© AUG 2026 | IRE Journals | Volume 10 Issue 2 | ISSN: 2456-8880
DOI: https://doi.org/10.64388/IREV10I2-1722266

Big Data Analytics for Cross-Domain Anomaly Detection to Identify Hidden Service-Impacting Patterns in Fiber Access Broadband Networks

RIZWAN FAROOQ KHAN
ORCiD: 0009-0004-7507-0246

Abstract- ***Fibre-to-the-x networks are observed through several operational systems, yet service-impacting degradation often remains hidden because each system describes only a fragment of the end-to-end service. Optical power, line errors, dynamic bandwidth allocation, IP-session behaviour, customer-premises telemetry, topology, alarms and complaints may each remain within local thresholds while their joint movement signals a developing fault. This review examines how cross-domain data correlation and anomaly detection can expose such weak, distributed evidence before it becomes a widespread outage. It synthesises research published from 2020 to 2025 on multivariate time-series modelling, graph learning, feature-domain interaction, optical-network failure management and explainable detection. The review proposes an evidence-centred architecture that aligns heterogeneous telemetry by service, topology and time; learns both within-domain signatures and between-domain dependencies; models propagation across shared network resources; and converts anomaly scores into service-impact hypotheses that engineers can verify. Particular attention is given to incomplete labels, changing baselines, class imbalance, encrypted traffic, topology reconfiguration and the distinction between statistical abnormality and operational harm. The analysis argues that effective FTTx assurance requires neither a single universal model nor indiscriminate data fusion. It requires a layered design in which deterministic rules, unsupervised detectors, graph-temporal models and human validation cooperate under explicit data-quality and governance controls. The resulting framework supports earlier detection, sharper fault-domain isolation and more defensible prioritisation while preserving operational explainability.***

Keywords: FTTx networks; passive optical networks; cross-domain correlation; anomaly detection; service assurance; graph learning; multivariate time series; root-cause analysis

I. INTRODUCTION

FTTx assurance is difficult not because operators lack telemetry, but because the evidence is fragmented across systems that were designed for different purposes. The optical access layer exposes receive and transmit power, loss of signal, forward-error correction, laser bias, ranging events and optical distribution network alarms. The Ethernet and IP layers expose packet loss, latency, queue occupancy, session churn, authentication failures and routing changes. Customer-premises equipment contributes Wi-Fi quality, device load and reboot history, while inventory, geographic information, tickets and contact-centre records provide topology and service context. A visible threshold breach in one source is easy to action. A weak pattern distributed across five sources is harder to recognise, even when it is more predictive of customer harm.

Conventional monitoring therefore produces two recurrent failure modes. First, isolated thresholds generate noisy alerts because normal behaviour varies by time of day, service tier, optical budget, access technology, firmware and local environment. Secondly, locally normal measurements can conceal globally abnormal relationships. A slowly falling optical level may still sit inside its permitted range, but become significant when accompanied by rising corrected errors, repeated optical-network-unit re-registration and clustered throughput complaints on a common splitter branch. The relevant signal is not a single value; it is the loss of consistency between values that should normally move together.

Recent anomaly-detection research increasingly addresses heterogeneous features, temporal dependence and relational structure. Domain-aware learning separates static fields, sequences and

aggregate statistics before modelling their interactions [1]. Multivariate detectors learn dependencies among variables rather than treating each metric independently [2-8]. Graph-based approaches represent entities and their relationships explicitly, supporting localisation and explanation when anomalies propagate through shared infrastructure [5,11,19]. Optical-network studies also show that machine learning can assist alarm classification, soft-failure localisation, branch identification and interpretation under sparse or imbalanced fault data [12-16,20-21,27-30]. These advances are relevant to FTTx, but they must be translated into an operational design that reflects service topology, asynchronous sampling and the practical cost of false escalation.

This review develops that translation. It treats hidden service impact as an evidence-correlation problem spanning physical, protocol, service and customer domains. The objective is not to claim that every correlated deviation is causal. Rather, it is to define a workflow that narrows plausible causes, quantifies uncertainty and retains the evidence needed for engineering verification. Such a workflow can detect emerging degradation earlier than alarm-only monitoring, distinguish access faults from home-network or application effects, and rank investigations by the number and criticality of services exposed.

II. AIM AND REVIEW OBJECTIVES

The aim is to synthesise contemporary methods for detecting hidden service-impacting patterns in FTTx networks and to formulate a deployable cross-domain assurance framework. Four objectives guide the review: to define the operational data domains and the relationships that make them informative; to compare statistical, representation-learning, graph and temporal approaches; to identify controls required for trustworthy deployment; and to establish a research agenda for service-impact reasoning rather than anomaly scoring alone. The review focuses on detection and investigation support, not autonomous network control.

III. REVIEW METHODOLOGY

A structured integrative review was undertaken across work published between January 2020 and December 2025. Searches combined terms relating to FTTx, passive optical networks, optical fault management, network anomaly detection, multivariate time series, graph neural networks, telemetry correlation, failure localisation and explainability. Peer-reviewed journal and conference papers were prioritised, together with technically substantive studies that introduced widely used anomaly-detection architectures. The attached domain-interaction study was used as a central reference because it explicitly separates heterogeneous feature groups, models their internal structure, learns interactions across groups and then applies temporal modelling [1].

