Sciences du Nord Nature Science and Technology
Vol. 2, No. 02, December 2025: pp. 117-131
E-ISSN: 3046-6504
P-ISSN: 3047-4922
DOI: https://doi.org/10.71238/snnst.v1i02.232

Self-Calibrating Mechanical Digital Twins for Multi-Physics Fault Localization in Industrial Gear Transmission Systems: A Systematic Review

Rupesh Rameshar Dhore [1]
[1] Senior Design Release Engineer Solize USA Corporation Canton, MI, USA
Coresponding Author: rupeshdhorey@gmail.com

ARTICLE HISTORY
Received Nov, 2025 **Revised Nov, 2025** **Accepted Dec, 2025**

ABSTRACT

The loading applied to an industrial gear transmission system is non-stationary, the nature of the signatures of any faults is complex and difficult to localise due to the simultaneous meshes, time varying transmission paths and progressive deterioration. A digital twin can be used to continuously diagnose the machine in real-time, but if the twin is calibrated at commissioning, it will be inaccurate if any part of the physical system changes, for example due to wear, pitting or cracking. This systematic review aims to summarize the information from 30 peer-reviewed publications spanning over the last 14 years (2011-2024) and serve as a basis for laying the foundations of a self-calibrating mechanical digital twin for multi-physics fault localization. The themes of the corpus were themed, and the formulations governing the corpus were incorporated into a general framework of Bayesian state updating, hybrid physics–data decomposition, modelling of mesh stiffness, modelling of vibration signal on planets, degradation laws, and discrepancy minimisation for calibration. The synthesis shows that 2021 or later, 56.7% of the corpus was available, and 23.3% of the sources in the corpus were from a single journal: physics-informed learning resulted in relative errors less than 6.7 × 10−4 with only 100 samples taken from the boundaries. This representative example of a planetary stage demonstrates small difference in fault frequency between planet and ring (1.0Hz) to enable a frequency only location even though the use of transfer path analysis and calibrated virtual data is encouraged. There was a mean bias of 30% from 27 samples of a self-calibrating twin and a mean bias of 32.7% from 27 samples of a static twin. Main challenges are hybrid, continuously calibrated twins, and computational cost, parameter identifiability, cyber security and validation requirements are the most balanced solution.

Keywords: *Digital Twin; Fault Localization; Gear Transmission; Model Updating; Multi-Physics Modelling; Physics-Informed Machine Learning; Planetary Gearbox; Prognostics and Health Management; Self-Calibration*

INTRODUCTION

Power conversion components in wind turbines, helicopters, mining and heavy industrial drives are the most common gear transmission failure types, and these failures result in long downtime and additional expenses. The complexity of planetary configurations stems from the fact that a number of sun–planet and planet–ring meshes are operating at the same time, the planets are rotating with the carrier, and the vibration transmission path from the gear to the sensor is changing periodically according to the position of the carrier [1]. These properties result in very modulated signals, where the fault signatures are hidden in the healthy meshing response, as well as background noise.

The following three types of existing diagnostic and prognostic approaches have been identified: model-based approaches, data-driven approaches and a hybrid approach. One disadvantage of the model based techniques is that they are not physically interpretable and they

are dependent on the changes in the value of the model parameters over time, which can change due to wear, temperature, assembly tolerances etc., while on the other hand, the data driven techniques can lead to high accuracy classification of faults, provided the corresponding fault data are acquired with the appropriate labels in large quantities [2]. The short run-to-failure histories, non-stationary operating conditions and high cost of a wrong maintenance action make family inappropriate to industrial practice; thus family, alone, is insufficient to satisfy the criteria for reliable fault localization.

The digital twin (DT) has been proposed as a single solution to overcome these challenges in the form of a virtual asset which is continuously linked to the physical asset via bidirectional data exchange. It was initially seen in the early 2000s as part of product lifecycle management (PLM), but has since been seen repeatedly as an important strategic technology trend (STT), and numerous industrial publications [3] were published on the topic. The three major application areas are manufacturing, energy and transportation.

Note that the difference between a simulation model and a twin is that a simulation model is needed to keep up a 'twin', and measured data needs to be updated when the asset changes, and the simulation model can be dynamically adapted to reflect changes in the asset [4]. This review will be for a self-calibrating mechanical DT (SCMDT), that is, a twin with re-estimation of the gear-mesh, contact and degradation parameters from the measurements, online. The goals include establishing the theoretical bases, comparing calibration/evaluation and fault localization approaches, quantifying the extent of research in this field and identification of challenges to be addressed for use in industry.

LITERATURE REVIEW

2.1 Digital Twin Foundations and Self-Calibration

A five-dimensional DT architecture for the physical entity, virtual entity, services, twin data and interactions between them was developed and applied to a wind turbine gearbox for the purpose of prognostics and health management (PHM) of complex equipment [5]. In this case, the virtual entity is continually compared to the physical entity, and the difference between them is regarded as a deficiency of the model of the physical entity, which should be corrected, and/or an indication of the onset of fault in the physical entity that should be diagnosed; with the conceptual foundations for self-calibration.

The values, challenges and enablers of DTs were then discussed from a modelling perspective by comparing modelling approaches namely physics-based, data-driven and hybrid analysis and modelling (HAM) [6]. Physics-based models were found to be opaque, but were found to be fast at inference and generalizing beyond the training range, while data-driven models were found to be fast at inference but opaque. A modelling approach that integrates different models was identified as the most feasible, if feasible also real-time model order reduction, uncertainty quantification, and secure data infrastructure.

Two kinds of twinning, virtual-to-physical (decisions are returned to the asset) and physical-to-virtual (states and parameters are inferred from the sensors) were also identified in a comprehensive analysis of modelling and twinning technologies [7]. The need for reliable updates was cited as very important which included uncertainty quantification. The reviewed literature can then be summarized in terms of the direction and level of automation of data flow, how the

parameters are changed, and the different classes of virtual representation can then be delineated as summarized in Table 1.

