IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

ISO 3297:2007 Certified ⁜ Impact Factor 7.918 ⁜ Vol. 11, Issue 12, December 2022

DOI: 10.17148/IJARCCE.2022.111261

Predictive Maintenance Models for Smart Manufacturing Systems

Madhu Sathiri

Independent Researcher, India

Abstract: Predictive Maintenance (PM) refers to the utilization of various forms of data for the timely anticipation of system failures. The objective is to schedule maintenance, instead of performing it according to a fixed pattern or after failure, thus maximizing uptime and resource efficiency. Modern PM systems are becoming popular within smart manufacturing as they contribute to automation and operational efficiency improvement, particularly when deployed near the equipment that generates and suffers the data. These models harness diverse forms of sensor data generated by the manufacturing process, and include techniques such as survival, reliability and hazard function estimation; time series analysis; and statistical, machine learning and deep learning methodologies. Such techniques are able to recognise and model 'normal' machine behaviour when sensors are not broken, and thus may fail to anticipate faults that lead to abnormal physical behaviours detectable by sensors. Such issues are of growing concern within smart factories, where maintenance modelling needs to remain accurate even when the operating environment or underlying machine behaviour changes. PM model deployment and operationalisation can thus be challenging, and requires effective instrumentation, data engineering and model management around these techniques, especially when real-time, low-latency inference at the edge is necessary.

Despite the challenges, there is considerable ongoing research work applying predictive maintenance solutions in production environments. Promising demonstrations have been reported across multiple domains, including automotive, semi-conductor, mining, electronics, and food processing. Use cases span A- and B- lines of automotive assembly, anticipation of failures in cooling units of A-lines, cooling balance modelling for wafer manufacturing and plasma etching, board cleaning process deviation alerting, run-to-failure estimation in semiconductor and electronic assembly, electric drive system on-condition servicing planning, web break prediction for textile manufacturing, and applying predictive and prescriptive analytics to food processing.

Keywords: Predictive maintenance,Smart manufacturing,Industrial Internet of Things (IIoT),Machine learning models,Condition-based monitoring,Equipment failure prediction,Time-series analysis,Sensor data analytics,Remaining useful life (RUL),Anomaly detection,Digital twin technology,Edge computing,Big data analytics,Fault diagnosis,Industry 4.0.

1. INTRODUCTION

Manufacturing strives to establish a truly integrated, intelligent manufacturing process. A smart factory is frequently compared to a human cell in which components are grouped, managed, from sensors, data, assets, business logic, and processes. Models based on data must connect real-time data in a physical system and cause mode data, e.g sensor data, in order to be useful for predictive maintenance.Predictive maintenance is considered a vital element for AI based smart factories and manufacturing process. Solutions and models for predictive maintenance and industrial development have been developed, for example in the automotive, semiconductor, electronics and food industry. Such models require their own sub-development.

The aim of predictive maintenance projects is to deliver trained and deployed working models, that detect when an asset is going to fail or degrade outside pre-determined thresholds within an acceptable time window. In contrast to simulation models, predictive maintenance models must reflect the actual status of assets over real time. A distinction can be made between three major time segments of these predictive maintenance projects: pre-processing of data, developing and training the predictive maintenance model, deploying and operationalising the predictive maintenance model. Data is at the core of predictive maintenance, and a lot of attention is devoted to data acquisition and pre-processing. The length and quality of the sensor data drives the robustness of the developed predictions. Multi-modal sensor data have to be combined and cleaned before they can be used for ML or deep learning predictions.

1.1. Background and Significance

Industry 4.0 refers to the smart digitization and integration of value-added processes within corporations, among stakeholders and supply-chain partners, and in collaborations with external ecosystems comprising on-demand service providers, such as data-hubfulfilling AI-based digital twins, analytic suppliers, and application developers. Smart manufacturing, a key action area of Industry 4.0, establishes fully and partly autonomous production systems capable of operating with little or no human intervention. The trusted, transparent, and reliable interplay of man and machine is supported by cyber-physical-social systems with embedded AI capabilities.

Routine machine and equipment maintenance has shifted to predictive maintenance—a timely prediction of the remaining useful life (RUL) of components and devices based on operational data. With increased frequency and scale of edge data being generated by sensor-equipped smart factories, predictive maintenance models are being deployed in manufacturing environments with real-time inference and monitoring of equipment reliability and performance. External factors, such as seasonal operating patterns, supply chain constraints, and economic cycles, further influence equipment reliability and performance. Therefore, data-generation and operational-use cycles are closely coupled in smart factories, necessitating the continual updating and real-time maintenance of predictive maintenance models.

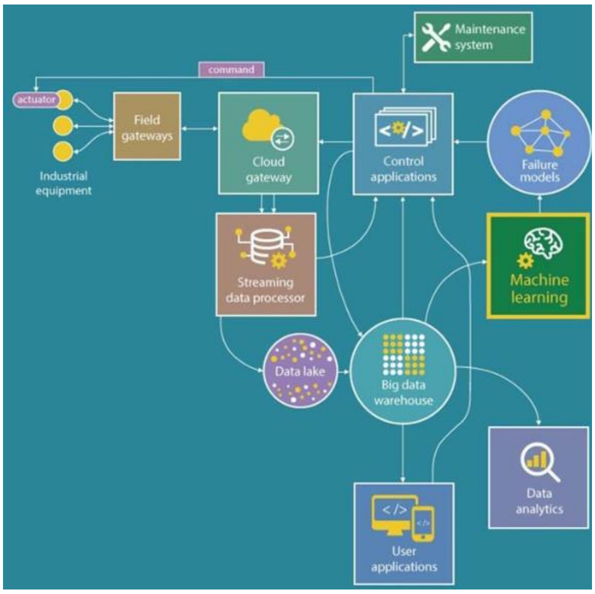


Fig 1: Predictive maintenance architecture in a smart manufacturing

2. CORE CONCEPTS IN PREDICTIVE MAINTENANCE

The operationalization of predictive maintenance in smart manufacturing environments requires supporting definitions, identification of data sources, feature engineering, evaluation metrics, and algorithms.

Predictive maintenance minimizes the effects of abrupt machine failures by detecting their precursory signs and hence by permitting timely corrective actions. Sufficiently advanced predictive maintenance algorithms are able to infer the cause and remaining useful lifetime of a malfunctioning machine component. Their deployment within a production environment thus moves maintenance away from the traditionally applied tBO strategy (i.e., where preventive interventions are carried out after a prespecified duration of tBO time) towards a tested and optimized tTd intervention strategy, which stipulates that repairs are only carried out upon detection of incipient failure. Such algorithms need to be embedded within a larger system—split into at least five blocks—to realize their predictive functionality in real time.

