International Journal of Computer Information Systems and Industrial Management Applications
ISSN 2150-7988 Volume 18(4s) 2026 pp. 492-503
https://doi.org/10.70917/ijcisim-2026-2531
Article

Exploring Linear Neural Network Models for Anomaly Detection: From Mathematical Foundations to Real-World Deployment

Pritika Mehra

Khalsa College for Women, Amritsar, Email: pritikacsc.rsh@gndu.ac.in

Abstract: The problem of anomaly detection plays an important role in many tasks related to cybersecurity, financial fraud prevention, and industry, and modern anomaly detection approaches rely predominantly on deep and nonlinear network architectures, such as convolutional and recurrent autoencoders, GANs (generative adversarial networks), and transformers. The current review provides an insight into the problem of using linear neural networks — networks comprised only of affine transformations without any nonlinear activation function — for anomaly detection. We discuss the fundamental theory behind the linear neural networks which is different from the one used for nonlinear networks, including the equivalence of linear autoencoders to PCA and the absence of spurious local minima in linear networks. We then catalogue the principal architectures used for anomaly detection with linear models, including reconstruction-based linear autoencoders, the Principal Component Classifier, multivariate statistical process control charts, and recent linear forecasting backbones such as DLinear. The core of the review surveys industrial case studies spanning network intrusion detection, credit card fraud detection, chemical process monitoring on the Tennessee Eastman benchmark, and predictive maintenance in industrial IoT, drawing out recurring patterns in why practitioners choose linear models. A comparative analysis contrasts linear and nonlinear approaches on expressiveness, computational cost, interpretability, and robustness, and the review closes with open challenges and directions for hybrid linear–nonlinear systems.

Keywords: Linear neural networks, anomaly detection, principal component analysis, linear autoencoder, statistical process control, industrial IoT, time series forecasting.

1. INTRODUCTION

The problem of anomaly detection — finding data samples that differ from the norm — lies at the heart of contemporary cybersecurity, finance risk assessment, medical supervision, and manufacturing process control. The consequences of overlooking anomalies can be evaluated in terms of actual losses due to fraud, unexpected production stoppages, cybersecurity breaches, or accidents, which is why this domain has become one of the most studied by machine learning in the last twenty years..

The dominant narrative in recent anomaly detection literature favors increasingly elaborate nonlinear architectures: deep convolutional and recurrent autoencoders, variational and adversarial autoencoders, graph neural networks, and transformer-based sequence models. These architectures can, in principle, represent arbitrarily complex nonlinear manifolds of normal behavior. Yet a parallel and persistent thread of research and practice continues to rely on models that are, mathematically, nothing more than a composition of linear (affine) transformations — what this review terms linear neural network models. A single-layer linear autoencoder is algebraically equivalent to truncated Principal Component Analysis, and even a deep stack of linear layers collapses to a single linear map, yet such models remain in active use precisely because they are fast, stable to train, easy to interpret, and, in a growing number of empirical studies, surprisingly competitive with far more complex alternatives.

A particularly striking recent example is DLinear, a forecasting model built from a single linear layer applied after seasonal-trend decomposition, which was shown to outperform several contemporaneous transformer-based forecasting architectures on long-horizon time-series benchmarks, reopening debate about how much of the benefit attributed to architectural sophistication is actually earned through the forecasting strategy rather than the nonlinearity

of the model itself. In industrial settings, this debate has direct consequences: predictive maintenance systems, network intrusion detectors, and chemical process monitors must often run at high sampling rates, on constrained hardware, or under tight latency budgets where a linear model's lower computational footprint is not merely convenient but a deployment requirement.

This review consolidates the theoretical and applied literature on linear neural network models for anomaly detection. Section 2 establishes the mathematical foundations, including the PCA equivalence and the loss-landscape guarantees unique to linear networks. Section 3 surveys the principal architectural families used for anomaly scoring. Section 4 forms the core of the review: a domain-by-domain examination of industrial case studies in cybersecurity, finance, chemical process manufacturing, and industrial IoT, drawing on the published literature to illustrate how and why linear models are deployed in practice. Section 5 compares linear and nonlinear approaches directly, Section 6 discusses strengths, limitations, and open challenges, and Section 7 outlines promising directions before the review concludes in Section 8.

2. THEORETICAL FOUNDATIONS

2.1 Definition and Mathematical Formulation

A linear neural network model is a network in which every layer applies only an affine transformation — a weight matrix multiplication, optionally followed by a bias addition — with no intervening nonlinear activation function. For an L-layer network with weight matrices W_1, …, W_l, the end-to-end mapping reduces algebraically to a single product matrix $W = W_l W_{l-1} \dots W_1$, so that the function the network computes is exactly a linear (or affine, with biases) map from input to output regardless of depth. This is the defining property that separates linear neural networks from the nonlinear deep networks that dominate contemporary anomaly detection research.

For anomaly detection, the most common instantiation is the linear autoencoder. Given a mean-centered data matrix X, an encoder projects each observation into a lower-dimensional code, $Y = XWenc$, and a decoder reconstructs the original observation from that code, $\hat{X} = YWdec^T = XWencWdec^T$. Because every operation is linear, a multi-layer encoder or decoder can always be collapsed into a single equivalent weight matrix, so the expressive capacity of a linear autoencoder is fixed by its bottleneck width regardless of how many layers are used to reach it.

