Applied Computer Science, vol. X, no. X, pp. X–X
doi: X

Submitted: YYYY-MM-DD | Revised: YYYY-MM-DD | Accepted: YYYY-MM-DD

Keywords: Industrial Manufacturing systems, Variational Autoencoder (VAE), Predictive Maintenance, interpretability, Developing countries

Richard NASSO TOUMBA [0009-0006-7596-1860]*, *Maxime MOAMISSOAL SAMUEL**, *Achille EBOKE**, *Wangkaké TAIWE***, *Timothée KOMBE* *

INTERPRETABLE VAE-BASED PREDICTIVE MODELING FOR ENHANCED COMPLEX INDUSTRIAL SYSTEMS DEPENDABILITY IN DEVELOPING COUNTRIES.

Abstract:

Rapid industrial growth in developing countries necessitates robust maintenance, with predictive maintenance (PdM) offering a vital solution to minimize downtime and costs. However, complex industrial systems and the acute scarcity of labeled data, particularly in African contexts, present significant implementation challenges for traditional PdM approaches. This research proposes a novel predictive maintenance approach utilizing a Variational Autoencoder (VAE), specifically designed to address data scarcity and enhance interpretability in complex industrial systems within developing countries. The VAE is trained on real-world operational data, learning intricate system patterns. Its interpretability is a core feature, achieved through visualization and analysis of the latent space, providing deeper insights into system behavior. The VAE model demonstrates strong and consistent performance in anomaly detection and data reconstruction, evidenced by low Mean Squared Error (MSE) and favorable R² values, and rigorously validated through cross-validation, which confirms its robustness and generalizability. This highlights its capacity to accurately model complex system dynamics across varied data subsets. This interpretable VAE model offers a powerful and promising predictive maintenance solution for enhancing the dependability of complex industrial systems in developing countries. By enabling early anomaly detection, synthetic data generation, and improved decision-making, this approach has the potential to significantly contribute to the growth and sustainability of industries in these regions through reduced downtime and optimized resource utilization.

---

[*] The university of Douala, Laboratory of technologies and Applied Sciences, Cameroon, richardnassotoumba4@gmail.com,moamissoalmaxime@gmail.com,ebokeachille@gmail.com, tkombe@yahoo.fr

[**] The university of Maroua , Energy and Environment Laboratory, Cameroon, wtaiwe@gmail.com

1. INTRODUCTION

1.1. Background and Motivation

Developing countries are experiencing a rapid industrialization process, with significant investments in manufacturing, mining, and energy sectors. This growth is crucial for economic development and social progress, offering pathways to improved livelihoods and sustainable growth. However, the sustained growth and competitiveness of these industries critically hinge on the reliability and efficiency of their industrial systems. Downtime due to equipment failures can lead to significant financial losses, not only through direct repair costs and lost production but also by disrupting supply chains, affecting market competitiveness, and potentially posing safety hazards to personnel. This unreliability can impede national economic targets and broader development goals.

Traditional maintenance strategies, often based on scheduled inspections and reactive repairs, are proving increasingly inadequate in addressing the complexities and dynamic nature of modern industrial systems (Geisbush & Ariaratnam, 2023; Nunes et al., 2023). These reactive approaches often result in unplanned downtime, inefficient resource allocation, and suboptimal operational performance.

Predictive Maintenance (PdM) offers a promising and advanced solution by leveraging data-driven techniques to anticipate potential failures and schedule maintenance proactively (Ma et al., 2024; Nunes et al., 2023). PdM systems utilize sensors to collect real-time data on equipment performance, which is then rigorously analyzed using sophisticated machine learning algorithms to identify anomalies and predict potential failures (Chen et al., 2023; Shahin et al., 2023). This proactive approach enables more efficient resource allocation, significantly reduces costly unplanned downtime, extends equipment lifespan, and fundamentally improves overall system reliability and operational efficiency. For developing countries, embracing such advanced methodologies is essential to maximize the return on industrial investments and foster sustainable growth.

1.2. Challenges in Applying PdM in Developing countries Contexts

Despite the transformative potential of PdM, its effective implementation in African contexts faces several distinct and formidable challenges that limit the applicability of conventional machine learning approaches.

Data Scarcity and Quality: Industrial data collection and management systems are often underdeveloped in Africa, leading to limited availability of high-quality, comprehensive, and labeled data for training robust machine learning models (Tapo et al., 2024; Mwanza, Telukdarie, & Igusa, 2023). This scarcity makes it difficult to train traditional supervised models effectively, which typically require large, diverse datasets.

Complex Industrial Systems**:** Many African industries operate intricate, often aging, systems with complex interdependencies, making it difficult to accurately model and predict their behavior using simpler statistical or linear models (Samuel, 2024; Schlüter et al., 2023). The underlying patterns are non-linear and high-dimensional, requiring advanced models capable of capturing these nuances.

Limited Expertise and Resources: The shortage of skilled personnel proficient in advanced data analytics, machine learning, and PdM technologies, coupled with financial constraints, can significantly hinder the adoption and successful implementation of sophisticated PdM solutions (Baroud, Yahaya, & Elzamly, 2024; Karippur, Balaramachandran, & John, 2024). This underscores the need for models that are not only effective but also interpretable and user-friendly for local teams.

Cybersecurity Concerns: The increasing reliance on digital technologies in industrial systems raises legitimate concerns about cybersecurity vulnerabilities and potential data breaches (Möller, 2023; Rahmanović et al., 2023). Ensuring the security and integrity of data within PdM systems is paramount to maintaining their reliability and trustworthiness.

1.3. Objectives and Contributions

This research directly addresses the aforementioned challenges of applying advanced PdM in African contexts by proposing a novel approach based on interpretable Variational Autoencoders (VAEs). VAEs are particularly well-suited for these environments due to their ability to learn complex data distributions, generate synthetic data, and perform anomaly detection even with limited labeled data, while also offering pathways to understanding their internal decision-making.

The primary aim of this study is to develop a robust, data-efficient, and interpretable VAE-based predictive maintenance model specifically tailored for African industrial settings, where traditional approaches often fall short. This involves effectively capturing the complex dynamics of industrial systems, even with limited or unlabeled data, thereby overcoming a significant hurdle in developing regions. Furthermore, the study aims to enhance the model's interpretability by analyzing its latent space representation, providing crucial insights into the system's underlying health status and facilitating informed, actionable decision-making by maintenance engineers, even those without deep AI expertise.

Through these efforts, this research seeks to rigorously evaluate the model's performance on real-world industrial data from an African context, specifically from a major industrial mill. By demonstrating its practical potential for significantly improving system dependability and reducing maintenance costs in challenging environments, this work offers a practical and transferable solution for enhancing the operational efficiency of complex industrial systems. Ultimately, this research contributes to the sustainable industrial growth and economic progress of African nations, with the interpretability feature empowering local teams and fostering greater adoption and trust in advanced maintenance technologies.

2. LITERATURE REVIEW

This section provides an overview of existing research on predictive maintenance (PdM) and the application of machine learning techniques, particularly Variational Autoencoders (VAEs), for enhancing industrial system dependability. We focus on recent advancements in the field, highlighting relevant studies from 2022, 2023, and 2024.

2.1. Predictive Maintenance in Industrial Systems

Predictive maintenance (PdM) has emerged as a key strategy for optimizing industrial operations and minimizing downtime (Abouelyazid, 2023; Dayo-Olupona, Genc, Celik, & Bada, 2023; Chen, Fu, Zheng, Tao, & Liu, 2023). Traditional maintenance approaches, based on scheduled inspections and reactive repairs, are often inefficient and costly (Dalhatu et al., 2023; Yazdi, 2024). PdM leverages data-driven techniques to predict potential failures and schedule maintenance proactively, leading to improved resource allocation, reduced downtime, and increased system reliability (Patil et al., 2023; Ucar et al., 2024; Meddaoui et al., 2023).

Machine learning (ML) algorithms have played a crucial role in the development of sophisticated PdM systems (Rosati et al., 2023; Ooko & Karume, 2024; Daoudi et al., 2023). These algorithms can analyze sensor data to identify patterns, detect anomalies, and predict future equipment behavior. Various ML techniques have been employed for PdM, including:

- Regression Models: Linear regression, support vector regression, and decision trees are commonly used to predict remaining useful life (RUL) or failure probability (Xu et al., 2020; Drakaki et al., 2022).
- Classification Models: Logistic regression, support vector machines, and random forests are used to classify equipment states as healthy or faulty (Yurek et al., 2022; Niyonambaza et al., 2020).
- Clustering Algorithms: K-means and hierarchical clustering can identify groups of similar equipment behavior, facilitating anomaly detection and condition monitoring (Carratù et al., 2023).

2.2. Deep Learning for Predictive Maintenance

Deep learning (DL) techniques, particularly deep neural networks (DNNs), have shown promising results in PdM applications due to their ability to handle complex data patterns and learn hierarchical representations (Wang et al., 2022; Khalil et al., 2021; Pandey et al., 2023). Several DL architectures have been explored for PdM, including:

- **Convolutional Neural Networks (CNNs)**: CNNs are effective in extracting spatial features from sensor data, making them suitable for applications involving image or time-series data (Moskolaï et al., 2021; Wang et al., 2021; Wang et al., 2023).

- **Recurrent Neural Networks (RNNs)**: RNNs are designed to handle sequential data, making them well-suited for analyzing time-series data from industrial systems (Weerakody et al., 2021; Fatima & Rahimi, 2024; Mienye et al., 2024).
- **Long Short-Term Memory (LSTM) Networks**: LSTMs are a type of RNN that can effectively capture long-term dependencies in time-series data, improving the accuracy of RUL predictions (Ma & Mao, 2020).

2.3. Variational Autoencoders for Predictive Maintenance

Variational Autoencoders (VAEs) are a powerful generative model that can learn a compressed representation of the input data, enabling efficient data reconstruction and anomaly detection (Oluwasanmi et al., 2021; Neloy & Turgeon, 2024; Ehrhardt & Wilms, 2022). VAEs have recently gained attention in the field of PdM due to their ability to:

- **Handle Data Scarcity**: VAEs can effectively learn from limited data by capturing the underlying data distribution and generating synthetic data for training other models (Akkem et al., 2024; Goyal & Mahmoud, 2024; Figueira & Vaz, 2022).
- **Improve Anomaly Detection**: VAEs can identify anomalies by reconstructing the input data and measuring the reconstruction error. Large reconstruction errors indicate potential deviations from normal behavior (Niu et al., 2020; Angiulli et al., 2020).
- **Enhance Interpretability**: The latent space representation learned by VAEs can provide insights into the system's behavior and facilitate the interpretation of anomaly detection results (Neloy & Turgeon, 2024; Costa & Sánchez, 2022).

3. MATERIALS AND METHODS

This section outlines the core concepts and develops the mathematical background for the reliability of complex industrial systems. To facilitate a clear understanding of the generative models central to this research, Fig.1 presents the basic architecture of a Variational Autoencoder (VAE)**,** which forms the foundation of our proposed predictive maintenance model.