The first two critical blocks—data acquisition and cleaning, and model selection and training—provide the wealth of relevant data from which a machine learning algorithm has to identify precursory and predictive patterns. These two blocks are the theme of the next two sections. The third block—the algorithm evaluation metrics—examines prediction quality through multiple dimensions. As part of the broader predictive maintenance implementation pipeline, the fourth block addresses information extraction from newly-built or refurbished models so that predictively useful disease signatures can be identified and delivered to users.

Equation 1: Classification metrics (Failure prediction)

Start with the **confusion matrix counts**:

- **TP**: predicted failure and actually failure
- **TN**: predicted normal and actually normal
- **FP**: predicted failure but actually normal

DOI: 10.17148/IJARCCE.2022.111261

- **FN**: predicted normal but actually failure

Let total samples be:

$$N = TP + TN + FP + FN$$

Accuracy

$$\text{Accuracy} = \frac{\#\text{ correct}}{\#\text{ total}} = \frac{TP + TN}{N}$$

Precision (how many predicted failures were truly failures)

$$\text{Precision} = \frac{TP}{TP + FP}$$

Recall / TPR (how many actual failures were detected)

$$\text{Recall} = \frac{TP}{TP + FN}$$

F1-score (harmonic mean of precision and recall)

$$F_1 = \frac{2}{\frac{1}{\text{Precision}} + \frac{1}{\text{Recall}}} = \frac{2 \cdot \text{Precision} \cdot \text{Recall}}{\text{Precision} + \text{Recall}}$$

FPR (false positive rate, needed for ROC)

$$\text{FPR} = \frac{FP}{FP + TN}$$

2.1. Definitions and Objectives

Models of predictive maintenance in smart manufacturing systems are anticipated to enhance production performance, equipment reliability, and product quality while minimizing maintenance costs. Such models aim to continually predict future machinery states through fault diagnosis and prognosis, supervised by labelled datasets that contain timestamps of previous failures. Machines of varying types typically possess different failure mechanisms, and transferring learnings across different types or brands may not be practical. Nevertheless, the underlying physics of failure mechanisms remains similar, and different processes or types of machines may benefit from sharing spatial or temporal information.

Conditions that cause failures are user-controllable, whereas features that lead to breakdown are not. Feature engineering, crucial to predictive maintenance models, often relies mostly on specific expertise, though some automated approaches are being developed. In smart manufacturing systems, sensors record an abundance of data from various sources. Components and machines alike are equipped with built-in sensors that are integrated into the same Industrial Internet of Things (IIoT) network. Nevertheless, data quality may not always satisfy modelling requirements, necessitating further data cleaning and preprocessing, and data from different sources must be synchronized.

2.2. Data sources and Feature Engineering

A wide choice of data sources and of features derived from them, including data from non-obvious sources and sensor-based Condition Monitoring (CM), are core to predictive maintenance. Data sources include those used for Failure Mode and Effects Analysis (FMEA), System Monitoring (SM), Diagnostics (D), Prognostics (P) and Cognitive systems. Feature Engineering (FE) impacts predictive maintenance profoundly.

Data sources commonly utilized for predictive maintenance comprise sensor data (and sensor-based CM), historical maintenance logs, preservation and lifecycle management tools, invoices for consumables and spare parts, logistics support services for spares, materials trapped in production, machine-specific resource extenders, test data at both time and product levels, product and resource-level specifications, floor management systems, cyber security systems, failure mode and effects analysis (FMEA) analyses, systems monitoring and diagnostics tools, machine deep-life-methods (acting as physical prognostics), data-words historical manuals and trial data from AI methods unsupported/untested by either physics or numerical measures. Used in isolation or in combination, features and classifiers from these data sources are trained using sensor-based CM, as these features are continuously updated on a timely basis. Predictive maintenance evaluates one or more product quality attributes with respect to quality specifications; availability with respect to service delivery and product requirements (at time level); health (an aggregated score of quality, availability and reliability) with respect to deep-life-method specifications; and cybersecurity compliance and security levels with respect to quality requirements.

2.3. Evaluation Metrics

Diverse metrics are employed at multiple levels to evaluate the usefulness of predictive maintenance methods in smart factories. First, at the model evaluation level, a combination of classification performance metrics (accuracy, precision, recall, F1-score, AUC) is related to assessment of the posterior maintenance scheduling. In particular, a late cautionary warning is generally preferred, since acute machinery failure or breakdown may have dire consequences. Second, at the operational level, the advantages are assessed by reduced Uptime and Cost indices. The Uptime index is defined in terms of Scheduled Maintenance Time, Unscheduled Maintenance Time, and Total Production Time, while the Cost index combines Production Cost, Maintenance Cost, and the Penalty Cost arising from missing customer demands. Third, at the business level, Return on Investment (ROI) is evaluated by considering the cost of generating the predictive maintenance service versus its positive impact on the business. Fourth, the effectiveness of predictive maintenance is assessed in terms of process yield and product quality parameters that are directly or indirectly linked to machine health. Finally, real-time prediction latency is also important for a predictive maintenance solution during its operation.

3. DATA ACQUISITION AND PREPROCESSING FOR SMART FACTORIES

The successful implementation of predictive maintenance (PdM) models depends largely on the quality of the underlying data. For real-world applications, this is often an arduous task because data are sourced from multiple devices that employ different protocols and have different manufacturers. Thus, seamless integration of data from these disparate sources is essential. Quality checks must be performed to eliminate corrupted data and to ensure that sensor readings are trustworthy before they are input to the predictive models. If manufactured components or subsystems have known failure modes or mechanisms, physical failure time indicators that accurately represent component or subsystem health status can be derived instead of applying complex machine-learning models to label the predictive data and reduce the amount of error flow into the final models.

Recent advances in industrial testing technologies and decentralization model training schemes, such as federated learning, reinforce the belief that real-time predictive maintenance leading to zero-downtime production systems is achievable. The model training, testing, validation, and updating phases of predictive maintenance development require accurate timestamping of data from all integrated sensors. Advanced timestamp-synchronization methods, based on cross-correlation of mutually influencing observable time series affected by common hidden variables, offer superior quality over conventional clock drift compensation schemes. The causal relationships among the data help the user identify the primary precursor data streams whose lead and lag relationships are supported by pre-existing knowledge of the affected systems. High-quality data-mining process methods such as these create PdM testing and validation outputs that inspire the production of real-time inference models at much reduced costs.