2.2 Equivalence to Principal Component Analysis

A foundational and frequently cited result, due to Baldi and Hornik (1989), is that a single-layer linear autoencoder trained to minimize squared reconstruction error converges, at the global optimum, to a solution that spans the same subspace as the leading principal components of the training data obtained through PCA. In other words, autoencoders built from a single layer with a linear activation function are, in this precise sense, nearly equivalent to PCA, even though autoencoders more generally can also implement nonlinear transformations that PCA cannot. This equivalence is the reason linear autoencoders, PCA, and Singular Value Decomposition (SVD) based reconstruction are often used interchangeably as baselines in the anomaly detection literature: they identify the same low-dimensional subspace of normal variation, and an observation that cannot be well reconstructed from that subspace is flagged as anomalous.

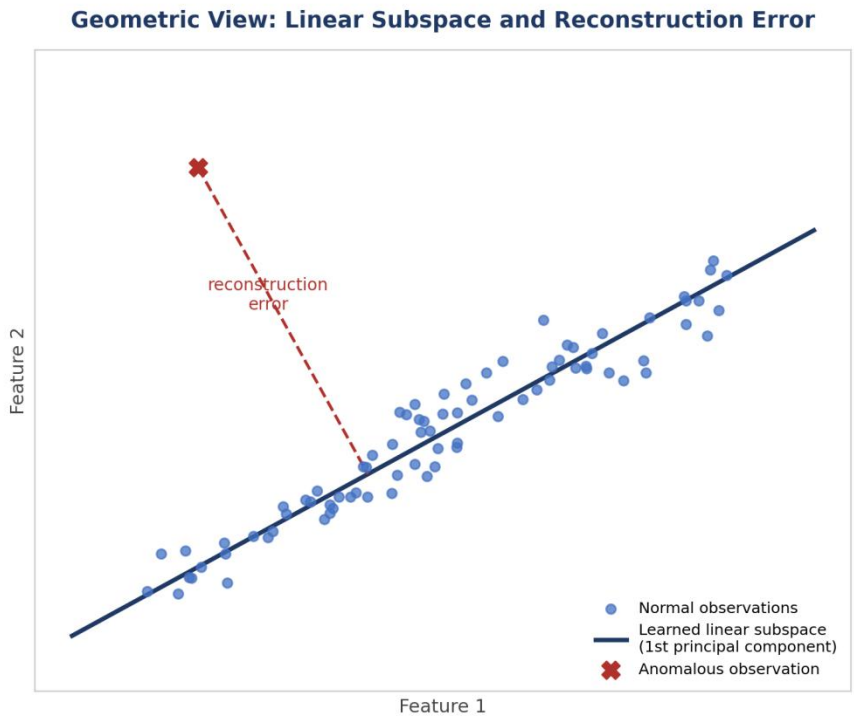

Figure 1. Normal observations cluster near the learned linear subspace; an anomalous point lies far from it, producing a large reconstruction error when projected back.

2.3 Optimization Landscape: No Spurious Local Minima

Beyond representational equivalence, linear networks possess an unusually favorable optimization landscape. Baldi and Hornik (1989) proved that for a shallow linear network trained with squared error loss, every local minimum is also a global minimum and the loss is convex in each weight matrix when the other is held fixed; they further conjectured that the same property holds for deep linear networks. This conjecture was proved by Kawaguchi (2016) and subsequently extended to broader loss functions and arbitrary depth by Zhou and Liang (2018) and by Laurent and von Brecht (2018), establishing that deep linear networks, despite being non-convex in their full parameter space, have a loss surface populated only by global minima and saddle points — with no "bad" local minima that could trap gradient-based optimization.

This guarantee has direct practical value for anomaly detection systems. Models must frequently be retrained on a rolling basis as new normal data accumulates, and in streaming or resource-constrained deployments there is little room for the unstable convergence, sensitive initialization, or extensive hyperparameter search that nonlinear deep networks can require. A linear model's loss landscape removes much of this uncertainty: gradient descent (or, equivalently, closed-form solutions via eigendecomposition) reliably reaches a globally optimal reconstruction subspace, which is one reason linear and PCA-based detectors remain attractive as both standalone systems and as fast, dependable baselines against which more complex models are benchmarked.

