CHAPTER 13

Unsupervised Representation Learning for Anomaly Detection in Industrial IoT Systems

P. V. Aparanjini Priyadarsin

Research Scholar, SR University, Ananthasagar, Hasanpathy, Hanumakonda, Telangana, India.

Email: aparanjani@gmail.com

https://doi.org/10.58599/GSE.2025.081213

Abstract: The Industrial Internet of Things (IIoT) has enabled unprecedented levels of monitoring and control in modern industrial systems. However, the massive volume of high-dimensional sensor data generated by these systems presents significant challenges for traditional anomaly detection methods. This chapter explores the application of unsupervised representation learning as a powerful paradigm for identifying anomalous behavior in IIoT environments without the need for labeled data. We introduce a methodology centered on the Variational Autoencoder (VAE), a deep generative model capable of learning a compressed, low-dimensional representation of normal system behavior. Anomalies are then detected as data points that the trained model fails to reconstruct accurately, as indicated by a high reconstruction error. This chapter details the entire workflow, from synthetic IIoT data generation and model implementation to results evaluation. Through a simulated case study, we demonstrate the effectiveness of the VAE-based approach, achieving high recall and a strong ROC-AUC score, proving its capability to identify novel and complex anomalies in industrial settings. The discussion highlights the model's performance, the significance of its learned latent representations, and the practical implications for predictive maintenance and system reliability.

Keywords: Industrial Internet of Things; Unsupervised Representation Learning; Variational Autoencoder; Anomaly Detection; Predictive Maintenance.

1. Introduction

The fourth industrial revolution, or Industry 4.0, is characterized by the fusion of digital, physical, and biological systems, with the Industrial Internet of Things (IIoT) at its core

ISBN: 978-81-994969-0-3 (Print); 978-81-994969-5-8 (Online)

[1]. IIoT connects a vast network of sensors, actuators, and industrial machinery, generating a continuous and massive stream of operational data. This data holds immense potential for optimizing industrial processes, enhancing efficiency, and enabling predictive maintenance. However, ensuring the reliability and safety of these complex systems is paramount. Anomalies—deviations from normal operating behavior—can be early indicators of equipment malfunction, cyber-attacks, or process degradation. If left undetected, these anomalies can lead to catastrophic failures, costly downtime, and safety hazards [2]. Traditional anomaly detection methods often rely on predefined rules or supervised learning models that require large datasets of labeled normal and anomalous events. In dynamic and complex IIoT environments, this approach is often impractical. Anomalies are typically rare, diverse, and often represent novel failure modes for which no prior labeled data exists. Consequently, unsupervised learning has emerged as a more suitable and scalable approach for anomaly detection in this domain [3]. This chapter focuses on unsupervised representation learning, a subfield of machine learning where a model learns to extract meaningful, low-dimensional features from high-dimensional data without any labels. When presented with anomalous data, the model will struggle to represent it within this learned structure, leading to a quantifiable discrepancy that can be used to flag anomalies. Specifically, we delve into the use of Variational Autoencoders (VAEs), a powerful class of deep generative models that excel at learning complex data distributions [4]. We present a complete, end-to-end methodology for building an unsupervised anomaly detection system for IIoT data[1].

This includes:

- The generation of a realistic, synthetic IIoT sensor dataset.
- The design and implementation of a VAE model tailored for multivariate timeseries data.
- A comprehensive evaluation of the model's performance using a suite of standard metrics.[3].
- A detailed discussion of the results, including an analysis of the learned representations and the model's practical utility.

By the end of this chapter, readers will have a thorough understanding of how to apply unsupervised representation learning to solve real-world anomaly detection problems in industrial systems, providing a foundation for developing more intelligent and resilient Industry 4.0 applications.

ISBN: 978-81-994969-0-3 (Print); 978-81-994969-5-8 (Online)