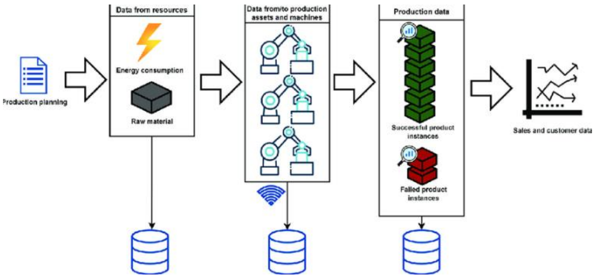


Fig 2: Data acquisition from different assets in a smart factory

3.1. Sensor Integration and Data Quality

The predictive maintenance strategy relies on modeling the risk of failure with sufficient lead time for corrective actions to be initiated. Data are the key enabler for realizing such modelling at scale and, hence, the underlying assumption is that modern smart factories are instrumented with a large number of heterogeneous sensors that capture and transcribe every aspect of their operations in real-time. The integration of sensor data for predictive maintenance is particularly advantageous in three different but interrelated aspects: (1) It allows capturing the system behaviour in a holistic, causal view by taking advantage of the natural structure of the data; (2) It enables utilizing a more extensive set of input variables that are potentially correlated to failure occurrences, thus improving prediction performance; (3) The modelling workload can be divided into sub-models, thus simplifying maintenance and improving inference performance.

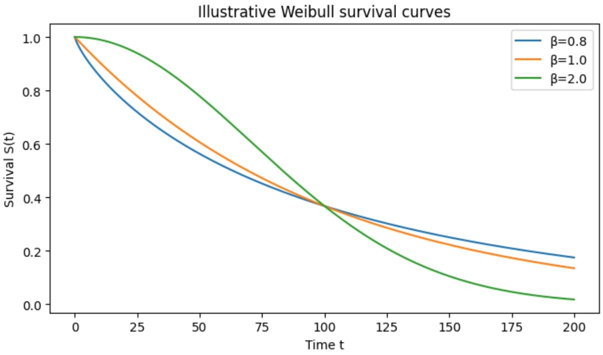
The actual value of predictive maintenance models is critically defined by their robustness and prediction quality. The introduction of sensors supporting a wide range of failure modes or the application of feature-selection techniques based on causal knowledge can contribute to more reliable solutions. Conversely, high-quality, available, and consistent historical records for all input processes are necessary to improve the supervised models' predictive power. Records are rarely complete, however, as data produced by commercial sensors put in place to monitor different processes for other reasons are often used. Quality problems, such as faulty sensors, outliers due to measurement errors, or sudden shifts in data distributions, are common in real-world applications. Consequently, data cleaning and imputation operations are typically required for restoring missing or corrupted samples.

3.2. Data Cleaning and Imputation Techniques

Data from sensors or acquired independently may contain noise, inaccuracies, and redundant information. Correctly processing these features prevents poor initial predictions, costly data drift, and premature failure. Filtering seeks to remove noise or other forms of interference, using algorithms such as moving averages or Kalman filters, with settings optimized through prowling.

New factors may arise that are independent of, but correlated with, sensing variables, such as the temperature in a four-season country without air conditioning. Missing values may occur due to faulty measurements or a data communication break. If the missing data constitutes a small part of the dataset, discarding it may suffice. Otherwise, more sophisticated interpolation, averaging, or specialized machine-learning techniques can estimate these values based on observed relationships. Trending up or down should enhance predictions.

Measurements of the same phenomenon across different sensors or even distinct kinds of sensors may differ for reasons such as calibration errors. At sensor placement, increased redundancy allows relative comparisons to spot spurious outliers, which can then be eliminated. At the modeling stage, sensor aequivalence predicts which sensors can be omitted when data is scarce. For example, if a temperature sensor across a production line fails, a valve position sensor can serve as a proxy for hot or cold conditions, provided the normal functioning of the other sensor is guaranteed.



3.3. Timestamp Synchronization and Causality

In multi-source data in predictive maintenance, timestamp-related issues play a vital role in data quality since they are compared in regard to the temporal dimension. Moreover, ensuring data causality is equally important. Events measured by sensors with different polling frequencies and events generated by automation equipment without time labels need to be synchronized based on a common time index. Stored data may contain features measured in milliseconds, seconds, and minutes, therefore the highest-polling-frequency feature can act as the time index. For data collected by automation with low time resolution, such as start/end of a job, robot pick/place, and robot malfunction, the event is repeated many times at the same time and doesn't need to be synchronized. However, features from real production lines contain independent and almost perfect discriminative labels.

The causality of events needs to be checked when using multi-source data for predictive maintenance. Interval–quality–quantity is a mathematical rule in model training that must be satisfied. Windows during normal and abnormal production should have correct labels. Remaining work in the predictive-maintenance task can be summarized as follows. Based on multi-sensor, multi-source, and multi-feature data, different predictive-maintenance models can be built, deployed, and verified in factory manufacturing environments, such as automotive, semiconductor, electronics, textile, and food processing. Major PdM tasks include failure prediction, remaining-useful-life estimation, maintenance-planning strategy, and remaining-component-life prediction.

DOI: 10.17148/IJARCCE.2022.111261

4. MODELING PARADIGMS IN PREDICTIVE MAINTENANCE

A predictive maintenance (PdM) model may be formulated based on either statistical principles or, more commonly in the Smart Manufacturing System literature, by employing machine learning or deep learning models. While these systems predominantly possess black-box characteristics, user trust may be heightened by integrating a physics-based component, which is often exploited in a hybrid framework.

Statistical Approaches

Statistical approaches may be appropriate when failure events are sparse, and the underlying hazard function is known. In such a case, a K–A model expressed in terms of the opening time h and a set of physical stress characteristics is suitable. Analogously, a reliability-based prediction model leverages real-world sensor data and a Weibull-based failure time distribution to establish a relationship between the lifetime of a product and associated sensor data, thus enabling the anticipation of product failure before it occurs.

Machine Learning and Deep Learning Methods

Given that the traditional approaches are not adaptable to the current trends of Industry 4.0, PdM models mainly use machine learning algorithms such as SVM, decision trees, random forests, and k-nearest neighbors, among others, to predict imminent failure or remaining useful life. Deep learning methods overcome the need for feature engineering in high-dimensional sensor data. For instance, an LSTM-based framework combines Long Short-term Memory (LSTM) and Physically-Based Models (PBMs) to sense putty content in a molding process. Furthermore, deep learning methods can also be employed for anomaly detection. For instance, contrastive learning-driven Deep Metric Learning (DML) framework uses embedded DML features instead of normal features extracted from an autoencoder to predict anomalies in vibration sensor data.

Physics-informed and Hybrid Models

The trustworthiness of predictions diminishes with the increase in the black-box nature of the applied machine learning technique. This largely limits acceptance by users in sensitive applications such as aerospace and healthcare, where even non-expert users should be able to comprehend the prediction process. Consequently, a hybrid framework is often adopted by first building a sparse physics-based model that requires only a few physical inputs.

