ORIGINAL ARTICLE

Operations-aware novelty detection framework for CNC machine tools: proposal and application

Marco Macchi[1] · Laura Cattaneo[2] · Adalberto Polenghi[1] · Danilo Bruno[3] · Sofia Mosci[3] · Fabio Ferriani[4] · Roberto Caneva[4]

Abstract
Digitisation offers manufacturing companies new opportunities to improve their operations and competitiveness in the market by unleashing potentialities related to real-time monitoring and control of operating machines. Through condition-based and predictive maintenance, the knowledge about the health state and probability of failure of the machines is improved for better decision-making. Amongst them, CNC machine tools do represent a complex case from a maintenance viewpoint as their operations are ever-changing and their high reliability brings to a lack, or limited set, of run-to-failure data. To address the problem, the research work proposes an operations-aware novelty detection framework for CNC machine tools based on already-in-place controllers. The framework is based on statistical modelling of the behaviour of the machine tools, namely through gradient boosting regression and Gaussian mixture models, to identify the health state considering varying operations through time. The proposed solution is verified on sixteen multi-axis CNC machine tools in a large manufacturing company. The results show that the proposed solution can effectively support maintenance decisions by informing on the health states while discerning between varying operations and abnormal/faulty states of interest. This solution represents a brick in a cloud-edge-based industrial information system stack that can be further developed for shop floor-integrated decision-making.

Keywords Novelty detection · Condition-based maintenance · Prognostics and health management · Machine tool · CNC · Manufacturing

1 Introduction

The digital transformation is bringing new opportunities to improve the business and operations of industrial companies. New business models, better engagement of suppliers and customers along the supply chain as well as enhanced operations management are some of the benefits companies could gather thanks to digitisation [1, 2]. Especially, on the shop floor, new ways for production and maintenance management are offered by advanced sensing solutions empowered by data analytics to extract useful information [3]. This, in turn, becomes essential to improve decision-making processes that are no more only reactive, but rather proactive so as to anticipate future possible scenarios and promptly adjust

✉ Marco Macchi
marco.macchi@polimi.it

✉ Adalberto Polenghi
adalberto.polenghi@polimi.it

Laura Cattaneo
lcattaneo@liuc.it

Danilo Bruno
danilo.bruno@camelotbio.com

Sofia Mosci
sofia.mosci@camelotbio.com

Fabio Ferriani
fabio.ferriani@ext.ansaldoenergia.com

Roberto Caneva
roberto.caneva@ansaldoenergia.com

1   Department of Management, Economics and Industrial Engineering, Politecnico di Milano, via Lambruschini 4/b, 20156 Milan, Italy

2   School of Industrial Engineering, Università Carlo Cattaneo – LIUC, C.so Matteotti 22, 21053 Castellanza, Italy

3   CAMELOT Biomedical Systems Srl, Via al Ponte Reale 2/20, 16124 Genova, Italy

4   Ansaldo Energia Spa, Via Nicola Lorenzi 8, 16152 Genova, Italy

the current situation, if necessary [4]. Specifically, maintenance is perceived as one of the most impacted areas by digitisation, given the shifts from corrective and time-based maintenance to condition-based and predictive maintenance [5]. These offer the possibility to have insights into the current working state of an operating machine, its health state and the remaining useful life (RUL) [6]. This information finally leads to improve maintenance performances in terms of reduced downtimes, improved availability, increased product quality as well as production performances, adapting to contingent situations [7, 8]. The body of knowledge that collects all the steps (referred also as "levels") necessary to build a system able to bring these benefits is often referred to as Prognostics and Health Management (PHM) [9]. As it is well-established, PHM can build on standards, such as the ISO 13374-1 [10], to determine a methodological approach for the PHM program development.

The cornerstones of PHM reside in suitable sensing technologies and data analytics, both fit for purpose. Sensors and transducers must be installed and configured to collect condition monitoring data from the machine components, which is key to improving knowledge about failure modes the machine may experience. Furthermore, work orders and manually inputted data are essential to combine condition monitoring data from the field with events recorded by operators. Once the data are available, proper multi-nature algorithms should be put in place to transform raw data into useful information for decision-making. Algorithms may be either physics-based, artificial intelligence (AI) or statistical [11]; recently, also hybrid models, mainly based on physics-informed AI algorithms, are gaining momentum [12].

Notwithstanding its potentialities, PHM suffers from some issues. Firstly, PHM program development should exploit run-to-failure (RTF) data: these mainly refer to event data that allow labelling condition monitoring data with respect to the real states of the operating machine, finally distinguishing faulty states (complete or partial faults) from healthy state; nevertheless, this potentiality is limited when RTF data are not, or scarcely, available [13, 14]. Besides, even condition monitoring data may be not available if a proper monitoring system fit for the purpose of maintenance decision-making is not put in place, or a monitoring campaign has not yet been established [15]. A further issue relates to the varying working conditions that may result from the intrinsic flexibility of the operations made by the operating machine, which could undermine the reliability of the information provided by PHM tasks [16].

All these issues are experienced in the case of extant CNC machine tools. Indeed, CNC machine tools represent a complex case where to apply PHM programmes, for different reasons [17–19]: (i) CNC machine tools are an ensemble of electro-mechanical and hydraulic systems that interact with each other; (ii) the part programmes CNC machine tools perform are time-varying in the light of the flexibility of the operations required to make multiple product types; (iii) CNC machine tools are highly reliable, which limits the availability of RTF data; and (iv) an integrated condition monitoring system fit for the purpose of maintenance decision-making is not originally in place in extant machines. These characteristics lead to a challenging PHM application, to be developed considering the following requirements: PHM solutions should take into account the interactions between subsystems and components of the machine tools in order to improve the maintenance decision-making [17]; PHM solutions should be able to adapt to ever-changing working conditions dictated by new part programmes introduced to satisfy continuously evolving production requirements [20]; and PHM solutions should be capable to cope with any limitations concerning RTF and condition monitoring data for maintenance decisions [14].

Overall, the reduction of options in terms of AI and statistical algorithms is a consequence, and this research work proposes to employ supervised learning models within the novelty detection framework, leaving aside diagnostic and prognostic analyses.

1.1 Research scope and objective

The research scope this work addresses with peculiar consideration is the so-called state detection (SD) level of the PHM process, where machine states need to be identified. In this scope, the assumption is to rely only upon the data fed by the operations run by the machine tool and made available by means of controllers (such as CNC or PLC) or sensors and transducers already-in-place for the machining process, without any need of integration with a condition monitoring system fit for maintenance tasks such as diagnostics or prognostics. Therefore, at the SD level, the research objective this work pursues is the development of a novelty detection (ND) framework for CNC machine tools enabling three major tasks: to identify when a new operation is carried out featuring different working conditions, to recognise if a failure is going to occur by means of the detection of an anomaly (*alias* partial fault), and, eventually, to classify the health state of the machine, either being in healthy or unhealthy state.

In detail, the developed solution should consider the capability to recognise ever-changing operations and to guarantee ND adequate for the maintenance decision-maker. The industrial viability is demonstrated through the application to a fleet of sixteen multi-axis CNC machine tools of a large manufacturing company in Italy. Indeed, the challenge faced in this research work is to improve the capabilities of old, less sensitised CNC machine tools so as to embrace the Machine tool 4.0 paradigm [21].

The remainder of the paper is so organised: section 2 revises current PHM approaches applied to CNC machine tools, giving a particular emphasis to ND; section 3 depicts the proposed solution, section 4 introduces the case company and section 5 describes the application to the case; section 6 then provides discussion and lessons learnt from the industrial experience; finally, section 7 draws some conclusions and envisions future works.

2 PHM program and novelty detection for CNC machine tools

The advent of PHM has been promoted by the availability of sensors and increased computational capability, enabling a shift from "fail-and-fix" to "predict-and-prevent" [22], whose development is strongly based on the ISO 13374-1 [9]. This norm structures the process into six levels: data acquisition, data manipulation, state detection, health assessment, prognostic assessment and advisory generation.

The applicability of the PHM process is independent of the specific asset or component under analysis, showing validity throughout different indenture levels [11, 23–25], including the equipment/asset itself, like a CNC machine tool, a centrifugal pump, or an industrial fan, and the subsystems/components, such as bearings, drive belts, or rotors. Moreover, the reviews propose specific insights on each level of the ISO 13374-1, sometimes showing different but compatible aggregation of the levels.

2.1 Role of novelty detection in the PHM program

ND can be defined as the "task of recognising that test data differ in some respect from the data that are available during training" [26]. With respect to the PHM process, as defined by ISO 13374-1, this task applies to the SD level. The states in which the machines could be are identified and then, during machine operations, if new data are significantly different from previously identified patterns, a "novelty" is alerted so that maintenance technicians and managers could take proper actions.

The importance of the SD level (and ND) in PHM program development is recognised for diagnostic as well as prognostic purposes:

- ND is relevant for diagnostics since the identified new pattern of data, i.e. a novelty, enables the diagnosis of what has happened [27, 28]. Also, in case of missing failure data, ND and abnormal state labelling almost coincide as anything new is a potentially damaging state [14].
- ND is necessary also for prognostics since the variables' prediction and related RUL estimation, highly depend on the current state information; prediction may be triggered as soon as a novelty comes across so as to identify future trends [11, 29].

Overall, ND represents a critical point in the development of effective PHM programmes since diagnostics and prognostics analyses are enabled by the identification of the actual states in which the equipment/asset or related subsystem/component is.

2.2 Novelty detection applied to CNC machine tools

A pool of scientific articles published in the last few years is hereinafter reported in order to shed light on the research trend on ND applied to CNC machine tools; the pool excludes those works focused on diagnosis and prognosis only.

Recently, in Schlagenhauf and Burghardt [30], an expert system is realised with the aim of detecting defects on a ball screw drive through convolutional neural network (CNN) on images. Also, the authors realised a prediction model able to forecast the area of the defect so as to anticipate future evolution. Also in Denkena et al. [31], ball screw anomalies are identified through a semi-supervised ND algorithm based on an ensemble approach, built on local outlier factor, k-nearest neighbours, angle-based outlier detection (LOF-kNN-ABOD). These are selected since no previous assumption about data distribution is needed. Experiments are carried out to understand the variation of performance of the ensemble approach with a varying set of features from the measured variables, and the results show that frequency domain-related features outperform time domain ones for ball screws condition monitoring.

Another relevant component for CNC machine tools, and in general for any machine including rotating parts, is the bearing. In the research by Wu et al. [32], bearing faults are analysed. To this end, CNN is used to detect and classify bearings faults together with variational mode decomposition (VMD) and autocorrelation function (ACF) as they allow for the reduction of the number of layers in the CNN. In Goyal et al. [33], kNN and WkNN (weighted kNN) are used to detect and diagnose bearings faults using non-contact laser-based vibration sensors instead of accelerometers. The work supports that vibration remains one of the main diagnostic signals for CNC machine tools even though improvements are foreseen both on the side of algorithms and on the physical sensors.

Regarding the hydraulic subsystem, Bernini et al. [34] set up an experimental campaign based on a digital twin that reproduces the behaviour of the targeted CNC machine tool subsystem to generate healthy and faulty data. A pool of several algorithms is used to detect, diagnose and predict hydraulic subsystem faults, like CNN, random forest classifier (RCF), quadratic discriminant analysis (QDA)

and AdaBoost classifier. (ADB). The application shows that there is no unique answer for ND, but some algorithms perform better than others for different components: QDA for a pump, and CNN for sensors and valve faults.

Instead of relying on AI for ND, Du et al. [35] adopt proportional hazards modelling to model failure events and condition monitoring data, namely vibration, for electrospindle. Proportional hazards modelling allows extracting health indicators that are controlled in real-time thanks to statistical process control (SPC) to trigger maintenance actions in case of anomalies.

