Svedberg Open
DISSEMINATION OF KNOWLEDGE

International Journal of Artificial Intelligence and Machine Learning

Publisher's Home Page: https://www.svedbergopen.com/

Research Paper Open Access

Distributional Shift Detection Algorithms for Proactive Model Maintenance in Production

Dr. G. Shanmugarathinam[1*], Dr. G Chandra Sekhar[2], Betty Elezebeth Samuel[3], Abdullo Nabiyev[4] Abdulkhamid Akbarov[5], Dilrabo Muqumova[6]

[1*]Professor, School of Computer Science and Engineering, Presidency University, Bengaluru, Karnataka, India.
Email: shanmugarathinam@presidencyuniversity.in

[2]Associate Professor, Computer Science and Engineering, Institute of Aeronautical Engineering, Dundigal, Hyderabad, India. Email: sekhar.gillala@gmail.com, ORCID: 0000-0002-3894-4205

[3]Department of Computer Science, College of Engineering & Computer Science, Jazan University, Jazan, Saudi Arabia.
Email: bsamuel@jazanu.edu.sa

[4]Samarkand state medical university, Uzbekistan. Email: tojik.sam@gmail.com, https://orcid.org/0009-0008-3433-8700

[5]Senior Lecturer, Tashkent State University of Economics, Andijan, Uzbekistan. E-mail: a.akbarov@tsue.uz, https://orcid.org/0000-0002-6478-8293

[6]Researcher, Tashkent Institute of Irrigation and Agricultural Mechanization Engineers, National Research University, Tashkent, Uzbekistan. E-mail: d.mukumova@tiiame.uz, https://orcid.org/0000-0002-7097-093X

*Corresponding author: Email: shanmugarathinam@presidencyuniversity.in

Abstract

ML models used in production settings are vulnerable to performance deterioration due to distributional shifts, in which case the statistical characteristics of incoming data differ from those of training data samples. Early detection of distributional shifts in ML models is important for ensuring their reliability and safety. This paper introduces an algorithm for early detection of distributional shifts based on latent projection and the calculation of divergence measures of projections from production data to the training data centroid in a low-dimensional latent space obtained by means of an autoencoder embedding. Once a shift exceeds a predetermined threshold, maintenance actions are automatically taken to mitigate the issue. Tests have been performed using benchmarks consisting of artificial distributional shifts on popular image datasets Fashion-MNIST and SVHN, where the proposed method was found to perform better than existing techniques using KL-divergence or autoencoder reconstruction measures, exhibiting higher detection accuracy and lower false positive rates at moderate latencies.

Keywords: Distributional Shift, Latent Projection, Proactive Maintenance, Autoencoder, Divergence Score, Model Monitoring, Real-Time Deployment

1. Introduction

ML algorithms are extensively used in production systems to automate important decision-making processes in various sectors such as financial services, healthcare, and autonomous machines. Nevertheless, their performance may deteriorate over time owing to distributional shift, which refers to alterations in the statistical properties of the input data relative to the training dataset. The negative impacts of distributional shift include errors in prediction, unreliability, and other hazards, which could affect business operations negatively [9]. The problem being addressed in this study involves inefficiency in monitoring changes in distributions in machine learning models while in use, which may affect model maintenance and lead to unexpected performance issues.

Proactive model maintenance is the practice of evaluating and improving ML models consistently without waiting until the model starts underperforming [6]. The inclusion of a distributional shift detector into the model helps maintain its reliable performance. This study proposes an innovative algorithm that can detect distributional

shift quickly and precisely, thus helping to implement machine learning algorithms effectively in real-life scenarios.

Key Contributions

1) Presented a new approach based on latent projections for identifying small and large-scale changes in distribution within ML models that are deployed for commercial purposes.
2) Better than any other method, like Kullback–Leibler Divergence and autoencoder approaches, because of its higher level of efficiency and low computational costs.
3) Facilitated quick incorporation in maintenance algorithms for immediate detection and retraining purposes.

