A Novel Data-Driven Method for Ball Bearing Impedance Modelling

Weiqian Li[1], *Student Member, IEEE*, Daniele De Gaetano[2], *Member, IEEE*, Wenjun Zhu[1], *Graduate Student Member, IEEE*, Xiao Chen[1], *Senior Member, IEEE*, and Xiangyu Sun[1], *Student Member*

***Abstract*—Nowadays, electrical machines are commonly driven by voltage source inverters in which the intrinsic switching phenomenon can result in common mode voltage which can further lead to high frequency bearing currents. To accurately model the bearing impedance and in turn bearing currents, this paper proposes a probabilistic framework for bearing impedance modeling in inverter-driven applications. The impedance distribution is decomposed using the chain rule into a phase model and an amplitude model. A multi-layer perceptron (MLP)-based network is employed to predict the impedance phase distribution under given conditions and the corresponding amplitudes for different phases. This approach effectively captures both the transition from capacitive to resistive states, through phase behavior, and the associated amplitude responses, making it applicable across a wide range of shaft speeds, voltage amplitudes, and excitation frequencies. This modular approach aligns well with the physical processes, underlying the bearing breakdown. Additionally, it could be readily extended to incorporate additional parameters such as temperature and lubrication condition.**

***Index Terms*— Bearing currents, impedance analysis, dielectric breakdown, multi-layer perceptron (MLP), conditional probabilistic modeling.**

I. Introduction

THE rapid evolution of inverter-driven electrical machines has brought renewed focus to the intricate interplay between high-frequency common-mode voltages and the dynamic electrical behavior of ball bearings. With the advent of wide bandgap (WBG) devices and faster switching speeds, the phenomena that lead to bearing degradation have become more pronounced [1] [2] [3]. Inverter switching generates a wide spectrum of voltage disturbances that can induce parasitic currents in bearings, accelerating both electrical breakdown and mechanical wear [4] [5]. Traditional lumped parameter models that treat bearing impedance as a fixed parameter are no longer sufficient for capturing the complex, variable nature of these interactions.

Experimental studies [6] [7] [8] have shown that bearing impedance is inherently stochastic: its phase and amplitude vary significantly with lubrication state, preload, temperature, shaft speed, and excitation frequency. Phase jumps typically indicate lubrication-film breakdown and the onset of resistive current paths, while amplitude distributions often exhibit multimodal peaks corresponding to different conduction regimes [1] [9]. To better reflect these dynamics, improved equivalent circuit models have introduced variable breakdown resistances and frequency-dependent distributed elements [10] [11]. Finite element analysis (FEA) has been used to calculate parasitic capacitances, inductances, and their coupling within the bearing structure [12] [13].

Despite these advances, existing models remain fragmented and largely deterministic. For example, reference [14] conducted a broad survey of inverter-induced bearing currents and highlighted the need for a unified probabilistic framework to capture variability across operating conditions. Building on this, reference [15] refined the equivalent-circuit approach by introducing a variable breakdown resistance term to model the transition from capacitive (non-breakdown) to resistive (breakdown) behavior. Earlier, reference [7] proposed a detailed electrical model for rolling-element bearings that incorporates dielectric constants, specific resistances, and lubrication-film effects. Although each addresses aspects of impedance variation or mitigation, no existing model provides a fully probabilistic description spanning both non-breakdown and breakdown regimes.

Accurate characterization of bearing impedance variation under different operating conditions is crucial for developing a model capable of predicting bearing currents [16] [17] [18]. In particular, the impedance amplitudes in the breakdown state are often difficult to predict. This paper proposes a machine learning-based probabilistic modeling framework to capture the nonlinear variations of bearing impedance under different shaft speeds, applied voltage amplitudes, and frequencies (up to 1 MHz), covering the higher order harmonic components of common-mode voltage (CMV) for PWM inverters. The main novelty of this framework lies in its ability to probabilistically describe bearing behavior across a substantial range of operating conditions, including both non-breakdown and breakdown regimes. This capability is achieved through a two-

Weiqian Li is with the University of Sheffield, Sheffield, S10 4DE, UK (e-mail: wli59@ sheffield.ac.uk).

Daniele De Gaetano is with the University of Sheffield Advanced Manufacturing Research Centre (AMRC), S60 5TZ, UK (e-mail: d.degaetano@amrc.co.uk).

Wenjun Zhu is with the University of Sheffield, Sheffield, S10 4DE, UK (e-mail: wzhu24@ sheffield.ac.uk).

Xiao Chen is with the University of Sheffield, Sheffield, S10 4DE, UK (e-mail: xiao.chen@sheffield.ac.uk).