As one of the most critical components, the spindle has received a lot of attention in CNC machine tools as it guarantees product quality [36]. In this regard, many attempts were made to improve monitoring and controlling of the spindle for chattering, tool wear prediction and thermal error compensation [37]. Vibration monitoring is widely adopted both in continuous and discrete checking modes, to later compare with predefined thresholds defined as a function of operations requirements or manufacturers' manual [38]. Recent reviews show current advancements and future outlooks for spindle-related applications. For example, a comprehensive review of AI and ML has been developed by Soori et al. [39], while the review by Kuntoğlu et al. [40] investigated which signals are currently used for advanced applications, especially for spindle monitoring.

Overall, the picture resulting from the selected scientific literature shows that CNC machine tools are widely addressed in all their units and components, ranging from ball screw and bearing diagnostics to the identification of anomalies for the hydraulic system and machining unit, focusing on the spindle. Nonetheless, most works rely on laboratory experiments so as to have the entire set of faults available, resulting in many applications being supervised in nature. In case these data are available, the proposed works enable deep investigation on the health state of the CNC machine tools, with information related to possible causes (diagnostics). On the other hand, in case of missing a priori knowledge, there are still gaps to overcome, with some tentative to transfer knowledge from one CNC machine tool to another, as in the work by Denkena et al. [41], which is focused on how to transfer ND thresholds specific for a certain CNC machine tool to similar machines

without the need to perform any simulation. Nonetheless, differences between machines like different operations to perform and different signal properties prevent success at the current state-of-the-art. Indeed, the review by Baur et al. [42] confirms the above-mentioned trends; they made clear that at the moment the implementation of PHM in all or some of its steps is subordinated to the knowledge of faults, the machine experiences over its history. This somehow prevents the full exploitation of such methods in industrial environments if no prior knowledge of faults and related measurements is available.

Table 1 summarises the literature findings, synthesising some relevant insights, grouped by relevant meta-variables, namely the indenture level targeted by the ND solutions, the models adopted in such solutions, and the application contexts where the solution is tested.

Overall, several are the research streams and approaches for ND, and PHM in general, regarding CNC machine tools. Firstly, given the complexity of such machines, specific subsystems, such as hydraulic or electrical subsystems, and components, such as bearings and ball screws, are usually analysed one at a time. There are efforts in relating failures of different subsystems/components, but this research stream is at its early stage [43]. Secondly, feature extraction and selection are vital to guarantee the downstream algorithms so that ND performs well. Hence, signals, especially vibrations, are transformed in time, frequency and time-frequency domains, and features are selected based on the current fault to be detected [44]. Lastly, AI algorithms are currently playing the lion's share for ND of machine tools. Specifically, CNN is often engaged due to its ability to approximate even complex phenomena without the need to know the underlying failure mechanism [45]. Ensemble and hybrid approaches are also used in order to combine the advantages of multiple algorithms and outperform the single, standalone application [17].

2.3 Concluding remarks and gaps

Despite the promising trends and approaches to ND evident in literature findings, there are some gaps that are needed to be filled in. Amongst the gaps, the following appear the most relevant to remark:

Table 1 Summary of selected literature findings

Meta-variable	Finding
Indenture level	Developed ND solutions work at subsystem or component as indenture level; also, the influence of faults between multiple subsystems/components is not well studied
Adopted model	Current ND solutions embrace both statistical and AI algorithms; however, as the phenomenon's complexity increases, most authors look for AI as a way to approximate the failure mechanism
Application context	Usually, ND solutions are developed and tested in controlled environment, as laboratories, or are developed based on numerically generated dataset via multi-physics simulation

1. AI algorithms, such as CNN, require a huge amount of data to be properly trained. However, most of the time, RTF data are not available in the industry [13] and not affordable to be replicated in practice or generated by simulations [34]. Then, opportunities offered by AI algorithms appear at least challenging in their full realisation when data are scarcely available.

2. There are few works that consider the wide set of available data from CNC machine tools. Condition monitoring data such as vibrations and temperatures are the most used, but CNC and PLC could provide valuable information regarding the part program and tool number that allow for improving algorithm performance [46, 47]. Thus, there is a lack of emphasis in the extant scientific literature on how to better tune PHM thanks to the knowledge of the working conditions of the machines.

3. The industrial viability of proposed solutions needs to be better investigated since, at present, datasets are generated numerically through multi-physics simulation or from laboratory tests. Few are applications considering CNC machine tools working in a real industrial environment.

4. Most of the work concentrates on diagnosis, usually referred to as fault detection and diagnosis (FDD) rather than prognosis. Indeed, given labelled or unlabelled data, many studies focus on the application of supervised or unsupervised approaches to identify which failure modes are evolving or which failure has been experienced. Prognostic assessment may be empowered in future developments.

In this research work, the focus is on the first three gaps, while the last is out of scope, but worth for the future scientific effort. Amongst the focused gaps, it is worth remarking that the industrial viability of the method is a key feature addressed by this work. Firstly, it relates to the need to demonstrate implementation of PHM in CNC machine tools; secondly, it implies the adoption of a real setting in the actual evolution of the production requirements, enabling an industrial testbed inclusive of the variations of the actual working conditions of the machines, which is a key item under concern in the present work.

3 Proposal of the operations-aware novelty detection framework for CNC machine tools

The evidence from the literature shows several approaches to ND for CNC machine tools, to further enable diagnostic and prognostic analyses. However, what emerged is that there are few examples that rely on CNC and PLC data only and that show the industrial viability of the proposed solutions, including other aspects rather than model performance, like reliability and usability of the provided information to the maintenance decision-makers. Therefore, stemming from the gaps identified in the scientific literature and according to the authors' experience in industrial contexts, the goal of this research work is to realise a framework with associated models, organised in an integrated, industrially viable solution for ND of CNC machine tools. This research considers CNC and PLC data also to tune the PHM with the working conditions of the machines, besides supporting the ND task aimed at fault detection as required by maintenance decision-making. It is worth recalling that diagnostic and prognostic analyses are considered out of scope in this research work and the proposed operations-aware ND framework, whereas fault detection is addressed with the main purpose to support information for the decision-maker that an anomaly arose and a functional failure is then incipient.

To tackle every aspect of the proposed solution, the remainder of the section is organised as follows: subsection 3.1 describes the proposed framework, the ground theory it refers to, the overall structure and the peculiarities that make it novel with respect to the extant state of art solutions; subsections 3.2 and 3.3 deal with the SD level of PHM, clearing out the novel part of it; subsection 3.4 presents the health assessment and related health indicator definition; and finally, subsection 3.5 depicts the working flowchart of the framework that enables it to be operative in an industrial environment.

3.1 Overview of the framework

The ground theory of the proposed framework is provided by the ISO 13374-1 [10], which allows structuring the solution around an agreed-upon flow to realise the PHM program. The hypotheses to which the framework and the related algorithms are constrained are listed hereinafter as long as the effects on the modelling decisions.

- RTF data are assumed not to be available (or to be scarcely available). It prevents embracing pure supervised learning built upon failure labels, as failures are not (or, at least, are poorly) recorded. Then, specifically for ND, this learning is sometimes referred to as "semi-supervised" [26] as, on the one hand, it leverages upon a kind of unsupervised modelling as there are no pre-established labels, but, on the other hand, the normality is assumed to represent the healthy state, which enables a labelling process, therefore a kind of supervised modelling. Within the ND framework, specific models—gradient boosting regression and Gaussian mixture models—are then implemented.

- No condition monitoring data are adopted at the SD level, as it is assumed a limited machine retrofitting, without the need to integrate a condition monitoring system fit for maintenance purposes; hence, "traditional" techniques for CNC machine tools, such as vibration analysis, are not applied.
- Multiple machining operations are performed by the CNC machine tool and the entire set of part programmes is supposed not to be entirely known a priori as product requirements may change; therefore, the developed models must consider the variability of part programmes to be realised so as to provide reliable information to maintenance decision-makers.
- CNC and PLC data are available so that operations, determining the working conditions, may be recognised. Particularly, CNC and PLC data available are those characterising the part programmes and operations, such as active program, program status, actual spindle speed, nominal spindle speed and actual current absorbed by the spindle. The models developed should integrate such CNC and PLC data to achieve relevant information for the maintenance decision-maker.

It is worth remarking that ND is hereinafter focused on the spindle and axes as critical equipment of the machine tool. Nevertheless, the framework, in its inherent logic, could be applied to other critical subsystems/components that are monitored and for which an ND model exists or could be developed, to finally come up with the definition of the health states of the CNC machine tool.

The framework is presented in Fig. 1, where the ISO 13374-1 is taken as a reference and extended in the level named "state detection". Health assessment is considered only as a matter of labelling the health state of the CNC machine tool, without any effort towards diagnosis; also, prognostic assessment is skipped as it is not within the scope of this work.

The first two levels entail the collection of data from sensors, transducers and manual inputs, and the pre-processing and processing of data to guarantee high-quality data flowing towards the state detection data-driven models. Then, the SD level is implemented, which is focused on the developed ND models of whatever nature to determine how many and which states the machine may experience. Within this level, the newly added step is designated "operations consistency checking" (righthand grey-coloured box in Fig. 1); this is combined with the development of "fault detection" (lefthand grey-coloured box in Fig. 1). Both are data-driven models, fed by only CNC and PLC data, namely part programmes, operations ID, tool number and other program-related variables from the former, and linear speed, rotational speed, current drive load and other process-related variables from the latter.

The "operations consistency checking" is in charge of assessing if the results derived by running the data-driven model could be used or not to judge maintenance-related decisions. In particular, only when the operations are comparable (i.e. training data are comparable with testing data), the fault detection and health assessment may take place to then support maintenance decisions. Otherwise, if the operations are not comparable, the results are not directly adopted for a
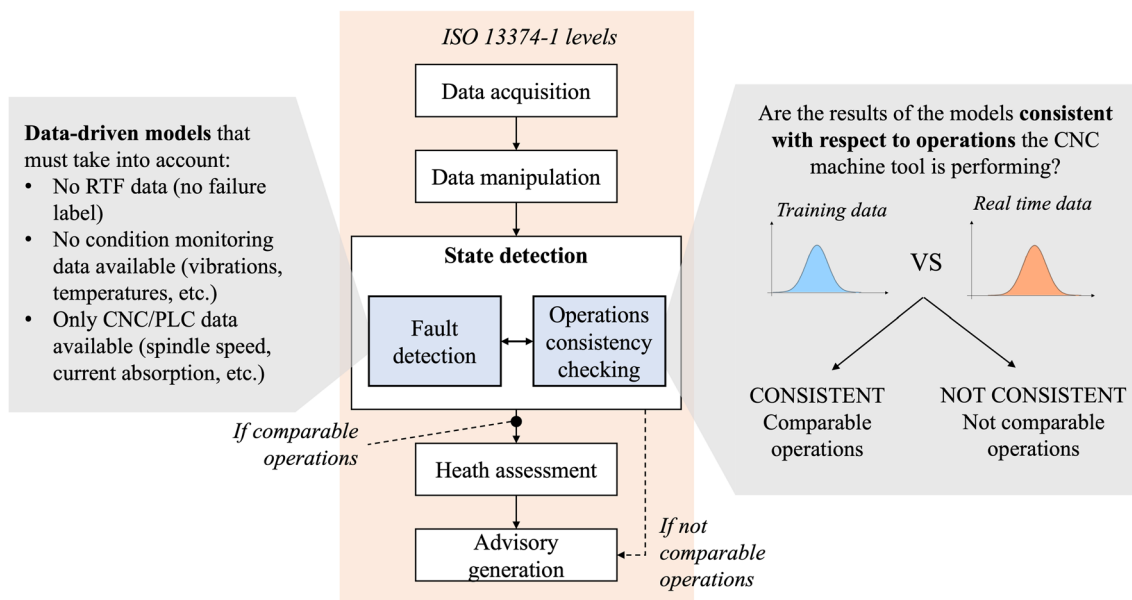

Fig. 1 Proposed novelty detection framework for CNC machine tools

maintenance reason; however, they are used to inform about a newly performed operation with respect to the history of the machine under analysis. In this way, failure-related information can be trusted by the decision-maker as they do really represent an incipient functional failure and not possible changes due to production requirements, e.g. in terms of material or processing parameters. This is assessed statistically as explained in subsection 3.2. Indeed, the "operations consistency checking" step is not in charge of assessing model accuracy and precision as this should be done during training and testing of the "fault detection" model itself. The novelty this framework claims does not refer to the single data-driven statistical model, but relies on the combination of those models for improved fault detection based on operations consistency checking, as better described hereinafter.

