion and exploration decisions, where machine learning can be introduced, and propagation of uncertainty through the stages is important (**Figure 1**)[13,14].

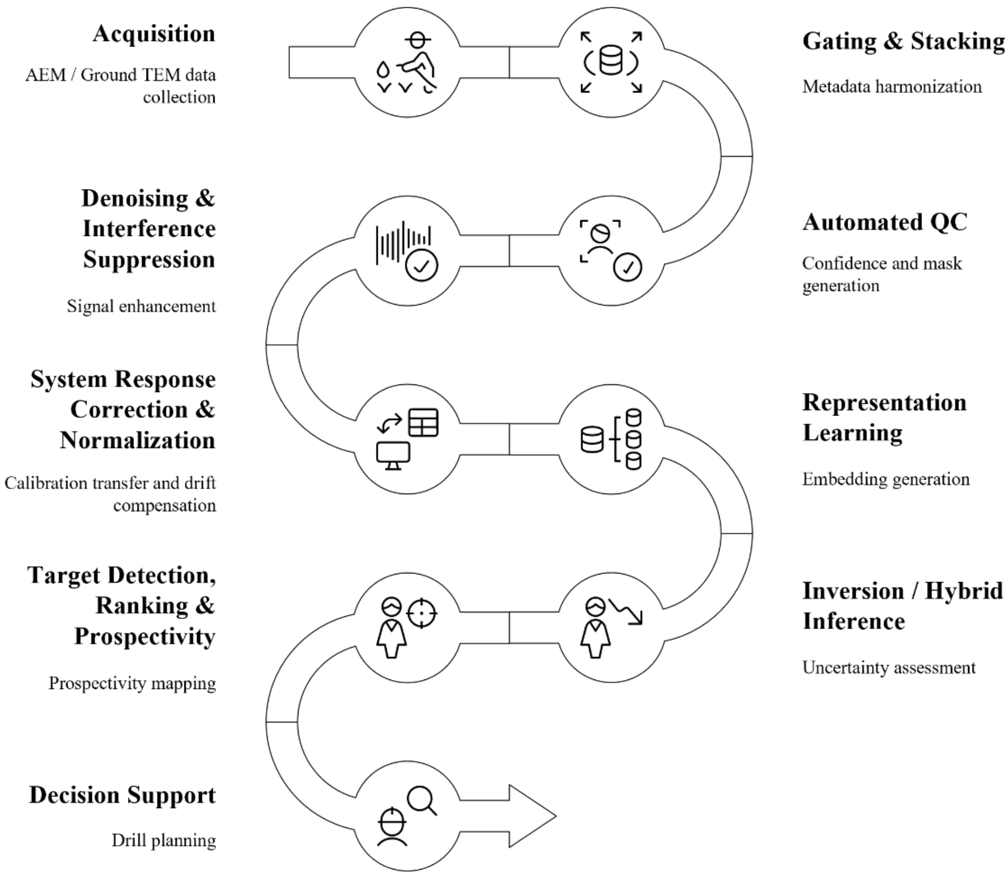

Figure 1. End-to-end TEM–ML workflow and review roadmap.

The role of ML in inversion is also a spectrum of relatively conservative to more discursive approaches[12]. The surrogate forward models are able to substitute costly physics calculations in the iterative process of inversion without altering the layout of well-established inversion procedures, but allow for computing cost reduction. Unrolled optimization or plug-and-play priors can be trained as hybrid strategies that learn useful regularization operators and preserve a structure of misfit to data, and allow the convenient control of stability. Duration-to-point predictive conductivity. Direct networks involving inversion based on transient predictive networks can achieve significant speed benefits, but must treat uncertainties and physical plausibility with significant care, particularly when extrapolating with network training. In all these decisions, exploration decision-making is focused on uncertainty quantification. Uncertainty is hardly one optimal model; practitioners require an exploratory, fueled uncertainty that explains the informative areas, bounds where the depth of exploration is important, and the existence of interpretations varies when substituted by alternative reasonable modeling.

The review is devoted to machine-learning-enhanced TEM signal processing in advanced mineral exploration with a specific emphasis on methods that have a significant effect on reliability, interpretability, and subsequent ultimate decisions, which will inform drilling and resource allocation. The aim, however, is to find a synthesis of how the components of learning can be used in supplementing physics-based processing, in such a way that it will be measurable, reproducible, and operationally safe. Special focus is put on the facts of exploration data: the sparsity and bias of labels, the temptation and the constraints of synthetic data, the abundance of domain shift across systems and regions, and the requirement of evaluation protocols that measure target-level performance, not just pointwise regression accuracy. Arranging the literature with the help of the pipeline taxonomy and focusing on measures based on late-time fidelity, depth-of-investigation awareness, and cost-weighted detection successfully to assess how much work has already yielded results, whereas the pending discoveries are still leading toward the edge, and what methodological and community requirements will be necessary to transform those cases successfully into consistent field application, the review will achieve its goal and help clarify when the ML has already proven a useful tool and when the findings are too young to conclude. Ultimately, TEM sits at a critical intersection of physics-rich sensing and high-stakes, operationally constrained decision-making. Machine learning introduces new capabilities for handling complexity and scale, but it also demands greater rigor in benchmarking, reporting, and risk management [15]. A credible path forward requires hybridization with physical knowledge, careful design of training and test splits that prevent spatial leakage, explicit attention to uncertainty and failure modes, and clear reproducibility standards that allow practitioners to judge whether a claimed improvement is likely to hold for their survey conditions. This review is intended to provide that structure, helping researchers and practitioners align method development with the practical realities of exploration and the scientific expectations of the Science Citation Index (SCI)-standard review article.

2. TEM Signal Characteristics and Classical Processing Baselines

2.1. Measurement Configurations and Data Representations in TEM

Transient electromagnetic methods infer subsurface conductivity by observing the temporal evolution of induced electromagnetic fields following a controlled change in transmitter current [16]. In most exploration implementations, the transmitter current is rapidly ramped down (or up) to approximate a step change, and the receiver measures the resulting transient response as eddy currents diffuse outward and downward through the earth. Because diffusion is scaling dependent, early-time measurements are typically influenced by shallow structure and system geometry, whereas late-time measurements increasingly reflect deeper conductivity contrasts. Ground-based TEM systems are found to have better signal fidelity and localized conductivity anomaly sensitization than airborne systems in some cases of exploration.

There are extensive variations in the survey configurations in ground TEM, downhole TEM, and airborne TEM (AEM). Instances Ground TEM generally applies loop or grounded-wire transmitters and measures time (in se) derivative of the magnetic field (dB/dt) with circular inductors (entering and leaving) or magnetic field B with fluxgate or SQUID-like sensors. Special cases: Ground TEM is typically used to measure the time (in seconds) derivative of the magnetic field (dB/dt) or the magnetic field with fluxgate or SQUID-type sensors. The AEM systems, unlike them, combine transmitter and receiver equipment in a moving platform, typically in towed bird geometries, with further kinematic and aerodynamic effects on the actual measurement. Even though the governing physics of both modalities are similar, the measurement footprint, coupling behavior, and motion artifact susceptibility vary significantly. These models can, however, be simplistic in the representation of the subsurface, whilst the geological conductivity structures can be highly spatially heterogeneous and multi-scale [17].

TEM measurements are normally documented in the form of time gates, averages, or integrals of the receiver signal in pre-defined windows during the transmitter current change. Gating not only reduces the transient to a set of values that are manageable but also ameliorates the signal-to-noise ratio by averaging in time, which oversimplifies the representation, which has been known to vary across instruments and contractors. The dense sampling of gate schedules can be done to sample fast decay or system effects at early times, and then to progressively increase the width of gate schedules at late times, to stabilize measurements at late times when the signal is small. Repetitive transmitting of the transmitter cycle and averaging or vigorous combination of the observed transients is often used to reduce uncorrelated noise by stacking. Gating and stacking collectively represent a viable data product that is neither a time-domain waveform nor a frequency-domain response, but rather a hybrid representation of these, the properties of which are affected by acquisition decisions [18].

In machine learning and current processing terms, metadata becomes as significant as the gated values [19]. The apparent transient can change with transmitter waveform characteristics, timing accuracy, receiver bandwidth, sensor orientation, platform altitude and attitude, and calibration parameters can change even with a subsurface response of the same. This means any effort to predict processing or learning between surveys should clearly consider such system-specific variables instead of assuming that gates are highly similar characteristics that can be generalized. It is further complicated by the fact that most archives are recorded with only partially documented preprocessing, and thus, it is hard to distinguish between geological variation and variances caused by instrumentation and workflow.

2.2. Signal Behavior across Time, Scale, and Geology

The fact that TEM is a diffusive medium means that the transient response is not merely a decelerating curve, but the expression of the spreading of induced currents along conductive pathways, and is perturbed by conductivity dissimilarities. With layered-earth environments, the transient typically has a comparatively smooth decay whose curvature corresponds with the amount of reserves of conductivity and thickness of the layers, whereas the response can possess a typical later-to-time behavior related to target conductance and architecture. Interpretation In exploration practice, subtle variations in the decay slope and the amplitude distribution across gates are often used as the basis of interpretation, as opposed to an individual channel, especially in the process of separating ore-associated conductors and conductive overburden or graphitic horizons [20].

Gates that operate early are particularly sensitive to near-surface conductivity, geometry between transmitters and receivers, and tuning of system response correction. They may harbor useful information on shallow cover, which will affect access and geological context, but they are also the area where instrument bandwidth constraints, transmitter turn-off character, and coupling effects can prevail. The late like gates (on the other hand) tend to be most applicable to deep exploration, but also the point at which the signal is approaching the noise ceiling, and thus susceptible to both random noise and systematic interference. This tradeoff forms a basic processing quandary: harsh noise abortive measures are more commonly used in locations that are pragmatic in need of late-time detail, when, however, fast-paced smoothing or surpassing mistakes might lose feeble mishaps included in the values of interest [21].

The science of geology adds even more confusion by providing the ambiguity of conductive fluids, clays, graphite, and sulfides, with all these having the ability to result in a strong TEM response [22]. On conductive regolith terrains or saline groundwater terrains, the presence of broad areas of conductors on a regional scale will obscure the localized ore-associated anomalies, thus decreasing their detectability and making target ranking more difficult. The conductor orientation, fragmentation, and connectivity in structurally complex areas can result in responses that vary very quickly along the line direction and therefore, the spatial context is important. This is a fundamental cause of the growing use of line-based or neighborhood-sensitive interpretation in the contemporary workflow that does not regard each sounding as an independent one. It is also an incentive to learning-based techniques that may include spatial continuity constraints or take advantage of trends of consistency among neighboring observations.

The system footprint and altitude, particularly in AEM, influence the effective sensitivity distribution with depth and lateral distance, which in turn affects how responses should be compared across surveys and conditions [23]. Two soundings over the same conductor can appear different if the footprint changes, if the platform altitude varies, or if the gating scheme emphasizes different time ranges. These factors are often treated as secondary in simplified analyses, but they become primary sources of apparent variability in large-scale survey datasets. Any robust signal-processing framework must therefore handle not only the transient shape but also the context in which it was measured.

2.3. Noise Sources, Interference Mechanisms, and Acquisition Artifacts

Noise in TEM data is best understood as a combination of environmental interference, platform or instrument artifacts, and processing-induced distortions [24]. Environmental interference includes harmonic contamination from power infrastructure and other cultural sources, which can introduce narrowband components that leak into time-domain gates depending on timing and receiver filtering. Impulsive transients such as sferics can contaminate multiple gates within a single transmitter cycle and may be intermittent in time and space, making them difficult to remove using stationary filtering assumptions. In remote regions, natural variations in the electromagnetic environment can still be significant, and in accessible regions cultural noise can dominate to the extent that late-time information is effectively unusable without advanced interference suppression.