The introduction of this paper is provided in Section I: Introduction. Section II: Related Work provides information about relevant literature related to the detection algorithms and their drawbacks. Section III: Proposed Algorithm explains the proposed latent projection algorithm. The performance of the algorithm is explained in Section IV: Benchmarking.

1. Related Work

Predictive maintenance is one of the most discussed topics within the fields of manufacturing, industries, and urban infrastructure systems because of its ability to save on expenses and avoid equipment failure. Federated learning for predictive maintenance, which exhibits resistance to the changes in the distribution of time-series data in manufacturing procedures, together with fuzzy classification techniques for detecting critical cardiac arrhythmias, is worth mentioning [1] [2][14]. Proactive maintenance model based on reinforcement learning in the rubber industry, with focus on adaptability to changes in operational circumstances [3]. Stock trend prediction using machine learning, which illustrates the wide scope of application of predictive models [4][16].

Proactive fault prediction for resource monitoring in IoT-based systems, while applications of AI in sustainable agriculture, considering both efficiency and equity, should be noted [5] [6]. Source component shift detection for improving remaining useful life estimation in alarm-based predictive maintenance [7][15]. Deep learning in combination with geographic information systems for automobile maintenance prediction and Gaussian mixture models and mean-shift clustering for wafer transfer robot maintenance [8] [9].

A predictive maintenance approach using IoT for smart manufacturing, along with probability distribution function and hierarchical clustering for condition-based scheduling, optimizing maintenance planning [10] [11]. An AI-based digital twin approach for urban infrastructure to generate real-time predictive insight. The AI-based predictive analytics frameworks for industrial equipment reliability, and process-based research on digital tools for vocabulary retention, focusing on feedback mechanisms, which could help in developing predictive maintenance training and education models for humans [12] [13].

2. Methodology

As illustrated in Figure 1 below, the general procedure of identifying distributional changes in machine learning models used in practical settings and performing maintenance activities consists of multiple stages. Firstly, data received from different sources undergoes preprocessing and is fed into the model in use. Next, the monitoring process identifies the difference scores between the distribution under observation and reference one, triggering the retraining process of the model.