Equation 2: Operational indices: Uptime and Cost

Let:

- T = Total Production Time
- S = Scheduled Maintenance Time
- U = Unscheduled Maintenance Time

A standard uptime fraction (consistent with the definition) is:

1. **Downtime**:

$$D = S + U$$

2. **Uptime time**:

$$T_{\mathrm{up}} = T - D = T - (S + U)$$

3. **Uptime index** (normalized 0–1):

$$\text{UptimeIndex} = \frac{T_{\mathrm{up}}}{T} = \frac{T - (S + U)}{T} = 1 - \frac{S + U}{T}$$

For costs, define:

DOI: 10.17148/IJARCCE.2022.111261

- C_p = Production cost
- C_m = Maintenance cost
- C_{pen} = Penalty cost (missed demand)

A direct combined cost expression matching the paper's components is:

$$C_{\text{total}} = C_p + C_m + C_{pen}$$

4.1. Statistical Approaches

An early widely applied statistical modeling strategy predicted the time until the subsequent failure of a physical asset by modeling the failure times as random samples from a last distribution. These distributions capture the reliability of physical assets and help estimate the probability of asset failure during the next fixed time period, enabling scheduling of routine maintenance for multiple assets while avoiding unnecessary costs. Practical implementation typically combines the model predictions with sampling from age distributions of the assets. When the prediction model uses Euclidean or generalized distances between monitoring samples of the physical asset and functional failure boundaries, it falls under the category of survival montage methods. The failure-time data can supplement a sensor-data-driven prognosis framework for a specific critical physical asset to mitigate the drawbacks of using only past data.

Machine-learning analogues of probability distribution of time-to-next-failure have also been proposed, with ideas motivated from markov processes. Given the complexity and diversity of smart factories, statistical models, predominantly survival analysis models, usually solve narrow-depth domain-specific PM problems. However, few PM tasks are frequently deployed. These shall eventually serve as templates for similar assets deployed elsewhere. A core proof-of-concept task for a PM solution shall hence be develop. This task and its related models shall guide further deployments and support templates that shall ease and accelerate future TM deployments.

4.2. Machine Learning and Deep Learning Methods

The use of statistical and handcrafted approaches to model predictive maintenance (PM) problems can become intractable in real-world applications due to the high number of features involved. Consequently, machine learning (ML) methods, especially in combination with deep learning (DL), have gained popularity for solving PM problems. Data-driven models are usually quick to build, require reduced domain knowledge on the problem, and can provide highly accurate results. However, they can lack interpretability, generalizability, and resilience to changes in the environment and system. These elements are essential to build trustworthy AI systems. Recent literature leverages these techniques to develop predictive maintenance models in a wide range of manufacturing domains for failure prediction, remaining useful life (RUL) estimation, and time-to-failure (TTF) prognosis. These models can be empirically classified according to the objective and feature set.

Automotive and aircraft systems sequentially integrate mechanical, electrical, and physical subsystems. Within automotive and aerospace assembly lines, predictive maintenance (PM) methods focus on forecasting failures, detecting anomalies, and analyzing remaining useful life (RUL) or time-to-failure (TTF). Structured electronic modules diagnose products at various assembly stages without requiring change points and supplementary sensor data from subsequent assembly stations. For aircraft production lines, an interpretable ML model for RUL prediction based on temporal behavior and features extracted from logistic customer support data has been applied. Elbow and silhouette methods combined with support vector machine classification identify the optimal cluster of sensor readings for TTF prognosis.

4.3. Physics-informed and Hybrid Models

Recent advances towards explainable artificial intelligence have prompted growing attention towards incorporating physical knowledge in predictive models. Physics-informed machine learning and deep learning algorithms have thus benefited from more interpretable results while still enjoying sufficient flexibility to tackle complex, high-dimensional tasks. By imbuing the learning model with information characterizing the data generating processes, such as differential equations, Hermite polynomials or other physics-informed priors, it is possible to improve extrapolation performance even when the training set is insufficiently extensive. For instance, safety-critical applications might still be supported under limited available data by embedding physical constraints that would otherwise jeopardise generalisation.

DOI: 10.17148/IJARCCE.2022.111261

A different approach consists in learning a predictive model for a specific domain amenable to data-rich inferring but insufficiently interpretable for reliable utilization in deployment, and subsequently augmenting explainability via an auxiliary physics consideration. Combining a physics model with learned residuals—hybrid modelling—has achieved substantial success. More specifically, physically based models can significantly reduce the global sample size requirement while still enabling the deployment of interpretability-hardened data-driven components.

5. MODEL DEPLOYMENT AND OPERATIONALIZATION

Model deployment and operationalization of predictive maintenance models encompass aspects such as inference operation, model updating as time elapses, and explanations of their decisions. During inference operation, a trained model is deployed to a suitable device to infer responses over new samples. Edge devices are becoming the norm for inference operation due to the motivation for real-time systems. However, deploying models at the edge is constrained by limited computational resources. Training and testing sets are usually from the same period, but data distributions might change with time, leading to data drift. Without proper handling, deployed models deteriorate, resulting in high business risks. Explanations imparting reasoning behind decisions increases trust towards AI-based systems, especially for critical applications. Therefore, the subcomponents of model deployment and operationalization include real-time inference and edge computing, model updating and lifecycle management, and explainability and trustworthiness.

Real-time inference and edge computing tackle the deployment of predictive maintenance models in real-time manufacturing scenarios. Model inference is executed at the time when new samples are available for prediction. Edge devices with limited computational resources are the deployment choice. Hence, low-cost models are requisite to achieve real-time prediction. Requisite constraints are imposed while developing predictive maintenance models for real-time inference applications. Model updating and lifecycle management address the temporal component of model deployment. Data distributions gradually change over time, resulting in performance drop of deployed models. Data distributions associated with testing and deployed datasets may vary along any given axis. Data streams are defined to reflect the format and organization of data samples acquired over time. Consequently, implementation methodologies enabling efficient model update, such as continuous learning, incremental learning, transfer learning, and active learning, are required.

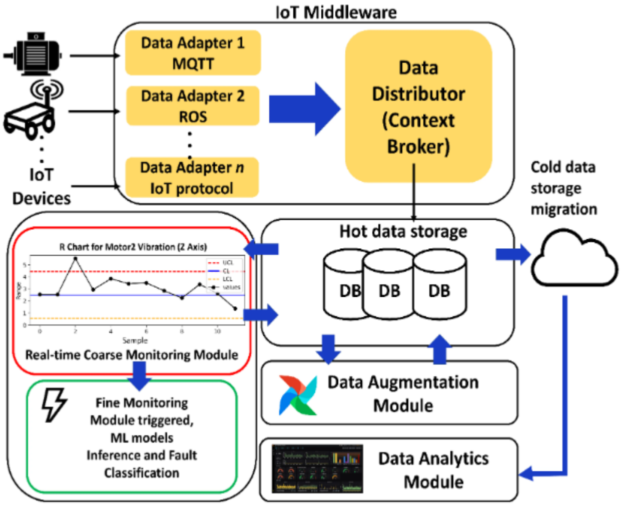

IoT Middleware
Data Adapter 1
MQTT
Data Adapter 2
ROS
Data Adapter n
IoT protocol
Data Distributor (Context Broker)
IoT Devices
Cold data storage migration
Hot data storage
DB
DB
DB
R Chart for Motor2 Vibration (2 Axis)
UCL
CL
LCL
values
Range
Sample
0
1
2
3
4
5
0
2
4
6
8
10
Real-time Coarse Monitoring Module
Fine Monitoring Module triggered, ML models Inference and Fault Classification
Data Augmentation Module
Data Analytics Module

Fig 3: A Predictive Maintenance System

5.1. Real-time Inference and Edge Computing

Models with low inference time may run in real time within the manufacturing plant, and can, for instance, provide short-term wear predictions and recommendations for maintenance work across the whole manufacturing line. Prediction results from the different models may then be collected at a central location and presented to the users in a human-usable way. The main concerns are the speed requirements for the predictions and the number of models that run at the same time and require maintenance work as a consequence. This is particularly relevant for models for which the algo- rithm is not 100% reliable or that provide an estimation of remaining useful life based on incomplete data. When such models are needed, less critical inferences may be relocated to private or public clouds, thus taking advantage of the increased computing power available there.