Xiangyu Sun is with the University of Sheffield, Sheffield, S10 4DE, UK (e-mail: xiangyu.sun@sheffield.ac.uk).

stage approach: First, the ball bearing impedance under various operating conditions is modeled as a joint conditional distribution. Then, following the chain rule of probability, it is factorized into two conditional distributions: the one captures how the impedance phase changes across different operating conditions, and the other one describes the impedance amplitude based on both the operating conditions and the corresponding phase states. As a result, this work presents several significant advancements compared to previous studies:

1) A probabilistic phase model using a bin-based neural network and kernel density estimation (KDE) smoothing to accurately represent complex, potentially multimodal distributions.
2) Assuming that the conditional amplitude model follows Gaussian distribution, a neural network is used to predict its parameters, thereby simplifying the modeling.
3) The chain-rule factorization confers a strong physical interpretation: a transition in phase often signals the onset of electric breakdown, and the corresponding amplitude distribution highlights the resistive or capacitive nature of the bearing at that operating point.
4) The framework uses conditional probability models which means it can be easily extended to other factors such as temperature or lubrication states.

This study focuses on the capacitive-to-resistive transition in bearing impedance, which is driven by high-frequency dielectric breakdown phenomena. The proposed probabilistic framework provides a quantitative characterization of the impedance behavior across breakdown and non-breakdown states, enabling improved understanding of the electrical processes underlying such transitions.

The paper is organized as follows. Section II details the development of the proposed probabilistic model and its underlying methodology. Section III and Section IV introduce the modeling of impedance phase and amplitude, respectively. Section V presents the experimental validation of the framework under various operating conditions. Finally, Section VI concludes the paper and outlines future research directions to further enhance bearing health monitoring and predictive maintenance strategies.

II. Impedance Probability Distribution Modelling

A. Problem Statement

Ball bearings typically have multiple rolling elements (balls) running between inner and outer raceways. As the rolling elements rotate, a thin lubrication film develops between the rolling elements and inner/ outer races, acting as a dielectric and thus isolating the balls from the raceways. Fig. 1 illustrates the schematic and electrical model of a ball bearing. In non-rotating conditions, the balls are in contact with the inner and outer raceways, allowing the bearing to behave almost as a short circuit. Under normal operating conditions, the bearing usually exhibits a predominantly capacitive response, often modeled as either a pure capacitor or a combination of capacitance and resistance [6] [15], where the impedance strongly depends on factors such as the lubricant relative permittivity, the contact area, and the film thickness. Once the voltage across the bearing exceeds a certain threshold, the dielectric experiences a breakdown, transitioning to a conductive state. This event typically triggers impulsive electrical discharge machining (EDM) currents that can significantly shorten the bearing lifespan [19] [20]. Based on these fundamental behaviors, the next section introduces a probabilistic modeling framework, designed to capture both the non-breakdown and breakdown characteristics of ball bearings under a wide range of operating conditions.

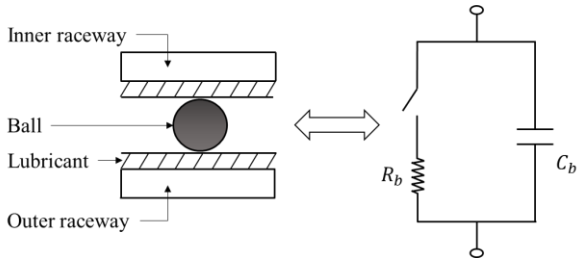


Fig. 1. Ball bearing schematic and its equivalent circuit model [6].

B. Probabilistic Model of Ball Bearing Impedance

The impedance of a ball bearing is affected by several operating parameters, including shaft speed, voltage amplitude, and excitation frequency. Due to inherent randomness, even under identical operating conditions, the impedance can exhibit varying behaviors [6] [7]. Consequently, it is challenging to describe impedance with a deterministic model based solely on these parameters. A more appropriate approach is to treat impedance as a conditional probability distribution, where the operating parameters define the probability of different impedance states.

This study aims to estimate the conditional probability distribution of bearing impedance, under various operating conditions characterized by rotational speed RPM, voltage amplitude V, and excitation frequency f. As illustrated in Fig. 2, the conditional probability distribution of impedance is represented as a collection of cloud points. Each dot in the figure corresponds to a specific impedance data point, where the horizontal axis represents the impedance amplitude, and the vertical axis denotes the phase angle. Those measured data are obtained by the work presented in [6]. Together, these dots form the distribution of impedance in the phase-amplitude plane. To better formulate the problem, the condition vector is denoted as $c = [RPM, V,\ f]$, x represents the measured phase angle and y represents the measured impedance amplitude. Then, the conditional probability distribution of impedance can be expressed as the joint distribution of phase and amplitude $p(x, y|c)$.

