Learning from Scarce Observations as Inverse Recovery of Compact Functional Structure

Self-Supervision, Manifold Saturation, and Emergent Operational Classes

Eduardo Di Santi

College of Engineering and Applied Science
eduardo.disanti@colorado.edu

June 24, 2026

Abstract

The central claim is that learning from scarce observations is not primarily a supervised classification problem, but an inverse recovery problem over compact functional structures. In this view, classes are not necessarily predefined labels. They may emerge as stable compact manifolds or submanifolds in an appropriate representation space. Once these structures stabilize, expert feedback can assign semantic or physical meaning to them: normal operation, friction, heat, harmless variation, degradation, emerging fault, or unknown regime.

1 Core Thesis

We propose that self-supervised learning and few-shot anomaly detection can be formulated as inverse problems over functional observations.

The classical supervised learning view assumes that classes are given in advance and that the main objective is to learn a decision function from labeled samples. In contrast, many real-world diagnostic settings do not provide a complete class taxonomy, abundant labels, or exhaustive examples of failure modes.

In such settings, the system observes signals generated through an unknown process and must recover the functional structure that explains them. The primary object is therefore not a supervised decision boundary, but the latent geometry of the observed phenomenon.

The central thesis is:

> *Learning from scarce observations is not primarily a supervised classification problem. It is an inverse recovery problem over compact functional structures.*

In this view, classes are not assumed a priori. They may emerge as stable compact manifolds, submanifolds, or functional regions in a suitable representation space.

2 Forward Generation and Inverse Recovery

Let a physical or functional system generate observations through an unknown process:

$$x = F(\theta, \eta), \tag{1}$$

where x is an observed signal or sample, θ represents the latent functional regime, and η represents nuisance variability such as noise, environment, operating conditions, sensor effects, or initial conditions.

The direct physical process is:

$$\theta \longrightarrow x. \tag{2}$$

However, learning and diagnosis require the opposite direction:

$$x \longrightarrow \theta. \tag{3}$$

More realistically, the system does not recover the full latent state θ, but stable invariants of it:

$$x \longrightarrow I(\theta), \tag{4}$$

where $I(\theta)$ denotes functional invariants of the latent regime.

This is an inverse problem. The concept of inverse problems in science refers to calculating from a set of observations the causal factors that produced them [4]. In machine learning, learning from examples has been formally connected to inverse problems, showing that the discrete problem solved in practice can be seen as the stochastic discretization of an infinite-dimensional inverse problem [2].

This is generally ill-posed because the forward map may be unknown, non-injective, noisy, high-dimensional, or only partially observed. Multiple latent causes may produce similar observations, while the same cause may generate different signals under different operating conditions.

According to Hadamard's definition, a problem is well-posed if: (1) a solution exists, (2) the solution is unique, and (3) the solution depends continuously on the input data [5]. If at least one condition fails, the problem is ill-posed [9]. Inverse problems are typically ill-posed because they lack uniqueness or stability [8].

Therefore, the learning system requires regularization. Regularization stabilizes ill-posed inverse problems by replacing them with nearby well-posed problems [9].

3 Compact Functional Manifolds

We assume that physically coherent regimes do not occupy arbitrary regions of observation space. Under a suitable representation,

$$\phi : X \longrightarrow Z, \tag{5}$$

observed samples concentrate around compact or approximately compact functional regions:

$$\phi(X) \approx \bigcup_{j=1}^{k} \mathcal{M}_j, \tag{6}$$

where each $\mathcal{M}_j$ is a compact manifold, submanifold, or stable functional region corresponding to a coherent operational regime.

The key point is that these regions need not be known in advance. As observations accumulate, the geometry of the representation begins to stabilize. The system does not first learn labels. It first recovers structure.

An operational class can then be defined as a stabilized functional manifold:

$$\mathcal{C}_j := \mathcal{M}_j. \tag{7}$$

A new observation is assigned to a class by geometric membership:

$$f(x) = L\left(\arg\min_j d\big(\phi(x), \mathcal{M}_j\big)\right), \tag{8}$$

provided that

$$\min_j d\big(\phi(x), \mathcal{M}_j\big) \leq \tau_j. \tag{9}$$

If no such manifold is sufficiently close, the system should abstain:

$$f(x) = \text{unknown/anomaly/new regime.} \tag{10}$$

Thus, anomaly detection is not forced classification. It is controlled non-membership with abstention.

