

RELIABILITY PAPER

A rule-based machine learning methodology for the proactive improvement of OEE: a real case study

Laura Lucantoni, Sara Antomarioni, Filippo Emanuele Ciarapica and Maurizio Bevilacqua

*Department of Industrial Engineering and Mathematical Science,
Università Politecnica delle Marche, Ancona, Italy*

Abstract

Purpose – The Overall Equipment Effectiveness (OEE) is considered a standard for measuring equipment productivity in terms of efficiency. Still, Artificial Intelligence solutions are rarely used for analyzing OEE results and identifying corrective actions. Therefore, the approach proposed in this paper aims to provide a new rule-based Machine Learning (ML) framework for OEE enhancement and the selection of improvement actions.

Design/methodology/approach – Association Rules (ARs) are used as a rule-based ML method for extracting knowledge from huge data. First, the dominant loss class is identified and traditional methodologies are used with ARs for anomaly classification and prioritization. Once selected priority anomalies, a detailed analysis is conducted to investigate their influence on the OEE loss factors using ARs and Network Analysis (NA). Then, a Deming Cycle is used as a roadmap for applying the proposed methodology, testing and implementing proactive actions by monitoring the OEE variation.

Findings – The method proposed in this work has also been tested in an automotive company for framework validation and impact measuring. In particular, results highlighted that the rule-based ML methodology for OEE improvement addressed seven anomalies within a year through appropriate proactive actions: on average, each action has ensured an OEE gain of 5.4%.

Originality/value – The originality is related to the dual application of association rules in two different ways for extracting knowledge from the overall OEE. In particular, the co-occurrences of priority anomalies and their impact on asset Availability, Performance and Quality are investigated.

Keywords Overall equipment effectiveness, Machine learning, Association rules, Priority anomaly detection
Paper type Research paper

1. Introduction

Industry 4.0 encompasses a multitude of digital technologies that affect manufacturing enterprises in different contexts (Zheng *et al.*, 2021). In particular, monitoring the status of production equipment, detecting anomalies and predicting and resolving them before they occur is possible through the use of advanced I4.0 technologies (Jasiulewicz-Kaczmarek and Katarzyna, 2022). In this context, Big Data often provides extensive information for continuous improvement, cost reduction and increased efficiency in manufacturing (Taghavi *et al.*, 2020), if properly analyzed.



New Machine Learning (ML) methods are thus under increasing development because of their relevant role in Big Data Analytics (BDA), particularly in achieving improved production and equipment productivity (Chan *et al.*, 2005). In this context, since Overall Equipment Effectiveness (OEE) is considered a standard for measuring the productivity of equipment as a function of its efficiency (Arun Prasad and Panse, 2022), studying the equipment through ML based on the evaluation of OEE can be beneficial for manufacturing process improvement.

An evaluation of data analysis techniques shows that traditional methodologies, such as the Pareto diagram (Abdelrahman and Keikhosrokiani, 2020), are widely used in the scientific literature for anomaly detection. For instance, Chong *et al.* (2016) analyzed how FMEA can provide a list of priority corrective actions for improving OEE in a real-world case study.

However, a recent literature review points out that academic research should strive to improve on existing theoretical deficiencies to carry out semi-automation and full automation analyses (Wu *et al.*, 2021). With this aim, Association Rules (ARs) and Network Analysis (NA) are proposed in the current industrial scenario as innovative methodologies for identifying corrective actions based on the exploration of hidden relationships among data (Antomarioni *et al.*, 2022). Considering the huge amount of data available in manufacturing, incorporating traditional methods with AR and NA can support the continuous improvement perspective of OEE by focusing on selecting useful data and then proceeding with their exploration, knowledge extraction and prediction (Maimon and Rokach, 2010).

In this context, despite valuable existing research, there is no methodology in the literature based on the application of traditional methods and multiple ARs and NA to structurally analyze a new scientific approach for identifying and prioritizing improvement actions in manufacturing from an OEE continuous improvement perspective.

Therefore, the proposed methodology aims to address this gap by proposing a new rule-based ML framework for improving OEE by experimenting with the automotive industry, which is constantly challenged by strong competitiveness in terms of product quality and process efficiency (Costa *et al.*, 2022). Specifically, the proposed approach is based on the integration of the analysis of the Six Big Losses and ARs for the detection, classification and prioritization of major anomalies through the Eisenhower Matrix. Thus, the simultaneous application of AR and NA techniques enables decision-makers to discover the hidden relationship between each priority anomaly and its influence on all OEE loss factors. The results obtained are used to identify and apply improvement actions in a continuous improvement perspective through the implementation of the Deming Cycle. The added value of this work is mostly based on the final objective: in (Antomarioni *et al.*, 2020), a framework based on traditional methods, ARs, and NA to define the relationships among failure modes and related characteristics that are likely to occur concurrently aiming to predict when to perform the inspection. Differently, this work aims to prioritize improvement actions based not only on failure modes and thus availability, but in general on anomalies that may affect all OEE loss factors (availability, performance and quality). An additional added value is the application of ARs in two different stages and modes of the framework in order to exploit the benefit of knowledge extraction for continuous improvement.

After this introduction, Section 2 presents the state of the art to explore current ML strategies available in the literature for OEE continuous improvement. Section 3 describes and details the rule-based ML methodology developed in this work, while Section 4 presents a real-life application with results and impact evaluation. Finally, Section 5 discusses conclusions and future developments.

2. State of the art

Research is still evolving toward developing new frameworks and methodologies that link I4.0 enabling technologies to specific goals and their impact on manufacturing enterprises

(Calabrese *et al.*, 2022). Regarding most of the goals of Industry 4.0 in the manufacturing sector, asset management and valuation are the key requirements to assist companies in the transition to I4.0 (Gökalp *et al.*, 2017).

Although averaging has been one of the most important ways to measure process performance, averages or any other aggregate measure do not provide a mechanism for identifying how to improve asset management (Jeong *et al.*, 2022). A common approach to deal with this issue is to use a key performance indicator (KPI)-based approach (Zhang *et al.*, 2018). As a result, an in-depth study of how to improve a production process by focusing mainly on KPIs that depend solely on component conditions (Deeskow *et al.*, 2008) in the field of asset monitoring and evaluation is discussed below. However, it should be noted that only ML methods were considered in this contest to investigate the effect of correlated index parameters on the asset.

2.1 ML techniques for KPI improvement

It should be pointed out that only ML techniques were considered to investigate KPI-based asset valuation. Specifically, 97 scientific contributions were found in the Scopus database regarding ML algorithms for asset valuation through the KPIs best suited to the manufacturing sector as in Figure 1.

Based on the evidence from Figure 1, MTBF (Main Time Between Failure) is the most popular KPI in the field of asset valuation due to its easy accessibility and understanding (Jittawiriyankoon and Srisarkun, 2022). In addition, asset performance monitoring is of great use in predicting any potential asset behavior, problems and losses (Márquez *et al.*, 2019), followed by asset reliability, availability and efficiency with the same objective (Liu *et al.*, 2022; Ruschel *et al.*, 2020; Yu *et al.*, 2019). The literature review revealed that ML algorithms are often correlated with MTBF, performance and reliability of assets, especially to predict priority in the maintenance schedule.

However, to improve asset evaluation in not just maintenance scheduling but also failure prevention and prediction, ML techniques are also focusing on improving OEE, although with few contributions, as well as MTTD (Main Time To Detection) which is still underutilized. Therefore, the current literature aims to study how OEE is emerging in existing contributions to overcome these problems, focusing mainly on real-world scientific applications.

2.2 DM models for OEE improvement

14 scientific contributions were found in the Scopus database by selecting (“machine learning” OR “machine learning” OR “ML”) AND “OEE”) in article titles, abstracts and keywords.