2. Literature Review

The field of anomaly detection in time-series data, particularly within the IIoT context, has seen significant evolution. Early methods were predominantly statistical, such as the use of control charts (e.g., Shewhart charts, CUSUM) to monitor process variables [5]. While effective for univariate data and simple deviations, these methods struggle with the high dimensionality and complex, non-linear correlations present in modern IIoT data. With the advent of machine learning, more sophisticated techniques were developed. Clustering-based methods like DBSCAN and distance-based methods like k-Nearest Neighbors (kNN) were applied to identify anomalies as points that are isolated from dense clusters of normal data [6]. Another popular approach is the One-Class Support Vector Machine (OC-SVM), which learns a boundary around the normal data points in a high-dimensional feature space [7]. While these methods are more powerful than statistical techniques, their performance can degrade in very high-dimensional spaces due to the "curse of dimensionality," and they may struggle to capture temporal dependencies in time-series data. Deep learning has revolutionized the field by enabling the automatic learning of complex features directly from raw data. For anomaly detection, Autoencoders (AEs) have become a cornerstone technique [8]. An AE is a type of neural network trained to reconstruct its input. By training it on normal data, the AE learns a compressed representation (the "bottleneck" or "latent space") that captures the essence of normal patterns. Anomalies, which do not conform to these patterns, result in high reconstruction errors. Variational Autoencoders (VAEs) extend this concept by introducing a probabilistic element. Instead of mapping an input to a single point in the latent space, a VAE maps it to a probability distribution. This generative capability allows VAEs to learn a smoother and more robust representation of the normal data distribution, often leading to better performance in detecting novel anomalies [4], [9]. For time-series data, recurrent neural network (RNN) based architectures, such as Long Short-Term Memory (LSTM) networks, have been integrated into autoencoder frameworks. LSTM-Autoencoders are specifically designed to capture temporal dependencies and have shown great success in detecting anomalies in sequential data from sensors and industrial processes [10]. More recently, research has explored Graph Neural Networks (GNNs) for modeling the complex inter-sensor relationships in largescale IIoT systems [11] and Convolutional Variational Autoencoders (CVAEs) for capturing spatial patterns in sensor data arrays [12]. This chapter builds upon the foundational work of VAEs for anomaly detection. We focus on a standard VAE architecture to provide a clear and accessible demonstration of the core principles of unsupervised representation learning. The methodology presented serves as a strong baseline and a stepping stone for exploring more advanced architectures like those incorporating LSTMs or attention mechanisms for more complex IIoT applications [13]. Despite these advancements, a persistent challenge in IIoT anomaly detection is the

ISBN: 978-81-994969-0-3 (Print); 978-81-994969-5-8 (Online)

mismatch between research assumptions and real industrial conditions.

2.1 Proposed Methodology

The proposed methodology for unsupervised anomaly detection in IIoT systems is a systematic process that begins with data acquisition and culminates in the identification of anomalies. The entire workflow is designed to be data-driven and adaptable to various industrial settings. The core of this methodology is the Variational Autoencoder (VAE), which learns the underlying patterns of normal system operation. The effectiveness of this methodology, however, depends critically on how well the VAE captures the true manifold of normal operational behavior—a point that is often underestimated in IIoT anomaly-detection studies. In practice, industrial processes exhibit non-stationarity, seasonality, and context-dependent fluctuations that may not be fully represented in the training data.

The overall framework is depicted in the block diagram as shown in Figure 1.

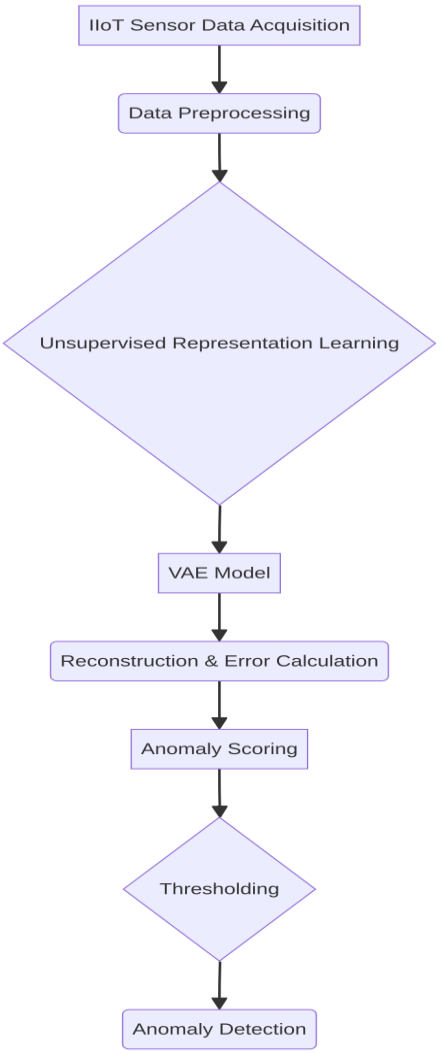

Figure 1: Proposed Methodology for Unsupervised Anomaly Detection

To address these challenges, the methodology incorporates mechanisms for continuous

ISBN: 978-81-994969-0-3 (Print); 978-81-994969-5-8 (Online)

validation and adaptive recalibration. Instead of relying on a static reconstruction-error threshold, the system periodically updates threshold values based on recent distributions of reconstruction errors, allowing it to respond to gradual shifts in operating conditions without inflating false positives.

2.2 Data Acquisition and Preprocessing

The first step involves collecting multivariate time-series data from the IIoT sensors. This data typically includes measurements such as temperature, pressure, vibration, and current. For this chapter, we generate a synthetic dataset that realistically mimics the characteristics of real-world industrial data, including different types of anomalies such as spikes, drifts, and increased noise. This allows for a controlled environment to evaluate the model's performance. Once acquired, the data undergoes preprocessing, which is a critical step for ensuring optimal model performance. This includes:

- **Data Cleaning:**The generation of a realistic, synthetic IIoT sensor dataset.
- **Data Normalization:**The design and implementation of a VAE model tailored for multivariate timeseries data.