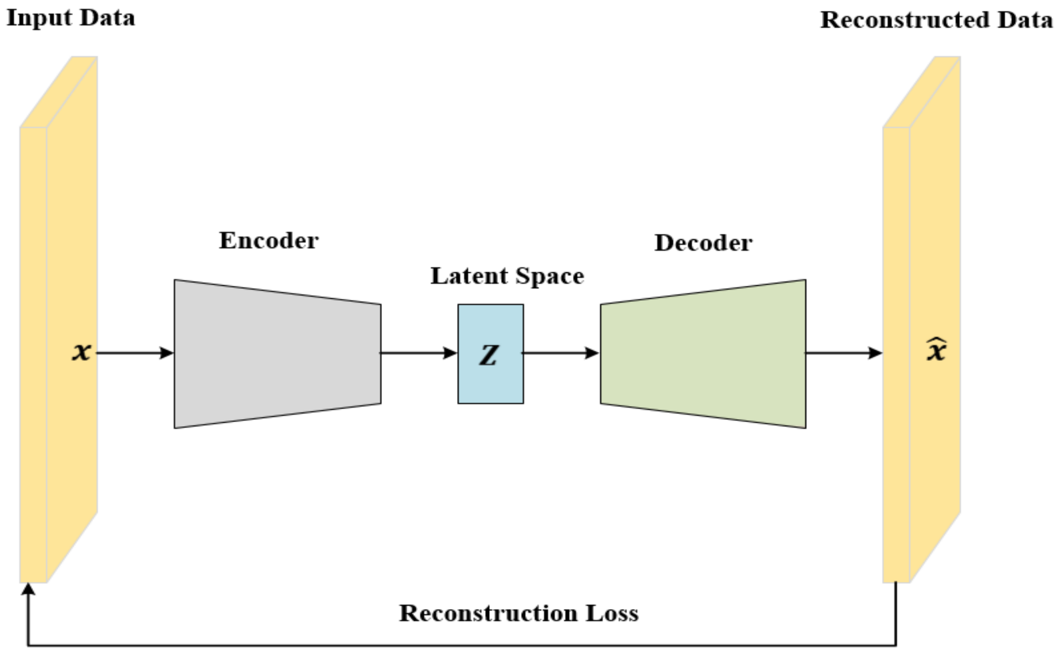


Fig.1. Basic architecture of a VAE

3. 1. Theoretical background

3.1.1 Variational Autoencoders.

Introduction

Variational Autoencoders (VAEs) represent a powerful class of generative models that combine the principles of deep learning with probabilistic inference. Unlike traditional autoencoders, which primarily focus on learning compact representations, VAEs introduce a probabilistic framework that enables them to generate new data samples and perform various tasks related to representation learning.

3.1.2 Core Concepts

The central concept behind VAEs is the use of latent variables, which are low-dimensional representations that capture the essential characteristics of the input data. These latent variables are assumed to be drawn from a prior distribution. The VAE architecture employs two neural networks:

1. **Encoder:** The encoder maps the input data to a distribution over the latent space. This distribution is parameterized by its mean and variance, which are the outputs of the encoder network.
2. **Decoder:** The decoder network takes a sample from the latent space and attempts to reconstruct the original input data. The decoder is trained to maximize the likelihood of the input data given the latent sample.

3.1.3. The Variational Objective

The training of a VAE revolves around optimizing a variational objective function, which consists of two terms:

1. **Reconstruction Loss:** This term measures the discrepancy between the reconstructed data and the original input. It encourages the VAE to learn meaningful representations that can be used to accurately reconstruct the input.
2. **KL Divergence:** This term quantifies the difference between the learned posterior distribution over the latent space and the prior distribution. It acts as a regularizer, encouraging the learned posterior to be close to the prior.

Notation

- D: the full data set ;
- X : represents the observed data or model input
- $X \sim D$ indicates that the observed data X is sampled from D
- Z : represents a latent representation or an abstract version of the input data X capturing its most important features
- $P_\theta(Z|X)$: The conditional probability of Z given X also called the posterior distribution. It encodes input data X into the latent space Z ; (inference).
- $P_\theta(X \mid Z)$: The conditional probability of X given Z also called the likelihood.
- $P_\theta(X)$: The marginal probability. It decodes or generate new data X from the latent representation Z (generation).

Problem Formulation

The core idea is to find a joint distribution:

$$P_\theta(z,x) = P_\theta(z).P_\theta(x|z) \tag{1}$$

where $P_\theta(z,\theta)$ is a multivariate unit Gaussian. The goal is to obtain the optimal parameters θ for the model such that $P_\theta(x) \approx P_\theta(x,\theta)$. This is achieved by maximizing the marginal likelihood $P_\theta(x,\theta)$.

It can be expanded by showing that the marginal likelihood can be expressed as an integral over the latent space z

$$\begin{aligned} P_\theta(x) &= \int_z P_\theta(z,x)dz \\ &= \int_z P_\theta(z).P_\theta(x|z)dz \end{aligned} \tag{2}$$

Equation (2) reflects the core idea of modeling the data x as generated from some underlying latent representation z. The joint distribution $P_\theta(z,x)$ is factorized into the product of a prior distribution over the latent space $P_\theta(z)$ and the likelihood $P_\theta(x|z)$, which describes how the observed data is generated from the latent variables.

However, directly computing this integral is often intractable due to the complexity of the latent space and the potentially high-dimensional nature of the data. To address this intractability, we resort to the Bayesian framework.

Bayes' theorem allows us to express the posterior distribution $P_\theta(z \mid x)$, which represents the probability of the latent representation given the observed data:

$$P_\theta(z \mid x) = \frac{P_\theta(z).P_\theta(x|z)}{P_\theta(x)} \tag{3}$$

However, both the marginal likelihood $P_\theta(x)$ and the posterior distribution $P_\theta(z|x)$ remain intractable.

To circumvent this challenge, we employ variational inference. This approach involves approximating the true posterior distribution $P_\theta(z|x)$ with a more manageable distribution $Q_\phi(z|x)$ parameterized by ϕ. The goal is to find the optimal parameters ϕ such that the approximation Q is as close as possible to the true posterior P.

Implying we should instead approximate the posterior through variational inference which is a process to approximate some target distribution P with an approximation Q parameterized by ϕ such that by optimizing ϕ the two distributions (P and Q) can be as close as possible.

This is mathematically expressed as:

$$Q_\phi(z|x) \approx P_\theta(z|x)$$

To measure the closeness between the two distributions, we utilize the Kullback-Leibler (KL) divergence, which serves as the objective function for optimization.

The KL divergence serves as a measure of dissimilarity between two probability distributions and is inherently non-negative.

Assuming that: $Q_\theta = Q_\phi(z|x)$ and $P_\theta = P_\theta(z|x)$

The KL divergence can be written as:

$$\begin{aligned}
D_{KL}(Q_\phi \| P_\theta) &= \int_Z Q_\phi.\log(\frac{Q_\phi}{P_\theta})dz \qquad (4)\\
&= E_{Q_\phi}\left[Log\left(\frac{Q_\phi}{P_\theta} \right) \right] \\
&= E_{Q_\phi}\left[LogQ_\phi \right] - E_{Q_\phi}\left[LogP_\theta \right] \\
&= E_{Q_\phi}\left[LogQ_\phi \right] - E_{Q_\phi}\left[Log\,\frac{P_\theta(z,x)}{P_\theta(x)} \right] \\
&= E_{Q_\phi}\left[LogQ_\phi \right] - E_{Q_\phi}\left[LogP_\theta(z,x) \right] + E_{Q_\phi}\left[LogP_\theta(x) \right]
\end{aligned}$$

$$= LogP_\theta(x) - E_{Q_\phi}\left[LogP_\theta(z,x) - LogQ_\phi\right]$$

$$LogP_\theta(x) = D_{KL}(Q_\phi \| P_\theta) + E_{Q_\phi}\left[\log(z,x) - \log Q_\phi\right] \tag{5}$$

Given that D_{KL} is non-negative,

$$LogP_\theta(x) \geq E_{Q_\phi}\left[\log(z,x) - \log Q_\phi\right] \tag{6}$$

The expression from equation (6):

$E_{Q_\phi}\left[\log(z,x) - \log Q_\phi\right]$ is referred to as the Evidence Lower Bound (ELBO).

Our objective is to maximize the ELBO. By doing so, we indirectly maximize the log-likelihood of the data, $logP_\theta(x)$, and minimize the KL divergence $D_{KL}(Q_\phi \| P_\theta)$. Maximizing the ELBO allows us to simultaneously optimize both the generative model and the inference model without needing to explicitly calculate $P_\theta(x)$.

To achieve this maximization, we employ stochastic gradient descent. We define the loss function $L(x)$ as the negative of the ELBO:

$$L(x) = -E_{Q_\phi}\left[Log\,\frac{P_\theta(z,x)}{Q_\phi(z|x)}\right] \tag{7}$$

$$\nabla_{\theta,\phi}L(x) = -\nabla_{\theta,\phi}\left(E_{Q_\phi}\left[Log\,\frac{P_\theta(z,x)}{Q_\phi(z|x)}\right]\right)$$

(8)

Taking the gradient of the loss function with respect to θ :

$$\nabla_\theta\left(E_{Q_\phi}(z|x)\left[LogP_\theta(z,x) - LogQ_\phi(z|x)\right]\right) = \nabla_\theta(\int_Z\left(Q_\phi(z|x)\left[LogP_\theta(z,x) - LogQ_\phi(z|x)\right]\right)dz$$

$$= \int_Z\left(Q_\phi(z|x)\nabla_\theta\left[LogP_\theta(z,x) - LogQ_\phi(z|x)\right]\right)dz$$

$$= E_{Q_\phi}(z|x)\nabla_\theta\left[LogP_\theta(z,x) - LogQ_\phi(z|x)\right] \tag{9}$$

This derivation demonstrates that when optimizing the ELBO with respect to θ, we can interchange the gradient and expectation operators, simplifying the computation and enabling the use of Monte Carlo estimation for efficient gradient updates.

$$E_{Q\phi}(Z \mid x)\nabla_\theta\left[LogP_\theta(z,x) - LogQ_\phi(Z \mid x)\right] \approx \frac{1}{L}\sum_{i=1}^{L}\nabla_\theta\left[LogP_\theta(z,x)\right] \tag{10}$$

Taking the gradient of the loss function with respect to ϕ :

$$\nabla_\phi\left(E_{Q_\phi}(z|x)\left[LogP_\theta(z,x) - LogQ_\phi(z|x)\right]\right) = \nabla_\phi(\int_Z\left(Q_\phi(z|x)\left[LogP_\theta(z,x) - LogQ_\phi(z|x)\right]dz\right) \tag{11}$$

$$= \int_Z \nabla_\phi\left[Q_\phi(z|x).ELBO\right]dz$$

$$= \int_Z Q_\phi(z|x).\nabla_\phi ELBO dz + \int_Z ELBO.\nabla_\phi Q_\phi(z|x) dz$$

$$= E_{Q_\phi}(z|x)\left[ELBO\right] + \int_Z ELBO.\nabla_\phi Q_\phi(z|x) dz \quad (12)$$