AEM introduces additional artifact classes associated with motion and changing geometry. Variations in altitude, bird swing, sensor orientation, and transmitter–receiver separation can imprint systematic patterns on the transient that resemble geological signals when projected along flight lines. These patterns are often non-stationary and can correlate with topography or flight dynamics, creating confounding relationships that are particularly problematic for data-driven models. Ground TEM avoids some of these motion-related issues but can exhibit coupling artifacts due to imperfect loop placement, variable contact conditions in grounded systems, and local near-surface heterogeneity. In both modalities, saturation, clipping, and timing jitter can distort parts of the transient, producing errors that may not be obvious without careful QC and system-response modeling [25].

Instrument response and calibration drift are persistent concerns because TEM interpretation assumes that the recorded transient accurately reflects the convolution of the earth response with the known system response [26]. If the transmitter waveform deviates from its assumed shape, if the receiver transfer function changes, or if timing calibration is imperfect, systematic biases can occur across gates. These biases may be small in absolute terms but large relative to late-time amplitudes, producing substantial errors after normalization or after transformation into apparent conductivity. Such effects can generate false anomalies or suppress true ones, and they are often difficult to diagnose using only the gated data without auxiliary system diagnostics.

Noise and artifacts also interact with processing choices. Gating and stacking can suppress uncorrelated noise but can smear impulsive events or allow structured interference to persist if it is coherent across cycles [27]. Preprocessing steps such as baseline removal, deconvolution, or filtering can introduce edge effects, phase distortions, or unintended attenuation of signal components if not carefully tuned. In practice, processing pipelines represent a series of compromises, and different organizations adopt different heuristics depending on survey objectives, typical noise conditions, and operational constraints. This variability is a major contributor to inconsistencies across public case studies and motivates the need for well-defined baselines when evaluating new methods.

2.4. Classical Preprocessing and Correction Workflows as Baselines

Classical TEM preprocessing typically begins with steps aimed at stabilizing the measurement: removal of DC offsets, correction for timing and waveform parameters, and stacking of repeated cycles to improve signal-to-noise ratio. Robust stacking strategies, including median-based estimators or outlier-rejecting means, are widely used to reduce the influence of occasional spikes or corrupted cycles. Filtering is often applied to suppress powerline contamination, either through notch filters tuned to expected frequencies or through adaptive schemes that estimate harmonic components and subtract them. For impulsive interference, detection and excision may be used, sometimes followed by interpolation across affected gates, though the extent to which such repairs preserve geological information varies by implementation and remains an important baseline question for ML comparisons [28,29].

System-response correction is a critical step that aims to remove the influence of the transmitter waveform and receiver transfer function so that the resulting transient more closely represents the earth response [30]. In practice, this can include deconvolution, calibration scaling, and correction for instrument-specific factors. Because deconvolution can amplify noise, especially at late times, classical workflows often employ regularization or smoothing to stabilize the corrected signal. The relationship between correction and noise suppression is thus the key one: excessively safe correction may leave artifacts of the system which bias inversion, whereas excessively aggressive correction may harm the already weak late-time signal. Soundings that are usually marked as saturated, dead gates, unusual decay shapes, or a repeat-to-repeat inconsistency are normally identified via quality control procedures, although the thresholds and criteria are usually survey-specific.

Normalization and transformation steps are commonly applied to facilitate interpretation and inversion. Depending on system and convention, data may be converted between dB/dt and B, scaled by transmitter moment, or transformed into apparent conductivity estimates. While these transformations can improve interpretability, they can also magnify systematic biases and noise if underlying assumptions are violated. In particular, apparent conductivity transforms can be unstable in low-SNR regimes and can create artifacts that appear geologically meaningful when they are not [31]. Consequently, classical practice often relies on a combination of visual inspection, heuristic rules, and comparative checks against neighboring soundings to decide which gates and soundings are reliable.

2.5. Classical Inversion Frameworks and Their Limitations under Non-Ideal Data

Inversion converts processed TEM responses into subsurface conductivity models, typically under assumptions about dimensionality and regularity [2]. One-dimensional layered-earth inversion remains widely used because it is computationally efficient and can be applied at scale, particularly in AEM, where the number of soundings can be very large. Laterally constrained inversion extends the 1D approach by encouraging continuity along lines or within grids, which can reduce noise-driven variability and improve geological plausibility. Full 3D inversion is increasingly feasible with modern computing, and it is often the most physically representative approach for structurally complex targets, but it can be computationally expensive and sensitive to preprocessing quality, particularly when late-time gates are noisy or when system-response uncertainties are significant.

All inversion approaches face non-uniqueness and limited resolution, issues that become more severe in the presence of noise and when survey geometry provides incomplete sensitivity. Regularization is therefore essential, but it introduces subjectivity: different regularization choices can yield different conductivity distributions that fit the data comparably well. In mineral exploration, where the goal is often to identify compact, high-conductance bodies rather than to recover a smooth regional conductivity field, standard smoothness-promoting regularization can produce models that under-represent sharp conductors or smear them laterally. Conversely, regularization that allows sharper features can become unstable when data quality is uneven. Depth of investigation further complicates interpretation because the absence of a response at late times may reflect insufficient sensitivity rather than true resistivity, and this ambiguity must be communicated in uncertainty estimates rather than hidden behind a single model [32].

These limitations define the classical baselines that ML-enhanced approaches must respect and surpass. A credible ML contribution cannot be assessed solely by regression error against a reference inversion, because the reference may itself be biased or regularization-dependent. Instead, ML methods should be judged against physics-consistent criteria such as data misfit parity, stability under perturbations, preservation of late-time diagnostic content, and performance on exploration-relevant tasks such as target detectability and ranking under realistic noise. Establishing these baselines is the purpose of this section: it clarifies what the data represent, why they are difficult, how classical pipelines address those difficulties, and where their failure modes create the opportunity and the responsibility for machine-learning-enhanced processing in subsequent sections. The dominant sources of TEM variability and their classical mitigation strategies define the baselines that ML methods must surpass without compromising geological fidelity (**Table 1**) [28,33].

Table 1. TEM data characteristics, artifacts, and classical baselines.

Category	Typical Manifestation in Gated TEM	Primary Causes	Common Classical Baseline Handling	Downstream Risk If Unmitigated
Early-time distortion	Unphysical curvature or offsets in early gates; inconsistent decay onset	transmitter turn-off, receiver bandwidth limits, coupling, timing error	instrument response correction, early-gate muting, robust gating	biased near-surface model; inversion instability; false shallow conductors
Late-time noise floor	Large variance, sign flips, non-monotonic late decay	weak signal amplitude, cultural EM, stacking limits	increased stacking, late-gate widening, smoothing/regularized deconvolution	missed deep conductors; depth bias; uncertain conductance estimates
Harmonic interference	Oscillatory contamination across multiple gates; line-dependent patterns	powerlines, industrial sources	notch filtering, harmonic modeling/subtraction, robust stacking	false anomalies; systematic bias along lines; overconfident interpretation

Table 1. *Cont.*

Category	Typical Manifestation in Gated TEM	Primary Causes	Common Classical Baseline Handling	Downstream Risk If Unmitigated
Impulsive events (sferics)	Spikes affecting subsets of cycles/gates; intermittent corruption	atmospheric discharges	cycle rejection, spike detection + interpolation/inpainting	patchy artifacts; instability in derived attributes; inversion artifacts
Motion/geometry artifacts (AEM)	Coherent line patterns correlated with topography/flight dynamics	altitude variation, bird swing, attitude changes	flight-path QC, smoothing with constraints, line-based QC rules	confounded anomalies; spurious lateral continuity; biased target ranking
Calibration/drift	Slow temporal trend in amplitude; survey-day offsets	temperature, electronics aging, gain drift	reference lines, calibration checks, baseline correction	cross-survey inconsistency; false regional trends; poor transferability
Saturation/clipping	Flattened or truncated gates; nonphysical plateaus	strong near-field response, receiver limits	saturation flagging, gate removal, conservative preprocessing	unrecoverable information; misleading apparent conductivity/inversion

3. Data, Supervision, and Benchmarking Foundations

3.1. Learning-Ready TEM Datasets and the Role of Metadata

Machine-learning performance in TEM signal processing is often determined less by network architecture than by how the data are represented, curated, and contextualized [34]. TEM measurements are not interchangeable vectors of gate amplitudes; they are outputs of a coupled measurement system in which transmitter waveform, receiver bandwidth, timing, geometry, and platform dynamics modulate the observed transient. For this reason, a learning-ready dataset must be treated as a structured object that includes both the gated response and the acquisition descriptors required to make responses comparable across time, lines, and surveys. At minimum, this structure should preserve the gate schedule and timing reference, the measured quantity (such as dB/dt or B), the transmitter moment or current waveform descriptors, and the receiver calibration and transfer-function information when available. In airborne applications, platform altitude, attitude proxies, transmitter–receiver separation, and navigation quality indicators are frequently necessary to distinguish geological variability from acquisition-induced variation.

In practice, many exploration archives are "processing products" rather than raw measurements, and documentation of prior preprocessing can be incomplete. This creates a recurrent challenge for ML: models may learn the idiosyncrasies of a contractor's processing chain rather than generalizable geophysical relationships. Even when nominally raw gates are provided, hidden steps such as baseline correction, proprietary filtering, or gain normalization can alter the statistical properties of the data. A robust review of ML-enhanced TEM must therefore emphasize data provenance. The value of an algorithm cannot be interpreted without knowing whether it was trained and tested on comparable representations, whether key metadata were available and used, and whether the evaluation conditions match the intended deployment setting [35].

The scale and granularity at which data are organized also matter. TEM soundings are strongly correlated along lines and across neighboring lines, especially when system settings remain constant and when geology varies smoothly at the footprint scale. Treating each sounding as independent can inflate apparent performance, because the model may simply exploit local spatial redundancy [17]. Conversely, organizing data at the line, block, or survey level allows evaluation under conditions that better approximate real operational transfer, such as applying a model trained on one area or campaign to another. This structural perspective shapes how training, validation, and testing should be defined, and it motivates the inclusion of spatial indices, line identifiers, and survey metadata as first-class fields rather than afterthoughts.

3.2. Supervision in Mineral Exploration: Labels, Uncertainty, and Bias

Supervision is the central bottleneck in TEM machine learning because "ground truth" is expensive, sparse, and observationally biased. The most reliable labels are typically derived from drill intersections, downhole electromagnetic measurements, conductivity logs, and detailed geological models tied to physical sampling. Yet such labels exist only for a small subset of surveyed locations, and they disproportionately represent anomalies that were considered worth drilling. This selection bias means that the absence of a drilled target cannot be treated as a negative label, and it complicates supervised learning for detection and ranking. In many settings, an ML model trained on drilled outcomes will be learning not only the geophysical signature of mineralization but also the historical decision policy that determined where drilling occurred [36].

As a result, much of the literature relies on weaker forms of supervision. One common approach uses proxy labels derived from legacy inversions, plate models, or manual interpretations by domain experts. While these sources can provide dense supervision, they encode the assumptions, regularization choices, and potential errors of the underlying interpretation workflow. If an ML model is trained to reproduce an inversion product, it may achieve low error while inheriting the same artifacts or smoothing biases that motivate ML enhancement in the first place. Moreover, manual interpretations are not static truths; they reflect uncertainty, alternative plausible models, and context-dependent preferences such as favoring compact conductors for drilling [37]. Treating such labels as deterministic can lead to overconfident models that understate uncertainty and mislead downstream decision-making.

Because exploration labels are sparse, biased, and often indirect, effective TEM ML depends on combining multiple supervision sources with different confidence and coverage; the practical supervision landscape is summarized in **Figure 2** [38].