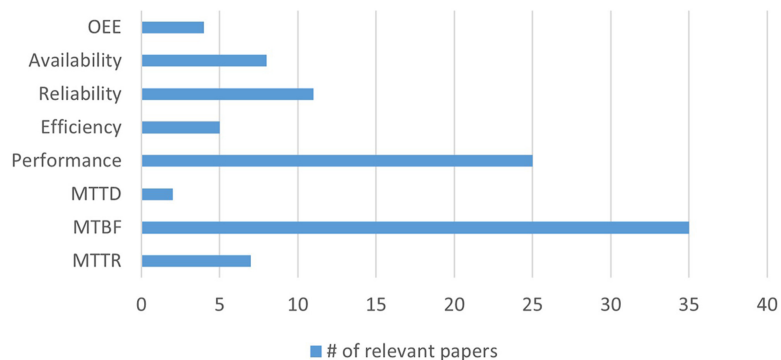


Figure 1.
ML scientific contributions for asset improvement actions discovered based on the most suitable KPIs

Source(s): Author's own creation/work

In line with the objective of the current paper, other scientific contributions in the literature emphasize the role of “ML for OEE improvement”, mostly based on improving availability as an OEE factor (Utz *et al.*, 2018) (Brodny and Tutak, 2018). In addition, in order to revolutionize the way of obtaining answers to asset performance questions (Black, 2014), the need for a more detailed and in-depth analysis of OEE loss factors is growing (see Table 1).

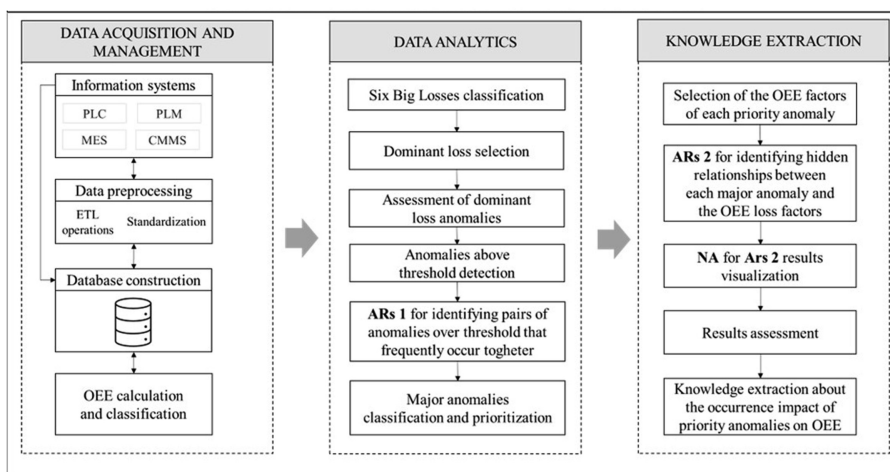
Therefore, despite the valuable research available in the literature, the integration of traditional methods, ARM, and NA to uncover the hidden relationships between priority anomalies and OEE loss factors is not developed in scientific research to the best of the author’s knowledge. Therefore, the present paper aims to fill this research gap.

3. Methodology

The proposed methodology consists of an ML framework and a Deming Cycle implementation, respectively for knowledge extraction and the selection and implementation of proactive actions and the final assessment. First, a three-step framework (Figure 2) is proposed for extracting useful information through rule-based ML methods with a view to the identification of priority anomalies major influencing the OEE losses. The proposed framework consists of the following: (I) data acquisition and pre-

DM models for OEE improvement	Reference
“Hybrid analysis using a combination of clustering and human analysis to identify bottleneck station at the automotive semi-automatic assembly line”	Dobra and J3osvai (2020)
“Neural Networks to estimate and to assess the efficiency loss by 15 related individual input factors, e.g. process time, useable tool, the standard deviation of lot size, etc.”	Yu <i>et al.</i> (2019)
“A Deep Learning (DL)-based approach and historical production performance data related to measurements, warnings, and alarms for production performance forecasting”	Brunelli <i>et al.</i> (2019)
“Formulation of ARM to find a rule that shows the well-computed relationship between measurable indicators of OEE”	Djatna and Alitu (2015)

Table 1.
Relevant ML methods
for OEE improvement
in manufacturing



Source(s): Author’s own creation/work

Figure 2.
Rule-based ML
framework for OEE
proactive improvement

processing for data collection, management and OEE classification; (II) major anomaly detection and prioritization; (III) priority anomalies and OEE loss hidden relationship investigation. Eventually, a Deming Cycle has been developed for implementing proactive actions from an OEE continuous improvement perspective.

3.1 *Data acquisition and pre-processing*

The data collection and management step is the starting point of the rule-based ML framework developed in this work for the OEE proactive improvement. The final aim is to identify the primary source of the proposed framework for supporting proactive actions and continuous improvement. Indeed, the data sources required to build a robust and complete information system include maintenance reports and interventions, historical parameters on assets' status and variables, preventive maintenance plans and Key Performance Indicators (KPI) for OEE calculation (i.e. availability, performance and quality).

Programmable Logic Controller (PLC) and telemetry system are also required in order to be able to continuously monitor machine status and production variables, by collecting the real-time signal from smart sensors and devices.

Data pre-processing is needed in order to transform data into knowledgeable information; in this context, OEE is calculated multiplying factors (i.e., performance, availability and quality) and ranges can be created in order to classify it.

Finally, previously collected data are integrated in one system. The implementation complexity of this step depends on data quantity and quality. Data pre-processing can be summarized as follows.

- (1) Data integration for saving the information gathered from data collection.
- (2) Data pre-elaboration through ETL operation Extraction, Transformation, and Loading to clean data from errors and repetitions, and provide as homogeneous a dataset as possible.
- (3) OEE calculation and *classification* in order to represent the AS-IS situation quantitatively.

3.2 *Data analytics*

The application of ML for anomaly detection and prioritization consists of preliminary analytics for major anomaly detection, which has proven problematic and should be easier to improve. With this aim, the Six Big Losses classification is required for the selection of the dominant loss class, as pointed out by several authors (e.g. [Setyawan et al. \(2021\)](#) and [Badiger and Gandhinathan \(2008\)](#)): this classification supports addressing the main OEE factor that will be affected by an improvement of the appointed cause (see [Table 2](#)).

Once the dominant loss class has been identified, a numerical assessment of each anomaly belonging to the class is needed for better classifying and prioritizing major anomalies. In this

Table 2.
Traditional six big losses classification based on OEE loss factors

Overall equipment effectiveness	Traditional six big losses
Availability loss	Breakdown
Performance loss	Setup and Adjustments
	Reduced Speed
	Idling Minor Stoppage
Quality loss	Rework
	Scrap or Yield

regard, there is no such thing as a global indicator, but a specific measure index should be computed based on the dominant loss class typology (a few examples of possible indicators are shown in Table 3).

Once both the dominant loss class and the measure index are selected, the second step of the framework aims to discover the pairs of major anomalies that frequently occur concurrently. Subsequently, an Eisenhower matrix is filled to classify the major anomalies and, most importantly, to prioritize them based on their urgency and importance (Jyothi and Parkavi, 2016). The urgent priority anomalies could affect even more than one OEE factor; hence, the relationships between them and the OEE loss factors should be investigated to identify the best proactive action.

3.2.1 Association rule mining. The Association Rule Mining aims to identify the relations among attributes and values stored in large datasets that frequently co-occur (Buddhakulsomsiri et al., 2006). The objective is to capitalize on the production process's data to extract information and knowledge from them.