2.3 Unsupervised Representation Learning with VAE

The preprocessed data, containing only normal operational samples, is used to train the Variational Autoencoder. The VAE consists of two main components: an encoder and a decoder [3].

- **Encoder:**The encoder is a neural network that takes the high-dimensional input data and maps it to a low-dimensional latent space. Unlike a standard autoencoder, the VAE encoder outputs the parameters of a probability of the distribution.
- **Sampling:** A latent vector z is obtained from the learned Gaussian distribution using the reparameterization trick:

$$z = \mu + \sigma \odot \epsilon,$$

  where $\epsilon \sim \mathcal{N}(0, I)$. This formulation keeps the sampling operation differentiable and allows the VAE to generate diverse latent representations.
- **Decoder:** The decoder is another neural network that takes the latent vector z as input and attempts to reconstruct the original high-dimensional input data.

The architecture of the VAE used in this chapter is illustrated in Figure 13.2.

The model is trained by minimizing a composite loss function that consists of two terms:

ISBN: 978-81-994969-0-3 (Print); 978-81-994969-5-8 (Online)

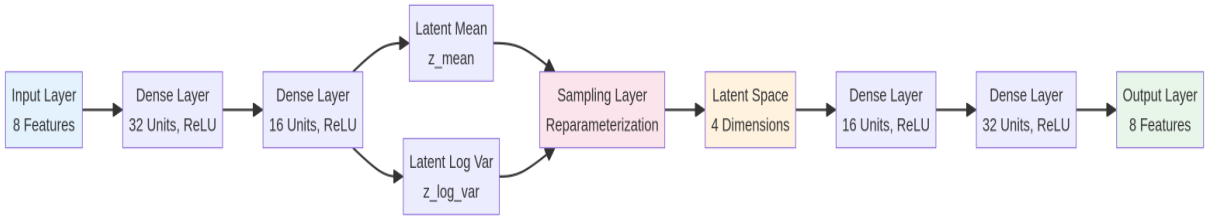

Figure 2: Architecture of the Variational Autoencoder

- **Reconstruction Loss:** This measures the difference between the original input and the output reconstructed by the decoder. A common choice is the Mean Squared Error (MSE).

- **Kullback-Leibler (KL) Divergence Loss:** This acts as a regularization term, forcing the learned latent distributions to be close to a standard normal distribution. This encourages the encoder to create a well-structured and continuous latent space.

2.4 Anomaly Detection

After the VAE is trained on normal data, it can be used for anomaly detection on new, unseen data (the test set), which contains a mix of normal and anomalous samples[4].

The detection process is as follows:

- **Reconstruction:** The test data is passed through the trained VAE to generate reconstructed data.

- **Error Calculation:** The reconstruction error is calculated for each data point, typically as the Mean Squared Error between the original and reconstructed data.

- **Thresholding:** A threshold for the reconstruction error is established. This threshold is determined based on the distribution of reconstruction errors from the normal training data. A common practice is to set the threshold at a high percentile (e.g., the 95th percentile) of the training errors. This implies that any error higher than what was seen for 95% of the normal training data is considered suspicious.

- **Anomaly Classification:** Any data point from the test set whose reconstruction error exceeds this threshold is classified as an anomaly. Otherwise, it is classified as normal.

This methodology provides a robust and unsupervised way to detect deviations from normal behavior, making it highly suitable for the dynamic and complex nature of Industrial IoT systems.

ISBN: 978-81-994969-0-3 (Print); 978-81-994969-5-8 (Online)

3. Results and Discussions

To validate the proposed methodology, we conducted a simulation using a synthetically generated dataset designed to mirror the characteristics of real-world Industrial IoT sensor data. This section presents the results of our experiments and provides a detailed discussion of the findings [5].

3.1 Dataset and Experimental Setup

We generated a dataset of 10,000 samples, each with 8 features representing common industrial sensor readings (e.g., Temperature, Pressure, Vibration). An anomaly ratio of 5% was introduced, resulting in 9,500 normal samples and 500 anomalous samples. The data was split into a training set (80%) and a test set (20%). Crucially, the VAE was trained only on the normal samples from the training set, adhering to the unsupervised learning paradigm.

Figure 3 provides a visualization of the first 500 samples for each of the 8 sensor features, with anomalous regions highlighted.