3.2 State detection level with a focus on operations consistency checking

The SD level represents the cornerstone of every well-conceived PHM program. The recognition of the state is subordinated to the identification of the machining operation the CNC machine tool is undergoing [48]. Nevertheless, given the specific goal of FDD, most studies from the literature concentrate on the development of well-performing algorithms for specific failure modes/causes, whereas few applications consider more than one machining operation at a time. Besides, the current state-of-the-art shows that most of the applications target ND, or equivalently fault detection in FDD, of CNC machine tools in a controlled environment or leveraging upon simulated datasets from multi-physics models. Therefore, to deal with the variations from a real industrial context, the current proposal aims at accompanying the application of the data-driven fault detection model with automatic checking of machining operations consistency; this aims to inform the decision-maker if the provided information is usable or not, for maintenance purposes.

Given the nature of the CNC machine tools, i.e. characterised by the ability to perform multiple operations, a relevant assumption must be stated: machining operations whose representative variables values are closer to each other have the same effect on the health state of the machine tools. The implications of this general assumption are hereinafter detailed:

1. Machining operations should be described with the distributions of the variables and machining parameters (such as rotational speed of the spindle, feed rate, and axis movement). These distributions can be known after one or more monitoring periods. They also do take into account exogenous variables not directly collected, such as material, which influences the operations, and so the variables, measured at runtime.

2. Machining operations could be compared by analysing the distance between the distributions of the representative variables. This enables to determine the statistical comparability.

3. Machining operations that are statistically comparable determine the same working condition, which is assumed to lead to the same effects on the CNC machine tools and their degradation, and so health state.

Therefore, as anticipated in Fig. 1, the operations consistency checking corresponds to proving the similarity between operations by comparing the distributions of the representative variables, enabling robust information about health state, and incipient functional failure, delivered to the decision-maker.

The idea is detailed in Fig. 2, which also introduces an index, between 0 and 1, that brings information about the operations consistency checking. This index is called "Affinity Index" and refers to the measurement of similarity between varying operations through time. The higher the Affinity Index, the higher the similarity between two operations, and vice versa. For the sake of simplicity, only two Gaussian distributions are drawn in Fig. 2, whereas the consistency check in its full deployment adopts a mixture of Gaussian distributions.

For a practical deployment, the solution relies on the definition of a threshold, designated as $TH_{affinity}$, in order to discriminate between comparable and not comparable operations; the reader should refer to subsection 5.3.1 for information on how to set $TH_{affinity}$. Being based on statistical analysis, the consistency check then verifies if similar machining operations are performed by the analysed machine over different monitoring periods, then bringing reliable and usable information for maintenance purposes.

3.3 State detection level with a focus on fault detection

The development of the fault detection model follows the approach by Pimentel et al. [26], which is inspired by statistical process control: if the difference between the predicted value by the ND model defined for fault detection, and the current measured value monitored from the running machine, exceeds a predefined threshold, then an anomaly has arisen. This is meant as a partial fault, activating a condition-based maintenance intervention.

To determine the predicted value, *alias* the response, a gradient boosting regression (GBR) is selected as the ND model. The selection of GBR comes from a comparison with other statistical models, namely, linear regression and support vector regression. GBR is indeed recognised as one of the most popular regression algorithm for tabular datasets. Its capability lies in its ability to effectively capture complex
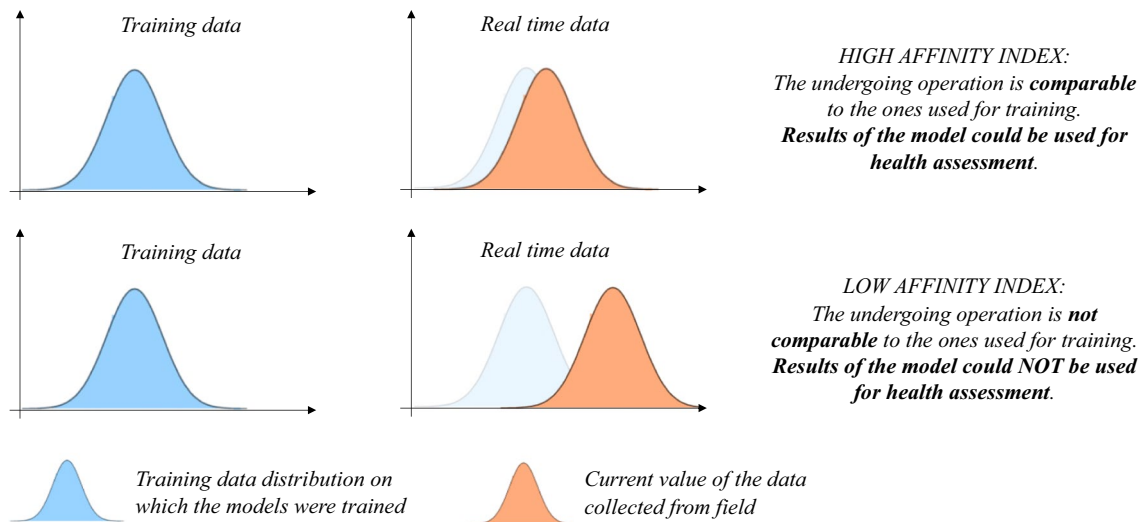
TBN: the example above is based on two gaussian distributions for which statistical testing may be performed (e.g. paired t-test). They are just used to explain the concept of Affinity Index, since each operation is characterised by multiple distributions and the comparison is based on mixture of gaussian distributions.

Fig. 2 Evaluation of Affinity Index by means of Gaussian mixture models

and nonlinear relationships between the target variable and the input features. Moreover, it offers remarkable usability by handling missing values, outliers, so mitigating the need for any special treatment [49, 50].

The goodness of the GBR model is validated via the coefficient of determination R^2, which is generally expressed as in Eq. 1.

$$R^2 = \left(1 - \frac{\sum (current\ value - response)^2}{\sum (current\ value - mean(current\ value))^2} \right)$$
(1)

The model is trained and then tested over historical data gathered during past operations of the machines in healthy state. Then, the model is used at runtime to evaluate the ND score, looking for anomalies, i.e. partial faults, by computing the difference between the predicted value (so called response) of the target variable and the measured actual value (*alias* the current value as it is measured), as in Eq. 2.

$$score = current\ value - response$$
(2)

In this way, the score acts as the ND metric later used to assess possible anomalies. Indeed, once the score has been evaluated on the test data, its distribution is computed, such as its standard deviation σ_S. These values will be used in the ND algorithm to define when an anomaly arose, i.e. when the actual value is different from the response more than 3 times the σ_S (which means the score, measured during the machine monitoring, is higher than 3 times the σ_S). This decision is commonly applied for novelty detection purposes

[26]. Grubbs identified that the mean plus three times the standard deviation represents a well-established definition for the novelty thresholds [51], as confirmed by industrial experiences [52, 14]. It is anyhow important to underline that the novelty threshold definition is an open challenge and depends on available a priori knowledge about probabilistic distributions of the signals [26].

Here in the present work, ND also comes the contribution of the Affinity Index: the ND score has validity only whenever data comes from comparable machining operations, as summarised in subsections 3.4 and 3.5.

3.4 Health assessment level with health indicator definition

Once the SD level has been formalized, it is possible to assess the health state of the CNC machine tool, or, more precisely, of the monitored subsystems and for which ND models are available. Since no RTF data are available, a ND learning approach is implemented [26]: only the healthy state is characterized and, as soon as a novelty (anomaly) arose, then a new degraded state (abnormal state) is experienced by the machine.

In practice, it is also worth observing that true-negative or false-positive may happen due to uncontrollable variations, as the operations on the machine are not deterministic and the production operators may change parameters due to contingent situations. As such, maintenance decision-makers may be mistakenly alerted along the day even for several reasons, not necessarily requiring

a maintenance intervention. Therefore, the modelling decision is to smooth this behaviour by setting a proper frequency of alerts as a relevant feature for decision-making, in the case under study daily frequency is decided. The daily frequency of alerts f_{alert} has been modelled as follows in Eq. 3:

$$f_{\text{alert}_i} = \frac{N_{\text{alert}_i}}{N_{\text{score}_i}} \tag{3}$$

where:

- N_{alert_i} is the daily number of alerts during the day i;
- N_{score_i} is the daily number of evaluated scores during the day i.

Building on the score definition (Eq. 2), the alert is released every time the score exceeds the 99.7% confidence interval of the response of the ND model for fault detection, which corresponds to the anomaly occurrence. It means that the alert is released whenever the difference between the measured and predicted values is higher than 3 times the standard deviation of the error evaluated on the test dataset.

The health state indicator (HI) is based on the daily alert frequency and is modelled as its moving average ($MA(k)$) as in Eq. 4:

$$HI = MA(k) = \frac{\sum_{i=0}^{k} f_{\text{alert}_i}}{k} \tag{4}$$

where k is the order of the moving average $MA(k)$.

To properly define the health indicator so as to be useful for decision-making, two additional open points need to be solved:

1. The assessment of the order k of the moving average. The selection depends on the required promptness of change of the indicator while avoiding useless spikes, thus, it should result after analysing real data available from the shop floor to properly run the health assessment.
2. The definition of the thresholds of the health indicators so as to label the values of HI in a discretized way, treatable for decision-making. According to the work-

ing hypotheses, no RTF data are available, hence, it is only possible to set these thresholds according to the risk the company would take. Therefore, the decision is to model these thresholds as follows, considering healthy, abnormal, or faulty states for discretizing the behaviour of the CNC machine tool:

a. Healthy state, when HI $< \mu + 3\sigma$, where σ is the standard deviation of HI and μ is the mean value.
b. Abnormal state, when HI $\geq \mu + 3\sigma$ and HI $< \mu + 4\sigma$.
c. Faulty state, elsewhere.

This builds on similar considerations expressed for the novelty threshold definition (subsection 3.3). Besides, it should be noted that the number of health states is case-dependent; the authors assume a three-state model so to include a "warning" period described by the abnormal state, before the failure (faulty state).

The information on the health state is then provided to the maintenance decision-maker in order to judge the decision to make. Indeed, the information is not provided directly, but is firstly filtered by joining together the Affinity Index as expressed in Table 2.

3.5 Working flowchart of the proposed framework

The working flowchart, characterized by the logical sequence of data management and computational steps, and decision points, completes the specification for the PHM program application according to the proposed approach in the case of CNC machine tools. Therefore, Fig. 3 reports the overall flowchart, also drawing, on its left-hand side, the ISO 13374-1 used as reference structure to build the framework.