4 Blueprint Saturation

Blueprints provide a way to measure whether the observed functional space has stabilized.

Given a growing sequence of observations,

$$x_1, x_2, \ldots, x_n, \tag{11}$$

we study the evolution of geometric descriptors such as entropy, radius, mean radius, volume, coverage, or neighborhood growth:

$$H(n), \qquad r_{\max}(n), \qquad r_{\text{mean}}(n), \qquad V(n). \tag{12}$$

When these quantities saturate,

$$H(n) \to H^*, \tag{13}$$

$$r_{\max}(n) \to r^*, \tag{14}$$

$$r_{\text{mean}}(n) \to \bar{r}^*, \tag{15}$$

$$V(n) \to V^*, \tag{16}$$

we interpret this as evidence that the system has captured a stable functional region rather than merely accumulating isolated samples.

This gives an operational criterion:

> *A regime becomes learnable when its functional geometry becomes stable under additional observations.*

This also reframes few-shot anomaly detection. With limited data, the objective is not to learn a complete supervised classifier, but to determine whether the observed samples already approximate a stable compact region. If they do, future deviations can be evaluated relative to that recovered structure.

5 Emergent Classification Through Expert Feedback

Once compact functional manifolds are recovered, expert feedback can be applied at the manifold level rather than at the sample level.

Instead of labeling thousands of individual points,

$$x_i \mapsto y_i, \tag{17}$$

the expert labels stabilized regions:

$$\mathcal{M}_j \mapsto y_j. \tag{18}$$

For example:

$$\mathcal{M}_1 \mapsto \text{normal}, \tag{19}$$

$$\mathcal{M}_2 \mapsto \text{friction}, \tag{20}$$

$$\mathcal{M}_3 \mapsto \text{heat}, \tag{21}$$

$$\mathcal{M}_4 \mapsto \text{harmless variation}, \tag{22}$$

$$\mathcal{M}_5 \mapsto \text{emerging degradation}. \tag{23}$$

The classifier is therefore not trained first and interpreted later. It is induced after geometric stabilization.

This gives a compact expert-in-the-loop formulation:

> *The domain expert does not label isolated samples. The expert assigns physical meaning to stabilized geometric structures.*

This is especially relevant for edge systems, where data arrive continuously, labels are scarce, and unknown regimes may appear after deployment.

6 Self-Supervised Learning as Inverse Regularization

Self-supervised learning can be interpreted as a mechanism for constructing the representation space Z in which the inverse problem becomes geometrically regular.

Historical precedent: Self-Organized Maps (SOM), introduced by Kohonen [6, 7], represent an early approach to recovering latent structure from functional observations without labels. SOM constructs a low-dimensional representation grid

$$\phi : X \longrightarrow Z_{\text{grid}}, \tag{24}$$

such that similar inputs x_i, x_j map to nearby grid positions. In the terminology of this paper, SOM can be interpreted as an inverse-structural mechanism: it does not recover the full latent state that generated the observations, but it organizes observations into a topology-preserving representation where latent regularities become visible.

SOM achieves this through competitive learning:

$$w_{\text{best}} \leftarrow w_{\text{best}} + \alpha \cdot (x - w_{\text{best}}), \tag{25}$$

where w_{best} is the winning neuron and neighboring units are updated accordingly. This mechanism preserves neighborhood structure and provides an early example of representation learning as unsupervised geometric regularization.

Joint-embedding predictive architectures: Modern joint-embedding predictive architectures [1] provide a recent neural counterpart to this principle. Rather than reconstructing raw observations, JEPA-style methods learn to predict representations of unobserved or future parts of the input in latent space. In I-JEPA, for instance, an image context block is used to predict the embeddings of target blocks, encouraging the representation to capture semantic structure while avoiding pixel-level reconstruction.

This distinction is important for the present framework. Predicting in embedding space implicitly suppresses nuisance variability that is irrelevant to the predictive relation, while preserving structure that remains stable across context and target views. Thus, JEPA-style objectives do not solve the inverse problem explicitly. Instead, they help construct a representation space in which the inverse problem becomes better conditioned: latent functional invariants become more compact, stable, and recoverable.