Table 1. Classes of Virtual Representation and Their Calibration Characteristics

Class	Physical-to-virtual flow	Virtual-to-physical flow	Parameter updating	Typical gear application	Sources
Simulation model (virtual twin)	Manual, offline	None	None or one-off	Design-stage dynamic analysis	[4], [9]
Digital shadow	Automatic	Manual	Periodic, offline	Condition-indicator trending	[4], [7]
Digital twin	Automatic	Automatic	Continuous, online	PHM of wind turbine gearboxes	[5], [7]
Hybrid twin	Automatic	Automatic	Physics model plus learned correction	Diagnosis with residual learning	[6], [9]
Predictive (probabilistic) twin	Automatic, Bayesian	Automatic, decision-theoretic	Probabilistic, with uncertainty	Fleet-scale health assessment	[8]

Source: Synthesized from the reviewed literature on twin definitions and architectures.

A mathematical framework of predictive DTs was proposed to make all the elements of the physical state, all the data observed, the digital state, and the control actions random variables, and to establish their relationship through a dynamic decision network [8]. As shown in Eq. (1), the digital state, ϕt, at time step t is updated in the process of recursive Bayesian inference, given the measurement history, y1:t, and health parameters. The formalism is illustrated for an unmanned aerial vehicle (UAV) employing a library of reduced-order structural models, and can be easily adapted to other gear-health states such as crack depth or wear depth.

$$p(\theta_t \mid y_{1:t}) \propto p(y_t \mid \theta_t) \int p(\theta_t \mid \theta_{t-1}) p(\theta_{t-1} \mid y_{1:t-1}) \mathrm{d}\theta_{t-1} \quad (1)$$

The hybrid approach of "twin" was developed to overcome the challenge of data and physics, by splitting the response into two parts: a physics-based prediction uP with some new parameters μ and a new data-based correction ΔuD of the deviation or ignorance term, as specified in Eq. (2) [9]. The virtual twins were rated as offline simulation, digital twins as calibrated simulation and hybrid twins as calibrated simulation with learned corrections, model order reduction for real time assessment.

$$u_{\mathrm{HT}}(x, t; \mu) = u_{\mathrm{P}}(x, t; \mu) + \Delta u_{\mathrm{D}}(x, t) \quad (2)$$

2.2 Physics of Gear Transmission Faults

Using time-varying mesh stiffness as the primary internal excitation, the different meshes of the gearbox fault dynamic models can be classified into lumped-parameter models, finite element models and combined models as shown in Eq. (3), where the generalized coordinate vector is represented by q, the mass, damping and stiffness matrices are represented by M, C and K, respectively. The analytical potential energy method is a combination of Hertzian compliance (C1),

bending compliance (C2) and shear compliance (C3), which is listed in Table 2 and is given in Eq. (4).

$$\mathbf{M}\ddot{\mathbf{q}} + \mathbf{C}\dot{\mathbf{q}} + \mathbf{K}(t,\theta)\mathbf{q} = \mathbf{F}(t) \tag{3}$$

$$\frac{1}{k_t} = \frac{1}{k_h} + \sum_{i=1}^{2}\left(\frac{1}{k_{b,i}} + \frac{1}{k_{s,i}} + \frac{1}{k_{a,i}} + \frac{1}{k_{f,i}}\right) \tag{4}$$

Table 2. Components of Gear Mesh Stiffness and Their Sensitivity to Local Faults

Component	Symbol	Physical origin	Most affected by	Calibration relevance
Hertzian contact	k_h	Local elastic deformation at contact	Pitting, spalling, wear	Contact width updated from surface state
Bending	k_b	Cantilever bending of the tooth	Root crack, tooth breakage	Crack depth and angle as parameters
Shear	k_s	Transverse shear of tooth section	Root crack	Updated jointly with bending term
Axial compressive	k_a	Compression along tooth centreline	Crack (minor influence)	Usually held fixed
Fillet foundation	k_f	Deflection of gear body at root	Rim crack, thin rims	Rim crack path parameter

Source: Synthesized from gearbox dynamic modelling and cracked gear reviews.

The cracked gear systems are especially investigated, since a crack is generated at the tooth root and gradually changes the bending stiffness and shear stiffness of the tooth, resulting in the stiffness of the mesh to be lower during contact with the faulty tooth, and the occurrence of impulsive excitation once per revolution [10], [11]. For cracks running through the tooth and through the rim, certain stiffness reductions occur leading to very different stiffness reductions as a function of crack depth. For an idealized waveform, with a contact ratio of 1.6, the cracked tooth is 10%, 25% and 40% less stiff, as indicated in Figure 1.

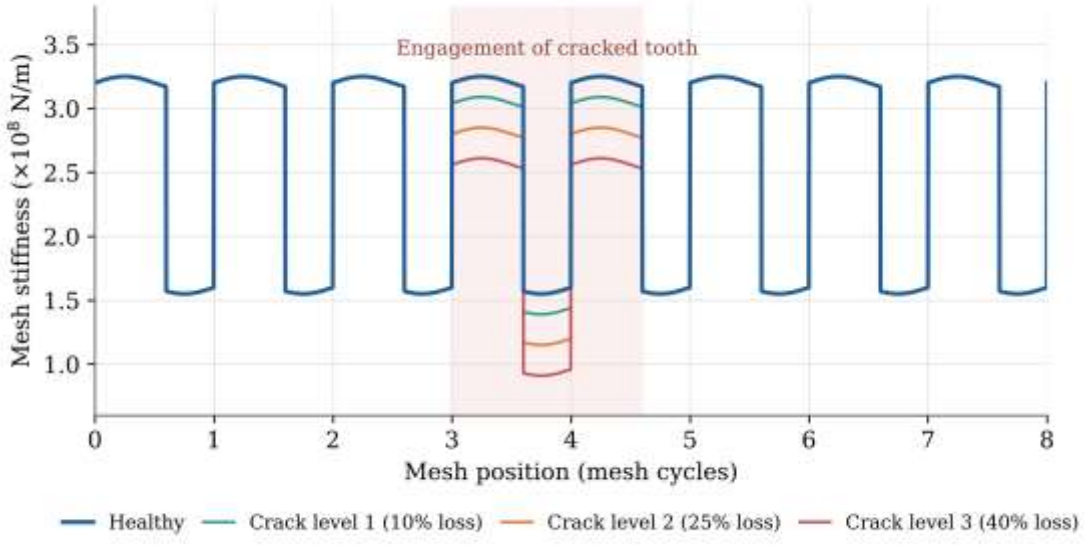


Figure 1. Idealized Time-Varying Mesh Stiffness of a Spur Gear Pair under Increasing Root Crack Severity

Source: Illustrative computation based on Eq. (4) and the mechanisms reported in the cracked gear literature.