Starting from the set of CNC and PLC data, the first step is the data pre-processing, integration, and homogenisation. This step is fundamental so that the data later analysed by the ND models are of high quality. Especially, data fusion is central as all data samples must be related to the very same timestamp. Later, the data are used to evaluate the score, for fault detection, and the Affinity Index, for operations consistency checking. Both results are used in the first two decisional points:

Table 2 Information provided to the maintenance decision-makers according to Affinity Index

		Health indicator (HI)		
		HI $< \mu + 3\sigma$	HI $\geq \mu + 3\sigma$ and HI $< \mu + 4\sigma$	elsewhere
Affinity Index	Not comparable operations ($< TH_{\text{affinity}}$)	The operation/s performed by the CNC machine tool is/are not comparable to those known a priori. No information on the health indicator is available.		
	Comparable operations ($\geq TH_{\text{affinity}}$)	The operation/s performed by the CNC machine tool is/comparable to those known a priori. Information on the health indicator is available.		
		Healthy	Abnormal	Faulty

Fig. 3 Working flowchart of the proposed framework
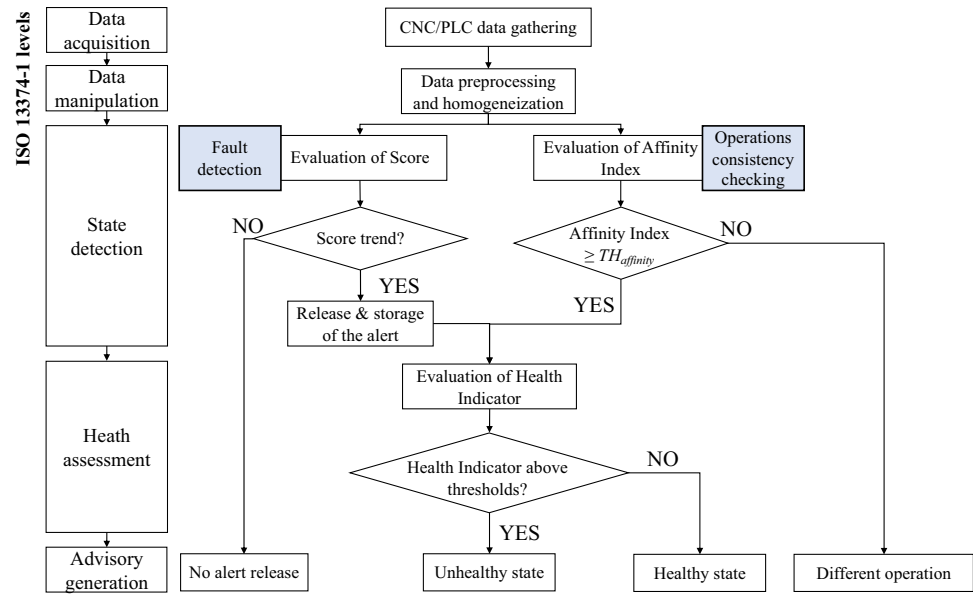
- The score for fault detection is analysed so to identify possible anomalies. If the score is within the pre-defined thresholds (three times the standard deviation, building on Eq. 2), then there is no release of the alert and the monitoring of the machine still goes on; otherwise, if the score exceeds thresholds (more than three times the standard deviation, building on Eq. 2), an alert is released and stored to be later used for the feature evaluation; the feature is then evaluated (both alert daily frequency, see Eq. 3, and moving average HI, see Eq. 4), this happens if and only if also the Affinity Index is verified.
- The Affinity Index is compared with a threshold $TH_{affinity}$ (in this case, set to 75%, that is the probability that the gathered data are similar to those used to train the model is higher than 75%, please see subsection 5.3.1 for further information). If the Affinity Index is higher, then the score could be used to evaluate the alert daily frequency and the health indicator for the decision-maker; otherwise, the information is not thrown away, but it is used to inform the decision-maker (maintenance technician or manager), that the current operation is too different from those within the training dataset of the model, so nothing could be said on the health state.

Every time both conditions are satisfied, an alert is released and stored to evaluate the alert daily frequency. Then, the health indicator is computed and, if the feature goes over the average value plus 3 times the standard deviation ($HI \geq \mu + 3\sigma$), the unhealthy state is identified, and an alarm is released to trigger a call out to the maintenance decision-maker. Otherwise, the healthy state is set, and no alarm is released.

4 Company overview and machines under study

The proposed framework for operations-aware ND of CNC machine tools, is tested in the real case of a large manufacturing company. This section 4 presents the company and the machines under study, while section 5 describes how the framework is applied in this context.

In fact, as industrial deployment, the solution has been developed during a three-year collaborative project based on action research where multiple parties are involved, such as an Italian manufacturing company, researchers, and small and medium-sized enterprises (SMEs) each with specific objectives. The Italian manufacturing company produces critical components for gas and steam turbines, later used in the power generation sector to generate electric energy for several applications. The company has a long-lasting story in this field, with more than 100 years of experience in the design, production, and maintenance of turbines. The products are sold all over the world, making the company one of the global leaders in this sector.

The Italian plant where the project was performed includes the machine tools under study, i.e. those in charge of producing the blades for the gas/steam turbines. The quality related to this kind of product must be at the highest possible standards; therefore, the health state of the machine tools producing the blades must be kept under control and any possible anomaly be noticed to avoid waterfall effects on the quality of the final assembly, i.e. the turbine. Specifically, the CNC machine tools under study are the following:

- 4-axis CNC machines tools and 5-axis CNC machine tools (5 and 1 machines, respectively, for a total of 6 machines) in charge of realising the blade roots;
- 5-axis CNC machine tools (total of 8 machines) in charge of realising the blades; and
- Newly bought 6-axis CNC machine tools (total of 2 machines) in charge of realising the blades.

The implementation of the proposed solution for operations-aware ND required the definition of a robust methodology, grounded on the ISO 13374-1 and PHM body of knowledge. Indeed, it was within the goal of the 3-year project to realise a PHM solution almost from greenfield, i.e. machines were not connected, and no platform was available for data storage and analysis. Therefore, the solution described in section 5 required some propaedeutic activities.

Amongst them, it is worth remarking on the need for criticality analyses, in this case, carried out by a failure mode effects and criticality analysis (FMECA), aimed to understand the machines and their critical items; moreover, the development of an ICT architecture was required in order to collect, elaborate and store data. In particular, it is worth pointing out that the results of the FMECA show that the spindle and the axes are those with the highest criticality in terms of risk priority number (RPN) considering occurrence, severity and detectability. Other details of the propaedeutic activities will not be treated as they are out of the scope of the insights provided in this research work.

5 Application of the operations-aware novelty detection framework for CNC machine tools

This section illustrates the development of the solution, following ISO 13374-1 and applied to the company case. Some insights on the data acquisition are provided in subsection 5.1; the data manipulation is described in subsection 5.2; the state detection is presented in subsection 5.3; and finally the health assessment in subsection 5.4.

5.1 Data acquisition

As one of the main assumptions of this research work, the only available data from the CNC machine tools are those coming from the CNC and related PLC controlling various subsystems. Hence, for streamlined data acquisition, those data must be automatically read and made available for further use, i.e. data storage and analysis. Therefore, the CNC and the PLC are connected via a gateway that is in charge of collecting the data and sending them to the cloud after some computations on edge, as represented in Fig. 4.

Given the constraint on the data effectively stored in the cloud, higher sampling frequencies required an aggregation computed on edge.

Specifically, the setup of a correct sampling frequency for each variable is of paramount importance to guarantee the correct functioning of the solution. This is combined with the identification of meaningful features from the measured variables, as obtained in the data manipulation phase.
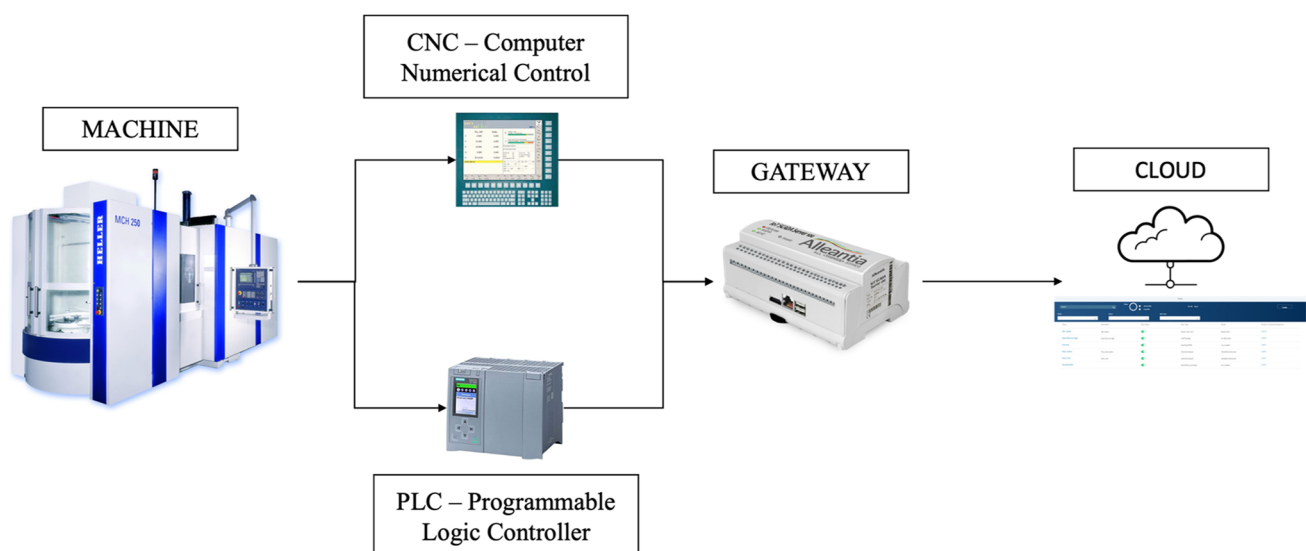

Fig. 4 Data acquisition system

Hence, an iterative approach was established between the data acquisition and the data manipulation, to identify the best trade-off between the number of variables/features and their sampling frequency.

As a result of the iterations, the sampling frequencies were fixed in the range from every second to every 2 min depending on the variable, while the aggregation was required at 2 min due to the constraint in the cloud storage.

5.2 Data manipulation

In PHM programmes, data manipulation typically entails data preparation, data cleaning and feature extraction and selection.

To enable high-quality data, an important data cleaning was made. Hereinafter, it is worth discussing one specific objective: the isolation of only active operation (i.e. metal cutting) by the CNC machine tool. In fact, during the execution of a part program, the machine tool undergoes both active and non-active operations from a product perspective: an active operation implies contact between the tooling head and the workpiece, while a non-active operation refers to correct positioning or tool changing. Given that non-active operation may impact the model performance and the constraint of 2-min aggregation time for cloud storage, it was decided to use an ad-hoc variable named "working state percentage" to estimate how much time, in a 2-min interval, the CNC machine tool spent in active operation or not. The working state percentage ranges from 0 to 1, where 1 means that all gathered data points within the 2 min represent active operation. The evaluation of the working state percentage works as follows:

1. For each data point, the "working state" is evaluated, which is based on the analysis of the CNC program string and its specific codes: if the CNC machine is in active operation, working state is equal to 1, 0 elsewhere.
2. The ratio between the number of data points with a working state equal to 1 over total data points provides the working state percentage.

The working state percentage threshold, above which the interval was considered good for the prediction of the health state, was 0.7, i.e. most of the time the CNC machine tool was in active operation. Furthermore, two additional constraints were superimposed to guarantee the correct functioning of the model: within the 2-min aggregation interval, there should be no more than 5 rapid movements and 2 tool-changing operations. These specific thresholds for rapid movements and tool-changing operations have been fixed after numerous tests and considering the manufacturing technology knowledge background.

Overall, the final value of the working state percentage aims at guaranteeing high-quality data for the actual machining operation, assuming that the higher the working state percentage, the more representative the description of the machining operation and its impact on the health state.

Building on high-quality data, the identification of meaningful variables/features requires typical methods such as dimensionality reduction looking for constant-value variables and highly correlated variables. It is worth mentioning that, for the correlation analysis, a combined physics/data-driven analysis was performed in this case, adopting: (i) a physical analysis based on empirical rules that relate manufacturing variables, like, e.g. the feed rate and the spindle speed; and (ii) a data-driven analysis based on the linear correlation coefficient. Overall, this enables the selection of relevant variables, for each CNC machine tool: the current of spindles and axes, the drive load of spindle and axis, the rotary speed of the spindle and the absolute velocity of the axes, together with the operation identifier.

5.3 State detection

State detection runs the ND data-driven models for operations checking and fault detection. These entail the following:

1. Definition of an index to measure the affinity/similarity between the operations; and
2. Definition of a score to measure the existence of an anomaly, i.e. a functional failure.

The following subsections 5.3.1 and 5.3.2 provide insights on the application, completing what was presented in subsections 3.2 and 3.3.