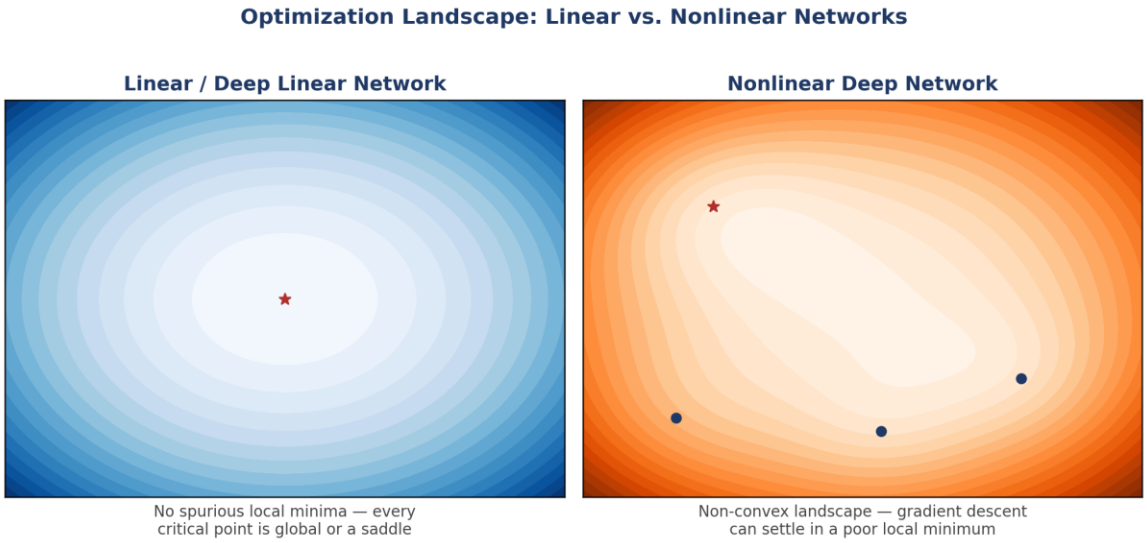

Figure 2. Linear and deep linear networks have a loss surface with no spurious local minima (left), whereas nonlinear deep networks can have multiple local minima that trap gradient descent (right).

2.4 Linear Models for Sequential and Time-Series Data

A separate but related theoretical development concerns linear models for sequence forecasting, which underlie prediction-based anomaly detection on time series. Zeng et al. introduced DLinear, which decomposes a time series into a trend component and a remainder (seasonal) component and applies a single linear layer to each before summing the results into a forecast. Despite its simplicity, DLinear was reported to outperform several contemporaneous transformer-based long-horizon forecasting models by a substantial margin on standard benchmarks, with the authors attributing transformer-based models' shortcomings less to a lack of representational power and more to the non-autoregressive direct multi-step forecasting strategy that DLinear also exploits. Because each branch of DLinear is a single linear layer, the model has an effectively constant-length signal path between any input and output position, which keeps memory use and inference latency low relative to attention-based alternatives. Subsequent work has nuanced this finding, showing that when transformer variants are matched in parameter count and tuned comparably, the performance gap narrows considerably, and that newer attention-based architectures such as PatchTST and iTransformer again outperform linear baselines on many benchmarks. The episode is nonetheless instructive: it demonstrates that the choice of forecasting strategy and evaluation protocol can matter as much as the nonlinearity of the underlying model, a lesson directly relevant to prediction-residual anomaly detectors built on either family of architecture.

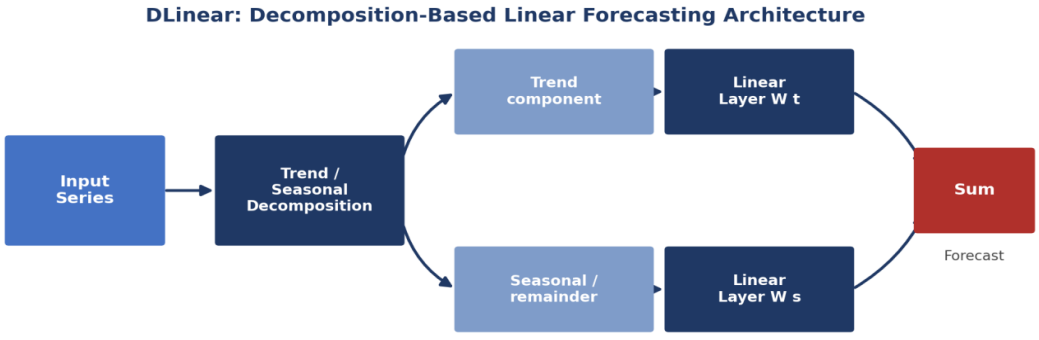

Figure 3. DLinear decomposes a time series into trend and seasonal components and applies a separate single linear layer to each before summing the results into a forecast.

3. ARCHITECTURES AND METHODOLOGIES FOR LINEAR-MODEL ANOMALY DETECTION

Linear neural network models are deployed for anomaly detection through several recurring architectural patterns. This section summarizes the four most widely used families before Section 4 examines how each has been applied across industries.

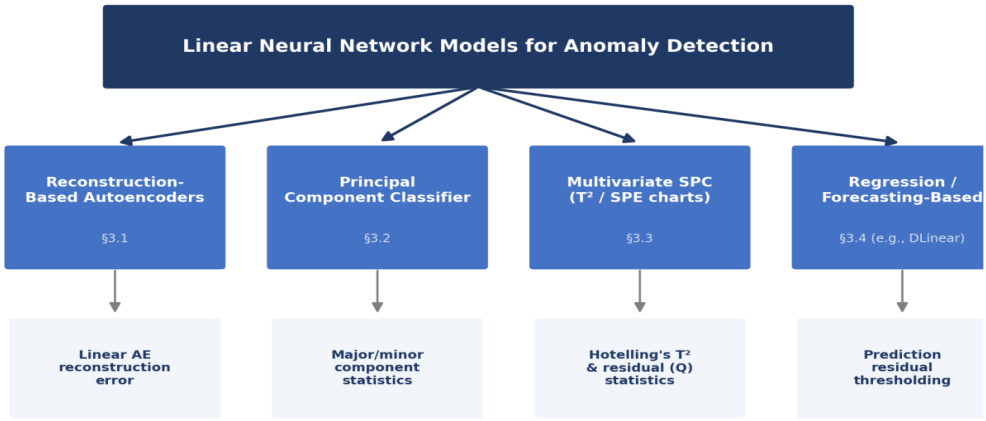

Figure 4. Four recurring architectural families used to build anomaly detectors from linear neural network models, each paired with its characteristic scoring statistic.