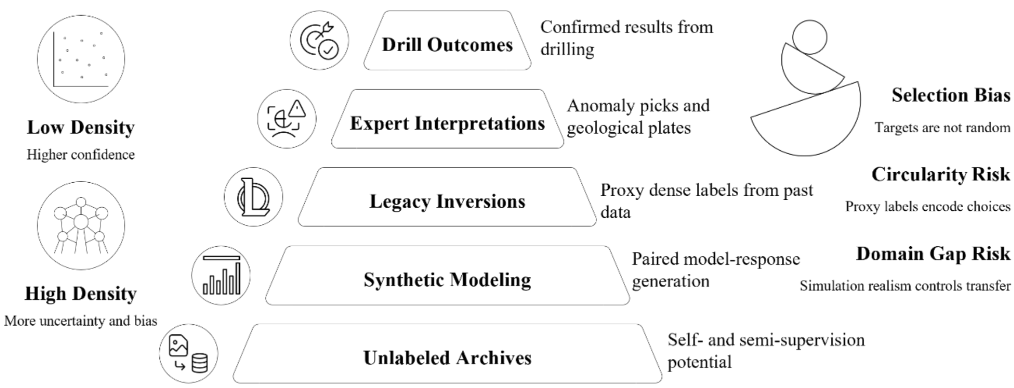

Figure 2. Data and supervision landscape for TEM ML.

A more defensible treatment recognizes labels as uncertain observations with varying reliability. For example, drill-confirmed conductors can be assigned high-confidence positive supervision, while interpreted anomalies can be treated as probabilistic or weak labels. In detection tasks, it is often more appropriate to define supervision at the "target" or "anomaly" level rather than at the single-sounding level, because mineralization is spatially coherent and interpretation is typically based on patterns across multiple soundings [39]. This motivates learning formulations that can handle ambiguous localization, such as weak supervision over segments or neighborhoods, and that can integrate multiple evidence sources without collapsing them into a single binary outcome. For inversion-related tasks, supervision may be better defined through physics-consistent constraints such as matching the observed transient within uncertainty rather than through direct regression to a single reference model.

The scarcity and uncertainty of labels also make semi-supervised and self-supervised learning particularly relevant. In TEM, large volumes of unlabeled data are routinely collected, and many learning objectives can be defined without explicit labels by exploiting structural properties of the measurements [40]. These include consistency across repeated cycles, spatial continuity along lines, invariances under physically plausible transformations, and the predictability of masked or corrupted gates from surrounding context. Such approaches are not merely convenience strategies; they can reduce dependence on biased labels and can promote representations that transfer across regions and system configurations.

3.3. Synthetic Data Generation and the Synthetic-to-Field Gap

Synthetic data are indispensable in TEM machine learning because they provide controllable training signals and allow exploration of rare target scenarios [41]. Forward modeling can generate transients for parameterized conductivity structures, including layered earth models, embedded plates, and volumetric conductors. Synthetic datasets can be expanded to cover a broad distribution of conductance, depth, geometry, and host conductivity, which is difficult to achieve through field labels alone. For inversion and surrogate modeling, synthetic data also enables paired examples of conductivity models and responses, facilitating supervised training that would otherwise be infeasible.

The principal risk is that synthetic datasets can produce misleading performance if they are not representative of the true measurement process. Real TEM data reflect the convolution of earth response with system response, the effects of discretization and approximations in the forward solver, and a combination of noise mechanisms that are seldom captured by simple additive Gaussian assumptions. If synthetic training and testing share the same forward model, discretization, and noise assumptions, models may exploit unrealistically consistent patterns and achieve excellent metrics while failing when confronted with field variability. This issue is often described as an "inverse crime," but in ML it can be more subtle: the crime may occur not only through identical physics between train and test, but through the shared simplifications that make the synthetic world easier than reality [42].

It must be bridged between the synthetic and field at several levels. The forward modeling should represent the dimensionality and complexity that can be anticipated during deployment, and especially in cases where 3D structures or anisotropies affect response. It should include system response, preferably including variability information about the calibration uncertainty, and instrument drift. Harmonic interference, impulsive events, artifacts related to motion (in AEM), non-stationary baseline variations, and non-random perturbations should be included in noise models. Just as important, the assignment of the geological backgrounds is supposed to be realistic. In the case that synthetic datasets are biased towards conductors that are isolated in resistive hosts, the model overfits to clean signature characteristics and fails to produce results in conductive cover or graphitic landscapes, where false positives are usually frequent [43].

A practical approach is to treat synthetic data not as a substitute for field data but as a scaffold for representation learning and robustness training, supplemented by field fine-tuning and domain adaptation. Synthetic datasets can be used to teach models the physics-consistent relationships between transient shape and conductivity structure, while limited field labels and unlabeled field data can be used to adapt to real noise and system variability [44]. Hybrid strategies that incorporate physics constraints during training, such as penalizing predictions that violate expected diffusion behavior or that fail to reproduce observed data within uncertainty, can further reduce dependence on perfectly realistic synthetic priors. Nonetheless, even physics-informed approaches must be evaluated against field behavior, because the most problematic discrepancies often arise from practical measurement effects and survey idiosyncrasies rather than from violations of Maxwellian physics. Because supervision in exploration is heterogeneous and often uncertain, we summarize the main data sources, label types, and associated biases that shape training and credible evaluation in **Table 2**.

Table 2. Data and supervision options for TEM ML.

Data/Supervision Type	What It Provides	Typical Availability	Strengths	Key Limitations/Bias	Best-Fit ML Use Cases
Raw or minimally processed gates + full metadata	Highest fidelity to the measurement process	mixed; often partial metadata	supports physics-consistent learning and transfer	heterogeneous formats; missing system response details	QC, denoising, system harmonization, hybrid inversion
Contractor-processed products	cleaned gates, derived attributes, sections	common in archives	convenient and standardized within a project	hidden preprocessing; may encode vendor-specific artifacts	anomaly detection on derived products; project-specific automation
Drill-confirmed targets/downhole EM	strong positive evidence of conductors/mineralization	sparse and biased	high reliability; operational relevance	selection bias (drilled anomalies); limited negatives	target ranking calibration; decision-aware evaluation

Table 2. Data and supervision options for TEM ML.

Data/Supervision Type	What It Provides	Typical Availability	Strengths	Key Limitations/Bias	Best-Fit ML Use Cases
Interpreted plates/anomaly picks	target locations/geometries interpreted by experts	moderate	dense supervision at the target level	subjective; style-dependent; may not be unique	segmentation/picking; weakly supervised detection
Legacy inversion models (proxy labels)	dense pseudo-ground-truth conductivity	common	cheap dense labels	circularity; inherits regularization bias; may hide uncertainty	pretraining; consistency learning; surrogate forward modeling
Synthetic forward-modeled datasets	paired (model, response) at scale	high (if modeling exists)	controllable; covers rare cases	synthetic-to-field gap; inverse crime risk	surrogate modeling; inversion priors; robustness training
Unlabeled field data	massive quantities for self-supervision	very high	captures true noise + system idiosyncrasies	no explicit targets; needs careful objectives	masked modeling; contrastive embeddings; domain adaptation

3.4. Data Splitting, Leakage Prevention, and Benchmark Task Design

Evaluation in TEM ML is frequently compromised by unintentional leakage driven by spatial correlation [45]. When training and testing samples are interleaved along the same line or within the same survey block, the model can exploit local redundancy and achieve inflated performance that does not translate to new areas. This problem is particularly acute in AEM datasets, where consecutive soundings may differ only slightly and where system settings remain constant across long line segments. Consequently, scientifically credible benchmarking must adopt spatially and operationally meaningful splits. Testing should ideally be performed on withheld regions, withheld lines, or withheld surveys that differ in geology and/or acquisition conditions, reflecting the intended use case of generalization.

Leakage may also be caused by preprocessing and label construction [46]. When labels are based on inversions utilizing all possible data, including data that will be considered as test data, then supervised learning can potentially inherit such information, which would not be obtainable in a realistic prospective environment. On the same note, global normalization statistics that are calculated on the entire dataset are also capable of introducing subtle leakage since a model can exploit the information of the test distribution. These problems are not just methodological technicalities, but they directly influence whether the improvements reported are likely to be applicable in the actual decision in exploration.

Benchmark design should align tasks with exploration value and with the stages of the TEM pipeline [47]. For denoising and interference suppression, benchmarks should assess not only generic error reduction but also preservation of late-time diagnostic content and the stability of downstream interpretation. For inversion acceleration or surrogate modeling, benchmarks should include physics-consistent criteria such as the ability to reproduce observed transients within uncertainty, depth-of-investigation-aware performance, and robustness under perturbations in system response. For detection and ranking, benchmarks should be defined at the target or anomaly level rather than the single sounding level, because exploration decisions are made on coherent features, not isolated samples. Cost-weighted evaluation is also essential, since the operational penalty of a false positive (unnecessary drilling) and a false negative (missing a deposit) are not symmetric and vary by project context.

A mature benchmarking ecosystem would include multiple domains and acquisition regimes, enabling explicit study of domain shift. For example, a benchmark suite might include a late-time recovery task under harmonic interference, an impulsive-event suppression task, a cross-platform transfer task where waveforms and gate schedules differ, and an end-to-end target ranking task that integrates TEM with auxiliary layers. The objective is not to produce a single leaderboard number, but to characterize tradeoffs, failure modes, and the conditions under which an approach is safe to deploy. This orientation is especially important in mineral exploration, where rare but high-value targets dominate project outcomes and where average-case metrics can obscure critical edge cases [48].

3.5. Reproducibility Expectations and Reporting Standards for SCI-Quality Evidence

Because exploration datasets are often proprietary and difficult to share, reproducibility in TEM ML must be supported by transparent reporting of data representations, preprocessing, and evaluation design [49]. SCI-standard evidence requires clarity about what constitutes the model input, which metadata were used, how gates were handled, how missing or corrupted values were treated, and what steps were applied before learning. It also requires explicit documentation of train–test separation, including spatial and operational boundaries, and an explanation of how proxy labels were constructed and what uncertainties they carry. Without this information, it is difficult to interpret whether reported improvements reflect true methodological advances or artifacts of a particular dataset and workflow.

Where open data cannot be released, reproducibility can still be advanced through partial releases such as synthetic benchmark generators, standardized data schemas, and code for preprocessing and evaluation that can be applied within proprietary environments. Reporting should emphasize not only performance but also stability, uncertainty calibration, and sensitivity to domain shift. These practices are especially important for a field where deployment decisions are consequential and where the same method can behave very differently across regions, systems, and noise regimes. Establishing rigorous data and benchmarking foundations is therefore not ancillary to ML-enhanced TEM processing; it is the prerequisite for distinguishing promising ideas from deployable solutions [49].

3.6. Machine Learning Techniques for TEM Signal Processing

Machine learning models are utilized progressively in the practical TEM data processing to solve issues associated with noise suppression and signal reconstruction, as well as feature extraction of transient decay curves. Convolutional neural networks (CNNs) have demonstrated a great ability to detect spatial-temporal variations in multi-channel TEM responses, whereas recurrent neural networks (RNNs) and long short-term memory (LSTM) models have proven useful in modeling the sequential nature of transient decay signals. More recently, time-gated transformer-based models have been investigated to learn long-range dependencies of time-gated electromagnetic responses. The architectures make it possible to extract fine features associated with conductivity that might be hard to see without the traditional signal processing techniques [50,51].

The physics-informed models of machine learning have also become a future of TEM data interpretation. These models can minimize the possibility of non-physical predictions and improve extrapolation to geological situations previously unobserved by incorporating Maxwell equations or forward electromagnetic simulation results in the training process.

4. Machine Learning Methods across the TEM Signal-Processing Pipeline

4.1. Automated Quality Control and Preprocessing

The earliest opportunity for machine learning to add value in TEM workflows is the automation of quality control, because QC decisions determine which gates, cycles, and soundings are trusted by all downstream steps [52]. In practice, QC has traditionally relied on rule-based thresholds, manual inspection of decay shapes, and contractor-specific heuristics that may not transfer well across systems or noise regimes. Machine-learning-based QC reframes this step as a classification or scoring problem in which the model learns to identify patterns associated with known failure modes such as saturation, clipping, timing jitter, coupling anomalies, abnormal baseline drift, or inconsistent stacking behavior. The principal advantage of learning-based QC is not merely speed, but the potential to model complex, context-dependent signatures of corruption that do not reduce to a single threshold on amplitude or variance.

