Intelligence & Robotics

A review of intelligent methods of health assessment technology

---Manuscript Draft---

<table>
<thead>
<tr>
<th>Manuscript ID:</th>
<th>IR-2023-021301</th>
</tr>
</thead>
<tbody>
<tr>
<td>Manuscript Title:</td>
<td>A review of intelligent methods of health assessment technology</td>
</tr>
<tr>
<td>Manuscript Type:</td>
<td>Review</td>
</tr>
<tr>
<td>Special Issue:</td>
<td></td>
</tr>
<tr>
<td>Keywords:</td>
<td>Prognostics and health management, health assessment, intelligent methods, industrial systems</td>
</tr>
<tr>
<td>Funding Agency and Grant Number:</td>
<td></td>
</tr>
</tbody>
</table>
A review of intelligent methods of health assessment technology

Diyi Liu¹, Linyuan Peng¹, Zhiyao Zhao¹,*

¹School of Artificial Intelligence, Beijing Technology and Business University, Beijing 100048, China.

Correspondence to: Prof. Zhiyao Zhao, School of Artificial Intelligence, Beijing Technology and Business University, No. 11/33 FuCheng Road, HaiDian District, City Beijing 100048, China. E-mail: zhaozy@btbu.edu.cn

Received: date month year [e.g., 1 Jan 2022]

Abstract

The core technology of prognostics and health management (PHM), a key technology that detects system anomalies, is health assessment, which analyzes and diagnoses the current system working status and quantitatively assesses the system health. This paper reviews health assessment technology’s development in recent years from three aspects: health definition, health assessment indicators, and health assessment approaches. In terms of health definition, this paper summarizes three common definition methods. Health assessment indicators are reviewed from four levels: process variables, data features, residuals, and fusion indicators. Finally, health assessment approaches are divided into model-based, data-driven, and fusion approaches. Concerning the data-driven approach, rapidly developing health assessment research based on an intelligent approach is discussed. The paper also compares various approaches and identifies the technology’s current challenges and development prospects.

Keywords: Prognostics and health management, health assessment, intelligent methods, industrial systems
1. INTRODUCTION

With the advent of Industry 4.0, complex systems such as industrial systems, aerospace equipment, vehicles, electricity, and ships are developing rapidly, followed by the reliability and safety assessment of diverse, complex systems. By the end of the twentieth century, prognostics and health Management (PHM)\(^1\) technology had become a key technology for realizing comprehensive safeguarding research in the Joint Strike Fighter (JSF) program. PHM can significantly reduce the cost of maintenance, use, and support; improve the safety and availability of aircraft; and achieve the “economic affordability” goal in the security of a large complex system.

Currently, PHM has not only been applied in the development of large-scale high-value equipment (e.g., aerospace and defense affairs) and achieved remarkable results, but has also been studied and applied in general complex systems (e.g., industrial systems, vehicles, mechanical exercises, electric power, and ships)\(^2\)–\(^7\).

Currently, defining the “health” of humans is not straightforward, as health may reflect an individual’s emotional status or the absence of known diseases. From the perspective of indicators, health can be reflected in statements like “I feel good/bad” or all normal indicators in medical examinations. Human health is concerned not only with the current condition of the human body but also with health prediction and prevention. For example, will someone catch a cold? What medications are needed to prevent colds? Therefore, how to quantify health status and take preventive measures to maximize the health level is desirable in the health industry. With the advancement of science and technology and the increase in system complexity, similar problems have been introduced into systems engineering, and the “health” of systems and products has become a topic of widespread concern. The field of systems engineering also needs to answer questions about human health based on what functions systems should fulfill and what standards systems should use to accomplish them. In this context, health can be described by the degree to which a system degrades or deviates from its expected normal working status\(^8\). In practical engineering, the health status of the system includes the health status of all devices in the system and the availability status of the functions provided by them, reflecting all the information of the devices, system structure, and functions in the system\(^9\).
Figure 1 shows that the plant, sensing equipment, and health management system constitute a closed-loop PHM framework for practical engineering systems. A common PHM system in the field of systems engineering mainly includes three main layers: monitoring, prediction, and management. PHM refers not only to detecting process anomalies but also diagnosing, analyzing, and predicting faults in addition to assessing the health status of a component or system, predicting its remaining useful life (RUL), and helping to develop corresponding maintenance and operation strategies to ensure that the system completes the expected function and realizes the status maintenance. The aforementioned are based on existing experience, cases, and model inference algorithms, using system-observed data with the help of models and related algorithms. This technology can find early failures as well as effectively guide the maintenance decision before the failure, avoid the occurrence of harmful accidents, and solve the problem of insufficient maintenance and excess maintenance in regular equipment maintenance [10–12]. Essentially, PHM uses a lot of condition monitoring data and prior knowledge, with the help of statistical algorithms or models. To assess the health status of equipment [13], this technology can predict the potential failure in advance and can combine various information to provide proactive maintenance decisions to achieve condition-based maintenance, improving the safety of the production process and reducing operating costs [14,15]. Health assessment, the core technology of PHM, from the perspective of system health, assesses whether the current working status of the system is normal and whether the system will undergo performance degradation within a certain period, which plays an important role in ensuring the security and reliability of the system. Unlike system failure detection and identification [16], health assessment does not focus on a critical component of the system or where a failure may occur, but rather on the overall system performance and the detection of abnormal symptoms [1,9].
This paper reviews the current development of health assessment technology in industrial systems and vehicles and reviews the health assessment technology from three levels: health definition, health assessment indicators, and health assessment approaches. Health definition refers to how to use known system models, operations data, and other information to characterize the health degree of the system. Health assessment indicators are quantitative indicators reflecting the overall working status or performance of the system. Moreover, health assessment approaches are used to assess whether the current operating status of the system is normal and whether the system will undergo performance degradation within a certain period. Through the reference analysis at three levels, the existing problems and possible prospects of the current research are summarized and discussed.