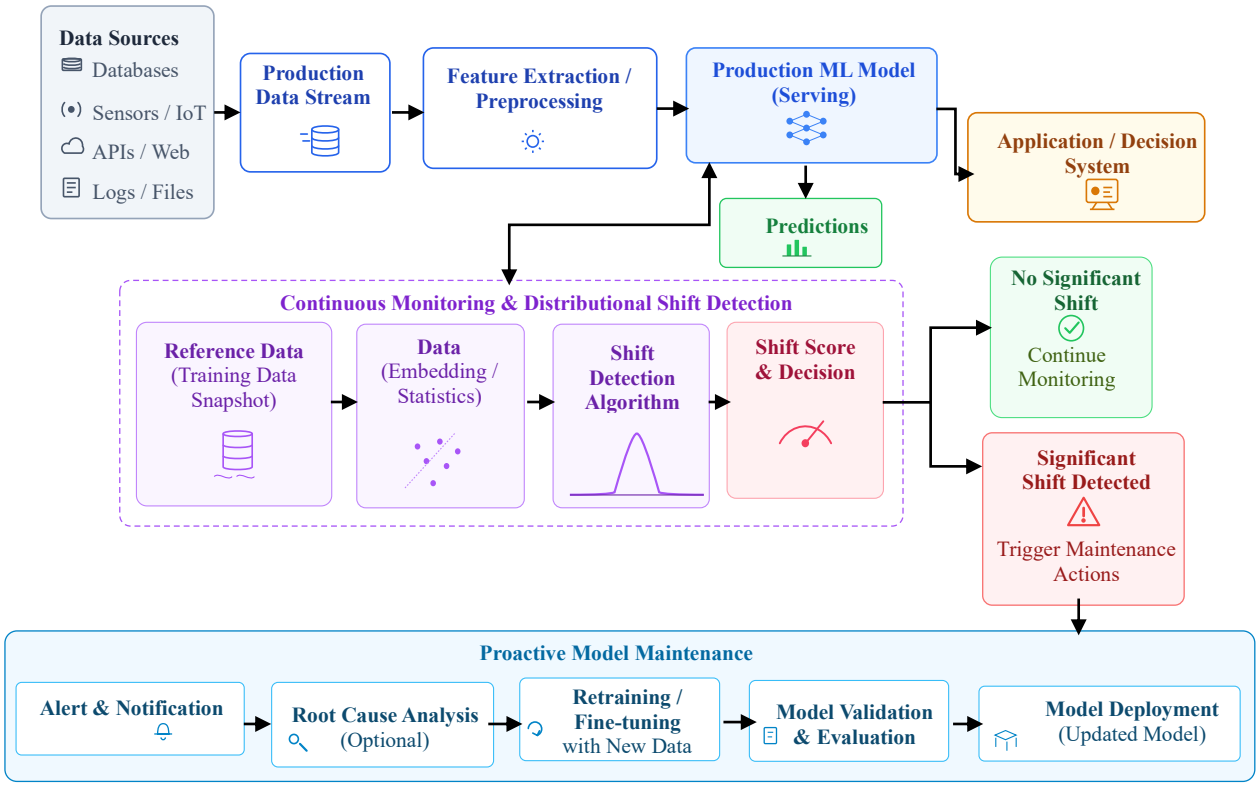

Figure 1: Proactive Model Maintenance Architecture with Distributional Shift Detection

3.1 Relevance of Proactive ML Model Maintenance in Production

Maintenance of models in a proactive way ensures the continuous effectiveness and reliability of machine learning technologies in application scenarios. In this way, potential threats that can occur when the model performance deteriorates can be detected, which leads to erroneous decision-making, time wastage, and dissatisfaction of customers. The reason behind the need for such an approach lies in production environments where distributions are constantly changing.

3.2 Methods of Re-training Models with Changing Distributions

A few methods can be suggested regarding the problem discussed. Retraining machine learning algorithms with new datasets can keep them up-to-date; online incremental learning will provide an opportunity to integrate new samples into a model continuously. Dynamic adjustment of detection thresholds depending on variance will decrease false positives and other types of errors. Ensemble approaches and data augmentation will contribute to improving the reliability of machine learning systems under changing distributions.

3.3 Integration of the Algorithm in Proactive Maintenance Process

The latent space distributional shift detection algorithm can be smoothly integrated within proactive maintenance workflows. The process involves continuous monitoring of the input stream, scoring the divergence measures, and identifying any shift occurring in real-time. Once the shift is recognized, automatic measures can be taken including training, tuning, or deploying a new model. In the process, human involvement will be minimized, ensuring low downtime and maintaining the accuracy and reliability of models, particularly where huge high-dimensional datasets are involved.

Algorithm: Latent Projection-Based Shift Detection

1. Input: $X_{train}, X_{prod}, \tau$
2. Train autoencoder on $X_{train} \rightarrow$ latent mapping $f(\cdot)$
3. Embed $X_{prod} \rightarrow f(X_{prod})$
4. Compute divergence score D using Equation (1)
5. If $D > \tau$, flag shift; else continue monitoring

6. Output: Shift detection signal for proactive model maintenance

Latent Projection-Based Shift Detection algorithm encodes the production data in a latent space through the autoencoder learned from the reference data, calculates a divergence score based on the difference from the training centroid, and triggers shift detection when it surpasses the threshold value to maintain the model proactively.

3. Result

4.1 Setting Up Experiments to Test the Algorithm

Benchmarked datasets, like Fashion-MNIST and SVHN (Street View House Numbers), are used to conduct the experiments with artificial distribution shifts simulated to represent a production environment. The new latent projection method is compared with KL-divergence and autoencoder reconstruction methods. Accuracy of shift detection (%), false positive rate (%), and latency (ms) were used to measure results of the comparison. All experiments were performed in Python 3.10, TensorFlow 2.12, and NVIDIA RTX 3060 GPU environments.