Given a set of items (i.e. Boolean data) $I = \{i_1, i_2, \dots, i_n\}$ and given a set of transactions $T = \{\tau_1, \tau_2, \dots, \tau_m\}$ each of whom is composed by an itemset included in I . An Association Rule (AR) $\alpha \rightarrow \beta$ can be defined as an implication between itemsets – α and β – belonging to I ($\alpha, \beta \subseteq I$) and having no elements in common ($\alpha \cap \beta = \emptyset$). ARs' quality is determined through the calculation of different metrics. Support (Supp) and Confidence (Conf) will be considered in this framework since they represent the most commonly used, according to the existing literature. The former (1) measures the statistical relevance of a rule, which is calculated as the ratio between the number of transactions containing both α and β over the cardinality of the transaction set T . Instead, the latter (2) is measured as the ratio between the support of the rule $\alpha \rightarrow \beta$ over the support of the itemset α . Indeed, it represents the conditional probability of having β in a transaction containing α .

$$Supp(\alpha \rightarrow \beta) = \frac{\#(\alpha, \beta)}{\#T} \tag{1}$$

$$Conf(\alpha \rightarrow \beta) = \frac{Supp(\alpha \rightarrow \beta)}{Supp(\alpha)} \tag{2}$$

The FP-growth algorithm has been chosen to mine the ARs in the proposed application (Han et al., 2007) since it is more efficient in time and memory requirements. Only the itemsets meeting the support and confidence constraints are mined. Accordingly, the procedure run for the ARM can be recapped as follows.

- (1) Definition of minimum *Supp* and *Conf* thresholds;
- (2) FP-growth algorithm running and generation of the Frequent Itemsets (i.e. itemset having support higher than the minimum *Supp* threshold);
- (3) Combination of the items belonging to the Frequent Itemsets to create the ARs having a *Conf* higher than the minimum *Conf*.

Dominant loss class	Anomalies	Measure index
Availability	Failures	Risk priority number
Performance	Minor stops	Cycle time
Quality	Waste and rejects	Minutes of production waste and reworks

Table 3.
Examples of anomalies
in numerical
assessment

3.2.2 *Application of association rule mining for the priority anomaly detection.* In order to explain the proposed approach, a simplified example shows the integration of measurement index assessment and ARM for the prioritization of failure modes.

From an improvement perspective, it is essential to identify the measurement index indicating the occurrence risk for each anomaly. Usually, an intervention is needed when the measurement index value exceeds the threshold.

Let a range from “1” to “10” be the possible anomalies managed by a manufacturing process. First, analyze the list of anomalies, identifying those with a measurement index over the threshold. Supposing we have ten as the threshold value and five anomalies over this, as represented in Figure 3 by five red points, they are considered the major anomalies. Then, the relationships between the major anomalies are studied with one of the leading data mining techniques, such as Association Rules. Indeed, the second diagram reported in Figure 3 contains the ARs mined from the exemplified list of the selected anomalies over the threshold. The red warning signals in the third diagram of Figure 3 represent the anomalies that need to be solved first due to being considered the “most priority”. The Eisenhower matrix is thus adopted to prioritize the anomalies which needed actions, as follows in Table 4.

Priority anomalies with the measurement index above the threshold value and present as both left and right sides of ARs are assigned to the “urgent and important” quadrant, while those present only as left or right sides are classified as “important and not urgent”. The remaining major anomalies with the measurement index above the threshold but no hidden relationship occurs are classified as “urgent and unimportant”. Finally, anomalies under the threshold value are not urgent or essential for OEE improvement; therefore, ARs are not investigated.

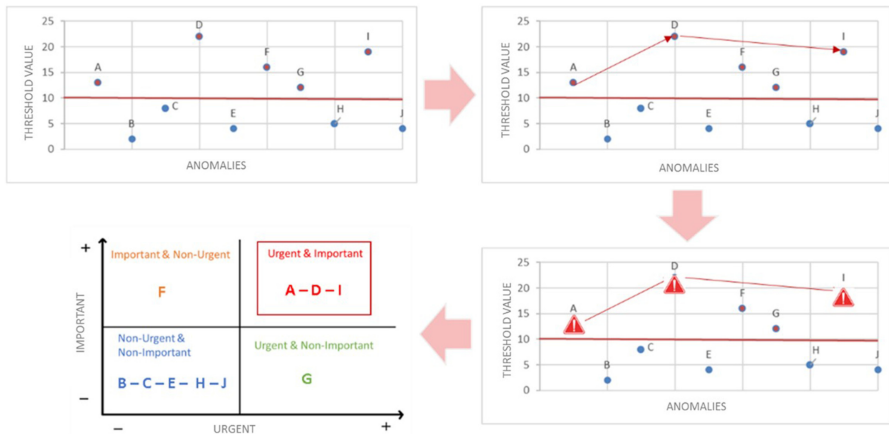


Figure 3. Example of measurement index assessment and ARM for anomalies prioritization

Source(s): Author’s own creation/work

Table 4. Criteria for the failure modes prioritization

Measurement index	ARs	Eisenhower matrix classification
Above threshold	“X” and “Y”	Urgent and Important
Above threshold	“X” or “Y”	Important and Non-Urgent
Above threshold	No	Urgent and Non-Important
Under threshold	Not analyzed in this work	Non-Urgent and Non-Important

Suppose the experts identify a new anomaly “A” in the future. In that case, this anomaly “A” must be solved at the same time as “D” and “T” because, from historical data, when the anomaly “A” occurs, usually “D” occurs as well, and therefore “T”. The process manager should repeat this analysis periodically. By following this procedure, repetitively, it is possible to prevent the number of anomalies from being solved after their occurrence.

3.3 Knowledge extraction

The last step addresses the knowledge level by identifying how the priority anomalies occurrences are influencing the OEE losses. Knowledge is thus the acquisition of helpful information for executing proactive improvement actions.

The ability to discover hidden relationships between the occurrence of each anomaly identified in the second stage of the framework and the OEE losses represents a knowledge extraction for managing improvement activities. For example, a rejection can be caused by the non-conformity of components, namely by non-controlled equipment (such as missing repairs, poor maintenance, out-of-control defect detection, etc.).

For each priority anomaly belonging to the “Urgent quadrant” of the Eisenhower Matrix, the Knowledge extraction stage concerns the simultaneous application of ARM and Network Analysis (NA) techniques.

Specifically, ARM and NA are used to identify priority anomalies frequently occurring in conjunction with a reduction of some of the OEE parameters. Thus, they can affect them and frequently cause effects the overall manufacturing process over time. It should be noted that this step is particularly important due to its ability to identify undiscovered relationships: for instance, a failure mode could affect not just the “availability” losses of the OEE but also the “quality”. In such a case, the proactive action to be taken should be able to act on both factors.

The relationships between each priority anomaly and the significant losses of the production process most influencing the OEE are identified from such results. These relationships are classified by rules assessment. The same procedure is thus repeated for all the anomalies of the Urgent quadrant according to their priority, quantity and experts’ intention. In such a way, the variables selected continually change over time, depending on the current Dominant loss class and its anomalies of the process and how they affect the production factors.

3.3.1 Application of association rules and network analysis for main losses identification. Starting from the critical events or anomalies identified, the relationships between them and the production performance metrics are assessed at this step. Specifically, the ARs are used to define such relationships, and, furtherly, they are represented using the NA formalism. In other words, the network is not used to represent the plant’s physical structure but rather to describe the probability links between the anomalies and their impact on the process.

Networks are used to describe relations occurring in social structures; they are usually represented by ordered pairs of vertices (V) and arcs (A): $N=(V, A)$ (Otte and Rousseau, 2002). Their implementation refers to analyzing exchanges and interactions, i.e. vertices, among actors, i.e. the network’s nodes. The V set is composed of the failure modes occurring in the production process under investigation and the metrics taken into consideration to analyze an anomaly impact on the whole system; the A set, instead, is obtained through the ARs relating such failure modes and the selected metrics. A weight is attributed to each arc, that is, the confidence of the rule represented in the network. Specifically, if a rule in the form *anomaly_i → metric_j* exists, it will be represented as shown in Figure 4. The meaning of such a relationship can be interpreted as follows: when *anomaly_i* occurs, the monitored *metric_j* will assume the value j with a probability equal to the confidence value ($Conf(anomaly_i \rightarrow metric_j)$).

NA is used to visualize the critical relationships between anomalies and metrics values. One of the metrics that should be inquired about is the OEE. Indeed, it is a sensitive parameter providing relevant insights into the status of the process and is monitored by most companies.