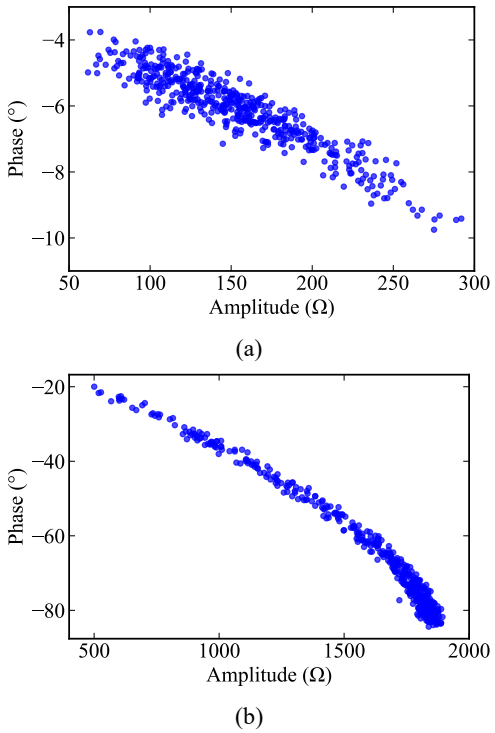


Fig. 2. Impedance distribution under different operating conditions. (a) Measured at 1400 r/min, 3 V, 1 MHz. (b) Measured at 1200 r/min, 2 V, 900 KHz.

Direct modeling of the impedance distribution $p(x, y|c)$ is difficult due to the complexity of capturing the joint dependencies between x and y as well as the variability introduced by c. To address this issue, an alternative approach is adopted, where the conditional probability distribution of the bearing impedance phase, $p(x|c)$, is modeled first. Subsequently, the amplitude distribution $p(y|x, c)$ is represented. By applying the probability chain rule, these two distributions are combined as follows:

$$p(x, y|c) = p(x|c) \cdot p(y|x, c) \quad (1)$$

One strength of the framework lies in its interpretability: the chain-rule factorization aligns well with physical intuition, since the bearing phase shift typically signals the onset of electric breakdown, and the impedance amplitude of the breakdown (purely resistive state) depends on the voltage amplitude and frequency. In addition, given that the phase angle often directly reflects the transition from capacitive to resistive behaviour in bearing electric breakdown phenomena, independent modelling $p(x|c)$ is crucial for monitoring and predicting breakdown events.

Rather than producing discrete health labels (e.g., “healthy”, “degraded”, “failed”), the proposed framework generates the complete conditional impedance distribution in the phase-amplitude plane for a given operating condition. This probabilistic representation can be regarded as a digital twin of the bearing’s electrical behavior, enabling downstream analysis such as classification into capacitive/transitional/resistive states using predefined phase thresholds, or direct computation of derived quantities like bearing current. In this way, the model can be applied to any particular case covered by the operating range of its training data.

III. Phase Distribution Modelling

This section focuses on a bin-based (discretized) formulation of conditional probability, a neural network training strategy, and kernel-based smoothing to reconstruct the predicted continuous distribution.

A. Discrete Distribution of Phase

To better understand the phase distribution characteristics of bearing impedance under various operating conditions, a histogram of the phase distribution under specified conditions can be plotted. This approach visually illustrates how the data are distributed across different intervals. Subsequently, employing the KDE method to draw a continuous probability distribution curve of the phase provides a more comprehensive depiction of the overall distribution and central tendency of the phase data within the selected range [21].

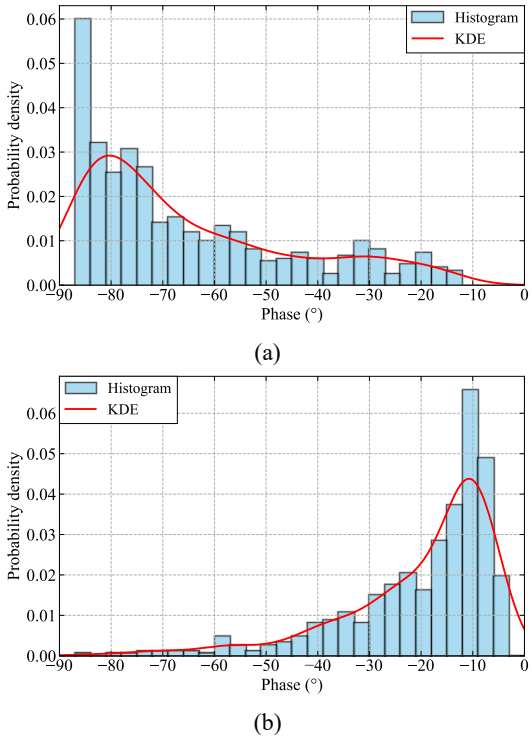


Fig. 3. Phase distribution under different operating conditions. (a) 800 r/min, 3 V, 500 KHz. (b) 600 r/min, 4 V, 300 kHz.

