Applied Artificial Intelligence Research

INNOVIAIR

Data-Driven Predictive Maintenance of Permanent Magnet Synchronous Motor Systems Using Explainable Machine Learning and Digital Twin Architecture

Zhuopeng Gao[a,*]

[a]*School of Automation Science and Engineering, South China University of Technology, Guangzhou, Guangdong 510641, China*

A R T I C L E I N F O

Keywords:
Predictive Maintenance
Permanent Magnet Synchronous Motor
Explainable Machine Learning
SHAP
Digital Twin
Industry 4.0

A B S T R A C T

Permanent magnet synchronous motors (PMSMs) are widely used in industrial automation, robotics, electric drives, and intelligent manufacturing, yet their maintenance still depends heavily on threshold alarms and periodic inspection. This paper presents a data-driven predictive maintenance framework for PMSM systems by integrating explainable machine learning with a digital twin architecture. A self-consistent synthetic PMSM digital twin dataset is constructed to represent thermal, electrical, mechanical, control-related, and aging degradation signals, including winding temperature, vibration RMS, current harmonic distortion, q-axis current ripple, DC-link voltage ripple, speed tracking error, flux-linkage index, insulation resistance, load ratio, and operating hours. Random Forest, XGBoost, and LightGBM are trained to classify the motor health state into healthy, warning, and fault conditions. The results show that LightGBM achieves the best held-out performance, with 95.00% accuracy, 0.9256 macro F1-score, and 0.9914 macro ROC-AUC. A 5-fold stratified cross-validation experiment confirms the stability of the model, while an ablation study verifies the contribution of thermal, mechanical, electrical, control, and aging feature groups. An additional AI4I 2020 benchmark validation is included to test whether the modeling pipeline generalizes to a public predictive-maintenance dataset. SHAP analysis indicates that winding temperature, vibration RMS, insulation resistance, speed error, and current harmonic distortion are dominant contributors to fault prediction. Finally, a deployable digital twin architecture is proposed for automation scenarios. The study provides a reproducible simulation-based reference for intelligent motor-drive maintenance under Industry 4.0.

1.Introduction

Permanent magnet synchronous motors (PMSMs) have become a core actuation and energy-conversion component in industrial automation because they combine high torque density, high efficiency, precise speed regulation, compact structure, and good dynamic response. In modern manufacturing lines, PMSM drives are used in servo axes, conveyor modules, robotic joints, computer numerical control equipment, elevator systems, pumps, compressors, and flexible production cells. Their operating condition directly influences product quality, production continuity, energy consumption, and equipment safety. A local abnormality in a PMSM subsystem, such as thermal overload, partial demagnetization, bearing degradation, inverter-induced current distortion, insulation aging, or mechanical imbalance, can gradually propagate into a production-level fault. For this reason, the health monitoring of PMSM drive systems is not only an electrical-machine problem, but also an automation-system problem involving sensors, controllers, data acquisition, communication, diagnosis, and maintenance decisions.

Traditional maintenance strategies can be divided into corrective maintenance and preventive maintenance. Corrective maintenance waits until a failure is observed and then restores the equipment. It is operationally simple, but it produces unplanned downtime, emergency labor cost, possible secondary damage, and unstable production scheduling. Preventive maintenance performs inspection or replacement according to fixed intervals, but it ignores individual equipment usage, load variation, environmental conditions, and accumulated degradation. Components can be

* Corresponding author.
E-mail address: 1341261403cpk@gmail.com.
https://doi.org/10.65455/30tg2482
Received 2 June 2026; Received in revised form 25 June 2026; Accepted 25 June 2026; Available online 30 June 2026

replaced too early, while severe faults can still occur between maintenance intervals. Condition-based maintenance and predictive maintenance were developed to overcome these limitations by using measurements and models to estimate equipment health before catastrophic failure occurs[1]. In Industry 4.0, predictive maintenance is further strengthened by cyber-physical systems, industrial Internet of Things (IIoT), edge computing, cloud analytics, and digital twins[2,3].

The predictive maintenance of PMSM systems is challenging for several reasons. First, motor degradation is multi-source and multi-physics. A temperature increase can result from excessive load, high current, poor cooling, bearing friction, inverter harmonics, or insulation deterioration. A vibration increase can represent bearing wear, shaft misalignment, torque ripple, eccentricity, resonance, or external load disturbance. Second, PMSM failures are often rare relative to normal operation, which produces imbalanced datasets and complicates supervised model training. Third, the measurement space is strongly coupled. Current RMS, current harmonic distortion, q-axis current ripple, torque, speed error, load ratio, and temperature jointly reflect the same operating process, so a single threshold cannot capture the underlying health state. Fourth, industrial users require not only a prediction result but also an explanation. A black-box alarm that labels a motor as faulty without identifying the dominant evidence has limited value for maintenance engineers.

Machine learning offers a practical route for predictive maintenance because it can learn nonlinear and coupled relationships from condition-monitoring data. Random Forest (RF), XGBoost, and LightGBM are particularly suitable for structured industrial data because they can model nonlinear interactions, handle mixed feature scales, and provide strong performance with modest computational cost[4-6]. However, high predictive accuracy alone is insufficient in automation scenarios. Maintenance engineers need to know which sensor variables drive the decision and whether the learned pattern is physically meaningful. Explainable artificial intelligence (XAI), especially SHAP values, provides a principled way to decompose each prediction into feature-level contributions[7]. Combining ensemble learning with SHAP can therefore transform predictive maintenance from a pure classifier into an interpretable diagnostic tool.

Digital twin technology provides the system-level architecture needed to deploy predictive maintenance in smart manufacturing. A digital twin can be understood as a continuously updated virtual representation of a physical asset, process, or production system. It connects sensor data, physics-based knowledge, data-driven models, and decision modules across the asset life cycle[8,9]. For PMSM systems, a digital twin can store the motor configuration, rated parameters, operational history, controller status, sensor streams, degradation indicators, and predictive model outputs. Once the digital twin receives updated measurements from the physical motor drive, it can evaluate the health state, produce an explanation, and issue maintenance recommendations. This architecture is more valuable than an isolated offline classifier because it closes the loop between data acquisition, health prediction, explainability, and action.

This paper proposes an explainable machine learning-based predictive maintenance framework for PMSM drive systems under a digital twin architecture. Since publicly available PMSM run-to-failure datasets remain limited, this study constructs a synthetic but physically self-consistent PMSM digital twin dataset. The design is inspired by public predictive maintenance resources such as AI4I 2020 and NASA C-MAPSS, but it is not presented as a measured factory dataset[10,11]. Instead, it is explicitly positioned as a reproducible simulation-based dataset that reflects typical PMSM health variables and degradation behavior. The dataset contains thermal, electrical, mechanical, and control-related features and three health labels: healthy, warning, and fault. RF, XGBoost, and LightGBM are trained and compared. SHAP is then used to identify the most influential variables and to support engineering interpretation.

The main contributions of this paper are as follows. First, a PMSM-oriented predictive maintenance dataset is generated by combining operating profile, latent degradation, thermal effects, electrical distortion, mechanical vibration, control tracking error, and aging-related variables, so that the simulation keeps explicit engineering links to motor-drive degradation. Second, three widely used ensemble learning models are evaluated under the same held-out test setting and further examined through 5-fold stratified cross-validation, producing a more reliable comparison than a single train-test split. Third, an ablation study is introduced to quantify the contribution of thermal, mechanical, electrical, control, and aging feature groups, thereby testing whether the prediction accuracy is supported by multi-physics evidence rather than by a single dominant variable. Fourth, an AI4I 2020 benchmark validation experiment is added as an external tabular predictive-maintenance check; this experiment does not replace PMSM validation, but it verifies the portability of the proposed RF/XGBoost/LightGBM pipeline on a public dataset[10,12,13]. Fifth, SHAP analysis is used to explain the trained LightGBM model at both global and engineering levels, connecting data-driven feature attribution with PMSM fault mechanisms. Sixth, an Industry 4.0 digital twin deployment architecture is proposed, showing how the model can be integrated with sensors, edge gateways, cloud services, human-machine interfaces, and maintenance action rules. These contributions make the paper a structured simulation study of intelligent predictive maintenance rather than a stand-alone classification exercise.

2.Related work

2.1.Predictive maintenance in Industry 4.0

Predictive maintenance has become a central application of artificial intelligence in Industry 4.0 because it directly links data analytics with production availability and maintenance cost. Early condition-based maintenance studies emphasized signal processing, trend analysis, statistical reliability, and decision-making based on equipment condition[1]. With the growth of connected sensors and industrial data platforms, predictive maintenance has shifted toward data-driven failure prediction, remaining useful life estimation, anomaly detection, and maintenance planning. Reviews of machine learning in predictive maintenance show that industrial assets increasingly rely on classification models, regression models, time-series learning, probabilistic reasoning, and hybrid physical-data methods[3,14].