2. HEALTH DEFINITION

This section introduces the research status of the system health definition. Notably, the primary issue in health assessment technology research is how to define health or how to characterize health through known information such as models and data. Most previous studies have not clearly defined health. Through investigation, we can divide the definition of health in existing health assessment studies into the following three categories: performance variables-based, residual-based, and reliability-based health definitions.
2.1. Performance variables-based health definition

For an actual engineering system, the health of the system at any time can be defined by the performance variables of the system, namely

\[
H_{\text{sys}}(t) = \begin{cases} 
1 & \text{if } \Sigma(x,t) \in S_H \\
0 & \text{if } \Sigma(x,t) \notin S_H 
\end{cases},
\]

(1)

where \( x \) represents the performance variable of the system; \( H_{\text{sys}} = 1 \) represents the healthy system; \( H_{\text{sys}} = 0 \) represents the unhealthy system; \( \Sigma(x,t) \) is a health function composed of the performance variables of the system; and \( S_H \) represents the corresponding health space.

i) If the health of the system can be fully characterized by a single performance variable without loss of generality, assuming that \( x_i \in x \) can fully characterize the health status of the system, we obtain

\[
H_{\text{sys}}(t) = \begin{cases} 
1 & \text{if } x_{i,\text{lower}} = x_i(t) = x_{i,\text{upper}} \\
0 & \text{otherwise}
\end{cases},
\]

(2)

where \( \Sigma(x,t) = x_i(t) \), \( S_H = [x_{i,\text{lower}}, x_{i,\text{upper}}] \). This definition means that the system is healthy when \( x_i \) changes within its health interval \([x_{i,\text{lower}}, x_{i,\text{upper}}]\); otherwise, the system is unhealthy.

ii) If the health of the system can be characterized by multiple performance variables Without loss of generality, assuming that the vector composed of multiple performance variables can characterize the health status of the system, we obtain

\[
H_{\text{sys}}(t) = \begin{cases} 
1 & \text{if } \|x - x_0\| \leq \varepsilon \\
0 & \text{otherwise}
\end{cases},
\]

(3)

where \( \Sigma(x,t) = \|x - x_0\| \), \( x_0 \) is the optimal operating point of the system; and \( S_H \) is a regular polyhedron with \( x_0 \) as the center point and the side length \( 2\varepsilon \). This definition means that the system is healthy when the distance between \( x \) and the optimal operating point is less than the tolerance threshold \( \varepsilon \); otherwise, the system is unhealthy.
For simplicity, health can also be defined as follows:

\[ H_{sys}(t) = \begin{cases} 1 & \text{if } x_{i,\text{lower}} \leq x_i(t) \leq x_{i,\text{upper}}, \forall x_i \in x, \\ 0 & \text{otherwise} \end{cases} \]  

This definition means that the system is healthy when each performance variable changes within its own healthy interval; otherwise, the system is unhealthy.

2.2. Residual-based health definition

A quantitative measure of the inconsistency between the actual behavior and the expected behavior of a system is called the residual [17]. For simplicity, the difference between the observed quantity (y) and the estimator (\(\hat{y}\)) output by the system can be called the residual:

\[ r = y - \hat{y} \]

Different forms of residuals can be sensitive to different types of exceptions that occur in the system. Therefore, the distribution of residuals can reflect the health status of the system. When the system is healthy, the residual distribution should be close to 0; when there are anomalies in the system, the residual distribution deviates from 0 and different anomalies correspond to different residual probability distribution forms. Based on this, we can define the health of the system in terms of the residuals:

\[ H_{sys}(t) = \begin{cases} 1 & \text{if } \forall |r_i(t)| \leq r_{i,T}, i = 1, \ldots, n, \\ 0 & \text{otherwise} \end{cases} \]

where \(r_i\) represents the \(i^{th}\) component of the residual, and \(r_{i,T}\) represents the corresponding tolerance threshold.

For complex systems, we can also calculate the corresponding health degree of each residual component according to Equation (6), and further calculate the health of the system:

\[ H_{sys}(t) = \sum_{i=1}^{n} w_i \cdot H_{r_i}(t) \]

where \(w_i > 0\) represents the weight of \(r_i\), which reflects the impact of \(r_i\) on the overall system health, \(\sum_{i=1}^{n} w_i = 1\).
2.3. Reliability-based health definition

Traditional reliability modeling methods are based on life test data and can reflect the common reliability features of similar systems under specified conditions. Therefore, the traditional definition of reliability is difficult to apply to the health assessment of a single product or system. However, the concept of performance reliability \[^{18}\] provides a healthy definition for practical engineering systems, especially dynamic systems.

The definition of interval performance reliability is given in the literature \[^{19,20}\]. Given a dynamic system,

\[
\begin{align*}
\dot{x} &= f(x,u) + \Gamma_w w \\
y &= h(x,u) + \Gamma_v v
\end{align*}
\]

(8)

where \( x \in \mathbb{R}^{p \times 1}, u \in \mathbb{R}^{r \times 1}, w \in \mathbb{R}^{q \times 1}, v \in \mathbb{R}^{q \times 1} \); \( f: \mathbb{R}^{p \times 1} \times \mathbb{R}^{r \times 1} \rightarrow \mathbb{R}^{p \times 1} \), and \( h: \mathbb{R}^{q \times 1} \rightarrow \mathbb{R}^{p \times 1} \). For the dynamic system shown in Equation (8), space \( \mathbb{R}^p \) is divided into a healthy space \( S_H \) and an unhealthy space \( S_F \) \( (S_H \cup S_F = \mathbb{R}^p) \). For a given time interval, \([t_0, t]\), the interval performance reliability is defined as

\[
R(t_0, t) = P\{x(\tau) \in S_H, \forall \tau \in [t_0, t] \mid x(t_0) \in S_H\}.
\]

(9)

This definition means that the health of a dynamic system in \([t_0, t]\) is the probability that the system status stays in the health space in this time interval. Equation (9) can also be written as

\[
R(t_0, t) = P\{x \text{ does not transfer to } S_F \text{ over } [t_0, t] \mid x(t_0) \in S_H\}.
\]

(10)

This definition means that the health of a dynamic system in \([t_0, t]\) is the probability that the system status does not undergo a status deterioration (transition to an unhealthy space) in that time interval.