Under gear wear condition that causes the amplitude modulation of vibration signal a(t) and the frequency modulation of vibration signal b(t), explicit models of vibration signals were obtained separately for planetary gearboxes [11] and under gear wear condition of sun, planet and ring gear, respectively, the fault characteristic frequencies were obtained [12]. Table 3 shows a simulated spectrum as a result of a localized fault on the sun gear, which is programmed into the relations in Figure 2, and is compared to a representative single stage gearbox.

$$x(t) = [1 + a(t)]\cos[2\pi f_m t + b(t) + \varphi] \tag{5}$$

$$f_m = Z_r f_c, \qquad f_{\text{sun}} = \frac{Nf_m}{Z_s}, \qquad f_{\text{planet}} = \frac{f_m}{Z_p}, \qquad f_{\text{ring}} = \frac{Nf_m}{Z_r} \tag{6}$$

Table 3. Fault Characteristic Frequencies for a Representative Planetary Stage

Quantity	Expression	Value (Hz)
Sun shaft input frequency	f_s	20.00
Carrier rotation frequency	$f_c = f_s Z_s/(Z_s + Z_r)$	4.35
Mesh frequency	$f_m = Z_r f_c$	313.04
Sun gear fault frequency	Nf_m/Z_s	46.96
Planet gear fault frequency (one flank)	f_m/Z_p	12.04
Planet gear fault frequency (both flanks)	$2f_m/Z_p$	24.08
Ring gear fault frequency	Nf_m/Z_r	13.04

Note: $Z_s = 20$, $Z_p = 26$, $Z_r = 72$, $N = 3$, ring gear fixed. Source: Computed from Eq. (6).

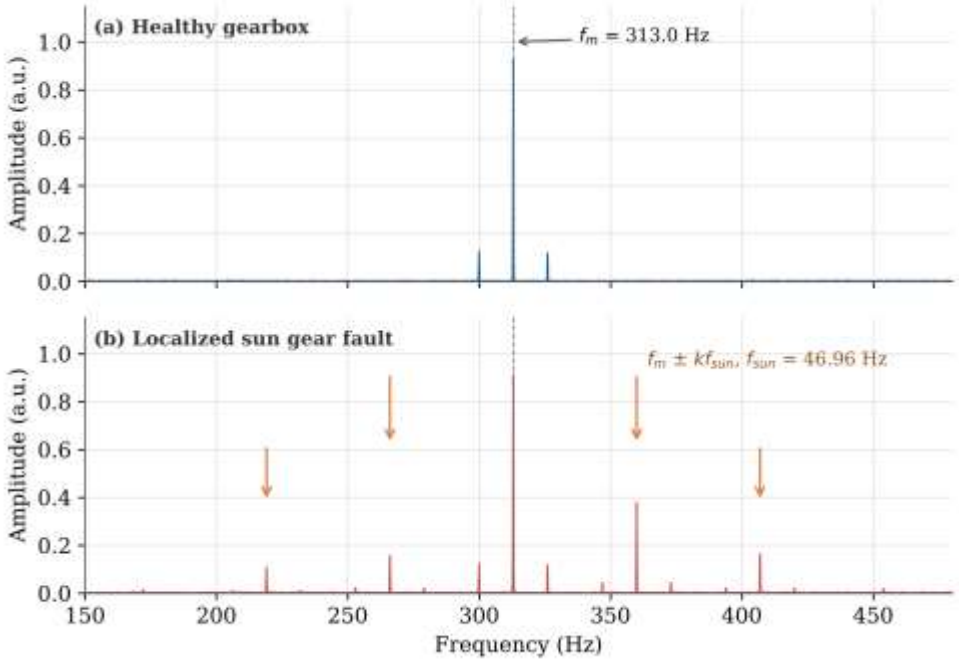


Figure 2. Simulated Vibration Spectra of the Planetary Stage in Table 3: (a) Healthy and (b) Localized Sun Gear Fault

Source: Illustrative computation based on Eqs. (5) and (6).

This type of model can be matched with measurements using a vibration based condition monitoring approach such as time synchronous averaging, which can separate individual shafts, envelope analysis, spectral kurtosis, which can identify impulsive faults, and cepstral analysis, which can identify sets of sidebands [13]. These operations are used as feature extractors for the DT and the outputs of these operations are the observation vector which is used to calibrate the model with the prediction results.

2.3 Data-Driven and Physics-Informed Learning

Machine learning based machine fault diagnosis was represented as three stages of evolution: Use of traditional machine learning with handcrafted features with ANNs or SVMs, automatic feature extraction with deep learning and transfer learning to transfer machine knowledge between machines and conditions [14]. A recurring difficulty in all stages is that sufficient data for faults is not available in the labeled form from real equipment and this leads to a tendency towards generating synthetic data in a physically consistent way with calibrated DTs.

Physics-informed neural networks (PINNs) insert the physics into the training loss, forcing the network with weights w to be both a good approximation of Nu observations and a good approximation of the physics at Nf collocation points as illustrated in Eq. (7) [15]. A good relative L2-accuracy of approximately 6.7 × 10−4 was achieved with only 100 initial and boundary samples and 10,000 collocation points in the case of the Burgers equation, while in the calibration problem of unknown equation coefficients, the same scheme was applied.

$$L(w,\theta) = \frac{1}{N_u}\sum_{i=1}^{N_u}|u(x_i;w) - u_i|^2 + \lambda\frac{1}{N_f}\sum_{j=1}^{N_f}\left|R\left(x_j;w,\theta\right)\right|^2 \quad (7)$$

Later on, physics-informed machine learning was proposed as the combination of three biases: observational, inductive and learning [16]. For gear DTs, it can be useful because the three kinds of meshing periodicity and the three kinds of kinematic constraints and energy balance can be used in any of the three ways, which will have an impact on the amount of data required to be trustworthy for the calibration.