A central design question is whether QC should be gate-level, sounding-level, or segment-level. Gate-level approaches aim to flag specific time windows as unreliable while preserving the rest of the transient, which is attractive for late-time–limited datasets where discarding entire soundings would be costly. Sounding-level approaches instead provide a global accept/reject decision or a quality score that can be used to weight soundings in inversion. Segment-level approaches exploit spatial continuity and platform context, recognizing that many artifacts appear as coherent patterns along lines, especially in airborne data. In review terms, these design choices should be interpreted through an operational lens: a gate-level QC model can preserve information but risks inconsistent masking that destabilizes inversion, while a sounding-level score is simpler but may throw away useful partial information. Increasingly, the literature trends toward probabilistic QC outputs, where each gate or sounding is assigned a confidence value that can be propagated into uncertainty-aware processing and inversion, rather than treated as a binary decision [53].

The supervision challenge is particularly prominent for QC because high-quality labels are rare and subjective. Many studies therefore, use weak supervision, where clear-cut corrupted cases provide positive examples, while the negative class is drawn from data judged "typical" rather than guaranteed clean. Self-supervised approaches also appear naturally: models can learn the expected consistency between repeated cycles, between adjacent soundings, or between overlapping time windows, and flag deviations as potential corruption. In a TEM context, the scientific contribution is strongest when QC models are evaluated not only on classification metrics, but on downstream stability, demonstrating that quality scoring improves denoising behavior, inversion convergence, or target detectability relative to classical rule-based QC [21].

4.2. Denoising and Interference Suppression

Denoising is the most visible application of ML in TEM because noise and interference directly limit late-time interpretability and depth sensitivity [54]. Unlike many generic denoising problems, TEM denoising must preserve subtle decay characteristics that encode conductor depth and conductance, and it must do so across an intrinsically non-stationary signal. Consequently, the central question is not simply whether noise variance is reduced, but whether the denoised transient remains physically plausible and operationally faithful to the underlying response.

Supervised denoising approaches typically require a notion of a "clean" target signal, which is difficult to obtain in the field. Some studies approximate clean signals through high-stack averages, through manual removal of corrupted cycles, or through synthetic forward-modeled data. Under these regimes, sequence models such as temporal convolutional networks, recurrent networks, and transformer encoders are used to map noisy gated sequences to clean estimates. Architectures that operate across time gates can exploit long-range dependencies, which is valuable when interference affects multiple gates in correlated ways. However, supervised denoisers must be treated with caution because they can learn to impose priors that over-smooth late-time behavior, effectively reducing variance at the cost of suppressing weak but meaningful responses [55].

Self-supervised and blind denoising methods reduce dependence on clean targets by exploiting redundancy and structure [56]. In TEM, redundancy arises from repeated transmitter cycles, from spatial continuity along lines, and from the predictability of masked gates given surrounding gates and context. Masked modeling approaches, in which the model learns to reconstruct withheld gates, are particularly natural for gated data and can be extended to inpainting tasks where impulsive events remove portions of the transient. Noise2Noise-style learning can be adapted when independent noisy realizations exist, such as different stacks or subsets of cycles; the model learns a mapping that converges to the expected clean signal under assumptions about noise independence. The review-level insight is that these methods are most compelling when they explicitly respect TEM-specific constraints, such as monotonic decay expectations in certain regimes, boundedness and causality constraints linked to system response, or consistency across adjacent soundings.

Interference suppression often benefits from explicitly modeling the interference mechanism rather than treating it as generic noise [57]. Harmonic contamination from power infrastructure, for example, is often more effectively handled by combining parametric harmonic estimation with ML residual learning than by end-to-end denoising alone. Impulsive sferics events similarly lend themselves to a two-stage approach: detection of corrupted cycles or gates, followed by repair via interpolation, learned inpainting, or robust aggregation across repeats. In airborne settings, motion-related noise is frequently correlated with auxiliary channels such as altitude, attitude proxies, or navigation quality; multi-input models that incorporate such context can outperform models that only see the gated transient. Across all categories, the key evaluation requirement is to demonstrate that denoising improves exploration-relevant outcomes late-time fidelity, inversion stability, and anomaly detectability rather than only pointwise error reduction against a proxy reference. To maintain a consistent comparison across studies, we map ML formulations, model families, outputs, and common failure modes to each processing stage in **Table 3**.

Table 3. ML methods mapped to the TEM pipeline.

Pipeline Stage	Typical ML Formulations	Representative Model Families	Outputs	What Success Should Mean (Practical Criteria)	Common Failure Mode
Automated QC & preprocessing	classification, anomaly detection, confidence scoring	gradient-boosted trees, CNN/TCN/Transformers, autoencoders	gate masks, sounding score, line-level flags	fewer corrupted gates entering inversion; improved repeat-line consistency	over-flagging late-time gates; discarding subtle targets
Denoising & interference suppression	supervised/weakly supervised denoising; self-supervised masking; detection+inpainting	TCN/UNet-1D, Transformers, diffusion/inpainting models	denoised gates, repaired segments, interference labels	late-time fidelity preserved; anomaly detectability improves	oversmoothing; hallucinated signal; bias under domain shift
System correction & normalization	learned deconvolution; domain-invariant mapping; conditional normalization	conditional nets, adversarial adaptation, normalizing flows	corrected response, harmonized representation	cross-survey comparability improves without removing geology	removal of true geology mistaken for drift; noise amplification
Representation learning	contrastive/self-supervised embeddings; spatial-context learning	Transformers, contrastive encoders, and graph neural nets	embeddings, cluster IDs, similarity measures	stable clustering of geology/targets across noise and campaigns	embeddings capture system identity instead of geology
ML-assisted inversion	surrogate forward models; direct inversion; hybrid/unrolled solvers	emulators, unrolled nets, plug-and-play priors	conductivity model, conductance/depth proxies, uncertainty	data-misfit parity + DOI-aware uncertainty; faster yet trustworthy results	confident but wrong models; poor extrapolation beyond training
Target detection & ranking	segmentation, MIL/weak supervision, multi-modal fusion	CNN/Transformer + attention fusion, GNNs	target probability, ranked anomalies	target-level precision/recall and drill-hit-rate improve	confounding with survey patterns; biased ranking due to labels

4.3. System-Response Correction, Drift Compensation, and Normalization

System-response correction occupies a delicate position in the TEM pipeline because it can substantially improve interpretability while also amplifying noise if applied aggressively [58]. Classical approaches treat correction as deconvolution or calibration scaling based on measured or assumed transfer functions. Machine learning enters here in several distinct ways. In one direction, ML is used to learn data-driven approximations to deconvolution operators, effectively learning to undo system filtering while regularizing the result through learned priors. In another direction, ML is used to harmonize data across instruments and campaigns, reducing variability induced by different gate schedules, waveforms, or calibration states. These objectives are closely related to domain adaptation: the goal is to map different measurement regimes into a common latent representation in which geological variability dominates and system variability is minimized.

The drift compensation and transfer of calibration are of particular interest in exploration where a survey can take weeks or months, where the instruments are reconfigured, or when the data is supplied by multiple contractors. A model of learning-based drift correction will usually either exploit repetitive patterns in background responses or make use of reference lines and calibration values whenever feasible. Drift in air data may have interactions with flight dynamics, and a stack, which may result in complex signatures. This encourages mixed approaches involving physical conditioning of expectations, including smooth drift conditioning with time, and adaptive residual models; these are learned using data. The scientific difficulty is to make it identifiable: drift correction must avoid weakly recreating artificial consistency or introducing falsely strong geological patterns. This means that strong evidence needs to be evaluated in such a way as to distinguish between temporal drift and spatial geology—e.g., by testing on withheld regions, or by showing that the corrected data are more consistent with other independent measurements (such as repeat lines, downhole EM, or other geophysical measurements). Normalization choices can also be a hidden source of inconsistency. Scaling by transmitter moment, converting between dB/dt and B, or transforming to apparent conductivity all alter the statistical properties of the input and can change the behavior of learning models. ML pipelines increasingly treat normalization as part of the model, either by learning invariant representations or by explicitly conditioning on system descriptors. This is an important shift: rather than forcing all surveys into an imperfect common scale through ad hoc preprocessing, modern approaches aim to let the model learn how to interpret data given a known system context, which can improve transfer but also increases the importance of metadata completeness [59,60].

4.4. Representation Learning and Feature Extraction for TEM

Beyond direct denoising and correction, ML contributes by learning representations of TEM responses that summarize the transient in ways useful for interpretation, clustering, and detection. Classical feature engineering often uses decay slopes, time constants, moments, or apparent conductivity transforms, which are interpretable but may be brittle under noise or under varying system configurations. Learned representations aim to capture discriminative information while being robust to nuisance variability such as mild calibration differences, gate schedule variations, or moderate interference [61].

Unsupervised and self-supervised representation learning is particularly well matched to TEM because unlabeled data are abundant and because there are natural "pretext" tasks that encourage meaningful embeddings [62]. Models can be trained to predict masked gates, to distinguish true neighboring soundings from shuffled pairs, or to enforce consistency between different stacks or augmentations of the same transient. Contrastive learning is especially attractive when spatial context is available: adjacent soundings over similar geology can be treated as positives, while distant soundings are negatives, encouraging embeddings that reflect geology rather than noise. In airborne surveys, line-based encoders can incorporate the fact that anomalies appear as coherent patterns along flight direction, and graph-based approaches can incorporate neighborhood connectivity across lines, which is relevant for interpreting targets with complex geometry or limited line spacing.

A persistent challenge is interpretability. Learned embeddings are valuable operationally only if they can be connected to physical meaning or at least to stable, explainable behavior. Methods such as saliency over time gates, prototype-based explanations, or latent-space traversals can help establish what aspects of the transient drive model decisions. In an SCI-standard review context, interpretability should not be treated as an optional add-on, because exploration decisions require traceability: practitioners must be able to understand whether a learned feature corresponds to late-time conductor response, early-time coupling, or a noise artifact. Representation learning is therefore most compelling when coupled with validation that the learned features correlate with independent physical indicators such as conductance estimates, depth proxies, or repeatability across survey conditions [63].

4.5. ML for Inversion, Surrogate Modeling, and Hybrid Inference

Inversion is where ML can offer the greatest potential speedups and the greatest risks. The conservative end of the spectrum uses ML as a surrogate forward model, replacing computationally expensive physics solvers with trained emulators that predict TEM responses for given conductivity models. These surrogates can accelerate iterative inversion by reducing the cost per forward evaluation, enabling larger-scale 3D workflows or more extensive uncertainty analysis. The primary scientific concern is error control: surrogates must generalize across the relevant model space, and their approximation error must be quantified so that inversion does not converge to models that fit an emulator rather than the true physics. In practice, this pushes research toward uncertainty-aware emulators, active learning strategies that refine training samples where the emulator is uncertain, and hybrid schemes that periodically verify emulator predictions with a high-fidelity solver [64].

More direct approaches attempt to learn inverse mappings from transients to conductivity structures [65]. One-dimensional inversion networks are the most common because 1D labels can be generated synthetically and because 1D inversions are widely used operationally. These models can provide rapid estimates of layered conductivity, depth-to-conductor proxies, or conductance-related parameters. However, because the inverse problem is non-unique, direct inversion networks must be interpreted as learning a conditional prior rather than recovering a unique truth. This is not inherently problematic; regularized classical inversion also encodes priors, but it makes uncertainty quantification essential. Without uncertainty, direct inversions can produce deceptively crisp models that understate ambiguity, particularly in conductive cover or when late-time data are noisy.