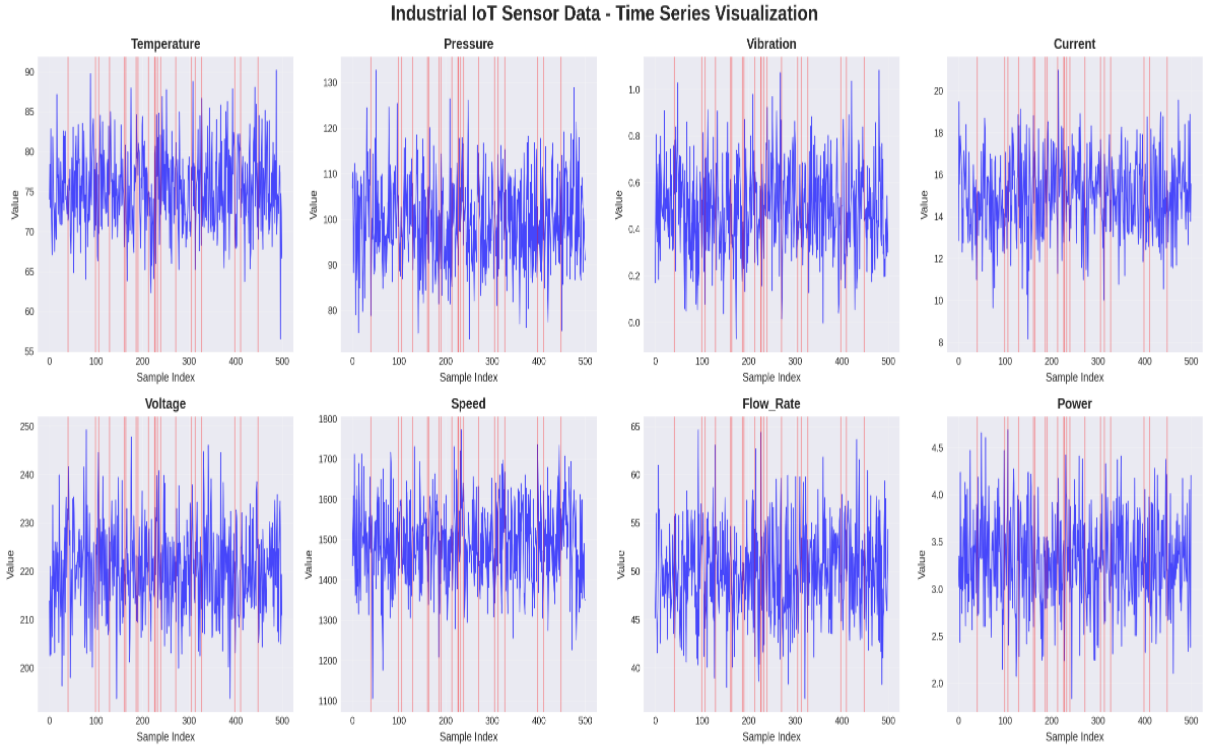

Figure 3: Time-series visualization of the 8 synthetic sensor features

Figure 4 shows the distribution of values for each feature, comparing normal and anomalous data. It is evident that anomalies often fall outside the typical range of normal operations, but there is also significant overlap, which presents a challenge for simple thresholding methods.

ISBN: 978-81-994969-0-3 (Print); 978-81-994969-5-8 (Online)