Studies were retained when they contributed at least one of five capabilities: learning normal multivariate behaviour; modelling relations among network entities or metrics; diagnosing optical or access-network failures; handling sparse, imbalanced or partially labelled data; or explaining and validating anomaly decisions. Papers limited to payload-only intrusion signatures, unrelated image recognition or purely theoretical optimisation without operational telemetry were excluded. The final evidence set contained thirty sources from 2020-2025. Findings were coded under data representation, correlation mechanism, temporal modelling, topology awareness, fault localisation, explainability, deployment constraints and evaluation practice.

The synthesis was interpretive rather than a statistical meta-analysis because datasets, fault taxonomies and reported metrics differ substantially. Accuracy values from laboratory datasets cannot be transferred directly to production FTTx environments, where labels are incomplete and normal behaviour changes after upgrades, migrations and seasonal shifts. Accordingly, the review emphasises design principles supported across studies: preserve domain semantics, model relationships explicitly, evaluate by incident and service impact, retain uncertainty, and combine model outputs with deterministic operational knowledge. The proposed architecture is therefore a reasoned synthesis

of the evidence, not an experimentally validated FTTx product.

IV. CROSS-DOMAIN OBSERVABILITY IN FTTX NETWORKS

A useful FTTx correlation model begins with the service path rather than the monitoring platform. The path links an optical line terminal port to feeder fibre, splitters, distribution branches, drop fibre, an ONU or ONT, the residential gateway, subscriber session, service profile and customer endpoint. Every observation should be attached, where possible, to this evolving service graph. Without that graph, a detector can recognise that several metrics changed together but cannot determine whether they share a physical resource, a logical policy or merely a clock interval.

The optical domain provides direct evidence of physical-layer health. Measurements include received optical power, transmit power, attenuation estimates, laser bias, temperature, corrected and uncorrected codewords, loss-of-frame events, ranging retries and OTDR-derived reflections. PON research demonstrates that branch localisation becomes difficult when reflections overlap or when models are trained on one topology and tested on another [15,21]. This supports two design choices: optical traces should be fused with inventory and branch geometry, and model confidence should fall when topology quality is poor. A power decline of two decibels has different meaning near the edge of the optical budget than on a short, stable branch.

The access and IP domains describe whether the physical path is delivering a usable service. OLT counters, dynamic bandwidth allocation behaviour, Ethernet errors, queue drops, BNG or BRAS sessions, DHCP and PPPoE events, RADIUS outcomes, DNS latency and throughput tests reveal different failure mechanisms. Feature standardisation matters because nominally similar counters may have different units, reset behaviour and collection intervals [10]. Encrypted application traffic also limits payload inspection, increasing the value of flow statistics, timing, directionality and relational patterns [9,11,25]. The customer and operational domains determine service impact. Gateway reboots, Wi-Fi retransmission, channel congestion and device counts help separate access degradation from in-home impairment. Trouble tickets, repeat calls, failed self-care tests, technician notes and compensation records provide delayed but highly valuable labels. Geographic and environmental data can reveal common exposure to excavation, cabinet heat, flooding, power instability or construction. These sources are noisy and socially mediated, but they show whether a technical deviation is perceptible. A model that detects unusual optical behaviour without estimating affected services risks prioritising harmless maintenance activity over a subtle fault affecting a hospital, enterprise or dense residential cluster.

Cross-domain alignment requires three forms of normalisation. Semantic normalisation maps vendor-specific names into common concepts and records provenance. Temporal normalisation reconciles polling, event and transactional streams without pretending that all samples are synchronous. Topological normalisation resolves changing relationships caused by port moves, splitter rebalancing, service migration and equipment replacement. Rather than force all data into a single flat vector, the framework should preserve feature domains, learn domain-specific encodings and then model interactions [1]. Table 1 summarises the principal domains and the hidden patterns that their combination can reveal.

Missingness is itself informative. A silent ONU, absent telemetry feed or delayed inventory update may signal a fault, a collector problem or a process failure. Imputation should therefore be accompanied by a missingness indicator and source-health score. Likewise, aggregation must not erase rare bursts. Five-minute averages can conceal second-scale packet loss, while raw event floods can overwhelm correlation. Multi-resolution windows are preferable: short windows for burst faults, medium windows for service degradation and longer baselines for drift, seasonality and repeated intermittency.

Identity resolution is often the least visible source of error. Customer identifiers, circuit identifiers, ONU serial numbers, MAC addresses and session identifiers change at different rates and may be stored in separate systems. Correlation should therefore use versioned

© AUG 2026 | IRE Journals | Volume 10 Issue 2 | ISSN: 2456-8880
DOI: https://doi.org/10.64388/IREV10I2-1722266

mappings with confidence levels rather than irreversible joins. When two records cannot be resolved confidently, the uncertainty should remain visible to the model. This prevents an attractive but false pattern from being created by an incorrect service-to-resource association. It also allows analysts to distinguish genuine network anomalies from inventory reconciliation failures, which are operational problems in their own right.