The challenge in computing the gradient of the loss function with respect to the variational parameters ϕ lies in the second term of equation (12), which involves an integral that is difficult to compute directly. The reparameterization trick addresses this by expressing the latent variable z as a deterministic function g of the input x, the variational parameters ϕ, and an auxiliary noise variable ε. The noise variable ε is sampled from a simple distribution $P(\varepsilon)$, typically a standard Gaussian. The function g is designed such that both ϕ and the input data x influence the output z deterministically, while the distribution of g itself remains constant throughout training. This separation allows us to propagate gradients through the deterministic part of the reparameterization, enabling efficient optimization of the variational parameters ϕ.

$$z = g(\phi, x, \varepsilon) \quad (13)$$

$$L(x) = -E_{P(\varepsilon)}\left[LogP_\theta(z,x) - LogQ_\phi(z|x)\right] \quad (14)$$

$$\nabla_{\theta,\phi} L(x) \approx \frac{1}{L}\sum_{i=1}^{L} \nabla_{\theta,\phi}\left[LogP_\theta(z,x) - LogQ_\phi(z|x)\right] \quad (15)$$

The reparameterization trick allows us to express the loss function and the ELBO as expectations over the noise variable ε. The ELBO is then further decomposed into the expected log-likelihood of the data given the latent representation and the KL divergence between the approximate posterior and the prior distribution. The gradient of the loss function with respect to the model parameters θ and ϕ can be efficiently estimated using Monte Carlo sampling.

The specific form of the variational distribution $Q_\phi(z|x,\theta)$ is assumed to be a Gaussian distribution with mean $g(x,\phi,\varepsilon)$ and variance σ^2. The KL divergence between this Gaussian distribution and the prior distribution $P_\theta(z)$, which is also a standard Gaussian, is then computed. The final expression for the KL divergence involves the logarithm of the variance σ^2, the trace of the covariance matrix Σ, and the squared Euclidean norm of the mean μ.

$$ELBO = E_{P(\varepsilon)}\left[LogP_\theta(z,x) - LogQ_\phi(z|x)\right] \quad (16)$$

$$= E_{P(\varepsilon)}\left[Log\left[P_\theta(x/z).P_\theta(z)\right] - LogQ_\phi(z|x)\right]$$

$$= E_{P(\varepsilon)}\left[LogP_\theta(x/z) + LogP_\theta(z) - LogQ_\phi(z|x)\right]$$

$$= E_{P(\varepsilon)}\left[LogP_\theta(x|z)\right] + E_{P(\varepsilon)}\left[Log\left(\frac{Q_\phi(z|x)}{P_\theta(z)}\right)\right]$$

The ELBO is approximated using Monte Carlo sampling, and the loss function over the entire dataset is defined as the average of the loss function computed on mini-batches. The number of data points in the dataset and the mini-batch are denoted by N and M, respectively.

$$\approx \frac{1}{L}\sum_{i=1}^{L}\nabla_{\theta}\left[LogP_{\theta}(z,x)\right]-D_{KL}(Q_{\phi}(z\,|\,x)\|P_{\theta}(z)) \tag{17}$$

In the next steps, we derive the KL divergence between two Gaussian distributions. We introduce the Gaussian probability distribution function and then apply the KL divergence formula.

$$Q_{\phi}(z|x)=g(\phi,x,\varepsilon)=\mathcal{N}(\mu,\sigma) \tag{18}$$

$$\mathcal{N}(z,\mu,\sigma)=\frac{1}{\sigma\sqrt{2\pi}}e^{\frac{1}{2}\left(\frac{z-\mu}{\sigma}\right)^{2}} \tag{19}$$

$$\begin{aligned}D_{KL}(Q_{\phi}(z|x)\|P_{\theta}(z)) &= \int_{Z}Q_{\phi}(z|x).Log\left(\frac{Q_{\phi}(z|x)}{P_{\theta}(z)}\right)dz\\ &= \int_{Z}(Q_{\phi}(z|x).Log\,\frac{1}{\sigma}e^{\frac{1}{2}\left(\frac{z-\mu}{\sigma}\right)^{2}}-z^{2})dz \qquad (20)\\ &= -\frac{1}{2}\int_{Z}(Q_{\phi}(z|x).\left[Log\sigma^{2}-z^{2}+\frac{1}{\sigma^{2}}(z-\mu)^{2}\right]dz\end{aligned}$$

$$=\frac{1}{2}\left[Log\sigma^{2}\int_{Z}Q_{\phi}(z|x)dz-\int_{Z}z^{2}Q_{\phi}(z|x)dz+\frac{1}{\sigma^{2}}\int_{Z}(z-\mu)^{2}Q_{\phi}(z|x)dz\right] \tag{21}$$

By taking into account the following identities:

$$\int_{Z}Q_{\phi}(z|x)=1\ ,\ \mu^{2}+\sigma^{2}=\int_{Z}z^{2}\mathcal{N}(z,\mu,\sigma)dz\,,\ \sigma^{2}=\int_{Z}(z-\mu)^{2}\mathcal{N}(z,\mu,\sigma)dz$$

Equation (21) becomes:

$$=-\frac{1}{2}\left[Log\sigma^{2}-\mu^{2}-\sigma^{2}+1\right] \tag{22}$$

And the ELBO becomes:

$$ELBO\approx\left[\frac{1}{L}\sum_{i=1}^{L}LogP_{\theta}(x,z)\right]+\frac{1}{2}\left[Log\sigma^{2}-\mu^{2}-\sigma^{2}+1\right] \tag{23}$$

Then we can compute the estimation of the $LogP_{\theta}(x,z)$

$$L(x)=-E_{P(\varepsilon)}LogP_{\theta}(x,z)\approx\left[\frac{1}{L}\sum_{i=1}^{L}LogP_{\theta}(x,z)\right] \tag{24}$$

Consequently the loss function over the entire data set is given by:

$$L(D)\approx\frac{N}{M}L(X^{M})-\frac{N}{ML}\sum_{j=1}^{L}\log P_{\theta}(x_{j}/z) \tag{25}$$

$L(X^M)$ being the loss function computed on a mini-batch X^M of size dataset D.

3.2. Proposed VAE-Based Predictive Maintenance Model

Fig.2 provides a schematic representation of this proposed VAE-based predictive maintenance model, illustrating the overall process and the interaction between its components.

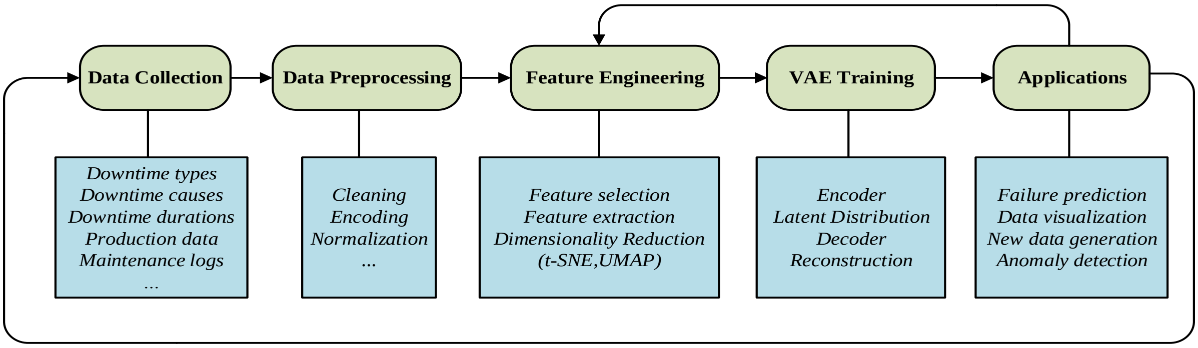


Fig.2. Schematic representation of the proposed VAE-based predictive maintenance model.

3.3. Framework Overview

The proposed framework, depicted in Fig.2, encompasses a series of interconnected stages designed to extract actionable insights from downtime data in industrial production systems.

- Data Collection**:** Gathers comprehensive downtime information, including types, causes, durations, and associated contextual factors, from diverse sources such as downtime logs, production data, maintenance records, and sensor readings.
- Data Preprocessing: Cleans and prepares the raw data through handling missing values, encoding categorical variables, normalizing numerical features, and removing outliers, ensuring data quality and consistency.
- Feature Engineering: Employs domain expertise and statistical methods to select, extract, and transform relevant features from the preprocessed data. Dimensionality reduction techniques may be applied to capture essential information while mitigating noise and complexity.
- VAE Training**:** Utilizes a Variational Autoencoder (VAE) to learn a compressed representation of the engineered features, capturing underlying structure and relationships within the data.
- Applications: Leverages the trained VAE for anomaly detection, failure prediction, data visualization, and generation of synthetic downtime scenarios, facilitating proactive maintenance and improved system understanding.

- Feedback Loops**:** Incorporates bi-directional feedback mechanisms between the "Applications" stage and the "Data Collection" & "Feature Engineering" stages. Insights gained from VAE applications inform and refine data collection strategies and feature engineering techniques, fostering continuous improvement.

This framework offers a systematic approach to harnessing the power of VAEs for downtime data analysis, enabling proactive maintenance, anomaly detection, and enhanced understanding of system behavior in industrial production environments.

The VAE architecture is designed to address the challenges of data scarcity and complexity in African industrial settings. By learning a compressed representation of the data, the model can effectively capture the intricate relationships within the system's operation, even with limited data availability.

3.4. Training Process

The VAE model is trained using a variational inference approach, which involves minimizing a loss function that balances two objectives:

Reconstruction Loss: This measures the difference between the reconstructed data and the original input data. The model aims to minimize this loss, ensuring accurate data reconstruction.

KL Divergence Loss: This measures the difference between the distribution of the latent representation and a standard normal distribution. Minimizing this loss encourages the latent space to have a well-defined and interpretable structure.

The training process involves iteratively feeding the model with sensor data and adjusting the model parameters to minimize the overall loss function. This iterative process allows the model to learn the complex dynamics of the industrial system and identify potential anomalies.

3.5. Model Functionalities

The proposed VAE-based model offers several key functionalities for enhancing predictive maintenance:

Anomaly Detection: The model can identify anomalies by comparing the reconstruction error of the input data with a predefined threshold. Large reconstruction errors indicate potential deviations from normal behavior, suggesting potential equipment failures.

Data Generation: The VAE can generate synthetic data that resembles the distribution of the real-world sensor data. This capability is particularly useful in scenarios where data is scarce, allowing for the augmentation of training datasets and improving model performance.

Interpretability: The latent space representation learned by the VAE provides insights into the system's behavior. By analyzing the latent variables, engineers and maintenance personnel can gain a deeper understanding of the system's dynamics and identify potential failure modes.