The hybrid inversion strategies are designed to compromise between speed, stability, and physical consistency [65]. Unrolled optimization, such as those that learn some part of an iterative solver (usually the regularization or proximal step), does not eliminate an explicit data misfit term. Plug-and-play approaches consider a learned denoiser to be an inversion, such that, in accordance with data-driven structure, it may be added without jeopardizing the readability of physics-based fitting. Physics-informed neural networks, which use the physical laws as a subset of the loss functional, are trained to produce physical phenomena that meet the physical relationship requirements. The most effective hybrid methods in TEM are generally those that do not ignore the measurement process, both system response and gating, and also those that do not depend on noises being implicitly clean. Across all inversion-related ML, evaluation should be grounded in criteria that reflect exploration use. Data misfit parity is important: a learned model should not only look plausible but should reproduce measured responses within estimated uncertainty. Depth-of-investigation awareness is similarly critical: models should communicate where the data constrain conductivity and where they do not. Finally, uncertainty calibration must be tested, not assumed. Ensembles, Bayesian approximations, or quantile regression approaches can provide predictive distributions, but their credibility depends on calibration checks against withheld field regions, repeat lines, or independent information such as downhole EM. In a review framing, these considerations separate demonstrative feasibility studies from methods with a realistic path to deployment [66].

4.6. Target Detection, Ranking, and Prospectivity Mapping

Detection and ranking are the points where TEM signal processing meets exploration economics [67]. Traditional workflows often involve interpreting anomalies on conductivity sections, fitting plates, and ranking targets based on conductance, geometry, and geological context. ML approaches can be positioned either as decision aids that operate on processed or inverted products, or as more end-to-end systems that predict target likelihood directly from transients with minimal intermediate interpretation. The former are generally easier to validate and integrate because they operate on familiar representations, while the latter can, in principle, learn to exploit subtle patterns that are lost during classical processing, but require careful safeguards against spurious correlations.

When ML operates on inverted sections or derived attributes, it often takes the form of anomaly segmentation, clustering, or classification. These approaches can standardize interpretation across large datasets and can incorporate contextual layers such as magnetics, gravity, mapped geology, or geochemistry. Multi-modal fusion is particularly relevant for prospectivity mapping, where TEM contributes conductivity constraints while other layers capture structure, alteration, or fertility indicators. Modern fusion approaches range from simple feature concatenation to attention-based models that learn which modalities are informative in different contexts, and to graph-based models that represent spatial relationships among soundings, geological units, and drillholes [68].

End-to-end detection models seek to map gated transients (and possibly line context) directly to target likelihood, depth proxies, or ranking scores. Their appeal lies in speed and in the possibility of learning task-specific representations optimized for discovery rather than for inversion fidelity. The primary risk is that the model may learn confounded signals correlated with historical drilling or survey conditions, especially when supervision is derived from biased drill outcomes [69]. For this reason, rigorous evaluation must emphasize target-level metrics and cross-region generalization. In exploration, the relevant unit is not the sounding but the target, and the relevant question is whether the model improves the hit rate of ranked targets under realistic drilling budgets. Cost-weighted evaluation is therefore central: a small number of high-confidence false positives may be acceptable if they lead to discoveries, while systematic suppression of subtle targets is not.

In prospectivity settings, uncertainty again becomes a core requirement. Ranking without uncertainty can encourage overconfident decision-making, whereas calibrated uncertainty enables risk-aware portfolio selection, where targets are chosen not only for expected value but for diversity and information gain [70]. Models that provide well-calibrated confidence intervals or that explicitly detect out-of-distribution conditions, such as new geological domains or new system configurations, offer clearer pathways to operational adoption. Across the detection and mapping literature, the strongest contributions are those that link algorithmic outputs to exploration decisions and demonstrate measurable improvement in workflows, rather than only reporting classification metrics on narrowly defined datasets.

Across the TEM pipeline, the most robust evidence for ML value tends to appear where the task aligns with abundant data and where evaluation can be grounded in repeatability and downstream stability [71]. Automated QC and interference suppression often meet these criteria because they can be validated against consistency across repeats, improvements in inversion convergence, and reductions in obvious artifacts. Representation learning is increasingly promising, particularly for organizing large survey volumes and supporting transfer, but its value depends on interpretability and on clear demonstrations that embeddings improve target identification or reduce false positives.

Inversion and end-to-end detection remain the most exciting and the most sensitive to pitfalls. Their success depends on realistic training distributions, careful handling of domain shift, explicit uncertainty estimation, and evaluation protocols that avoid leakage and proxy-label circularity. A consistent message across emerging work is that ML is most effective when it is integrated into physics-based workflows rather than positioned as an opaque replacement. This integration can take many forms confidence-weighted inversion using ML QC scores, hybrid denoising with physics constraints, or surrogate modeling inside conventional inversion loops but the underlying principle is the same: TEM is a physics-defined measurement, and learning-based methods must be assessed by how well they preserve or enhance that physics in the service of exploration decisions [72].

5. Validation, Generalization, and Deployment in Exploration Workflows

5.1. Evaluation Objectives Aligned with Exploration Decisions

Validation in machine-learning-enhanced TEM processing must be framed around the decisions the data are meant to support [73]. Unlike many ML applications, where average predictive accuracy is the central criterion, mineral exploration is dominated by asymmetric costs and rare, high-value outcomes. A denoiser that reduces mean error but occasionally suppresses weak late-time responses may be unacceptable if those rare failures correspond to subtle but economically meaningful conductors. Similarly, an end-to-end target-ranking model that performs well on a curated test set may fail prospectively if it has learned project-specific confounders. For these reasons, evaluation must be multi-layered, connecting signal-level metrics to interpretation stability, and interpretation outputs to operational outcomes such as target detectability and ranking efficiency.

At the signal level, improvements should not be assessed solely by pointwise deviation from a reference transient, because reference signals are often proxies produced by stacking, manual cleaning, or proprietary processing. Instead, a credible evaluation should test whether processing preserves the physically meaningful degrees of freedom in the transient, particularly those expressed in late-time gates that control depth sensitivity. Repeatability gives a very good empirical anchor i.e., when two traverses through the same line or repetitions of measurements on the same station give similar results after the processing, this is a very big confirmation that what the method is doing is not removing the signal, but rather removing noise. Where repeat data are unavailable, consistency can be approximated through internal redundancy, such as comparisons between independent stacks, between adjacent soundings in geologically quiet zones, or between responses conditioned on stable system settings [74].

At the inversion and interpretation levels, evaluation should be grounded in physics-consistent criteria [75]. A processed transient that looks plausible but yields an inversion that does not match measured data within uncertainty is not operationally trustworthy. Conversely, a method that improves inversion stability and reduces nonphysical artifacts in conductivity sections is valuable even if it changes the transient relative to a proxy "clean" reference. Interpretation-level evaluation must also consider depth of investigation, because apparent improvements can be an artifact of smoothing or regularization that hides uncertainty. The most informative validation, therefore combines data misfit parity, sensitivity-aware measures, and stability tests under perturbations of preprocessing, noise, or system response assumptions.

Finally, at the exploration decision level, performance should be evaluated in terms of target-level outcomes [2]. This includes whether ranked anomaly lists improve hit rates under a fixed drilling budget, whether false positives are reduced in conductive cover settings, and whether subtle targets become more consistently detectable across surveys. Such evaluation is inherently context dependent, but a review of the field should emphasize that meaningful success in exploration is measured by improved decision quality, not only by improved numerical metrics on intermediate representations.

5.2. Metrics and Protocols for Denoising, Correction, and Inversion

For denoising and interference suppression, the central requirement is to quantify both noise reduction and signal preservation [76]. Noise reduction can be expressed through measures such as variance reduction in geologically quiet regions, suppression of known interference signatures, or improvements in stack consistency. Signal preservation is harder but more important; it can be assessed by the stability of late-time decay characteristics, by the persistence of anomalies across adjacent soundings, and by agreement with independent evidence such as downhole EM, repeat lines, or known conductors. Because TEM responses are non-stationary, evaluation should be gate-aware and should explicitly focus on late-time behavior, where processing often has the strongest effect and where exploration value is frequently highest.

System-response correction and normalization require evaluation protocols that separate true geological variability from instrument-induced differences. Cross-campaign harmonization can be assessed through repeat lines flown with different system configurations, through calibration sites, or through comparison of background responses in stable lithologies. Drift correction should be tested by demonstrating improved temporal stability without introducing spatial smoothing that could remove geological contrasts. Because deconvolution and correction can amplify noise, evaluation should include robustness tests under varying noise levels, and should report how correction affects the distribution of uncertainty across gates [77].

For inversion-related methods, evaluation must acknowledge non-uniqueness and the role of priors. Comparing a learned inversion directly to a classical inversion does not establish superiority unless the comparison is controlled for data misfit and regularization strength. A more defensible protocol evaluates whether methods achieve comparable or improved data fit while producing models that are more consistent with independent information and more stable under perturbations. Depth-of-investigation-aware evaluation is essential: errors should be reported preferentially in regions where sensitivity is demonstrably non-negligible, and uncertainty should increase where sensitivity is low. When ground truth is available via drilling or downhole EM, evaluation should emphasize target-relevant attributes such as depth to top of conductor, conductance proxies, and lateral continuity, rather than voxel-wise error against an uncertain reference model [78].

Uncertainty quantification must be validated explicitly. Predictive intervals and confidence scores should be calibrated, meaning that stated probabilities correspond to observed frequencies under withheld test conditions. In TEM, calibration is particularly challenging because test conditions often differ from training conditions, and because label noise is substantial. Nevertheless, without calibrated uncertainty, ML-enhanced workflows risk producing outputs that appear precise and thereby over-influence decisions. A strong evaluation, therefore, includes reliability diagrams or coverage tests for predicted intervals, as well as stress tests under domain shift [79].

5.3. Generalization under Domain Shift: Systems, Settings, and Geology

Domain shift is the primary obstacle to deploying ML in TEM, because real exploration workflows encounter variability across instruments, waveforms, gate schedules, survey geometries, and geological environments [80]. Even within a single region, system settings may change due to operational constraints, instrument upgrades, or contractor differences. In airborne surveys, changes in altitude, footprint, and flight dynamics can alter responses in ways that are difficult to distinguish from geological signals unless the model is explicitly conditioned on relevant metadata. In ground surveys, loop geometry, cultural noise conditions, and contact variability can introduce similarly consequential shifts. Geological domain shift is often even more challenging: a model trained in resistive terranes with isolated conductors may fail in conductive cover where broad background conductivity dominates, and anomalies are subtle.

Strict methods of generalization start at the evaluation design. Withheld regions or withheld surveys that vary meaningfully with training data should be tested, but not random bifurcations that maintain spatial correlation. Where feasible, leave-one-survey-out/leave-one region-out/evaluation is a more articulate demonstration of future performance. Comparison across waveform and gating variations should also be considered important, as they occur in operational systems and may invalidate models that implicitly assume that the representation is fixed [81].

The combination of explicit conditioning, domain-invariant representation learning, and adaptation can be used to enhance generalization methodologically. Conditioning supports the model with system descriptors, such as gate times or waveform parameters, and altitude or other metadata, to tell the transient in context instead of making the gates a fixed semantic location. Domain-invariant learning tries to generate embeddings that become the least reliant on system identity without losing geological data; however, that needs to be done with care since the suppression of domain information can also lead to the suppression of actual geology-domain interactions. Strategies of adaptation are fine-tuning small-sized labeled datasets in the new domain, semi-supervised adaptation with unlabeled field data. Test-time adaptation, along with normalization-based methods, could also be used to reduce the mismatch, although difficulties on both operational and instability issues can be experienced unless well supervised [82].

A particularly important deployment consideration is out-of-distribution (OOD) detection. Even strong models will encounter conditions outside their training envelope, such as new noise regimes, unusual geology, or unexpected instrument behavior. OOD detection mechanisms that flag low-confidence predictions or distributional anomalies can prevent silent failures and encourage human review. In exploration, such safeguards can be more valuable than marginal gains in average accuracy, because they protect against rare but costly misinterpretations [83]. Given pervasive domain shift across instruments, waveforms, noise environments, and geology, we recommend a layered evaluation and deployment framework that progresses from signal fidelity to physics consistency, transfer testing, uncertainty calibration, and decision-level utility (**Figure 3**) [84].