In [17] the authors presented one concrete application wherein the physics informed hyperparameter optimization of long short-term memory (LSTM) networks was performed for a gearbox based on its physical properties and not on an exhaustive search. Correlation of the receptive field of the detector with physically meaningful periodicities of faults helped in tuning and showed an extremely simple form of physics guidance that could be used in an embedded real-time application.

METHODS

3.1 Review Protocol

The sources included in the corpus are 30 peer reviewed sources published between the years 2011 and 2024 and analyzed using a systematic review protocol. Coding was based on DT-driven diagnosis of rotating machinery where the model parameters were updated based on sensitivity analysis before the diagnosis, studies were retained if they discussed the dynamics of

gears or rotating machinery, DT architectures, physics-informed learning, degradation modelling or DT-based diagnosis. They were themed, year coded, venue coded, modelling paradigm coded, calibrated parameters coded and localization capability coded for each source. As indicated in Table 4 and Figure 3, there were seven sources (23.3%) that appeared prior to 2021, but 17 sources appeared after that year.

Table 4. Thematic Profile of the Reviewed Corpus

Theme	Sources	Share (%)	Span	Dominant contribution
DT foundations and architectures	7	23.3	2018–2022	Definitions, architectures, Bayesian and hybrid formalisms
Gear dynamics and vibration physics	6	20.0	2011–2021	Mesh stiffness, signal models, diagnostic reviews
Physics-informed and data-driven learning	4	13.3	2019–2022	PINNs, learning biases, ML roadmap, PI-LSTM
Multi-physics degradation modelling	5	16.7	2022–2024	Wear, pitting, and crack growth twins
DT-based diagnosis and localization	6	20.0	2019–2023	Model updating, data fusion, TPA, multibody models
Drivetrain field monitoring	2	6.7	2022	Virtual sensing, online parameter tracking
Total	30	100.0	2011–2024	—

Source: Coding of the reviewed corpus.

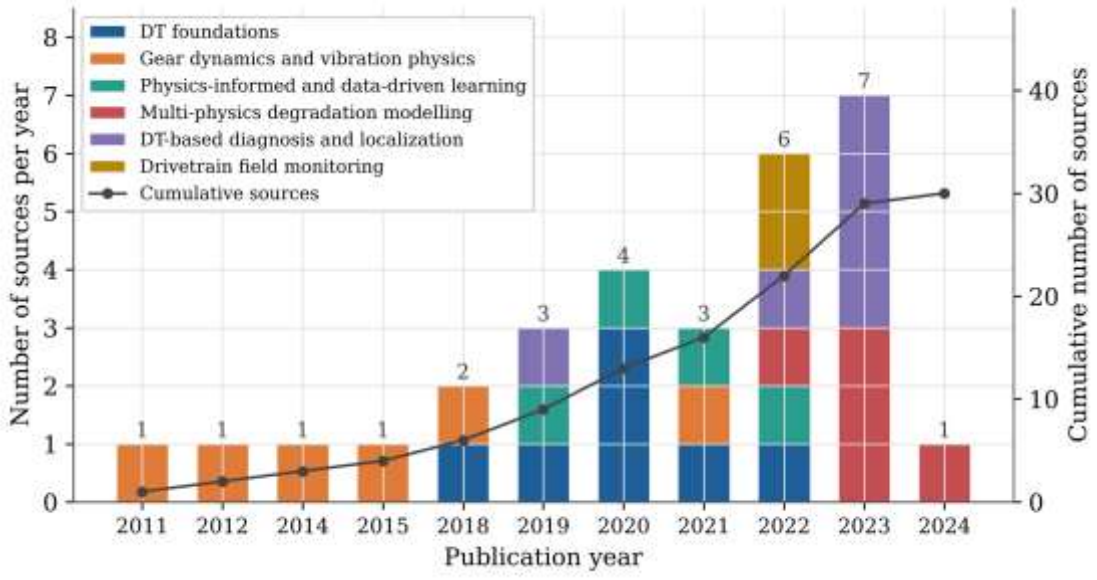


Figure 3. Annual and Cumulative Distribution of the Reviewed Corpus by Theme

Source: Coding of the reviewed corpus.

3.2 Self-Calibration Formulation

A synthesized framework where the minimization of the weighted difference between the measurements yk and the prediction of the model k with the prior parameter values θ0 is formulated as a self-calibration. This method has been used with gearboxes, where the subset of parameters are first selected and tuned to get simulated and measured responses to agree and then fault samples are generated by the updated model during the diagnosis. The closed loop architecture obtained is illustrated in Figure 4. In the example recursive estimation in Figure 5, after reducing the stiffness of the object by 30%, the estimation error is kept in the range of -3% to +3% in 27 samples and the RMS of estimation error is 0.017 and the mean estimation error of the static twin is 32.7%.

$$\hat{\theta} = \arg\min_{\theta} \left[\sum_{k=1}^{K} w_k \left(y_k - \hat{y}_k(\theta) \right)^2 + \beta \|\theta - \theta_0\|^2 \right] \quad (8)$$