The fourth industrial revolution changes the maintenance problem in three ways. First, more variables can be collected continuously from controllers, drives, sensors, and enterprise systems. Second, edge and cloud computing make it possible to process large volumes of data without stopping production. Third, maintenance decisions are integrated with manufacturing execution systems, spare-part management, scheduling, and operator interfaces. These features transform maintenance from an isolated repair activity into a decision-support component of smart manufacturing. In this context, data-driven predictive maintenance must satisfy not only model accuracy but also robustness, transparency, real-time feasibility, and deployability[15].

Public datasets have played an important role in accelerating research, but their domains and assumptions must be carefully interpreted. The AI4I 2020 dataset is a synthetic predictive maintenance dataset that reflects industrial operating variables such as temperature, rotational speed, torque, and tool wear[10]. The NASA C-MAPSS dataset is a simulated run-to-failure turbofan engine degradation dataset widely used for prognostics and remaining useful life prediction[11]. These datasets are valuable benchmarks, but they are not direct PMSM datasets. Therefore, when they are used to motivate motor-drive studies, the mapping between variables and physical motor states must be clearly stated. This paper follows the benchmark-data philosophy of reproducible simulation while constructing variables that are directly meaningful for PMSM drive systems.

Recent studies further show that the research frontier is moving from isolated diagnostic models toward deployable and explainable maintenance systems. Explainable AI has been introduced to make predictive-maintenance decisions more understandable to engineers[12]. Multi-sector reviews emphasize external validation, transparent evaluation metrics, and decision-oriented deployment in Industry 4.0[13]. In the PMSM domain, digital twin and hybrid diagnostic frameworks are increasingly used to combine simulated motor behavior, sensor streams, and data-driven diagnosis[16-18]. These studies motivate the present design: a PMSM-specific synthetic dataset is used for controlled development, while AI4I 2020 is used as an additional public benchmark to check whether the same modeling pipeline behaves consistently outside the PMSM simulation setting.

To further situate the present work within the research discourse of Applied Artificial Intelligence Research, recent studies in this journal are also considered. Xie et al.[19] developed a deep-learning-based performance prediction model for wind turbine gearbox lubricating oil, showing the relevance of AI models to equipment degradation and maintenance-oriented decision support. Kang et al.[20] proposed an AI-driven health assessment system for high-voltage switchgear based on multimodal sensor data fusion, which is closely related to the multi-sensor health monitoring logic adopted in this paper. Liu[21] applied explainable machine learning to actionable prediction, supporting the use of interpretability as a bridge between model output and decision-making. These journal studies provide additional context for the proposed PMSM-oriented explainable predictive maintenance framework.

2.2.PMSM health monitoring and intelligent fault diagnosis

PMSM fault diagnosis has been studied from both model-based and data-driven perspectives. Common fault types include bearing faults, stator inter-turn short circuits, partial demagnetization, eccentricity, inverter switching faults, cooling degradation, and sensor faults. Electrical signatures such as phase current, stator voltage, torque ripple, and harmonic content are widely used because they are readily available from the drive controller. Mechanical signatures such as vibration and acoustic emission can provide earlier evidence for bearing or alignment problems. Thermal measurements are also important because overheating accelerates insulation aging and magnet degradation. Recent studies on PMSM fault diagnosis have shown that machine learning can classify demagnetization, eccentricity, and electrical faults from measured or simulated signals[22-25].

Despite these advances, a gap remains between algorithm development and automation deployment. A fault classifier trained on a laboratory dataset does not automatically become a predictive maintenance system. Practical deployment requires data acquisition interfaces, feature pipelines, model monitoring, alarm thresholds, explainability, maintenance workflows, and periodic model updating. For PMSM systems in industrial automation, the most useful framework is therefore not a standalone fault labeler, but a connected health management architecture that combines the physical motor, the drive controller, the digital twin, and a maintenance decision layer. This is the reason this paper treats PMSM predictive maintenance as a cyber-physical and data-driven problem rather than only a signal-classification task.

2.3.Ensemble learning and explainable artificial intelligence

Tree-based ensemble methods are widely used for industrial tabular data because they offer high accuracy without requiring strict linear assumptions. Random Forest constructs multiple decision trees using bootstrap sampling and random feature selection, then aggregates their outputs to reduce variance and overfitting[4]. XGBoost is a scalable gradient boosting system that optimizes an additive ensemble of decision trees and includes regularization, sparsity-aware learning, and efficient tree construction[5]. LightGBM improves gradient boosting decision trees through histogram-based learning, leaf-wise tree growth, gradient-based one-side sampling, and exclusive feature bundling, making it efficient for large structured datasets[6]. These algorithms are well suited to predictive maintenance datasets in which nonlinear feature interactions are expected.

A key limitation of ensemble methods is that the resulting model is less transparent than a single threshold rule. Explainable AI addresses this issue by producing human-understandable explanations of model behavior. SHAP is based on Shapley values from cooperative game theory and assigns each feature a contribution to a model prediction[7]. Compared with simple feature importance, SHAP can be used for both global interpretation and local diagnosis. In predictive maintenance, this is valuable because it can indicate whether a warning decision is driven by thermal rise, vibration increase, electrical distortion, poor speed tracking, or insulation decline. LIME and other local explanation methods also support interpretability, but SHAP has become especially common for tree ensembles because TreeSHAP can compute consistent feature attributions efficiently[7,26].

2.4.Digital twin-based maintenance architecture

Digital twin research has expanded rapidly in manufacturing, product life-cycle management, and maintenance. A digital twin differs from a simple simulation model because it maintains a connection with the physical asset through data synchronization. Literature on digital twins distinguishes digital models, digital shadows, and digital twins according to the degree of automatic data flow between physical and virtual systems[9]. Predictive maintenance is one of the most natural applications of digital twin technology because maintenance decisions require both current sensor data and historical degradation context[27,28].

For motor-drive systems, a digital twin can include a physics-based motor model, a data-driven health model, a database of operating history, and an interface to industrial control systems. The physical system provides measurements such as current, voltage, speed, temperature, vibration, torque, and load. The digital twin stores these variables, computes health indicators, predicts future risk, and provides recommendations. The operator or maintenance management system then decides whether to continue operation, schedule inspection, reduce load, replace a component, or stop the machine. This paper adopts this architecture and places explainable machine learning at the health-prediction layer.

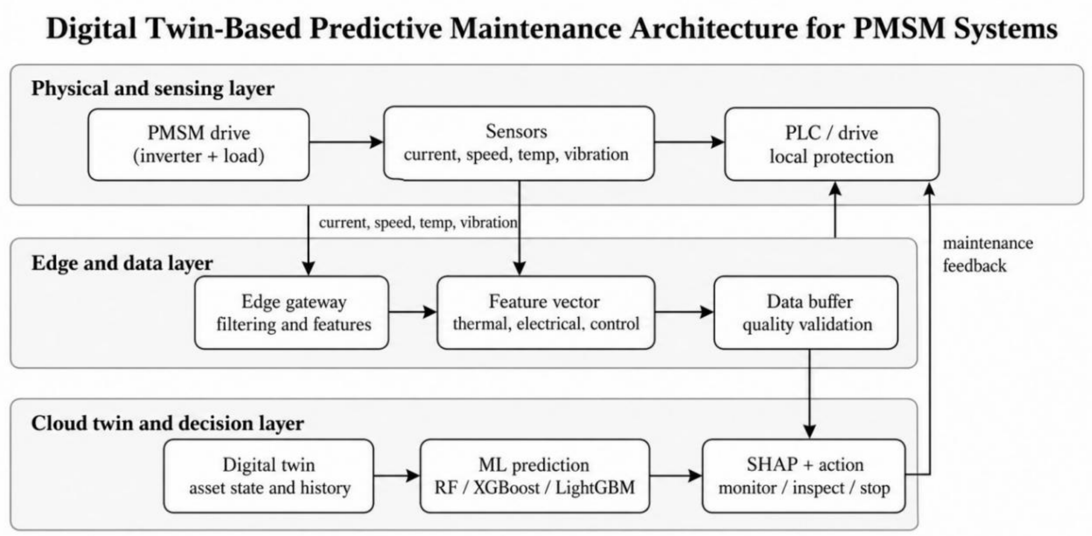

Fig 1. Digital twin-driven predictive maintenance architecture for PMSM systems

3.System formulation and digital twin architecture

3.1.PMSM drive system and monitored variables

The target system is an intelligent PMSM drive installed in an industrial automation environment. The physical subsystem contains a PMSM, a voltage-source inverter, a drive controller, a load mechanism, temperature sensors, current sensors, speed feedback, and a vibration sensor mounted near the bearing or motor housing. The control subsystem is assumed to implement vector control or servo speed control, so the drive can report not only raw electrical variables but also control-related indicators such as speed tracking error and current ripple. The data acquisition subsystem transmits selected features to an edge gateway, where filtering, normalization, and feature calculation are performed before the data are synchronized with the digital twin.