The definition of instantaneous performance reliability is given in the literature \[^{18,21}\]. For the dynamic system shown in Equation (8), space \( \mathbb{R}^p \) is divided into a healthy space \( S_H \) and an unhealthy space \( S_F \) \( (S_H \cup S_F = \mathbb{R}^p) \). For a given time, the instantaneous performance reliability of the dynamic system at that time can be defined as
\[ R(t) = P\{ x(t) \in S_H \} = \int_{S_H} p(x,t) \, dx, \]  

where \( p(x,t) \) is the joint probability density function of the process variable \( (x) \) of the dynamic system at time \( t \) in \( \mathbb{R}^p \) space. This definition means that the instantaneous health of a dynamic system at time \( t \) is the probability that the system status is in health space \( S_H \) at that time. Moreover, according to the literature \(^{[18]}\), the instantaneous performance reliability of the system at time \( t \) can also be written as

\[ R(t) = R(t_0, t) \cdot R(t_0). \]  

Combined with Equations (9)–(11), Equation (12) can be written as

\[ R(t) = P\{ x(\tau) \in S_H, \forall \tau \in [t_0, t] | x(t_0) \in S_H \} \cdot P\{ x(t) \in S_H \}, \]  

or

\[ R(t) = P\{ x \text{ does not transfer to } S_F \text{ over } [t_0, t] | x(t_0) \in S_H \} \cdot P\{ x(t) \in S_H \}. \]

3. HEALTH ASSESSMENT INDICATORS

Health assessment indicators are quantitative indicators that reflect the overall working status or performance of the system. This section details the health assessment indicators proposed in previous research to describe system health. Health assessment indicators can be divided into three categories: process variables, data features, and residuals \(^{[22]}\).

3.1. Process variables

In power electronics, many health assessment indicators are directly aided by process variables. In the health assessment and prediction of batteries, the capacity and internal resistance of batteries are the most common health assessment indicators. Reference \(^{[23]}\) used the battery capacity of Li-ion batteries as a health assessment indicator and used interacting multiple model (IMM) particle filtering to determine the RUL of Li-ion batteries and the probability distribution function of related uncertainties to assess the health status. Reference \(^{[24]}\) achieved a more accurate prediction of the RUL of batteries by using particle filtering and autoregressive time-series models to track the capacity...
decline process of Li-ion batteries. Reference [25] compared the capacity decay rate of 16 batteries and gave a hybrid method by a sparse Bayesian learning module and recursive Bayesian filtering module to realize the health assessment of batteries. Reference [26] developed a comprehensive health assessment indicator to predict the RUL of batteries by combining capacitance, resistance, and constant current charging time. Reference [27] determined the degradation behavior of batteries in different statuses and predicted its RUL by using the long short-term memory model to learn the relationship between the battery voltage, current, and battery temperature.

In other areas, many studies have used process variables directly as indicators of health assessment. Reference [28] used machine learning to establish a classifier and then used the change of frequency modulation when the nuclear system performs the task to detect the fault of the nuclear system in the early stage. Reference [29] performed a reliability assessment and health prediction of production equipment by collecting information on systems or components installed in the field and using the failure mode and impact analysis and hazard analysis techniques.

However, in many cases, process variables cannot directly reflect the health of the system, so we need to filter the data of process variables through certain algorithms and use the information extracted from process variables as health assessment indicators, mainly including data features and residuals.

### 3.2. Data features

A data feature obtains a process variable from the original sensor’s measured data by using a specific mathematical algorithm, which can reflect the hidden features of the data. This method is often used in the health assessment [6] of mechanical systems and equipment, wherein we can obtain health assessment indicators by extracting data features.

In the study of mechanical system health assessment, many studies have achieved the
health assessment and prediction of different equipment based on the features shown in
Table 1. For example, researchers \cite{30} used the method of wavelet transform to extract
the frequency spectrum and energy contained in the generator vibration signal to
perform effective fault diagnosis. Reference \cite{31} conducted quantitative analysis and
diagnosis of rotor vibration fault features by combining the wavelet correlation filter
method and information entropy theory. Reference \cite{32} used time domain analysis and
fast Fourier transform (FFT) to obtain features from vibration data, and they used the
particle swarm algorithm for feature selection. Time domain analysis is used to obtain
the statistical scalar features from vibration analysis data. FFT decomposes the
waveform signal into its spectrum, which contains component frequencies and their
amplitudes. The energies (defined as the sum of the squares of amplitudes) over the
frequency bands centered on specific frequencies (e.g., rotating frequency and
harmonics and/or bearing defect frequencies) are calculated as features. Moreover, the
design of the fitness function of the particle swarm algorithm considers three aspects:
monotonicity, predictability, and trend. Reference \cite{33} extracted the root-mean-square,
kurtosis, peak factor, skewness, and other features from bearing vibration signals, and
they used them to construct and train a hidden Markov model (HMM) to realize health
status assessment and prediction. Reference \cite{34} used the Hilbert-Huang transform to
extract intrinsic mode functions from vibration signals and then used them as
multidimensional health assessment indicators of the system to construct a
multidimensional health space. Further, a support vector machine (SVM) is used to
assess the health status of ball bearings, and support vector regression (SVR) is used to
predict the RUL. Reference \cite{35} also used the cutting force wavelet transform analysis to
realize the status assessment of micro-milling, and the waveform and waveform
singularity generated by the wavelet transform differ in different health statuses. Based
on this principle, the method analyzes the cutting force signal under the working status
of micro-milling to realize the health assessment.
3.3. Residuals

Residuals are the most commonly used health assessment indicator in model-based health assessment methods, which are often generated by inconsistencies between the actual and expected process variable values or features of the system.