Connection to current work: While SOM constructs a fixed low-dimensional grid and modern self-supervised learning typically learns ϕ through neural networks, both can be understood as mechanisms for regularizing inverse recovery. Their common role is to transform raw observations into representation spaces where latent functional regimes are more likely to appear as compact and separable structures.

Self-supervised objectives impose constraints such as

$$\phi(x) \approx \phi(Tx), \tag{26}$$

where T is an admissible transformation;

$$x_{\text{partial}} \longrightarrow x_{\text{missing}}, \tag{27}$$

for masked or incomplete observations;

$$x_t \longrightarrow x_{t+\Delta}, \tag{28}$$

for temporal prediction; or

$$\phi(x_1) \approx \phi(x_2), \tag{29}$$

when x_1 and x_2 are different views of the same underlying process.

These objectives do not merely train a model to solve artificial pretext tasks. They constrain the representation to preserve stable invariants across transformations, corruptions, partial observations, or temporal continuations. This is consistent with regularization theory for ill-posed inverse problems, where regularization compensates for inherent ill-conditioning [9].

Therefore:

> *Self-supervised learning regularizes inverse recovery by constructing representation spaces in which latent functional regimes become compact, stable, and separable.*

7 Consequences for Classification

In the proposed framework, classification occurs at two levels. The first is *structural classification*: a new observation is assigned to a recovered functional region, to an emerging unsaturated region, to a transition between regions, or to an unknown state. This classification does not require semantic labels. It is induced by the geometry of the representation space.

The second is *operational classification*: once a recovered region has stabilized, an expert or downstream policy may assign semantic meaning to it, such as normal operation, harmless variation, friction, heat, degradation, or a known fault regime.

Thus, classification is not absent before labels are assigned. Rather, it first appears as geometric membership, and only later becomes semantic classification.

Let $\{\mathcal{M}_j\}_{j=1}^{k}$ denote a family of recovered compact functional regions in representation space. A semantic or operational label is assigned at the manifold level through a map

$$L : \{\mathcal{M}_j\}_{j=1}^{k} \longrightarrow \mathcal{Y}, \tag{30}$$

with

$$L(\mathcal{M}_j) = y_j. \tag{31}$$

Here, y_j may denote an operational class such as normal operation, harmless variation, friction, heat, degradation, or a known fault regime.

Given a new observation x, classification is then defined by geometric membership:

$$j^*(x) = \arg\min_j d\big(\phi(x), \mathcal{M}_j\big), \tag{32}$$

provided that

$$d\big(\phi(x), \mathcal{M}_{j^*(x)}\big) \leq \tau_{j^*(x)}. \tag{33}$$

The predicted class is therefore

$$\hat{y}(x) = L\left(\mathcal{M}_{j^*(x)}\right). \tag{34}$$

If no manifold is sufficiently close, the system abstains:

$$\hat{y}(x) = \text{unknown}. \tag{35}$$

This differs from classical supervised classification. A supervised classifier assumes a predefined label set and learns a decision boundary between labeled samples. In contrast, the present framework first recovers compact functional structures and only then assigns meaning to them. Classification is therefore not forced assignment to a fixed taxonomy, but expert-grounded membership in stabilized regions.

This has several consequences:

- a class may correspond to a compact manifold rather than to an arbitrary label;
- multiple manifolds may share the same semantic label if they represent operational variants of the same regime;
- a single nominal label may be decomposed into several submanifolds when the underlying behavior is structurally heterogeneous;
- unknown observations need not be misclassified into the closest known class;
- expert feedback can label entire manifolds rather than individual samples;
- classification can evolve as new stable regimes are discovered after deployment.

Thus, classification becomes an interpretive layer over recovered geometry:

> *The system does not learn labels first. It first recovers stable functional regions; classification begins when those regions receive operational meaning.*

8 Consequences for Anomaly Detection

Anomaly detection follows naturally from the classification mechanism above. Once functional regions have been recovered, an anomaly is not merely a point with a high score, but a specific form of failure, instability, transition, or deformation relative to the recovered geometry.

In this framework, an anomaly may take several forms.