Fig. 3 depicts the phase distribution of bearing impedance under different operating conditions, and these distributions under given conditions typically exhibit multimodal or heavy-tailed behavior. It is challenging to adopt a single parametric distribution (e.g., uniform, gaussian, exponential) model to describe these distributions, making it difficult to directly fit or model the phase distribution in a purely parametric manner. Therefore, a binning approach is adopted to discretize the phase angle into a finite set of bins. Since both the condition variables c and the phase x can span different scales, each dimension is first standardized separately. This transformation is performed to ensure zero mean and unit variance for each dimension. Specifically, the standardization is applied as follows:

$$x' = \frac{x - \mu_x}{\sigma_x}, c_i' = \frac{c_i - \mu_{c_i}}{\sigma_{c_i}} \quad (2)$$

where μ_x and σ_x are the mean and standard deviation of x, while μ_{c_i} and σ_{c_i} are the mean and standard deviation of the i-th component of c. This standardization removes scale disparities among features (e.g., voltage amplitude in the range $1-5$ versus frequency in the range 10^5-10^6), accelerates convergence, and prevents any single variable from dominating the learning process. After standardization, the phase x' is normalized to lie within the range $[\alpha_{min}, \alpha_{max}]$. The normalized phase is then discretized into B bins, which are defined by boundaries $\{\beta_0, \beta_1, \dots, \beta_B\}$. The number of phase bins was set to $B = 30$, which provides a balanced trade-off between resolution and statistical reliability: too few bins would smear sharp phase peaks, while too many bins would lead to sparsity and noisier probability estimates. Each observed phase x_n is assigned to an integer bin index k_n based on the condition:

$$k_n = argmax_k \mathbf{1}\{\beta_k \le x_n < \beta_{k+1}\} \quad with\ k = 0,1,\dots,B-1 \quad (3)$$

where $\mathbf{1}$ is the indicator function. Instead of modeling $p(x|c)$ directly, the discrete probabilities are expressed as:

$$p_k(c) = p(x \in [\beta_k, \beta_{k+1}|c) \quad with\ k = 0,1,\dots,B-1 \quad (4)$$

By adopting this binning approach, the training objective is simplified. The potential complexity of the phase distribution is preserved within the network outputs in the form of predicted discrete probabilities.

B. Neural Network Probability Model

To obtain $p_k(c)$, a feedforward neural network multi-layer perceptron (MLP) was designed to map the condition vector c to the probability of each bin, as illustrated in Fig. 4. An MLP is a type of artificial neural network composed of multiple layers of neurons, where each layer applies a weighted transformation followed by a nonlinear activation function to learn complex relationships in the data.

In this phase model, three standardized features (Let $c \in \mathbb{R}^3$ be standardized into $\tilde{c}$), i.e. rotational speed, voltage and frequency, are served as input. The input vector passes through three fully connected hidden layers with 128, 64, and 64 neurons respectively, each employing a ReLU activation function. To mitigate overfitting, Dropout is applied within these hidden layers. The network then outputs a B-dimensional vector z, whose component $\{z_k\}$ are the logits for each bin k. A softmax layer is subsequently used to convert these logits into probabilities:

$$p_k(c) = \frac{\exp(z_k)}{\sum_{j=0}^{B-1} \exp(z_j)} \quad (5)$$

If the observed phase for the n-th sample falls into bin k_n, the network can be trained by minimizing the cross-entropy loss:

$$L = -\frac{1}{N}\sum_{n=1}^{N}\sum_{k=0}^{B-1} \delta(k, k_n) \ln p_k(c_n) \quad (6)$$

where $\delta(k, k_n)$is the Kronecker delta, which equals 1 if $k = k_n$ and 0 otherwise. Once the network is trained, for any new condition vector c^*, it directly outputs $\{p_k(c^*)\}$, a discrete probability distribution over phase bins.

C. Kernel-Based Distribution Reconstruction

Since the output of MLP is a discrete histogram representing the impedance phase distribution under given conditions, a KDE method is applied to transform this discrete distribution into a continuous probability density function for data sampling.

To achieve this, a large number of data are first drawn from the histogram bins, where each selected bin corresponds to a normalized phase value. These sampled phase values are then mapped back to the original scale via inverse normalization, forming a sample set $\{x_m\}_{m=1}^{M}$. A kernel function $K(\cdot)$, typically Gaussian, is then used to reconstruct a smooth continuous distribution:

$$\hat{f}(x) = \frac{1}{Mh}\sum_{m=1}^{M} K\left(\frac{x - x_m}{h}\right) \quad (7)$$

where h is the bandwidth parameter, which is fixed at 0.2 based on empirical evaluation. A smaller value produces overfitted, spiky phase distributions that amplify noise, while a larger value oversmooths the curves and masks narrow breakdown-related peaks. This KDE approach effectively smooths the probability density, capturing multiple peaks or sharp transitions in the distribution. It also provides a clearer understanding of the likelihood that the bearing transitions from a capacitive (non-breakdown) state to a resistive (breakdown) state at different phase angles.