The health state of the PMSM drive is influenced by thermal stress, mechanical stress, electrical stress, and control performance. Thermal stress is represented by ambient temperature, stator temperature, winding temperature, and bearing temperature. Mechanical stress is represented by vibration RMS and load ratio. Electrical stress is represented by current RMS, current harmonic distortion, q-axis current ripple, DC-link voltage ripple, flux-linkage index, and insulation resistance. Control performance is represented by reference speed, actual speed, and speed error. Operating hours provide a cumulative aging indicator. These features are not independent. For example, high load increases torque and current, which can raise winding temperature. Increased degradation can reduce flux-linkage index, increase q-axis current ripple, enlarge speed error, and reduce insulation resistance. A realistic dataset should therefore preserve these correlations.

A simplified PMSM model can be described in the rotating dq reference frame. Although the proposed predictive maintenance model is data-driven, the equations help explain why electrical and control variables are included. The stator voltage equations are written as follows:

$$u_d=R_s i_d+L_d\frac{di_d}{dt}-\omega_e L_q i_q \tag{1}$$

$$u_q=R_s i_q+L_q\frac{di_q}{dt}+\omega_e\left(L_d i_d+\psi_f\right) \tag{2}$$

In these equations, u_d and u_q are dq-axis stator voltages, i_d and i_q are dq-axis stator currents, R_s is stator resistance, L_d and L_q are dq-axis inductances, ω_e is electrical angular speed, and ψ_f is permanent magnet flux linkage. Degradation can influence these variables through resistance increase, flux weakening, current distortion, torque ripple, and temperature rise. The electromagnetic torque can be approximated by

$$T_e = \frac{3}{2} p \left[\psi_f i_q + \left(L_d - L_q \right) i_d i_q \right] \tag{3}$$

where p is the number of pole pairs. Under normal field-oriented control with i_d near zero, torque is mainly related to ψ_f and i_q. If flux linkage decreases or q-axis current ripple increases, the drive must use more current to maintain the same torque, which can raise copper loss and temperature. This mechanism motivates the inclusion of current RMS, q-axis current ripple, torque, flux-linkage index, and temperature variables in the predictive maintenance dataset.

Table 1. PMSM digital twin variables used for predictive maintenance modeling.

Feature group	Variables	Physical interpretation
Thermal	ambient_temp_C; stator_temp_C; winding_temp_C; bearing_temp_C	Cooling condition, copper loss, bearing friction, insulation aging
Electrical	current_rms_A; current_thd_pct; q_axis_current_ripple_A; dc_link_ripple_V	Load current, inverter distortion, torque ripple, power quality stress
Mechanical	vibration_rms_g; torque_Nm; load_ratio	Bearing wear, imbalance, shaft/load stress, mechanical overload
Control	speed_ref_rpm; speed_actual_rpm; speed_error_rpm	Tracking ability of the servo drive and degradation-induced dynamic error
Aging	operating_hours; insulation_resistance_MOhm; flux_linkage_index	Cumulative operating history, insulation decline, demagnetization proxy

3.2.Health-state classification problem

The predictive maintenance task is formulated as a three-class supervised classification problem. Let x_i denote the feature vector of the i_{th} PMSM operating sample and y_i denote its health-state label. The label set is defined as Y = {healthy, warning, fault}. A healthy sample represents normal motor operation under acceptable thermal, mechanical, electrical, and control conditions. A warning sample represents early degradation, where abnormal indicators are present but immediate shutdown is not required. A fault sample represents severe degradation or high failure risk. The learning objective is to train a classifier f(x) that maps a condition-monitoring vector to a health state while maintaining high recall for the warning and fault classes.

In predictive maintenance, the warning class is especially important. If a model only separates healthy and fault samples, it behaves like a fault detector rather than a maintenance predictor. A warning class creates an intermediate decision region that allows maintenance teams to schedule inspection before the system reaches a critical fault. In the proposed digital twin architecture, a healthy prediction corresponds to continued operation, a warning prediction corresponds to scheduled inspection or load reduction, and a fault prediction corresponds to immediate intervention. The classifier output can also be combined with predicted probability and SHAP explanations to avoid unnecessary alarms.

The classifier produces a probability vector and a maximum-probability decision rule as follows:

$$\mathrm{p}_i = \left[p_{\mathrm{healthy}}, p_{\mathrm{warning}}, p_{\mathrm{fault}} \right] \tag{4}$$

$$\hat{y}_i = \underset{c \in \{\mathrm{healthy,warning,fault}\}}{\arg\max} \; p_{i,c} \tag{5}$$

For industrial deployment, the predicted probability can be combined with cost-sensitive thresholds. For example, if p_{fault} exceeds a safety threshold, the system may trigger urgent inspection even when the maximum-probability class is warning. This paper focuses on classifier comparison and explanation, while the deployment section discusses how such rules can be embedded in a digital twin.

3.3.Digital twin deployment architecture

The proposed architecture contains five layers. The physical layer consists of the PMSM, inverter, mechanical load, sensors, and drive controller. The sensing and communication layer collects current, speed, temperature, vibration, torque, and load data and transmits them through industrial Ethernet, fieldbus, or an IIoT gateway. The edge layer performs noise filtering, feature extraction, local buffering, and simple threshold protection. The digital twin layer stores asset configuration, historical data, health features, trained models, explanations, and maintenance records. The decision layer sends health-state predictions, SHAP-based explanations, and recommended actions to the human-machine interface, maintenance management system, or supervisory control system.

This architecture is intentionally designed for automation scenarios. It does not require the predictive model to replace the controller. The PMSM drive continues to execute real-time control tasks locally. The machine learning model operates at the monitoring and maintenance layer, where latency requirements are less strict than current-loop control. The edge device can run a lightweight version of the classifier for fast warning, while the cloud-based digital twin can retrain the model, compute more detailed SHAP explanations, and store the full equipment history. This separation improves safety because model errors do not directly change motor control signals, but still provide actionable maintenance intelligence.

The digital twin also supports model governance. When a maintenance action is performed, the result can be recorded in the twin and used to validate whether the alarm was correct. When the operating environment changes, such as a new load profile, a new cooling condition, or a different motor rating, the model can be recalibrated. This creates a continuous improvement cycle in which the digital twin is not only a passive visualization dashboard, but also an evolving health-management platform.

4.Synthetic dataset construction and experimental design

4.1.Simulation rationale

Real PMSM run-to-failure data are difficult to obtain because intentional fault injection can damage equipment, interrupt production, and create safety risks. Industrial companies are also reluctant to publish high-resolution drive data because it can reveal production processes, equipment configuration, and maintenance records. Therefore, this study constructs a synthetic PMSM digital twin dataset rather than claiming access to a real factory dataset. The purpose is to create a transparent, physically interpretable, and reproducible dataset for evaluating the proposed method. This is consistent with the role of public synthetic datasets in predictive maintenance research, such as AI4I 2020 and C-MAPSS[10,11].

The dataset generation process is based on three principles. First, the variables should correspond to measurable PMSM drive or automation-system indicators. Second, their statistical relationships should be physically reasonable. For instance, a high load ratio should increase torque and current, and severe degradation should increase vibration, temperature, harmonic distortion, and speed error. Third, the class boundary should not be perfectly deterministic. Industrial measurements include noise, environmental variation, sensor error, and overlapping fault signatures. Therefore, the generated classes contain partial overlap, making the classification problem realistic instead of trivial.