Some models may also implement a “hot model” strategy, which consists of a temporary duplication of a model and its relocation on an enterprise server that has more computing power. It is mainly used in situations where large amounts of data have been generated, such as simulation data during production downtime, and where these data cannot be processed in the normal time window of the model. The duplication can be implemented if there is the required additional enterprise computing power and if no specific data privacy is at risk.

With the growing adoption of Industrial Internet of Things technologies in the Smart Manufacturing systems, the volume of sensor data has significantly increased. Therefore, moving the inference and analytics of predictive models from the Cloud to the Edge becomes crucial for performance. Edge computing, as an enabling technology, reduces latency and improves response time of data mining and predictive analytics. Despite these advantages, Edge computing raises a number of challenges, such as resource constraints and data privacy. Accordingly, the research community has proposed various Edge-Cloud collaborative frameworks for different types of applications.

5.2. Model Updating and Lifecycle Management

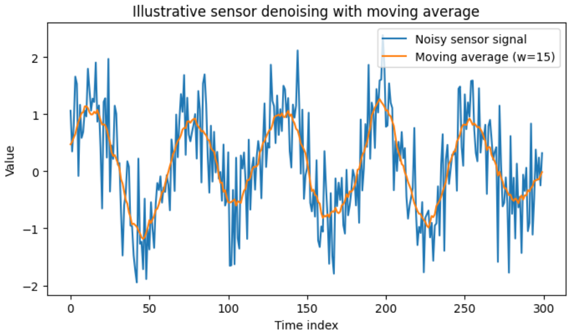
Among the aspects essential to the successful operationalization of predictive maintenance models are their continual updating, monitoring, and maintenance throughout the different stages of their lifecycles. Like all machine-learning models, predictive maintenance models are subject to degradation — which occurs when a model's performance decreases over time — for many reasons. Important among these is a change in the spatiotemporal distribution of the training or test data sets, termed data or covariate drift. Modern manufacturing systems are complex, data-intensive environments where many factors can cause data drift. For predictive maintenance, one of the major sources of risk is the effect of user-defined parameters, such as sensor locations and fault indicators.

Data drift, nevertheless, is not the only factor for which predictive maintenance models must be periodically updated. Over the courses of their lifecycles, such models will also undergo natural performance degradation due to recent changes in the system's behaviour. These aspects require that predictive maintenance models encompass lightweight monitoring systems capable of detecting these two types of decay and that they accommodate model retraining and revalidation. Indeed, the aim of model monitoring is to detect a decrease in model performance so that retraining can occur before the model becomes ineffective.

5.3. Explainability and Trustworthiness

Predictive maintenance problems are often high-stakes applications that directly affect business operations (e.g., profitability, stock prices) and user safety (e.g., aviation, autonomous driving). Given the black-box nature of machine learning and deep learning methods, their inferences must be evaluated critically and not blindly trusted. Furthermore, the perception of machine learning as a “science” may inhibit human curiosity and investigation of its decisions. To this end, explainability (also known as interpretability, transparency, or trustworthiness) methods analyze model behavior toward improved comprehension and transparency, ultimately increasing the trustworthiness of predictions and decision-making processes.

As models become increasingly influential and non-biased decisions are essential, trust has emerged as an active area of research. Particularly, explainability and trust must be discussed in conjunction with security, privacy, and responsibility. Moreover, generative models based on reasoning and knowledge formalized in ontologies or knowledge bases are highly interpretable, whereas knowledge distillation provides explanation capabilities from deep networks. Ai-generated design flows have been proposed to ensure that both model quality and generated designs are trustworthy. Careful design decisions, especially for neural architecture search, can also increase trust.



6. CASE STUDIES ACROSS SMART MANUFACTURING DOMAINS

Evidence of predictive-maintenance capabilities covers various manufacturing domains, including automotive, electronics, semiconductor, textile, and food processing. The examples show how predictive models have been integrated or deployed into real-life smart factories. First, a cloud-based architecture using 5G-enabled edge-cloud systems for the A-line assembly of an automotive manufacturing industry exemplifies the end-to-end predictive-maintenance framework. The industry uses industrial excepted edge 5G and an edge-cloud system while accelerating sensor implementation. Wireless sensors such as vibration sensors attached to the robotic arms generate real-time data for cloud processing and predictive analysis. Second, for a semiconductor manufacturing factory, forecasting model results are evaluated and presented with a parallel-edge-edge-cloud system and dry-etching-maintainable-pilot-production-room-related manufacturing-resource-forecasting-MRF model. Pilot production rooms (PPR) of a semiconductor manufacturing factory provide maintenance services for the entire semiconductor line. For a semiconductor manufacturing factory using a parallel cumulative incidence intensity-forecasting model for factory-pilot-production-room-related-resources, proper maintenance reduces delay time, providing better service and maximizing factory-resources profit.

Third, using two dedicated predictive models, a solution is applied to a textile manufacturing plant producing garment and upholstery textiles with robotic-laser-cut technology. Mistral dyeing machines consume long periods because they run slow to ensure fabric color continuity. The high-energy consumption of Mistral and other machines enables modelling of electricity consumption and water level to optimize power and minimize cost. Last, to predict shutdown when devoting power, inspecting the quality of manufactured foods, and optimizing production orders in real factory-dataset conditions, an explainable ML approach uses the AdaBoost ensemble learning method as a classifier. All cases not only solve production problems but provide crucial business-related decisions, minimizing transitions to transparent operations.