First, it may be a failure of membership:

$$\phi(x) \notin \bigcup_j N_{\tau_j}(\mathcal{M}_j), \tag{36}$$

where $N_{\tau_j}(\mathcal{M}_j)$ denotes a tolerated neighborhood around manifold $\mathcal{M}_j$.

Second, it may be an unstable emerging region that has not yet saturated.

Third, it may be a transition between known manifolds.

Fourth, it may be a deformation of a known manifold, indicating drift, degradation, or changing operating conditions.

This is richer than binary anomaly detection. The system can distinguish between:

- known normality;
- known harmless variation;
- known fault regimes;
- unknown but stable new regimes;
- unstable emerging regimes;
- true outliers;
- distributional drift.

The important point is that "unknown" is not an error. It is a valid output of a responsible diagnostic system.

9 Main Contribution

The proposed perspective reverses the usual order of learning.

The classical supervised view is:

$$\text{labels} \longrightarrow \text{classifier}. \tag{37}$$

The proposed inverse-geometric view is:

$$\text{observations} \longrightarrow \text{functional geometry} \longrightarrow \text{operational classes}. \tag{38}$$

This extends the functional-topological framework of [3] by providing explicit inverse problem formalization and operational saturation criteria.

In this view, classes emerge as stabilized compact functional manifolds. Expert feedback assigns meaning to these manifolds. Classification then follows from geometric membership, while anomaly detection follows from non-membership, instability, or the appearance of new unsaturated regions.

The central principle can be summarized as follows:

> *Classes are not assumed. They emerge as expert-interpreted compact functional manifolds recovered from scarce observations.*

More generally:

> *Self-supervised learning and few-shot anomaly detection are regularized inverse problems whose objective is the recovery of stable functional invariants from incomplete observations of an unknown generative process.*

10 Preliminary Positioning

This framework extends and refines the functional-topological foundation presented in [3], where compact perceptual manifolds were empirically validated across five real-world domains.

Related work: Self-Organized Maps (SOM) by Kohonen[6, 7] represent an earlier approach to inverse problem recovery over functional structure. SOM's competitive learning preserves topological structure without labels, achieving self-supervision [7]. Our work differs by:

1. Using **geometric clustering over extracted features** (not fixed grid);
2. Providing explicit inverse problem formalization ($\theta \to x$, $x \to \theta$);
3. Introducing blueprint saturation as operational learnability criterion;
4. Extending to manifold-level labeling (not sample-level).

The theoretical connection between learning and inverse problems has been established in prior work: De Vito et al. formally connected learning from examples to inverse problems, showing consistency approaches in learning theory align with stability convergence in ill-posed inverse problems [2]. Rosasco et al. translated learning problems into regularization theory for ill-posed inverse problems, deriving consistency results and optimal regularization parameter choice [8].

The current work advances that foundation by:

1. Reformulating self-supervised learning and few-shot anomaly detection as **inverse problems** over functional observations;

2. Providing explicit mathematical formalization of the direct versus inverse directions:

$$\theta \longrightarrow x \quad \text{(direct)}, \qquad x \longrightarrow \theta \quad \text{(inverse)};$$

3. Introducing the **blueprint saturation criterion** as an operational test for learnability:

$$H(n) \to H^*, \quad r_{\max}(n) \to r^*, \quad V(n) \to V^*;$$

4. Extending the human-in-the-loop formulation to **manifold-level** labeling rather than sample-level labeling:

$$\mathcal{M}_j \mapsto y_j \quad (\text{not } x_i \mapsto y_i).$$

11 Illustrative Example: MNIST Operational Manifold Coverage

To illustrate the proposed principle in a controlled setting, consider MNIST. Each image can be represented as a vector

$$x_i \in [0,1]^{784}. \tag{39}$$

In this example, MNIST is not used as a supervised classification benchmark, nor is the objective to propose a new classifier for handwritten digits. Instead, MNIST is used as a simple and familiar observation space in which semantic labels can be withheld during geometric analysis and introduced only afterward as expert feedback.

Handwritten digits do not occupy arbitrary regions of $\mathbb{R}^{784}$. They concentrate around recurring morphological structures corresponding to digit forms. In a suitable representation space,

$$\phi : X \longrightarrow Z, \tag{40}$$

the empirical observation space may therefore be approximated as

$$\phi(X_{\text{MNIST}}) \approx \bigcup_{d=0}^{9} \mathcal{M}_d, \tag{41}$$

where each $\mathcal{M}_d$ denotes an operational manifold associated, a posteriori, with one digit class.