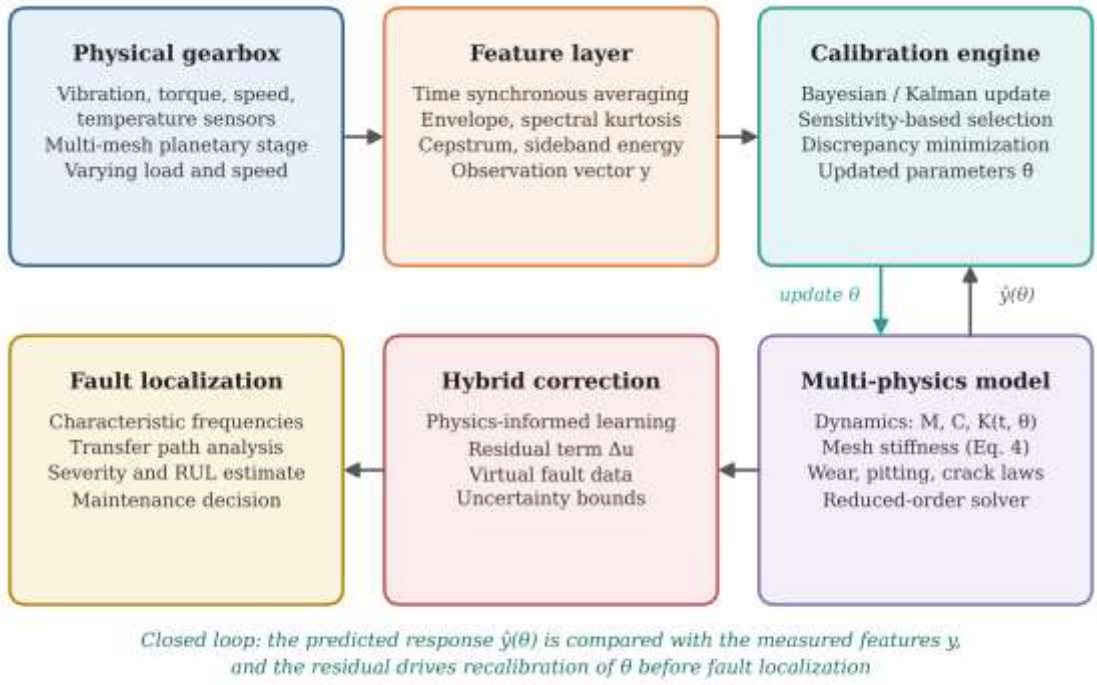


Figure 4. Closed-Loop Architecture of a Self-Calibrating Mechanical Digital Twin for Gear Fault Localization

Source: Synthesized framework.

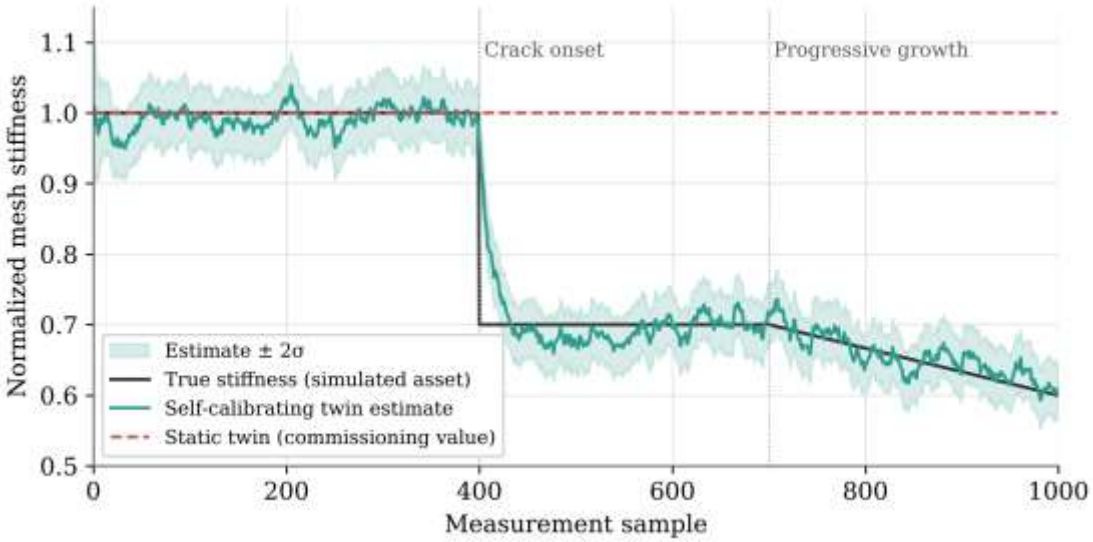


Figure 5. Recursive Mesh Stiffness Calibration Compared with a Static Twin during Crack Onset and Growth

Source: Illustrative computation based on Eqs. (1) and (8).

RESULTS AND DISCUSSION

4.1 Multi-Physics Degradation Twins

The vibration based gear wear monitoring and prediction methods were reviewed and the general expression of wear was shown as Eq. (9) [18] if the coefficient of wear is kw, contact pressure is p and the distance of sliding is s. Tribology and dynamics are mutually coupled, which refers to the variation of meshes' stiffness and transmission error as the wear occurs, and needs to be fine-tuned in an iterative manner during service operation.

$$h_{n+1}(x) = h_n(x) + k_w p_n(x) s_n(x) \tag{9}$$

The method is based on a model of tribo-dynamic pitting evolution of planetary gear sets introduced, which includes dynamic contact pressure calculated using the dynamic model as a source of fatigue damage, and formation of pits on the surface of the tooth, followed by updating the topography and using the updated topography for calculation of the mesh stiffness and dynamic response [19].

A life-cycle dynamic wear degradation model of planetary gear systems was developed that modeled running-in, steady and accelerated running-in regimes, in which the wear depth and dynamic response co-evolved [20]. The three stages of wear are illustrated in Figure 6 calculated from Eq. (9) where the coefficients are stage-dependent; in the final 25% of the gearbox's life the accelerated stage accounts for 5.3 μm (43.3%) of the total wear.

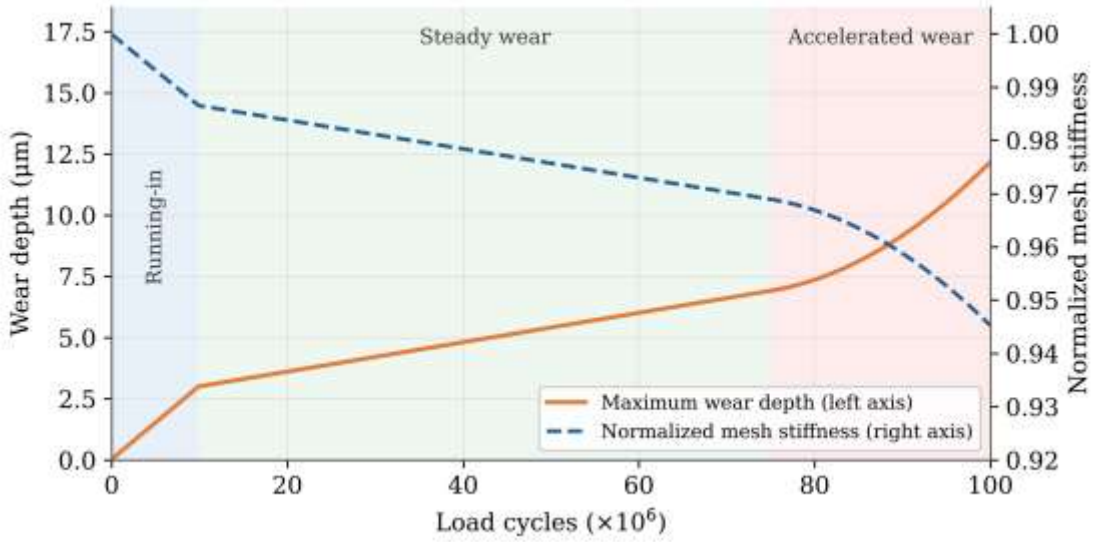


Figure 6. Three-Stage Life-Cycle Wear Trajectory and Associated Mesh Stiffness Reduction
Source: Illustrative computation based on Eq. (9) and the three-stage behaviour reported for planetary gears.