3.6. VAE Model Architecture and Training Details

The Variational Autoencoder (VAE) implemented in this study is specifically designed for time series analysis, explicitly modeling sequential dependencies. Its architecture utilizes 1D convolutional layers for both the encoder and decoder to process sequential windows of data. The model takes a window of 10 observations as input, where each time step contains 5 features. The latent space, a key hyperparameter, is generally set to a latent_dim of 16 for the robust evaluation, but was 8 for the initial evaluation.

The Encoder takes the input sequence and processes it through a series of 1D convolutional layers combined with pooling operations. These layers progressively extract features and reduce the sequence length. The processed output is then flattened and projected to the mean and log-variance parameters of the latent distribution through dedicated dense layers, utilizing linear activation.

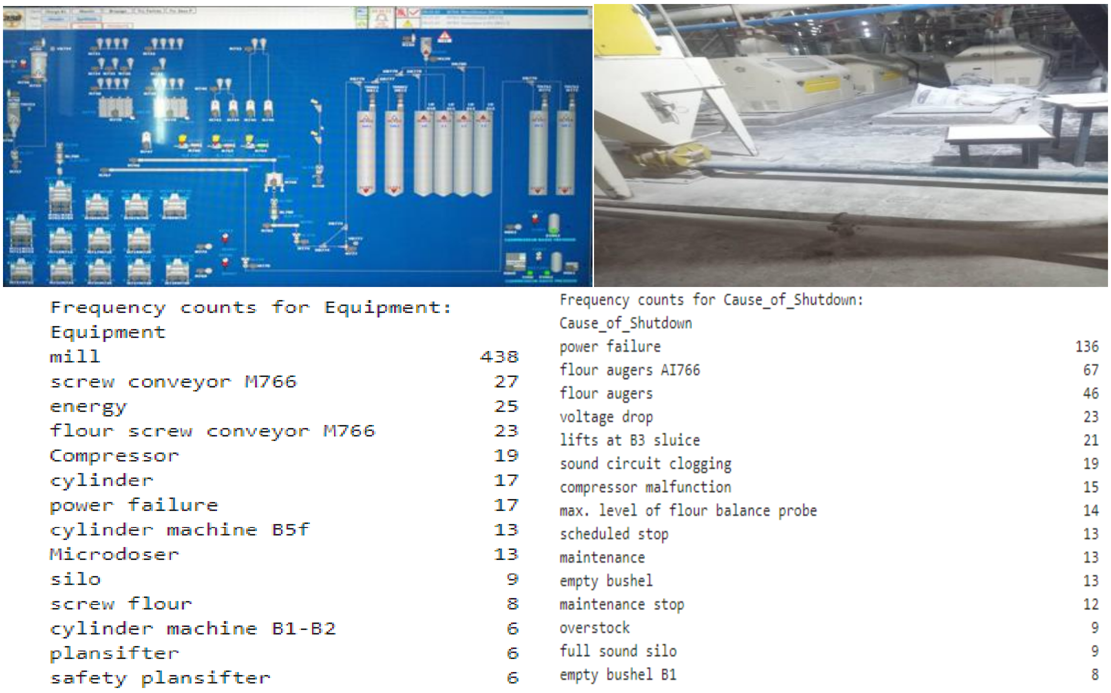
The Reparameterization Trick is employed to enable backpropagation through the sampling process. A latent vector is sampled from these parameters. This involves a random component from a standard normal distribution, which is scaled by the standard deviation derived from the log-variance, and then shifted by the mean. This sampled vector (with dimensions varying based on the specific evaluation phase) is subsequently passed to the decoder.

The Decoder takes the latent sample, expands it through an initial dense layer, and then reshapes it into a suitable sequence format. It utilizes 1D deconvolutional (Conv1DTranspose) layers to progressively reconstruct and upsample the data back to the original input sequence length. The final output layer is a Conv1D layer, reconstructing the original 5 features for each time step in the sequence with a linear activation function.

For training, the model employs the Adam optimizer (gradient descent-based optimization algorithm) with a fixed learning rate of 0.001 for the initial evaluation, 2e-05 for the robust evaluation, and a batch size of 64 for both evaluations. The VAE's total loss, based on the Evidence Lower Bound (ELBO), combines a Reconstruction Loss (Mean Squared Error) and a KL Divergence Loss (which regularizes the latent distribution against a standard normal prior), with a KL Divergence Weight of 1.0 for initial evaluation and 0.005 for the robust evaluation. The model is trained for 1000 epochs for initial evaluation and 2000 **epochs** for the robust evaluation. Additionally, Dropout with a rate of **0.2** is applied within the network to prevent overfitting.

3.7. System description

To understand the operational challenges and improvement opportunities within a typical industrial mill, consider the visual data presented in Fig.3. It offers a concise overview of of the real plant, its monitoring panel, the main equipments failure frequencies and the leading causes of production shutdowns.



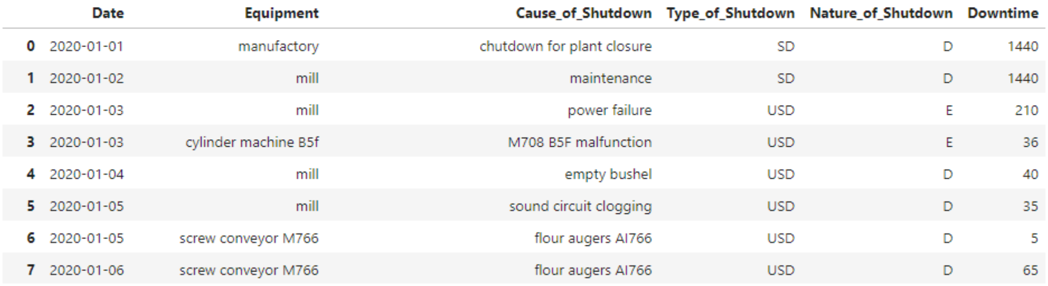
```
Frequency counts for Equipment:
Equipment
mill                                   438
screw conveyor M766                     27
energy                                  25
flour screw conveyor M766               23
Compressor                              19
cylinder                                17
power failure                           17
cylinder machine B5f                    13
Microdoser                              13
silo                                     9
screw flour                              8
cylinder machine B1-B2                   6
plansifter                               6
safety plansifter                        6
```

```
Frequency counts for Cause_of_Shutdown:
Cause_of_Shutdown
power failure                                                 136
flour augers AI766                                             67
flour augers                                                   46
voltage drop                                                   23
lifts at B3 sluice                                             21
sound circuit clogging                                         19
compressor malfunction                                         15
max. level of flour balance probe                              14
scheduled stop                                                 13
maintenance                                                    13
empty bushel                                                   13
maintenance stop                                               12
overstock                                                       9
full sound silo                                                 9
empty bushel B1                                                 8
```

Fig.3. The real industrial plant with key components failures frequencies

3.7.1. Dataset

The model was evaluated using a real-world dataset collected from an industrial mill LA PASTA situated in central Africa specifically in Douala Cameroun. The industrial system under study comprises 76 distinct pieces of equipment and compônents. The original dataset shows frequent occurrences across 750 recorded instances, spanning five critical dimensions: the implicated equipment, the nature of the shutdown, the underlying cause, and the resulting downtime. The Original data is shown in the fig.4.

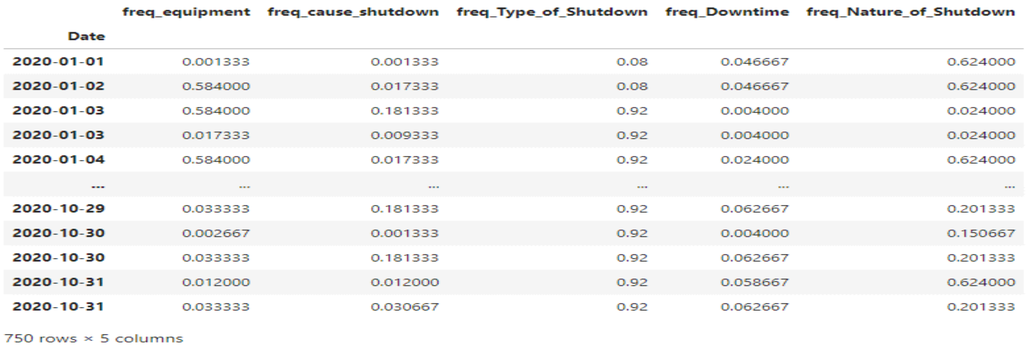


	Date	Equipment	Cause_of_Shutdown	Type_of_Shutdown	Nature_of_Shutdown	Downtime
0	2020-01-01	manufactory	chutdown for plant closure	SD	D	1440
1	2020-01-02	mill	maintenance	SD	D	1440
2	2020-01-03	mill	power failure	USD	E	210
3	2020-01-03	cylinder machine B5f	M708 B5F malfunction	USD	E	36
4	2020-01-04	mill	empty bushel	USD	D	40
5	2020-01-05	mill	sound circuit clogging	USD	D	35
6	2020-01-05	screw conveyor M766	flour augers AI766	USD	D	5
7	2020-01-06	screw conveyor M766	flour augers AI766	USD	D	65

Fig.4. Original dataset

Data Preprocessing:

Before training the model, the dataset was preprocessed to handle missing values, normalize the data, and prepare it for the VAE architecture. The DataFrame presents frequencies associated with various aspects of equipment operation and shutdowns, spanning from January 1st, 2020, to October 31st, 2020 (Fig.5). It comprises 750 rows (data points) and 5 columns, each representing a specific frequency metric.



Date	freq_equipment	freq_cause_shutdown	freq_Type_of_Shutdown	freq_Downtime	freq_Nature_of_Shutdown
2020-01-01	0.001333	0.001333	0.08	0.046667	0.624000
2020-01-02	0.584000	0.017333	0.08	0.046667	0.624000
2020-01-03	0.584000	0.181333	0.92	0.004000	0.024000
2020-01-03	0.017333	0.009333	0.92	0.004000	0.024000
2020-01-04	0.584000	0.017333	0.92	0.024000	0.624000
...	...	...	...	...	...
2020-10-29	0.033333	0.181333	0.92	0.062667	0.201333
2020-10-30	0.002667	0.001333	0.92	0.004000	0.150667
2020-10-30	0.033333	0.181333	0.92	0.062667	0.201333
2020-10-31	0.012000	0.012000	0.92	0.058667	0.624000
2020-10-31	0.033333	0.030667	0.92	0.062667	0.201333

750 rows × 5 columns

Fig.5. Encoded dataset

freq_equipment: The range of values signifies the relative prevalence of equipment-related events. Higher values indicate more frequent occurrences, which could guide maintenance prioritization or root cause investigations.

freq_cause_shutdown: This reveals the proportion of equipment events leading to shutdowns. Lower values compared to freq_equipment are still positive, showcasing system resilience, but areas with higher ratios might need attention to improve fault tolerance.

freq_Type_of_Shutdown: The dominance of one shutdown type (0.92) is crucial. This type likely represents the most common reason for shutdowns, making it a prime target for process optimization or preventive measures.

freq_Downtime: These values now indicate the proportion of time the mill experiences downtime. Even though they're relatively low, the economic impact of downtime in an industrial mill can be substantial, warranting further analysis to identify improvement opportunities.

freq_Nature_of_Shutdown: The most frequent values (0.624000) represents the predominant nature or category of shutdowns. Understanding this category's root causes could lead to targeted interventions to minimize their occurrence.