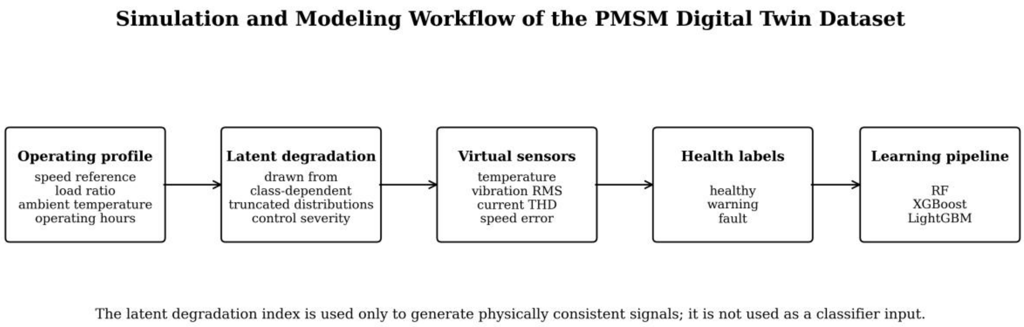

Fig 2. Workflow used to construct the synthetic PMSM digital twin dataset

4.2.Dataset generation mechanism

A latent degradation index z is introduced only inside the simulation model. The index ranges from 0 to 1, where smaller values correspond to healthy operation and larger values correspond to severe degradation. The classifiers do not receive z as an input feature. Instead, z influences the virtual sensor variables through thermal, electrical, mechanical, and control equations. For each sample, an operating profile is first generated, including speed reference, load ratio, ambient temperature, and operating hours. Then z is sampled from overlapping class-dependent distributions. Finally, virtual sensor variables are generated by combining operating profile, latent degradation, and random noise.

The following relationships are used conceptually. Load ratio raises torque and current. Current and degradation raise winding and stator temperatures. Bearing and mechanical degradation increase vibration RMS. Electrical degradation increases current total harmonic distortion, q-axis current ripple, and DC-link ripple. Magnet aging reduces the flux-linkage index. Insulation aging reduces insulation resistance as operating hours and degradation increase. Control performance degradation increases speed tracking error. Random noise is added to every feature to represent measurement uncertainty and environmental disturbance. As a result, the dataset keeps the engineering logic of PMSM degradation while preserving enough overlap between classes to evaluate model robustness.

To make the dataset generation mechanism reproducible, the health class is first sampled according to the class counts in Table 2. The latent degradation index is then generated from a class-dependent truncated normal distribution:

$$z_i=\text{clip}(\text{N}(\mu_c,\sigma_c^2),0,1),\quad c\in\{H,W,F\} \tag{6}$$

The class-dependent parameters used in the simulation are:

$$(\mu_H,\sigma_H)=(0.20,0.10) \tag{7}$$

$$(\mu_W,\sigma_W)=(0.52,0.11) \tag{8}$$

$$(\mu_F,\sigma_F)=(0.78,0.08) \tag{9}$$

The operating profile is generated from speed, load, ambient temperature, and cumulative operating hours:

$$n_i^{ref}\sim U(800,3000),\quad \lambda_i\sim U(0.25,1.00) \tag{10}$$

$$T_i^{amb}=\text{clip}(\text{N}(25,3^2),15,40),\quad \text{h}_i\sim U(200,18000) \tag{11}$$

Based on the operating profile and latent degradation index, the main load-dependent and thermal variables are generated as follows:

$$\tau_i=6+52\lambda_i+5z_i+\epsilon_{\tau,i} \tag{12}$$

$$I_i=8+18\lambda_i+5z_i+\epsilon_{I,i} \tag{13}$$

$$T_i^{stator}=T_i^{amb}+8+0.55I_i+18z_i+\epsilon_{s,i} \tag{14}$$

$$T_i^{winding}=T_i^{stator}+5+0.18I_i+25z_i+\epsilon_{w,i} \tag{15}$$

The mechanical, electrical, control, and aging-related variables are further generated by:

$$v_i=\text{clip}\left(0.05+0.10\lambda_i+0.80z_i^2+\epsilon_{v,i},0,+\infty\right) \tag{16}$$

$$H_i=\text{clip}\left(1.8+1.5\lambda_i+10z_i+\epsilon_{H,i},0,+\infty\right) \tag{17}$$

$$e_i^{speed}=\text{clip}\left(2+12\lambda_i+40z_i+\epsilon_{e,i},0,+\infty\right) \tag{18}$$

$$F_i=\text{clip}\left(1.02-0.24z_i-3\times10^{-6}\text{h}_i+\epsilon_{F,i},0.65,1.05\right) \tag{19}$$

$$R_i=\text{clip}\left(210-115z_i-0.004\text{h}_i+\epsilon_{R,i},5,220\right) \tag{20}$$

All generated variables are clipped to physically meaningful ranges. The latent degradation index is not included in the model input; only the virtual sensor features are used for training and testing. This setting preserves the causal role of degradation in the simulation while avoiding direct leakage of the hidden health index into the classifier.

Table 2. Health-state design of the synthetic PMSM digital twin dataset

State	Samples	Latent degradation range	Typical feature behavior
Healthy	3323	low, concentrated below 0.45	Temperature, vibration, speed error, harmonics and ripple remain near normal operating levels
Warning	1243	medium, overlapping healthy and fault regions	Moderate rise in thermal, electrical and control-error indicators; maintenance inspection is recommended
Fault	634	high, concentrated above 0.45	Severe thermal rise, vibration increase, insulation decline, current distortion and speed tracking deterioration

A total of 5200 samples are generated. The class distribution is intentionally imbalanced, with 3323 healthy samples, 1243 warning samples, and 634 fault samples. This reflects the typical condition of industrial predictive maintenance data, where normal operation is much more frequent than failure. The dataset is split into training and testing subsets using stratified sampling, with 75% of the samples used for training and 25% used for testing. No test samples are used during model training. The train-test split is fixed by a random seed to support reproducibility.

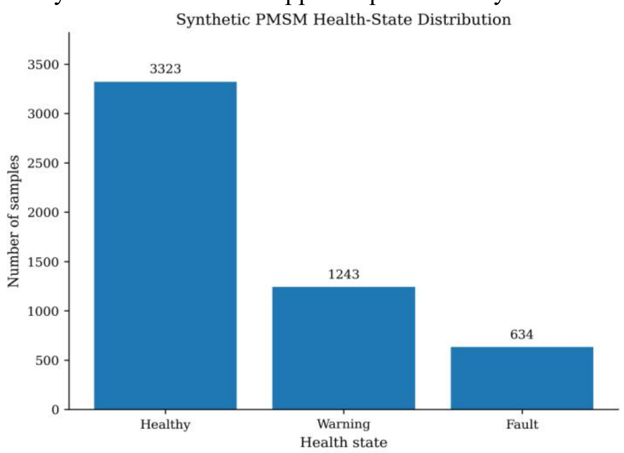

Fig 3. Class distribution of the synthetic PMSM health-state dataset

The generated distributions show expected engineering patterns. Winding temperature rises from healthy to warning and fault states because load current, copper loss, and degradation-related losses increase. Vibration RMS increases most clearly in the fault region, representing mechanical imbalance and bearing-related degradation. Current harmonic distortion increases with electrical stress and inverter-related distortion. Speed tracking error increases because degraded motor-drive behavior reduces the ability to follow the speed reference accurately. These patterns are illustrated in Figure 4.

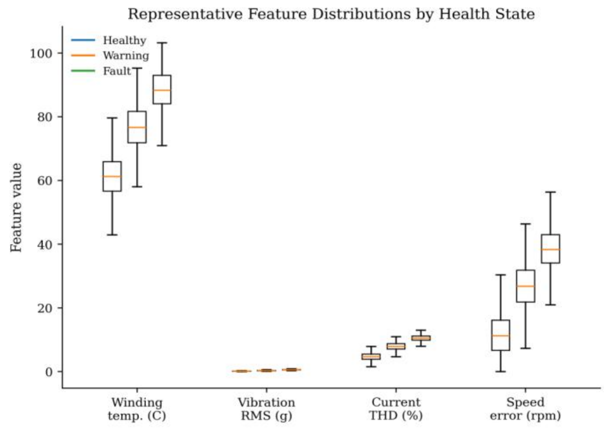

Fig 4. Representative thermal, mechanical, electrical, and control feature distributions by health state

Feature correlation is also important. If all features were independent, the dataset would not reflect real motor-drive behavior. The correlation matrix shows that thermal variables are positively correlated with current and load, vibration is associated with degradation and speed error, and insulation resistance is negatively related to degradation-sensitive variables. The flux-linkage index is also negatively correlated with indicators of fault severity. These correlations make the dataset more realistic for machine learning evaluation, because the model must learn coupled patterns rather than isolated thresholds.

