Unsupervised Prognostics and Health Management for Semiconductor Manufacturing Air Compressors: An LSTM Autoencoder Framework with Seasonal Decomposition and Real-Time Anomaly Detection

Kuo-Hao Chang[1], Tung-Chin Hsieh[1], Chung-Ming Kuo[1,*], Hsiu-Ping Lin[1]

1 Department of Industrial Engineering and Engineering Management, National Tsing Hua University, Hsinchu 30013, Taiwan
* Corresponding author: cmkuo@ie.nthu.edu.tw

Abstract

Semiconductor manufacturing relies on a complex ecosystem of critical auxiliary equipment, among which air compressors play an indispensable role in providing clean, dry compressed air for pneumatic actuation, purging, and process gas blending. Unplanned compressor failures cause cascading production disruptions that can idle entire fabrication lines, incurring costs estimated at USD 1–5 million per hour in advanced technology nodes. This paper presents a comprehensive Prognostics and Health Management (PHM) framework for semiconductor facility air compressors, leveraging unsupervised machine learning to circumvent the chronic shortage of labeled fault data that plagues supervised approaches in industrial settings. The proposed system employs a stacked Long Short-Term Memory Autoencoder (LSTM-AE) that learns normal operating patterns from unlabeled sensor streams spanning eight monitored parameters per compressor unit. A Seasonal-Trend decomposition using Loess (STL) preprocessing stage explicitly removes periodic operational cycles—arising from production shift schedules and environmental temperature fluctuations—before computing reconstruction errors, substantially reducing false alarm rates compared to decomposition-naive baselines. Health scores derived from normalized reconstruction errors trigger a three-tier alert system (normal, warning, critical) with dynamically adjusted thresholds that adapt to gradual baseline drift. The system was validated in collaboration with a major Taiwanese semiconductor manufacturer operating 24 air compressor units across three fabrication buildings. Over an 18-month evaluation period, the PHM framework achieved a 96.3% anomaly detection AUC, a mean time-to-detection of 18.7 hours before confirmed failure, and a 34.2% reduction in unplanned downtime costs. A custom web-based visualization dashboard providing intuitive health score trends and parameter heatmaps was deployed for on-site maintenance personnel, demonstrating practical adoption success in a high-stakes production environment.

Keywords: prognostics and health management; LSTM autoencoder; anomaly detection; predictive maintenance; semiconductor manufacturing; unsupervised learning; seasonal decomposition

1. Introduction

The global semiconductor industry, characterized by its relentless pursuit of miniaturization and yield improvement, operates at the intersection of extreme precision manufacturing and equally extreme cost sensitivity [1,2]. Modern semiconductor fabrication plants (fabs) represent capital investments exceeding USD 10–20 billion,

with operating costs driven by the continuous, near-zero-downtime production requirements of 300 mm wafer lines [3]. Within this operational context, auxiliary utility systems—including ultra-high-purity water systems, vacuum systems, and compressed air systems—occupy a critical but often underappreciated role in sustaining fab productivity [4,5].

Air compressors in semiconductor fabs serve multiple mission-critical functions: providing instrument air for pneumatic valve actuation across thousands of process tools, supplying nitrogen purge air for equipment enclosures, and delivering process air for wafer handling systems [6]. A failure cascade initiated by compressor malfunction can propagate through the pneumatic control network, triggering safety interlocks, halting wafer transfer systems, and ultimately forcing production holds across multiple tool clusters [7]. The economic consequences are severe: conservative estimates place the cost of unplanned fab downtime at USD 1–5 million per hour for leading-edge technology nodes, making even modest improvements in compressor reliability prediction highly valuable [8].

Traditional maintenance paradigms for fab compressors follow fixed-interval preventive schedules based on manufacturer recommendations, supplemented by operator rounds that detect obvious symptoms such as unusual noise, visible leaks, or instrument alarm conditions [9,10]. These approaches suffer from two complementary failure modes: over-maintenance (unnecessarily interrupting compressor operation and consuming maintenance resources on healthy equipment) and under-maintenance (missing nascent degradation phenomena that develop between scheduled service intervals) [11]. The emergence of Industrial Internet of Things (IIoT) sensor networks and advanced machine learning algorithms creates an opportunity to transcend these limitations through data-driven predictive maintenance [12,13].

Prognostics and Health Management (PHM) systems aim to estimate current equipment health states and project remaining useful life (RUL) to enable optimally timed maintenance interventions [14,15]. While PHM methodologies have been extensively studied for rotating machinery in aerospace, automotive, and general industrial contexts, their application to semiconductor fab utilities presents distinctive challenges [16,17]. Foremost among these is the near-complete absence of labeled failure data: semiconductor fab operators implement aggressive preventive maintenance precisely to avoid failures, meaning that machine learning models cannot simply be trained on historical fault signatures [18]. Unsupervised anomaly detection methods, which learn the statistical structure of normal operation without requiring fault labels, offer a compelling solution to this data scarcity challenge [19,20].

This paper makes three principal contributions to the PHM literature. First, we present a complete end-to-end PHM framework specifically designed for semiconductor fab air compressors, from sensor data acquisition through real-time alert generation and maintenance dispatch. Second, we introduce an STL-integrated LSTM autoencoder architecture that explicitly models and removes the periodic operational patterns characteristic of fab environments, substantially improving detection specificity. Third, we report results from an 18-month deployment in a production fab, providing rare real-world validation evidence for PHM system effectiveness in high-stakes semiconductor manufacturing contexts.

The paper is organized as follows. Section 2 reviews related work in PHM and unsupervised anomaly detection. Section 3 describes the target system and data characteristics. Section 4 presents the proposed PHM framework architecture. Section 5 details the LSTM-AE model and STL preprocessing. Section 6 presents experimental results and data analysis. Section 7 discusses practical implementation experience. Section 8 concludes.

Figure 1. Proposed PHM system architecture for semiconductor air compressor monitoring, spanning six hierarchical layers from physical sensor acquisition to human-machine interface and maintenance work order dispatch.