3.7.2. Evaluation Metrics

The following metrics were used to evaluate the model's performance:
Reconstruction Error: This metric measures the difference between the reconstructed data and the original input data. Lower reconstruction error indicates better model performance

in capturing the underlying data distribution. The specific reconstruction error metric used, is Mean Squared Error (MSE), R-squared (R2) has also been evaluated

3.7.3. Experimental Procedure

The following steps were taken to evaluate the model's performance:

Data Splitting:

The dataset was split into training, validation, and testing sets. The training set was used to train the VAE model, the validation set was used to tune the model's hyperparameters, and the testing set was used to evaluate the model's final performance. The dataset was split into 70% training, 10% validation, and 20% testing sets. Early stopping technics that monitor the model’s performance on a held-out portion of training data during training have been used.

Model Training:

The VAE model was trained on the training set using the variational inference approach described in Section 3.1. The training process was continued until the model converged, achieving a satisfactory level of performance on the validation set. However, Early stopping technic that monitor the model’s performance on a held-out portion of training data during training have been used.

Model Evaluation:

The trained model was evaluated on the testing set using the metrics described in Section 3.2. The model's performance was compared to other benchmark models in the litterature, such as traditional machine learning algorithms and other deep learning architectures.

4. RESULTS, INTERPRETATION AND VALIDATION

This section presents the results of the proposed interpretable VAE-based predictive maintenance model, focusing on its performance in terms of reconstruction error, anomaly detection, and data generation. We also discuss the model's interpretability by analyzing the latent space representation.

4.1. Initial Performance Evaluation (Regression aspect of the autoencoder)

While VAEs are primarily known for generative tasks, their ability to learn a compressed representation and reconstruct data makes them suitable for applications with a regression component. This justifies the use of regression metrics like Mean Squared Error (MSE) and R-squared (R^2) in evaluating VAE performance. Specifically, MSE is valuable for assessing the accuracy of data reconstruction, which is crucial in applications like dimensionality reduction and denoising. Furthermore, when VAEs are applied to time series prediction like in this study, MSE measures the accuracy of forecasting future values. Even in anomaly detection, where the primary goal isn't regression, MSE can quantify the reconstruction error, with higher values indicating potential anomalies. Therefore, considering the inherent reconstruction capabilities of VAEs and their applicability to tasks with regression elements, employing MSE and R^2 provides a comprehensive evaluation of VAE performance.

For this initial evaluation, the VAE model was trained for 1000 epochs. The model employs the Adam optimizer (gradient descent-based optimization algorithm) with a fixed learning rate of 0.001, and a batch size of 64. The VAE's total loss, based on the Evidence Lower Bound (ELBO), combines a Reconstruction Loss (Mean Squared Error) and a KL Divergence Loss (which regularizes the latent distribution against a standard normal prior), with a KL Divergence Weight of 1.0**.** Additionally, Dropout with a rate of 0.2 was applied within the network to prevent overfitting.

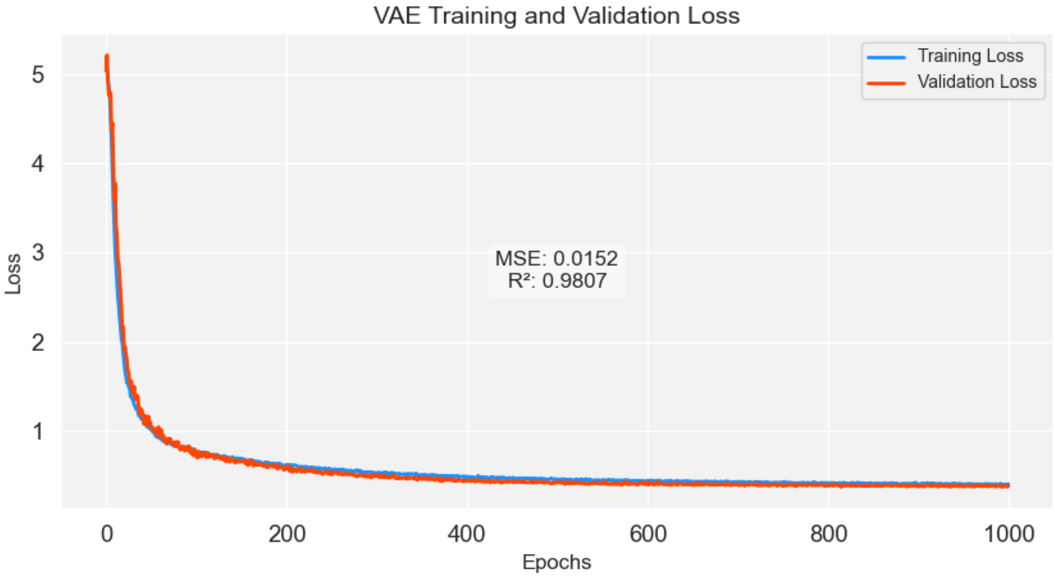


Fig.6. VAE Training Loss vs Validation Loss

The plot demonstrates the model's learning process, showing a rapid initial decrease in both training and validation losses, followed by a gradual convergence. This suggests the model's potential for both timely anomaly detection and long-term, reliable deployment.

Tab.1. Comparison of Model Performance with Literature

Model	**Dataset**	**MSE**	**R^2**	
Proposed VAE	industrial mill LA PASTA	0.0152	0.9807	
XGBRegressor	mechanical properties of rubberized concrete	0.33	0.976	[51]

Model	Dataset	MSE	R²	
CNN-LSTM	Tensile Strength of Friction Stir Welded AA7075-T651 Aluminum Alloy5 (Vibration sensor data from bearings under different operating conditions.)	0.002	0.976	[52]

Table 1 compares the performance of the proposed VAE model with other relevant models from the literature. It showcases different machine learning approaches, including ensemble methods like XGBRegressor and deep learning models such as CNN-LSTM.

It is important to note that the performance metrics presented for these baseline models are drawn from diverse studies in the existing literature. For example:

- The results for the **XGBRegressor** are sourced from SenthilVadivel et al. (2024), which focuses on predicting static mechanical properties of rubberized concrete using experimental data. This represents a distinct problem domain and dataset characteristic from continuous time-series data.
- The **CNN-LSTM** performance is cited from Song et al. (2023), which addresses a task related to vibration sensor data from bearings under different operating conditions for predicting tensile strength. While this involves time-series data, the specific dataset, problem formulation (e.g., direct strength prediction versus anomaly detection/RUL), and characteristics may still differ from our primary industrial system data.

Consequently, these external studies utilized datasets with varying characteristics, which often included differences in raw data dimensionality, specific feature engineering, and temporal horizons. Therefore, this table serves as an illustrative summary of the general landscape of methods and their typical performance ranges in different application contexts across the field, rather than a direct, controlled benchmark comparison on a single, unified dataset. This approach aligns with a qualitative benchmarking perspective, offering contextual understanding of the VAE's standing in the broader field. Furthermore, it is crucial to highlight that the VAE is a generative model, a fundamental distinction from the discriminative nature of many of these baseline models, enabling unique capabilities pertinent to industrial system monitoring such as learning the underlying data distribution for robust future anomaly detection.

4.2. Robustness Evaluation through Cross-Validation

This section delves into the rigorous evaluation of the Variational Autoencoder (VAE) model, specifically focusing on demonstrating its robustness and consistent performance across varied data subsets. By employing a comprehensive cross-validation strategy, we aim to establish the model's reliability and its ability to generalize effectively to unseen operational data.

4.2.1 Cross-Validation Methodology

To ensure a robust and generalized assessment of the Variational Autoencoder (VAE) model's performance, a K-Fold Cross-Validation strategy was implemented. This approach is particularly crucial for datasets of moderate size, such as the 750 instances utilized in this study, as it provides a comprehensive evaluation of model stability and mitigates the risk of reporting results influenced by a single, arbitrary data split.

The VAE architecture employed a sequential design, processing data through windows of 10 timesteps, each containing 5 features**.** The model projects these inputs into a latent space of 16 dimensions, balancing reconstruction quality with a KL divergence weight of 0.005 **.**Training was conducted with an Adam optimizer at a learning rate of 2e-05, using a batch size of 64 and applying a dropout rate to prevent overfitting.

The dataset was divided into 5 folds with the data randomly shuffled prior to splitting to ensure representative subsets. In each iteration, the VAE model was independently initialized and trained on the data from four folds, while the remaining fold served as the test set. This process was repeated five times, ensuring that every data instance contributed to both the training and evaluation phases. Each model instance was trained for 2000 epochs to facilitate convergence and thorough learning of the underlying data patterns. The training set for each fold comprised approximately 592-593 instances, providing sufficient data for model training.

4.2.2 Reconstruction Performance Analysis

The primary metric for evaluating the VAE's performance was the Mean Squared Error (MSE) of reconstruction. This metric quantifies the average squared difference between the input sequential data and its reconstruction by the VAE. A lower MSE indicates a higher fidelity in reconstruction, signifying the model's effectiveness in capturing and representing the "normal" patterns within the time-series data. This capability is fundamental for subsequent anomaly detection tasks, where significant deviations from this learned normal reconstruction error indicate anomalous behavior.

The cross-validation yielded the following aggregated reconstruction MSE results:

- Mean Reconstruction MSE: 0.0074
- Standard Deviation of Reconstruction MSE: 0.0002
- 95% Confidence Interval for Mean Reconstruction MSE: (0.0072, 0.0076)

The individual reconstruction MSE for each of the five folds was observed as: Fold 1: 0.0076, Fold 2: 0.0074, Fold 3: 0.0074, Fold 4: 0.0071, and Fold 5: 0.0074. These results, including the number of training instances per fold, are visually presented in Fig.7.

The results demonstrate the VAE model's remarkable stability and consistent performance across diverse subsets of the data. The exceptionally low standard deviation of 0.0002 highlights minimal variance in reconstruction performance across different data partitions, affirming the model's robustness and generalization capabilities. Furthermore, the very narrow 95% confidence interval of (0.0072, 0.0076) provides a precise statistical estimate of the expected mean reconstruction MSE, reinforcing confidence in the model's predictive performance on unseen, similar industrial system.