4.2 Results of Experiments Comparing the Algorithm to Existing Approaches

Experiments performed show the efficiency of the proposed latent projection algorithm as a method to detect shifts in comparison with traditional techniques. Figure 1 shows a comparative analysis between the proposed latent projection algorithm and baseline algorithms of KL-divergence and autoencoders reconstruction. The metrics used in evaluation are accuracy of detection, false positives, and latency of computation. The latent projection algorithm shows the highest level of accuracy, low false positives, and a moderate level of computation latency, thus being applicable for real-time use. The baseline approaches demonstrate high efficiency in detecting large shifts but lack sensitivity to small distribution changes.

Table 1: Performance Comparison of Distributional Shift Detection Algorithms

Algorithm	Accuracy (%)	False Positive Rate (%)	Latency (ms)
KL-Divergence	78.4	15.2	12
Autoencoder Reconstruction	84.7	12.3	25
Latent Projection (Proposed)	92.1	7.6	18

Figure 2 illustrates the relationship between shifted accuracy and reference accuracy for the baseline model. Each point represents a dataset instance, with the linear fit highlighting overall trends. The dashed line $y = x$indicates ideal performance, showing deviations caused by distributional shifts in production data.

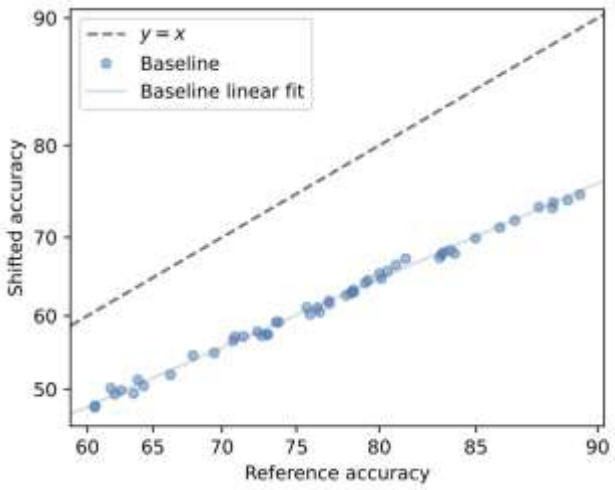

Figure 2: Shifted Accuracy versus Reference Accuracy for Baseline Model

4.3 Discussion of the Implications of the Experimental Results

The experiments demonstrate that the latent projection-based approach outperforms traditional methods in detecting both subtle and significant distributional changes with little or no false positives. Low latency allows

for deploying the solution in real time to production pipelines. Thus, it is expected that using the algorithm within proactive maintenance systems will help prevent system downtime, performance degradation, and ensure model accuracy. In summary, the findings of the experiments show that latent-space divergence scoring is a scalable and reliable tool for real-world applications.

4. Conclusion

This work discussed a latent-projection based distributional shift detection algorithm for implementing proactive maintenance of ML models used in production. The method involves projecting the production data set into a lower-dimensional latent space using autoencoders and measuring divergences from the data centroid obtained during the training phase. The experiment results showed that the latent projection algorithm performed better compared to KL-divergence and autoencoder-reconstruction-based approaches in terms of accuracy and moderate latency.

Embedding the algorithm in the process of maintaining the model will lead to automation in the process of retraining, fine-tuning, and deployment, which will reduce any form of manual labor or downtime. Future areas of research for this project include the implementation of this approach on larger datasets using reinforcement learning for dynamic thresholds and the efficiency of implementation in distributed/edge computing environments. Other potential areas for investigation may involve uncertainty estimation for detecting shift signals, increasing model interpretability, and analyzing the algorithm's effects on drift for extended periods of time.

Declaration Statement

Conflict of Interest: The authors declare no conflicts of interest associated with this research.

Funding: This research received no specific grant from any funding agency, commercial entity, or not-for-profit organization.

Data Availability: The datasets used in this study are publicly available, including Fashion-MNIST (https://github.com/zalandoresearch/fashion-mnist) and SVHN (http://ufldl.stanford.edu/housenumbers/).