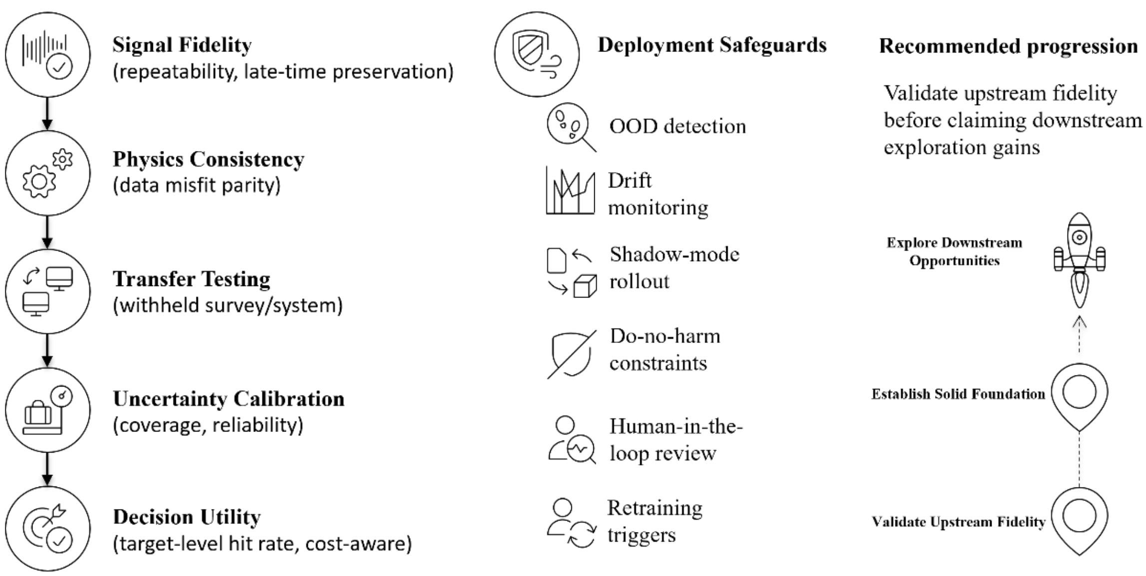

Figure 3. Evaluation and deployment framework under domain shift.

5.4. Reproducibility, Transparency, and SCI-Standard Reporting

SCI-standard evidence in this field depends on clear reporting that allows readers to assess whether a method is likely to transfer beyond the reported case studies. Because proprietary constraints often prevent open release of full field datasets, transparency must be achieved through detailed descriptions of data representation, preprocessing, and evaluation splits. Inputs should be specified unambiguously, including whether data are raw gates, contractor-processed products, or transformed attributes; what system-response corrections have already been applied; how missing or corrupted gates were handled; and what metadata were used. Evaluation design should be fully described, including the spatial boundaries between train and test, the rationale for those boundaries, and any differences in system configuration or geology between splits [85].

Reporting must also avoid circularity. If labels are derived from inversions that used the entire dataset, or if the method is evaluated against proxy labels generated by a workflow similar to the method itself, then the evaluation may measure consistency rather than correctness. In such cases, results should be framed as reproducing an interpretation style rather than recovering ground truth, and the limitations should be explicit. Similarly, when synthetic data are used, authors should report how synthetic realism was assessed and how train/test separation avoids shared simplifications that inflate performance [86].

A mature reporting standard should include ablation studies that isolate the contribution of each component, comparisons to classical baselines under equivalent conditions, and uncertainty characterization [87]. Computational cost should also be reported, because practical deployment often requires processing millions of soundings and integrating outputs into existing software ecosystems. Without these details, even promising methods remain difficult to evaluate and adopt. Given the high risk of performance inflation from spatial leakage and domain shift, we propose a minimum evaluation and deployment checklist suitable for SCI-standard evidence (**Table 4**).

Table 4. Evaluation protocol and deployment.

Evaluation Dimension	Recommended Protocol	Metrics That Matter	Minimum Reporting Requirement	Why It Matters for Exploration
Train/test separation	spatially blocked splits (by region/survey), not random soundings	performance drop under withheld-survey tests	explicit split map + rationale	prevents inflated scores from spatial leakage
Late-time preservation	test on low-SNR regimes; compare anomaly detectability pre/post	gate-wise fidelity, repeatability, anomaly persistence	late-time-focused analysis, not only aggregate error	late-time gates control depth sensitivity and drill targeting
Physics consistency	verify corrected/denoised outputs can be forward-modeled/inverted with comparable misfit	data misfit parity, residual structure	misfit plots and residual statistics	avoids "pretty" signals that violate measurement physics
Domain shift robustness	cross-waveform/gate schedule and cross-instrument tests	performance under transfer, OOD rate	description of system differences in test data	most field failures occur under a distribution shift
Uncertainty calibration	coverage tests; reliability diagrams; stress tests under shift	calibration error, interval coverage, ensemble spread	calibrated uncertainty, not only the point estimate	enables risk-aware target selection
Operational utility	target-level evaluation with cost-aware decision metrics	target PR, hit-rate at K, cost-weighted utility	definition of "target" unit + budget assumptions	aligns evaluation with drilling economics
Deployment monitoring	drift detection + retraining triggers; shadow-mode validation	shift statistics, alert rate, post-deployment KPI	monitoring plan and failure handling	prevents silent degradation in long campaigns

5.5. Deployment Constraints and Human-in-the-Loop Integration

Field deployment introduces constraints that are not always visible in academic studies. In airborne operations, near-real-time QC and processing can be valuable for adjusting survey plans, detecting instrument failures, and prioritizing areas for infill lines. In such settings, latency and compute footprint matter, and models must be robust to incomplete information and evolving noise conditions. Offline processing, by contrast, can leverage heavier compute and more elaborate inversion, but it must still integrate with established interpretation tools and deliver outputs in formats that geophysicists can trust and interrogate [88].

The general configuration was that in order to build a practical deployment architecture, physics-based baselines with ML modules are used, but in the form of a decision assist as opposed to black-box modules. As an example, the ML-based QC can generate gate-level scores of confidence that regulate the robust stacking/weighting of the inversion. ML-based denoising can be limited to physically plausible constraints and can be accompanied by diagnostics indicating the extent to which the signal was distorted. The hybrid inversion schemes are able to give a fast ML estimate in addition to a slow physics-based verification step of high-value anomalies. These trends can be seen as an extension of a larger rule: in high-stakes exploration processes, ML can be adopted in the most acceptable way when it increases efficiency and consistency, maintains interpretability, and provides clear indicators of failure [89].

It is also necessary to have model monitoring. The data drift may arise when the instruments become old, when there are varying conditions of the surveys, or when the exploration enters new geological areas. The distribution of inputs can be monitored, and the shift in QC flag rate, or the shift in the model uncertainty, can be monitored, and activate review or retraining when the threshold is met. The retraining processes should be structured such that bias is not reinforced, especially where new labels are based on the outputs of the model. This risk can be addressed by human-in-the-loop workflows through expert review as well as active learning to emphasize the most informative new labels instead of retraining indiscriminately on model-selected data [90].

5.6. Risk Management and Failure Modes in ML-Enhanced TEM Processing

One of the major differences between deployable TEM ML and proof-of-concept studies is the management of risk. Over-suppression of the weak late-time signal, which can obscure the small conductors and bias depth measures, is the most frequent mode of failure in denoising. In system correction, system failures can include the elimination of geological trends that have been confused with drift or the enhancement of noise by excessively vigorous deconvolution [91]. In inversion, failures may take the form of confident but wrong structural formations of conductivity that are consistent with proxy goals but are not representative of the actual span of realistic models. Confounding factors, e.g., learning correlating with line spacing, altitude, or historical drilling patterns, rather than conductor physics, are often the cause of failure in detection and ranking.

These risks need both methodological and procedural safeguards to prevent them [92]. Uncertainty estimation, OOD identification, and physics-consistency checks can be considered as methodological protective measures that limit the results in such a way that these results are still consistent with measured data, and also per the dark-side effects of the system. Procedural protections involve staged deployment and shadow-mode testing, whereby the outputs of the ML are compared against known workflows without affecting decisions until performance has been established, and the incorporation of so-called do-no-harm constraints into the model, restricting the amount of modification the model is permitted to perform on a transient. On an SCI-standard review, these safeguards are to be viewed as being part of the core of the field, since the way between better metrics and better discoveries is whether or not the models act safely in the messy and changing conditions of actual exploration.

To conclude, validation and deployment cannot be considered downstream issues but are some of the characteristics of ML-based TEM signal processing. The most attractive ones are those that show beneficial behavior in realistic domain shift, quantify uncertainty, prevent evaluation leakage, and make their integration into workflows in a manner that complements and does not replace expert reasoning. These necessarily precondition the final synthesis, in which the capabilities of the field will be opposed by the existing gaps and the priorities of further study.

6. Discussion

Combining machine learning with transient electromagnetic exploration offers considerable potential in enhancing the interpretation of subsurface, although there are many technical issues. In contrast to most traditional ML applications, TEM data have high physical constraints subject to electromagnetic diffusion, signal decayed by time, and artifacts depending upon acquisitions. Consequently, ML approaches should be constructed in a way that does not interfere with the physics of the problem.

The key weakness is posed by the TEM measurements structure. System turn-off effects, transmitter-receiver coupling, and near-surface conductivity contrasts are commonly known to affect the early-time gates, whereas the late-time gates have lower signal-noise ratios but are more sensitive to more deeply-conducted structures. The ML models that are trained on full-time series data should therefore learn to discriminate between physically relevant behavior of decay and acquisition artifacts. In the absence of relevant preprocessing or physics-informed architectures, models can mistakenly predict that noise or instrumentation effects are present in the subsurface.

The other challenge is with regard to the quality and availability of training data. Field TEM datasets are usually incompletely labeled, have little ground truth, or are sparsely represented with boreholes or other geophysical surveys. This problem can be partially overcome by synthetic training data produced by forward modeling, though overly simplistic geological models can be biased and restrictive to any form of generalization. One promising way to go seems to be the hybrid training strategies involving integration of stimulated responses with field measurements.

Adaptable model interpretability is also highly essential in the adoption of exploration workflows. Although deep neural networks are able to learn highly nonlinear dependencies in the relationship between TEM responses and subsurface conductivity distributions, they should behave in a similar manner to the electromagnetic theory. The current approaches that are being developed are physics-informed neural networks, constrained inversion networks, and surrogate forward models, which are trying to incorporate the physical laws directly into the learning.

Practically, it is unlikely that in the near future, the traditional inversion will be substituted by ML methods. Rather, they can be considered as tools that are complementary and help with data processing faster, detecting anomalies, initializing inversion, or screening large survey data rapidly. ML methods can be used to enhance efficiency, but at the same time, be physically reliable when introduced thoughtfully into current geophysical workflows.

Future studies are thus suggested to be directed into three main directions: (1) formation of physics-driven ML structures that can be adjusted to TEM data properties, (2) formation of standard benchmark datasets that comprise both artificial and field measurements, and (3) systematized assessment structures that can compare ML-based interpretation techniques with conventional inversion procedures. The solution of these issues will be crucial to transferring the achievements in methodology into effective exploratory gains.

7. Conclusion

Machine learning techniques are also considered as a means of analyzing transient electromagnetic data and enhancing the characterization of the subsurface conductivity of the ground. This review discussed the key types of ML methods used in the exploration of TEM, such as supervised learning, deep neural networks, and physics-guided hybrid models. It is analyzed that ML can be beneficial in data classification, anomaly detection, fast inversion approximation, and fully automated interpretation of massive data. Nevertheless, the quality of their work heavily relies on the training data and their approach to acquisition artifacts, as well as integration of the physical boundaries of the electromagnetic diffusion processes. Instead of unifying the traditional geophysical inversion, the ML approaches can be viewed as complementary tools that facilitate efficiency and aid interpretation. Further developments will need a better training dataset, a model design that is sensitive to physics, and mathematical validation with existing geophysical procedures. Machine learning could play a crucial role in future TEM exploration processes with these developments and assist in quicker and more dependable subsurface characterizations.

Funding

This work received no external funding.