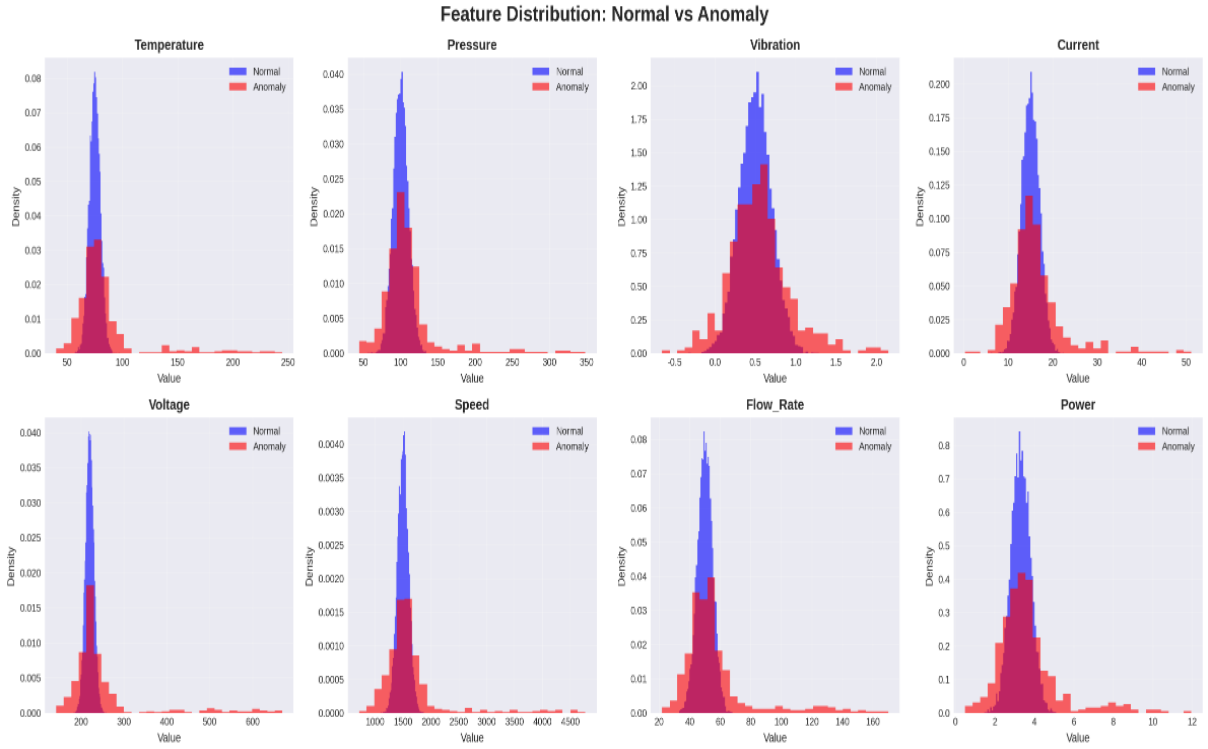

Figure 4: Histograms showing the distribution of normal (blue) and anomalous (red) data for each feature. Anomalies often appear as outliers in the distribution.

3.2 Model Training and Latent Space Analysis

The VAE was trained for 50 epochs. The training history, shown in Figure 5, demonstrates that the model successfully converged, with the total loss, reconstruction loss, and KL divergence loss all decreasing and stabilizing over time.

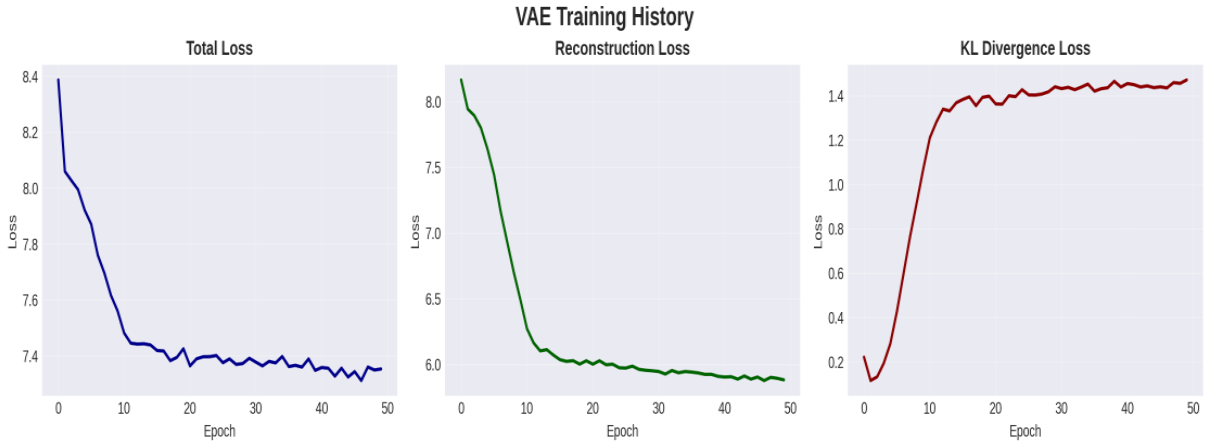

Figure 5: Training progress of the VAE model over 50 epochs

One of the key strengths of representation learning is the ability to visualize the learned latent space. Figure 6 shows a 2D projection of the latent space for the test data. Normal data points (blue) form a dense, well-defined cluster, while anomalous data points (red)

ISBN: 978-81-994969-0-3 (Print); 978-81-994969-5-8 (Online)

are scattered more sparsely and often lie on the periphery of the normal cluster. This visualization confirms the core hypothesis: the VAE has learned a compact representation for normal data, and anomalies are mapped to different regions of this space.