Equation 3: Reliability / Survival / Hazard and Weibull model

Define:

- T = random variable "time-to-failure"
- **CDF**: $F(t) = P(T \leq t)$
- **Survival**: $S(t) = P(T > t) = 1 - F(t)$
- **PDF**: $f(t) = \frac{d}{dt}F(t)$
- **Hazard** (instantaneous failure rate given survival to t):

$$h(t) = \lim_{\Delta t \to 0} \frac{P(t \leq T < t + \Delta t \mid T \geq t)}{\Delta t}$$

Derive hazard formula:

2. Conditional probability:

$$P(t \leq T < t + \Delta t \mid T \geq t) = \frac{P(t \leq T < t + \Delta t)}{P(T \geq t)}$$

3. Substitute:

- $P(T \geq t) = S(t)$
- $P(t \leq T < t + \Delta t) \approx f(t)\Delta t$

So:

$$P(t \leq T < t + \Delta t \mid T \geq t) \approx \frac{f(t)\Delta t}{S(t)}$$

4. Divide by Δt, take limit:

$$h(t) = \frac{f(t)}{S(t)}$$

Weibull distribution with scale η and shape β:

1. CDF:

$$F(t) = 1 - e^{-(t/\eta)^{\beta}}$$

2. Survival:

$$S(t) = 1 - F(t) = e^{-(t/\eta)^{\beta}}$$

3. PDF:
Differentiate $F(t)$:

$$f(t) = \frac{d}{dt}\left(1 - e^{-(t/\eta)^{\beta}}\right) = e^{-(t/\eta)^{\beta}} \cdot \frac{d}{dt}\left((t/\eta)^{\beta}\right) \frac{d}{dt}(t/\eta)^{\beta} = \beta(t/\eta)^{\beta-1} \cdot \frac{1}{\eta}$$

So:

$$f(t) = \frac{\beta}{\eta}\left(\frac{t}{\eta}\right)^{\beta-1} e^{-(t/\eta)^{\beta}}$$

4. Hazard:

$$h(t) = \frac{f(t)}{S(t)} = \frac{\frac{\beta}{\eta}\left(\frac{t}{\eta}\right)^{\beta-1} e^{-(t/\eta)^{\beta}}}{e^{-(t/\eta)^{\beta}}} = \frac{\beta}{\eta}\left(\frac{t}{\eta}\right)^{\beta-1}$$

This shows:

- $\beta < 1$: decreasing hazard (infant mortality)
- $\beta = 1$: constant hazard (exponential)
- $\beta > 1$: increasing hazard (wear-out)

6.1. Automotive and A manufacturing lines

A well-documented application of data-driven predictive maintenance to decrease the downtime of an automotive assembly line provides insights into the data sources appropriate for real-world applications. A flap closing and bow-tying process on a production line was monitored with fifty-eight sensors and one-hundred seventy-five process features related to heating and pressing cycles. An unsupervised gradient boosting model trained on fault-free intervals detected several faults and provided time to failure alerts. Causality analysis using reinforcement learning model-agnostic interpretability identified the critical influence of motors on the closing and bow-tying operations.

An integrated physics-informed model-predictive control approach for an A manufacturing line aims to minimize both setup and production costs by employing predictive maintenance based on a kernel method for the handle unit. Set-up datasets were obtained from angle scenarios by an intrusion detection system for an A product on an A manufacturing line, and feature importance detected extreme environments and noised data. The visualized subtle variations of the setup dataset provided conditions for preventing poor-quality products and setups. The kernel method with elevated penalty accompanied by physical knowledge for unit purposes, objectives, and consequences recovered predictive performance.

6.2. Semiconductor and Electronics Manufacturing

Semiconductor and electronics fabrication plants encompass a wide spectrum of high- and low-volume production lines. The fabrication of non-memory and memory chips relies on a large amount of complex manufacturing processes and requires an extremely efficient failure detection capability. However, sudden failures may still occur in various fabrication machines and adversely affect production performance. The failure of the Chemical Mechanical Polishing (CMP) process, for instance, may lead to significant expense. Moreover, the failure detection and prediction time window of the CMP process is considered very small. As such, a hybrid logistic regression model is developed to detect future failures of the CMP process, based on the Time-to-Event (TTE) data items collected from the relevant equipment.

DOI: 10.17148/IJARCCE.2022.111261

Several other predictive maintenance applications exist in the semiconductor fabrication sector, such as the repairs of sensors in photolithography equipment, disturbance-adjusted life prediction for extreme ultra violet lithography machines, and remaining useful life (RUL) prediction for dry etching equipment. In electronics manufacturing, time-dependent production data, including assembly machine failure records, are harnessed to optimize smart factory building preconditions and business key performance indicators (KPIs) using fuzzy logic. Furthermore, research investigates the concurrent interactive effect of historical manufacturing KPIs and machine status on production quality in an electronics assembly manufacturing line.

6.3. Textile and Food Processing

The textile and food manufacturing sectors are also leveraging predictive maintenance to continuously monitor asset condition and minimize unanticipated downtimes. In the textile industry, knitting and weaving machines are often equipped with vibration sensors for early fault detection. Fourier transform analysis of vibration data from these sensors enables the identification of common fault types, such as specific bearing defects. To address potential false alarms using a signal-based vibration monitoring approach, Jiang et al. use zonal vibration data as features for an LSTM model, resulting in classified defect types. Moreover, Xu et al. conduct a failure prediction study on air compressors at a cotton spinning factory. Based on their failure history, they build a random forest classifier that uses the operating hours, running time, temperature, vibration, and oil temperature data of the compressors as features.

In the food and beverage industry, prediction services help avoid equipment breakdowns, which could cause economic losses and affect sales dynamics. Rao et al. investigate the feasibility of predicting the remaining useful life of a washing machine used in an automated bottle washer to remove bacteria and soils. The original equipment manufacturer provides the three RUL observations in service logs, and Wuxi E-Winea provides the feature-engineered dataset, where the two features are the elapsed time and the count of wash cycles. The RUL prediction is performed using extreme gradient boosting, and a so-called gap metric indicates the prediction accuracy of the three instances.

7. CHALLENGES, RISKS, AND MITIGATION STRATEGIES

Although predictive maintenance promises significant advantages over standard practices, there are a number of ecosystem-wide challenges, risks, and failure modes that must be managed effectively. This section outlines these issues and discusses possible mitigation measures.

Data privacy and security risks arise from the collection and sharing of large amounts of data with external cloud computing and model-hosting services. Appropriate measures must be taken throughout the lifecycle of data assets, from acquisition through model deployment and risk management. Combining locally maintained models with only the aggregated data that contains valuable insights reduces data exposure and presents considerable advantages. Synthetic data generation, GANs, and CR-CDAs complement the previous strategy, helping maintain privacy while supporting model training and validation.

Data drift occurs as underlying patterns change over time, leading to declining model performance. Standard testing processes for updated models mitigate the impact, especially on less frequently used deployed models. A variety of techniques enable unsupervised analysis of drifting features, data distribution, and relationship shift detection through the estimation of the Wasserstein distance, thereby aiding model refresh. Causality-based relationships through graphical models or deep learning also help maintain reliable operation and induce spatio-temporal forecasting.

Scalability and resource constraints impede implementation of an ecosystem-wide strategy. New solutions need to be developed at all levels, from edge to cloud systems, in a resource-efficient way combined with resource-efficient models.

