

Subscribe to DeepL Pro to edit this document.  
Visit [www.DeepL.com/pro](https://www.deepl.com/pro?cta=edit-document) for more information.

In a context marked by the accelerated development of AI, industry is being forced to upgrade.

The use of all this recent technology has modernised industry. This development is marked, among other things, by the autonomy of systems such as autonomous buildings or autonomous cars and a multiplication of data sources thanks to connected objects (iot INTERNET OF THINGS): we talk about Industry 4.0. This situation has led to the availability of data that can be useful for decision-making, particularly in the context of maintenance. Maintenance helps to keep the components of a system, such as machinery, in optimum working condition. In fact, it helps prevent system breakdowns, which can be very costly for industries [ran2019survey].

There are generally three main maintenance strategies according to \citep{ ran2019survey}. The first, known as proactive or reactive maintenance, is a strategy where intervention will only take place when a breakdown is observed. This can lead to fairly high costs associated with unexpected interruptions or breakdowns. The second strategy is called preventive maintenance. This is a strategy that consists of planning interventions on a regular basis. It can be costly when the intervention is planned even though there is no breakdown. Predictive maintenance, on the other hand, is a strategy that attempts to strike a balance between the first two approaches. It is based on predicting the health of the machine, such as the residual machine life (RUL), using modelling to help plan repair work. In research, three types of model are generally used to predict RUL.

These are physical models based on an understanding of the degradation process of the machines, which is formulated using mathematical models to calculate the RUL. The second type of model is called a hybrid model, which combines physical models and knowledge acquired (empirical data) during machine operation to update the parameters of the mathematical model used to calculate the RUL.

The data-driven approach uses available historical data to predict the RUL using learning models such as neural networks. In this way, it exploits the mass of information available from sensors to predict the residual machine life (RUL), which is essential for managing breakdowns.

In this approach, the literature shows that DEEP LEARNING AI models often provide the best performance (see references). This can be explained by the fact that sensors often provide complex data such as images or signals, which are multivariate time series. Because of their complexity, the patterns they contain are better taken into account by Deep Learning models.

However, these types of deep learning models, despite their good performance in this area, provide results that are difficult to interpret. This does not empower the decision-makers who rely on the results of these AIs. In addition, it is important to know which sensors are most involved in predicting the RUL in order to optimise maintenance intervention. DARPA has also introduced the concept of Explainable AI (XAI), for the purposes of trustworthy AI and greater transparency.

This concept has become a field of AI in its own right. It promotes better consideration of explainable AI through so-called explainability methods (XAI method). To meet this need, authors have proposed a variety of approaches to explain the results provided by less interpretable learning models such as ensemble models or deep learning models, often referred to as black boxes. [Linardatoss] has presented more than 20 explicability methods.

However, these methods use algorithms that are often independent of learning (e.g. so-called post-hoc methods). This makes it impossible to have a vision of the consequences (direct or indirect) of the modelling aspects on the quality of the explanation provided by the XAI methods.

This modelling can be broken down into 3 phases: pre-processing of the data, choice and training of the model and evaluation of the results. In each phase, we have to make choices about methodology. For example, during pre-processing in the context of multivariate time series, we need to make sure that we select the 'right features', an optimal way of smoothing, etc.; during training, we need to choose the model's hyper-parameters, set the evaluation criteria, etc. All these choices of methods (pre-processing, training and evaluation) are made in the same way. Thus, all these choices of methods (e.g. standardisation method, window size in the case of time series) can have a direct or indirect influence on the quality of the explanations obtained using XAI methods.

Faced with this variety of explicability methods, there appears to be a need for a comparison framework that will enable arbitration between XAI methods for a better explanation of the results of black box models. This is what has led some authors to propose different approaches for evaluating XAI methods. These approaches can be grouped into three groups: human grounded, application grounded and functionnal grounded.

The first one,

The second

The third is based on property and nnaa.

A