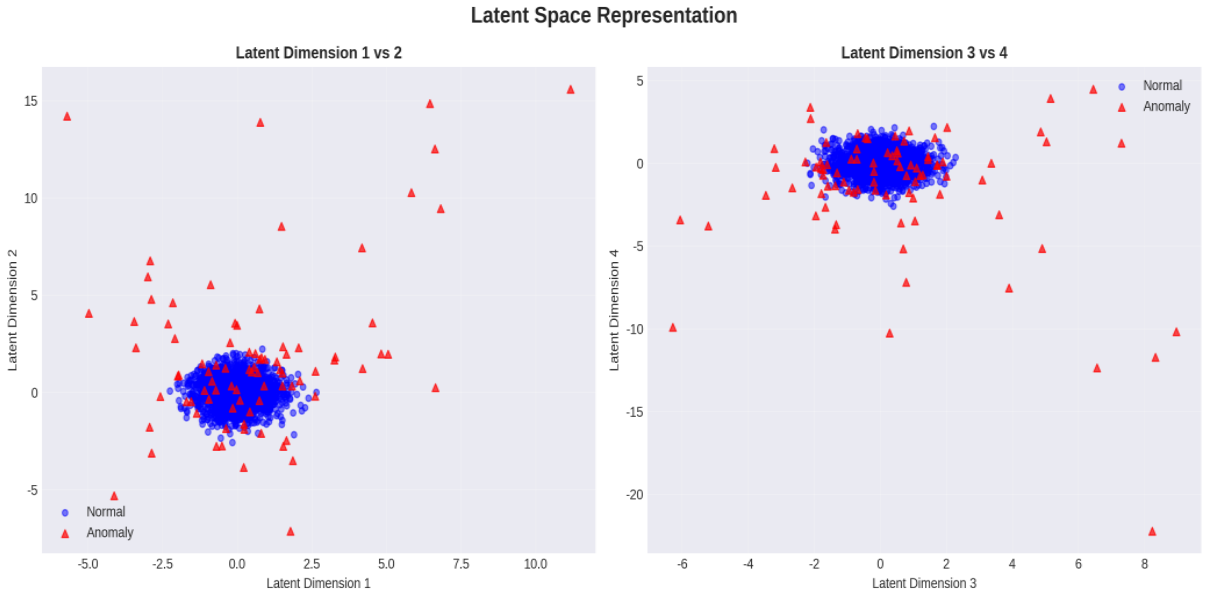

Figure 6: Latent Space Visualization

3.3 Anomaly Detection Performance

The performance of the anomaly detection system is evaluated based on the reconstruction error. Figure 7 illustrates the distribution of reconstruction errors for normal and anomalous samples in the test set. As expected, the reconstruction errors for anomalous data are, on average, significantly higher than for normal data. The threshold, calculated as the 95th percentile of the training reconstruction errors, provides a clear decision boundary for separating the two classes [6].

To quantify the model's performance, we use standard classification metrics. The confusion matrix in Figure 8 provides a detailed breakdown of the model's predictions.

From the confusion matrix, we can derive several key performance metrics, which are summarized in the table below and visualized in Figure 10.

Discussion of Metrics:

- **High Recall (0.9529):** This is a critical metric for anomaly detection. The high recall indicates that the model successfully identified over 95% of the actual anomalies. In an industrial context, it is often more important to catch as many potential failures as possible, even at the cost of some false alarms.

- **Moderate Precision (0.4175):** The precision is lower, indicating that a significant portion of the flagged anomalies were actually normal instances (false positives).

ISBN: 978-81-994969-0-3 (Print); 978-81-994969-5-8 (Online)