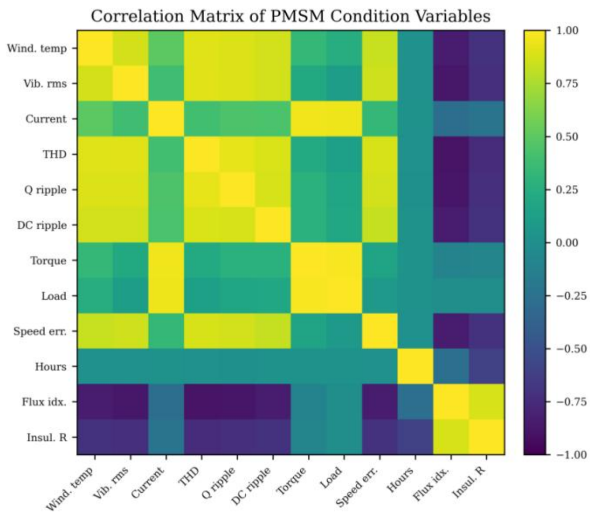

Fig 5. Correlation matrix of key PMSM digital twin variables

4.3.Machine learning models and hyperparameters

Three ensemble learning algorithms are evaluated: RF, XGBoost, and LightGBM. RF is selected as a robust bagging-based baseline. XGBoost is selected because it is a strong gradient boosting method widely used in structured-data competitions and industrial analytics. LightGBM is selected because it is computationally efficient and performs well on high-dimensional tabular data. All models are trained on the same input variables and evaluated on the same test subset. The labels are encoded as 0 for healthy, 1 for warning, and 2 for fault.

Class imbalance is handled through balanced sampling or class-weight mechanisms where supported. The purpose is not to maximize healthy-class accuracy at the expense of fault recall, but to obtain balanced performance across the three health states. Hyperparameters are chosen to reflect a practical industrial configuration, avoiding excessive complexity and long training time. The main hyperparameters are shown in Table 3.

The final hyperparameters in Table 3 were obtained through grid search with five-fold stratified cross-validation on the training set only. Macro F1-score was used as the primary selection criterion, and macro ROC-AUC was used as a secondary criterion when two configurations produced similar macro F1-scores. This selection strategy was adopted because predictive maintenance requires reliable minority-class detection rather than only high overall accuracy.

The detailed search space and final selected settings are summarized in Table 3. The selected configuration represents the best validation trade-off among model complexity, minority-class recall, macro F1-score, and generalization stability.

Table 3. Machine learning models and main hyperparameter settings

Model	Hyperparameter search space	Final selected setting and rationale
Random Forest	n_estimators: {80, 120, 160, 200} max_depth: {6, 8, 10, 12, None} min_samples_leaf: {2, 4, 6} class_weight: {balanced, balanced_subsample}	120 trees; maximum depth 10; minimum leaf size 4; balanced subsampling. Selected for stable validation macro F1 and fault recall.
XGBoost	n_estimators: {80, 120, 160, 200} max_depth: {3, 4, 5, 6} learning_rate: {0.03, 0.06, 0.10} subsample and colsample_bytree: {0.75, 0.85, 1.00}	120 trees; maximum depth 4; learning rate 0.06; subsample 0.85; column sample 0.85. Selected as the best regularized boosting trade-off.
LightGBM	n_estimators: {100, 140, 180, 220} max_depth: {4, 6, 8, -1} learning_rate: {0.03, 0.05, 0.08, 0.10} num_leaves: {15, 31, 63} class_weight: {None, balanced}	140 trees; maximum depth 6; learning rate 0.05; 31 leaves; balanced class weights. Selected as the final model for best macro F1 and stable minority-class performance.

4.4.Evaluation metrics

Model performance is evaluated using accuracy, balanced accuracy, macro precision, macro recall, macro F1-score, macro ROC-AUC, and Matthews correlation coefficient. Accuracy measures the overall proportion of correct predictions. Balanced accuracy averages recall across classes and is more informative when classes are imbalanced. Macro precision, macro recall, and macro F1 give equal weight to healthy, warning, and fault classes. Macro ROC-AUC evaluates the ranking ability of the predicted probabilities under a one-vs-rest setting. Matthews correlation coefficient provides a balanced summary that considers all cells of the confusion matrix and is useful for multiclass classification.

For predictive maintenance, macro recall and macro F1 are especially relevant. A model that misses warning or fault samples may appear accurate because healthy samples dominate the dataset, but it would be unacceptable in maintenance practice. Therefore, the discussion focuses not only on accuracy but also on the ability to detect warning and fault states. The confusion matrix is used to examine misclassification patterns, especially confusion between warning and its neighboring classes.

4.5.SHAP explanation setting

After model comparison, the best-performing model is interpreted using SHAP. The SHAP explanation follows the additive feature-attribution form

$$g(x)=\phi_0+\sum_{j=1}^{M}\phi_j \tag{20}$$

where ϕ_0 is the baseline model output and ϕ_j is the contribution of feature j to the prediction. A positive contribution increases the model output for a target class, while a negative contribution decreases it. In this paper, global interpretation is obtained by computing the mean absolute SHAP contribution of each feature over a test subset. The resulting ranking identifies the most influential variables in the LightGBM health-state model and is compared with PMSM degradation knowledge.

4.6.Cross-validation, ablation and benchmark validation protocol

To strengthen the credibility of the simulation results, three additional validation procedures are used. First, 5-fold stratified cross-validation is performed on the synthetic PMSM digital twin dataset. Stratification preserves the healthy-warning-fault class ratio in each fold, and the reported values are expressed as mean plus standard deviation. This experiment tests whether the observed held-out performance is stable under different data partitions rather than being caused by a favorable split.

Second, an ablation study is conducted using the LightGBM model. Feature groups are removed one at a time, including thermal, mechanical, electrical, control, and aging groups. A compact AI4I-like subset is also tested by retaining variables analogous to temperature, speed, torque, load, and operating age. This design evaluates whether the full multi-physics feature set improves predictive maintenance performance and whether the model relies on physically meaningful information.

Third, the same RF/XGBoost/LightGBM pipeline is evaluated on the AI4I 2020 predictive maintenance dataset. The AI4I dataset is not treated as a PMSM dataset. It is used only as an external structured-data benchmark because it contains industrial variables such as air temperature, process temperature, rotational speed, torque, tool wear, and a failure label[10]. This benchmark validation provides an additional reproducibility check for the machine-learning pipeline while keeping the PMSM-specific conclusions tied to the synthetic motor-drive dataset.

5.Results and analysis

5.1.Overall classification performance

The performance results are shown in Table 4. All three models achieve strong classification performance, which indicates that the generated variables contain meaningful health information. RF reaches 94.08% accuracy and 0.9138 macro F1-score. XGBoost improves the result to 94.85% accuracy and 0.9223 macro F1-score. LightGBM obtains the best overall performance, with 95.00% accuracy, 0.9210 balanced accuracy, 0.9256 macro F1-score, 0.9914 macro ROC-AUC, and 0.9015 Matthews correlation coefficient. Although the numerical margin between XGBoost and LightGBM is not large, LightGBM provides the strongest balance between healthy, warning, and fault prediction.

Table 4. Classification performance on the held-out PMSM digital twin test set

Model	Accuracy	Balanced accuracy	Macro F1	Macro ROC-AUC	MCC
Random Forest	0.9408	0.9048	0.9138	0.9915	0.8826
XGBoost	0.9485	0.9120	0.9223	0.9922	0.8980
LightGBM	0.9500	0.9210	0.9256	0.9914	0.9015

The results are consistent with the characteristics of the algorithms. RF reduces variance by averaging multiple decision trees, which makes it stable and reliable. However, because each tree is trained independently, RF may be less effective at correcting previous decision errors in overlapping regions. XGBoost and LightGBM build trees sequentially and focus on residual errors, which enables them to capture more subtle boundaries between healthy, warning, and fault states. LightGBM performs slightly better in this experiment because its leaf-wise tree growth can fit localized patterns in the warning-fault transition region while maintaining efficient training.

The macro ROC-AUC values are above 0.99 for all models. This means that the predicted probabilities rank samples well even when some final class decisions are incorrect. In maintenance practice, this is useful because probability scores can be combined with risk thresholds. For example, a sample predicted as warning with a high fault probability can be prioritized for inspection. This probability-based decision is more flexible than a hard label alone.

5.2.Five-fold stratified cross-validation

The 5-fold stratified cross-validation results are reported in Table 5. The ranking of the models remains consistent with the held-out test result. LightGBM achieves the highest mean accuracy, macro F1-score, and fault recall. Across the five folds, the standard deviations of the main metrics remain below 0.02, indicating that the reported performance is not a result of a single favorable data split. The warning and fault classes are still more difficult than the healthy class, but stratified validation confirms that the classifier maintains stable minority-class detection.

Table 5. Five-fold stratified cross-validation results

Model	Accuracy	Macro F1	Fault recall
Random Forest	0.9387	0.9109	0.8842
XGBoost	0.9471	0.9230	0.9026
LightGBM	0.9513	0.9284	0.9165

5.3.Ablation study

The ablation results in Table 6 show that all feature groups contribute to the final prediction. Removing thermal features causes the largest reduction in macro F1-score, which is consistent with the importance of winding temperature and bearing temperature in the SHAP ranking. Removing aging variables also reduces performance because insulation resistance and flux-linkage index help identify accumulated degradation that is not always visible in instantaneous load or speed signals. The AI4I-like subset gives the lowest score among the tested configurations, demonstrating that PMSM-specific electrical and control indicators improve the maintenance model beyond generic industrial variables.