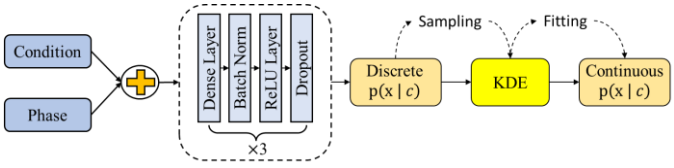


Fig. 4. Architecture of phase distribution model

IV. Amplitude Distribution Modelling

A. Amplitude Modeling under a Gaussian Assumption

Modeling the conditional amplitude distribution $p(y|x,c)$ is essential for capturing the complete bearing impedance behavior. The phase x is already learned through $p(x|c)$, but the amplitude y depends strongly on both the operating conditions $c = [RPM, V, f]$ and the phase x. A bin-based approach similar to $p(x|c)$ would require discretizing a higher-dimensional space (phase plus conditions), leading to excessive data sparsity and model complexity.

To address this, a more compact strategy was adopted by assuming that, within a small phase interval, the amplitude measurements can be approximated by a Gaussian distribution. This approximation balances modeling tractability with empirical and theoretical validity. Within a narrow 1° phase window, amplitude variations exhibit relative smoothness and are primarily influenced by numerous small, partially independent perturbations, such as oil film thickness fluctuations, contact resistance variations, micro-discharge

energy, and sensor noise. By the Central Limit Theorem, the aggregate effect of these bounded disturbances tends toward a normal distribution. This is consistent with electro-tribo-dynamic analyses of rolling element bearings [22], which attribute local electrical variation to stochastic thermal and mechanical factors, and with statistical studies on oil insulation breakdown [23], which report near-Gaussian behavior under fixed physical conditions despite globally skewed failure distributions. Observations of electrically-induced bearing damage [24] further link local impedance fluctuations to microscale discharge events and tribological instabilities, reinforcing the Gaussian assumption at the local scale.

For each operating condition, all amplitude samples are pooled to estimate a single, phase invariant standard deviation σ_c. The experimental phase spans from $-90°$ to $10°$, uniformly divided into 100 one-degree segments. For each segment, the corresponding amplitudes are collected and their sample mean $\bar{\mu}_i$ is computed. These statistics were then mapped to the center point of each segment (for instance, $-10°$ for the segment $[-11°, -10°)$), forming a coarse lookup of amplitude behavior. It is important to note that this segment-based statistical 3approach primarily serves as an initial reference for local Gaussian parameters. Ultimately, the neural network described below learns these parameters across the continuous space of phase and operating conditions, thus mitigating data sparsity issues that might arise from purely segment-based discretization.

B. Neural Network Probability Model

Real-world impedance amplitude measurements can span several orders of magnitude, especially during breakdown events, where ball bearings transition from a capacitive to a resistive state. This variability poses challenges for transformations based on Gaussian distributions. To address these issues, a small shift (e.g., 10^{-6}) is added to the raw amplitude data to ensure all values are positive. Subsequently, a Box-Cox transformation is applied, which is advantageous for reducing skewness, stabilizing variance, and transforming the data to better approximate a Gaussian distribution [25]. The Box-Cox transformation is defined as:

$$y_{BC} = \begin{cases} \dfrac{\tilde{y}^{\lambda} - 1}{\lambda}, & \lambda \neq 0 \\ \ln \tilde{y}, & \lambda = 0 \end{cases} \tag{8}$$

where $\tilde{y}$ is the shifted amplitude. Following the transformation, the data is normalized to have zero mean and unit variance, which accelerates network convergence and enhances model stability.

After data preprocessing, a feedforward neural network (MLP) is employed to model the conditional probability distribution $p(y|x,c)$. The architecture of the amplitude model is presented in Fig. 5. The network takes four input features: rotational speed, voltage, frequency, and phase. These inputs are processed through three fully connected hidden layers with 64, 32, and 32 neurons, respectively. The output of each hidden layer is activated by ReLU activation function. At the output stage, the network splits into two parallel heads:

1) Mean head: predicts the conditional mean amplitude $\hat{\mu}(u)$, where $u = [c, x]$.
2) STD head: predicts the phase-invariant standard deviation $\hat{\sigma}(u)$, a soft-plus activation ensures strictly positive.

The conditional distribution is interpreted as Gaussian:

$$p(y|x,c) \approx \mathcal{N}(y; \hat{\mu}(u), \hat{\sigma}(u)^2) \tag{9}$$

The network is trained by minimizing a mean-squared-error (MSE) objective instead of the full negative log-likelihood to avoid numerical instabilities when $\hat{\sigma}$ is close to zero.