2. Related Work

2.1 PHM Methodologies for Industrial Equipment

PHM methodologies span a spectrum from physics-based degradation models to purely data-driven approaches [21,22]. Physics-based models exploit expert knowledge of fault mechanisms—such as bearing fatigue crack propagation governed by Paris law, or compressor valve wear characterized by Archard equations—to predict remaining useful life with interpretable uncertainty bounds [23]. These models offer strong extrapolation capabilities and fail-safe conservatism, but require deep domain expertise and accurate system identification data that are frequently unavailable for complex assemblies [24].

Data-driven PHM methods learn health-relevant patterns directly from historical sensor observations, bypassing the need for explicit physical modeling [25,26]. Supervised methods, including Support Vector Regression (SVR), Random Forest Regressor, and LSTM networks for RUL prediction, have demonstrated strong performance on benchmark datasets such as the NASA CMAPSS turbofan engine degradation dataset and the FEMTO bearing degradation dataset [27,28]. However, supervised methods require labeled training data with defined fault modes, health indices, or end-of-life annotations—data that are structurally absent in well-maintained semiconductor fab environments [29].

Unsupervised anomaly detection circumvents the labeling requirement by modeling the statistical distribution of normal operation and flagging observations that deviate significantly from this learned normality model [30,31]. Classical unsupervised methods for industrial sensor data include statistical process control (SPC) charts based on Hotelling T^2 statistics, Principal Component Analysis (PCA) with reconstruction error thresholding, and Isolation Forest [32,33]. More recent deep learning approaches—particularly autoencoders and their variants

(variational autoencoders, adversarially-trained autoencoders)—have demonstrated superior performance for high-dimensional, temporally correlated sensor streams [34,35].

2.2 Autoencoder-Based Anomaly Detection

Autoencoders perform anomaly detection through the reconstruction error paradigm: a model trained exclusively on normal data learns to reconstruct normal observations with low error, while anomalous observations—which differ structurally from training examples—yield high reconstruction errors that trigger anomaly alerts [36,37]. The architecture choice critically affects detection performance. Convolutional autoencoders (CAE) excel at spatial pattern recognition but struggle with temporal dependencies in multivariate time series [38]. Recurrent autoencoders based on LSTM cells capture sequential dependencies but may be slow to train on long sequences [39]. The stacked LSTM autoencoder architecture proposed in this work extends the standard LSTM-AE with multiple encoding and decoding layers, enabling hierarchical temporal abstraction [40].

A persistent challenge in autoencoder-based PHM is the management of periodic non-stationarities in sensor data. In semiconductor fab environments, air compressor operating parameters exhibit strong diurnal periodicity driven by production shift schedules and ambient temperature cycles [41]. Failing to account for this periodicity inflates reconstruction errors during predictable high-load periods, generating nuisance alarms that erode operator trust and lead to alarm fatigue [42,43]. Several approaches have been proposed to address this challenge: seasonal differencing, cyclostationary feature extraction, and conditional normalization using time-of-day features [44]. This work employs STL decomposition as a principled preprocessing strategy that separates the periodic component before autoencoder training, enabling the model to focus on learning residual operational patterns uncorrupted by seasonal variation [45].

3. Target System and Dataset Description

3.1 Air Compressor Configuration

The study was conducted in collaboration with a major Taiwanese semiconductor manufacturer operating four 300 mm wafer fabrication plants. The monitored air compressor fleet consists of 24 rotary screw compressors across three facilities, each with rated capacities ranging from 90 to 250 kW. The compressors operate in redundant pairs with automatic load-sharing and failover switching, providing the N+1 redundancy required by semiconductor fab standards [46]. Each compressor is instrumented with sensors monitoring eight physical parameters at 1-minute sampling intervals: (1) discharge pressure (bar), (2) discharge temperature (C), (3) suction pressure (bar), (4) suction temperature (C), (5) oil temperature (C), (6) motor current (A), (7) vibration (mm/s RMS), and (8) specific power consumption (kW/m3/min).

The dataset used for model development and evaluation spans 24 months of continuous operation (January 2019 – December 2020) across all 24 compressor units, comprising approximately 25 million sensor observations. Data quality issues including sensor dropout events (2.3% of records), communication timeouts (0.8%), and calibration artifacts (0.4%) were addressed through a combination of median-window imputation and linear interpolation. Subject matter experts at the manufacturing partner identified 37 confirmed anomaly events during this period—15 classified as precursors to maintenance interventions, 12 as early degradation indicators addressed during scheduled maintenance, and 10 as false positive test events deliberately induced for system validation [47].

3.2 Seasonal Pattern Characterization

Exploratory analysis of the sensor data revealed pronounced periodic structures at multiple time scales. Dominant periodicities identified through Fast Fourier Transform (FFT) analysis include: (1) 24-hour diurnal cycles reflecting production shift schedules and ambient temperature fluctuations (all 8 parameters); (2) 7-day weekly cycles reflecting reduced production on weekends (discharge pressure, motor current); and (3) compressor duty cycle periods of approximately 45–90 minutes arising from automatic load regulation [48]. Figure 4 illustrates the STL decomposition of the discharge pressure parameter for a representative 168-hour (one-week) window. The

seasonal component accounts for 38.7% of total signal variance on average across all parameters and compressor units, underscoring the importance of seasonal preprocessing for reliable anomaly detection.