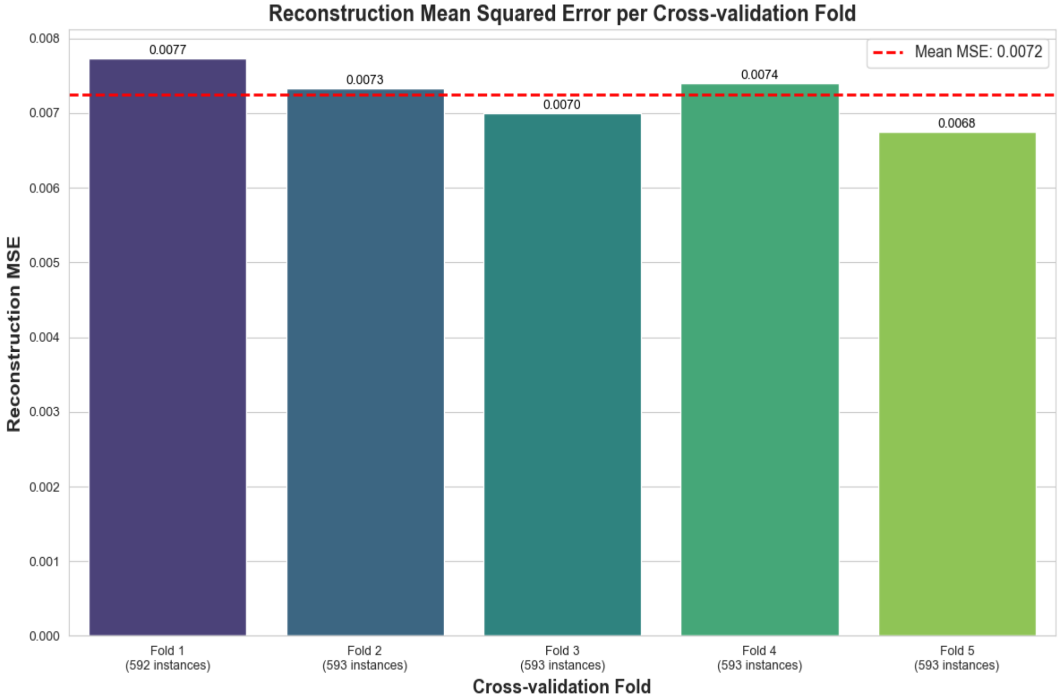


Fig.7. Reconstruction Mean Squared Error per Cross-validation Fold.

4.3. Anomaly Detection and Data Generation

The VAE's ability to reconstruct the input data with low error allows for effective anomaly detection. Large reconstruction errors indicate potential deviations from normal behavior, signaling potential equipment failures.

4.3.1. Unveiling System Dynamics: A Temporal and Predictive Analysis of Downtime Frequency

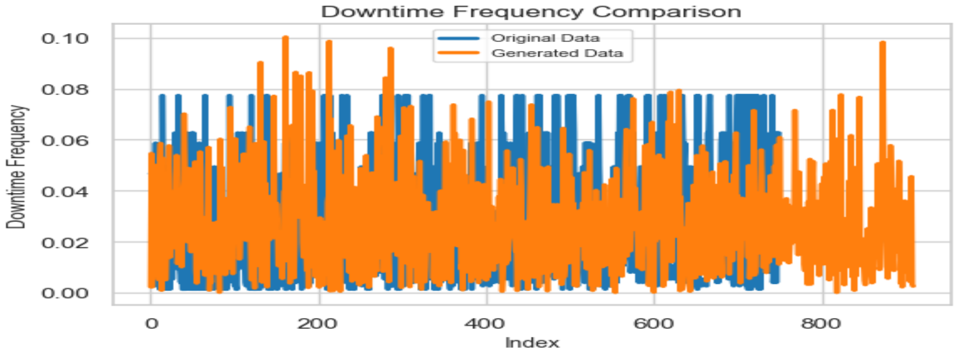


Fig.8. Downtime Frequency Comparison

The "Downtime Frequency Comparison" graph (fig.8) serves as a powerful tool for understanding both historical and potential future downtime patterns in an industrial system. By charting the evolution of downtime frequency and comparing actual data to a predictive model's output, it provides actionable insights for enhancing system reliability and performance.

The temporal analysis reveals the inherent variability of the system, with both datasets showcasing significant fluctuations in downtime occurrences. This dynamic behavior underscores the system's sensitivity to a spectrum of influencing factors, from external conditions to component wear. While the generated data generally aligns with the original, discrepancies exist, emphasizing the need for continuous model refinement to enhance predictive accuracy. Peaks observed in both datasets pinpoint periods of heightened vulnerability, prompting focused investigations into their root causes and enabling proactive preventive measures.

From a predictive standpoint, the model's capacity to mimic original data trends, despite its imperfections, underscores its potential for forecasting future downtime patterns. This predictive capability is pivotal for implementing proactive maintenance strategies, allowing organizations to anticipate and address periods of elevated downtime risk before costly disruptions occur. Additionally, insights gained from the model's analysis of downtime drivers can inform operational optimization efforts, contributing to reduced downtime and improved system availability.

Delving deeper, the seemingly random nature of the fluctuations in downtime frequency hints at the stochastic nature of the industrial system, where unpredictable events and factors can significantly impact its operation. This underscores the inherent complexity of such systems, where countless components and processes interact, often with cascading effects that defy precise prediction.

The occasional peaks in both datasets potentially represent critical downtime points or thresholds, signifying increased system vulnerability. Identifying these critical points empowers maintenance teams to proactively implement preventive measures, such as predictive maintenance or adjustments to operating parameters, thus minimizing downtime and its associated costs.

The model's ability to closely replicate the statistical characteristics of the original data speaks to its effectiveness in capturing the system's essential dynamics. This opens doors to various applications, including simulations, scenario testing, and data augmentation, all of which can contribute to a deeper understanding of the system's behavior.

However, even subtle discrepancies between the original and generated data hold valuable information. These differences may point to anomalies or unexpected behaviors not fully captured by the model. Scrutinizing these nuances can lead to new insights, potentially uncovering hidden system vulnerabilities or opportunities for optimization.

In conclusion, this graph transcends its visual simplicity, offering a profound window into the complexities and dynamics of an industrial system. By meticulously analyzing its patterns and variations, we can extract invaluable insights into system behavior, identify critical vulnerabilities, and optimize both maintenance and operational strategies. The resulting improvements in system reliability, efficiency, and overall performance ultimately translate into substantial cost savings and enhanced productivity.

4.3.2. Analysis of Shutdown Cause Frequency Data: Implications for a Complex Industrial System.

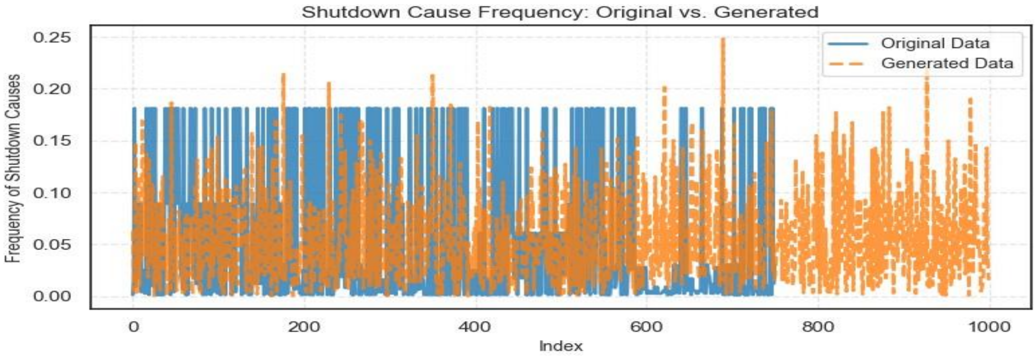


Fig.9. Shutdown Cause Frequency Comparison

The line chart (Fig.9.) compares the frequency of shutdown causes between original and generated data of the industrial mill, offering insights into the operational dynamics and potential failure modes of a complex industrial system. The original data demonstrates a relatively stable pattern, suggesting well-established operational procedures and maintenance practices. Occasional spikes, however, indicate underlying vulnerabilities or recurring issues that warrant further investigation.

In contrast, the generated data, intended to predict future failures, exhibits a wider range of frequencies, suggesting potential for increased instability or unforeseen events. This variability could intentionally represent a broader spectrum of scenarios for risk assessment and proactive maintenance.

While the generated data broadly aligns with the overall range of observed frequencies, its increased volatility underscores the potential for rare but high-impact events. This discrepancy highlights opportunities for model refinement and emphasizes the importance of considering both frequent and infrequent shutdown causes in risk management strategies. In the context of a complex industrial system, understanding and predicting shutdown cause frequencies is critical for optimizing operations, minimizing downtime, and ensuring system reliability. The presented analysis suggests that the generative model, despite its limitations, holds promise as a tool for proactive maintenance and risk mitigation. By further refining the model and integrating it into decision-support systems, the industrial system can enhance its resilience and achieve greater operational efficiency.

4.3.3 Analysis of Equipment-Related Event Frequencies in a Complex Industrial System

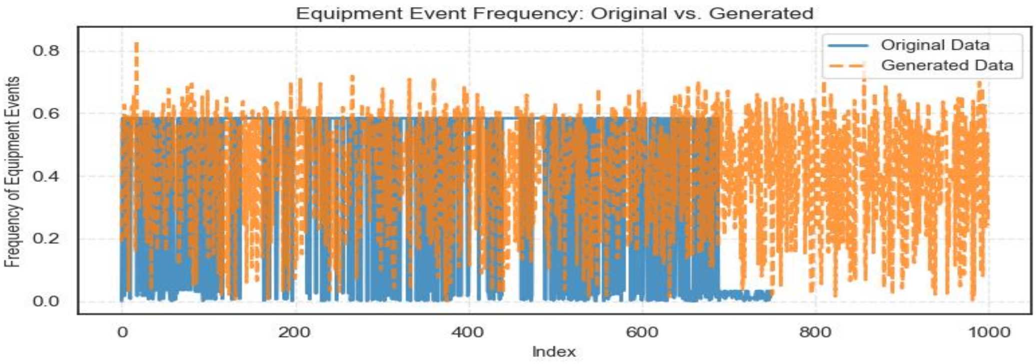


Fig.10. Equipment Event Frequency: Original vs. Generated

This section analyzes the graph comparing the frequency of equipment-related events in a complex industrial system illustrated by Fig.10, using both original and generated data. The original data displays relative stability, mostly oscillating around 0.05, suggesting a well-controlled system. However, occasional spikes indicate periods of disturbance or problems requiring further investigation. Most frequencies fall below 0.2, implying that equipment events are relatively rare during normal operation.

In contrast, the generated data exhibits a much wider range of frequencies (0 to 0.8), indicating the model's exploration of various scenarios, including low-probability, high-impact events. The generated frequencies are more uniformly distributed, suggesting consideration of various factors and their complex interactions. This wider range could help identify potential failure scenarios or situations with increased equipment-related events.