Identifying the influence of an anomaly across the network is helpful to verify if, over time and through appropriate interventions, they have been reduced in terms of occurrence probability; on the other hand, it allows to visualize whether their occurrence is associated with an increase of the OEE – or the other monitored metrics, alternatively.

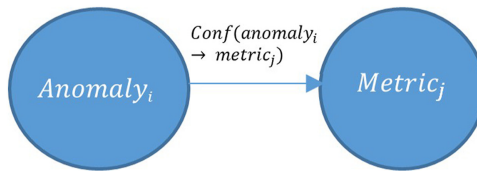
In the following sub-section, the rationale behind the approach’s cyclical implementation is presented to outline the main benefits derivable.

3.4 Deming Cycle implementation

The results are achieved through continuous process optimization by applying the PDCA 4.0 (Plan-Do-Check-Act), namely Deming Cycle as in Figure 5.

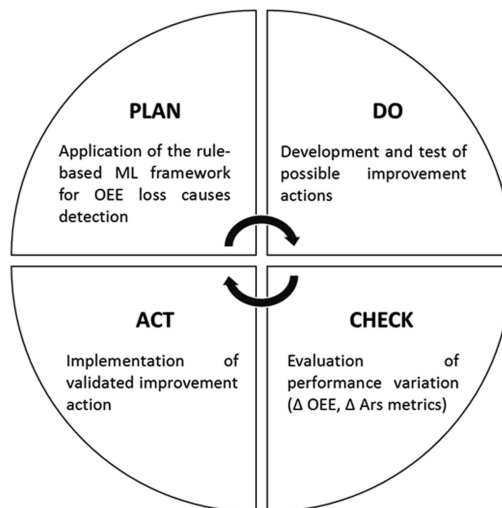
The “PLAN” step is designed to apply the rule-based ML framework for OEE proactive improvement developed in this paper. The “DO” is intended for the identification and testing of improvement solutions in order to define the most suitable proactive strategies (rejects maximum number reduction, quality sensor parameters re-assessment, early tools

Figure 4.
Representation of the relationship investigated through ARs



Source(s): Author’s own creation/work

Figure 5.
Deming Cycle application for proactive actions implementation from an OEE continuous improvement perspective



Source(s): Author’s own creation/work

replacement). While the “CHECK” and “ACT” steps evaluated the proactive actions’ effectiveness and the OEE’s actual variation after several months, by the following.

- (1) Proactive improvement activities are implemented according to schedule.
- (2) The variation in OEE is achieved.
- (3) The variation in ARs metrics is achieved.

The same procedure should be applied periodically, aiming to achieve continuous improvement through automatic data collection processes and periodical data analytics.

4. Methodology application

The framework proposed in this work (Figure 2) has been implemented through the Deming Cycle process (Figure 5) in an automotive company during its transition from a preventive strategy to a proactive philosophy for OEE improvement.

Specifically, the production line in the exam is a fully automated assembly line consisting of twelve stations. The time frame of the analysis is two years (Figure 6), and the real application of the first six months has been described in this paper just to explain the methodology. However, the overall results and impact of the methodology are detailed in the conclusions section.

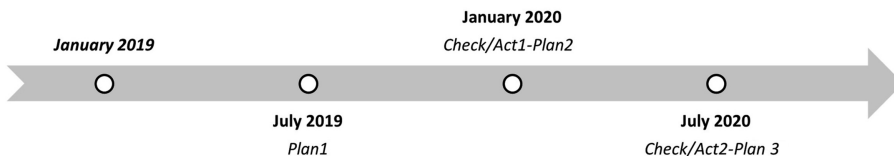
During the “Plan1” phase data from six months – from January 2019 to June 2020 – were analyzed through the application of the rule-based ML framework for proactive OEE improvement presented in this paper. Hence, the selected improvement actions were tested during the “Do1” phase. Next, during the “Check/Act1” phases the effectiveness of the improvement actions was verified in order to validate and standardized the most appropriate ones. As these activities have a cyclical pattern over time, each planning phase is followed by a new testing activity, while a new validation and implementation activity follows each check: each time frame is six months in order to correctly evaluate results.

How the rule-based framework and Deming Cycle have been implemented is described in the following sections.

4.1 Data acquisition and management

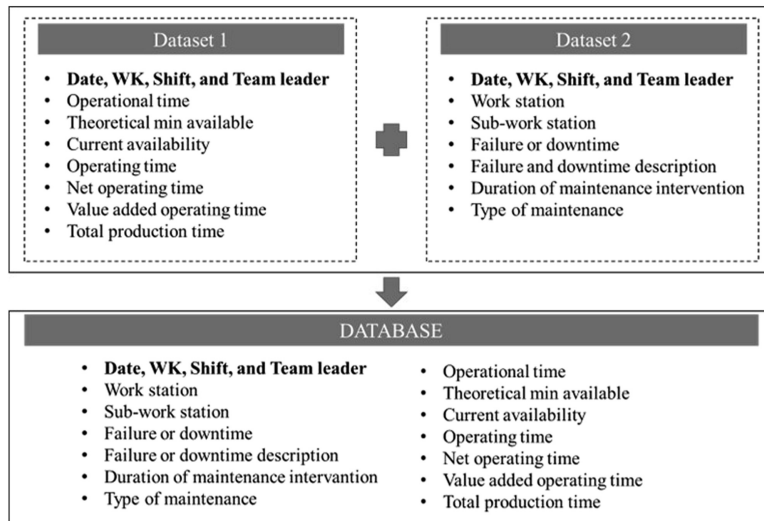
A thorough data acquisition is required to get information on the occurrence of anomalies. In particular, data and information are accessed from both PLM and MES systems to be aware of the quality, production and performance parameters to be monitored; consequently, it is necessary to access the technical information of machines – stored in the CMMS – of such station and equipment of the assembly process. It should be noted that PLM and MES are cloud-based systems, while the CMMS is server-based.

Two different datasets have been managed as in Figure 7 for the construction of the final database. The first dataset includes daily production data, such as production line structure, losses information, times, production and quality data, and performance measurement, aggregated by date and work shift. On the other hand, the second one concerns the detail of



Source(s): Author’s own creation/work

Figure 6.
Timeline of the Deming
Cycle application to the
case study



Source(s): Author's own creation/work

Figure 7.
Database construction

fault events, such as workstation structure, failure, downtimes typologies, causal and description, failure duration, and type of maintenance. Hence, both datasets have been aggregated for a faster application of the DM framework developed in this work. In this way, the final database consists of 3,854 instances throughout the application period of the framework. It should be also noted that the aggregation key for the final databases regards the date of a failure event, the week, the number of shifts and the name of the team leader as well.

After the aggregation of both datasets, all information has been cleaned from inconsistencies (i.e. attributes with negative values) and standardized. Once the final aggregation of both datasets has been completed, the final version of the database has been created by calculating and classifying OEE.

At this stage, the analysis described by the last two steps in the framework is first performed considering failure events and downtimes from 01/01/2019 to 30/06/2019. Each instance of the final database represents a specific anomaly of the process, while all the information such as date and shift of the occurrence, typology, description and production factors parameters are reported in the following columns.

During the first six months of 2019, the company collected 919 failure types/downtime and 28,417 min for maintenance interventions. The manufacturing company initially introduced a preventive maintenance strategy for early equipment replacement, i.e. for every 1,000 pieces produced. Subsequently, this strategy will be improved by acting on the causes of the events from a proactive maintenance perspective. The availability, performance and quality losses are calculated for each anomaly occurrence through the database in Figure 7. Once the final database has been defined, the OEE was chosen as the global indicator of the production line effectiveness. It is defined by determining the availability, performance and quality losses identified in a single anomaly occurrence, i.e. when a system operates at the standard production rate. The OEE values provided as input are classified by date and work shift by considering the following ranges.