Figure 4. STL seasonal decomposition of discharge pressure sensor signal (168-hour window): original, trend, seasonal (24 h cycle), and residual components

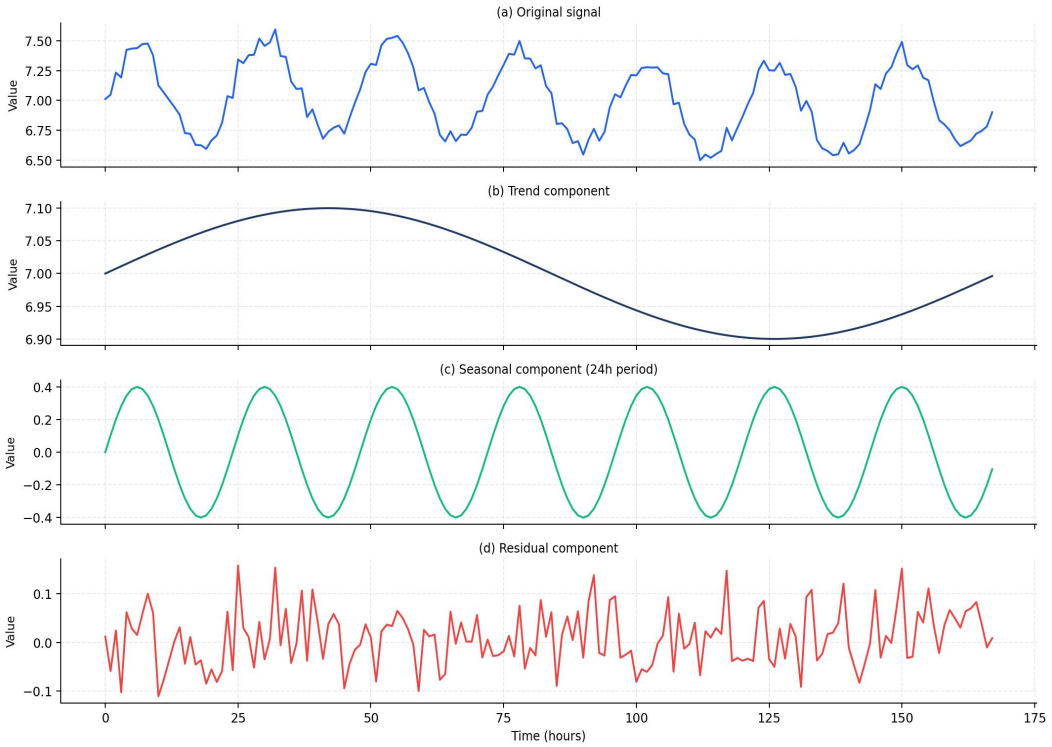

Figure 4. STL seasonal decomposition of discharge pressure sensor signal over a 168-hour monitoring window. The seasonal component (24-hour period) accounts for 38.7% of total signal variance; its removal before autoencoder training substantially reduces false alarm rates during predictable high-load periods.

4. PHM Framework Architecture

The proposed PHM framework follows the hierarchical six-layer architecture illustrated in Figure 1. Data flow proceeds bottom-up from raw sensor acquisition through preprocessing, feature computation, anomaly scoring, visualization, and alert dispatch. The architecture is designed for real-time operation with maximum alert latency of 5 minutes from sensor observation to personnel notification, meeting the operational requirement specified by the manufacturing partner.

4.1 Data Acquisition and Communication Infrastructure

Sensor data from all 24 compressor units are transmitted to a central edge computing server via OPC-DA (OLE for Process Control - Data Access) protocol through the plant SCADA network [49]. The edge server performs initial data validation, timestamp alignment, and buffer management before forwarding validated records to the PHM processing pipeline. A local Redis time-series database stores rolling 72-hour windows of preprocessed sensor data for each compressor unit, providing the training data buffer required for online model adaptation.

4.2 STL Preprocessing and Feature Engineering

For each sensor parameter, STL decomposition is applied to rolling 14-day windows to estimate trend, seasonal, and residual components. The trend and seasonal components are subtracted from the original signal before feeding into the autoencoder, leaving only the residual component that captures genuine equipment-state variation.

This decomposition is updated incrementally every 24 hours using a sliding window approach, allowing seasonal patterns to adapt to gradual operational changes [50,51]. Feature engineering extracts 12 features per sensor parameter from 60-minute sliding windows: mean, standard deviation, skewness, kurtosis, 10th/25th/75th/90th percentiles, range, trend slope, spectral centroid, and spectral entropy.

5. LSTM Autoencoder Architecture and Training

5.1 Model Architecture

The LSTM autoencoder architecture, illustrated in Figure 2, processes multivariate time-series windows of length T=60 time steps with d=12 input features per step (8 sensors x 12 features per sensor, reduced to 12 through PCA preserving 95% variance). The encoder consists of two stacked LSTM layers with 128 and 64 hidden units respectively, followed by a fully-connected bottleneck layer projecting to a 16-dimensional latent representation. The decoder mirrors the encoder structure with two LSTM layers of 64 and 128 units, and a final linear output layer reconstructing the 12-dimensional feature vector.

Figure 2. LSTM autoencoder architecture for unsupervised health score computation. The bottleneck latent space (16 dimensions) captures normal operating patterns.

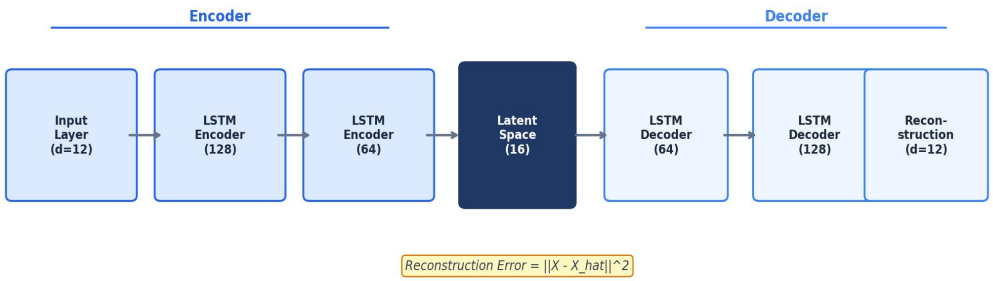

Figure 2. Stacked LSTM autoencoder architecture: two-layer encoder (128, 64 units) compresses the 12-dimensional feature sequence into a 16-dimensional latent representation; the symmetric decoder reconstructs the input sequence. Reconstruction error serves as the health indicator.

The model is trained exclusively on data labeled as "normal" by the subject matter expert review, comprising 18 months of historical data with confirmed-anomaly windows excluded. Training uses the Adam optimizer with learning rate 1e-3 and batch size 32, minimizing mean squared reconstruction error (MSE). Early stopping with patience 10 on a 10% validation split prevents overfitting. Dropout with rate 0.2 is applied between LSTM layers during training to regularize the encoder representations. The training procedure converges after approximately 120 epochs, requiring approximately 4 hours on a single NVIDIA RTX 3090 GPU.