In this, the intelligent assessment of gear surface degradation was applied by using the dynamic model and the assessment module of surface damage, and matching the vibration signals produced by the gear degradation to the field signals [21]. The phenomena of degradation, controlling laws and new variables found in the literature are listed below and shown in the table below (Table 5) to show that geometric state updating is needed for surface degradation and stiffness state updating is needed for crack degradation.

Table 5. Multi-Physics Degradation Mechanisms in Gear Digital Twins

Mechanism	Governing relation	Coupled physics	Updated twin state	Observable	Sources
Sliding wear	Archard law, Eq. (9)	Contact, tribology, dynamics	Tooth profile, k_h	Transmission error, mesh harmonics	[18], [20]
Pitting and spalling	Fatigue damage with topography update	Contact pressure, lubrication, dynamics	Surface topography	Modulation sidebands	[19], [21]
Root crack	Paris law, Eq. (10)	Fracture mechanics, dynamics	Crack depth, k_b, k_s	Once-per-revolution impulses	[11]
Combined faults	Superposed stiffness changes	Multibody dynamics	Several stiffness functions	Mixed sideband families	[22]
Drivetrain fatigue	Palmgren–Miner rule, Eq. (11)	Structural dynamics, external loads	Accumulated damage	Virtual sensor loads	[23]

Source: Synthesized from the degradation and drivetrain literature.

The developed DT for gear root crack prognosis was a dynamic model of the cracked gear pair (2) coupled with a crack growth law (10) (canonical form of the crack growth law) and used to propagate the crack length a estimated from vibration over load cycles N to estimate the remaining life [24]. The accuracy of the prognosis has been shown to depend not only on the state but also on the re-calibration of the state and of the still uncertain growth parameters C and m with new measurements.

$$\frac{\mathrm{d}a}{\mathrm{d}N} = C(\Delta K)^m \tag{10}$$

4.2 Fault Localization Capability

The issue of localization is split into two parts: Firstly the twin needs to tell that damage has been done, and secondly he needs to mark the damage point. The method used for tracing the fault within the gear system was developed based on an in-situ measurement based transfer path analysis method, in which the effect of the individual excitation sources and the effect of transmission paths on the response in the housing were separated in order to be able to trace the faulty component back to the source [25]. This is in addition to localization by frequency, which can be ambiguous if characteristic frequencies are the same; in Table 3 the planet fault frequency is only 1.0 Hz away from the ring fault frequency.

A multi-body dynamics model of planetary gearbox combination failures was built using a DT and it was found that simultaneous failure of several components leads to interacting modulation patterns which are difficult to be separated through single failure templates [22]. Two responses from the twin were therefore simulated for combining fault analysis, to enable the identification of compound faults as opposed to single faults, and to enable fault localization from one component to multiple damage sites.

In [26] a DT driven approach was taken to develop virtual data from a refined twin and combine them with a few physical measurements to train a diagnostic network, due to the lack of fault samples. The approach is based on the "step of the refinement" as samples from a not-calibrated model are not distributed in the same way as real samples and may decrease rather than enhance the generalization abilities of the classifiers.

The idea of fault diagnosis for gearboxes using a fusion of the physical model and data-driven approach was also similar: The physical model would provide a set of interpretable fault mechanisms, which were to be captured by the data-driven model as the remaining complexity [27]. Table 6 and Figure 7 show a comparative analysis, where the self-calibrating hybrid twins are the most promising option in terms of data demand, localization granularity and robustness, but require more modelling and computational work compared to frequency matching.

Table 6. Comparative Analysis of Fault-Localization Strategies for Gear Transmissions

Strategy	Physics required	Labelled data need	Localization granularity	Computational cost	Interpretability	Sources
Characteristic-frequency matching	Kinematics only	None	Component	Low	High	[12], [13]

Transfer path analysis	Path transfer functions	None	Component and path	Medium	High	[25]
Deep learning classifier	None	High	Fault class	Low inference, high training	Low	[14], [17]
Calibrated dynamic twin with virtual data	Full dynamic model	Low	Component and severity	High	Medium to high	[19], [27]
Self-calibrating hybrid twin	Reduced model plus correction	Low to medium	Component, severity, compound faults	Medium	Medium	[9], [28]

Source: Qualitative synthesis of the reviewed literature.