- (1) Low values for OEE are included in the interval [0–0.333];
- (2) Medium values for OEE are included in the interval [0.333–0.667];
- (3) High values for OEE are included in the interval [0.667–1].

4.2 Data analytics

To analyze the OEE of the assembly line, the specific losses affecting the current case study were classified based on the traditional Six Big Losses as well as the total hours of events (Table 5), and the occurrence percentage through the Pareto Diagram (Figure 8). In this way, the dominant losses that need to be proactively addressed were identified.

As shown in the Pareto Diagram, the dominant loss class regards the “Breakdown loss” and the Risk Priority Number (RPN) has been selected as the measurement index used to better classify failures and breakdowns. Hence, it is essential to develop an FMEA process for identifying the root causes of maintenance, so that the occurrence of the specific failure modes (FM) can be considered as the anomalies to be monitored. The FMEA analysis is summarized by the RPN index ($RPN = G \times P \times R$), indicating the occurrence risk for each failure mode according to its probability (P), severity (G) and detectability (R). Usually, an intervention is needed when the RPN value exceeds the threshold.

In this sense, an FMEA analysis was developed in order to identify any process FM. In this sense, all failure events data are reorganized, and an ID number of failure modes is identified. Indeed, the probability of an FM occurring is estimated by assessing the degree of risk through historical data provided by the company. Therefore, a failure mode in the present study scenario may have the following effects: products non-conforming, downtime, maintenance failure, poor production performance and costs. The severity indexes (G) of the failure mode are instead identified according to the time required for maintenance, the number of pieces produced and the importance of each workstation. Thus, MTTR, MTBF and MTMB values are calculated to estimate the severity index. On the other hand, about 1,000 product units are counted for the probability factor (P) assignment, calculating the frequency and probability of an event concerning the total number of failures recorded during about 180 working days. Hence, the RPN index provides the detectability factor of the failure modes. In this regard, a numerical limit has been defined, beyond which the identified criticality needs improvement interventions. The RPN values are thus classified as follows.

Traditional six big losses	Case-study specific losses	Time (h)
Breakdown loss	Faults solved AM	99.12
	Faults solved PM	374.5
Set up and adjustment loss	PM scheduled maintenance	35.08
	Tool regeneration	7.67
	AM calendars stop	0
	Technical cleaning	0.93
	Waiting PM intervention	0
	Stop automation exclusion	0
	Set-up	3.25
	Restart	0
	Reduced speed loss	Different cycle time from standard
Micro-stops		18.17
Idling minor stoppage	Emergency actions	0
	Intentional cycle stops	0
	Lack of personnel	0
	Lack of direct materials	8.15
	Power failure from the upstream location	20.18
	Lack of absorption from the downstream station	0
	Rework loss	Minutes production rework
Scrap or yield loss	Minutes production waste	11.76
	Slow down quality problems	0
	Stop quality problem	3

Table 5.
Classification of the
case-study-specific
losses

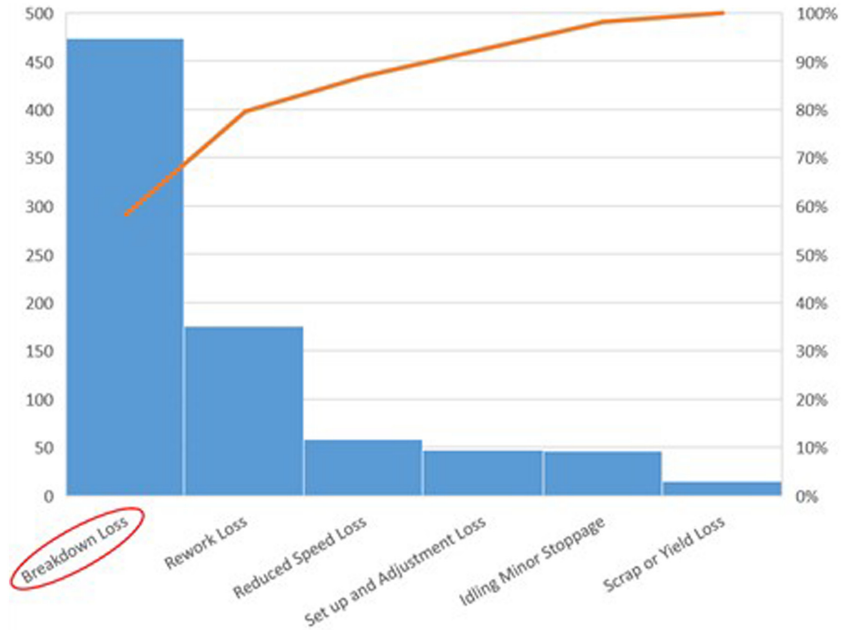


Figure 8.
Pareto diagram

Source(s): Author's own creation/work

- (1) Excellent-good values between 1 and 10
- (2) Good-sufficient values between 10 and 100
- (3) Sufficient-poor values between 100 and 1,000

Setting the upper limit of RPN to 100, the associated FM is identifiable as a significant anomaly. According to the results, 53% of the fault events showed a fairly high RPN compared to the threshold value and required further investigation in order to reduce the range (see Figure 9).

At this stage, a detailed analysis is thus performed through the ARM in order to identify relationships between the related failure events and select the group of critical FMs that require urgent and important improvement actions to prioritize interventions.

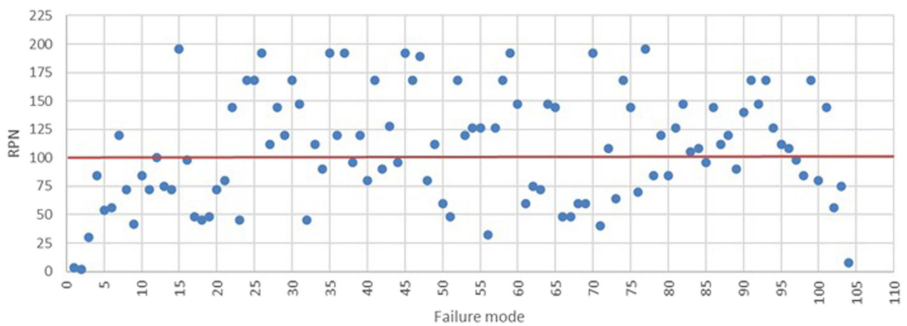


Figure 9.
RPN

Source(s): Author's own creation/work

The support of the rule is set at 0.3, indicating the statistical significance of the sample under consideration. As far as confidence is concerned, it was considered appropriate to choose a threshold value of 0.3, and all those rules with a confidence value lower than the threshold are not considered. A table with 141 rows representing tLD transactions as working days and 55 columns representing failure modes has been created. The association of Boolean variables, namely variables that assume TRUE/FALSE values, is then carried out. Specifically, value 1 indicates that the failure modes have occurred, 0 otherwise. The following association rules are achieved (see Table 6).

As described in Section 3.3.2, based on the integration of traditional methodologies such as FMEA, and ARs, all major FMs have been classified in the Eisenhower matrix in Figure 10.

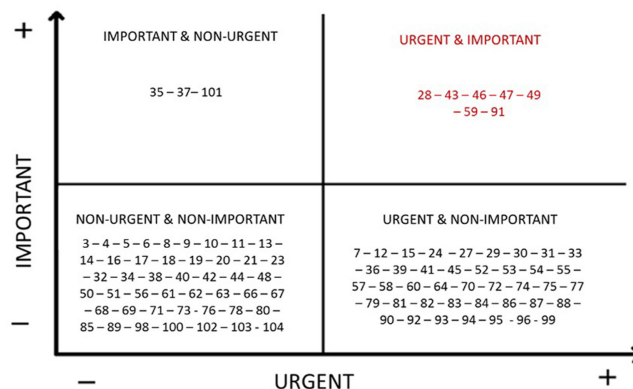
In the upper right area, urgent and important anomalies are prioritized, namely, those with an RPN index above the threshold revealed by the FMEA analysis and closely related to each other as revealed by the ARMs. Thus, the next stage is needed to analyze their relationships with production factors and identify proactive actions.