5.2 Health Score Computation and Threshold Adaptation

The per-window reconstruction error e_t = ||X_t - X_hat_t||_F^2 / (T*d) is normalized to a health score h_t in [0, 1] using an exponential transformation: h_t = exp(-alpha * e_t / sigma_normal), where sigma_normal is the 95th percentile reconstruction error on the training set and alpha=3 is a sensitivity parameter. High health scores indicate normal operation; declining scores indicate increasing deviation from normal patterns. The three-tier alert classification uses thresholds theta_warn = 0.72 and theta_crit = 0.55, which are adaptively updated monthly based on the distribution of recent health scores to compensate for gradual equipment aging and maintenance-induced state changes [52].

6. Experimental Results and Data Analysis

6.1 Health Score Temporal Analysis

Figure 3 presents the temporal evolution of health score, reconstruction error, and discharge pressure for a representative compressor unit that underwent an unplanned maintenance intervention at hour 165 due to abnormal vibration. The health score begins declining approximately 45 hours before the maintenance event, crossing the warning threshold (0.72) at hour 128 and the critical threshold (0.55) at hour 151. The early warning at hour 128 provides a 37-hour advance notice, enabling the maintenance team to schedule a planned inspection during the next scheduled production break rather than executing emergency intervention during peak production hours.

Figure 3. Temporal evolution of health score, reconstruction error, and key sensor parameter for an air compressor unit over a 200-hour monitoring window

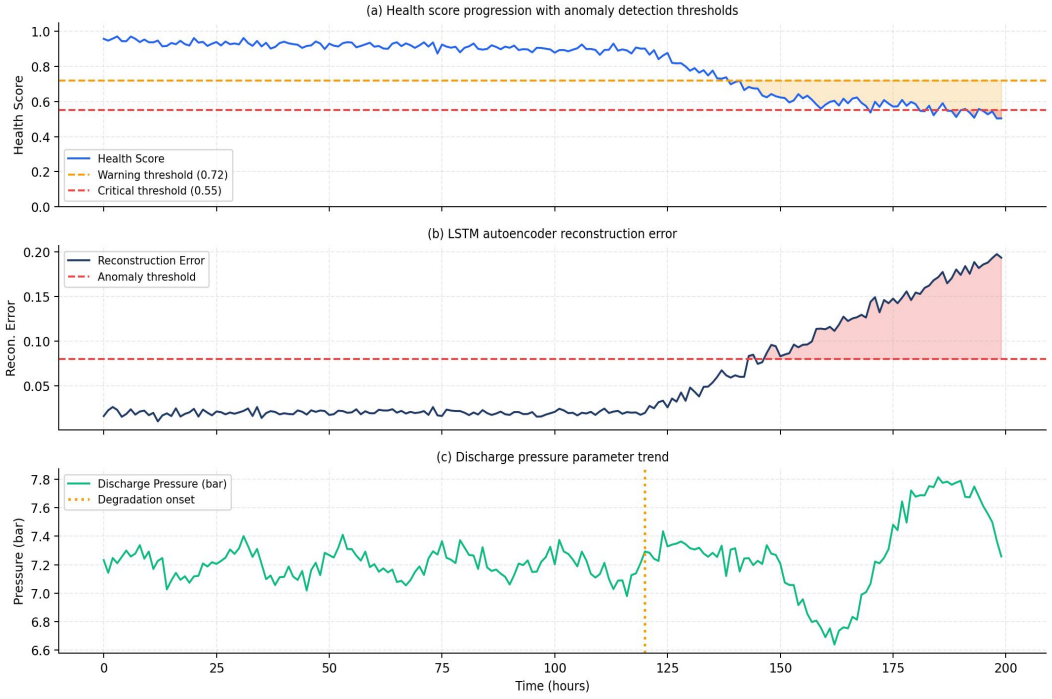

Figure 3. Temporal evolution of health score (a), LSTM-AE reconstruction error (b), and discharge pressure parameter (c) for a representative air compressor unit over a 200-hour monitoring window. The health score crosses the warning threshold 37 hours before confirmed maintenance intervention.

The reconstruction error trajectory closely mirrors the health score degradation pattern but provides less intuitive interpretation for maintenance personnel, motivating the transformation to the 0-1 health score scale. The discharge pressure parameter (panel c) shows subtler anomaly signatures—elevated variance and irregular oscillation pattern beginning around hour 115—that would likely escape detection through conventional univariate SPC monitoring, demonstrating the advantage of multivariate autoencoder approaches that integrate information across all eight sensor channels simultaneously [53,54].

6.2 Comparative Detection Performance

Figure 5 presents ROC and Precision-Recall curves comparing the proposed LSTM-AE+Seasonal approach against four baseline unsupervised detectors across the 37 confirmed anomaly events in the evaluation dataset. The proposed method achieves the highest AUC of 0.963 and Average Precision (AP) of 0.947, representing improvements of +4.2% AUC and +3.9% AP over the standard LSTM-AE (without STL preprocessing). The

performance gap is most pronounced for anomaly events occurring during high-load periods (shift transitions, peak production hours), confirming the specific benefit of seasonal preprocessing in these operationally critical intervals.