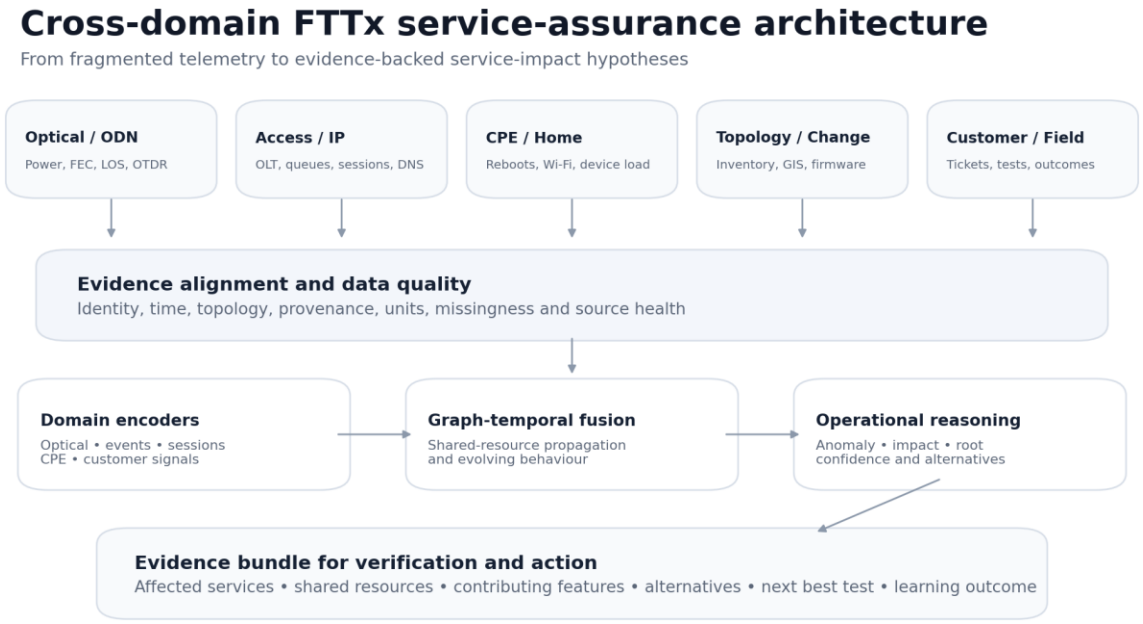

Figure 1. Proposed evidence-centred architecture for cross-domain anomaly detection in FTTx operations.

Table 1. Cross-domain evidence sources and examples of hidden service-impacting patterns.

Data domain	Representative evidence	Hidden cross-domain pattern	Operational value	Primary caution
Optical and ODN	Rx/Tx power, FEC, LOS, ranging, OTDR, laser bias	Small power drift + corrected errors + ONU re-registration on a shared branch	Detects soft faults and narrows physical fault domain	Optical budgets and device calibration vary
Access and Ethernet	OLT counters, DBA, queues, CRC, port state	Peak-hour queue loss + upstream demand + stable optical layer	Separates congestion from fibre impairment	Counter resets and polling intervals differ
IP and session	DHCP, PPPoE, RADIUS, BNG, DNS, flow statistics	Session churn + authentication delay + OLT event cluster	Links access events to service establishment failures	Encrypted traffic limits payload visibility
CPE and home network	Gateway reboot, Wi-Fi quality, device load, firmware	Stable access + Wi-Fi retransmission + local channel congestion	Avoids misclassifying in-home faults as network faults	Customer devices are heterogeneous and intermittently online

Data domain	Representative evidence	Hidden cross-domain pattern	Operational value	Primary caution
Topology and change	Inventory, GIS, splitter tree, port moves, firmware, work orders	Symptoms align with a shared resource or recent change cohort	Supports propagation analysis and candidate-root ranking	Stale inventory creates false relationships
Customer and field	Tickets, self-tests, calls, technician notes, recovery	Low-level telemetry drift + geographically clustered complaints	Converts abnormality into evidence of perceived impact	Labels are delayed, subjective and process-dependent

V. ANALYTICAL APPROACHES FOR HIDDEN PATTERN DETECTION

No single algorithm covers the full assurance problem. Deterministic rules remain essential for known conditions, safety limits and contractual thresholds. Robust statistics can identify deviations from peer groups or seasonal baselines with low computational cost. These methods are transparent, but they struggle when normality depends on many interacting variables or when the same symptom has several causes. Their best role is to provide guardrails, features and corroborating evidence rather than the sole decision mechanism.

Unsupervised representation learning is attractive because labelled FTTx faults are scarce. Autoencoder-based methods learn to reconstruct normal observations and use reconstruction error as an anomaly score. USAD improves stability through adversarially trained autoencoders [4], while early traffic models show that unsupervised learning can detect abnormal behaviour without exhaustive attack labels [2]. In access assurance, reconstruction models can learn joint profiles of optical power, errors, session activity and throughput. However, a low reconstruction error does not guarantee operational normality when the training data already contain chronic faults. Training windows therefore require data-quality screening, and models need periodic review against incident outcomes.

Graph learning is especially well matched to FTTx because topology constrains how faults propagate. A graph can represent OLTs, ports, splitters, ONUs, gateways, sessions and customer accounts as nodes, with physical, logical and geographic relationships as edges. Graph deviation networks learn expected relations among variables and expose which dependencies have broken [5]. Graph-attention and learned-structure models capture both feature and temporal interactions [3,6], while graph anomaly analytics provides a broader basis for node, edge and subgraph reasoning [19]. In optical networks, graph-based localisation has been used to combine monitoring data with network structure and to produce interpretable prototypes or dynamic localisation under reconfiguration [29-30].