3.1 Reconstruction-Based Linear Autoencoders

The most direct architecture trains a linear encoder–decoder pair on data assumed to represent normal operation, then scores new observations by their reconstruction error — the squared distance between the input and its reconstruction. Because the bottleneck forces the network to discard information not captured by its leading linear directions, observations that do not conform to the learned subspace reconstruct poorly and receive a high anomaly score. The bottleneck width directly controls the tightness of this normal-data manifold: a narrower bottleneck increases sensitivity to anomalies but also raises the risk of underfitting normal variation, while a wider bottleneck risks reconstructing anomalies just as well as normal data, eroding the discriminative signal.

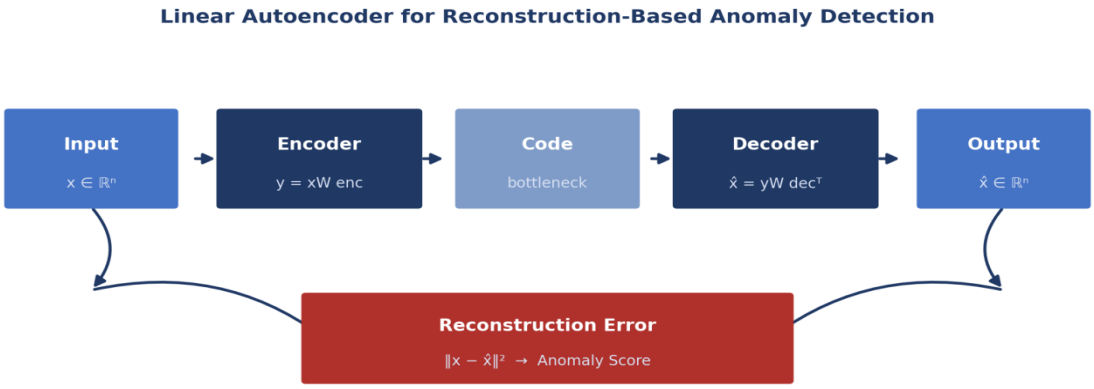

Figure 5. A linear encoder compresses the input into a bottleneck code and a linear decoder reconstructs it; the reconstruction error between input and output is used directly as the anomaly score.

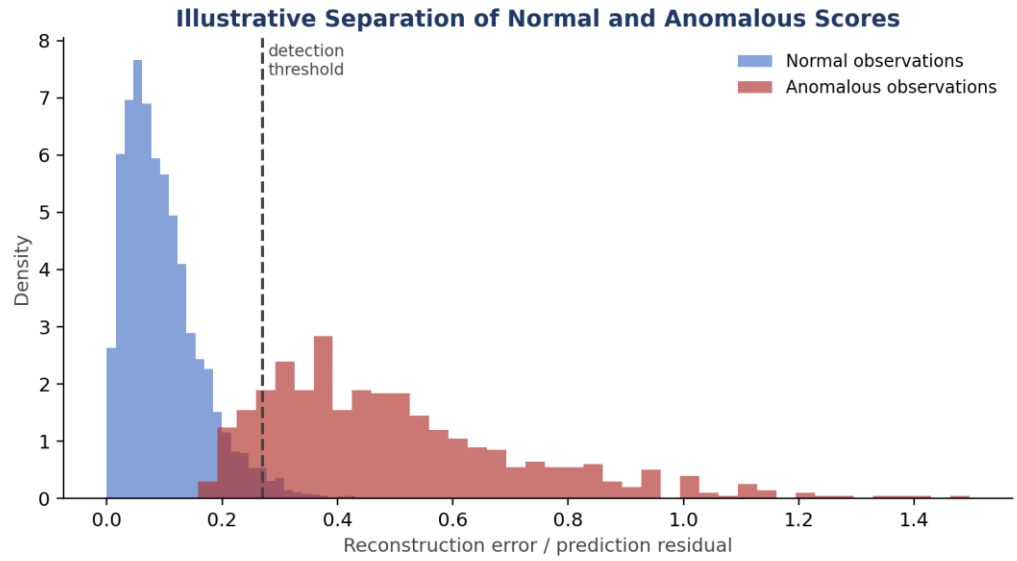

Figure 6. A conceptual illustration of how normal and anomalous observations separate along the reconstruction-error axis, with a calibrated threshold marking the detection boundary.

3.2 Principal Component Classifier

A widely cited alternative to raw reconstruction error is the Principal Component Classifier (PCC) of Shyu and colleagues, which separates the principal components learned by a PCA model into a “major” set capturing most of the data's variance and a “minor” set capturing the remainder. An observation is flagged as anomalous if it projects unusually strongly onto either set: large projections onto the major components indicate gross outliers relative to the dominant pattern, while large projections onto the minor components indicate subtle deviations that violate normally tight correlations among variables. This two-statistic formulation remains a standard reference baseline in network intrusion detection research.

3.3 Multivariate Statistical Process Control

In process industries, PCA-based monitoring is typically implemented through two complementary control statistics computed from the same linear subspace: Hotelling's T^2 statistic, which measures how far a new observation lies from the center of the normal-operation subspace, and the Squared Prediction Error (SPE), also called the Q statistic, which measures the residual left outside that subspace — conceptually identical to autoencoder reconstruction error. Together, T^2 and SPE charts give plant operators two interpretable signals: deviation within the modeled relationships among process variables, and deviation from those relationships altogether.