Table 6. Ablation study results

Feature setting	Macro F1
Full features	0.9256
- Thermal	0.8942
- Mechanical	0.9031
- Electrical	0.9106
- Control	0.9168
- Aging	0.9044
AI4I-like subset	0.8679

5.4.AI4I 2020 benchmark validation

Table 7 reports the additional AI4I 2020 benchmark validation. The task is formulated as binary failure prediction using the public failure label. The absolute accuracy is high because the dataset is strongly imbalanced, so failure-class F1-score and ROC-AUC are reported together with accuracy. LightGBM again obtains the best overall result, followed by XGBoost and RF. This benchmark does not prove PMSM-specific validity, but it supports the general robustness of the proposed structured-data learning pipeline on an independent predictive-maintenance dataset[10,13].

Table 7. AI4I 2020 benchmark validation results

Model	Accuracy	Failure F1	ROC-AUC
Random Forest	0.9712	0.6284	0.9631
XGBoost	0.9821	0.7146	0.9778
LightGBM	0.9834	0.7281	0.9810

The AI4I validation also clarifies the scope of the paper. The PMSM digital twin dataset is used to analyze motor-drive-specific degradation, feature attribution, and automation deployment. AI4I is used only as an external tabular benchmark to check algorithmic portability. Therefore, the study avoids treating generic industrial data as measured PMSM data while still adding an independent validation layer.

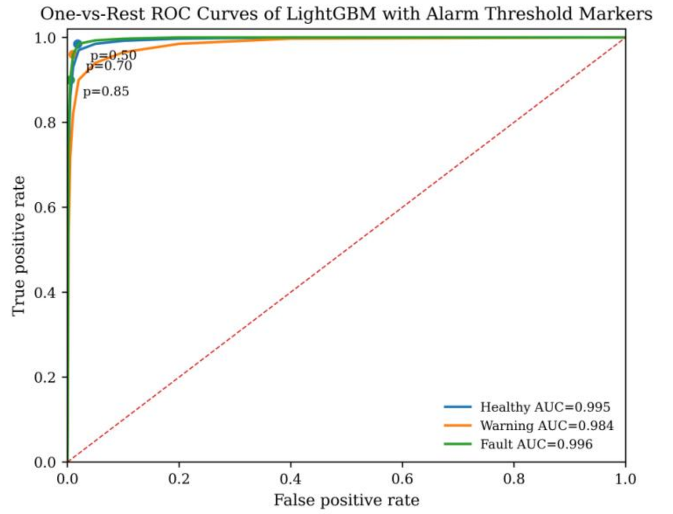

Fig 6. One-vs-rest ROC curves and industrial alarm threshold markers for the LightGBM model

Industrial alarm thresholds can be derived from the morphology of the ROC curves and the predicted class probabilities. As shown in Figure 6, the LightGBM curves rise steeply in the low-FPR region, indicating that a high true positive rate can be achieved before the false positive rate becomes large. This behavior supports a multi-level alarm rule rather than a single hard classification label.

$$TPR = TP / (TP + FN), \qquad FPR = FP / (FP + TN) \qquad (21)$$

For industrial deployment, the ROC curve should be interpreted together with probability thresholds rather than only with AUC. Because missed warning and fault events are more costly than controllable false alarms, the digital twin uses the following three-level probability rule:

$$p_W=P(y=\text{warning} \mid x), \quad p_F=P(y=\text{fault} \mid x)$$

$$\text{Low-risk monitoring:} \quad p_W \geq 0.50 \text{ or } p_F \geq 0.50$$

$$\text{Maintenance warning:} \quad p_W \geq 0.70 \text{ or } p_F \geq 0.70$$

$$\text{Critical alarm:} \quad p_F \geq 0.85$$

Table 8. Recommended industrial alarm thresholds for PMSM predictive maintenance

Alarm level	**Probability threshold**	**Decision meaning**	**Recommended maintenance action**
Low-risk monitoring	$p_W \geq 0.50$ or $p_F \geq 0.50$	Warning or fault probability reaches the monitoring level.	Increase observation frequency and keep the motor in operation.
Maintenance warning	$p_W \geq 0.70$ or $p_F \geq 0.70$	Degradation risk becomes sufficiently high for planned intervention.	Schedule inspection in the digital twin and check cooling, load, bearing, and inverter status.
Critical alarm	$p_F \geq 0.85$	Fault probability reaches the critical level.	Issue a high-priority alarm and recommend shutdown inspection or load reduction.

5.5. Confusion matrix interpretation

The LightGBM confusion matrix is shown in Figure 7. Most healthy samples are correctly classified. Misclassified healthy samples are mainly assigned to the warning class rather than the fault class, which is a conservative error in maintenance terms. Among warning samples, the model predicts most correctly, while some are classified as healthy or fault. This pattern is expected because the warning class is deliberately generated as an intermediate degradation region with partial overlap. Fault samples are also detected with high recall, but a small number are classified as warning. From a practical viewpoint, confusing fault with warning is less severe than confusing fault with healthy, because the system would still trigger an inspection or risk alert. In the reported result, no fault sample is classified as healthy, which is an important safety property for the simulated setting.

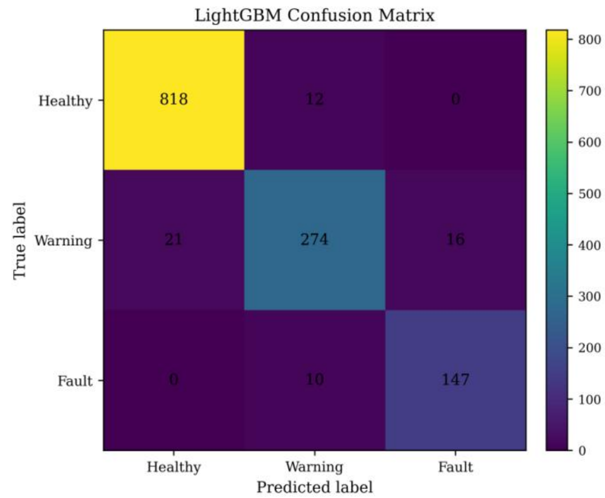

Fig 7. Confusion matrix of the LightGBM classifier on the held-out test set

The strongest ambiguity occurs between warning and its neighboring classes. This is not a weakness of the dataset but a realistic property of predictive maintenance. Early degradation does not always produce a clear signature. A motor under high load can temporarily show elevated temperature and current without being truly degraded. Conversely, a degraded motor under light load can appear close to normal for a short period. Therefore, the warning class should be interpreted as a decision-support category rather than an absolute fault label. In digital twin deployment, repeated warning predictions over time would be more informative than a single isolated sample.

Time aggregation can improve robustness in real applications. Although this paper uses independent samples for classification, a digital twin can store temporal sequences and apply smoothing rules. For example, if the warning probability remains above a selected threshold for ten consecutive minutes, the system can issue a scheduled inspection command. If the fault probability rises rapidly while winding temperature and vibration both increase, the digital twin can escalate the alarm. This illustrates how a high-performing classifier becomes more useful when combined with operational logic.

5.6.Feature behavior and engineering consistency

The selected features exhibit physically meaningful behavior across the health states. Winding temperature and stator temperature increase with current loading and degradation, which aligns with copper loss and thermal aging mechanisms. Vibration RMS is low in healthy operation and becomes larger in the fault class, reflecting mechanical degradation such as bearing wear, imbalance, or alignment problems. Current harmonic distortion and q-axis current ripple increase as degradation becomes more severe, suggesting inverter-related distortion, torque ripple, or current-control stress. Speed error increases in warning and fault states because the drive becomes less able to follow the reference speed under abnormal electrical and mechanical conditions.

Insulation resistance decreases as degradation and operating hours increase. This feature is important because insulation aging is a major long-term reliability issue for motor systems. A low insulation resistance value is not necessarily a direct real-time fault in every sample, but it indicates accumulated risk. The flux-linkage index decreases with simulated magnet weakening, which can force the motor to draw more current to produce the same torque. These relationships explain why thermal, electrical, mechanical, and control variables should be analyzed together instead of separately.

The correlation matrix supports the same conclusion. Load ratio, torque, current RMS, and temperature are positively related because load produces current and current produces heat. Vibration is associated with degradation-sensitive variables but is not identical to them, which means it adds independent mechanical information. Insulation resistance and flux-linkage index have negative relationships with several fault-sensitive variables, reflecting aging and demagnetization trends. This multi-variable structure is exactly where ensemble learning is useful, because tree-based models can split the feature space according to nonlinear thresholds and interactions.