4.3 Knowledge extraction

The first step for knowledge extraction consists of mining the ARs relating the urgent and important failure modes with the corresponding OEE. In Table 7, an excerpt of the results is

X	Y	Supp (%)	Conf (%)
28	101	5	23
47	101	6	22
35	22	5	39
46	22	8	46
47	22	5	19
91	22	6	47
28	26	6	26
37	26	5	41
46	26	5	29
59	26	5	24
101	28	5	33
26	28	6	36

Table 6.
Some of the association
rules among failure
events with RPN above
the threshold value



Source(s): Author's own creation/work

Figure 10.
Failure modes
classification and
prioritization during
the first six months of
analysis

reported. An emblematic example is constituted by the occurrence of FM 91. It can be observed, in fact, that when the head of the rule is FM = 91, in the majority of cases the OEE is included in the range [0.333–0.667] since the confidence is 58.8%. The other two relevant cases, i.e. [0–0.333] and [0.667–1] both occur with a confidence of 17.6%. The occurrence of FM = 91 represents the only case in which the OEE belongs to the most critical case, i.e. OEE < 0.333.

Presenting the same results through a network analysis (Figure 11) this aspect is immediately evident since there are no other edges connecting the node OEE = [0–0.333]. Hence, preventing the occurrence of FM 91 will be a priority with a view to not penalizing the OEE.

In order to identify which are the OEE factors that are more impacted by the occurrence of the urgent and important FMs, as well as the availability, the relationships between them and Availability, Quality and Performance ranges are represented in Figure 12. As shown, the lowest range of OEE depends of course on the low value of Availability but also a little bit on Quality losses. In particular, FM 91 is the only one connected to the low value of Availability by the rule FM = 91 and Availability = [0–0.333]. Hence, it can be said that, in order to prevent the occurrence of FM 91, the preponderating factor to take into consideration is Availability with a focus on Quality, rather than Performance. In this sense, the improvement action undertaken, from an OEE continuous improvement perspective, will focus on better maintenance scheduling, for instance, programmed maintenance interval reduction.

Table 7.
Excerpt of the results
of ARs mining

X	Y	Supp	Conf
FM = 46	OEE = range3 [0.333–0.667]	0.028353326	0.565217
FM = 46	OEE = range4 [0.667–1]	0.019629226	0.391304
FM = 59	OEE = range3 [0.333–0.667]	0.011995638	0.323529
FM = 59	OEE = range4 [0.667–1]	0.023991276	0.647059
FM = 47	OEE = range4 [0.667–1]	0.022900763	0.552632
FM = 47	OEE = range3 [0.333–0.667]	0.014176663	0.342105
FM = 91	OEE = range3 [0.333–0.667]	0.010905125	0.588235
FM = 91	OEE = range4 [0.667–1]	0.003271538	0.176471
FM = 91	OEE = range2 [0–0.333]	0.003271538	0.176471

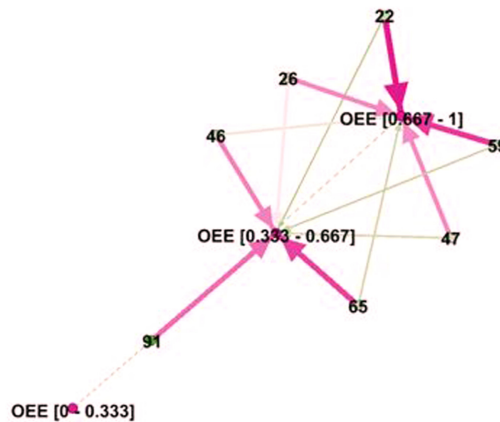
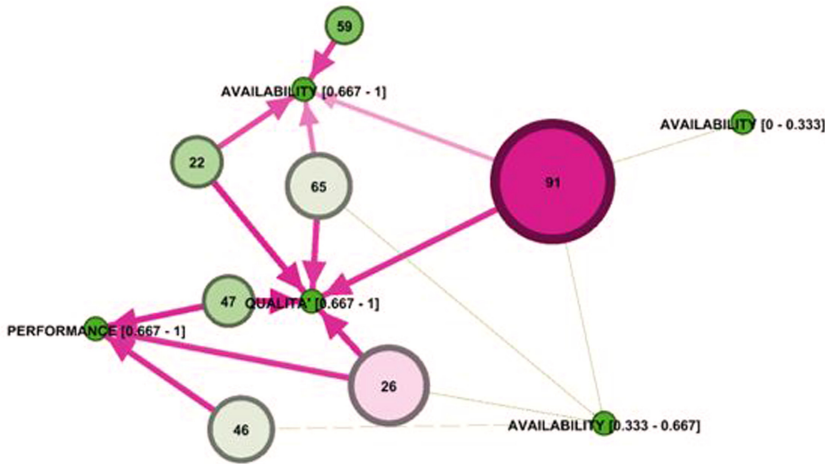


Figure 11.
Network relating the
ARs presented in
Table 4

Source(s): Author's own creation/work



Source(s): Author’s own creation/work

Figure 12. Investigation of the impact of the occurrence of the critical failure on the OEE factors

The defined solution has been implemented for six months and, in January 2020 during the check phase of the Deming Cycle, the results obtained were assessed: the number of FM 91 occurrences has been reduced by about 40%; the average OEE measured on the days in which FM 91 occurred has improved of 6.4%. This is also justified by the ARs reported in Table 8: when FM 91 occurs, indeed, in 20% of cases, the OEE is included in the range [0.667–1], in 70% of cases it belongs to [0.333–0.667]. The occurrence of FM 91 resulting in a low OEE (i.e. belonging to the range [0–0.333]) does not constitute relevant ARs, since its support is lower than the min_sup threshold. In order to assess the validity of the results, it has been verified that this event occurred once in the monitored period. The reduction of programmed maintenance can be considered validated and, hence, it will be applied even in the following period.

At this point, according to the presented research approach, a new “Plan” is carried out to define the new priority anomaly for addressing and continuing to solve them. At the end of the monitored time interval, i.e. July 2020, all the urgent and important anomalies have been addressed with appropriate improvement actions (Table 9). For each of them, the starting date and the variation of the OEE have also been reported, as well as the main OEE factor addressed by the selected action.

5. Discussion

Since the current paper represents a real case study, both theoretical and practical implications can be highlighted. From a theoretical perspective, the proposed methodology can be considered a novelty due to its accuracy in selecting and prioritizing anomalies by integrating traditional methodologies and ARM, then investigated in terms of OEE losses

X	Y	Supp	Conf
Failure Type/Downtime = FM 91	OEE = range4	0.001919	0.2
Failure Type/Downtime = FM 91	OEE = range3	0.006718	0.7

Table 8. Extraction of the ARs between FM 91 and OEE ranges

Table 9.
Improvement actions
proposed to reduce or
eliminate the most
critical failure modes of
the process

Anomaly code	Proactive action	Start of the proactive action	Influence on OEE factor	Δ OEE (at validation)
28	Production line refilling improvement	July 2019	Availability and Performance	+7.2%
43	Substitution of robot sensors	July 2019	Availability	+7.3%
46	Additional tool insertion	January 2020	Availability and Performance	+3.1%
47	Workplace reorganization	January 2020	Availability and Performance	+5.4
49	Robot picking strategy review	January 2020	Availability	+5.5
59	Programmed maintenance interval reduction	January 2020	Availability	+5.8%
91	Programmed maintenance interval reduction	July 2019	Availability and Quality	+6.4%

through ARM and NA; while others in the literature are focused on using just traditional methodologies (such as the Pareto diagram) for anomaly selection, then investigating OEE losses through ARM (Djatna and Alitu, 2015). However, when dealing with high amount of data, a thorough selection and analytics is required not to lose relevant information.