5.3.1 Definition of the Affinity Index

Knowing that the CNC machine tool is in active operation—as obtained from the working state percentage—does not prevent obtaining a result that is not coherent if the new operation is significatively different from those in the training dataset. As reported in Fig. 5, the probability density functions of three relevant variables collected by the CNC machines, namely, the spindle drive load, the spindle actual speed value and the current on one of the axes, depend on the specific operation the CNC machines performed over the acquisition time.

Therefore, it is of utmost importance to compare, time by time, if the gathered values are comparable or not with those present in the training dataset as this should allow reliable information for maintenance decision-making: if the operations are comparable, information on how the machine behaves could be extracted in order to make the health
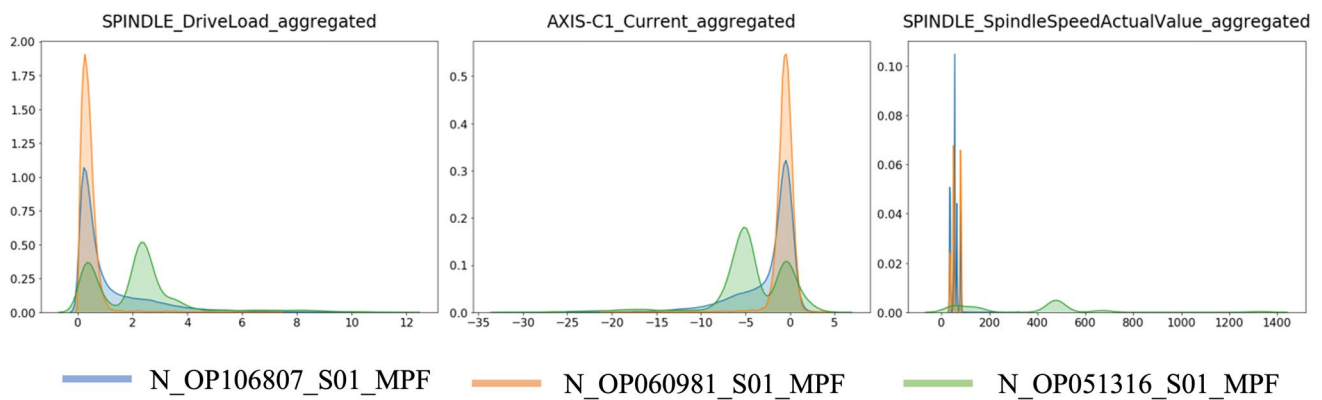
Fig. 5 Distribution of values for three relevant variables (same legend for all three plots, see bottom part)

assessment; otherwise, it is known only that the current operation is different from historical ones, whereas it is not possible to draw any conclusion on the health state of the machine.

Therefore, building on the general idea already presented in Fig. 2, the Affinity Index has been implemented through the following steps:

- A statistical model of the possible states that the CNC machine goes through during the training period is created from the data.
- Each state is represented as a multivariate Gaussian in the space of the variables describing the system (current, drive load, speed, etc.), with mean μ_i and covariance matrix Σ_i.
- The model has learnt automatically from the training set by using a Gaussian mixture model. The number of states K (i.e., the number of components of the mixture model) is inferred from the data by using the Bayesian information criterion (BIC).
- For each new test point x, the probability of belonging to the state i that the CNC machine has visited during the training process is evaluated as in Eq. 5:

$$p_i(x) = \exp\left[-\frac{1}{2}\left(x - \mu_i\right)^T \Sigma_i^{-1} \left(x - \mu_i\right)\right] \quad (5)$$

- The Affinity Index of the point x is defined as the highest probability amongst the states composing the model as defined in Eq. 6 (where A.I. stands for Affinity Index):

$$A.I.(x) = \max_{i=1..K} p_i(x) \quad (6)$$

The application of the Affinity Index enables discerning between already known operations, different operations, and similar operations, as noticeable in Fig. 6, based on the statistical comparison of collected variables. On the y-axis of the upper plots, there is the Affinity Index value for each

two-interval data point (x-axis). Instead, on the y-axis of the bottom plots, there are operation IDs for each two-interval data point (x-axis). Therefore, it is possible to read, for each operation, the related Affinity Index, thus shedding light on the comparability of the operations.

Indeed, it is possible to see that:

- Operation _N_N93Z99_S01_MPF is present both in the training and in the test dataset and the Affinity Index is high; then, it is a known operation already experienced in the past.
- Operation _N_N93Z99_S11_MPF is not present in the training dataset and the Affinity Index is indeed very low (around 0.6) in the test dataset; then, it is considered a different operation, with respect to the past experience, and is not comparable.
- Operation _N_OP006569_S02_MPF is not present in the training, but the Affinity Index is high as the related variable values are similar to those already seen in the training dataset; even if not experienced, it appears similar, i.e. statistically comparable, with past operations.

Therefore, if operations are similar (such as the Operation _N_N93Z99_S01_MPF and the Operation _N_OP006569_S02_MPF), then the ND model results can be trusted and, in case of novelty thresholds overshoot, an incipient functional failure is foreseeable. Otherwise, if operations are not similar (e.g. the Operation _N_N93Z99_S11_MPF), then the only information retrievable refers to the machine performing operations never seen and not comparable with respect to those observed during the period used to train the algorithm.

The scaled-up evaluation of the Affinity Index involved the adoption of the equal error rate (EER) assessment to identify the best threshold ($TH_{affinity}$). Specifically:

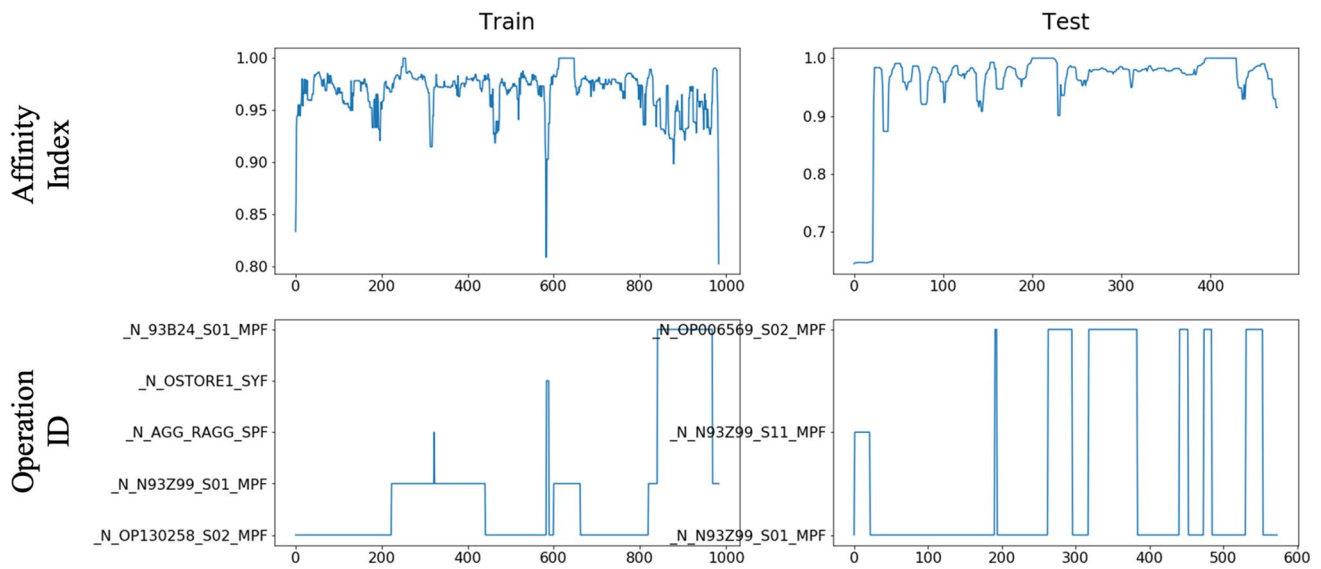1. The available dataset of operations was divided into training set and test set.

Fig. 6 Affinity Index as a function of the performed operations

2. The operations consistency checking was carried out as a function of $TH_{affinity}$; those operations with Affinity Index higher than $TH_{affinity}$ were consistent, and were not otherwise.
3. The confusion matrix was calculated considering the ground truth given by the labelling of the operations.
4. The false acceptance rate (FAR) and false rejection rate (FRR) were evaluated and the EER showed that 0.75 as $TH_{affinity}$ was the best option as visible in Fig. 7.

5.3.2 Definition of the fault detection score

Building on the approach by Pimentel et al. [26], the difference between the predicted value due the ND model and the current measured value is computed and, if the correspondent score (see Eq. 2) exceeds a predefined threshold, then an anomaly has arisen. In particular, GBR is the ND model (as in subsection 3.3) and is used at the spindle and axis level so as to be able not only to notify an anomaly but also to locate the anomaly within the machine tool. As a matter of fact, a CNC machine tool with 3 axes, has 4 GBR models, one for the spindle, one for the x-axis, one for the y-axis and one for the z-axis. Table 3 summarises the predictors for each type of model (spindle or axis) as well as the response. Predictors and responses have been identified starting from the variables/features identified as a result of data manipulation.

Each model is trained on a 2-month period and tested on a 1-month period. The dataset is selected based on feedbacks from the company about the continuous work of the machine

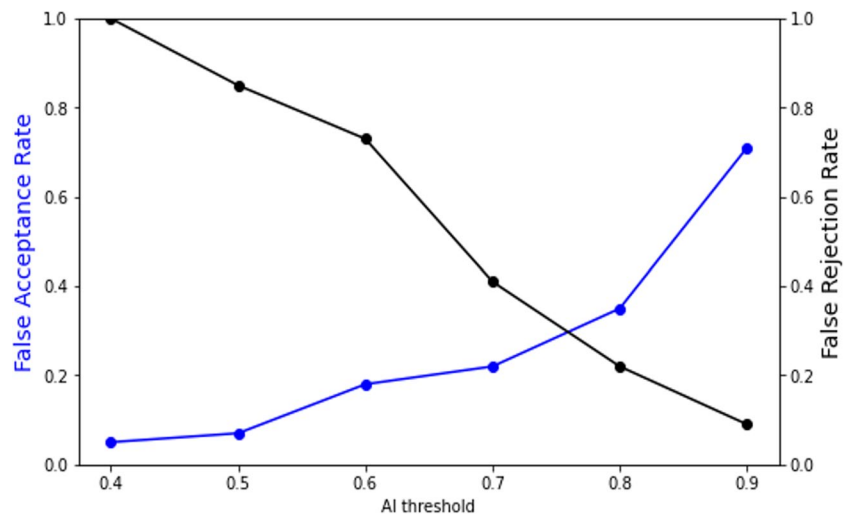**Fig. 7** EER for $TH_{affinity}$ definition (example for the 4-axis and 5-axis CNC machines tools)

Table 3 Predictors and response for the spindle and axis models

	Predictors	Response
Spindle model	Drive load Actual rotary speed Working state percentage	Actual current
Axis model	Drive load of all axis except the one under analysis Working state percentage Absolute velocity of all axes except the one under analysis	Actual current

along the considered monitored period. Also, the dates are compared with the maintenance interventions historian to avoid considering weeks in which maintenance revisions took place. Figure 8 shows the results for a 4-axis CNC machine tool; specifically, the value of the score is expressed for the training dataset (named "train"), the testing dataset without any filter (named "test"), and the testing dataset after filtering based on Affinity Index (named "test filtered").

Using the box plots, and assuming the healthiness of the machine over the monitoring period during which the training and testing have been performed in this case, the reader may be sensitive to the improvement: leaving out the variability induced by the different/not comparable operations, the box plot should be closer to the one from the training phase. In Fig. 7, this happens for all axes.

Below are the comparison plots with the results for the 4-axis CNC machine tools (Fig. 9), the 5-axis CNC machine tools (Fig. 10) and the 6-axis machine tools (Fig. 11). On the *y*-axis of each figure, there is the average score, while on the *x*-axis is the specific axis or spindle to which the score refers to.