Given a training set $\{(c_n, x_n, \bar{\mu}_n, \sigma_{c_n})\}_{n=1}^{N}$, where $\bar{\mu}_n$ is the empirical mean of the n-th phase segment and σ_{c_n} is the global standard deviation for condition c_n, the loss function is

$$L = \frac{1}{N}\sum_{n=1}^{N}\left[(\bar{\mu}_n - \hat{\mu}(u_n))^2 + \left(\sigma_{c_n} - \hat{\sigma}(c_n)\right)^2\right] \tag{10}$$

This criterion independently penalizes deviations in the predicted mean and predicted standard deviation, driving the network to regress both statistics simultaneously.

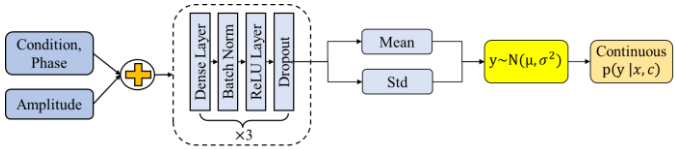


Fig. 5. Architecture of amplitude distribution model

C. Generation of Impedance Distribution

For a complete bearing impedance model, the approach for $p(x|c)$ can be combined with a complementary model $p(y|x,c)$ to represent how the amplitude y depends on both the phase x and the condition vector c. Once the joint distribution $p(x,y|c)$ is learned, sampling from it enables more comprehensive simulation of bearing breakdown processes under various speeds, voltages, and frequencies. Since the learned results are consistent with underlying electrical and oil-film thickness models, this probabilistic framework offers valuable insights for accurate bearing current modelling.

To generate the complete impedance distribution, a large number of phase samples $\{x_m\}$ were first drawn from the kernel-smoothed distribution of $p(x|c)$. Each sampled phase x_m was then passed to the trained neural network together with the condition vector c to predict $\hat{\mu}_m$ and $\hat{\sigma}_m$, producing amplitude samples $\{y_m\}$. As a result, a synthetic cloud of points $\{(x_m, y_m)\}$ was generated, providing a probabilistic view of the bearing impedance under the specified RPM, V, and f. This cloud not only revealed expected amplitude-phase relationships in non-breakdown conditions (highly capacitive states) but also captured potential shifts toward resistive behaviors at higher speeds or voltages, enabling more accurate fault detection and risk assessment for bearing breakdown events.

V. Experiment Validation

A. Experimental validation method

To validate the effectiveness of the proposed method, ball bearing impedance data under different working conditions was obtained through experiments. The schematic and the platform of experimental setup are illustrated in Fig. 6 and Fig. 7 [6],

respectively. The bearing under test is an SKF 61 905 2RS1. An impedance analyzer applies different voltage amplitudes and frequencies between the shaft and the bearing holder. A carbon brush, contacting the inner raceway of the bearing, and the bearing holder, in contact with the outer raceway, establishes the electrical connection. To reduce measurement noise, the rig incorporates a dielectric plate with capacitive coupling, and dielectric foils are inserted between the main platform and the carbon-brush holder. By varying the shaft speed, voltage amplitude and frequency, the system measures and records the impedance (both amplitude and phase). A full factorial grid of 500 unique operating conditions is used, comprising 10 speed levels (0, 100 r/min, 200 r/min, 400 r/min to 1600 r/min), 5 voltage levels (1 V to 5 V), and 10 frequency levels (100 kHz to 1 MHz). The distribution across levels is uniform by design. For each combination of condition c (rotational speed, voltage, frequency), the system collects 500 impedance measurements (including phase x and amplitude y), with a sampling time of approximately 6 ms. In total, this yields 250,000 impedance records. The observed phase distribution may fall into purely capacitive (non-breakdown), transitional (mix of breakdown and non-breakdown), or purely resistive (breakdown) states.

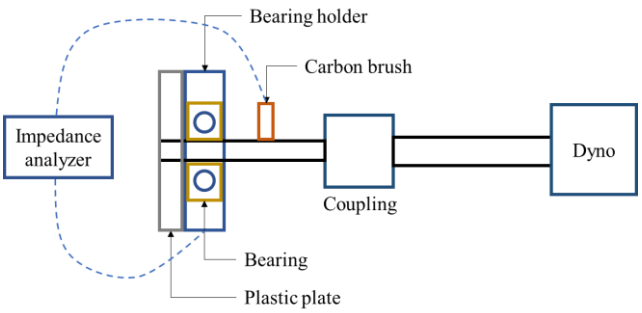


Fig. 6. Schematic of the experiment setup [6]

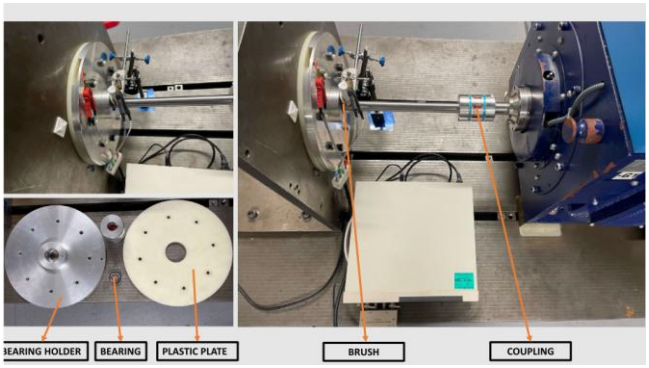


Fig. 7. Experiment platform [6]

A condition-based data partitioning strategy, defined by combinations of rotational speed, voltage, and frequency is employed to evaluate the model accuracy and generalization capability. First, the entire dataset is divided into distinct groups according to these operating conditions, with 80% allocated to the development pool and the remaining 20% reserved for an unseen test set.