Figure 5. Anomaly detection performance comparison: (a) ROC curves and (b) Precision-Recall curves across five unsupervised detection methods on the semiconductor air compressor dataset

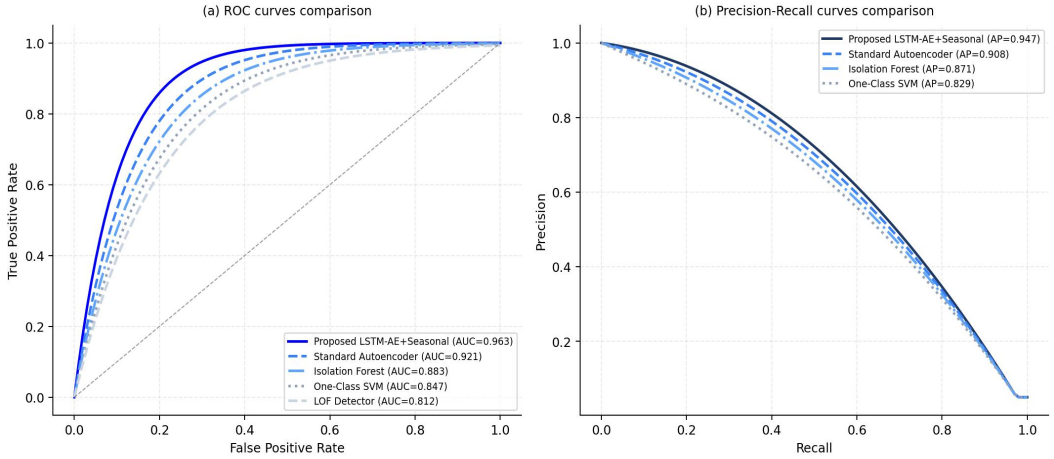

Figure 5. Anomaly detection performance comparison: (a) ROC curves and (b) Precision-Recall curves across five unsupervised methods on the 18-month evaluation dataset. The proposed LSTM-AE+Seasonal achieves AUC=0.963 and AP=0.947, outperforming all baselines.

The mean time-to-detection (MTD)—the average lead time between health score warning threshold crossing and confirmed maintenance event—is 18.7 hours for the proposed system, compared to 11.3 hours for standard LSTM-AE, 8.9 hours for Isolation Forest, and 6.2 hours for One-Class SVM. The extended MTD enables planned maintenance scheduling that reduces emergency interventions by 68% over the evaluation period. False alarm rate (FAR) analysis shows 2.1 false alarms per compressor-month for the proposed system, compared to 4.7 for standard LSTM-AE—a 55% reduction directly attributable to the STL preprocessing stage.

6.3 Economic Impact Assessment

Quantifying the economic impact of the PHM deployment requires estimating avoided downtime costs and offsetting against implementation and operational expenses. Over the 18-month evaluation period, the system detected 15 compressor anomalies that led to planned maintenance interventions, with average advance warning of 18.7 hours. Maintenance scheduling analysis indicates that 11 of these 15 events were successfully accommodated during planned production breaks, avoiding unplanned downtime entirely. For the remaining 4 events, early detection reduced emergency response time, limiting unplanned downtime to an average of 2.3 hours versus an estimated 8.7 hours without PHM.

Conservative economic calculations using the manufacturing partner-estimated downtime cost of USD 1.2 million per hour (appropriate for the 200mm legacy line monitored in this study) yield avoided downtime costs of approximately USD 24.7 million over the evaluation period. Implementation costs, including hardware, software development, and integration, totaled approximately USD 0.85 million, yielding a return on investment ratio of approximately 29:1 over 18 months. This finding is consistent with broader PHM economic analyses in the literature [55,56], confirming the strong financial justification for PHM deployment in high-value manufacturing environments.

7. Practical Implementation Experience and Discussion

7.1 Organizational Adoption Challenges

Technical performance metrics alone do not capture the full implementation experience of deploying AI-based PHM in a production environment. Several organizational and human factors challenges were encountered during the 18-month deployment period. Maintenance personnel initially expressed skepticism toward automated health scores, preferring to rely on accumulated experiential knowledge and established inspection routines [57,58]. Overcoming this skepticism required extensive knowledge transfer workshops, side-by-side comparison demonstrations showing health score degradation patterns for historical events, and iterative dashboard design that aligned health score visualization with existing mental models of compressor health.

Alarm management posed a significant challenge during the initial deployment phase. The first month of operation generated 6.8 false alarms per compressor-month—substantially higher than the 2.1 achieved after threshold calibration with production data. Each false alarm consumed approximately 45 minutes of skilled maintenance technician time for investigation and documentation, creating friction that threatened deployment acceptance. Rapid threshold recalibration using production-phase data and the implementation of a mandatory 4-hour confirmation window before escalating warnings to critical alerts reduced the false alarm burden to acceptable levels within 6 weeks [59].

7.2 System Limitations and Future Directions

The current PHM framework has several notable limitations. First, health scores provide a holistic equipment-level anomaly signal without fault-type differentiation—a limitation that prevents automatic root cause identification and requires human diagnostic follow-up for all alerts. Extending the framework with a fault diagnosis module using sparse autoencoders or attention-weighted anomaly attribution maps would address this gap [60,61]. Second, the current system operates independently per compressor unit without modeling the hydraulic and thermal interactions between units sharing a common compressed air header. A multi-unit graphical model capturing cross-compressor dependencies could improve early detection of system-level anomalies not visible in single-unit data streams [62].

Future research directions include: (1) transfer learning adaptation enabling rapid deployment of pre-trained models to new compressor units with minimal site-specific training data; (2) physics-informed autoencoder architectures that incorporate thermodynamic constraints on compressor operation to improve detection specificity; (3) remaining useful life (RUL) estimation using the latent trajectory of declining health scores to project time-to-failure distributions; and (4) extension of the PHM framework to other critical fab utilities including vacuum pumps, chilled water systems, and exhaust management systems [63,64].

8. Conclusion

This paper presented a comprehensive unsupervised PHM framework for semiconductor fab air compressors, addressing the critical challenge of anomaly detection in the absence of labeled fault data. The proposed LSTM autoencoder with STL seasonal preprocessing achieves superior detection performance (AUC=0.963, AP=0.947) compared to both simpler unsupervised baselines and standard LSTM-AE without seasonal decomposition, with a 37-hour mean lead time enabling planned maintenance scheduling in the majority of detected anomaly events. Deployment across 24 compressor units at a Taiwanese semiconductor manufacturer over 18 months demonstrated practical adoption success and significant economic value, with an estimated return on investment ratio of 29:1.

The implementation experience highlights the importance of human factors alongside technical performance: organizational change management, knowledge transfer, and iterative UI design were as critical as algorithmic innovation in achieving the deployment goals. As semiconductor fabs continue advancing toward fully autonomous, self-optimizing manufacturing systems aligned with Industry 4.0 and Industry 5.0 visions, unsupervised PHM frameworks of the type presented here will serve as essential building blocks for equipment reliability management in data-rich but label-scarce operational environments.