Detailed steps of fault detection and diagnosis based on analytical redundancy and residuals are given in the literature [17,36]. However, residual-based health assessment has been more focused on fault detection and identification. For example, a team [37] designed an adaptive digital twin for bearings and used the difference between the estimated signal output and the actual signal output of the original system as the residual signal, which was used as the basis for bearing fault classification and crack size identification. Moreover, researchers [38] designed a partial differential equation observer to obtain the system output residual, which was used as the residual signal indicating the fault, and they used the threshold value of the residual signal to assess the health status of the system. Reference [39] designed the signal prediction model of a fiber current sensor based on a long short-term memory network and obtained the residual signal through the predicted and observed signals. Using the features of the residual signal, the fault diagnosis model based on SVM is established. Another team [40] achieved the comprehensive estimation of the circulating current and output current of the multilevel converter according to the Kalman filter (KF) algorithm, used the predicted current value and the measured current value to achieve the residual signal generation, built the residual assessment function and its threshold based on the obtained residual, and achieved the fault detection of a multilevel converter. The fault detection and health assessment of hybrid systems are also mostly based on residuals. For example, researchers [41] used a hybrid bond graph to model the hybrid system, constructed different residuals to be sensitive to different faults, and used the size of the residual to represent the health degree of each component of the system. Then, the adaptive hybrid differential evolution algorithm was used to realize the fault identification and health assessment of the steering system of electric vehicles (refer to
3.4. Fusion indicators

The fusion of health assessment indicators aims to combine different process variables, features, or residuals, in addition to using indicators with lower dimensions and better monotonicity and trends to reflect the health information that the original scale indicates [47].

3.4.1. Fusion of process variables and features

Principal component analysis [48] (PCA) is the most commonly used method in the fusion of health assessment indicators.

Reference [49] extracted the time domain and frequency domain features from the motor torque, stator current, and velocity signals, using the PCA to fuse the time domain and frequency domain features for the health status prediction of the motor rotor. Another team [50] used the PCA to establish the principal component model and then carried out dimensionality and noise reduction on the original complete dataset of the brake system of the hoist, achieved feature extraction, and completed the fusion of the data layer, which is used as the basis for locating the fault site and completing the fault diagnosis of the hoist. Another team [51] used the PCA to reduce the characteristic quantity of the motor bearing, and then the obtained low-dimensional characteristic quantity was input into the back-propagation neural network for fault classification. References [52,53] used PCA to reduce the data dimension of the observed information set for timely fault diagnosis and health prediction of complex systems. Reference [54] conducted a PCA on 21 measured data points of turbofan engines and took the first principal component as the health assessment indicator, which was further used to achieve health assessment and residual RUL prediction. Another team [55] used a global feature extractor based on kernel PCA to reduce the redundant attributes of vibration data, providing a reference for other studies using residuals as health assessment indicators) [42–46].
for deep learning fault diagnosis of rotating machinery.

In addition to PCA, researchers [22] used genetic programming and took monotonicity as the optimization objective to find the optimal combination from various features of bearings and spindles and then constructed health assessment indicators. Another team [56] used the fuzzy support vector data description to fuse various characteristics such as the mean square root value and peak state of the bearing into a fuzzy monitoring coefficient. This coefficient is sensitive to the early defects of the system, and with the development of faults, it can grow steadily. Considering that this coefficient produces oscillations, the runtime is introduced, and a monotonous health indicator is established. Reference [57] used the random forest classifier to obtain deep representative feature values from multichannel data and then constructed a multi-deep belief network fusion model for fault diagnosis of the main reducer. Reference [58] established a multiparameter regular fault diagram and conducted a weighted analysis of the historical operating parameters of the injection well to realize the health assessment of the operating status of the injection well. Moreover, researchers [59] proposed a data fusion model to fuse multiple sensor data in a multi-physical field measurement system for fault diagnosis of battery systems. Reference [60] proposed the Mahalanobis distance (MD) to characterize system health. This definition converts multidimensional sensor data into an MD value, which can effectively characterize the health status of the system.

3.4.2. Fusion of residuals

In the above studies, the fusion of health assessment indicators referred to the fusion of process variables and features. However, the fusion of residuals has also been a common method in health assessment studies. Reference [61] used the status estimation algorithm based on IMM to realize the fault detection and status assessment of F/A-18 aircraft. The algorithm realizes fault
detection according to the filtering residuals of multiple models and performs probability fusion of the filtering values of each model status based on the residuals to achieve status estimation. References [62,63] improved the algorithm to increase the accuracy of fault detection. Reference [62] modified the status probability in each iteration. Reference [63] used particle filtering to replace KF, extended KF, and unscented KF. Reference [64–66] used this algorithm for fault detection of a satellite attitude control system. Similarly, researchers [67,68] used improved multiple model adaptive estimation (MMAE) to achieve fault detection of actuators and sensors of unmanned aerial vehicles (UAVs). The algorithm also performs a probabilistic fusion of the status variables of a UAV based on residuals. Reference [69] used the status estimation algorithm based on IMM to realize a multiple fault diagnosis and a status assessment of Li-ion batteries.

4. HEALTH ASSESSMENT APPROACHES

Health assessment approaches can be divided into model-based, data-driven, and fusion approaches [4,70].

4.1. Model-based approaches

Model-based health assessment approaches can be divided into three categories according to different models: physics-of-failure (PoF), mathematical, and qualitative model-based methods.