The operational benefit of a graph is not merely higher classification accuracy. It enables service-cone reasoning. When several ONUs degrade, the model can test whether they share a splitter, feeder, OLT card, cabinet, power source or software version. It can distinguish a branch-level optical event from a system-wide policy change and can rank candidate roots by the proportion of observed symptoms they explain. The graph must nevertheless be time-aware. Using today's topology to analyse an incident from last month can create false causal paths, so relationship histories and effective dates are required.

Temporal models address the fact that many service-impacting patterns develop as sequences rather than isolated points. Recurrent, convolutional and Transformer architectures capture dependencies across different horizons. The Anomaly Transformer uses association discrepancy to distinguish abnormal time points [7], and TranAD combines attention, self-conditioning and adversarial training for multivariate detection [8]. Network-specific Transformer models have also been applied to encrypted traffic [9], while NetMamba explores efficient sequence modelling for high-volume classification [24]. In FTTx, temporal

context is needed to identify progressive attenuation, intermittent flaps, post-maintenance instability, peak-hour congestion and recurring weather-related degradation.

Pure sequence modelling can still overlook feature heterogeneity. A packet-loss series, a categorical alarm, a topology identifier and a customer complaint are not interchangeable tokens. Domain-interaction architectures address this by extracting local patterns within each feature group before attention is used across groups [1]. A practical FTTx model might use separate encoders for optical telemetry, performance counters, session events, CPE state, topology and customer signals. Cross-domain attention would then learn, for example, that a modest optical decline becomes important only when error correction, re-registration and complaint density move together.

Hybrid models are consequently more defensible than monolithic ones. One layer can apply rules for hard faults and data validation; unsupervised models can identify previously unseen deviations; graph components can localise shared infrastructure; and temporal models can distinguish bursts from trends. Optical failure-management studies report benefits from attention mechanisms, partial telemetry, alarm classification, data augmentation and experimental comparison of generative and conventional models [12-16,20]. These findings suggest that the model portfolio should be selected by use case: immediate loss-of-signal, covert degradation, branch localisation, congestion, customer-experience divergence or post-change assurance.

Class imbalance is a central constraint because normal records vastly outnumber verified faults. Data augmentation can improve learning when some failures are rarely observed [14], while ensemble and undersampling strategies can preserve sensitivity under imbalance [27]. Synthetic data should nevertheless be treated as a training aid, not as evidence that production coverage is complete. Simulated optical traces or injected counter patterns may omit human, environmental and configuration interactions. Evaluation should report precision, recall, false alarms per monitored service, time-to-detection and localisation quality, with separate results for common, rare and unseen fault families.

Anomaly scoring must be separated from service-impact scoring. Statistical rarity answers whether behaviour is unusual; impact estimation asks whether users are likely to experience degradation and how many are exposed. The second score can combine affected-service count, service criticality, symptom persistence, complaint evidence, redundancy and confidence in the shared root. A planned migration may be highly anomalous but harmless, whereas a small correlated drift across a high-value enterprise branch may warrant urgent investigation. This distinction reduces alert fatigue and aligns detection with operational priorities.

Explainability is required because engineers must decide whether to test, dispatch, reconfigure or observe. Feature attribution can show which metrics drove a decision, as demonstrated by explainable tree-based optical fault diagnosis [27]. Graph prototypes can expose topology and fault-class similarities [29]. For FTTx, explanations should be written as evidence chains: which services deviated, when the deviation began, which relationships are shared, which measurements changed, what alternative causes were considered, and which verification step would most reduce uncertainty. A confidence score without this chain is insufficient for high-cost field action.

Incident formation is a separate analytical task from point detection. Thousands of anomalous samples may belong to one developing fibre event, while a single sample may represent a short and harmless transition. Event clustering should consider temporal proximity, shared topology, symptom similarity and recovery behaviour. The resulting incident object can maintain a start estimate, affected subgraph, supporting observations and confidence trajectory. This reduces duplicate alerts and permits escalation rules based on persistence and spread. It also creates a stable unit for evaluation because engineers resolve incidents, not isolated rows in a telemetry table.

Calibration is as important as raw discrimination. An anomaly score from one model is not directly comparable with a score from another model or network segment. Probability calibration, peer-relative ranking and confidence intervals can convert heterogeneous outputs into operationally interpretable bands. Ensembles should combine evidence only

© AUG 2026 | IRE Journals | Volume 10 Issue 2 | ISSN: 2456-8880
DOI: https://doi.org/10.64388/IREV10I2-1722266

when their errors are sufficiently independent; otherwise, several models trained on the same counters can create false confidence. A conservative fusion policy can require agreement between a domain detector and a topology or customer-impact signal before a high-severity case is raised, while still retaining low-confidence observations for trend analysis.

Table 2. Comparative suitability of anomaly-detection approaches for FTTx service assurance.