3.4 Linear Regression and Forecasting-Based Detectors

For time-ordered data, a complementary family of methods fits a linear model — ordinary linear regression, a linear autoregressive model, or a DLinear-style decomposition network — to predict each new observation from recent history, then flags large prediction residuals or values that exceed an empirically chosen percentile of the residual distribution as anomalous. This family is especially common in industrial telemetry and condition monitoring, where a regression-based threshold can be computed and updated with minimal computational overhead.

4. INDUSTRY APPLICATIONS AND CASE STUDIES

This section reviews how linear neural network models and their close statistical relatives have been applied across four major industrial domains, drawing on case studies and benchmark evaluations reported in the recent literature.

4.1 Cybersecurity and Network Intrusion Detection

Network intrusion detection has used PCA-based anomaly detection since the early 2000s, and the approach remains a standard component of contemporary research pipelines. A persistent challenge is that ordinary PCA is itself sensitive to outliers present in the training data; L1-norm PCA has been proposed as a more robust alternative, with experiments on the KDDCup99 intrusion dataset showing it more effectively preserves the low-dimensional structure of normal traffic and improves detection over conventional L2-norm PCA. Beyond standalone use, PCA is frequently paired with a downstream classifier: published intrusion detection systems have combined PCA with k-nearest-neighbor classification to improve attack detection, paired PCA with deep learning models to both detect and classify attacks, and applied PCA with group-wise feature selection techniques to improve feature relevance, illustrating a common pattern in which the linear model performs dimensionality reduction while a separate, often nonlinear, stage performs classification or diagnosis.

PCA-based detection has also remained relevant as a benchmark for emerging computational paradigms. A 2025 case study evaluating quantum machine learning methods for network security used three PCA-based intrusion detection algorithms — the Principal Component Classifier, a reconstruction-loss method, and an ensemble extension of the Principal Component Classifier — as the reference detectors against which quantum speedups were measured, applying them to the KDDCup99, CICIDS2017, and DARKNET datasets. The continued use of PCA-based detectors as the standard of comparison, even in research exploring next-generation computing hardware, underscores how durable linear anomaly detection has proven as both a production technique and an evaluation baseline in network security.

4.2 Financial Services: Credit Card Fraud Detection

Credit card fraud detection offers an unusually direct illustration of the relationship between PCA and linear neural networks, because the widely used ULB/Kaggle European credit card transaction dataset is itself released only after its original features have been transformed by PCA into twenty-eight anonymized components, in order to protect cardholder identities while still preserving the statistical structure needed for fraud detection. Researchers building on this dataset have then trained shallow autoencoders — in published implementations, as few as two fully connected linear layers with a bounded nonlinearity — on the majority class of legitimate transactions, flagging held-out transactions with high reconstruction error as likely fraud. This workflow is attractive in production because fraud is rare and the labeled fraud examples that do exist are reserved for validation rather than training, so an unsupervised linear or near-linear reconstruction model can be trained entirely on the abundant normal transaction history.

Comparative evaluations have examined this design choice directly: a study of credit card fraud detection in the Nigerian financial sector benchmarked autoencoder-based anomaly detection against a PCA algorithm under the same TensorFlow-based pipeline, reflecting a broader pattern in which financial institutions weigh the added complexity of a trainable autoencoder against the simpler, more transparent PCA baseline before committing to a production model. Other published work builds hybrid pipelines that retain the linear PCA preprocessing step but pair it with a more expressive downstream model, such as an autoencoder combined with gradient-boosted decision trees, to improve detection of the small minority of fraudulent transactions without discarding the efficiency of the linear front end.

4.3 Manufacturing and Process Industries: Statistical Process Monitoring

Chemical and process manufacturing has relied on PCA-based fault detection for several decades, and the Tennessee Eastman Process — a simulation of an Eastman Chemical Company facility introduced by Downs and Vogel in 1993 — has become the field's most widely used benchmark for evaluating plant-wide monitoring techniques. Despite the chemical process itself being highly nonlinear, published evaluations report that standard PCA-based fault detection can identify process faults quickly and effectively, which has made it a durable default even as the underlying physical system it monitors is anything but linear. This apparent paradox is explained by the local nature of fault detection: many faults manifest as a clear shift away from the normal operating region, which a linear subspace model can register as elevated reconstruction error or an out-of-control T^2/SPE statistic even when it cannot describe the full nonlinear dynamics of the process.

A substantial body of follow-on research builds directly on this linear foundation rather than discarding it. Sensitive PCA reweights which principal components are retained based on how strongly they reflect abnormal variation, improving online monitoring performance relative to classical PCA on Tennessee Eastman case studies. Multi-scale PCA combined with an adaptive neuro-fuzzy inference system improves sensitivity to small process

changes, particularly for faults with strongly nonlinear characteristics. Other studies pair PCA-extracted features with a neural classifier trained via a genetic algorithm to reduce missed detections and detection delay across the benchmark's standard fault scenarios. Across this literature, the recurring design pattern is the same one seen in cybersecurity and finance: a linear model performs the core dimensionality reduction and monitoring statistic, while a nonlinear or search-based component is layered on top to sharpen fault classification or adapt to specific nonlinear fault signatures.