It is worth seeing that the model performance (test after AI versus test) for the 6-axis CNC machine tools, even acceptable, is overall worse than the 4- and 5-axis ones. As a reasonable hypothesis, this evidence could be motivated by the fact that the formers perform more complex and heterogenous operations compared with the latters.

Overall, the results obtained by the application of the GBR models confirmed first of all the goodness of the model
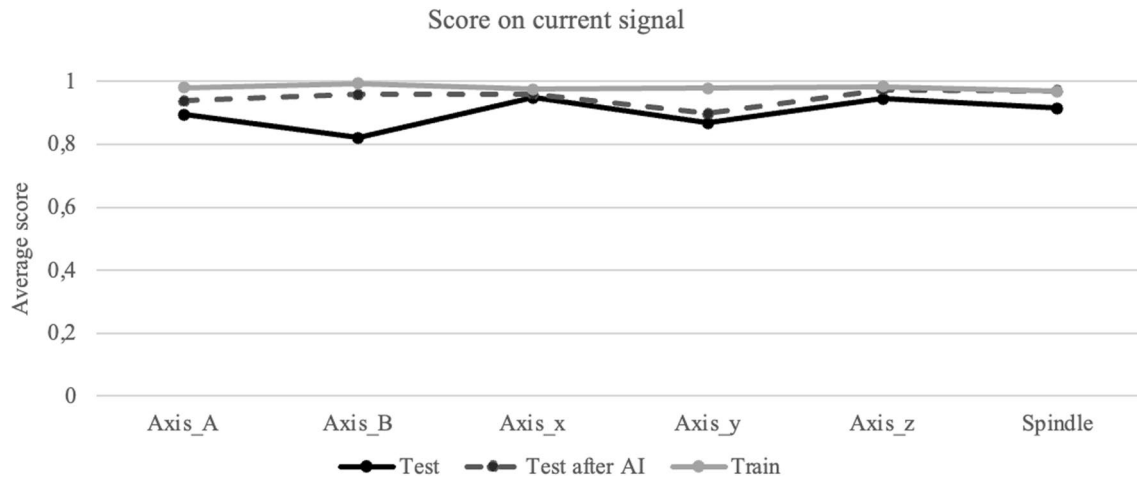

Fig. 8 Scores for the 4-axis CNC machine tool
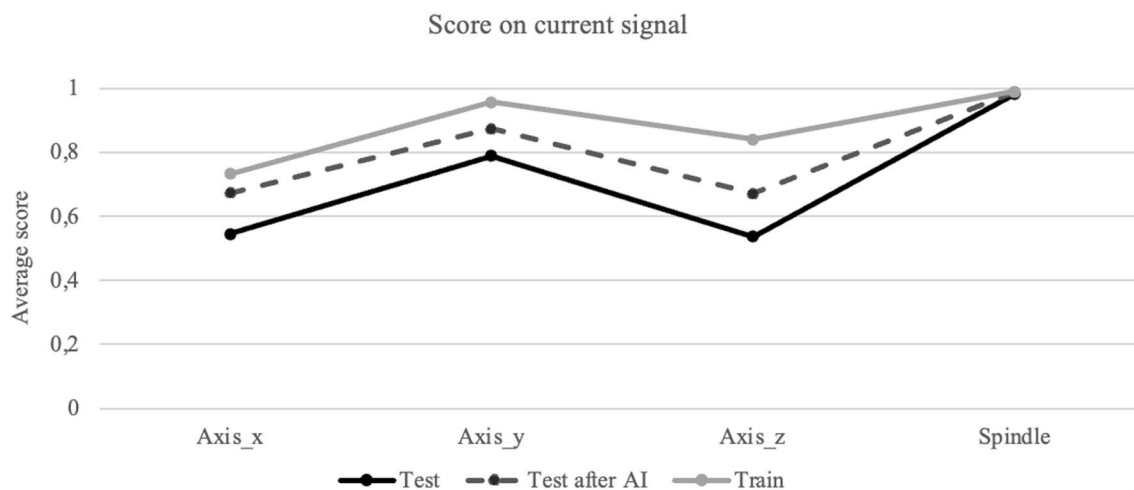
Fig. 9 Model performance for the 4-axis CNC machine tools

Score on current signal

Fig. 10 Model performance for 5-axis CNC machine tools
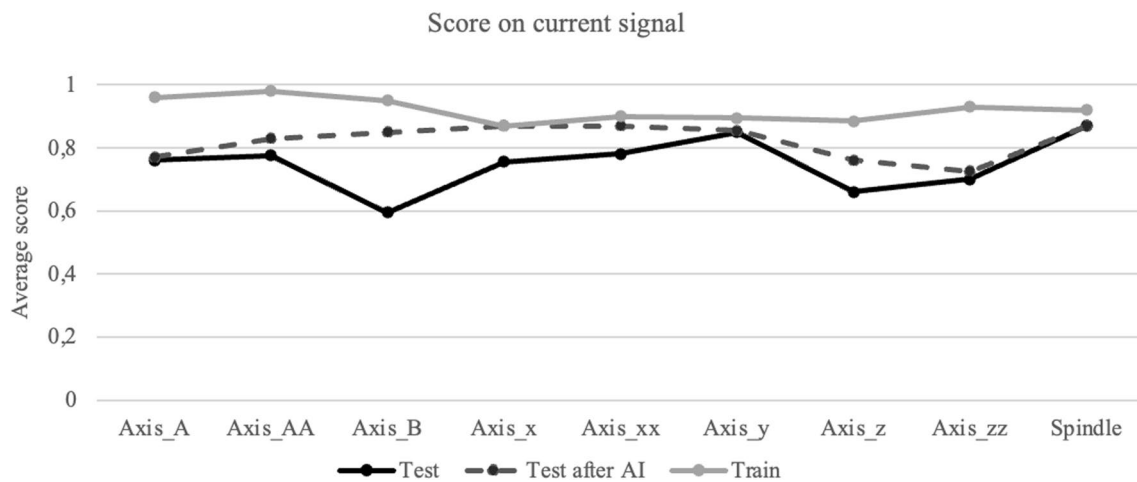
Score on current signal

Fig. 11 Model performance for the 6-axis CNC machine tools

in predicting the value of the current given the historical trend within the healthy state and, also, the capability of the Affinity Index to improve model performance by avoiding the evaluation for those operations different from the ones in the training dataset.

5.4 Health assessment

Running the GBR models allows evaluating the health indicators HI as expressed in subsection 3.4: the health indicator is the moving average of daily alert frequency. Figure 12 shows the HI of the spindle for five 4-axis CNC machine tools, named 144 (blue line), 145 (orange line), 242 (green line), 243 (red line) and 244 (purple line) according to the company's internal identification number.

It is possible to observe a constant trend for machine 145, whose HI clearly deviates from the almost stable

behaviour of the other HIs. However, to provide meaningful information to the maintenance decision-maker, the health state should be determined. Hence, two thresholds should be defined, one that sets the limit between the healthy and the abnormal states, and the second one that separates the abnormal state from the faulty state.

The decision is to adopt the same two thresholds for all machines of the same type. This is established so as to have a reference baseline that the decision-maker could use to compare the HI of more machines. For the 4-axis CNC machine tools, the two thresholds are set at 0.10 and 0.12, which respectively, means that 10% and 12% of the calculated scores over the day bring an alert. Figure 13 shows the HI of the five monitored machines with highlighted the two thresholds that define the health states (healthy is green, abnormal is yellow and faulty is red).
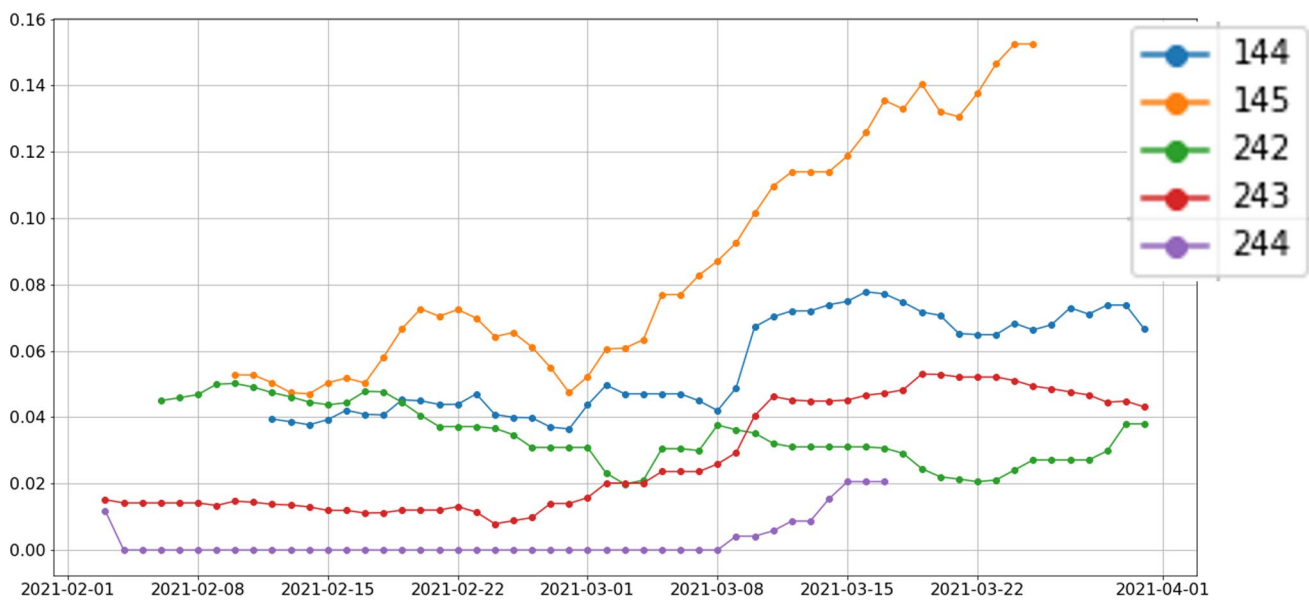
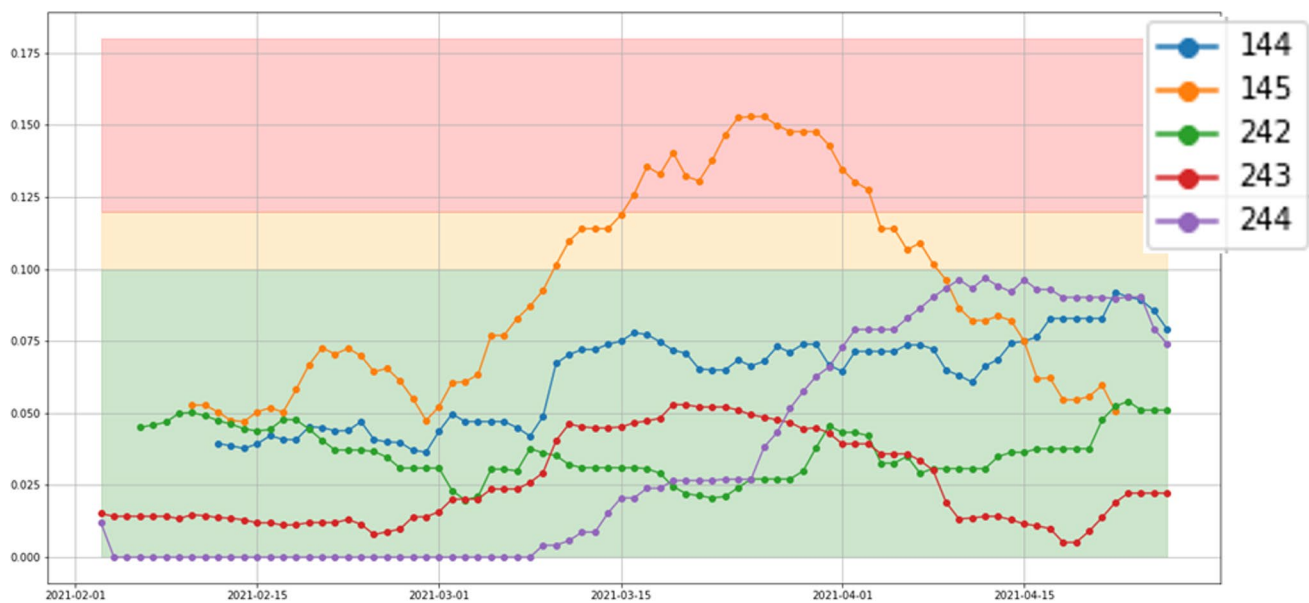**Fig. 12** Health indicators of the spindle for the 4-axis CNC machine tools

Fig. 13 Health indicators of the spindle for the 4-axis CNC machine tools with highlighted thresholds

As Fig. 13 shows, machine 145 is experiencing abnormal behaviour bringing it to a faulty state. In order to offer a shred of consistency to the results and given that, according to the company's experience, CNC machine tools are very reliable, hence, no failures are expected at their components, an alternative way to validate the proposed method was defined. Specifically, an independent maintenance provider (IMP) expert in vibration analysis was asked to install its condition monitoring solution to measure anomalies of the spindles and axes. Therefore, along a few months, the two solutions—the one proposed in this work and the condition monitoring leading to vibration analysis—worked concurrently, assuming the traditional and reliable vibration analysis as the state-of-the-art and ground truth. Indeed, vibration analysis as a meaningful technique to understand the health state comes from the experts' knowledge and it is supported by previous scientific works as shown in section 2. It is important to point out that the installed condition monitoring

solution worked under the same condition of no failure data as the proposed operations-aware ND framework.