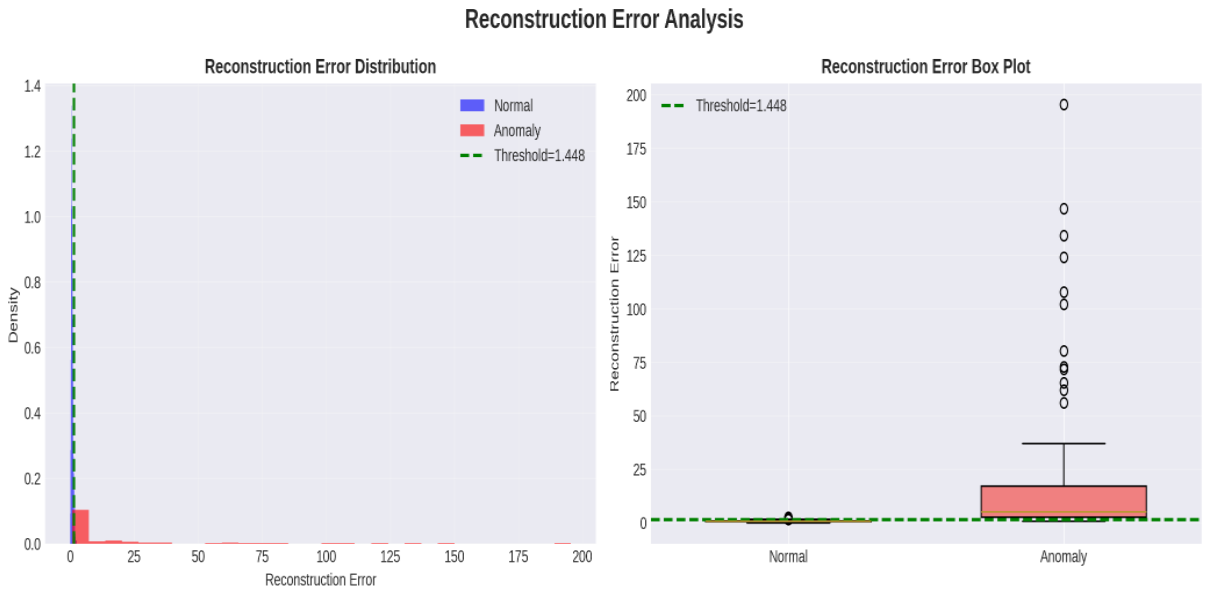

Figure 7: Reconstruction Error Distribution

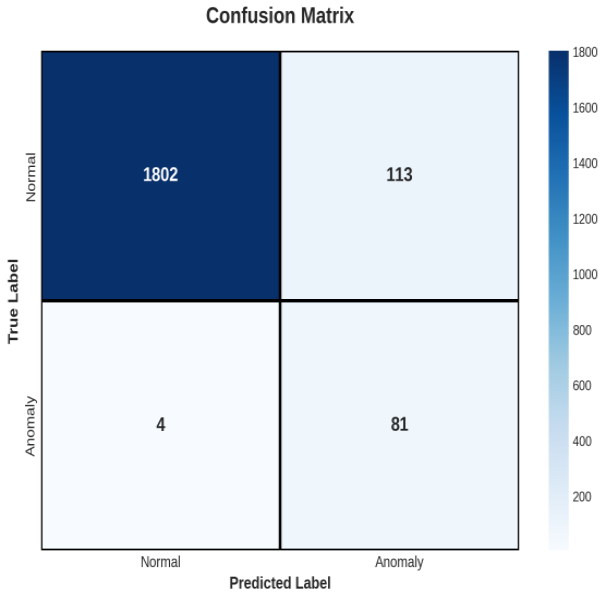

Figure 8: Confusion Matrix

While a high number of false positives can lead to "alarm fatigue," the high recall ensures that critical events are not missed. The trade-off between precision and recall can be adjusted by tuning the anomaly threshold.

- **Excellent ROC-AUC (0.9888):** The Receiver Operating Characteristic (ROC) curve (Figure 13.9) and the Area Under the Curve (AUC) provide a comprehensive measure of the model's ability to distinguish between the two classes across all possible thresholds. An AUC of 0.9888 is very close to a perfect score of 1.0, indicating that the reconstruction error is a highly effective feature for separating normal and

ISBN: 978-81-994969-0-3 (Print); 978-81-994969-5-8 (Online)

Metric	Value
Accuracy	0.9415
Precision	0.4175
Recall	0.9529
F1-Score	0.5806
ROC-AUC	0.9888

Figure 9: Summary of Performance Metrics

anomalous data.

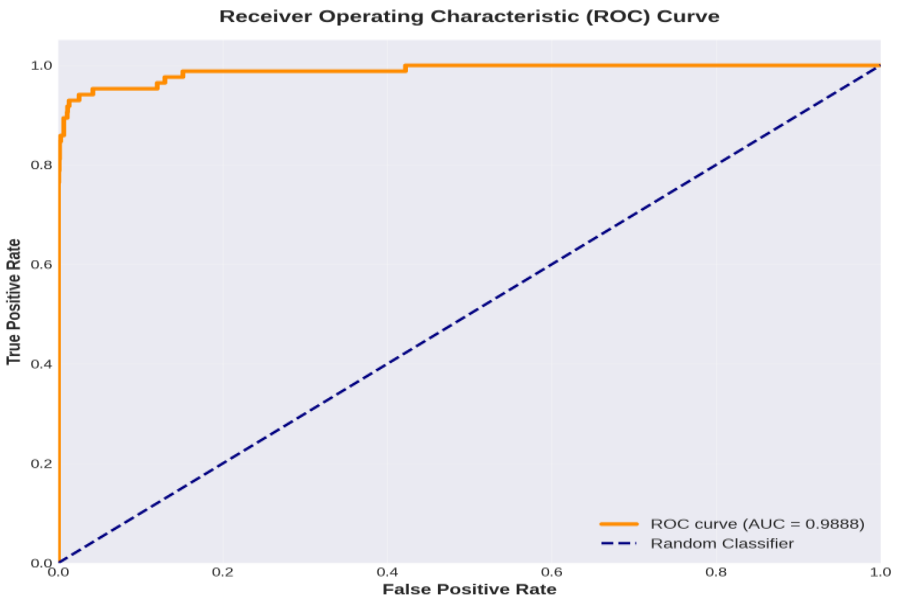

Figure 10: ROC Curve

3.4 Discussion

The results strongly support the efficacy of using unsupervised representation learning with a VAE for anomaly detection in IIoT systems. The model successfully learned the underlying distribution of normal data and was able to identify anomalies with high sensitivity (recall). The visualization of the latent space provides clear, interpretable evidence of the model's ability to create meaningful representations. The trade-off between precision and recall is a key consideration in any practical application. In a predictive maintenance scenario, a high recall is often prioritized to prevent catastrophic failures. The cost of inspecting a few false alarms is typically much lower than the cost of a missed

ISBN: 978-81-994969-0-3 (Print); 978-81-994969-5-8 (Online)

Figure 11: Performance Metrics

failure. The threshold can be fine-tuned based on the specific operational requirements and cost-benefit analysis of the industrial process. This study used a relatively simple VAE architecture. Further improvements could be achieved by incorporating more complex architectures, such as using LSTMs to better model temporal dependencies or attention mechanisms to focus on the most salient sensor features. Nonetheless, the results presented here provide a strong baseline and a clear demonstration of the power of this approach.