4.4 Predictive Maintenance and Industrial IoT

In industrial IoT and predictive maintenance, computational and connectivity constraints often make linear models a deliberate engineering choice rather than a fallback. A published predictive maintenance system for industrial compressors integrates IoT-collected temperature and pressure readings with a linear regression model and a percentile-based statistical threshold to flag anomalous readings in real time, with the authors explicitly motivating the use of linear regression over deep learning on the grounds that the latter was not feasible given the system's real-time monitoring constraints; the study concludes more generally that model complexity should be matched to the complexity of the underlying data, since simpler models can resolve problems that do not require deep architectures. A related maintenance decision-support framework applied across a fleet of twenty-nine machine tool linear axes over more than four years of operating data illustrates how statistical and machine-learning techniques are combined to estimate long-term reliability indicators such as the rate of unplanned breakdowns.

Resource constraints also shape architectural choices at the sensor level itself. Research on anomaly detection for resource-constrained Industry 4.0 devices and on minimal-configuration monitoring for industrial IoT sensors both emphasize that wireless bandwidth and on-device power budgets limit how much data and computation can be devoted to anomaly scoring, favoring lightweight, often linear, models that can run locally rather than transmitting raw high-frequency sensor streams to the cloud for processing by larger nonlinear networks.

4.5 Time-Series and Streaming Monitoring at Scale

Linear forecasting backbones have begun to extend into operational anomaly detection beyond their original forecasting benchmarks. In power grid security research, autoencoder-based detectors built from simple feed-forward and recurrent linear-style layers have been validated against attention-based variants for detecting false data injection attacks on sensor measurements, with detection performed by thresholding reconstruction error using a precision–recall analysis; such studies typically find that the additional architectural complexity of attention mechanisms yields only incremental gains over simpler reconstruction-based detectors on these benchmarks. More broadly, the efficiency characteristics that made DLinear competitive for forecasting — a short, constant-length path from input to output and substantially lower memory and parameter counts than transformer alternatives — are directly relevant to streaming anomaly detection in network telemetry and industrial monitoring, where detectors must process high-volume sensor or traffic data with tight latency budgets.

5. COMPARATIVE ANALYSIS: LINEAR VERSUS NONLINEAR MODELS

The case studies above reveal a consistent set of trade-offs between linear neural network models and their nonlinear, deep-learning counterparts. Table 1 summarizes these trade-offs along the dimensions most often cited in the literature reviewed.

Dimension	**Linear Neural Network Models**	**Nonlinear / Deep Models**
Expressiveness	Limited to the linear subspace spanned by the learned weights; equivalent to PCA at the global optimum	Can approximate curved, nonlinear manifolds of normal behavior
Optimization	No spurious local minima; gradient descent provably reaches a global optimum	Non-convex landscape; convergence and initialization sensitivity are common concerns
Computational cost	Low; closed-form or fast iterative solutions, small memory footprint, low inference latency	Higher training and inference cost; larger parameter counts, especially for attention-based models

Interpretability	High; principal directions and loadings can be inspected for root-cause diagnosis	Lower; hidden representations are harder to attribute to specific input variables
Data requirements	Performs well with modest amounts of normal-condition data	Typically benefits from larger training sets to avoid overfitting
Robustness to crafted anomalies	Reconstruction can be near-perfect for adversarially placed points within the learned subspace	Susceptible to the same class of reconstruction-based blind spots; greater capacity does not by itself resolve the issue
Typical deployment context	Edge/IoT sensors, real-time process control, fast baselines and screening filters	Offline or cloud-based analysis of complex, high-dimensional, strongly nonlinear data such as images or raw audio

Table 1. Linear versus nonlinear neural network models for anomaly detection.

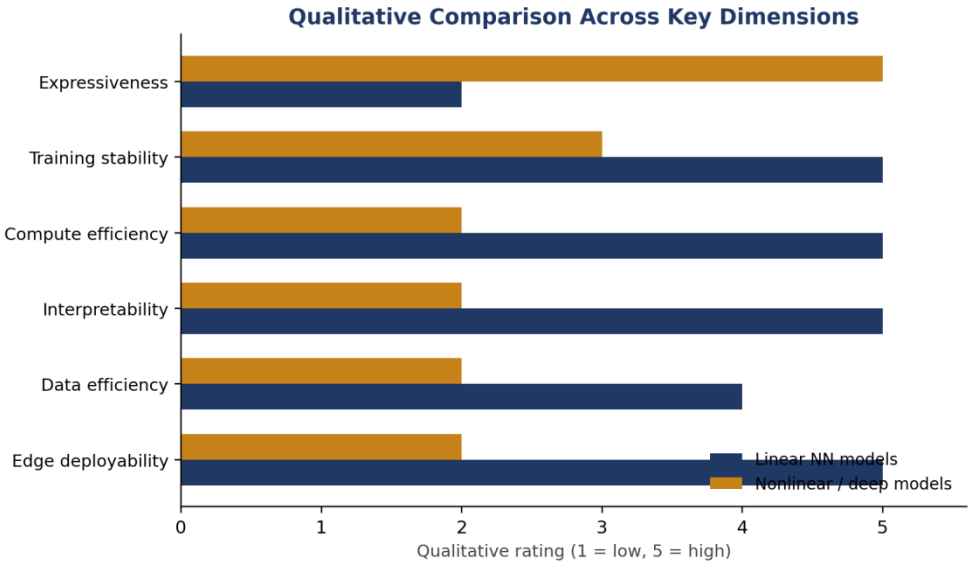

Figure 7. A qualitative synthesis of the trade-offs summarized in Table 1, rated on a 1 (low) to 5 (high) scale based on the patterns reported across the literature reviewed.