For the specific monitoring period reported in Fig. 12, the time domain and frequency domain analyses from the accelerometers by the IMP performed on a daily basis, confirmed that the spindle of the CNC machine tool 145 was experiencing an important degradation trend as highlighted by the proposed solution. The observation of the physical state of the spindle confirmed that it was close to sudden breakdown. Hence, the double-check from the IMP and the company's maintenance technicians confirmed that the proposed solution correctly represented the health state of the monitored spindle of the CNC machine tools. If already in place, the proposed operations-aware ND solution could have been able to foresee the spindle failure 10 days in advance with no additional sensor installation within the related architecture.

6 Discussion and lessons learnt

This section 6 aims at reflecting on the developed operations-aware ND framework and the obtained results, in order to discuss the practical relevance and the managerial implications of the work.

The results obtained have been reached always considering the need to achieve a good balance between the model performance, the industrial viability and the cost-effective applicability and usability of the solution. In particular, it is worth summing up some considerations regarding the requirements and constraints requested for the development of the framework, before analysing the contributions this research work aims at bringing to the scientific literature.

Lessons learnt gained during the development and deployment of the proposed framework could be indeed summarised as follows, retracing the PHM model:

1. Data acquisition:

(a) The identification of items towards which modelling effort is worth being directed, comes from the results of the FMECA as well as the availability of already-in-place monitoring solutions, as a kind of digital readiness of the machines. Consequently, the spindle and the axes were identified as those of high criticality and already connected to CNC and PLC; vice versa, high-critical items like swarf removal and refrigeration units were excluded due to the need for important retrofitting activities, in spite of their relevance in terms of impacts on downtime. They may be the target of the following projects.

(b) The sampling frequency for data collected from CNC and PLC is influenced by constraints on cloud storage. Setting a sampling time equal to 10 ms requires 560 GB per machine per month, scaling up to 9 TB of required storage per month for the 16 CNC machine tools. The cost-benefit analysis was, therefore, core in defining the final architecture with on-edge processing with 2-min aggregation for cloud storage and subsequent elaboration.

2. Data manipulation:

(a) The selection of the relevant features used to train and run the models has required various interactions with the project partners, especially with the personnel working daily on the CNC machine tools; this enabled the exploitation of domain knowledge in terms of manufacturing technology as well as industrial automation, thus reducing the "time-to-deploy" solution.

3. State detection:

(a) The choice not to go for artificial intelligence algorithms, e.g. neural network or deep learning, was based on the request of having interpretable model outputs by maintenance technicians and managers. In managers' minds, statistical modelling of the behaviour of the machines could have promoted adequate trustworthiness by the future users of the solution, while artificial intelligence could be included in a roadmap of next project developments.

(b) The selected statistical model for fault detection, i.e. GBR, was identified based on (i) the previous experience the company in charge of algorithm development had, and (ii) the results from a trade-off between the scalability options and easiness in model maintenance in the long-term.

4. Health assessment:

(a) The selection of the health indicator was based on the trade-off between the model's capability to identify a deviation and the "suitable" notification frequency so as not to overload the maintenance process. Therefore, the best health indicator was built based on the daily frequency of alerts, not the point alerts.

The final result is a cloud-based solution that informs the maintenance decision-maker about the health states of the monitored subsystems of the CNC machine tools on the

manufacturing shop floor. The availability of an alternative CBM solution based on vibrations signals allows the validation of the results obtained with the operations-aware ND models and health assessment. This proves it effectively detects anomalies in the spindle and potentially in the CNC machine tool axes, even though one failure only has been recorded during the monitored period. The information is accessible to multiple users with different privileges and can be integrated with the company's CMMS for better decision-making in preventive and condition-based maintenance actions.

Stemming from these results, it is worth digressing over the adaptation of the proposed framework to other CNC machine tools. From the technological point of view, the framework is able to adapt to multi-axis CNC machine tools; indeed, it was already applied to 4-, 5- and 6-axis CNC machine tools. Furthermore, if the CNC machine tool under study does perform always the same set of part programmes and operations, then the Affinity Index could be skipped as there is no need to dynamically adapt to ever-changing and unpredictable part programmes dictated by varying production requirements. If run-to-failure data are available, the evaluation of the score as novelty detection could be skipped as it may be effective to move towards a complete supervised learning. In all other cases, the suggested framework may find applicability as it targets the most critical step in PHM that is state detection in the particular case of limited availability of failure data and limited knowledge of upcoming operations resulting from changing production requirements.

Discussing the advancements with respect to the scientific state-of-the-art, this research work proposes an ND framework based on the concurrent use of two statistical models, each for a specific goal: the first one looks for automatic operations consistency checking, and does represent the main novelty that this work brings about as it enables industrial viability in cases of different working conditions of the operating machines; the second one is focused on anomaly detection, once similar working conditions are identified. The joint evaluation of both in terms of Affinity Index and fault detection score allows robust information to be transferred to the decision-makers. In particular, it is worth remarking that the experimental observations (see section 5.3.2) confirmed that the effect of variability due to the changing operations, was absorbed, finally leading to a reliable fault detection score and, subsequently, health indicator to support decisions.

Considering the decision-makers' perspective, the solution offers usability with limited retrofitting. It utilizes variables measured by the existing controllers for the machining process, allowing the maintenance decision-maker to make informed decisions by employing the ND and health assessment on a regular basis, i.e. while the machining operations are running. Further condition investigation can be delegated to other technicians (i.e., experts in specific techniques such as vibration analysis). This cost-effective solution provides reliable information on the ongoing degradation risks during the machine runtime while optimising the usage of experts when required in further condition investigation.

Overall, the project demonstrates that developing, deploying and releasing a stable CBM solution for industrial machines like CNC machine tools requires looking for a good balance between model performance, industrial viability, cost-effective applicability and usability.

7 Conclusions and outlooks for future research

The research work aims at developing a framework for operations-aware ND for CNC machine tools with available data from already-in-place CNC and PLC. As the scientific state-of-the-art confirms, most of the developed ND solutions rely on vibration or temperature signals so as to perform FDD of subsystems of machine tools. However, the availability of such data does often require an important retrofitting activity on the machines under study, while CNC and PLC data could be easier to gather as they are already installed. Therefore, the research tackles this gap by proposing an ND solution based on CNC and PLC data, by assuming no vibration monitoring system available; indeed, the developed solution is able to cope with varying operations the machine tools could realise to enable reliable and usable information for maintenance decision-making.

The proposed framework is based on the PHM process guidelines and extends the SD level by adding the so-called "operations consistency checking" step, towards the Machine Tool 4.0 paradigm. This additional step is in charge of guaranteeing that the data-driven models for ND are running under similar conditions between the runtime data gathered from the machines and the training dataset. Specifically, an index called Affinity Index, is evaluated, which compares the current data with the historical one on which the models were trained, and measures the degree of similarity of the operations. If the Affinity Index is high enough, then the model response could be used to judge the health state of the CNC machine tools, triggering eventual maintenance actions. Otherwise, the maintenance decision-makers are informed about the different operations the machine is performing with respect to its previous history and, through their expertise, a decision could be eventually judged.

The proposed solution has been developed and deployed during a 3-year industrial project, which allowed testing of the operations-aware ND framework over several CNC machine tools. After proper modelling of a GBR algorithm for each of the interested subsystems, namely the machining unit that includes the spindle and the axes, the solution

effectively identified an abnormal state of the spindle of one of the CNC machine tools. If already operating, the model could have identified a real failure 10 days in advance. Also, this solution has been tested and compared via traditional monitoring solutions based on vibrations, which are recognised as the state-of-the-art in research and industry for this kind of machine, showing comparable health assessment performance. Overall, the obtained results improved the information available to maintenance technicians and managers as the cloud-based system is accessible by multiple users even remotely, highlighting the health state of the monitored subsystems of the CNC machine tools.

Concerning the scientific state-of-the-art, the proposed operations-aware ND framework overcomes current limitations related to the need for RTF data to implement an effective condition-based maintenance solution. Also, the claim this research brings about is the possibility to establish a well-performing ND solution even in case of no condition monitoring data, e.g. vibrations or temperatures, but relying only on CNC and PLC already-in-place. Finally, besides extensive data manipulation to enable data quality for the next PHM levels, this research claims that CNC/PLC data are a key element to finally achieve reliable information, as they form the basis for adapting to the changing working conditions as dictated by new part programmes introduced to satisfy continuously evolving production requirements; this is an essential concept when multiple operations, determining varying working conditions, impact on the health state evolution of the machine.

From an industrial point of view, the proposed framework could be adapted, even made simple, to various situations. As described in section 6, according to the variability in part programmes and operations and/or in the availability of failure data, some step of the framework could be skipped (e.g. Affinity Index) or modified (e.g. complete supervised learning), respectively. Hence, the research could be of interest to any company willing to implement a condition-based maintenance solution without important retrofitting activity. Indeed, this may represent the first implementation of Industry 4.0 technologies that could be then pushed forward by leveraging also on condition monitoring data and integration with information systems, such as CMMS, to label events thanks to the register of maintenance work orders.

Current and main limitation of this work relates to the assessment of the solution with one failure only available. This has been counteracted by the concurrent evaluation of health by means of vibration analysis, somehow certifying the correct health state definition for the sixteen machines explored in the study. However, the capability of the framework and related algorithms to identify evolving failures should be better assessed, beyond the only one failure occurred during the project timeline, which represent per se a future endeavour. Furthermore, other operational contexts resulting from different production requirements, could be helpful to extend the validation of the automatic operations consistency checking and of its experimental proofs as herein reported (in subsection 5.3.1 and, concerning the effect on fault detection, in subsection 5.3.2). Overall, this may inform future experimental works both in industrial settings and in laboratory settings. Due to controlled environment, laboratory tests are helpful to extend the experimentation to explore different influent factors, while variations that could result in specific industrial settings, are also relevant in order to align to challenges in real settings as well as needs and requirements from the decision-makers and other stakeholders involved in the industrial development.

Furthermore, additional future research works could first test the application of other algorithms for ND, searching for AI solutions to improve prediction performance. Also, the Affinity Index is currently based on a statistical comparison between historical operations the machine performed and the current ones; also for this purpose, AI may be tested to verify any potential advantage in classifying similar and different operations. Secondly, the proposed solution does not consider any feedback to be included automatically to improve model performance; hence, an effort to establish a comeback of information from experts to be included is envisioned as a potential breakthrough to optimise the performance of the overall solution. Finally, integration with other data may be useful to extend diagnostics and prognostics capabilities. This may come out in cloud-edge-based solutions capable to combine the computational potentialities at different levels of the cyber-physical production system built as a result of digitisation. Indeed, the proposed operations-aware ND framework is conceived as a core part of an industrial information system stack able to inform multiple actors for integrated decision-making on the shop floor.