To ensure model robustness and stability, key hyperparameters were selected through lightweight grid and random search procedures conducted solely within the development pool. For both the phase and amplitude models, we explored network depth (2–4 layers), width (32–128 neurons per layer), and dropout rate (0–0.5). Additionally, for the phase model, we optimized the number of discretized phase bins B in the range 10–50. The final selections balanced resolution, generalization, and computational cost. The full search space and selected optimal values are summarized in Table I.

TABLE I
HYPERPARAMETERS TO BE OPTIMZED

Component	Hyperparameter	Search space	Optimal
Phase model	Hidden layers	2,3,4	3
	Number of neurons	32, 64, 128	128, 64, 64
	Dropout	0, 0.1, 0.2	0.2
	Bins	10, 20, 30, 40, 50	30
Amplitude model	Hidden layers	2,3,4	3
	Number of neurons	32 64 128	64, 32, 32
	Dropout	0, 0.1, 0.2	0.2
Training	Learning rate	5e-4, 1e-3, 2e-3	1e-3
	Optimizer	Adam	Adam
	Batch size	32	32

Given the relatively small dataset (500 instances), Group K-Fold cross-validation (e.g., K = 5) is adopted in the development pool: in each fold, samples from K−1 condition groups are used for training, while the remaining group serves as the validation fold for hyperparameter tuning and optimal model selection. By ensuring that every validation fold contains operating conditions never encountered during the corresponding training phase, and by keeping the test set completely out of the cross-validation loop, the approach provides a realistic assessment of the model performance in previously unobserved operating regions. While in practical scenarios the entire dataset covering all operational conditions is typically utilized to maximize predictive performance, this condition-based Group K-Fold partitioning provides critical insights into the robustness and generalizability of the proposed method when confronted with states beyond the training range.

Model validation is performed in three main steps. The flowchart is shown in Fig. 8. First, under the specified operating conditions, the condition vector is fed into the trained phase model to obtain the conditional phase probability distribution. Next, random samples are drawn from this phase distribution and, along with the same operating conditions, are used as inputs to the trained amplitude model to produce the corresponding amplitude probability distribution. Finally, additional samples are drawn from the amplitude distribution and paired with the previously sampled phases to form the overall impedance distribution, providing a comprehensive evaluation of bearing impedance under the given conditions.

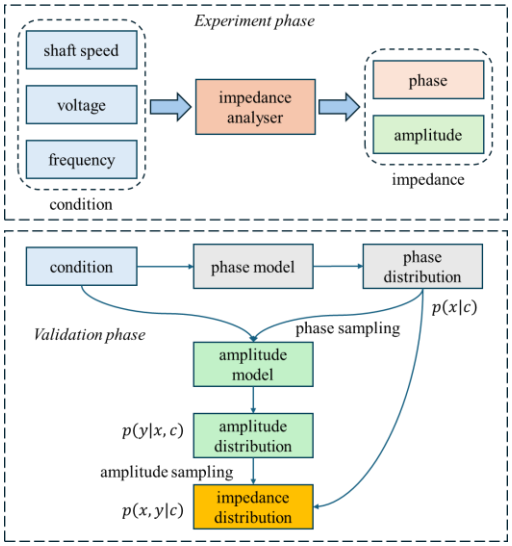
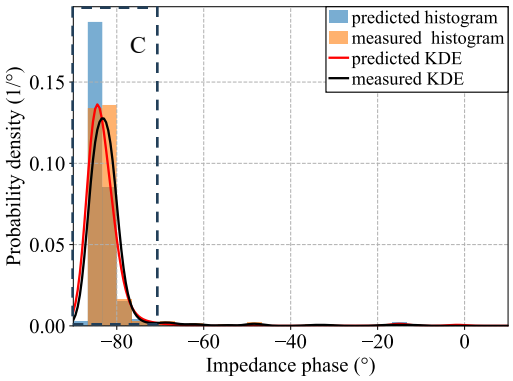


Fig. 8. Experiment validation flowchart.