In addition, the current approach deals with the identification of anomalies most influencing the whole loss factors of OEE, while most of the literature just focuses on the widely investigated Availability losses (El kihel *et al.*, 2022). In particular, focusing on all loss factors of OEE, i.e. Availability, Performance and Quality, allows for a broader view of the overall anomalies of the process.

On the other hand, data homogeneity is a critical issue due to data not being collected in real time. The pre-processing phase, especially during standardization, is quite time-consuming and can influence data veridicality. For this reason, a periodic review of the data homogeneity level is needed in order to limit the propagation of misinformation.

From a practical point of view, the novel application of ARM in two different steps and ways within the same data-driven methodology of traditional approaches represents a promising approach to handling anomalies in manufacturing. In particular, the proposed methodology provides.

- (1) A specific rule-based ML framework for knowledge extraction regarding priority anomalies, their co-occurrences and their major influence on the OEE loss factors.
- (2) The definition and implementation of improvement actions from a proactive philosophy perspective.
- (3) The continuous data analytics and framework application based on the Deming Cycle implementation.

It should be noted that the proposed methodology will provide better improvements if data are continuously updated. Hence, actions on anomalies can be taken frequently based on the following concept "A huge number of small improvements are more effective than a few improvements of large value" (Singh *et al.*, 2013).

Nevertheless, during the implementation phase, one of the main criticisms is related to usability due to the difficulty of operators in completely understanding the methodology. In order to solve this issue, good training of the operators is needed to make them independent in the application. However, an accurate selection of trainers by the company is required to be as effective as possible in terms of practical results.

6. Conclusions and future developments

Machine learning approaches can support the maintainer's decision-making process in managing a large amount of data and reducing the occurrence of anomalies in a manufacturing process, influencing equipment effectiveness. When referring to anomalies, existing contributions often refer only to failure analysis thus showing particular attention to indicators such as MTBF, asset performance, reliability and predictive maintenance actions.

In this work, an innovative perspective is provided due to all anomalies that may affect OEE are considered, as well as improvement actions affecting more than just predictive maintenance. To this end, a novel methodology for the deeper analysis of the OEE factors is provided by combining traditional techniques (such as the Pareto diagram and Deming Cycle) with a double application of the Association Rule Mining and Network Analysis in different steps of the methodology. Based on the Deming Cycle perspective, the Plan phase consists of a rule-based framework application for knowledge extraction about the occurrence impact of priority anomalies on OEE. In this regard, the framework proposed is organized into three steps, respectively referred to as data acquisition and management, data analytics and knowledge extraction.

The first goal of the current approach is to provide a decision-making strategy in order to help operators leverage all available data. Then, the major dominant anomalies analysis and their prioritization both through association rule mining and network analysis. A real application of the rule-based whole methodology is proposed by implementing the real solution and providing a real evaluation of the methodology and its impact on the OEE. More in details, the results of the framework's implementation in an automotive company highlighted that the breakdown losses represent the major cause impacting the OEE, with specific reference to the availability factor. FMEA was carried out and its results were analyzed to identify the failure modes frequently occurring together, as well as the most critical ones in terms of impact on the OEE. Jointly implementing association rule mining and network analysis provided a clear vision of the interrelations hidden in data, as well as supporting the definition of improvement actions. Following a Deming Cycle perspective, the Plan phase can be recognized in the definition of the improvement actions for OEE loss reduction, which have been implemented in the Do phase, and validated in the Check and Act phases.

Due to the OEE being considered the major key performance indicator in the Total Productive Maintenance (TPM) philosophy, further, development involves the expansion of the methodology from a TPM perspective also including the training of operators who are missing in the current methodology. An innovative implementation for the actual case depends greatly on the needs of the company, however, a new methodology can certainly be implemented theoretically.

Taking into account the transition from a Pre-Covid to a Post-Covid industrial environment should also involve careful consideration of factors supporting safety, efficiency and adaptability of operations. Through the proposed framework, remote monitoring and the ability to automate data collection and analytics phases surely represent valuable support in case of operations disruption since maintenance managers can observe the collected data and analytics results remotely and make data-driven decisions accordingly. In this way, process and maintenance schedules can be optimized, and failures and downtimes can be sensibly reduced, proactively decreasing the need for reactive, on-site interventions.

References

Abdelrahman, P. and Keikhosrokiani, O. (2020), "Assembly line anomaly detection and root cause analysis using machine learning", *IEEE Access*, Vol. 8, pp. 189661-189672, doi: [10.1109/access.2020.3029826](https://doi.org/10.1109/access.2020.3029826).

- Antomarioni, S., Bellinello, M.M., Bevilacqua, M., Ciarapica, F.E., Favarão da Silva, G.F. and Renan Martha de Souza (2020), "A data-driven approach to extend failure analysis: a framework development and a case study on a hydroelectric power plant", *Energies (Basel)*, Vol. 13 No. 23, p. 6400, doi: [10.3390/en13236400](https://doi.org/10.3390/en13236400).
- Antomarioni, S., Ciarapica, F.E. and Bevilacqua, M. (2022), "Data-driven approach to predict the sequence of component failures: a framework and a case study on a process industry", *International Journal of Quality and Reliability Management*, Vol. 40 No. 3, pp. 752-776, doi: [10.1108/IJQRM-12-2020-0413](https://doi.org/10.1108/IJQRM-12-2020-0413).
- Arun Prasad, G.K. and Panse, C. (2022), "Predictive maintenance in forging industry", *Proceedings of 2nd International Conference on Innovative Practices in Technology and Management*, pp. 794-800.
- Badiger, A.S. and Gandhinathan, R. (2008), "A proposal: evaluation of OEE and impact of six big losses on equipment earning capacity", *International Journal of Process Management and Benchmarking, Inderscience Enterprises*, Vol. 2 No. 3, pp. 234-248.
- Black, J.H. (2014), "A new Era in mill analytics: easy and effective ways to improve production management decisions", *Paper Conference and Trade Show*, pp. 410-418.
- Brodny, J. and Tutak, M. (2018), "Analysis of availability of longwall-shearer based on its working cycle", *IOP Conference Series: Earth and Environmental Science*, doi: [10.1088/1755-1315/95/4/042020](https://doi.org/10.1088/1755-1315/95/4/042020).
- Brunelli, L., Masiero, C., Tosato, D., Beghi, A. and Susto, G.A. (2019), "Deep learning-based production forecasting in manufacturing: a packaging equipment case study", *Procedia Manufacturing*, Vol. 38, pp. 248-255, doi: [10.1016/j.promfg.2020.01.033](https://doi.org/10.1016/j.promfg.2020.01.033).
- Buddhakulsomsiri, Y., Siradeghyan, J., Zakarian, A. and Li, X. (2006), "Association rule-generation algorithm for mining automotive warranty data", *International Journal of Production Research*, Vol. 44 No. 14, pp. 2749-2770, doi: [10.1080/00207540600564633](https://doi.org/10.1080/00207540600564633).
- Calabrese, A., Dora, M., Levaldi Ghiron, N. and Tiburzi, L. (2022), "Industry's 4.0 transformation process: how to start, where to aim, what to be aware of", *Production Planning and Control*, Vol. 33 No. 5, pp. 492-512, doi: [10.1080/09537287.2020.1830315](https://doi.org/10.1080/09537287.2020.1830315).
- Chan, F.T.S., Lau, H.C.W., Ip, R.W.L., Chan, H.K. and Kong, S. (2005), "Implementation of total productive maintenance: a case study", *International Journal of Production Economics*, Vol. 95 No. 1, pp. 71-94, doi: [10.1016/j.ijpe.2003.10.021](https://doi.org/10.1016/j.ijpe.2003.10.021).
- Chong, K.E., Ng, K.C. and Goh, G.G.G. (2016), "Improving overall equipment effectiveness (OEE) through integration of maintenance failure mode and effect analysis (maintenance-FMEA) in a semiconductor manufacturer: a case study", *IEEE International Conference on Industrial Engineering and Engineering Management*, pp. 1427-1431.
- Costa, R.D.F.S., Barbosa, M.L.S., Silva, F.J.G., Sá, J.C., Ferreira, L.P. and Pinto, B. (2022), "Improving procedures for production and maintenance control towards industry 4.0 implementation", *2nd International Conference Innovation in Engineering, ICIE 2022*, pp. 58-67, doi: [10.1007/978-3-031-09382-1_6](https://doi.org/10.1007/978-3-031-09382-1_6).
- Deeskow, P., Steinmetz, U. and Hay, M. (2008), "Data mining and statistical process control for condition bases maintenance", *VGB PowerTech*, Vol. 88 No. 10, pp. 84-87+6-7.
- Djatna, T. and Alitu, I.M. (2015), "An application of association rule mining in total productive maintenance strategy: an analysis and modelling in wooden door manufacturing industry", *Procedia Manufacturing*, Vol. 4, pp. 336-343, doi: [10.1016/j.promfg.2015.11.049](https://doi.org/10.1016/j.promfg.2015.11.049).
- Dobra, P. and Jósvali, J. (2020), "Enhance of OEE by hybrid analysis at the automotive semi-automatic assembly lines", *Procedia Manufacturing*, Vol. 54, pp. 184-190, doi: [10.1016/j.promfg.2021.07.028](https://doi.org/10.1016/j.promfg.2021.07.028).
- El kihel, E.M., El kihel, Y. and Bouyahrouzi, A. (2022), "Contribution of maintenance 4.0 in sustainable development with an industrial case study", *Sustainability*, Vol. 14 No. 17, p. 11090, doi: [10.3390/su141711090](https://doi.org/10.3390/su141711090).