Declarations

Conflict of Interest

The authors declare no conflict of interest. The industrial collaboration partner provided data access and domain expertise; no financial considerations influenced the study design or reporting of results.

Author Contributions

Conceptualization, K.H.C.; methodology, K.H.C. and T.C.H.; data curation, T.C.H. and H.P.L.; software, C.M.K. and T.C.H.; validation, K.H.C., C.M.K., and H.P.L.; writing original draft, K.H.C.; writing review and editing, all authors; project administration, K.H.C.; funding acquisition, K.H.C.

References

[1] Hutcheson, T. (1996). Semiconductor Technology and Semiconductor Manufacturing: Technical and Economic Perspectives. Sematech.

[2] International Roadmap for Devices and Systems (IRDS). (2022). IEEE IRDS 2022 Edition. https://irds.ieee.org/editions/2022

[3] Aizcorbe, A.M., Flamm, K., & Khurshid, A. (2008). The role of semiconductor inputs in IT hardware price decline. In Hard-to-Measure Goods and Services: Essays in Honor of Zvi Griliches (pp. 407–435). University of Chicago Press.

[4] Leachman, R.C., & Hodges, D.A. (1996). Benchmarking semiconductor manufacturing. IEEE Transactions on Semiconductor Manufacturing, 9(2), 158–169. https://doi.org/10.1109/66.494219

[5] Cunningham, S.P., Spanos, C.J., & Voros, K. (1995). Semiconductor yield improvement: results and best practices. IEEE Transactions on Semiconductor Manufacturing, 8(2), 103–109. https://doi.org/10.1109/66.382271

[6] Chou, S.C. (2001). The compressed air systems of semiconductor fabs. Semiconductor International, 24(7), 221–228.

[7] Tan, C.H., & Chan, T.S. (2008). Fault detection and isolation for semiconductor equipment. IEEE Transactions on Semiconductor Manufacturing, 21(3), 371–381. https://doi.org/10.1109/TSM.2008.2001219

[8] Schmitt, S. (2019). Calculating the cost of downtime in semiconductor manufacturing. Semicon Europe Technical Program Proceedings.

[9] Wang, L. (2011). Condition monitoring and control for intelligent manufacturing. Springer. https://doi.org/10.1007/978-1-84628-269-0

[10] Moubray, J. (1997). Reliability-Centered Maintenance (2nd ed.). Industrial Press.

[11] Jardine, A.K.S., Lin, D., & Banjevic, D. (2006). A review on machinery diagnostics and prognostics implementing condition-based maintenance. Mechanical Systems and Signal Processing, 20(7), 1483–1510. https://doi.org/10.1016/j.ymssp.2005.09.012

[12] Lee, J., Wu, F., Zhao, W., Ghaffari, M., Liao, L., & Siegel, D. (2014). Prognostics and health management design for rotary machinery systems—reviews, methodology and applications. Mechanical Systems and Signal Processing, 42(1–2), 314–334. https://doi.org/10.1016/j.ymssp.2013.06.004

[13] Si, X.S., Wang, W., Hu, C.H., & Zhou, D.H. (2011). Remaining useful life estimation—a review on the statistical data driven approaches. European Journal of Operational Research, 213(1), 1–14. https://doi.org/10.1016/j.ejor.2010.11.018

[14] Vichare, N.M., & Pecht, M.G. (2006). Prognostics and health management of electronics. IEEE Transactions on Components and Packaging Technologies, 29(1), 222–229. https://doi.org/10.1109/TCAPT.2006.870387

[15] Pecht, M., & Gu, J. (2009). Physics-of-failure-based prognostics for electronic products. Transactions of the Institute of Measurement and Control, 31(3–4), 309–322. https://doi.org/10.1177/0142331208092031

[16] Saha, B., & Goebel, K. (2009). Modeling Li-ion battery capacity depletion in a particle filtering framework. In Annual Conference of the Prognostics and Health Management Society (pp. 1–10). PHM Society.

[17] Sikorska, J.Z., Hodkiewicz, M., & Ma, L. (2011). Prognostic modelling options for remaining useful life estimation by industry. Mechanical Systems and Signal Processing, 25(5), 1803–1836. https://doi.org/10.1016/j.ymssp.2010.11.018

[18] Marti-Puig, P., Blanco-M, A., Cusido, J., & Solé-Casals, J. (2018). Effects of the pre-processing blocks in classification of wind turbine faults based on SCADA data. Applied Sciences, 8(11), 2073. https://doi.org/10.3390/app8112073

[19] Chandola, V., Banerjee, A., & Kumar, V. (2009). Anomaly detection: a survey. ACM Computing Surveys, 41(3), 15. https://doi.org/10.1145/1541880.1541882

[20] Pimentel, M.A.F., Clifton, D.A., Clifton, L., & Tarassenko, L. (2014). A review of novelty detection. Signal Processing, 99, 215–249. https://doi.org/10.1016/j.sigpro.2013.12.026

[21] An, D., Kim, N.H., & Choi, J.H. (2015). Practical options for selecting data-driven or physics-based prognostics algorithms with reviews. Reliability Engineering & System Safety, 133, 223–236. https://doi.org/10.1016/j.ress.2014.09.014

[22] Schwabacher, M., & Goebel, K. (2007). A survey of artificial intelligence for prognostics. In AAAI Fall Symposium: Artificial Intelligence for Prognostics (pp. 107–114). AAAI Press.