Approach	Strength in FTTx assurance	Typical use	Main limitation	Recommended role
Rules and robust statistics	Transparent and inexpensive	Hard faults, threshold breaches, peer deviation	Weak for nonlinear, evolving relationships	Guardrails, validation and corroboration
Autoencoders and unsupervised models	Works with limited labels; learns joint baseline	Unknown deviations, covert degradation	Can learn chronic faults as normal	Entity and domain-level anomaly scoring
Graph neural networks	Represents topology and shared resources	Branch/card/cabinet localisation, propagation	Depends on accurate time-aware topology	Subgraph detection and candidate-root ranking
Transformers and sequence models	Captures long-range and multi-scale behaviour	Drift, intermittency, post-change instability	Costly; may blur heterogeneous feature meaning	Temporal context after domain encoding
Domain-interaction models	Preserves feature semantics before fusion	Optical-IP-CPE-customer correlation	Requires careful domain design and missingness handling	Cross-domain evidence fusion
Explainable ensembles and prototypes	Shows feature or pattern contributions	Engineer verification and high-cost dispatch	Explanation may be locally faithful, not causal	Evidence chains and decision support

VI. PROPOSED EVIDENCE-CENTRED ARCHITECTURE AND WORKFLOW

Figure 1 presents a layered architecture. Source adapters ingest streaming counters, alarms, traces, session events, CPE telemetry, topology, change records and customer signals. A data-quality layer validates units, timestamps, counter resets, missingness and collector health. An identity and topology service resolves every observation to the relevant resource and service path. Domain encoders then preserve the structure of each source family, while a graph-temporal fusion layer learns cross-domain and propagation patterns. The final reasoning layer produces anomaly, impact and candidate-root scores together with an evidence bundle.

Detection should operate at several scopes. Entity models monitor individual ONUs, ports, gateways and sessions against their own baselines. Peer models compare similar services by technology, profile, optical budget and geography. Subgraph models search for correlated deviations within branches, cards, cabinets or software cohorts. Network-wide models identify broad policy, routing or platform changes. Agreement across scopes strengthens confidence; disagreement is diagnostically useful. For example, an ONU anomaly without peer or branch evidence points toward a drop fibre or premises issue, while aligned anomalies across a splitter subgraph point upstream.

Figure 2 converts the architecture into an investigation workflow. The process begins by framing the affected service and observation window, then aligning evidence by identity, time and topology. Models generate hypotheses, but targeted verification follows before action. Verification may request an OTDR

© AUG 2026 | IRE Journals | Volume 10 Issue 2 | ISSN: 2456-8880
DOI: https://doi.org/10.64388/IREV10I2-1722266

trace, compare neighbouring branches, inspect change records, run an active service test or ask the gateway for fresh diagnostics. The workflow ends with controlled remediation and outcome capture. Engineer disposition, recovery behaviour and customer confirmation become new labels, while false positives are recorded rather than silently discarded.

This closed loop prevents the model from becoming detached from operations. A detector trained only on historical alarms will reproduce historical blind spots. Incorporating verified incidents, no-fault-found visits, maintenance changes and customer outcomes creates a more realistic incident memory. Nevertheless, learning should be governed: labels need provenance, automated retraining requires approval thresholds, and model versions must be linked to the data and topology snapshots used for each decision.

The verification stage can be optimised as an information-gain problem. Instead of requesting every available diagnostic, the system can recommend the test most likely to separate competing hypotheses. A fresh OTDR trace may distinguish fibre loss from congestion; a neighbouring-ONU comparison may separate branch and premises faults; a configuration diff may reveal a change-induced pattern. Each test has cost, delay and potential service risk. Ranking tests by expected uncertainty reduction makes the workflow faster and more disciplined, while ensuring that automated reasoning remains anchored to measurements that an engineer can inspect.

Scalability requires selective computation. Lightweight entity baselines can run continuously, while more expensive graph-temporal inference is triggered when several weak signals align or when a high-value service deviates. Feature stores can cache stable topology and peer descriptors, and streaming joins can maintain short-lived evidence windows. This tiered design avoids applying the largest model to every counter update. It also supports graceful degradation: if customer or CPE data are temporarily unavailable, the optical and access layers can continue scoring with reduced confidence rather than stopping or silently assuming normality.