4.1.1. Physics-of-failure model-based method

The PoF-based health assessment method uses the PoF model and failure mechanism knowledge of a system or product to describe system degradation and achieve health assessment [71,72]. This kind of method is commonly used in the health assessment of electronic products. Reference [73] used the AdaBoost classifier to adjust the sample distribution and then establish the physical degradation model of the gearbox to characterize the damage
status of the gearbox. By analyzing the acoustic emission signals of the gearbox
damage process, the authors could achieve an effective fault diagnosis of the gearbox.
Reference [74] achieved an accurate description of air spring features by establishing
physical models of an electronic control air suspension system, and they established
physical degradation models of different sensors with different fault types for fault
identification of sensors in an electronic control air suspension system. Reference [75]
established the competitive failure model of the vehicle integrated transmission and
used the maximum likelihood estimation method to estimate the parameters of the
model, realizing the prediction of the RUL of integrated transmission. Reference [76]
established the fault model of a single-stage fixed shaft gear system based on the
dynamic model of a single-stage fixed shaft gear system combined with the energy
method, and they analyzed the dynamic response of the single-stage gear system when
the single-stage gear was normal, cracked, and peeled. The system’s spectral features
were analyzed, providing a theoretical basis for the health monitoring and fault
diagnosis of the gear system. Reference [77] used the exponential model to model the
degradation process of bearings and the observed data to estimate the model parameters,
and the maximum likelihood estimation and particle filtering algorithm were used to
predict the RUL. Reference [78] represented a second-order exponential model as the
physical degradation of battery capacity, and the model parameters were considered
invariant owing to Gaussian noise. Furthermore, the unscented particle filtering was
used to achieve the real-time estimation of the model parameters, and the physical
degradation curve of capacitance was obtained to predict the RUL. Reference [79]
modeled the physical degradation of battery capacitance and used particle filtering to
achieve the estimation of model parameters and the prediction of the RUL of batteries.

The biggest advantage of the PoF-based health assessment method is that it can use the
known PoF information and fault mechanism to achieve high-precision health
assessment and prediction. However, in practical engineering, we often can only obtain
the PoF information of a certain system component or parameter (e.g., the bearing in
the mechanical system or the capacitance in the battery), not the physical degradation model of the whole dynamic system. Especially, the dynamic system pays more attention to the macroscopic performance of the system; its system model often does not contain physical degradation information. Therefore, using PoF to achieve a health assessment must meet the following requirements: 1) health and performance changes of dynamic systems can be fully reflected by key components and parameters; 2) the physical degradation model of key components and parameters is known, and the model parameters can be updated in real time through sensor data. More importantly, in health assessment studies, this model is used more to solve the health “prediction” problem after the health “monitoring” phase has been completed, rather than to study the health “monitoring” of dynamic systems.

4.1.2. Mathematical model-based method

The mathematical model-based health assessment method refers to the use of the mathematical model of the dynamic system for systematic health assessment. Among them, the mathematical model of the dynamic system can be existing or be obtained by the system identification method. Then, filters, observers, parameter estimation, and other methods are used to realize the system state estimation or residual generation, determine the system health state according to the system state estimation value, or determine the system anomaly according to the size of the residual [17].

Reference [80] constructed the power system model of a fuel cell vehicle and the mathematical model of hydrogen consumption cost and degradation cost, and then they adopted the adaptive moving average filtering and the equivalent cost minimum strategy to optimize the output power of the fuel cell. The nonlinear control strategy was used to control the energy status of a supercapacitor in a reasonable range to achieve the health assessment of the fuel cell vehicle. In the fault detection and health assessment of aircraft, the most common method is to establish the mathematical model of the aircraft and then use the observer or filter to achieve fault detection and status
estimation. Reference [81] used the status estimation based on the IMM algorithm for the fault diagnosis of UAV control systems, established the global and local fault models of sensors and actuators, and applied the residuals of each model for filtering and probabilistic fusion, which realized the accurate diagnosis of the local and global faults of sensors and actuators of UAVs. Another team [82] added the update of transition probability based on the classical IMM algorithm to solve the problem of pattern recognition error and intermittent false alarms caused by the constant transition probability to accurately estimate the real-time distribution of the hybrid state of multirotors and assess their health status. Reference [83] also proposed a particle filtering algorithm to assess the flight performance of multirotors, and they used particle filtering instead of KF in the IMM algorithm to estimate the real-time probability distribution of the hybrid state of the multirotor model. The comparison showed that the IMM algorithm based on particle filtering could effectively reduce the estimation error and improve the accuracy of the health assessment of multirotors. Another team [68] used the improved multi-model adaptive estimation algorithm to achieve the fault detection of UAV sensors, which also performs probability fusion of the state variables of UAVs based on residual errors.

General dynamic systems have known or partially known mathematical models. Therefore, mathematical model-based methods are suitable for studying the health assessment of a dynamic system. However, most mathematical model-based methods focus more on fault detection and diagnosis in the field of PHM, and there have been relatively few studies on health assessment using this method. This is mainly because 1) the fault of the dynamic system can be intuitively added to the dynamic system model, which is easy to understand and study; 2) the health of dynamic systems is not clearly defined, and it is difficult to combine model parameters with health.

4.1.3. Qualitative model-based method

The qualitative model is usually based on the understanding of the system mechanism
or its physical and chemical process, taking this kind of information as a priori knowledge to establish a qualitative system model [84].

Among qualitative model-based methods, the graph model is a common method for dynamic system modeling. Reference [85] encoded fault features through the observation of the residual in the bond graph model to locate the fault source, which effectively overcomes the shortcomings of the analytical redundancy method that relies on mathematical linear models. Another team [86] introduced interval analysis theory into the traditional linear difference change technique of a bond graph, modeled parameter uncertainty and measurement uncertainty in a unified manner, extended the bond graph model to an uncertainty bond graph model, and the deduced interval analytic redundancy relation; this was done to calculate interval analytic redundancy relation using the interval mathematical operation method and obtain the diagnostic threshold, which has been successfully applied to the fault diagnosis of parametric faults and sensor faults of electro-hydrostatic actuators. Based on the bond graph model, a team [87] obtained the system’s dual-causal bond graph model by using the method to separate causality. By analyzing the causal relationship of each node in the dual causal bond graph model, the authors examined the system analytical redundancy relationship and fault feature matrix and achieved the system fault detection and isolation according to the analytical redundancy relationship and the fault feature matrix. Thus, the set of possible faults was obtained, and the health assessment and prediction of nonlinear electromechanical systems could be carried out.