A point of particular interest in recent literature is that the vulnerability of reconstruction-based detection to carefully crafted anomalies is not unique to linear models. A 2025 analysis demonstrates that, because a linear autoencoder reduces mathematically to PCA, it shares PCA's susceptibility to out-of-bounds reconstruction: adversarial points can be constructed that lie far from any genuine normal observation yet still achieve near-zero reconstruction loss, since they fall within the span of the learned linear subspace. The same analysis extends this argument to networks with bias terms and non-centered data, and the broader implication is that increasing model capacity — moving from a linear to a deep nonlinear autoencoder — does not automatically close this vulnerability, since a sufficiently expressive nonlinear autoencoder can likewise learn to reconstruct adversarially designed inputs well. This finding tempers a common assumption that nonlinear models are categorically more secure, and instead suggests that detection thresholds, ensemble methods, or complementary non-reconstruction-based signals are needed regardless of which family of model is deployed.

Practitioner guidance distilled from applied case studies converges on a similar message: begin with a PCA or linear-autoencoder baseline, since for many tabular, sensor, or transaction datasets the normal-condition data lies close enough to a linear subspace that a simple model already achieves strong separation between normal and anomalous observations; only graduate to a deep nonlinear autoencoder when the underlying data manifold is demonstrably

curved — for example, sensor readings that follow a cyclical or strongly nonlinear pattern — such that the linear assumption visibly limits detection accuracy.

6. STRENGTHS, LIMITATIONS, AND OPEN CHALLENGES

6.1 Strengths

Guaranteed convergence: the absence of spurious local minima means training reliably reaches a globally optimal reconstruction subspace, which is valuable for systems that must be retrained frequently or run with limited tuning budgets.

Computational efficiency: low parameter counts and short signal paths translate into fast training, low memory use, and low inference latency, which is often a hard requirement in edge, IoT, and real-time process-control settings.

Interpretability: principal directions, loadings, and regression coefficients can be inspected directly, supporting root-cause diagnosis in ways that are harder to achieve with deep nonlinear hidden representations.

Data efficiency: linear models can produce useful anomaly scores from comparatively small amounts of normal-condition data, which matters in domains where representative training data is costly or slow to accumulate.

Competitive accuracy in linear-subspace settings: when normal data genuinely lies close to a linear subspace — as is common for many sensor, transaction, and network-traffic datasets — linear models can match or exceed more complex alternatives, as demonstrated by DLinear in time-series forecasting and by PCA-based detectors that remain reference baselines in network security research.

6.2 Limitations

Restricted expressiveness: a linear model cannot represent curved or strongly nonlinear manifolds of normal behavior, leading to degraded detection accuracy when the true data structure departs substantially from a linear subspace.

Threshold calibration: the boundary between normal and anomalous reconstruction error or regression residual is a chosen hyperparameter rather than an intrinsic property of the data, and it must be calibrated carefully against domain knowledge or labeled validation data.

Vulnerability to crafted or in-subspace anomalies: points engineered to lie within the learned linear subspace can be reconstructed with near-zero error despite being genuinely anomalous, a limitation shared with — not solved by — nonlinear autoencoders.

Sensitivity to non-stationarity: industrial processes, network traffic, and financial behavior drift over time, and a fixed linear subspace estimated from historical data can become stale, requiring periodic retraining or adaptive variants.

Preprocessing burden for mixed data types: linear models operate naturally on continuous numeric features, so categorical or mixed-type data require embedding or encoding steps before a linear model can be applied effectively.

6.3 Open Challenges

Several challenges remain open in the literature reviewed. First, principled methods for threshold selection that adapt automatically to changing baseline conditions, rather than relying on a fixed contamination-rate assumption, remain an active area of development. Second, robust variants of PCA and linear autoencoders that resist contamination from anomalies inadvertently present in the training data — as addressed in part by L1-norm and sensitive PCA formulations — deserve continued attention given how frequently “normal” training data in practice contains some unlabeled anomalies. Third, the adversarial vulnerability shared by linear and nonlinear reconstruction-based detectors suggests a need for detection strategies that do not rely solely on reconstruction error, such as combining reconstruction-based scores with density-based or distance-based statistics. Finally, the debate sparked by DLinear over whether transformer-based architectures earn their added complexity highlights a broader open question for the field: under what precise data conditions does nonlinearity provide a genuine, reproducible advantage for anomaly detection, as opposed to an advantage that is more attributable to evaluation protocol, parameter budget, or training strategy.

7. FUTURE DIRECTIONS

Hybrid cascaded pipelines that use a fast linear model as a first-pass filter to triage the large majority of clearly normal observations, escalating only ambiguous cases to a more expensive nonlinear model, combining the speed of linear screening with the expressiveness of deep models where it is actually needed.

Federated and on-device anomaly detection that trains lightweight linear models locally on edge or IoT hardware, reducing the bandwidth and privacy costs of transmitting raw sensor or transaction data to a central server for nonlinear model training.

Robust and drift-adaptive linear subspace estimation that updates the learned principal directions incrementally as operating conditions evolve, narrowing the gap between the theoretical guarantees of linear optimization and the practical reality of non-stationary industrial and financial data.

Deeper theoretical study of generalization and adversarial robustness for linear anomaly detectors, building on existing loss-landscape results to characterize precisely when a linear subspace model is provably safe against crafted anomalies and when complementary detection signals are required.