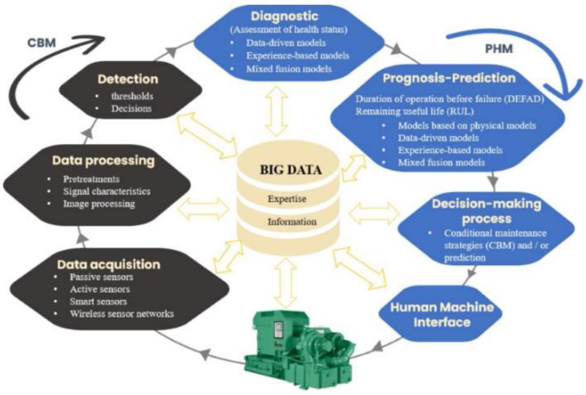


Fig 4: Overview, Models, and Challenges

7.1. Data Privacy and Security

Over the last two decades, predictive maintenance (PdM) models have been proposed for various production-oriented applications. Most of these proposals have focused on a proof-of-concept level rather than the industrialization of the models or developing an operational deployment process. Recent trends in Industry 4.0 have combined the concept of smart factories with the building of suitable predictive models that can be operationalized at scale. The implications of this transition provide interesting challenges, risks, and a range of potential mitigation strategies. Most of the potential risks originate from the underlying use of data, both during developing and deploying models.

The availability of sensitive data during the development and deployment phases of PdM models is a safety and privacy concern for an industrial organization. Unless the organization has a strong data security and privacy policy in place – for example, a formal nondisclosure agreement (a legal document preventing data theft, misuse, and leaking of trade secrets) with all parties involved – the data will be available for attack from a malicious insider or by a dedicated attack team. The mitigating strategy is to distinctively limit the availability of sensitive data. Data management approaches, such as Multilayer Perceptrons, Fuzzy logic, Support Vector Machines, and K-Nearest Neighbors, have been shown to overcome sensitive-data-smuggling problems effectively by masking the data attributes. Proposing suitable methods and tools for sensitive-data masking that are precise for PdM models would, therefore, contribute toward risk mitigation.

7.2. Data Drift and Model Degradation

The statistical and machine learning models for predictive maintenance and related use cases follow the same principle they are trained on the historical data and the real-time sensors data are fed to the model for inference. The statistical models have simpler structures, requiring fewer parameters and demand less data than machine learning/dl models. In a low-availability data situation, the lifetime of the predictive maintenance model can be longer than complex ML/DL models. During inference time, the statistical models generate signals separately and feed them to the exceedance prediction model allowing the ML/DL models directly inferred on the key CMParts.

Machine learning and deep learning-based predictive maintenance methods require large sizes of high-quality labelled data. In the initial phase, the demand for such high-quality labelled data for model training may not be fulfilled. The early ML/DL predictive maintenance models may use the labelled data of a select few CMParts and extend the capability to cover the remaining CMParts of the line through transferrable learning. During operation, data drift can be a concern. A set of strategies for a low-cost signal-based concept-drift detection and customized ML model decomposition for timely updating of concept-drifted models can be considered to mitigate the risk. Models can thus be updated on-the-fly to suit real-time signals without further additional costly retagging of data.

7.3. Scalability and Resource Constraints

Business concerns about resource engagement, service or expertise prices, train and equip needs, technology adaptability and overhead costs hinder adoption of predictive maintenance solutions. The Service-oriented Architecture concept can counterbalance these concerns letting the required service be provided by a third party, only when needed. The dealing party may provide a substitute and perform a replace or repair operation at a distant site, or deliver scheduled preventive actions. Data privacy risks emerge, as data produced by a platform owner are shared with a Service Provider that does not belong to the supply chain. Nevertheless, the advantages of such PAes significantly outshine the risks.

Analytics engine scalability assumes great importance when deploying data-driven models since both the number of devices and the volume of generated data increase quickly. Normally, Managed Services offered by cloud providers cope with such efficiency needs and allow the required elasticity, but concerns about data ownership and privacy may prevent their adoption. Edge Computing may provide a partial solution, processing and storing part of the data in own nodes or in nodes distributed by the supply chain. However, on-premises solutions, such as data lakes deployed in dedicated infrastructures nowadays potentially represent the most suitable answer. blue edge, hybrid edge and other solutions should also be investigated. In any case, solutions for automatic monitoring of data drifts at various levels need to be developed, enabling smart cross-validation of the produced models, fast calculation of new replacement function and hypothesis tests validation whenever required.

8. CONCLUSION

A decade ago, predictive maintenance—its definitions and objectives, data sources and feature engineering, evaluation metrics, modeling paradigms, acquisition, and preprocessing—revealed the concept's broad relevance. Recent developments and examples close the discussion. Future trends include the promise of digital twins and federated learning for privacy-aware industrial predictive maintenance.

A confluence of forces drives the need for increased efficiency in industry. Characteristics of upcoming industries are directed toward real-time communication, standardization, flexibility, rapid response, integrated management, coexistence with the world environment, modular design, and friendly links related to intelligent manufacturing technology, cyberspace, the Internet of Things, and smart grids. Breakthroughs in deep learning accelerate smart

DOI: 10.17148/IJARCCE.2022.111261

manufacturing research. Advanced intelligent semiconductor processes facilitate next-generation technologies—opto-electronic, ultraviolet, nanometer-scale artificial intelligence integrated circuit, radio-frequency, and power electronic. New intelligent textile technologies promise sustained progress in the next decade for printing and dyeing, silk, cotton, wool, knitwear, fur, and small home textiles. In the food industry, systems that can mimic human taste and smell are replacing human sensory analysis, requiring efficient, accurate, and connection-oriented food product process design.

8.1. Future Trends

For the automotive industry's production lines—such as assembly and painting—different models for different sub-processes usually involve predictive algorithms embedded in cloud systems. In contrast, semiconductor and electronics manufacturing lines require a different approach. These processes are highly repetitive; hence, testing is hardly executed at all. Pre-defined sequences of function calls serve as state machine representations of these processes, known to operate for a specific duration. For such repetitive processes, hybrid models combining inferred patterns of probed variables with support vector classifiers have performed best.

The lifecycle of Predictive Maintenance models involves regular updating of the data used for training and inference execution. In automotive and A-models, these models are computation-heavy and not executed continuously. Performing inferences only when further operations are about to become critical—paired with maintenance operations in case of detected malfunctions—decreases the processing power needed in the edge devices. Consequences of a predicted failing mode are usually more severe than costly but unnecessary maintenance actions. Furthermore, the design of the inference engines accounts for explainability—not due to regulation, but to enhance trust within the users.

However, even large multilayer-perceptron Models are not sufficient to infer the state of remaining life. The so-called semiconductor corner is defined by very low yields in the process. Δy conclusion: Yp(best submodel) < Yp(worst submodel) always. All parts of the process suffer from different types of wear. The final conclusion is not to predict remaining lifetime but to keep the mode not worn out.