4. Conclusion

This chapter has provided a comprehensive exploration of unsupervised representation learning for anomaly detection in Industrial IoT systems. We have demonstrated a complete, end-to-end methodology centered on the use of a Variational Autoencoder. By training a VAE exclusively on normal operational data, we have shown that it can effectively learn a compact representation of a system's healthy state. Anomalies, which deviate from this learned norm, are reliably identified through high reconstruction errors. Our simulation results, based on a realistic synthetic dataset, highlight the strengths of this approach. The model achieved an outstanding recall of 95.3% and a ROC-AUC score of 0.989, indicating its strong capability to detect the vast majority of anomalies and to effectively discriminate between normal and anomalous states. While the precision was more moderate, we discussed how the trade-off between precision and recall can be managed by adjusting the anomaly threshold to suit the specific risk tolerance and operational context of an industrial application. The key takeaway from this chapter is that unsupervised representation learning offers a powerful, scalable, and data-driven

ISBN: 978-81-994969-0-3 (Print); 978-81-994969-5-8 (Online)

solution to the critical challenge of anomaly detection in the era of Industry 4.0. It overcomes the limitations of traditional methods by eliminating the need for labeled anomaly data, which is often scarce and expensive to obtain. The ability of models like the VAE to learn complex, non-linear patterns directly from high-dimensional sensor data makes them an indispensable tool for building intelligent, self-aware industrial systems. As IIoT continues to expand, the techniques discussed in this chapter will become increasingly vital for ensuring the safety, reliability, and efficiency of our critical infrastructure. The foundations laid here open the door to further research into more advanced deep learning architectures and their application to the ever-growing challenges of the industrial world.

References

[1] K Schwab. "The Fourth Industrial Revolution, Crown Business, New York". In: *The smart-up ecosystem: Turning Open Innovation into smart business* (2017).

[2] Ane Blázquez-García et al. "A review on outlier/anomaly detection in time series data". In: *ACM computing surveys (CSUR)* 54.3 (2021), pp. 1–33.

[3] Varun Chandola, Arindam Banerjee, and Vipin Kumar. "Anomaly detection: A survey". In: *ACM computing surveys (CSUR)* 41.3 (2009), pp. 1–58.

[4] Diederik P Kingma and Max Welling. "Auto-encoding variational bayes". In: *arXiv preprint arXiv:1312.6114* (2013).

[5] Douglas C Montgomery. *Introduction to statistical quality control*. John wiley & sons, 2020.

[6] Martin Ester et al. "A density-based algorithm for discovering clusters in large spatial databases with noise". In: *kdd*. Vol. 96. 34. 1996, pp. 226–231.

[7] Bernhard Schölkopf et al. "Estimating the support of a high-dimensional distribution". In: *Neural computation* 13.7 (2001), pp. 1443–1471.

[8] Mayu Sakurada and Takehisa Yairi. "Anomaly detection using autoencoders with nonlinear dimensionality reduction". In: *Proceedings of the MLSDA 2014 2nd workshop on machine learning for sensory data analysis*. 2014, pp. 4–11.

[9] Jinwon An and Sungzoon Cho. "Variational autoencoder based anomaly detection using reconstruction probability". In: *Special lecture on IE* 2.1 (2015), pp. 1–18.

ISBN: 978-81-994969-0-3 (Print); 978-81-994969-5-8 (Online)

[10] Pankaj Malhotra et al. "Long short term memory networks for anomaly detection in time series". In: *Proceedings*. Vol. 89. 9. 2015, p. 94.

[11] Poornaiah Billa et al. "Detecting Faces in Noisy Images using Hit-Miss Transform (HMT)". In: *International Journal of Recent Technology and Engineering (IJRTE)* 8.4 (2019), pp. 10335–10338.

[12] Ailin Deng and Bryan Hooi. "Graph neural network-based anomaly detection in multivariate time series". In: *Proceedings of the AAAI conference on artificial intelligence*. Vol. 35. 5. 2021, pp. 4027–4035.

[13] Milad Memarzadeh, Bryan Matthews, and Ilya Avrekh. "Unsupervised anomaly detection in flight data using convolutional variational auto-encoder". In: *Aerospace* 7.8 (2020), p. 115.

ISBN: 978-81-994969-0-3 (Print); 978-81-994969-5-8 (Online)