The difference between the original and generated data highlights the system's complexity. The generative model captures a broader range of potential behaviors, crucial for anticipating unusual situations or failures. This data can be used for predictive maintenance by identifying critical frequency thresholds, enabling alerts and preventive actions. Additionally, it allows for risk assessment by exploring high-frequency scenarios to assess system resilience and identify weaknesses. Finally, the model can simulate the impact of different operational strategies or system modifications on event frequency, optimizing overall performance and reliability.

4.4. Interpretability and Explainability: Unveiling the VAE's Inner Workings and System Dynamics

The ability of t-SNE and UMAP to preserve non-linear relationships in the projections provides a window into how the Variational Autoencoder (VAE) has learned to model the complex interactions inherent in the industrial system. This visualization offers valuable insights into both the model's behavior and its understanding of the underlying data. The distribution of variables within the projections pinpoints those that have the most significant impact on cluster formation and data separation. For instance, the central position of freq_equipment in multiple clusters suggests its pivotal role in the VAE's understanding of system behavior. The quality of clusters in the projections, particularly the clear separation observed in t-SNE, indicates that the VAE has learned robust representations capable of handling unseen scenarios, suggesting good generalization capability. Furthermore, the consistent and accurate representation of different data groups across both projections suggests that the VAE has learned a fair and unbiased representation, minimizing the risk of discriminatory or misleading predictions in industrial applications. The subsequent detailed

analysis in Section 4.5, accompanied by Fig. 11, further elaborates on how these abstract projections are interpreted to reveal deep insights into the VAE's learned representations.

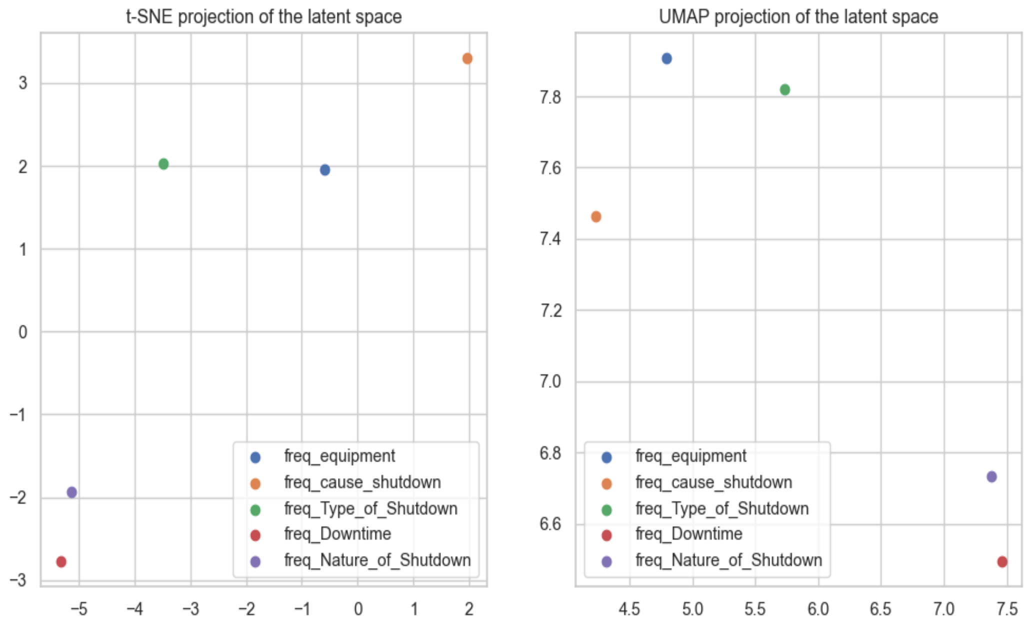


Fig.11. Latent Space Visualization with t-SNE and UMAP

4.5 Deeper Insights into the VAE Model's Behavior: Beyond Descriptive Analysis of the industrial system.

Fig. 11 presents two visualizations of the latent space using t-SNE and UMAP. These techniques are powerful non-linear dimensionality reduction algorithms that allow for the projection of high-dimensional data (in this case, the VAE's learned latent representations) into a lower, more interpretable two-dimensional space. This visualization facilitates the interpretation of complex relationships between variables as understood by the VAE.

It is important to clarify the interpretation of these plots, particularly regarding the axes. The numerical scales on the axes of t-SNE and UMAP visualizations are abstract and do not represent specific physical units or directly correspond to original input features. Furthermore, the specific ranges and numerical values on the t-SNE axes will inherently differ from those on the UMAP axes. This difference arises because each algorithm employs a distinct mathematical approach to construct its low-dimensional embedding:

- **t-SNE** focuses on preserving local neighborhoods, using a probability distribution to map distances, and its scale is influenced by a 'perplexity' parameter which can stretch or compress the final output. This often results in plots where distinct clusters are well-separated but the absolute distances between clusters can be less meaningful.
- **UMAP** aims to preserve both local and global structure by constructing a fuzzy simplicial complex. Its optimization process also yields an arbitrary scale. While UMAP tends to preserve global structure better than t-SNE, its axes, like t-SNE's, are scaled in a way that is unique to its embedding process and not directly comparable to other projections or real-world units.

Therefore, the exact numerical values on the axes for either plot, or the difference in these values between the two plots, hold no direct interpretability.

The interpretability of these visualizations stems primarily from the relative positions of the data points and the formation of clusters within this two-dimensional projection, not from the absolute values on the axes. Points that are close to each other in this projected space are considered highly similar by the VAE in its high-dimensional understanding of the industrial system. Conversely, points that are far apart represent distinct or dissimilar characteristics. The color-coding of points by specific frequency-encoded variables (e.g., freq_equipment, freq_cause_shutdown) is crucial for this interpretation, allowing for immediate visual identification of which variables contribute to particular clusters or occupy specific regions of the latent space.

By analyzing these clusters and patterns within the latent space, engineers and maintenance personnel can gain a deeper understanding of the system's dynamics and identify potential failure modes. Technically, T-SNE and UMAP projections reveal how VAEs understand complex industrial data, especially when frequency encoded. Distinct clustering patterns (such as the tight coupling of freq_equipment and freq_cause_shutdown in t-SNE) highlight the VAE's ability to prioritize key features for distinguishing system states. Furthermore, contrasting distributions in t-SNE (characterized by clear, often discrete, clusters) and UMAP (showing smoother transitions) suggest a balance within the VAE's learned representation between discrete categorization and the capture of subtle variations. Well-defined clusters, particularly evident in t-SNE, indicate robust representations and suggest good generalization capability for unseen data. The consistent and accurate representation of different data groups across both projections further suggests that the VAE has learned a fair and unbiased representation, minimizing the risk of discriminatory or misleading predictions in industrial applications. These insights empower targeted maintenance (e.g., prioritizing freq_Type_of_Shutdown based on potential impact), anomaly detection, and root cause analysis.

4.6 Discussion

The results demonstrate the effectiveness of the proposed interpretable VAE-based predictive maintenance model in capturing the complex dynamics of industrial systems with limited data. The model's ability to reconstruct the input data with low error, generate synthetic data, and provide insights into the latent space representation highlights its potential for enhancing the dependability of complex industrial systems in the context of developping countries..

The model's interpretability is a key advantage, enabling engineers and maintenance personnel to understand the system's behavior and make informed decisions. This approach contributes significantly to the growth and sustainability of developping countries industries by reducing downtime, optimizing resource utilization, and promoting a culture of proactive maintenance.

However, it is important to note that the model's performance is dependent on the quality and quantity of the available data. Further research is needed to investigate the model's generalizability to other type of industrial systems and datasets.

The proposed VAE-based model offers a promising solution for enhancing the dependability of complex industrial systems in developping countries. The model's interpretability, coupled with its ability to handle data scarcity and complexity, makes it a valuable tool for predictive maintenance and the optimization of industrial operations.

5. Conclusion and Future Perspectives

This research proposed a novel predictive maintenance approach leveraging a Variational Autoencoder (VAE) specifically designed to enhance complex industrial system dependability, particularly addressing challenges posed by data scarcity in developing countries. The developed VAE model, with its carefully tuned architecture and optimized parameters, demonstrates a robust capability to learn intricate normal operational patterns from real-world time-series data.

The comprehensive K-fold cross-validation study unequivocally validated the model's high stability and generalization performance. This rigorous evaluation provides robust confirmation of the VAE's effectiveness, reinforcing the promising capabilities observed in initial assessments and definitively establishing its reliability across diverse data subsets. The consistently low Mean Reconstruction MSE and its exceptionally minimal standard deviation across all folds signify that the model's performance is remarkably consistent. This strong evidence of robustness directly addresses concerns about overfitting and variability, establishing the VAE as a reliable tool for accurately characterizing normal system behavior and, by extension, identifying deviations indicative of potential faults. The low and stable reconstruction error foundational to this approach positions it as a highly effective method for anomaly detection in continuous industrial monitoring.

Building upon the robust foundation established in this work, several promising avenues for future research emerge:

1. Enhanced Interpretability of Latent Space**:** Further efforts will be directed towards deepening the interpretability of the VAE's latent space. This could involve developing novel visualization techniques to represent complex feature relationships or leveraging advanced machine learning interpretability methods (e.g., SHAP, LIME) to better understand which specific features or combinations of features contribute most significantly to normal and anomalous patterns. This will provide richer, actionable insights for maintenance engineers.
2. Real-world Deployment and Edge Computing: Investigating the deployment of the VAE model in real-time, industrial settings, potentially on edge computing devices. This involves optimizing the model for computational efficiency and exploring its integration with existing IoT infrastructures to enable on-site, rapid anomaly detection without constant cloud connectivity.
3. Multi-source Data Fusion: Expanding the model to integrate and leverage data from multiple heterogeneous sensors or data sources (e.g., vibration, temperature, pressure, electrical signals) to build a more holistic understanding of system health and detect more complex, multi-modal anomalies.

Acknowledgment

The authors would like to thank LA pasta for the operational data support for this study.

Funding Sources

The authors declare that no funds, grants, or other support were received during this work.

Conflict of Interest

The authors report there are no competing interests to declare.

REFERENCES

Geisbush, J., & Ariaratnam, S. T. (2023). Reliability centered maintenance (RCM): literature review of current industry state of practice. *Journal of Quality in Maintenance Engineering*, *29*(2), 313-337.

Ma, S., Flanigan, K. A., & Bergés, M. (2024). State-of-the-art review and synthesis: A requirement-based roadmap for standardized predictive maintenance automation using digital twin technologies. *Advanced Engineering Informatics*, *62*, 102800.

Nunes, P., Santos, J., & Rocha, E. (2023). Challenges in predictive maintenance–A review. CIRP Journal of Manufacturing Science and Technology, 40, 53-67.

Chen, C., Fu, H., Zheng, Y., Tao, F., & Liu, Y. (2023). The advance of digital twin for predictive maintenance: The role and function of machine learning. *Journal of Manufacturing Systems*, *71*, 581-594.