-
- Gökalp, E., Umut, Ş. and Eren, P.E. (2017), "Development of an assessment model for industry 4.0: industry 4.0-MM", *International Conference on Software Process Improvement and Capability Determination*, pp. 128-142.
- Han, J., Cheng, H., Xin, D. and Yan, X. (2007), "Frequent pattern mining : current status and future directions", *Data Mining and Knowledge Discovery*, Vol. 15 No. 1, pp. 55-86, doi: [10.1007/s10618-006-0059-1](https://doi.org/10.1007/s10618-006-0059-1).
- Jasiulewicz-Kaczmarek, M. and Katarzyna, A. (2022), "Industry 4.0 technologies for maintenance management – an overview", *2nd International Conference Innovation in Engineering, ICIE 2022*, doi: [10.1007/978-3-031-09382-1_7](https://doi.org/10.1007/978-3-031-09382-1_7).
- Jeong, C., Yu, Y., Patino, D., Venkatakrishnan, D. and Mansour, S. (2022), "Behavior anomalies detection in drilling time series through feature extraction", *SPE - International Association of Drilling Contractors Drilling Conference Proceedings*, doi: [10.2118/208676-MS](https://doi.org/10.2118/208676-MS).
- Jittawiriyankoon, V. and Srisarkun, C. (2022), "Simulation for predictive maintenance using weighted training algorithms in machine learning", *International Journal of Electrical and Computer Engineering*, Vol. 12 No. 3, pp. 2839-2846, doi: [10.11591/ijece.v12i3.pp2839-2846](https://doi.org/10.11591/ijece.v12i3.pp2839-2846).
- Jyothi, N.S. and Parkavi, A. (2016), "A study on task management system", *International Conference on Research Advances in Integrated Navigation Systems, RAINS 2016*, doi: [10.1109/RAINS.2016.7764421](https://doi.org/10.1109/RAINS.2016.7764421).
- Liu, N., Hu, M., Wang, J., Ren, Y. and Tian, W. (2022), "Fault detection and diagnosis using Bayesian network model combining mechanism correlation analysis and process data: application to unmonitored root cause variables type faults", *Process Safety and Environmental Protection*, Vol. 164, pp. 15-29, doi: [10.1016/j.psep.2022.05.073](https://doi.org/10.1016/j.psep.2022.05.073).
- Maimon, O. and Rokach, L. (2010), "Introduction to knowledge discovery and data mining", in *Data Mining and Knowledge Discovery Handbook*, pp. 1-15.
- Márquez, A.C., de la Fuente Carmona, S. and Antomarioni, A. (2019), "A process to implement an artificial neural network and association rules techniques to improve asset performance and energy efficiency", *Energies (Basel)*, Vol. 12 No. 18, p. 3454, doi: [10.3390/en12183454](https://doi.org/10.3390/en12183454).
- Otte, E. and Rousseau, R. (2002), "Social network analysis: a powerful strategy, also for the information sciences", *Journal of Information Science*, Vol. 28 No. 6, pp. 441-453, doi: [10.1177/01655515022800601](https://doi.org/10.1177/01655515022800601).
- Ruschel, E., Santos, E.A.P. and Loures, E.F.R. (2020), "Establishment of maintenance inspection intervals: an application of process mining techniques in manufacturing", *Journal of Intelligent Manufacturing*, Vol. 31 No. 1, pp. 53-72, doi: [10.1007/s10845-018-1434-7](https://doi.org/10.1007/s10845-018-1434-7).
- Setyawan, W., Sutoni, A., Munandar, T. and Mujiarto (2021), "Calculation and analysis of overall equipment effectiveness (OEE) method and six big losses toward the production of corter machines in Oni Jaya motor", *Journal of Physics: Conference Series*, Vol. 1764 No. 1, p. 012162, doi: [10.1088/1742-6596/1764/1/012162](https://doi.org/10.1088/1742-6596/1764/1/012162).
- Singh, R., Gohil, A.M., Shah, D.B. and Desai, S. (2013), "Total productive maintenance (TPM) implementation in a machine shop: a case study", *Procedia Engineering*, Vol. 51, pp. 592-599, doi: [10.1016/j.proeng.2013.01.084](https://doi.org/10.1016/j.proeng.2013.01.084).
- Taghavi, V., Vahidtaghaviensetsmtlca, E. and Beauregard, Y. (2020), "The relationship between lean and industry 4. 0: literature review", *5th North American Conference on Industrial Engineering and Operations Management*, pp. 808-820.
- Utz, F., Neumann, C. and Tafreschi, O. (2018), "How to discover knowledge for improving availability in the manufacturing domain?", *Proceedings of the Annual Hawaii International Conference on System Sciences*, pp. 4380-4389.
- Wu, Z., Liu, W. and Nie, W. (2021), "Literature review and prospect of the development and application of FMEA in manufacturing industry", *International Journal of Advanced Manufacturing Technology*, Vol. 112 Nos 5-6, pp. 1409-1436, doi: [10.1007/s00170-020-06425-0](https://doi.org/10.1007/s00170-020-06425-0).

- Yu, C.-M., Kuo, C.-J., Chiu, C.-L., Wen, W.-C. and Zhang, M. (2019), "Unveil the black box for performance efficiency of OEE for semiconductor wafer fabrication", *IEEE International Symposium on Semiconductor Manufacturing Conference Proceedings*, doi: [10.1109/ISSM.2018.8651146](https://doi.org/10.1109/ISSM.2018.8651146).
- Zhang, K., Shardt, Y.A.W., Chen, Z., Yang, X., Ding, S.X. and Peng, K. (2018), "A KPI-based process monitoring and fault detection framework for large-scale processes", *ISA Transactions*, Vol. 68, pp. 276-286, doi: [10.1016/j.isatra.2017.01.029](https://doi.org/10.1016/j.isatra.2017.01.029).
- Zheng, T., Ardolino, M., Bacchetti, A. and Perona, M. (2021), "The applications of Industry 4.0 technologies in manufacturing context: a systematic literature review", *International Journal of Production Research*, Vol. 59 No. 6, pp. 1922-1954, doi: [10.1080/00207543.2020.1824085](https://doi.org/10.1080/00207543.2020.1824085).

Corresponding author

Laura Lucantoni can be contacted at: l.lucantoni@pm.univpm.it