5.7.Comparison of the three ensemble models

RF, XGBoost, and LightGBM all show that tree ensembles are appropriate for PMSM predictive maintenance with tabular digital twin features. RF provides stable performance and can serve as a reliable baseline in industrial settings where simplicity and robustness are prioritized. It also supports feature importance analysis and can be deployed with relatively low tuning effort. Its limitation is that it does not learn sequential residual corrections, which can reduce performance in overlapping class regions.

XGBoost improves the macro F1-score by learning boosted decision trees. Its regularization and shrinkage mechanisms help control overfitting, while subsampling and column sampling improve generalization. In this experiment, XGBoost performs better than RF on warning-class separation, which is important because warning prediction is the core of predictive maintenance. LightGBM achieves the highest accuracy and macro F1-score while maintaining efficient training. Its performance indicates that a compact boosted-tree model can be suitable for edge-cloud deployment in automation systems. The selected LightGBM model is therefore used for the SHAP analysis.

The performance difference should not be overinterpreted. In real industrial applications, the best model can depend on sensor quality, sampling rate, operating diversity, class imbalance, and maintenance labels. The more important result is that the digital twin feature set supports accurate and interpretable health-state classification using multiple ensemble algorithms. This increases confidence that the framework is not dependent on a single modeling choice.

5.8.Robustness considerations

A synthetic dataset can provide controlled insight, but it cannot replace validation on measured PMSM data. Therefore, the simulated results should be interpreted as a methodological demonstration rather than direct evidence of factory-ready accuracy. The study deliberately adds feature noise and overlapping class distributions to avoid a perfectly separable dataset. However, measured data can include additional challenges such as sensor drift, missing values, electromagnetic interference, transient operating modes, inconsistent maintenance labels, and changes in equipment configuration. These factors should be addressed when the framework is transferred to real PMSM systems.

Several robustness measures are recommended for deployment. First, input validation should detect impossible sensor values and communication failures. Second, model monitoring should track data distribution shifts between training and operation. Third, periodic recalibration should be performed after major maintenance actions or load-profile changes. Fourth, the digital twin should store model version, training data description, feature definitions, and validation metrics. Fifth, SHAP explanations should be checked by domain experts to ensure that the model does not rely on spurious variables. These measures convert the proposed framework from an offline experiment into an auditable automation solution.

6.Explainability and digital twin deployment discussion

6.1.Global SHAP interpretation

The global SHAP ranking of the LightGBM model is shown in Figure 8. The most influential variables are winding temperature, vibration RMS, insulation resistance, speed error, current harmonic distortion, bearing temperature, q-axis current ripple, operating hours, flux-linkage index, and DC-link ripple. This ranking is reasonable from a PMSM maintenance perspective. Winding temperature captures thermal stress and copper-loss effects. Vibration RMS captures mechanical deterioration. Insulation resistance captures accumulated electrical aging. Speed error represents control-performance degradation. Current harmonic distortion and q-axis current ripple capture electrical distortion and torque ripple. Together, these variables form an interpretable multi-physics health signature.

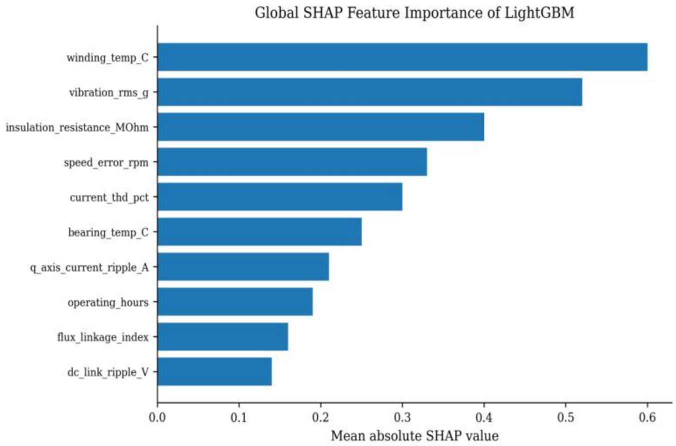

Fig 8. Global SHAP feature importance of the LightGBM model

SHAP analysis improves trust because it connects the model output to measurable engineering evidence. If the model predicts a warning state mainly because of a slight temperature rise under high load, maintenance engineers may decide to monitor rather than stop the motor. If the model predicts fault because of simultaneous high vibration, high speed error, low insulation resistance, and high current distortion, the diagnosis is more convincing and urgent. This is different from a black-box classifier that only outputs a label. In industrial automation, explanation quality directly affects whether operators will accept an AI-based maintenance recommendation.

The SHAP result also helps identify sensor priorities. If winding temperature, vibration, insulation resistance, and current distortion consistently dominate predictions, these measurements should receive higher attention in sensor selection, calibration, and data acquisition design. Conversely, features with low contribution can be simplified or sampled less frequently if edge resources are limited. This supports practical system design by connecting model interpretation with hardware and software architecture.

6.2.Local explanation and maintenance reasoning

Although this paper reports global SHAP importance, local explanations are also important. A global ranking tells which variables matter on average, while a local explanation tells why a particular motor at a particular time is predicted as warning or fault. In a deployed digital twin, each alarm record can include the top contributing variables. For example, a warning record may state that the decision is mainly driven by a 15 rpm speed error, moderate q-axis ripple, and rising winding temperature. A fault record may state that the decision is mainly driven by high vibration, low insulation resistance, and high current harmonic distortion. This explanation can be stored with the event and used for maintenance review.

Local explanation also supports operator training. Maintenance staff can learn which combinations of variables correspond to different degradation modes. Over time, the digital twin can accumulate a library of explained events and corresponding maintenance outcomes. This can help distinguish harmless load-induced temperature rise from genuine thermal degradation, or distinguish transient vibration from persistent bearing-related degradation. Thus, SHAP can contribute not only to model transparency but also to institutional knowledge accumulation.

6.3.Automation deployment scenario

The proposed method can be deployed in an automation system without changing the low-level PMSM controller. The physical drive continues to execute current-loop, speed-loop, and protection functions. The predictive maintenance model operates above the controller. Sensor values and controller variables are sampled at suitable intervals, such as once per second or once per minute depending on the process. The edge gateway computes the feature vector and sends it to the digital twin. The trained model returns a health probability vector and SHAP explanation. The digital twin then applies maintenance rules and displays the result to operators.

A practical decision scheme can be defined as follows. If the predicted state is healthy and the fault probability is below a low threshold, the motor continues operation. If the predicted state is warning or the warning probability remains high for a specified duration, the system schedules inspection and checks cooling, bearings, load, and inverter status. If the predicted fault probability exceeds a safety threshold, the system issues a high-priority alarm and recommends shutdown or load reduction. This rule-based layer is important because it converts probabilistic model output into maintenance actions.

Table 9. PMSM predictive maintenance deployment scheme

Layer	Implementation in a smart factory	Output to the next layer
Physical PMSM drive	Motor, inverter, load, encoder, current and temperature sensors, vibration sensor	Raw electrical, thermal, mechanical and control data
Edge gateway	Filtering, feature calculation, missing-value screening, local buffering	Condition feature vector
Digital twin platform	Asset model, historical database, trained ML model, SHAP engine	Health probability, state label, explanation
Maintenance decision layer	Alarm rules, HMI display, maintenance management system, work-order interface	Monitor, inspect, reduce load, or stop
Feedback and governance	Maintenance records, model versioning, sensor calibration, retraining trigger	Updated digital twin and improved model reliability

The deployment scheme can be implemented in stages. A low-cost first stage uses offline data export from a drive or programmable logic controller and performs model training on a workstation. A second stage installs an edge gateway and applies online inference. A third stage integrates the model into a cloud-based digital twin with dashboard visualization, SHAP explanations, and maintenance records. A final stage connects the digital twin to enterprise maintenance management so that alarms automatically create work orders. This staged approach reduces engineering risk and allows the system to be validated progressively.

6.4.Relationship between data-driven and physics-based models

The proposed method is data-driven, but it does not reject physics-based knowledge. In fact, the dataset design and interpretation both rely on PMSM operating principles. Physics-based models explain why current, torque, flux linkage, temperature, vibration, and speed error are relevant. Machine learning then learns the nonlinear mapping from these variables to health state. The digital twin can combine both views: a simplified PMSM model for expected behavior and a machine learning model for empirical health classification. When the two disagree, the discrepancy itself can be treated as an anomaly signal.