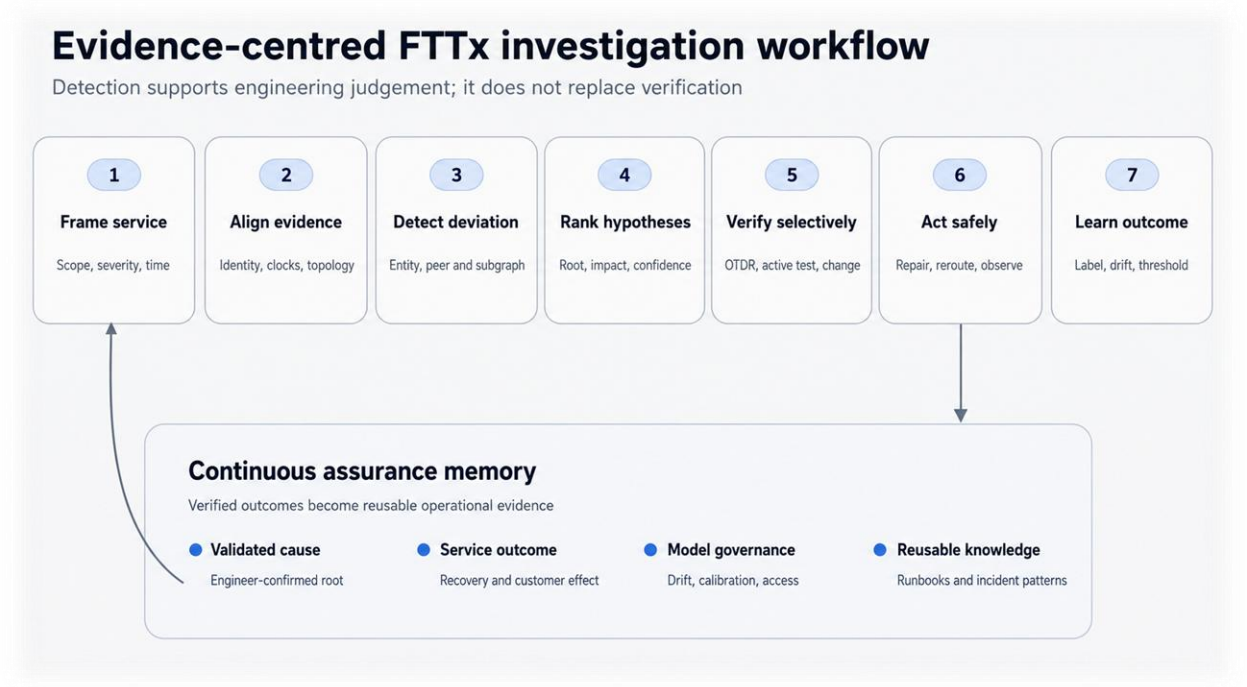

Figure 2. Closed-loop workflow for evidence alignment, anomaly interpretation, verification and controlled learning.

VII. DEPLOYMENT, VALIDATION AND GOVERNANCE

Production validation should use time-separated and topology-separated testing. Random record splits allow information from the same incident or branch to leak into both training and test sets, overstating performance. More credible tests hold out complete time periods, exchanges, OLTs or fault families. Optical studies also warn that models may not generalise when topology or monitoring conditions change [15,28,30]. Shadow deployment is therefore advisable: models score live events without controlling operations, and outputs are compared with engineer findings before escalation rules are enabled. Thresholds should reflect operational capacity and fault cost. A single global threshold is unlikely to suit mass-market broadband, business access and critical services. Calibrated thresholds can be set by service class, model confidence and corroborating evidence, with hysteresis to prevent repeated alerting. Drift monitoring should track input distributions, missingness, prediction confidence, alert yield and incident confirmation. Sudden drift after firmware or inventory changes may indicate a data-pipeline problem rather than network degradation.

Data governance is equally important. Customer and device telemetry should be minimised, access-controlled and retained only as long as operationally justified. Feature definitions, transformations and lineage must be documented so that engineers can reproduce an alert. Vendor-specific counters should not be merged merely because their labels appear similar. The model registry should record training period, network scope, assumptions, performance by fault class and known limitations. Human override and appeal routes are necessary when model outputs affect dispatch, customer communication or service credits.

Operational success should be measured beyond model accuracy. Relevant outcomes include mean time to detect, mean time to isolate, avoided repeat visits, reduction in no-fault-found cases, percentage of incidents detected before complaints, false escalations per thousand services, and restoration time. These measures connect analytical quality to customer and engineering value. They also reveal whether a model is merely detecting obvious outages already visible to existing systems.

Transfer across vendors, regions and PON generations should be assessed explicitly. A model trained on one OLT family may exploit counter definitions or alarm behaviour that do not exist elsewhere. Domain adaptation can reuse general temporal and relational patterns, but local calibration and feature contracts remain necessary. Deployment reports should state which features are portable, which are vendor-specific and which require retraining. A small decline in benchmark accuracy may be acceptable if the model remains stable across equipment changes, because operational robustness is more valuable than a fragile peak result obtained on one static dataset.

Human factors also shape performance. Engineers may distrust an opaque model, over-trust a consistently confident one, or change their ticket language after automation is introduced. Training should explain how evidence is assembled, where the model is weak and how feedback is used. Interfaces should permit disagreement with a predicted cause and record the reason without penalising the user. Monitoring should look for automation bias, confirmation bias and feedback loops in which model-generated categories become the labels used to prove the same model correct.

VIII. DISCUSSION AND FUTURE RESEARCH

The literature supports a clear direction: hidden FTTx problems are more likely to be exposed when feature semantics, topology and time are modelled together. Yet several gaps remain. Public anomaly datasets rarely contain the combined optical, IP, CPE, customer and change-management evidence required for end-to-end service-impact research. The field therefore needs privacy-preserving benchmark datasets with realistic missingness, topology evolution, repeated intermittent faults and verified customer outcomes. Synthetic generation can supplement these datasets, but benchmark design should disclose which relationships are simulated and which are measured.

Causal reasoning is another open problem. Correlation can narrow the fault domain but may confuse common cause, propagation and coincidence. Future work should combine graph-temporal detection with intervention evidence from maintenance, controlled tests and configuration changes. Counterfactual questions are particularly useful: would the complaint cluster still be expected if optical conditions had remained stable, or if the gateway firmware had not changed? Such reasoning can improve root-cause ranking without pretending that observational data alone proves causality.

Adaptive domain discovery also deserves attention. Manually defined domains reflect engineering knowledge but may miss interactions that emerge with new access technologies and service platforms. Data-driven clustering, attention-guided partitioning and meta-learning can propose new feature groups, while engineers retain authority to reject semantically misleading combinations [1,26]. Efficient models are required for high-cardinality estates; compression, distributed inference and event-triggered scoring may reduce cost without flattening all evidence into a coarse aggregate.