Qualitative model-based methods are based on a deep understanding of the system structure and causality and are used for fault detection and diagnosis as well as health assessment by establishing qualitative models and logical reasoning. This method does not require accurate quantitative models, which can avoid modeling difficulties and provide a more intuitive explanation of the causes of failures. However, it has two problems: 1) the qualitative model generated based on residuals is usually for fault
detection and diagnosis, rather than health assessment; and 2) when the qualitative model based on the knowledge base and expert reasoning is used for health assessment, its accuracy is relatively poor.

4.2. Data-driven approaches

Data-driven health assessment approaches model and analyze the observed historical data, extract the data performance patterns under normal working conditions and abnormal working conditions, and compare them with the observed data at the current moment to achieve health assessment [72]. Generally, data-driven health assessment approaches can be divided into statistics-based, stochastic process-based, and artificial intelligence-based methods.

4.2.1. Statistics-based method

The statistics-based method for prediction of the RUL is based on the theoretical basis of probability statistics and involves collecting the failure life data of equipment or some degradation data representing the failure of equipment, modeling according to a statistical model, estimating parameters according to monitoring data, and then achieving prediction of the RUL of equipment.

Reference [88] proposed a hybrid mechanical fault diagnosis method based on probabilistic box theory and an improved grey wolf optimization (GWO) algorithm to optimize SVM. The feature vector set for fault diagnosis was constructed by directly establishing the probability box and using the cumulative uncertainty measurement method to extract its features. The improved GWO algorithm is used to optimize the SVM to achieve classification and diagnosis of the feature set and complete the fault diagnosis of the rolling bearing. Reference [89] used the error integral of the output probability density function of the complex system as the driving information so that the system status and fault can be estimated by the adaptive fault diagnosis observer. Reference [90] established the probability density function of mainshaft bearing
vibration by the guided maximum entropy method; introduced the Lagrange multiplier and empirical coefficient; and established the upper and lower limit function of probability density, the truth function model of fault probability, and the estimated value of reliability interval to achieve the health assessment of bearings. Reference \cite{91} used the hierarchical sampling method based on particle filtering to obtain the importance probability density function of the status of autonomous underwater vehicles and achieved real-time status estimation and trend prediction of the motion state of autonomous underwater vehicles.

4.2.2. Stochastic process-based method

The stochastic process-based method is used more in studying the system degradation process. This is because in the actual engineering application of all kinds of equipment, the degradation is often random owing to the internal structure instability, the sudden situation in the manufacturing process, the change in operating conditions, or the change in the external environment. Under the guidance of mathematical statistics knowledge, the stochastic process model is created to describe the degradation process. Commonly used stochastic processes are the Wiener, Gamma, and inverse Gaussian processes, which are used to characterize different cases of the degradation process.

Reference \cite{92} took the meta-action unit as the object of study and used the Wiener process to describe the performance degradation of a meta-action unit. Among them, the meta-action unit is a unified whole composed of all parts in accordance with the assembly relation to ensure that the most basic action of electromechanical products can be realized. Reference \cite{93} extracted the original health indicator representing the system operating status and used the Gamma process to model its change process to perform fault diagnosis of complex systems. In another work \cite{94}, there were two main problems in the analysis of the time-varying failure probability of a corroded pipeline: 1) the determination of the failure status of a corroded pipeline; 2) the accuracy of the simulation of the corrosion degradation process of a pipeline. Based on the blasting
data of a pipeline test, the residual internal pressure-bearing capacity of a pipeline was
selected according to the minimum loss function value. In addition, the Gamma, inverse Gaussian, and Wiener processes were introduced into the calculation of the failure probability of the corroded pipeline, and then the time-varying failure probability of the corroded pipeline was calculated in combination with the Monte Carlo method, which provided accurate and reliable results for predicting the time-varying failure probability of the corroded pipeline.

### 4.2.3. Artificial intelligence-based method

As a self-learning method, artificial intelligence does not need to estimate parameters in the estimation process, but the definition of the content and objective function of input and output are needed, followed by iterative estimation through a training learning algorithm. Finally, the estimation result is obtained and the training stops. Artificial intelligence is essentially a simulation of the information process of the human brain thinking, which establishes a neural network for data analysis and learning by simulating the mechanism of the human brain. Collecting a large amount of data is the foundation of artificial intelligence. The artificial intelligence–based method does not need to analyze the mechanism model of the device, nor does it need to estimate the parameters of the model. Moreover, it has a strong nonlinear mapping ability, which can estimate the status of various models, and it has a strong fault tolerance ability. Common artificial intelligence methods include SVM, artificial neural networks, and deep learning.

Reference [95] used the spiking neural network (SNN) as an intelligent fault diagnosis tool for rotating machinery bearings. The features extracted from original vibration signals are encoded as spikes through local average decomposition, and the SNN is trained by learning rules to achieve fault diagnosis of bearings. Another team [96] proposed a rolling bearing fault diagnosis model based on a dual-stage attention-based recurrent neural network (DA-RNN), which was used to expand unbalanced datasets in
actual fault diagnosis cases. Then, the improved convolutional neural network model was used for fault classification. Reference [97] enhanced the sensitivity of extracting sensor fault features through the combination of wavelet transform and the temporal convolutional network.

The advantage of a data-driven health assessment approach is that it can learn the system’s dynamic behavior from historical data without relying too much on the system’s physical and mathematical models. Considering the large amount of measured data in dynamic systems, data-driven approaches are more widely used than model-based approaches. However, the physical meaning of the health assessment results obtained by this method is sometimes difficult to interpret. Also, data-driven approaches often require a long learning time, suffer from overfitting, and make it difficult to determine the unhealthy threshold.

### 4.3. Fusion approach

Considering the advantages and disadvantages of both model-based and data-driven approaches, many studies have attempted to integrate the two approaches to improve the accuracy of health assessment.