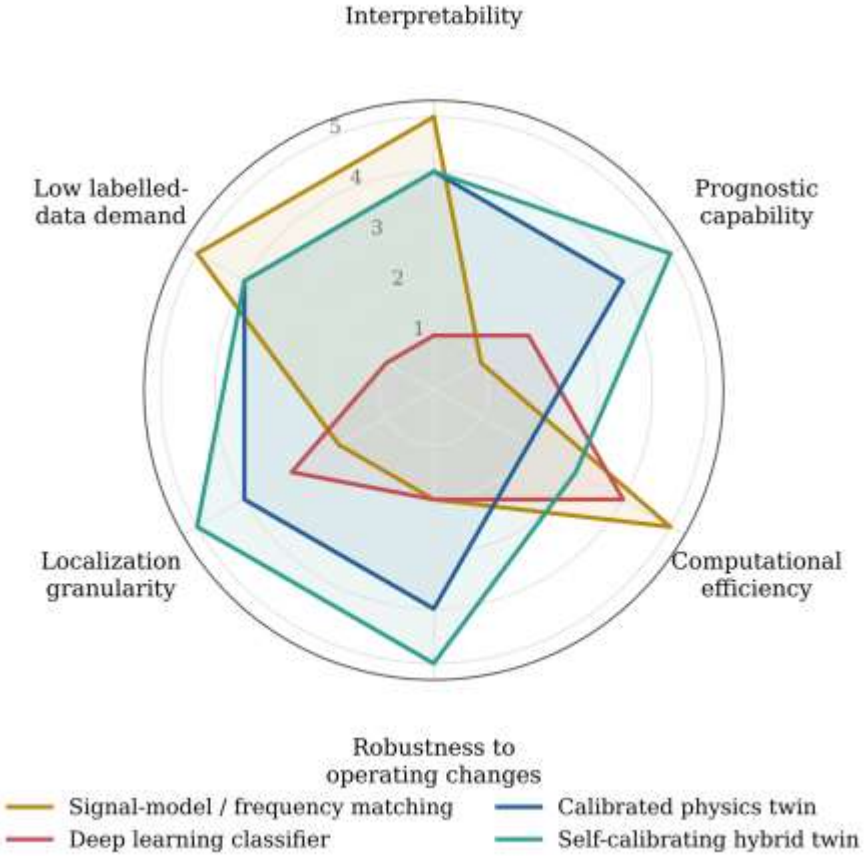


Figure 7. Qualitative Comparison of Fault-Localization Paradigms across Six Criteria (Scale 1–5)

Source: Qualitative synthesis of Table 6.

4.3 Deployment, Scalability, and Open Challenges

The most evident fields of use for the field deployment are the wind turbine drivetrains. Application of a DT-based virtual sensor for the offshore wind turbines led to the estimation of internal loads of a wind turbine drivetrain that cannot be measured directly, and the online

calculation of fatigue damage that the wind turbine suffered from using the Palmgren–Miner rule in Eq. (11) by setting ni and Ni as the number of applied and allowable cycles for load level i, respectively [23]. Apart from the benefits of the virtual sensing, selfcalibrating twins would also be useful for diagnosis, but with the introduction of loadaware prognosis in the future, the need for continuous calibration is further emphasized.

$$D = \sum_{i=1}^{n} \frac{n_i}{N_i} \quad (11)$$

In [30] a method for online condition monitoring of floating wind turbine drivetrains by using an estimated DT derived from measured signals to monitor parameters of the drivetrain (e.g., torsional stiffness) and deviations of these parameters from their estimated values as fault indicators was proposed. The evolution, shown in Table 7, shows that the models are moving from offline signal models to probabilistic, online and multi-physics models for twins. But these models are still expensive, many of these parameters are still not measurable, placement of sensors is still an issue, cyber security and data ownership with connected assets are still to be sorted, and there is no validation of safety critical certification.

Table 7. Evolutionary Trend of Gear Digital Twin Capability in the Reviewed Corpus

Period	Sources	Dominant capability	Representative developments	Calibration mode
2011–2015	4	Signal models and diagnostics	Planetary signal models, cracked gear dynamics, monitoring practice	None (offline models)
2018–2019	5	DT architectures and first updating	Five-dimensional DT, fault modelling review, PINNs, sensitivity-based updating	Offline, deterministic
2020–2021	7	Hybrid and probabilistic formalisms	Hybrid twins, HAM, model–twin distinction, Bayesian networks, learning biases	Online concepts, probabilistic
2022	6	Multi-physics and field twins	Pitting topography updating, PI-LSTM, virtual sensing, floating wind DT	Online parameter tracking
2023–2024	8	Integrated degradation and fusion twins	Surface degradation DT, crack prognosis, TPA, data fusion, life-cycle wear	Online, multi-physics, fused data

Source: Coding of the reviewed corpus by publication year.

CONCLUSION

This systematic review has demonstrated that, for the industrial world, a coherent way forward for multi-physics fault localisation can be provided by self-calibrating mechanical digital twins. These different formulations were combined in a single closed-loop through synthesis. These were Bayesian updating of the state, hybrid decomposition, modelling of mesh stiffness, planetary signal models, degradation laws and discrepancy minimizing calibration. The analysis indicated that: a) static twins degrade over time as they age, wear, pit and crack; b) In the frequency domain, two similar characteristic frequencies are hard to interpret and have a high number of labelled-datas

required; c) continuously calibrated hybrid twins are interpretable and have low numbers of labelled-datas required. The literature has been increasing rapidly and most of the sources mentioned in this literature review are published since 2021, although the literature on the field validation of wind turbine drivetrains has been focused. Efficient reduced-order multi-physics solvers, identifiability analysis of coupled degradation parameters, uncertainty aware calibration and distinguishing model error and degradation, benchmark datasets with known Run to Failure history and on-going validation protocols for certification of twins in safety critical transmissions should be a future research priority.

REFERENCES

[1] Y. Lei, J. Lin, M. J. Zuo, and Z. He, "Condition monitoring and fault diagnosis of planetary gearboxes: A review," *Measurement*, vol. 48, pp. 292–305, 2014, doi: 10.1016/j.measurement.2013.11.012.

[2] P. Kundu, A. K. Darpe, and M. S. Kulkarni, "A review on diagnostic and prognostic approaches for gears," *Struct. Heal. Monit.*, vol. 20, no. 5, pp. 2853–2893, 2021, doi: 10.1177/1475921720972926.

[3] F. Tao, H. Zhang, A. Liu, and A. Y. C. Nee, "Digital twin in industry: State-of-the-art," *IEEE Trans. Ind. Informatics*, vol. 15, no. 4, pp. 2405–2415, 2019, doi: 10.1109/TII.2018.2873186.

[4] L. Wright and S. Davidson, "How to tell the difference between a model and a digital twin," *Adv. Model. Simul. Eng. Sci.*, vol. 7, no. 1, 2020, doi: 10.1186/s40323-020-00147-4.

[5] F. Tao, M. Zhang, Y. Liu, and A. Y. C. Nee, "Digital twin driven prognostics and health management for complex equipment," *CIRP Ann.*, vol. 67, no. 1, pp. 169–172, 2018, doi: 10.1016/j.cirp.2018.04.055.

[6] A. Rasheed, O. San, and T. Kvamsdal, "Digital twin: Values, challenges and enablers from a modeling perspective," *IEEE Access*, vol. 8, pp. 21980–22012, 2020, doi: 10.1109/ACCESS.2020.2970143.