REFERENCES

[1] Rongali, S. K. (2022). AI-Driven Automation in Healthcare Claims and EHR Processing Using MuleSoft and Machine Learning Pipelines. Available at SSRN 5763022.

[2] Ahn, J., Park, S., Kim, J., & Lee, S. Federated learning for predictive maintenance and anomaly detection using time series data distribution shifts in manufacturing processes. Sensors, 23(17), 7331. doi:10.3390/s23177331

[3] Varri, D. B. S. (2022). AI-Driven Risk Assessment And Compliance Automation In Multi-Cloud Environments. Journal of International Crisis and Risk Communication Research , 56–70. https://doi.org/10.63278/jicrcr.vi.3418.

[4] Baur, M., Albertelli, P., & Monno, M. (2020). A review of prognostics and health management of machine tools. The International Journal of Advanced Manufacturing Technology, 107, 2843–2863.

[5] Vadisetty, R., Polamarasetti, A., Guntupalli, R., Raghunath, V., Jyothi, V. K., & Kudithipudi, K. (2022). AI-Driven Cybersecurity: Enhancing Cloud Security with Machine Learning and AI Agents. Sateesh kumar and Raghunath, Vedaprada and Jyothi, Vinaya Kumar and Kudithipudi, Karthik, AI-Driven Cybersecurity: Enhancing Cloud Security with Machine Learning and AI Agents (February 07, 2022).

[6] Cheng, X., Chen, W., & Li, J. (2022). Systematic literature review on visual analytics of predictive maintenance: Approaches, applications, and future directions. Sensors, 22(17), 6321. doi:10.3390/s22176321

[7] Inala, R. Advancing Group Insurance Solutions Through Ai-Enhanced Technology Architectures And Big Data Insights.

[8] Gascón, A., Hernández, J., & Galar, D. (2022). Providing fault detection from sensor data in complex industrial environments: A layered structure for predictive maintenance. Sensors, 22(2), 586. doi:10.3390/s22020586

[9] Garapati, R. S. (2022). A Web-Centric Cloud Framework for Real-Time Monitoring and Risk Prediction in Clinical Trials Using Machine Learning. Current Research in Public Health, 2, 1346.

[10] He, W., Chen, J., Zhou, Y., Liu, X., Chen, B., & Guo, B. (2022). An intelligent machinery fault diagnosis method based on GAN and transfer learning under variable working conditions. Sensors, 22(23), 9175. doi:10.3390/s22239175

[11] Nagabhyru, K. C. (2022). Bridging Traditional ETL Pipelines with AI Enhanced Data Workflows: Foundations of Intelligent Automation in Data Engineering. Available at SSRN 5505199.

[12] Hermansa, M., Kozielski, M., Michalak, M., Szczyrba, K., Wróbel, Ł., & Sikora, M. (2022). Sensor-based predictive maintenance with reduction of false alarms—A case study in heavy industry. Sensors, 22(1), 226. doi:10.3390/s22010226

[13] Avinash Reddy Aitha. (2022). Deep Neural Networks for Property Risk Prediction Leveraging Aerial and Satellite Imaging. International Journal of Communication Networks and Information Security (IJCNIS), 14(3), 1308–1318. Retrieved from https://www.ijcnis.org/index.php/ijcnis/article/view/8609.

[14] Li, Z., & X A survey of deep learning-driven architecture for predictive maintenance. Engineering Applications of Artificial Intelligence, 132, Article 108xxx.

[15 Gottimukkala, V. R. R. (2022). Licensing Innovation in the Financial Messaging Ecosystem: Business Models and Global Compliance Impact. International Journal of Scientific Research and Modern Technology, 1(12), 177-186

[16] Liu, W., & X (2022). Three-stage Wiener-process-based model for remaining useful life prediction of cutting tools. Sensors, 22(13), 4763. doi:10.3390/s22134763

[17] Avinash Reddy Segireddy. (2022). Terraform and Ansible in Building Resilient Cloud-Native Payment Architectures. International Journal of Intelligent Systems and Applications in Engineering, 10(3s), 444–455. Retrieved from https://www.ijisae.org/index.php/IJISAE/article/view/7905.

[18] Liu, Y., Yu, W., Dillon, T., Rahayu, W., & Li, M. (2022). Empowering IoT predictive maintenance solutions with AI: A distributed system for manufacturing plant-wide monitoring. IEEE Transactions on Industrial Informatics, 18(2), 1345–1354. doi:10.1109/TII.2021.3091774

[19] Chava, K., Chakilam, C., & Recharla, M. (2021). Machine Learning Models for Early Disease Detection: A Big Data Approach to Personalized Healthcare. International Journal of Engineering and Computer Science, 10(12), 25709–25730. https://doi.org/10.18535/ijecs.v10i12.4678.

[20] Mazzei, D., Betti, A., Fantoni, G., & Iannone, R. (2022). Machine learning for Industry 4.0: A systematic review using deep learning-based topic modelling. Sensors, 22(23), 9251.

[21] Amistapuram, K. (2022). Fraud Detection and Risk Modeling in Insurance: Early Adoption of Machine Learning in Claims Processing. Available at SSRN 5741982.

[22] Pollak, A., & X (2021). Prediction of belt drive faults using vibration-based anomaly detection for predictive maintenance. Applied Sciences, 11(21), 10307.

[23] Chakilam, C., Suura, S. R., Koppolu, H. K. R., & Recharla, M. (2022). From Data to Cure: Leveraging Artificial Intelligence and Big Data Analytics in Accelerating Disease Research and Treatment Development. Journal of Survey in Fisheries Sciences. https://doi.org/10.53555/sfs.v9i3.3619.

[24] Pugalenthi, K., & X (2022). Remaining useful life prediction of lithium-ion batteries: A data-driven approach. Sensors, 22(10), 3803. doi:10.3390/s22103803

[25] Annapareddy, V. N. (2022). AI-Driven Optimization of Solar Power Generation Systems Through Predictive Weather and Load Modeling. Available at SSRN 5265881.

[26] Soori, M., & X Internet of things for smart factories in Industry 4.0: A review. Internet of Things and Cyber-Physical Systems, 3, 1–xx.

[27] Muthusamy, S., Kannan, S., Lee, M., Sanjairaj, V., Lu, W. F., Fuh, J. Y., ... & Cao, T. (2021). Cover Image, Volume 118, Number 8, August 2021. Biotechnology and Bioengineering, 118(8), i-i.

[28] van Dinter, R., Tekinerdogan, B., & Catal, C. (2022). Predictive maintenance using digital twins: A systematic literature review. Information and Software Technology, 151, 107008. doi:10.1016/j.infsof.2022.107008

[29] Sriram, H. K. (2022). Advancements in Credit Score Analytics using Deep Learning and Predictive Modeling Techniques. Available at SSRN 5255128.