Shahin, M., Chen, F. F., Hosseinzadeh, A., & Zand, N. (2023). Using machine learning and deep learning algorithms for downtime minimization in manufacturing systems: An early failure detection diagnostic service. *The International Journal of Advanced Manufacturing Technology*, *128*(9-10), 3857-3883.

Fordal, J. M., Schjølberg, P., Helgetun, H., Skjermo, T. Ø., Wang, Y., & Wang, C. (2023). Application of sensor data based predictive maintenance and artificial neural networks to enable Industry 4.0. *Advances in Manufacturing*, *11*(2), 248-263.

Tapo, A. A., Traoré, A., Danioko, S., & Tembine, H. (2024). Machine Intelligence in Africa: a survey. *arXiv preprint arXiv:2402.02218*.

Mwanza, J., Telukdarie, A., & Igusa, T. (2023). Impact of industry 4.0 on healthcare systems of low-and middle-income countries: a systematic review. *Health and Technology*, *13*(1), 35-52.

Samuel, A. T. (2024). The Justification of Complex Systems Analysis in Better Informing Project Decisions: A Study of the us Surface Transportation Board. *International Journal of Research and Scientific Innovation*, *11*(8), 263-280.

Schlüter, M., Brelsford, C., Ferraro, P. J., Orach, K., Qiu, M., & Smith, M. D. (2023). Unraveling complex causal processes that affect sustainability requires more integration between empirical and modeling approaches. Proceedings of the National Academy of Sciences, 120(41), e2215676120.

Baroud, S. Y., Yahaya, N. A., & Elzamly, A. M. (2024). Cutting-Edge AI Approaches with MAS for PdM in Industry 4.0: Challenges and Future Directions. *Journal of Applied Data Sciences*, *5*(2), 455-473.

Karippur, N. K., Balaramachandran, P. R., & John, E. (2024). Data driven predictive maintenance for large-scale asset-heavy process industries in Singapore. *Journal of Manufacturing Technology Management*, *35*(3), 544-567.

Möller, D. P. (2023). Cybersecurity in digital transformation. In *Guide to Cybersecurity in Digital Transformation: Trends, Methods, Technologies, Applications and Best Practices* (pp. 1-70). Cham: Springer Nature Switzerland.

Rahmanović, A., Hakić, E., Saračević, M., & Međedović, E. (2023). Application and Development of Embedded Systems with IoT Components: Aspect of Safety and Reliability. *SAR Journal (2619-9955)*, *6*(4).

Abouelyazid, M. (2023). Advanced Artificial Intelligence Techniques for Real-Time Predictive Maintenance in Industrial IoT Systems: A Comprehensive Analysis and Framework. *Journal of AI-Assisted Scientific Discovery*, *3*(1), 271-313.

Dayo-Olupona, O., Genc, B., Celik, T., & Bada, S. (2023). Adoptable approaches to predictive maintenance in mining industry: An overview. *Resources Policy*, *86*, 104291.

Chen, C., Fu, H., Zheng, Y., Tao, F., & Liu, Y. (2023). The advance of digital twin for predictive maintenance: The role and function of machine learning. *Journal of Manufacturing Systems*, *71*, 581-594.

Dalhatu, A. A., Sa'ad, A. M., Cabral de Azevedo, R., & De Tomi, G. (2023). Remotely operated vehicle taxonomy and emerging methods of inspection, maintenance, and repair operations: An overview and outlook. *Journal of Offshore Mechanics and Arctic Engineering*, *145*(2), 020801.

Yazdi, M. (2024). Maintenance strategies and optimization techniques. In *Advances in Computational Mathematics for Industrial System Reliability and Maintainability* (pp. 43-58). Cham: Springer Nature Switzerland.

Patil, S. A., Sable, N. P., Mahalle, P. N., & Shinde, G. R. (2023). Intelligent Mechanisms for PdM in Automotive Machinery: A Comprehensive Analysis using ML/DL. *Journal of Electrical Systems*, *19*(2).

Ucar, A., Karakose, M., & Kırımça, N. (2024). Artificial intelligence for predictive maintenance applications: key components, trustworthiness, and future trends. *Applied Sciences*, *14*(2), 898.

Meddaoui, A., Hain, M., & Hachmoud, A. (2023). The benefits of predictive maintenance in manufacturing excellence: a case study to establish reliable methods for predicting failures. *The International Journal of Advanced Manufacturing Technology*, *128*(7-8), 3685-3690.

Rosati, R., Romeo, L., Cecchini, G., Tonetto, F., Viti, P., Mancini, A., & Frontoni, E. (2023). From knowledge-based to big data analytic model: a novel IoT and machine learning based decision support system for predictive maintenance in Industry 4.0. *Journal of Intelligent Manufacturing*, *34*(1), 107-121.

Ooko, S. O., & Karume, S. M. (2024). Application of Tiny Machine Learning in Predicative Maintenance in Industries. *Journal of Computing Theories and Applications*, *2*(1), 132.

Daoudi, N., Smail, Z., & Aboussaleh, M. (2023, February). Machine Learning Based Predictive Maintenance: Review, Challenges and Workflow. In *International Conference on Artificial Intelligence & Industrial Applications* (pp. 71-88). Cham: Springer Nature Switzerland.

Xu, X., Mo, R., Yin, X., Khosravi, M. R., Aghaei, F., Chang, V., & Li, G. (2020). PDM: Privacy-aware deployment of machine-learning applications for industrial cyber–physical cloud systems. *IEEE Transactions on Industrial Informatics*, *17*(8), 5819-5828.

Drakaki, M., Karnavas, Y. L., Tziafettas, I. A., Linardos, V., & Tzionas, P. (2022). Machine learning and deep learning based methods toward industry 4.0 predictive maintenance in induction motors: State of the art survey. *Journal of Industrial Engineering and Management (JIEM)*, *15*(1), 31-57.

Yurek, O. E., Birant, D., & Kut, A. (2022). T-PdM: a tripartite predictive maintenance framework using machine learning algorithms. *International Journal of Computational Science and Engineering*, *25*(3), 325-338.

Niyonambaza, I., Zennaro, M., & Uwitonze, A. (2020). Predictive maintenance (Pdm) structure using internet of things (iot) for mechanical equipment used into hospitals in Rwanda. *Future Internet*, *12*(12), 224.

Carratù, M., Gallo, V., Iacono, S. D., Sommella, P., Bartolini, A., Grasso, F., ... & Patrizi, G. (2023). A novel methodology for unsupervised anomaly detection in industrial electrical systems. *IEEE Transactions on Instrumentation and Measurement*.

Wang, H., Zhang, W., Yang, D., & Xiang, Y. (2022). Deep-learning-enabled predictive maintenance in industrial internet of things: methods, applications, and challenges. *IEEE Systems Journal*, *17*(2), 2602-2615.

Khalil, R. A., Saeed, N., Masood, M., Fard, Y. M., Alouini, M. S., & Al-Naffouri, T. Y. (2021). Deep learning in the industrial internet of things: Potentials, challenges, and emerging applications. *IEEE Internet of Things Journal*, *8*(14), 11016-11040.

Pandey, R., Uziel, S., Hutschenreuther, T., & Krug, S. (2023). Towards deploying DNN models on edge for predictive maintenance applications. *Electronics*, *12*(3), 639.

Moskolaï, W. R., Abdou, W., Dipanda, A., & Kolyang. (2021). Application of deep learning architectures for satellite image time series prediction: A review. Remote Sensing, 13(23), 4822.

Wang, Y., Zhang, Z., Feng, L., Ma, Y., & Du, Q. (2021). A new attention-based CNN approach for crop mapping using time series Sentinel-2 images. *Computers and electronics in agriculture*, *184*, 106090.

Wang, J., Wang, P., Tian, H., Tansey, K., Liu, J., & Quan, W. (2023). A deep learning framework combining CNN and GRU for improving wheat yield estimates using time series remotely sensed multi-variables. *Computers and Electronics in Agriculture*, *206*, 107705.

Weerakody, P. B., Wong, K. W., Wang, G., & Ela, W. (2021). A review of irregular time series data handling with gated recurrent neural networks. Neurocomputing, 441, 161-178.

Fatima, S. S. W., & Rahimi, A. (2024). A Review of Time-Series Forecasting Algorithms for Industrial Manufacturing Systems. *Machines*, *12*(6), 380.

Mienye, I. D., Swart, T. G., & Obaido, G. (2024). Recurrent Neural Networks: A Comprehensive Review of Architectures, Variants, and Applications. *Information*, *15*(9), 517.

Ma, M., & Mao, Z. (2020). Deep-convolution-based LSTM network for remaining useful life prediction. IEEE Transactions on Industrial Informatics, 17(3), 1658-1667.

Oluwasanmi, A., Aftab, M. U., Baagyere, E., Qin, Z., Ahmad, M., & Mazzara, M. (2021). Attention autoencoder for generative latent representational learning in anomaly detection. *Sensors*, *22*(1), 123.

Neloy, A. A., & Turgeon, M. (2024). A comprehensive study of auto-encoders for anomaly detection: Efficiency and trade-offs. *Machine Learning with Applications*, 100572.

Ehrhardt, J., & Wilms, M. (2022). Autoencoders and variational autoencoders in medical image analysis. In *Biomedical Image Synthesis and Simulation* (pp. 129-162). Academic Press.

Akkem, Y., Biswas, S. K., & Varanasi, A. (2024). A comprehensive review of synthetic data generation in smart farming by using variational autoencoder and generative adversarial network. Engineering Applications of Artificial Intelligence, 131, 107881.

Goyal, M., & Mahmoud, Q. H. (2024). A Systematic Review of Synthetic Data Generation Techniques Using Generative AI. *Electronics*, *13*(17), 3509.

Figueira, A., & Vaz, B. (2022). Survey on synthetic data generation, evaluation methods and GANs. *Mathematics*, *10*(15), 2733.

Niu, Z., Yu, K., & Wu, X. (2020). LSTM-based VAE-GAN for time-series anomaly detection. *Sensors*, *20*(13), 3738.

Angiulli, F., Fassetti, F., & Ferragina, L. (2020, October). Improving deep unsupervised anomaly detection by exploiting VAE latent space distribution. In *International Conference on Discovery Science* (pp. 596-611). Cham: Springer International Publishing.

Neloy, A. A., & Turgeon, M. (2024). A comprehensive study of auto-encoders for anomaly detection: Efficiency and trade-offs. *Machine Learning with Applications*, 100572.

Costa, N., & Sánchez, L. (2022). Variational encoding approach for interpretable assessment of remaining useful life estimation. *Reliability Engineering & System Safety*, *222*, 108353.

SenthilVadivel, T., Suseelan, A., Karthick, K., Safran, M., & Alfarhood, S. (2024). Experimental Investigation and Machine Learning Prediction of Mechanical Properties of Rubberized Concrete for Sustainable Construction. *Scientific Reports*, *14*(22725)

A Hybrid Deep Learning Prediction Method of Remaining Useful Life for Rolling Bearings Using Multiscale Stacking Deep Residual Shrinkage Network" by Song et al. (2023)