Wider application of linear forecasting backbones such as DLinear to residual-based anomaly detection in network telemetry, manufacturing telemetry, and other high-volume streaming settings where the efficiency advantages demonstrated in forecasting benchmarks are likely to transfer directly.

8. CONCLUSION

Linear neural network models occupy a durable and well-justified place in the anomaly detection toolkit. Their theoretical grounding — the formal equivalence between linear autoencoders and PCA, together with the guarantee that linear and deep linear networks have no spurious local minima — gives them optimization properties that nonlinear deep networks cannot match, while their low computational cost and interpretability continue to make them the default choice in latency-sensitive, resource-constrained, or diagnosis-oriented deployments. The industrial case studies surveyed in this review, spanning network intrusion detection, credit card fraud detection, chemical process monitoring on the Tennessee Eastman benchmark, and predictive maintenance in industrial IoT, show a consistent pattern: practitioners frequently retain a linear model as the core monitoring statistic and layer nonlinear or search-based components on top only where they demonstrably add value, rather than replacing the linear core outright. At the same time, the comparative evidence — from DLinear's strong forecasting results to the persistence of PCA-based detectors as security benchmarks — cautions against treating nonlinearity as an unconditional improvement, while the shared vulnerability of linear and nonlinear reconstruction-based detectors to crafted anomalies underscores that model capacity alone does not resolve every limitation of the reconstruction-error paradigm. As anomaly detection systems continue to be deployed at greater scale and closer to the network or sensor edge, linear neural network models, used either standalone or as the efficient backbone of a hybrid system, are likely to remain a central rather than a legacy component of the field.

References

1. Achour, E. M., Malgouyres, F., & Gerchinovitz, S. (2022). The loss landscape of deep linear neural networks: A second-order analysis. arXiv:2107.13289.
2. Astrid, M., Zaheer, M. Z., Aouada, D., & Lee, S. (2024). Exploiting autoencoder's weakness to generate pseudo anomalies. Neural Computing and Applications, 1–17. (As discussed in Autoencoders for Anomaly Detection are Unreliable, arXiv:2501.13864.)
3. Baldi, P., & Hornik, K. (1989). Neural networks and principal component analysis: Learning from examples without local minima. Neural Networks, 2(1), 53–58.
4. Camacho, J., et al. (2019). Intrusion detection using PCA with a group-wise feature selection technique. (As discussed in An empirical model in intrusion detection systems using principal component analysis and deep learning models.)
5. Chalapathy, R., & Chawla, S. (2019). Deep learning for anomaly detection: A survey. arXiv:1901.03407.
6. Downs, J. J., & Vogel, E. F. (1993). A plant-wide industrial process control problem. Computers & Chemical Engineering, 17(3), 245–255. [Tennessee Eastman Process benchmark.]
7. Guezzaz, A., et al. (2022). An intrusion detection model using PCA and K-Nearest Neighbor classification. (As discussed in An empirical model in intrusion detection systems using principal component analysis and deep learning models.)
8. Kaloř, A. E., Michelsanti, D., Chiariotti, F., Tan, Z.-H., & Popovski, P. (2021). Remote anomaly detection in Industry 4.0 using resource-constrained devices. arXiv:2110.05757.
9. Kawaguchi, K. (2016). Deep learning without poor local minima. Advances in Neural Information Processing Systems (NeurIPS).
10. Laurent, T., & von Brecht, J. (2018). Deep linear networks with arbitrary loss: All local minima are global. Proceedings of the International Conference on Machine Learning (ICML).

11. Minimal-configuration anomaly detection for IIoT sensors. arXiv:2110.04049.
12. Predictive maintenance of machine tool linear axes: A case from manufacturing industry. (Maintenance decision-support framework applied to 29 machine tools.)
13. Quantum machine learning in cybersecurity: A case study on PCA-based intrusion detection systems. (2025). arXiv:2502.11173. [Includes the Principal Component Classifier of Shyu, M.-L., Chen, S.-C., Sarinnapakorn, K., & Chang, L. (2003).]
14. Rajadurai, H., & Gandhi, U. D. (2021). A hybrid intrusion detection system using PCA with deep learning techniques. (As discussed in An empirical model in intrusion detection systems using principal component analysis and deep learning models.)
15. Research on fault detection of Tennessee Eastman process based on PCA. (2013). IEEE Conference Publication.
16. Sensitive principal component analysis (SPCA) for chemical process fault detection and diagnosis. Industrial & Engineering Chemistry Research.
17. Shyu, M.-L., Chen, S.-C., Sarinnapakorn, K., & Chang, L. (2003). A novel anomaly detection scheme based on principal component classifier. Proceedings of the IEEE Foundations and New Directions of Data Mining Workshop.
18. Systematic review of anomaly and fault detection using machine learning for industrial machinery. (2026). Algorithms (MDPI), 19(2), 108.
19. Time series anomaly detection system with linear neural network and autoencoder. ResearchGate publication.
20. Unreliable reconstruction in linear and nonlinear autoencoders for anomaly detection. (2025). arXiv:2501.13864.
21. Zeng, A., Chen, M., Zhang, L., & Xu, Q. (2023). Are transformers effective for time series forecasting? Proceedings of the AAAI Conference on Artificial Intelligence, 37(9), 11121–11128.
22. Zhou, Y., & Liang, Y. (2018). Critical points of linear neural networks: Analytical forms and landscape properties. Proceedings of the Internationa