For example, researchers [25] used a fusion approach based on data-driven/model-based approaches to assess the RUL of the battery. The sparse Bayesian learning module was used to infer the capacity from charge-related features, and the recursive Bayesian filtering module was used to update the empirical capacity decay model and predict the RUL, which effectively achieved the prediction of the RUL of batteries. Reference [98] used existing failure models and advanced fault prediction algorithms to locate and predict faults of guided munitions systems, which provided a way to achieve fault prediction and health assessment of typical guided munitions. Reference [99] proposed a fault prediction method based on the fusion of the physical failure and data-driven models. Reference [100] proposed a device fault diagnosis method based on the fusion of
knowledge and data, which can not only classify the equipment operation data through
the optimized bidirectional long short-term memory network model but also make
auxiliary decisions through the knowledge graph based on the fusion fault chain,
realizing the construction of the equipment fault domain graph driven by the fusion of
mechanism knowledge and data. In another work \cite{101}, aiming at the health assessment
problem of diesel engine air management system, the flow model and inflation
coefficient model were established by the mechanism modeling method, and the
inflation coefficient model and intake pressure fluctuation amplitude model were
established by the data-driven modeling method. The odd-even equation method was
used to design the residual generator and generate three residual signals. The mapping
matrix between the fault and the residual value could be obtained by simulation
analysis, and the fuzzy inference method is used to diagnose the fault to carry out an
effective health assessment.

5. CONCLUSIONS AND PROSPECTS
Through the above research work, as an important part of the PHM framework, health
assessment technology has been widely researched. However, deficiencies remain in
current health assessment studies.

1) Currently, most existing health assessment studies have focused on RUL prediction,
fault detection and identification studies, or the timely estimation of system process
variables studies, but there has been a lack of health assessment studies in a real sense.
Owing to the unclear and unified definition of health, health assessment research has
deviated from the correct development direction to a certain extent, such as a lack of
systematic health quantitative measurement and subsequent health predictions. Studies
of RUL prediction have often used performance variables to characterize system health
and have achieved prediction of RUL by predicting the variation trend of performance
variables. Furthermore, fault detection and identification studies have used residuals to
characterize system health and achieved fault detection and identification by assessing
the characteristics of residuals. Research on the timely estimation of system process variables has generally used the current system observations to estimate the timely system process variables as well as to define whether the system is healthy or not. In fact, studies on RUL prediction, fault detection and identification, and timely estimation of system process variables are not equivalent to real health assessment studies. This is primarily because it is unreasonable to directly define health based on performance variables and residuals; it is also insufficient and inaccurate to define the system’s health only using the current system’s working status.

2) Most existing health assessment studies have only focused on a certain level of the health assessment system; the input and output studies often do not consider the output results and input information studied at other levels. For example, when studying the problem of health assessment, researchers have not considered how to predict the RUL of the system after detecting the abnormality and then plan the corresponding health assessment approaches. When studying the problem of RUL prediction, researchers have only assumed that the abnormality has been detected through the health assessment approach in advance, and they have then directly conducted the degradation modeling and RUL prediction based on the health assessment results. These kinds of modular research situations arise due to the lack of an available theoretical basis for health assessment technology, which is not conducive to the overall development of health assessment technology.

3) At present, most health assessment techniques for industrial vehicles, such as unmanned autonomous systems, have been studied for individual systems, and the assessment results are easily affected by the external environment. There has been a lack of health assessment studies for isomorphic systems. For the health assessment of an individual system, it is difficult to effectively overcome the noise interference of the external complex environment, and it is difficult to accurately assess the system’s health status. Especially when the system process variables are selected as health
assessment indicators, the assessment results of the system health degree are more significantly affected by the external environment. Recently, the application of unmanned system formation has become increasingly widespread. For example, unmanned system formation is widely used in performance, inspection, agricultural plant protection, and other aspects, so research on the health assessment technology of isomorphic systems is expected.

Future studies on health assessment technology should consider the following development trends.

1) Health assessment technology based on the model and data-driven fusion approaches: considering that both model-based and data-driven methods have unique advantages and disadvantages, combining the two methods is an important development trend of health assessment technology and even the whole PHM to exploit each other's strengths and compensate for their shortcomings. For example, with the wide application of intelligent algorithms, the identification of system models through system-observed data can overcome the problem that complex systems are difficult to model.

2) Health assessment technology based on the fusion of quantitative and qualitative knowledge: system knowledge includes quantitative measured data, which can objectively reflect the system operating status, as well as people’s subjective qualitative cognition and experience. Therefore, how to integrate these objective quantitative data and subjective qualitative information in the studies of health assessment technology would be an important direction of a health assessment study in the future.

3) Health assessment technology for isomorphic systems: with the progress of science and technology, more unmanned autonomous systems have entered people’s lives. The health assessment of the isomorphic system can not only conform to ongoing trends but also skillfully overcome the influence of the external environment on the assessment
results of each vehicle in formation. For example, the comparison of system residuals between objects in the isomorphic system is used to effectively distinguish between health states and abnormal states. This represents an important direction for the future development of health assessment technology.

4) PHM study integration: most existing studies have only focused on one level of the PHM system, and the input and output studies often do not consider the output results and input information studied at other levels. This situation is due to the lack of an available theoretical basis for PHM, which makes it difficult to form a complete research chain including health assessment, RUL prediction, and emergency decision-making. Therefore, providing a solid theoretical basis for PHM and achieving the integration study of each module of PHM on this basis is another key direction of future research.

DECLARATIONS

Authors’ contributions
Project administration: Zhao Z
Writing-original draft: Liu D, Peng L
Writing-review and editing: Zhao Z

Availability of data and materials
Not applicable.

Financial support and sponsorship
This work was supported by Beijing Natural Science Foundation under Grant 4222042, and National Natural Science Foundation of China under Grant 61903008.

Conflicts of interest
All authors declare that they have no conflicts of interest.

Ethical approval and consent to participate
REFERENCES


7. Ahmad R, Kamaruddin S. An Overview of Time-Based and Condition-Based


11. Zhao ZY, Quan Q, Cai KY. A Profust Reliability Based Approach to Prognostics and Health Management. *IEEE Transactions on Reliability* 2014;63:26-41. [DOI: 10.1109/TR.2014.2299111]


51. Zou YJ, Tian MQ, Qiao JQ, Ma B, Song JC, Zhang WJ. A time-frequency image based feature extraction method for motor bearing faults in toothed roller crushers.