References

1. Ahn, J., Lee, Y., Kim, N., Park, C., & Jeong, J. (2023). Federated learning for predictive maintenance and anomaly detection using time series data distribution shifts in manufacturing processes. *Sensors, 23*(17), 7331. https://doi.org/10.3390/s23177331
2. Khanaa, V., Thooyamani, K. P., & Udayakumar, R. (2013). Categorization and forecast of critical cardiac arrhythmias using filter bank and fuzzy classification approach. *Middle-East Journal of Scientific Research, 18*(12), 1798–1802.
3. Senthil, C., & Sudhakara Pandian, R. (2022). Proactive maintenance model using reinforcement learning algorithm in rubber industry. *Processes, 10*(2), 371. https://doi.org/10.3390/pr10020371
4. Alavi, S. E., Sinaei, H., & Afsharirad, E. (2015). Predict the trend of stock prices using machine learning techniques. *International Academic Journal of Economics, 2*(2), 1–11.
5. Chowdhury, A., Raut, S., & Pal, A. (2022). Internet of Things resource monitoring through proactive fault prediction. *Computers & Industrial Engineering, 169*, 108265. https://doi.org/10.1016/j.cie.2022.108265
6. Narayanan, L., & Rajan, A. (2024). Artificial intelligence for sustainable agriculture: Balancing efficiency and equity. *International Journal of SDG's Prospects and Breakthroughs, 2*(1), 4–6.
7. Fathi, K., Sadurski, M., Kleinert, T., & van de Venn, H. W. (2023, October). Source component shift detection and classification for improved remaining useful life estimation in alarm-based predictive maintenance. In *2023 23rd International Conference on Control, Automation and Systems (ICCAS)* (pp. 975–980). IEEE.
8. Mishra, N., Haval, A. M., Mishra, A., & Dash, S. S. (2024). Automobile maintenance prediction using integrated deep learning and geographical information system. *Indian Journal of Information Sources and Services, 14*(2), 109–114. https://doi.org/10.51983/ijiss-2024.14.2.16
9. Jeon, J. E., Song, W. S., Hong, S. J., & Han, S. S. (2024). Predictive maintenance system for wafer transfer robot using Gaussian mixture model and mean-shift clustering. *Procedia Computer Science, 237*, 453–460. https://doi.org/10.1016/j.procs.2024.05.245
10. Novak, P., & Vacek, T. (2023). An IoT-driven predictive maintenance model for smart manufacturing systems. *International Academic Journal of Innovative Research, 10*(3), 9–15. https://doi.org/10.71086/IAJIR/V10I3/IAJIR1019

11. Kordestani, M., Rezamand, M., Orchard, M. E., & Saif, M. (2025). Condition-based maintenance scheduling using probability distribution function and agglomerative hierarchical clustering approaches: AI-driven predictive maintenance mapping. *IEEE Systems, Man, and Cybernetics Magazine, 11*(3), 61–72.
12. Prasath, C. A. (2025). AI-enabled digital twin framework for predictive maintenance in smart urban infrastructure. *Journal of Smart Infrastructure and Environmental Sustainability, 2*(1), 1–10.
13. Rony, M. A. (2025). AI-enabled predictive analytics and fault detection frameworks for industrial equipment reliability and resilience. *ASRC Procedia: Global Perspectives in Science and Scholarship, 1*(1), 705–736.
14. K P Uvarajan. (2026). Design of a Low-Power VLSI Architecture for Real-Time Image Processing Using Optimized DSP Algorithms. *Journal of Integrated VLSI and Signal Processing*, 1–10.
15. D. Barhani and P. Kharabi, "A Hardware–Software Co-Design Approach for Adaptive Embedded Systems Using Reconfigurable Logic", *Journal of Reconfigurable Hardware Architectures and Embedded Systems*, vol. 1, no. 1, pp. 43–51, Dec. 2024.
16. K. Maidanov, & Jeon Sungho. (2025). AI-Integrated System-on-Chip (SoC) Architectures for High-Performance Edge Computing: Design Trends and Optimization Challenges. *Journal of Integrated VLSI, Embedded and Computing Technologies* , *2*(3), 47-55.