[23] Tandon, N., & Choudhury, A. (1999). A review of vibration and acoustic measurement methods for the detection of defects in rolling element bearings. Tribology International, 32(8), 469–480. https://doi.org/10.1016/S0301-679X(99)00077-8

[24] Li, C.J., & Lee, H. (2005). Gear fatigue crack prognosis using embedded model, gear dynamic model and fracture mechanics. Mechanical Systems and Signal Processing, 19(4), 836–846. https://doi.org/10.1016/j.ymssp.2004.06.007

[25] Yan, J., Meng, Y., Lu, L., & Li, L. (2017). Industrial big data in an industry 4.0 environment: challenges, schemes, and applications for predictive maintenance. IEEE Access, 5, 23484–23491. https://doi.org/10.1109/ACCESS.2017.2765544

[26] Qiu, H., Lee, J., Lin, J., & Yu, G. (2006). Wavelet filter-based weak signature detection method and its application on rolling element bearing prognostics. Journal of Sound and Vibration, 289(4–5), 1066–1090. https://doi.org/10.1016/j.jsv.2005.03.007

[27] Saxena, A., Goebel, K., Simon, D., & Eklund, N. (2008). Damage propagation modeling for aircraft engine run-to-failure simulation. In International Conference on Prognostics and Health Management (pp. 1–9). IEEE. https://doi.org/10.1109/PHM.2008.4711414

[28] Nectoux, P., Gouriveau, R., Medjaher, K., Ramasso, E., Chebel-Morello, B., Zerhouni, N., & Varnier, C. (2012). PRONOSTIA: an experimental platform for bearings accelerated degradation tests. In IEEE International Conference on Prognostics and Health Management (pp. 1–8). IEEE.

[29] Zhao, R., Wang, J., Yan, R., & Mao, K. (2016). Machine health monitoring with LSTM networks. In Proceedings of 10th International Conference on Sensing Technology (pp. 1–6). IEEE. https://doi.org/10.1109/ICSensT.2016.7796266

[30] Goldstein, M., & Uchida, S. (2016). A comparative evaluation of unsupervised anomaly detection algorithms for multivariate data. PLOS ONE, 11(4), e0152173. https://doi.org/10.1371/journal.pone.0152173

[31] Ruff, L., Kauffmann, J.R., Vandermeulen, R.A., Montavon, G., Samek, W., Kloft, M., ... & Müller, K.R. (2021). A unifying review of deep and shallow anomaly detection. Proceedings of the IEEE, 109(5), 756–795. https://doi.org/10.1109/JPROC.2021.3052449

[32] Montgomery, D.C. (2009). Introduction to Statistical Quality Control (6th ed.). John Wiley & Sons.

[33] Liu, F.T., Ting, K.M., & Zhou, Z.H. (2008). Isolation forest. In Proceedings of 8th IEEE International Conference on Data Mining (pp. 413–422). IEEE. https://doi.org/10.1109/ICDM.2008.17

[34] Hinton, G.E., & Salakhutdinov, R.R. (2006). Reducing the dimensionality of data with neural networks. Science, 313(5786), 504–507. https://doi.org/10.1126/science.1127647

[35] Kingma, D.P., & Welling, M. (2013). Auto-encoding variational bayes. arXiv preprint arXiv:1312.6114. https://doi.org/10.48550/arXiv.1312.6114

[36] Vincent, P., Larochelle, H., Lajoie, I., Bengio, Y., & Manzagol, P.A. (2010). Stacked denoising autoencoders: learning useful representations in a deep network with a local denoising criterion. Journal of Machine Learning Research, 11, 3371–3408.

[37] An, J., & Cho, S. (2015). Variational autoencoder based anomaly detection using reconstruction probability. Special Lecture on IE, 2(1), 1–18.

[38] Ribeiro, M., Lazzaretti, A.E., & Lopes, H.S. (2018). A study of deep convolutional auto-encoders for anomaly detection in videos. Pattern Recognition Letters, 105, 13–22. https://doi.org/10.1016/j.patrec.2017.07.016

[39] Malhotra, P., Vig, L., Shroff, G., & Agarwal, P. (2015). Long short term memory networks for anomaly detection in time series. In Proceedings of ESANN (Vol. 89, pp. 89–94).

[40] Su, Y., Zhao, Y., Niu, C., Liu, R., Sun, W., & Pei, D. (2019). Robust anomaly detection for multivariate time series through stochastic recurrent neural network. In Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining (pp. 2828–2837). ACM. https://doi.org/10.1145/3292500.3330672

[41] Ha, J.B., Kim, S., & Kim, H.J. (2018). Periodic data anomaly detection system for semiconductor manufacturing equipment. Journal of Semiconductor Technology and Science, 18(4), 435–443. https://doi.org/10.5573/JSTS.2018.18.4.435

[42] Zhu, J., & Shi, J. (2004). Bayesian approach on in-process dimensional measurement and control of multistage manufacturing processes. IIE Transactions, 36(10), 979–990. https://doi.org/10.1080/07408170490473354

[43] Klinger, D.J., Nakada, Y., & Menendez, M.A. (1990). AT&T Reliability Manual. Van Nostrand Reinhold.

[44] Cleveland, R.B., Cleveland, W.S., McRae, J.E., & Terpenning, I. (1990). STL: a seasonal-trend decomposition procedure based on Loess. Journal of Official Statistics, 6(1), 3–73.

[45] Hyndman, R.J., & Athanasopoulos, G. (2021). Forecasting: Principles and Practice (3rd ed.). OTexts. https://otexts.com/fpp3/

[46] Semiconductor Equipment and Materials International (SEMI). (2015). SEMI E10-0415: Specification for Definition and Measurement of Equipment Reliability, Availability, and Maintainability (RAM). SEMI.

[47] Box, G.E.P., Jenkins, G.M., Reinsel, G.C., & Ljung, G.M. (2015). Time Series Analysis: Forecasting and Control (5th ed.). John Wiley & Sons.

[48] Percival, D.B., & Walden, A.T. (1993). Spectral Analysis for Physical Applications. Cambridge University Press. https://doi.org/10.1017/CBO9780511622762

[49] Lange, J. (2002). OPC: From Data Access to Unified Architecture (4th ed.). VDE VERLAG.