In the present illustration, the labels are used only to identify the operational manifolds after the representation has been constructed. They play the role of expert feedback:

$$\mathcal{M}_d \mapsto d. \tag{42}$$

The experiment then asks a different question from standard classification: once operational manifolds have been identified, do additional observations primarily expand the manifold support, or do they mostly densify regions that are already covered?

For a new observation x, let $z = \phi(x)$. Given a partially observed manifold memory built from the first n observations, we define the distance to the correct expert-identified manifold as

$$d_{\text{in}}(z) = d\left(z, \mathcal{M}y^{(n)}\right), \tag{43}$$

where y is the expert label of x, and $\mathcal{M}y^{(n)}$ is the currently observed support of that operational manifold. We also define the distance to the nearest competing manifold as

$$d_{\text{out}}(z) = \min_{c \neq y} d\left(z, \mathcal{M}c^{(n)}\right). \tag{44}$$

The membership margin is then

$$\gamma(z) = d\text{out}(z) - d_{\text{in}}(z). \tag{45}$$

A positive margin indicates that the observation is geometrically closer to its own operational manifold than to any competing one. This does not require perfect classification. Rather, it measures whether the operational manifold structure is becoming geometrically separable as observations accumulate.

We evaluate three quantities as the number of observations n increases. First, we measure the fraction of observations closer to their own operational manifold than to any competing manifold. Second, we measure the median membership margin $\gamma(z)$. Third, we measure the empirical radius of each operational manifold, using the mean p95 class radius as a robust estimate of manifold extent.

Figure 1 shows the early coverage of expert-identified operational manifolds. The grey background denotes the final embedded domain, while colored points show only the observations available up to $n = 500$, $n = 1000$, and $n = 5000$. Already by $n = 500$–1000, the main operational regions are visibly represented. Later observations primarily densify the same regions rather than creating a qualitatively different geometry.

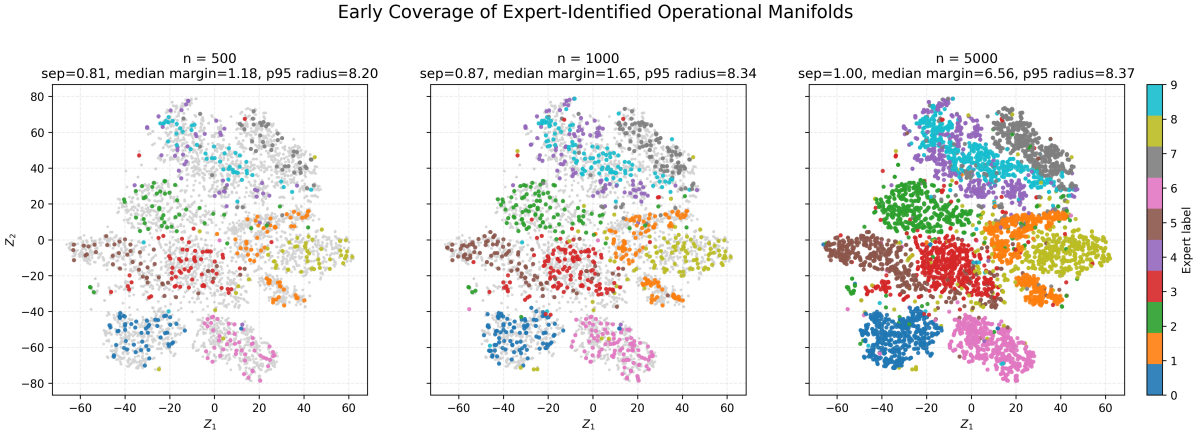

Figure 1: Early coverage of expert-identified operational manifolds in MNIST. The grey background shows the final embedded observation domain, while colored points show the observations available up to each value of n. The figure illustrates that the main operational regions are already represented at relatively small sample sizes; additional observations mainly densify the recovered support.

Figure 2 quantifies this effect. The fraction of geometrically separable observations increases rapidly, the membership margin becomes positive and grows as the manifolds are covered, and the p95 class radius stabilizes early. This indicates that the empirical extent of the operational manifolds is largely recovered before the full dataset is observed.