Institutional Review Board Statement

Not applicable.

Informed Consent Statement

Not applicable.

Data Availability Statement

No new data were created during the research.

Conflicts of Interest

The authors declare no conflict of interest.

References

[1] Guo, Q., Mao, Y., Yan, L., et al., 2024. Key Technologies for Surface-Borehole Transient Electromagnetic Systems and Applications. Minerals. 14(8), 793. DOI: https://doi.org/10.3390/min14080793

[2] Cheng, M., Yang, D., Luo, Q., 2022. Interpreting Surface Large-Loop Time-Domain Electromagnetic Data for Deep Mineral Exploration Using 3D Forward Modeling and Inversion. Minerals. 13(1), 34. DOI: https://doi.org/10.3390/min13010034

[3] West, G.F., Macnae, J.C., 1991. Physics of the electromagnetic induction exploration method. In Electromagnetic Methods in Applied Geophysics: Volume 2, Application, Parts A and B. Society of Exploration Geophysicists: Houston, TX, USA.

[4] Smith, R., 2014. Electromagnetic Induction Methods in Mining Geophysics from 2008 to 2012. Surveys in Geophysics. 35(1), 123–156. DOI: https://doi.org/10.1007/s10712-013-9227-1

[5] Riurean, S.M., Leba, M., Ionica, A.C., 2021. Conventional and Advanced Technologies for Wireless Transmission in Underground Mine. In Application of Visible Light Wireless Communication in Underground Mine. Springer International Publishing: Cham, Switzerland. pp. 41–125. DOI: https://doi.org/10.1007/978-3-030-61408-9_2

[6] Kaur, M., Kakar, S., Mandal, D., 2011. Electromagnetic interference. In Proceedings of the 2011 3rd International Conference on Electronics Computer Technology, Kanyakumari, India, 8–10 April 2011; pp. 1–5. DOI: https://doi.org/10.1109/ICECTECH.2011.5941844

[7] Maticevic, S., 2011. Optimised Design of Isolated Industrial Power Systems and System Harmonics [PhD Thesis]. Curtin University: Perth, Australia.

[8] Brereton, W., Eng, P., 1986. Report on Reverse Circulation Drilling on the Enjalran-Carheil Joint Venture Properties. Minlstère de PÉnergie et des Ressources: Québec, QC, Canada.

[9] Jiang, C., Lin, J., Duan, Q., et al., 2011. Statistical stacking and adaptive notch filter to remove high-level electromagnetic noise from MRS measurements. Near Surface Geophysics. 9(5), 459–468. DOI: https://doi.org/10.3997/1873-0604.2011026

[10] Shen, C., Appling, A.P., Gentine, P., et al., 2023. Differentiable modelling to unify machine learning and physical models for geosciences. Nature Reviews Earth & Environment. 4(8), 552–567. DOI: https://doi.org/10.1038/s43017-023-00450-9

[11] Karpatne, A., Ebert-Uphoff, I., Ravela, S., et al., 2019. Machine Learning for the Geosciences: Challenges and Opportunities. IEEE Transactions on Knowledge and Data Engineering. 31(8), 1544–1554. DOI: https://doi.org/10.1109/TKDE.2018.2861006

[12] Woodhead, J., Landry, M., 2021. Harnessing the Power of Artificial Intelligence and Machine Learning in Mineral Exploration—Opportunities and Cautionary Notes. SEG Discovery. (127), 19–31. DOI: https://doi.org/10.5382/Geo-and-Mining-13

[13] Huang, Q., Wu, S., Xue, J., 2025. Data Science and Machine Learning in Geo-Electromagnetics: A Review. Surveys in Geophysics. DOI: https://doi.org/10.1007/s10712-025-09904-9

[14] Zhou, X., Shi, P., 2025. Machine learning based subsurface modelling using geological exploration data: a comprehensive review. Georisk: Assessment and Management of Risk for Engineered Systems and Geohazards. 19(4), 871–891. DOI: https://doi.org/10.1080/17499518.2025.2581567

[15] Estay, H., Lois-Morales, P., Montes-Atenas, G., et al., 2023. On the Challenges of Applying Machine Learning in Mineral Processing and Extractive Metallurgy. Minerals. 13(6), 788. DOI: https://doi.org/10.3390/min13060788

[16] Everett, M.E., 2012. Theoretical Developments in Electromagnetic Induction Geophysics with Selected Applications in the Near Surface. Surveys in Geophysics. 33(1), 29–63. DOI: https://doi.org/10.1007/s10712-011-9138-y

[17] Sapia, V., Viezzoli, A., Jørgensen, F., et al., 2014. The Impact on Geological and Hydrogeological Mapping Results of Moving from Ground to Airborne TEM. Journal of Environmental and Engineering Geophysics. 19(1), 53–66. DOI: https://doi.org/10.2113/JEEG19.1.53

[18] Rainieri, C., Fabbrocino, G., 2014. Data Acquisition. In Operational Modal Analysis of Civil Engineering Structures. Springer New York: New York, NY, USA. pp. 59–102. DOI: https://doi.org/10.1007/978-1-4939-0767-0_3

[19] Safder, I., Hassan, S.-U., Visvizi, A., et al., 2020. Deep Learning-based Extraction of Algorithmic Metadata in Full-Text Scholarly Documents. Information Processing & Management. 57(6), 102269. DOI: https://doi.org/10.1016/j.ipm.2020.102269

[20] Edwards, R., Atkinson, K., 1986. Ore Deposit Geology and its Influence on Mineral Exploration. Springer Netherlands: Dordrecht, The Netherlands. DOI: https://doi.org/10.1007/978-94-011-8056-6

[21] Christiansen, A.V., Auken, E., Sørensen, K., 2006. The

transient electromagnetic method. In: Kirsch, R. (Ed.). Groundwater Geophysics. Springer-Verlag: Berlin/Heidelberg, Germany. pp. 179–225. DOI: https://doi.org/10.1007/3-540-29387-6_6

[22] Nieto García, F., 2008. TEM in Geology. Basics and Applications. Sociedad Española de Mineralogía: Madrid, Spain.

[23] Beamish, D., 2002. An assessment of inversion methods for AEM data applied to environmental studies. Journal of Applied Geophysics. 51(2–4), 75–96. DOI: https://doi.org/10.1016/S0926-9851(02)00213-6

[24] Al Marashly, O., Dobróka, M., 2023. Applying the Most Frequent Values Assisted Hilbert Transform into Seismic Attributes. Multidiszciplináris Tudományok. 13, 81–98.

[25] Christensen, N.B., Lawrie, K.C., 2012. Resolution analyses for selecting an appropriate airborne electromagnetic (AEM) system. Exploration Geophysics. 43(4), 213–227. DOI: https://doi.org/10.1071/EG12005

[26] Auken, E., Foged, N., Andersen, K.R., et al., 2020. On-time modelling using system response convolution for improved shallow resolution of the subsurface in airborne TEM. Exploration Geophysics. 51(1), 4–13. DOI: https://doi.org/10.1080/08123985.2019.1662292

[27] Spaulding, A., Middleton, D., 1977. Optimum Reception in an Impulsive Interference Environment–Part I: Coherent Detection. IEEE Transactions on Communications. 25(9), 910–923. DOI: https://doi.org/10.1109/TCOM.1977.1093943

[28] Binetti, M.S., Massarelli, C., Uricchio, V.F., 2024. Machine Learning in Geosciences: A Review of Complex Environmental Monitoring Applications. Machine Learning and Knowledge Extraction. 6(2), 1263–1280. DOI: https://doi.org/10.3390/make6020059

[29] Wedge, D., Hartley, O., McMickan, A., et al., 2019. Machine learning assisted geological interpretation of drillhole data: Examples from the Pilbara Region, Western Australia. Ore Geology Reviews. 114, 103118. DOI: https://doi.org/10.1016/j.oregeorev.2019.103118

[30] Mihaylov, A., El Naggar, H., 2021. A comparison of instrument response correction methods: Post-processing and real-time methods. Results in Geophysical Sciences. 8, 100033. DOI: https://doi.org/10.1016/j.ringps.2021.100033

[31] Hürlimann, M.D., Heaton, N.J., 2015. NMR Well Logging. In: Johns, M.L., Fridjonsson, E.O., Vogt, S.J., et al. (Eds.). New Developments in NMR. Royal Society of Chemistry: Cambridge, UK. pp. 11–85. DOI: https://doi.org/10.1039/9781782628095-00011

[32] Li, A., Parsekian, A.D., Grana, D., et al., 2025. Quantification of measurement uncertainty in electrical resistivity tomography data and its effect on the inverted resistivity model. Geophysics. 90(3), WA275–WA291. DOI: https://doi.org/10.1190/geo2024-0466.1

[33] Colombo, D., Turkoglu, E., Sandoval-Curiel, E., et al., 2023. Machine-learning inversion via adaptive learning and statistical sampling: Application to airborne micro-TEM for seismic sand corrections. Geophysics. 88(3), K51–K68. DOI: https://doi.org/10.1190/geo2022-0407.1

[34] Dahrouj, H., Alghamdi, R., Alwazani, H., et al., 2021. An Overview of Machine Learning-Based Techniques for Solving Optimization Problems in Communications and Signal Processing. IEEE Access. 9, 74908–74938. DOI: https://doi.org/10.1109/ACCESS.2021.3079639

[35] Schelter, S., Böse, J.-H., Kirschnick, J., et al., 2017. Automatically tracking metadata and provenance of machine learning experiments. In Proceedings of the Machine Learning Systems Workshop at NIPS 2017, Long Beach, CA, USA, 8 December 2017.

[36] Goldstein, D., Aldrich, C., Shao, Q., et al., 2025. Unlocking Subsurface Geology: A Case Study with Measure-While-Drilling Data and Machine Learning. Minerals. 15(3), 241. DOI: https://doi.org/10.3390/min15030241

[37] Olofsson, T., 2020. Mining Futures: Predictions and Uncertainty in Swedish Mineral Exploration [PhD Thesis]. Uppsala University: Uppsala, Sweden.

[38] Daruna, A., Zadorozhnyy, V., Lukoczki, G., et al., 2024. Enabling Scalable Mineral Exploration: Self-Supervision and Explainability. In Proceedings of the 2024 IEEE International Conference on Big Data (BigData), Washington, DC, USA, 15–18 December 2024; pp. 2090–2099. DOI: https://doi.org/10.1109/BigData62323.2024.10825956

[39] Ali Memon, F., Li, H., Hussain, M., et al., 2025. Technical Advancements of Laterally Constrained Inversion for Geophysical Datasets. IEEE Access. 13, 160602–160618. DOI: https://doi.org/10.1109/ACCESS.2025.3608240

[40] Cerqueira, F.G., 2024. Exploring Label Efficiency with Semi-Supervision and Self-Supervision Methods [Master's Thesis]. Universidade do Porto: Porto, Portugal.

[41] Rangel DaCosta, L., Sytwu, K., Groschner, C.K., et al., 2024. A robust synthetic data generation framework for machine learning in high-resolution transmission electron microscopy (HRTEM). Npj Computational Materials. 10(1), 165. DOI: https://doi.org/10.1038/s41524-024-01336-0

[42] Kerim, A., 2023. Synthetic Data for Machine Learning: Revolutionize Your Approach to Machine Learning with This Comprehensive Conceptual Guide. Packt Publishing Ltd.: Birmingham, UK.