Finally, evaluation must move from point anomalies to incidents and consequences. A useful detector should group related events, estimate the fault's temporal boundaries, identify exposed services, rank shared roots and suggest verification. Research should compare these outputs against engineer investigations, not only against per-sample labels. Explainable and dynamically topology-aware optical models provide promising foundations [27-30], but their integration with customer experience and operational workflow remains limited. This integration is the most important step toward genuinely service-centred assurance.

Service-impact labels need richer structure than a binary outage flag. Useful labels can include degradation type, affected application class, duration, intermittency, customer visibility, business criticality and whether redundancy masked the fault. Weak supervision may combine complaint bursts, failed active tests, technician findings and post-repair recovery to estimate these labels. Research should quantify label confidence and disagreement rather than presenting operational records as perfect ground truth. This would support models that distinguish a technically abnormal condition from one that is genuinely harmful, and would make evaluation more aligned with service-level objectives.

A further priority is resilience to incomplete observability. Production estates frequently lose one telemetry stream during the same event that operators need to diagnose. Models should be trained and tested with realistic source dropouts, delayed messages and contradictory measurements. Confidence should decline according to the importance and reliability of the missing domain, rather than through a penalty. Redundant evidence paths can then be designed: customer sessions may corroborate access counters, peer ONUs may substitute for a failed branch sensor, and active tests may replace absent CPE reports. This approach treats observability architecture as part of fault tolerance, ensuring that correlation remains useful during collector, platform or communications failures.

IX. CONCLUSION

Hidden service-impacting patterns in FTTx networks arise when individually weak observations become meaningful through their relationship. Effective detection therefore requires cross-domain alignment of optical, access, IP, CPE, topology, operational and customer evidence. The reviewed work shows that unsupervised learning can discover unfamiliar deviations, graph methods can represent shared infrastructure and propagation, temporal models can capture evolving behaviour, and explainable optical fault techniques can support localisation and trust. Their operational value increases when they are combined rather than treated as competing universal solutions.

The proposed framework preserves domain semantics, validates data quality, resolves evidence to a time-aware service graph, applies layered graph-temporal detection, separates anomaly from impact, and requires targeted verification before remediation. Its closed learning loop records outcomes and uncertainty, while governance controls preserve reproducibility and human authority. This approach offers a practical route from fragmented monitoring to

evidence-centred assurance: earlier warning, narrower fault domains, fewer avoidable dispatches and prioritisation based on actual service exposure. The central principle is simple but demanding: the network should be judged not by isolated counters, but by whether its physical, logical and customer-facing signals remain mutually consistent.

REFERENCES

[1] Sun, W., Zhang, F., Guo, L. and Liu, X. (2025). A multi-feature domain interaction learning framework for anomalous network detection. Electronics, 14, 3729. https://doi.org/10.3390/electronics14183729.

[2] Hwang, R.H., Peng, M.C., Huang, C.W., Lin, P.C. and Nguyen, V.L. (2020). An unsupervised deep learning model for early network traffic anomaly detection. IEEE Access, 8, 30387-30399.

[3] Zhao, H., Wang, Y., Duan, J., Huang, C., Cao, D., Tong, Y., Xu, B., Bai, J., Tong, J. and Zhang, Q. (2020). Multivariate time-series anomaly detection via graph attention network. Proceedings of the IEEE International Conference on Data Mining, 841-850. https://doi.org/10.1109/ICDM50108.2020.00093.

[4] Audibert, J., Michiardi, P., Guyard, F., Marti, S. and Zuluaga, M.A. (2020). USAD: UnSupervised anomaly detection on multivariate time series. Proceedings of the 26th ACM SIGKDD Conference on Knowledge Discovery and Data Mining, 3395-3404. https://doi.org/10.1145/3394486.3403392.

[5] Deng, A. and Hooi, B. (2021). Graph neural network-based anomaly detection in multivariate time series. Proceedings of the AAAI Conference on Artificial Intelligence, 35(5), 4027-4035. https://doi.org/10.1609/aaai.v35i5.16523.

[6] Chen, Z., Chen, D., Zhang, X., Yuan, Z. and Cheng, X. (2022). Learning graph structures with transformer for multivariate time-series anomaly detection in IoT. IEEE Internet of Things Journal, 9(12), 9179-9190. https://doi.org/10.1109/JIOT.2021.3100509.

[7] Xu, J., Wu, H., Wang, J. and Long, M. (2022). Anomaly Transformer: Time series anomaly detection with association discrepancy. International Conference on Learning Representations.

[8] Tuli, S., Casale, G. and Jennings, N.R. (2022). TranAD: Deep transformer networks for anomaly detection in multivariate time series data. Proceedings of the VLDB Endowment, 15(6), 1201-1214.

[9] Zheng, W., Zhong, J., Zhang, Q. and Zhao, G. (2022). MTT: An efficient model for encrypted network traffic classification using multi-task transformer. Applied Intelligence, 52, 10741-10756.

[10] Sarhan, M., Layeghy, S. and Portmann, M. (2022). Towards a standard feature set for network intrusion detection system datasets. Mobile Networks and Applications, 27, 357-370.

[11] Huoh, T.L., Luo, Y., Li, P. and Zhang, T. (2023). Flow-based encrypted network traffic classification with graph neural networks. IEEE Transactions on Network and Service Management, 20, 1224-1237.