Figure 3 further shows that this saturation is not only a global effect. Individual digit manifolds exhibit different empirical radii, reflecting different morphological variability, but most of them stabilize early. This supports the interpretation of operational classes as compact functional regions with class-specific geometric extent.

This example should be interpreted as a limiting case in which the observation space is already highly structured. MNIST images are centered, normalized, low-clutter, and morphologically regular. In more complex domains, such as industrial edge diagnostics, raw sensor observations or raw images may not separate directly. In those cases, the representation map $\phi : X \to Z$ becomes essential. Self-supervised learning, temporal embedding, or domain-specific feature construction can be understood as mechanisms for building a representation space in which analogous compact functional manifolds become visible.

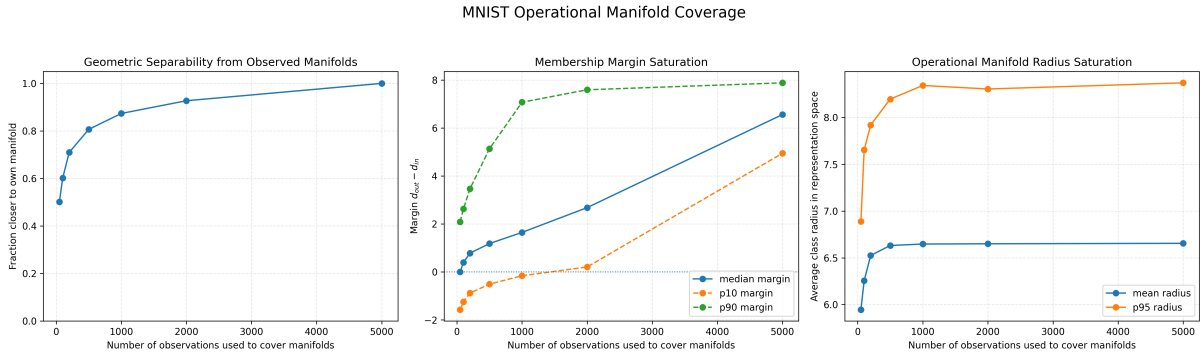

Figure 2: Operational manifold coverage on MNIST. Left: fraction of observations closer to their own expert-identified manifold than to any competing manifold. Middle: membership margin $\gamma = d_{\text{out}} - d_{\text{in}}$. Right: mean and p95 operational manifold radius. The p95 radius stabilizes early, while separability and margin continue to improve as the already recovered manifolds are densified.

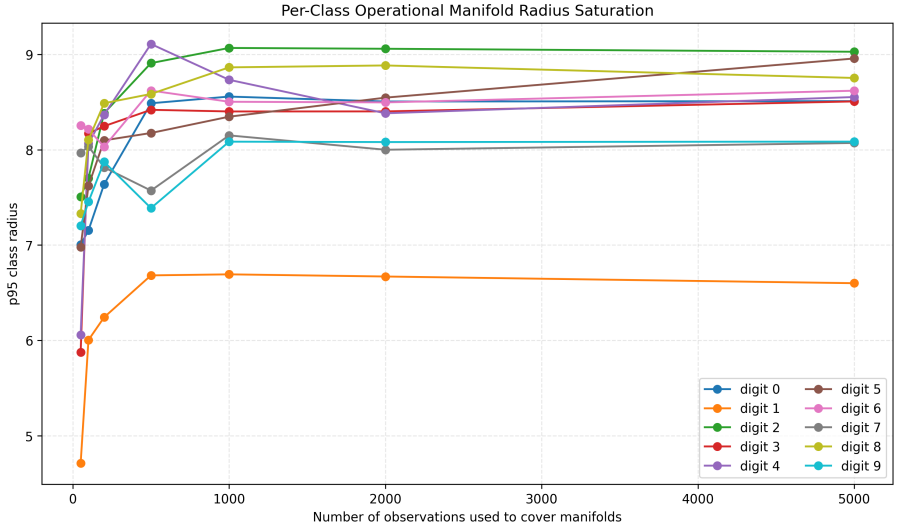

Figure 3: Per-class operational manifold radius saturation on MNIST. Each curve shows the p95 radius of one expert-identified digit manifold as observations accumulate. Most operational manifolds reach a stable empirical extent by approximately $n = 500$–1000, after which additional observations mainly densify the already recovered regions rather than expanding the class support.