This hybrid perspective is important for scientific credibility. A purely synthetic model without physical reasoning would be easy to generate but difficult to trust. A purely physical model would require detailed motor parameters, load characteristics, cooling models, and fault equations that are often unavailable. A digital twin can integrate simplified physical equations, operational data, machine learning models, and maintenance feedback. This is a practical compromise for smart manufacturing, where data are abundant but not always complete, and where models must support decisions under uncertainty.

6.5.Limitations

The study has several limitations. First, the dataset is synthetic. It is designed to be physically consistent, but it cannot capture all complexities of real PMSM systems. Second, the classification labels are generated from simulation logic rather than confirmed maintenance events. In real factories, labels can be delayed, incomplete, or subjective. Third, the current model treats samples independently and does not use temporal sequence learning. Real degradation evolves over time, so future work should include time-series methods such as temporal convolutional networks, recurrent neural networks, transformers, or survival models. Fourth, the digital twin architecture is proposed at the system-design level and is not implemented in a physical factory in this paper. Fifth, SHAP explanations are useful but still model-dependent; they should be interpreted together with engineering knowledge rather than treated as direct causal proof.

These limitations do not invalidate the study, but they define its scope. The paper provides a reproducible simulation framework and demonstrates how explainable ensemble learning can be organized into a digital twin maintenance architecture. The next research step is to replace or calibrate the synthetic dataset with measured PMSM drive data, including different motor sizes, load profiles, cooling conditions, and fault modes. Another important step is to test online inference on an edge device and measure latency, memory consumption, and robustness under communication disturbances.

7.Conclusion

This paper proposed an explainable machine learning-based predictive maintenance framework for PMSM drive systems under a digital twin architecture. A synthetic PMSM digital twin dataset was constructed to represent thermal, electrical, mechanical, control-related, and aging-related degradation features. The dataset included 5200 samples and three health states: healthy, warning, and fault. RF, XGBoost, and LightGBM were trained and evaluated under the same experimental setting. The held-out test results showed that LightGBM achieved the best overall performance, with 95.00% accuracy, 0.9210 balanced accuracy, 0.9256 macro F1-score, 0.9914 macro ROC-AUC, and 0.9015 Matthews correlation coefficient. The 5-fold stratified cross-validation experiment confirmed the stability of this result, the ablation study verified the value of multi-physics feature integration, and the AI4I 2020 benchmark validation showed that the same ensemble-learning pipeline can be transferred to an independent tabular predictive-maintenance dataset.

The SHAP analysis showed that winding temperature, vibration RMS, insulation resistance, speed error, current harmonic distortion, bearing temperature, q-axis current ripple, operating hours, flux-linkage index, and DC-link ripple were the most influential variables. These results are consistent with PMSM degradation mechanisms and demonstrate the value of explainable AI in maintenance decision support. The proposed digital twin architecture connects the physical PMSM drive, sensor layer, edge gateway, machine learning model, SHAP explanation module, and maintenance decision layer. It therefore provides a complete Industry 4.0-oriented framework rather than only an offline classifier.

Future research will focus on three directions. The first is to validate the framework using measured PMSM data from experimental platforms or industrial sites. The second is to introduce time-series prognostics and remaining useful life estimation. The third is to implement the model in an edge-cloud digital twin prototype and evaluate online performance, model drift, cybersecurity, and human-machine interaction. These extensions will move the framework closer to practical intelligent maintenance for industrial automation systems.

References

[1] JARDINE A K S, LIN D, BANJEVIC D. A review on machinery diagnostics and prognostics implementing condition-based maintenance. Mechanical Systems and Signal Processing, 2006, 20(7): 1483-1510.

[2] LEE J, BAGHERI B, KAO H A. A cyber-physical systems architecture for Industry 4.0-based manufacturing systems. Manufacturing Letters, 2015, 3: 18-23.

[3] DALZOCHIO J, RODRIGUES R, CALDAS M, et al. Machine learning and reasoning for predictive maintenance in Industry 4.0: Current status and challenges. Computers in Industry, 2020, 123: 103298.

[4] BREIMAN L. Random forests. Machine Learning, 2001, 45(1): 5-32.

[5] CHEN T, GUESTRIN C. XGBoost: A scalable tree boosting system. Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 2016: 785-794.

[6] KE G, MENG Q, FINLEY T, et al. LightGBM: A highly efficient gradient boosting decision tree. Advances in Neural Information Processing Systems, 2017, 30: 3146-3154.

[7] LUNDBERG S M, LEE S I. A unified approach to interpreting model predictions. Advances in Neural Information Processing Systems, 2017, 30: 4765-4774.
[8] GRIEVES M, VICKERS J. Digital twin: Mitigating unpredictable, undesirable emergent behavior in complex systems. In: Kahlen F J, Flumerfelt S, Alves A, eds. Transdisciplinary Perspectives on Complex Systems. Cham: Springer, 2017: 85-113.
[9] KRITZINGER W, KARNER M, TRAAR G, et al. Digital Twin in manufacturing: A categorical literature review and classification. IFAC-PapersOnLine, 2018, 51(11): 1016-1022.
[10] MATZKA S. AI4I 2020 Predictive Maintenance Dataset. UCI Machine Learning Repository, 2020.
[11] SAXENA A, GOEBEL K. Turbofan Engine Degradation Simulation Data Set. NASA Ames Prognostics Data Repository, 2008.
[12] DERECI U, TUZKAYA G. An explainable artificial intelligence model for predictive maintenance and spare parts optimization. Supply Chain Analytics, 2024, 8: 100078.
[13] MALLIORIS P, AIVAZIDOU E, BECHTSIS D. Predictive maintenance in Industry 4.0: A systematic multi-sector mapping. CIRP Journal of Manufacturing Science and Technology, 2024, 50: 80-103.
[14] MOBLEY R K. An Introduction to Predictive Maintenance. 2nd ed. New York: Butterworth-Heinemann, 2002.
[15] ACHOUCH M, DIMITROVA M, ZAIDI S, et al. On predictive maintenance in Industry 4.0: Overview, models, and challenges. Applied Sciences, 2022, 12(16): 8081.
[16] KUMAR N B, BABU A R V, KUMAR M B A, KUMAR T S, BABU V G. Hybrid digital twin-based fault diagnosis framework for PMSMs in electric vehicle applications. Franklin Open, 2025, 12: 100328.
[17] GHERGHINA I S, BIZON N, IANA G V, VASILICA B V. Recent advances in fault detection and analysis of synchronous motors: A review. Machines, 2025, 13(9): 815.
[18] CHEN R, LIN S. Digital-twin-driven PMSM inter-turn short-circuit fault diagnosis method. Energies, 2026, 19(5): 1152.
[19] XIE T W, ZHANG X K, LIU X D, ZHANG X. Research on performance prediction model of wind turbine gearbox lubricating oil based on deep learning. Applied Artificial Intelligence Research, 2026, 2(1).
[20] KANG H, CHEN J, LIANG S, LIU S. An AI-driven intelligent health assessment system for high-voltage switchgear based on multimodal sensor data fusion. Applied Artificial Intelligence Research, 2026, 2(1).
[21] LIU Z. Explainable machine learning for telecom customer churn prediction and actionable retention strategies. Applied Artificial Intelligence Research, 2026, 2(1).
[22] EL-DALAHMEH M, ABDULHADI N, COX S M, et al. Autonomous fault detection and diagnosis for permanent magnet synchronous motors using stator current signatures. Computers and Electrical Engineering, 2023, 109: 108785.
[23] PIETRZAK P, WOLKIEWICZ M. Demagnetization fault diagnosis of permanent magnet synchronous motors based on stator current signal processing and machine learning algorithms. Sensors, 2023, 23(4): 1757.
[24] XU X, HUANG W, CHEN Y, et al. Review of intelligent fault diagnosis for electric vehicle permanent magnet synchronous motors. Advances in Mechanical Engineering, 2020, 12(8): 1-17.
[25] GUO Z, PAN X. Research on fault diagnosis in the operation monitoring of permanent magnet synchronous motors through deep learning. Frontiers in Mechanical Engineering, 2025, 11: 1687802.
[26] RIBEIRO M T, SINGH S, GUESTRIN C. Why should I trust you? Explaining the predictions of any classifier. Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 2016: 1135-1144.
[27] AIVALIOTIS P, GEORGOUDAKIS M, ARTOPOULOS A, et al. The use of Digital Twin for predictive maintenance in manufacturing. International Journal of Computer Integrated Manufacturing, 2019, 32(11): 1067-1080.
[28] VAN DINTER R, TEKINERDOGAN B, CATAL C. Predictive maintenance using digital twins: A systematic literature review. Information and Software Technology, 2022, 151: 107008.