[12] Mayer, K.S., Soares, J.A., Pinto, R.P., Rothenberg, C.E., Arantes, D.S. and Mello, D.A.A. (2021). Machine-learning-based soft-failure localization with partial software-defined networking telemetry. Journal of Optical Communications and Networking, 13(10), E122-E131.

[13] Babbar, J., Triki, A., Ayassi, R. and Laye, M. (2022). Machine learning models for alarm classification and failure localization in optical transport networks. Journal of Optical Communications and Networking, 14(8), 621-628.

[14] Khan, L.Z., Pedro, J., Costa, N., De Marinis, L., Napoli, A. and Sambo, N. (2023). Data augmentation to improve performance of neural networks for failure management in optical networks. Journal of Optical Communications and Networking, 15(1), 57-67. https://doi.org/10.1364/JOCN.472605.

[15] Abdelli, K., Tropschug, C., Griesser, H. and Pachnicke, S. (2023). Faulty branch identification in passive optical networks using

machine learning. Journal of Optical Communications and Networking, 15(4), 187-196. https://doi.org/10.1364/JOCN.475882.

[16] Zeng, C., Zhang, J., Wang, R., Zhang, B. and Ji, Y. (2023). Multiple attention mechanisms-driven component fault location in optical networks with network-wide monitoring data. Journal of Optical Communications and Networking, 15(7), C9-C19.

[17] Fosic, I., Zagar, D., Grgic, K. and Krizanovic, V. (2023). Anomaly detection in NetFlow network traffic using supervised machine learning algorithms. Journal of Industrial Information Integration, 33, 100466.

[18] Han, D., Zhou, H., Weng, T.H., Wu, Z., Han, B., Li, K.C. and Pathan, A.S.K. (2023). LMCA: A lightweight anomaly network traffic detection model integrating adjusted MobileNet and coordinate attention mechanism for IoT. Telecommunication Systems, 84, 549-564.

[19] Ren, J., Zhang, F., Wang, Z., Li, J., Zhang, Y., Pan, S. and Yu, P.S. (2023). Graph learning for anomaly analytics: Algorithms, applications, and challenges. ACM Transactions on Knowledge Discovery from Data, 17. https://doi.org/10.1145/3570906.

[20] Kruse, L.E., Kuhl, S., Dochhan, A. and Pachnicke, S. (2024). Experimental investigation of machine-learning-based soft-failure management using the optical spectrum. Journal of Optical Communications and Networking, 16(2), 94-103. https://doi.org/10.1364/JOCN.500930.

[21] Straub, M., Reber, J., Saier, T., Borkowski, R., Li, S., Khomchenko, D., Richter, A., Farber, M., Kafer, T. and Bonk, R. (2024). ML approaches for OTDR diagnoses in passive optical networks-event detection and classification: Ways for ODN branch assignment. Journal of Optical Communications and Networking, 16(7), C43-C50.

[22] Han, X., Xu, G., Zhang, M., Yang, Z., Yu, Z., Huang, W. and Meng, C. (2024). DE-GNN: Dual embedding with graph neural network for fine-grained encrypted traffic classification. Computer Networks, 245, 110372.

[23] Yang, Z., Ma, Z., Zhao, W., Li, L. and Gu, F. (2024). HRNN: Hypergraph recurrent neural network for network intrusion detection. Journal of Grid Computing, 22, 52.

[24] Wang, T., Xie, X., Wang, W., Wang, C., Zhao, Y. and Cui, Y. (2024). NetMamba: Efficient network traffic classification via pre-training unidirectional Mamba. Proceedings of the IEEE International Conference on Network Protocols, 1-11.

[25] Azab, A., Khasawneh, M., Alrabaee, S., Choo, K.K.R. and Sarsour, M. (2024). Network traffic classification: Techniques, datasets, and challenges. Digital Communications and Networks, 10, 676-692.

[26] Liu, Z., Huang, X., Zhang, J., Hao, Z., Sun, L. and Peng, H. (2024). Multivariate time-series anomaly detection based on enhancing graph attention networks with topological analysis. arXiv:2408.13082.

[27] Zhang, C., Chen, Y., Zhang, M., Liu, Z. and Wang, D. (2025). SHAP-assisted EE-LightGBM model for explainable fault diagnosis in practical optical networks. Journal of Optical Communications and Networking, 17(2), 81-94. https://doi.org/10.1364/JOCN.527872.

[28] Musumeci, F. and Tornatore, M. (2025). Failure management in optical networks with ML: A tutorial on applications, challenges, and pitfalls. Journal of Optical Communications and Networking, 17(8), C144-C155. https://doi.org/10.1364/JOCN.551910.

[29] Chen, X., Chen, X. and Zhu, Z. (2025). Interpretable optical network fault detection and localization with multi-task graph prototype learning. Journal of Optical Communications and Networking, 17(9), D73-D82. https://doi.org/10.1364/JOCN.562633.

[30] Karunakaran, V., Reyes, R.R., Shariati, B., Fischer, J.K., Autenrieth, A. and Bauschert, T. (2025). Dynamic network-aware soft failure localization using machine learning in optical networks. Journal of Optical Communications and Networking, 17(11), 1032-1046. https://doi.org/10.1364/JOCN.564177.