Observations n	Geometric separability	Median margin	Mean p95 class radius
500	0.81	1.18	8.20
1000	0.87	1.65	8.34
5000	1.00	6.56	8.37

Table 1: MNIST operational manifold coverage. Labels are used only as expert feedback to identify operational manifolds. Geometric separability denotes the fraction of observations closer to their own expert-identified manifold than to any competing manifold. The final point $n = 5000$ corresponds to the full observed support and is included as a reference endpoint.

The point of the experiment is therefore not that MNIST can be classified geometrically. That is already well known. The point is that, once expert feedback identifies operational manifolds, their empirical support can be seen to stabilize early: additional observations mostly densify already recovered functional regions. This illustrates the inverse-problem view proposed in this paper. The primary object is not a supervised decision boundary, but the recovery and stabilization of compact functional structures to which operational meaning can later be assigned.

12 Conclusion

This paper has proposed a geometric inverse-problem view of autonomous edge diagnostics. The central claim is that, in settings where labels are scarce, failure modes are incomplete, and observations arrive from unknown physical or functional processes, learning should not be understood primarily as supervised classification. Rather, it should be understood as the inverse recovery of stable functional structure from observations.

Under this view, diagnostic classes are not necessarily assumed in advance. They may emerge as compact or approximately compact manifolds in a suitable representation space. Once these manifolds stabilize, they define operational regimes. Expert semantic assignment can then attach physical meaning to these regimes, while classification follows from geometric membership and anomaly detection follows from non-membership, instability, or the emergence of new unsaturated regions.

The role of self-supervised learning is therefore not merely to solve pretext tasks, but to construct representations in which the inverse problem becomes geometrically regular. By enforcing invariance, temporal consistency, reconstruction, or view agreement, self-supervised objectives help reveal compact functional manifolds that may otherwise be hidden in raw observation space.

Blueprint saturation provides an operational criterion for this process. A regime becomes diagnostically meaningful when its geometric descriptors stabilize under additional observations. This shifts the central question from whether a classifier can be trained from labeled examples to whether the functional geometry has stabilized enough to support reliable membership, abstention, and expert interpretation.

The resulting framework reframes few-shot anomaly detection, self-supervised diagnostics, and edge intelligence as instances of the same principle: learning from limited observations is inverse recovery of compact functional structure. Future work should formalize the saturation criteria, characterize the stability of recovered manifolds under representation maps, and evaluate the approach across visual, temporal, and industrial diagnostic domains.

Learning from limited observations is inverse recovery of compact functional structure.

References

[1] Mahmoud Assran, Quentin Duval, Ishan Misra, Piotr Bojanowski, Pascal Vincent, Michael Rabbat, Yann LeCun, and Nicolas Ballas. Self-supervised learning from images with a joint-embedding predictive architecture. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2023.

[2] Ernesto De Vito, Lorenzo Rosasco, Andrea Caponnetto, Umberto De Giovannini, and Francesca Odone. Learning from examples as an inverse problem. *Journal of Machine Learning Research*, 6(30):883–904, 2005.

[3] Eduardo Di Santi. The blueprints of intelligence: A functional-topological foundation for perception and representation, April 2026. 6 figures.

[4] H. W. Engl, M. Hanke, and A. Neubauer. Inverse problems: Associated with the reversal of the cause-effect sequence. *Nature Publishing*, 1996.

[5] Jacques Hadamard. Leçons sur les propriétés des fonctions analityiques. *Gauthier-Villars*, 1932.

[6] Teuvo Kohonen. Self-organized formation of topologically correct feature maps. *Biological Cybernetics*, 43:59–69, 1982.

[7] Teuvo Kohonen. *Self-Organizing Maps*. Springer, 3rd edition, 1995.

[8] Lorenzo Rosasco, Ernesto De Vito, and Andrea Caponnetto. Learning, regularization and ill-posed inverse problems. In *COLT*, 2004.

[9] A. N. Tikhonov and V. Y. Arsenin. *Solutions of Ill-Posed Problems*. Wiley, 1977.