60. Kumar S, Pecht MG. Health Monitoring of Electronic Products Using Symbolic Time Series Analysis. AAAI Fall Symposium: Artificial Intelligence for Prognostics; 2007, pp. 73-80


87. Li H. Fault diagnosis and prediction of nonlinear electromechanical systems based on particle filtering and limit learning machine [M.S.]: Hefei University of


96. Li J, Liu Y, Li Q. Intelligent Fault Diagnosis of Rolling Bearings under Imbalanced


To Editor

Dear Editor,

Many thanks for your effort in handling the review process of our manuscript. We highly appreciate the reviewers and editor-in-chief for their carefulness, conscientious and the broad knowledge on the relevant research fields, since they have given us a number of beneficial suggestions. After a careful study of all comments, we have carefully revised the manuscript and enriched it. The main revisions are as follows:

1. Some explanations and statements and related analysis are added to make the paper more clarified to readers.
2. A few confusing sentences are modified.
3. The overall structure of this manuscript has been improved, such as deleting Table 1.

To Reviewer 1

Reviewer #1: A review on PHM were conducted in this paper. Some of my concerns are as following :

1) The abstract should be revised. There are too much things on the background with only few words on the work of this paper.

Response: Thanks. The content of the work of this paper is supplemented.

2) The language should be polished completely. There are still errors in the paper. For example, “which is mainly to analyze and diagnose the current working status of the system”. There is grammar error here. Line 381, “this kind of method” should be “this kind of methods”. There are also some errors like these two.

Response: Thanks. They have been correctly modified.
3) Why there are only industrial systems and vehicles were considered in this paper?

Response: Because the review scope of this paper is aimed at industrial systems and vehicles. From the perspective of health management, it is also used in other fields, such as the food and medical industry. But the health management methods of other industries and the industrial systems and vehicles industry are not the same system. Therefore, this paper only considers industrial systems and vehicles.

4) I don’t think it is clear in Section. In fact, this part is not on health definition but on health measurement. In my opinion, these are different things. Furthermore, where does the method in Section 2.1 come from? This is a review paper; the things discussed in this paper should focus on the existing works.

Response: Definition of health: System health can be considered when the indicator is normal. How to determine whether the indicator is normal or not is to see whether the indicator is within the specified threshold. If the threshold is exceeded, it is considered that the system is unhealthy, which can also be understood as abnormal detection. The health status of the system is divided into two states: "health" and "unhealthy". We define health in such a way. Health measurement refers to how to measure health systematically. They are different. In addition, the method in Section 2.1 is the result of our integration of existing ways of dealing with health definitions.

5) In my opinion, this review paper was not well organized. For example, what is the purpose to present the materials in Table 1.

Response: Thanks. The overall structure of this manuscript has been improved, such as deleting Table 1.

We appreciate your valuable comments very much, which make the paper clearer to readers. Thank you!
To Reviewer 2

Reviewer #2: Comments:

1) Line 253-254: Please clarify what is meant by “fast Fourier change”.

Response: Thanks. Details on the fast Fourier Transform have been supplemented.

2) Line 546: Please elaborate on what a “meta-action unit” is.

Response: Thank you. The explanation of the "meta-action unit" has been fully supplemented.

3) Verb tenses are not always correct. Example: Line 426: “to obtain”

Response: Thanks. Due to the adjustment of the word order, the tense of the word has been adjusted correctly.

4) Some sentences are run on sentences. Example: Line 426-432

Response: Thanks. The word order has been adjusted appropriately.

5) Some nouns are missing proper articles. Example: Line 434: “a” missing before “fuel cell vehicle”. Line 462: “a” missing before “dynamic system”. Line 470: “a” missing before “Qualitative model”.

Response: Thanks. The missing articles have been added appropriately.

Thank you for your comments. They have been correctly modified.
To Reviewer 3

Reviewer #3: According to the author’s response, some PHM ideas may be incorrect. The authors should think more about the conclusion.

1) According to response 1, can the author provide the theoretical basis of health management technology for all the applications? I don’t think it exists.

Response: Thanks. The appropriate modifications have been made. As far as I am concerned, for a part of applications, the theoretical basis of available health management technology may be developed. For example, the theoretical concept of quantitative measurement and prediction of system health. And we have made appropriate modifications to Conclusion (1).

2) According to response 2, the author may be mistaken for "real-time". In actual application, nobody can achieve "real-time". "real-time" depends on how much time lag you can accept. Each “current state” results from the previous state based on the basic assumption of “Continuity in the physical world”.

Response: Thanks. It has been changed from “real-time” to “timely”.

3) According to response 3, what is the so-call “unified theoretical basis for health management technology?”

Response: Thanks. It has been changed from “unified” to “available”.

4) According to response 4, how does it work for reducing the effect of the external environment by using “the health assessment technology of cluster systems.”

Response: Thanks. It has been changed from “cluster” to “isomorphic”. For example, the comparison of system residuals between objects in the isomorphic system is used
to effectively distinguish between health states and abnormal states. And we have made appropriate modifications to conclusion (1) and future studies (3).

Thank you for your perceptive comments. We have thought about it carefully and revised this paper. Thank you very much!
Abstract:
The core technology of prognostics and health management (PHM), a key technology that detects system anomalies, is health assessment, which analyzes and diagnoses the current system working status and quantitatively assesses the system health. This paper reviews health assessment technology’s development in recent years from three aspects: health definition, health assessment indicators, and health assessment approaches. In terms of health definition, this paper summarizes three common definition methods. Health assessment indicators are reviewed from four levels: process variables, data features, residuals, and fusion indicators. Finally, health assessment approaches are divided into model-based, data-driven, and fusion approaches. Concerning the data-driven approach, rapidly developing health assessment research based on an intelligent approach is discussed. The paper also compares various approaches and identifies the technology’s current challenges and development prospects.