Declarations

Conflict of interest The authors declare no competing interests.

References

1. Cohen Y, Faccio M, Pilati F, Yao X (2019) Design and management of digital manufacturing and assembly systems in the Industry 4.0 era. Int J Adv Manuf Technol 105:3565–3577. https://doi.org/10.1007/s00170-019-04595-0

2. Ibarra D, Ganzarain J, Igartua JI (2018) Business model innovation through Industry 4.0: a review. Procedia Manuf 22:4–10. https://doi.org/10.1016/j.promfg.2018.03.002

3. Lee J, Davari H, Singh J, Pandhare V (2018) Industrial artificial intelligence for Industry 4.0-based manufacturing systems. Manuf Lett 18:20–23. https://doi.org/10.1016/j.mfglet.2018.09.002

4. Negri E, Pandhare V, Cattaneo L et al (2021) Field-synchronized Digital Twin framework for production scheduling with uncertainty. J Intell Manuf 32:1207–1228. https://doi.org/10.1007/s10845-020-01685-9

5. Ahmad R, Kamaruddin S (2012) An overview of time-based and condition-based maintenance in industrial application. Comput Ind Eng 63:135–149. https://doi.org/10.1016/j.cie.2012.02.002

6. Zonta T, da Costa CA, da Rosa RR et al (2020) Predictive maintenance in the Industry 4.0: a systematic literature review. Comput Ind Eng 150:106889. https://doi.org/10.1016/j.cie.2020.106889

7. Li K, Deng Q, Zhang L et al (2021) An effective MCTS-based algorithm for minimizing makespan in dynamic flexible job shop scheduling problem. Comput Ind Eng 155:107211. https://doi.org/10.1016/j.cie.2021.107211

8. Polenghi A, Roda I, Macchi M, Pozzetti A (2021) Ontology-augmented Prognostics and Health Management for shopfloor-synchronised joint maintenance and production management decisions. J Ind Inf Integr 100286. https://doi.org/10.1016/j.jii.2021.100286

9. Guillén AJ, Crespo A, Macchi M, Gómez J (2016) On the role of Prognostics and Health Management in advanced maintenance systems. Prod Plan Control 27:991–1004. https://doi.org/10.1080/09537287.2016.1171920

10. ISO 13374-1 (2003) Condition monitoring and diagnostics of machines — data processing, communication and presentation— Part 1: General guidelines. British Standards Institution 3:31

11. Lei Y, Li N, Guo L et al (2018) Machinery health prognostics: a systematic review from data acquisition to RUL prediction. Mech Syst Signal Process 104:799–834. https://doi.org/10.1016/j.ymssp.2017.11.016

12. Yucesan YA, Dourado A, Viana FAC (2021) A survey of modeling for prognosis and health management of industrial equipment. Adv Eng Inform 50:101404. https://doi.org/10.1016/j.aei.2021.101404

13. Gittler T, Glasder M, Öztürk E et al (2021) International Conference on Advanced and Competitive Manufacturing Technologies milling tool wear prediction using unsupervised machine learning. Int J Adv Manuf Technol 117:2213–2226. https://doi.org/10.1007/s00170-021-07281-2

14. Cattaneo L, Polenghi A, Macchi M (2021) A framework to integrate novelty detection and remaining useful life prediction in Industry 4.0-based manufacturing systems. Int J Comput Integr Manuf 35:388–408. https://doi.org/10.1080/0951192X.2021.1885062

15. de Jonge B, Teunter R, Tinga T (2017) The influence of practical factors on the benefits of condition-based maintenance over time-based maintenance. Reliab Eng Syst Saf 158:21–30. https://doi.org/10.1016/j.ress.2016.10.002

16. Saari J, Odelius J (2018) Detecting operation regimes using unsupervised clustering with infected group labelling to improve machine diagnostics and prognostics. Oper Res Perspect 5:232–244. https://doi.org/10.1016/j.orp.2018.08.002

17. Luo W, Hu T, Ye Y et al (2020) A hybrid predictive maintenance approach for CNC machine tool driven by Digital Twin. Robot Comput Integr Manuf 65:101974. https://doi.org/10.1016/j.rcim.2020.101974

18. He X (2016) Recent development in reliability analysis of NC machine tools. Int J Adv Manuf Technol 85:115–131. https://doi.org/10.1007/s00170-015-7926-0

19. Liu W, Kong C, Niu Q et al (2020) A method of NC machine tools intelligent monitoring system in smart factories. Robot Comput Integr Manuf 61:101842. https://doi.org/10.1016/j.rcim.2019.101842

20. Filev DP, Chinnam RB, Tseng F, Baruah P (2010) An industrial strength novelty detection framework for autonomous equipment monitoring and diagnostics. IEEE Trans Industr Inform 6:767–779. https://doi.org/10.1109/TII.2010.2060732

21. Xu X (2017) Machine Tool 4.0 for the new era of manufacturing. Int J Adv Manuf Technol 92:1893–1900. https://doi.org/10.1007/s00170-017-0300-7

22. Lee J, Ni J, Djurdjanovic D et al (2006) Intelligent prognostics tools and e-maintenance. Comput Ind 57:476–489. https://doi.org/10.1016/J.COMPIND.2006.02.014

23. Jardine AKS, Lin D, Banjevic D (2006) A review on machinery diagnostics and prognostics implementing condition-based maintenance. Mech Syst Signal Process 20:1483–1510. https://doi.org/10.1016/j.ymssp.2005.09.012

24. Kan MS, Tan ACC, Mathew J (2015) A review on prognostic techniques for non-stationary and non-linear rotating systems. Mech Syst Signal Process 62–63:1–20. https://doi.org/10.1016/j.ymssp.2015.02.016

25. Xia T, Dong Y, Xiao L et al (2018) Recent advances in prognostics and health management for advanced manufacturing paradigms. Reliab Eng Syst Saf 178:255–268. https://doi.org/10.1016/j.ress.2018.06.021

26. Pimentel MAF, Clifton DA, Clifton L, Tarassenko L (2014) A review of novelty detection. Signal Process 99:215–249. https://doi.org/10.1016/j.sigpro.2013.12.026

27. Galar D, Thaduri A, Catelani M, Ciani L (2015) Context awareness for maintenance decision making: a diagnosis and prognosis approach. Measurement 67:137–150. https://doi.org/10.1016/J.MEASUREMENT.2015.01.015

28. Li R, Verhagen WJC, Curran R (2020) A systematic methodology for Prognostic and Health Management system architecture definition. Reliab Eng Syst Saf 193:106598. https://doi.org/10.1016/j.ress.2019.106598

29. Peng Y, Dong M, Zuo MJ (2010) Current status of machine prognostics in condition-based maintenance: a review. Int J Adv Manuf Technol 50:297–313. https://doi.org/10.1007/s00170-009-2482-0

30. Schlagenhauf T, Burghardt N (2021) Intelligent vision based wear forecasting on surfaces of machine tool elements. SN Applied Sciences 3:858. https://doi.org/10.1007/s42452-021-04839-3

31. Denkena B, Dittrich M-A, Noske H et al (2021) Data-based ensemble approach for semi-supervised anomaly detection in

machine tool condition monitoring. CIRP J Manuf Sci Technol 35:795–802. https://doi.org/10.1016/j.cirpj.2021.09.003

32. Wu Y, Liu L, Qian S (2021) A small sample bearing fault diagnosis method based on variational mode decomposition, autocorrelation function, and convolutional neural network. Int J Adv Manuf Technol. https://doi.org/10.1007/s00170-021-08126-8

33. Goyal D, Dhami SS, Pabla BS (2021) Vibration response-based intelligent non-contact fault diagnosis of bearings. J Nondestruct Eval Diagn Progn Eng Syst 4. https://doi.org/10.1115/1.4049371

34. Bernini L, Waltz D, Albertelli P, Monno M (2021) A novel prognostics solution for machine tool sub-units: the hydraulic case. Proc Inst Mech Eng B J Eng Manuf. https://doi.org/10.1177/09544054211064682

35. Du X, Gai J, Chen C (2020) Condition-based maintenance optimization for motorized spindles integrating proportional hazard model with SPC charts. Math Probl Eng 2020:7618376. https://doi.org/10.1155/2020/7618376

36. Fan J, Xue L, Liu Y, Li W (2021) Reliability analysis of spindle system of CNC grinder based on fault data. Int J Adv Manuf Technol 117:3169–3183. https://doi.org/10.1007/s00170-021-07552-y

37. Cao H, Zhang X, Chen X (2017) The concept and progress of intelligent spindles: a review. Int J Mach Tools Manuf 112:21–52. https://doi.org/10.1016/j.ijmachtools.2016.10.005

38. Rastegari A, Archenti A, Mobin M (2017) Condition based maintenance of machine tools: vibration monitoring of spindle units. In: 2017 Annual Reliability and Maintainability Symposium. RAMS, pp 1–6

39. Soori M, Arezoo B, Dastres R (2023) Machine learning and artificial intelligence in CNC machine tools, a review. Sustain Manuf Service Econ 100009. https://doi.org/10.1016/j.smse.2023.100009

40. Kuntoğlu M, Salur E, Gupta MK et al (2021) A state-of-the-art review on sensors and signal processing systems in mechanical machining processes. Int J Adv Manuf Technol 116:2711–2735. https://doi.org/10.1007/s00170-021-07425-4

41. Denkena B, Bergmann B, Stiehl TH (2021) Transfer of process references between machine tools for online tool condition monitoring. Machines 9. https://doi.org/10.3390/machines9110282

42. Baur M, Albertelli P, Monno M (2020) A review of prognostics and health management of machine tools. Int J Adv Manuf Technol 107:2843–2863. https://doi.org/10.1007/s00170-020-05202-3

43. Zhang Y, Mu L, Shen G et al (2019) Fault diagnosis strategy of CNC machine tools based on cascading failure. J Intell Manuf 30:2193–2202. https://doi.org/10.1007/s10845-017-1382-7

44. Cheng Y, Zhu H, Hu K et al (2019) Multisensory data-driven health degradation monitoring of machining tools by generalized multiclass support vector machine. IEEE Access 7:47102–47113. https://doi.org/10.1109/ACCESS.2019.2908852

45. Aghazadeh F, Tahan A, Thomas M (2018) Tool condition monitoring using spectral subtraction and convolutional neural networks in milling process. Int J Adv Manuf Technol 98:3217–3227. https://doi.org/10.1007/s00170-018-2420-0

46. Herranz G, Antolínez A, Escartín J et al (2019) Machine tools anomaly detection through nearly real-time data analysis. J Manuf Mater Process 3. https://doi.org/10.3390/jmmp3040097

47. Huang X, Zhao F, Sun Z et al (2020) A novel condition monitoring signal analysis method of numerical control machine tools in varying duty operation. IEEE Access 8:72577–72584. https://doi.org/10.1109/ACCESS.2020.2988028

48. Li W, Liu T (2019) Time varying and condition adaptive hidden Markov model for tool wear state estimation and remaining useful life prediction in micro-milling. Mech Syst Signal Process 131:689–702. https://doi.org/10.1016/j.ymssp.2019.06.021

49. Friedman JH (2001) Greedy function approximation: a gradient boosting machine. Ann Stat 29:1189–1232. https://doi.org/10.1214/aos/1013203451

50. Friedman JH (2002) Stochastic gradient boosting. Comput Stat Data Anal 38:367–378. https://doi.org/10.1016/S0167-9473(01)00065-2

51. Grubbs FE (1969) Procedures for detecting outlying observations in samples. Technometrics 11:1–21

52. Fumagalli L, Cattaneo L, Roda I et al (2019) Data-driven CBM tool for risk-informed decision-making in an electric arc furnace. Int J Adv Manuf Technol 1–14. https://doi.org/10.1007/s00170-019-04189-w