[7] A. Thelen *et al.*, "A comprehensive review of digital twin—Part 1: Modeling and twinning enabling technologies," *Struct. Multidiscip. Optim.*, vol. 65, no. 12, 2022, doi: 10.1007/s00158-022-03425-4.

[8] M. G. Kapteyn, J. V. R. Pretorius, and K. E. Willcox, "A probabilistic graphical model foundation for enabling predictive digital twins at scale," *Nat. Comput. Sci.*, vol. 1, no. 5, pp. 337–347, 2021, doi: 10.1038/s43588-021-00069-0.

[9] F. Chinesta, E. Cueto, E. Abisset-Chavanne, J. L. Duval, and F. El Khaldi, "Virtual, digital and hybrid twins: A new paradigm in data-based engineering and engineered data," *Arch. Comput. Methods Eng.*, vol. 27, no. 1, pp. 105–134, 2020, doi: 10.1007/s11831-018-9301-4.

[10] X. Liang, M. J. Zuo, and Z. Feng, "Dynamic modeling of gearbox faults: A review," *Mech. Syst. Signal Process.*, vol. 98, pp. 852–876, 2018, doi: 10.1016/j.ymssp.2017.05.024.

[11] H. Ma, J. Zeng, R. Feng, X. Pang, Q. Wang, and B. Wen, "Review on dynamics of cracked gear systems," *Eng. Fail. Anal.*, vol. 55, pp. 224–245, 2015, doi: 10.1016/j.engfailanal.2015.06.004.

[12] Z. Feng and M. J. Zuo, "Vibration signal models for fault diagnosis of planetary gearboxes," *J. Sound Vib.*, vol. 331, no. 22, pp. 4919–4939, 2012, doi: 10.1016/j.jsv.2012.05.039.

[13] R. B. Randall, *Vibration-Based Condition Monitoring: Industrial, Aerospace and Automotive Applications*. Wiley, 2011. doi: 10.1002/9780470977668.

[14] Y. Lei, B. Yang, X. Jiang, F. Jia, N. Li, and A. K. Nandi, "Applications of machine learning to machine fault diagnosis: A review and roadmap," *Mech. Syst. Signal Process.*, vol. 138, no. 106587, 2020, doi: 10.1016/j.ymssp.2019.106587.

[15] M. Raissi, P. Perdikaris, and G. E. Karniadakis, "Physics-informed neural networks: A deep learning framework for solving forward and inverse problems involving nonlinear partial differential equations," *J. Comput. Phys.*, vol. 378, pp. 686–707, 2019, doi: 10.1016/j.jcp.2018.10.045.

[16] G. E. Karniadakis, I. G. Kevrekidis, L. Lu, P. Perdikaris, S. Wang, and L. Yang, "Physics-informed machine learning," *Nat. Rev. Phys.*, vol. 3, no. 6, pp. 422–440, 2021, doi: 10.1038/s42254-021-00314-5.

[17] Y. Chen, M. Rao, K. Feng, and M. J. Zuo, "Physics-informed LSTM hyperparameters selection for gearbox fault detection," *Mech. Syst. Signal Process.*, vol. 171, no. 108907, 2022, doi: 10.1016/j.ymssp.2022.108907.

[18] K. Feng, J. C. Ji, Q. Ni, and M. Beer, "A review of vibration-based gear wear monitoring and prediction techniques," *Mech. Syst. Signal Process.*, vol. 182, no. 109605, 2023, doi: 10.1016/j.ymssp.2022.109605.

[19] Y. Huangfu, X. Dong, K. Chen, G. Tu, X. Long, and Z. Peng, "A tribo-dynamic based pitting evolution model of planetary gear sets: A topographical updating approach," *Int. J. Mech. Sci.*, vol. 220, no. 107157, 2022, doi: 10.1016/j.ijmecsci.2022.107157.

[20] Y. Huangfu, X. Dong, Y. Cao, Z. Li, Z. Peng, and Y. Sun, "A life-cycle dynamic wear degradation model of planetary gear systems," *Wear*, vol. 542–543, no. 205281, 2024, doi: 10.1016/j.wear.2024.205281.

[21] K. Feng, J. C. Ji, Y. Zhang, Q. Ni, Z. Liu, and M. Beer, "Digital twin-driven intelligent assessment of gear surface degradation," *Mech. Syst. Signal Process.*, vol. 186, no. 109896, 2023, doi: 10.1016/j.ymssp.2022.109896.

[22] D. Zhu, Z. Li, and N. Hu, "Multi-body dynamics modeling and analysis of planetary gearbox combination failure based on digital twin," *Appl. Sci.*, vol. 12, no. 23, 2022, doi: 10.3390/app122312290.

[23] F. C. Mehlan, A. R. Nejad, and Z. Gao, "Digital twin based virtual sensor for online fatigue damage monitoring in offshore wind turbine drivetrains," *J. Offshore Mech. Arct. Eng.*, vol. 144, no. 6, 2022, doi: 10.1115/1.4055551.

[24] O. Matania, E. Bechhoefer, and J. Bortman, "Digital twin of a gear root crack prognosis," *Sensors*, vol. 23, no. 24, 2023, doi: 10.3390/s23249883.

[25] Y. Huangfu, X. Dong, X. Yu, K. Chen, Z. Li, and Z. Peng, "Fault tracing of gear systems: An in-situ measurement-based transfer path analysis method," *J. Sound Vib.*, vol. 553, no. 117610, 2023, doi: 10.1016/j.jsv.2023.117610.

[26] J. Xia, R. Huang, Z. Chen, G. He, and W. Li, "A novel digital twin-driven approach based on physical-virtual data fusion for gearbox fault diagnosis," *Reliab. Eng. Syst. Saf.*, vol. 240, no. 109542, 2023, doi: 10.1016/j.ress.2023.109542.

[27] J. Yu, S. Wang, L. Wang, and Y. Sun, "Gearbox fault diagnosis based on a fusion model of virtual physical model and data-driven method," *Mech. Syst. Signal Process.*, vol. 188, no. 109980, 2023, doi: 10.1016/j.ymssp.2022.109980.