[43] Viezzoli, A., Dauti, F., Wijns, C., 2021. Robust scanning of AEM data for IP effects. Exploration Geophysics. 52(5), 563–574. DOI: https://doi.org/10.1080/08123985.2020.1856624

[44] Willard, J., Jia, X., Xu, S., et al., 2023. Integrating Scientific Knowledge with Machine Learning for Engineering and Environmental Systems. ACM Computing

Surveys. 55(4), 1–37. DOI: https://doi.org/10.1145/3514228
[45] Apicella, A., Isgrò, F., Prevete, R., 2025. Don't push the button! Exploring data leakage risks in machine learning and transfer learning. Artificial Intelligence Review. 58(11), 339. DOI: https://doi.org/10.1007/s10462-025-11326-3
[46] Bouke, M.A., Zaid, S.A., Abdullah, A., 2024. Implications of Data Leakage in Machine Learning Preprocessing: A Multi-Domain Investigation. Research Square Preprint. DOI: https://doi.org/10.21203/rs.3.rs-4579465/v1
[47] Mörbe, W., Yogeshwar, P., Tezkan, B., et al., 2020. Deep exploration using long-offset transient electromagnetics: interpretation of field data in time and frequency domain. Geophysical Prospecting. 68(6), 1980–1998. DOI: https://doi.org/10.1111/1365-2478.12957
[48] Young, R.E., Gann, G.D., Walder, B., et al., 2022. International principles and standards for the ecological restoration and recovery of mine sites. Restoration Ecology. 30(S2), e13771. DOI: https://doi.org/10.1111/rec.13771
[49] Radha, S.K., Kuehlkamp, A., Nabrzyski, J., 2025. Advancing Transparency and Responsibility in Machine Learning: The Critical Role of FAIR Principles - A Comprehensive Review. ACM Journal on Responsible Computing. 2(4), 1–38. DOI: https://doi.org/10.1145/3768151
[50] Yao, L., Chen, Q., 2023. Machine learning in nanomaterial electron microscopy data analysis. In Intelligent Nanotechnology. Elsevier: Amsterdam, The Netherlands. pp. 279–305. DOI: https://doi.org/10.1016/B978-0-323-85796-3.00010-X
[51] Yu, M., Huang, Q., Li, Z., 2024. Deep learning for spatiotemporal forecasting in Earth system science: a review. International Journal of Digital Earth. 17(1), 2391952. DOI: https://doi.org/10.1080/17538947.2024.2391952
[52] Asif, M.R., Maurya, P.K., Foged, N., et al., 2022. Automated Transient Electromagnetic Data Processing for Ground-Based and Airborne Systems by a Deep Learning Expert System. IEEE Transactions on Geoscience and Remote Sensing. 60, 1–14. DOI: https://doi.org/10.1109/TGRS.2022.3202304
[53] Talpini, J., 2025. Uncertainty Quantification and Distributed Models to Enhance ML-Based Network Security [PhD Thesis]. University of Milano-Bicocca: Milan, Italy.
[54] Yu, S., Shen, Y., Zhang, Y., 2023. CG-DAE: A noise suppression method for two-dimensional transient electromagnetic data based on deep learning. Journal of Geophysics and Engineering. 20(3), 600–609. DOI: https://doi.org/10.1093/jge/gxad035
[55] Young, S.I., Dalca, A.V., Ferrante, E., et al., 2025. Supervision by Denoising. IEEE Transactions on Pattern Analysis and Machine Intelligence. 47(9), 7279–7291. DOI: https://doi.org/10.1109/TPAMI.2023.3299789
[56] Seo, H.-j., 2024. A Survey on Self-Supervised Methods for Image Denoising in Deep Learning [Master's Thesis]. Seoul National University Graduate School: Seoul, Republic of Korea.
[57] Oyedare, T., Shah, V.K., Jakubisin, D.J., et al., 2022. Interference Suppression Using Deep Learning: Current Approaches and Open Challenges. IEEE Access. 10, 66238–66266. DOI: https://doi.org/10.1109/ACCESS.2022.3185124
[58] Tuck, J., 2016. Field Application of Transient Analysis Methods for Pipeline Condition Assessment [Master's Thesis]. University of Canterbury: Christchurch, New Zealand.
[59] Gibert, K., Sànchez–Marrè, M., Izquierdo, J., 2016. A survey on pre-processing techniques: Relevant issues in the context of environmental data mining. AI Communications. 29(6), 627–663. DOI: https://doi.org/10.3233/AIC-160710
[60] García, S., Ramírez-Gallego, S., Luengo, J., et al., 2016. Big data preprocessing: methods and prospects. Big Data Analytics. 1(1), 9. DOI: https://doi.org/10.1186/s41044-016-0014-0
[61] Song, H., Kim, M., Park, D., et al., 2023. Learning From Noisy Labels With Deep Neural Networks: A Survey. IEEE Transactions on Neural Networks and Learning Systems. 34(11), 8135–8153. DOI: https://doi.org/10.1109/TNNLS.2022.3152527
[62] Ericsson, L., Gouk, H., Loy, C.C., et al., 2022. Self-Supervised Representation Learning: Introduction, advances, and challenges. IEEE Signal Processing Magazine. 39(3), 42–62. DOI: https://doi.org/10.1109/MSP.2021.3134634
[63] Audebert, N., 2025. Learning Representations from Observations [PhD Thesis]. Université Paris-Est Sup: Paris, France.
[64] An, Z., 2025. Towards Uncertainty-Aware Model-Based Reinforcement Learning [PhD Thesis]. UC Merced: Merced, CA, USA.
[65] Colombo, D., Turkoglu, E., Li, W., et al., 2021. Physics-driven deep-learning inversion with application to transient electromagnetics. GEOPHYSICS. 86(3), E209–E224. DOI: https://doi.org/10.1190/geo2020-0760.1
[66] Hernandez Mejia, J.L., 2024. Enhanced Subsurface Estimation and Uncertainty Modeling through Data Science and Engineering Physics [PhD Thesis]. The University of Texas at Austin: Austin, TX, USA.
[67] Wang, X., Wang, P., Zhang, X., et al., 2022. Target Electromagnetic Detection Method in Underground Environment: A Review. IEEE Sensors Journal. 22(14), 13835–13852. DOI: https://doi.org/10.1109/JSEN.2022.3175502

[68] Zhao, F., Zhang, C., Geng, B., 2024. Deep Multimodal Data Fusion. ACM Computing Surveys. 56(9), 1–36. DOI: https://doi.org/10.1145/3649447

[69] Islamov, S., Grigoriev, A., Beloglazov, I., et al., 2021. Research Risk Factors in Monitoring Well Drilling—A Case Study Using Machine Learning Methods. Symmetry. 13(7), 1293. DOI: https://doi.org/10.3390/sym13071293

[70] Nouri, A., Davis, P.E., Subedi, P., et al., 2021. Exploring the role of machine learning in scientific workflows: Opportunities and challenges. arXiv preprint. arXiv:2110.13999.

[71] Kalinin, S.V., Mukherjee, D., Roccapriore, K., et al., 2023. Machine learning for automated experimentation in scanning transmission electron microscopy. Npj Computational Materials. 9(1), 227. DOI: https://doi.org/10.1038/s41524-023-01142-0

[72] Humphreys, D., Kupresanin, A., Boyer, D., et al., 2019. Advancing Fusion with Machine Learning Research Needs Workshop. Office of Science, U.S. Department of Energy: Washington, DC, USA.

[73] Medico, R., Lambrecht, N., Pues, H., et al., 2019. Machine Learning Based Error Detection in Transient Susceptibility Tests. IEEE Transactions on Electromagnetic Compatibility. 61(2), 352–360. DOI: https://doi.org/10.1109/TEMC.2018.2821712

[74] Melo, A.T., 2018. Integrated Quantitative Interpretation of Multiple Geophysical Data for Geology Differentiation. Colorado School of Mines: Golden, CO, USA.

[75] Wenxue, Z., Shikun, D., Hongjun, T., et al., 2025. An Overview Study of Deep Learning in Geophysics: Cross-Cutting Research to Advance Geoscience. IEEE Access. 13, 124364–124388. DOI: https://doi.org/10.1109/ACCESS.2025.3586693

[76] Vaseghi, S.V., 2008. Advanced Digital Signal Processing and Noise Reduction, 1st ed. Wiley: Hoboken, NJ, USA. DOI: https://doi.org/10.1002/9780470740156

[77] Randa, J., Dunsmore, J., Gu, D., et al., 2011. Verification of Noise-Parameter Measurements and Uncertainties. IEEE Transactions on Instrumentation and Measurement. 60(11), 3685–3693. DOI: https://doi.org/10.1109/TIM.2011.2138270

[78] Manrique, I.I., 2023. Quantitative Interpretation of Geophysical Data in Anthropogenic Waste Deposits and Disposal Sites for Resource Recovery and Remediation. Universite de Liege: Liège, Belgium.

[79] Segundo-Ramirez, J., Bayo-Salas, A., Esparza, M., et al., 2020. Frequency Domain Methods for Accuracy Assessment of Wideband Models in Electromagnetic Transient Stability Studies. IEEE Transactions on Power Delivery. 35(1), 71–83. DOI: https://doi.org/10.1109/TPWRD.2019.2927171

[80] Balaji, C., Srinivasan, B., 2026. Machine learning approach to inverse problems. In Machine Learning and Bayesian Methods in Inverse Heat Transfer. Elsevier: Amsterdam, The Netherlands. pp. 197–285. DOI: https://doi.org/10.1016/B978-0-44-336791-5.00012-9

[81] He, H., Li, J., Stoica, P., 2012. Waveform Design for Active Sensing Systems: A Computational Approach, 1st ed. Cambridge University Press: Cambridge, UK. DOI: https://doi.org/10.1017/CBO9781139095174

[82] Kang, J., Kim, N., Ok, J., et al., 2025. MemBN: Robust Test-Time Adaptation via Batch Norm with Statistics Memory. In: Leonardis, A., Ricci, E., Roth, S., et al. (Eds.). Computer Vision – ECCV 2024, Lecture Notes in Computer Science. Springer Nature: Cham, Switzerland. pp. 467–483. DOI: https://doi.org/10.1007/978-3-031-73390-1_27

[83] Rose, P.R., 1987. Dealing with Risk and Uncertainty in Exploration: How Can We Improve? AAPG Bulletin. 71(1), 1–16. DOI: https://doi.org/10.1306/94886D30-1704-11D7-8645000102C1865D

[84] Fink, O., Nejjar, I., Sharma, V., et al., 2025. From Physics to Machine Learning and Back: Part II-Learning and Observational Bias in PHM. arXiv preprint. arXiv:2509.21207.

[85] Cracknell, M.J., Reading, A.M., 2014. Geological mapping using remote sensing data: A comparison of five machine learning algorithms, their response to variations in the spatial distribution of training data and the use of explicit spatial information. Computers & Geosciences. 63, 22–33. DOI: https://doi.org/10.1016/j.cageo.2013.10.008

[86] Van Breugel, B., Qian, Z., Van Der Schaar, M., 2023. Synthetic data, real errors: How (not) to publish and use synthetic data. In Proceedings of the 40th International Conference on Machine Learning, Honolulu, HI, USA, 23–29 July 2023.

[87] Fakour, F., Mosleh, A., Ramezani, R., 2024. A structured review of literature on uncertainty in machine learning & deep learning. arXiv preprint. arXiv:2406.00332.

[88] Virieux, J., Operto, S., 2009. An overview of full-waveform inversion in exploration geophysics. Geophysics. 74(6), WCC1–WCC26. DOI: https://doi.org/10.1190/1.3238367

[89] Kovalerchuk, B., 2024. Interpretable AI/ML for High-stakes Tasks with Human-in-the-loop: Critical Review and Future Trends. Research Square Preprint. DOI: https://doi.org/10.21203/rs.3.rs-3989807/v1

[90] Kumar, S., Datta, S., Singh, V., et al., 2024. Applications, Challenges, and Future Directions of Human-in-the-Loop Learning. IEEE Access. 12, 75735–75760. DOI: https://doi.org/10.1109/ACCESS.2024.3401547

[91] Dimri, V., 2013. Deconvolution and Inverse Theory: Application to Geophysical Problems. Elsevier Science: Amsterdam, The Netherlands.

[92] Katragadda, S.R., Tanikonda, A., Peddinti, S.R., et al., 2025. Machine Learning-Enhanced Root Cause Analysis for Accelerated Incident Resolution in Complex Systems. Journal of Science & Technology. 2(4), 253–275. DOI: https://doi.org/10.2139/ssrn.5104444