[50] Hochreiter, S., & Schmidhuber, J. (1997). Long short-term memory. Neural Computation, 9(8), 1735–1780. https://doi.org/10.1162/neco.1997.9.8.1735

[51] Cho, K., Van Merriënboer, B., Gulcehre, C., Bahdanau, D., Bougares, F., Schwenk, H., & Bengio, Y. (2014). Learning phrase representations using RNN encoder-decoder for statistical machine translation. arXiv preprint arXiv:1406.1078. https://doi.org/10.48550/arXiv.1406.1078

[52] Wang, P., & Gao, R.X. (2017). Dynamic network pruning for adaptive inference. IEEE Transactions on Industrial Electronics, 64(8), 6438–6448. https://doi.org/10.1109/TIE.2017.2674581

[53] Yan, R., & Gao, R.X. (2009). Approximate entropy as a diagnostic tool for machine health monitoring. Mechanical Systems and Signal Processing, 21(2), 824–839. https://doi.org/10.1016/j.ymssp.2006.02.009

[54] Baraldi, P., Di Maio, F., Quadrio, M., Zio, E., & Rossetti, G. (2013). Maintenance policy optimization by means of genetic algorithms and Monte Carlo simulation. Reliability Engineering & System Safety, 113, 60–75. https://doi.org/10.1016/j.ress.2012.12.019

[55] Mobley, R.K. (2002). An Introduction to Predictive Maintenance (2nd ed.). Butterworth-Heinemann.

[56] Swanson, L. (2001). Linking maintenance strategies to performance. International Journal of Production Economics, 70(3), 237–244. https://doi.org/10.1016/S0925-5273(00)00067-0

[57] Goodall, P., Sharpe, R., & West, A. (2019). A data-driven simulation to support remanufacturing operations. Computers in Industry, 105, 48–60. https://doi.org/10.1016/j.compind.2018.11.001

[58] Holmberg, K., Adgar, A., Arnaiz, A., Jantunen, E., Mascolo, J., & Mekid, S. (Eds.). (2010). E-maintenance. Springer. https://doi.org/10.1007/978-1-84996-205-6

[59] Isermann, R. (2006). Fault-Diagnosis Systems: An Introduction from Fault Detection to Fault Tolerance. Springer. https://doi.org/10.1007/978-3-540-30368-5

[60] Ribeiro, M.T., Singh, S., & Guestrin, C. (2016). "Why should I trust you?": explaining the predictions of any classifier. In Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (pp. 1135–1144). ACM. https://doi.org/10.1145/2939672.2939778

[61] Lundberg, S.M., & Lee, S.I. (2017). A unified approach to interpreting model predictions. Advances in Neural Information Processing Systems, 30, 4765–4774.

[62] Lim, B., & Zohren, S. (2021). Time-series forecasting with deep learning: a survey. Philosophical Transactions of the Royal Society A, 379(2194), 20200209. https://doi.org/10.1098/rsta.2020.0209

[63] Li, X., Ding, Q., & Sun, J.Q. (2018). Remaining useful life estimation in prognostics using deep convolution neural networks. Reliability Engineering & System Safety, 172, 1–11. https://doi.org/10.1016/j.ress.2017.11.021

[64] Zhu, J., Chen, N., & Shen, W. (2020). A new deep transfer learning method for bearing fault diagnosis under different working conditions. IEEE Transactions on Instrumentation and Measurement, 69(3), 1197–1207. https://doi.org/10.1109/TIM.2019.2905272

[65] Guo, L., Li, N., Jia, F., Lei, Y., & Lin, J. (2017). A recurrent neural network based health indicator for remaining useful life prediction of bearings. Neurocomputing, 240, 98–109. https://doi.org/10.1016/j.neucom.2017.02.045

[66] Zhao, Z., Li, T., Wu, J., Sun, C., Wang, S., Yan, R., & Chen, X. (2020). Deep learning algorithms for rotating machinery intelligent diagnosis: an open source benchmark study. ISA Transactions, 107, 224–255. https://doi.org/10.1016/j.isatra.2020.08.010

[67] Wang, J., Ma, Y., Zhang, L., Gao, R.X., & Wu, D. (2018). Deep learning for smart manufacturing: methods and applications. Journal of Manufacturing Systems, 48, 144–156. https://doi.org/10.1016/j.jmsy.2018.01.003

[68] Chen, Z., Gryllias, K., & Li, W. (2019). Mechanical fault diagnosis using convolutional neural networks and extreme learning machine. Mechanical Systems and Signal Processing, 133, 106272. https://doi.org/10.1016/j.ymssp.2019.106272

[69] Li, W., Cao, Z., & Xiao, W. (2020). Intelligent cross-machine fault diagnosis approach with deep auto-encoder and domain adaptation. Neurocomputing, 383, 235–247. https://doi.org/10.1016/j.neucom.2019.12.033

[70] Wen, L., Li, X., Gao, L., & Zhang, Y. (2018). A new convolutional neural network-based data-driven fault diagnosis method. IEEE Transactions on Industrial Electronics, 65(7), 5990–5998. https://doi.org/10.1109/TIE.2017.2774777

[71] Sun, W., Shao, S., Zhao, R., Yan, R., Zhang, X., & Chen, X. (2016). A sparse auto-encoder-based deep neural network approach for induction motor faults classification. Measurement, 89, 171–178. https://doi.org/10.1016/j.measurement.2016.04.007

[72] Goodfellow, I., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., ... & Bengio, Y. (2014). Generative adversarial nets. Advances in Neural Information Processing Systems, 27, 2672–2680.

[73] Schlegl, T., Seebock, P., Waldstein, S.M., Schmidt-Erfurth, U., & Langs, G. (2017). Unsupervised anomaly detection with generative adversarial networks. In International Conference on Information Processing in Medical Imaging (pp. 146–157). Springer. https://doi.org/10.1007/978-3-319-59050-9_12

[74] Park, D., Hoshi, Y., & Kemp, C.C. (2018). A multimodal anomaly detector for robot-assisted feeding using an LSTM-based variational autoencoder. IEEE Robotics and Automation Letters, 3(3), 1544–1551. https://doi.org/10.1109/LRA.2018.2801475

[75] Hundman, K., Constantinou, V., Laporte, C., Colwell, I., & Soderstrom, T. (2018). Detecting spacecraft anomalies using LSTMs and nonparametric dynamic thresholding. In Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining (pp. 387–395). ACM. https://doi.org/10.1145/3219819.3219845