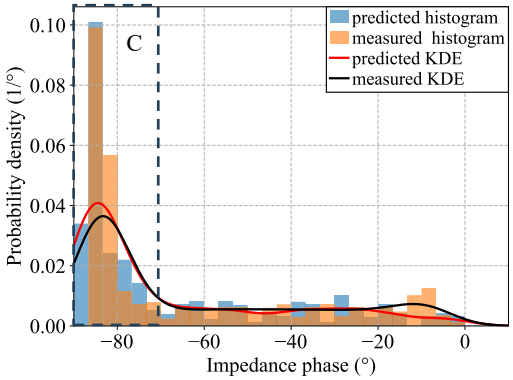
B. Phase distribution prediction

Accurately estimating the impedance breakdown probability, under various operating conditions, is crucial for predicting EDM bearing currents. Fig. 9 compares four bearing impedance states under four different conditions. These states are categorized based on the impedance phase angle: a phase greater than –20° typically indicates a resistive (breakdown) state, between –20° and –70° indicates a transitional state, and less than –70° corresponds to a capacitive (non-breakdown) state. In Fig. 9 (a), the impedance is in a capacitive state (the bearing is essentially at no risk of breakdown). In Fig. 9 (b), the impedance is in a transitional state leaning toward capacitive (the bearing has a relatively low breakdown probability). In Fig. 9 (c), the impedance is in a transitional state leaning toward resistive (the bearing has a relatively high breakdown probability). In Fig. 9 (d), the impedance is in a fully resistive state (the bearing experiences breakdown).

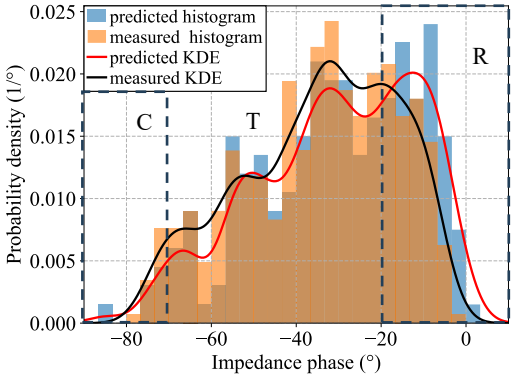
Experimental results show that the predicted impedance phase distributions align well with the actual measurements across a range of operating conditions, demonstrating the effectiveness of the proposed model.



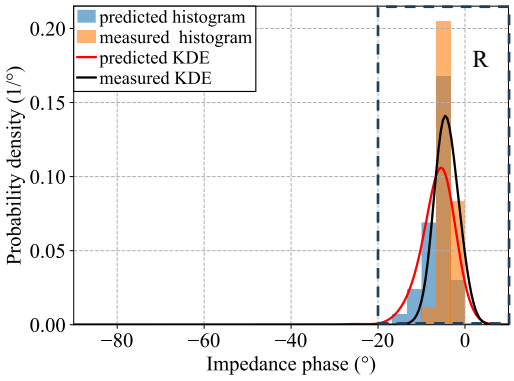
(a)



(b)



(c)



(d)

Fig. 9. Predicted phase distribution versus measured distribution under four operating conditions. (a) 200 r/min, 1 V, 500 kHz; (b) 400 r/min, 3 V, 300 kHz; (c) 1400 r/min, 2 V, 100 kHz; (d) 1000 r/min, 4 V, 700 kHz. C = capacitive state; T = transition state; R = resistive state.

C. Impedance distribution prediction

The phase model evaluates the bearing breakdown probability distribution under various operating conditions, while the amplitude model provides the impedance magnitude for a given phase, both of which are needed in a bearing current calculation. To validate the accuracy of the amplitude model and the effectiveness of the overall framework, Fig. 10 offers a visual comparison between the predicted and actual impedance point clouds across various operating conditions. In particular, the real impedance shown in Fig. 10(a) is largely concentrated in the non-breakdown region, whereas Figs. 10(b) and (c) exhibit a distinctly multi-modal distribution. Other sub-figures reveal

that the impedance amplitude can vary significantly across a broad phase range. Fig. 10(d) illustrates a scenario where the entire distribution falls within the breakdown region, and Fig. 10(e) similarly shows a large portion of the distribution in breakdown. Finally, Fig. 10(f) spans from breakdown to non-breakdown states, displaying a pronounced long-tail distribution.

To quantify the accuracy of the model, the Earth Mover's Distance (EMD) was employed to assess the overall discrepancy between the predicted and the real distributions. EMD measures the minimal transportation cost required to shift each discrete point in the predicted distribution to the corresponding location in the real distribution within the two-dimensional (amplitude-phase) space. To eliminate discrepancies arising from different measurement scales or coordinate ranges under varied experimental conditions, the amplitude and phase were both constrained to their respective minimum and maximum values for each operating condition. Subsequently, EMD was normalized to the $[0, 1]$ range to define the point-cloud similarity:

$$d_{max} = \sqrt{(x_{max} - x_{min})^2 + (y_{max} - y_{min})^2} \tag{11}$$

$$S_{EMD} = 1 - \frac{EMD}{d_{max}} \tag{12}$$

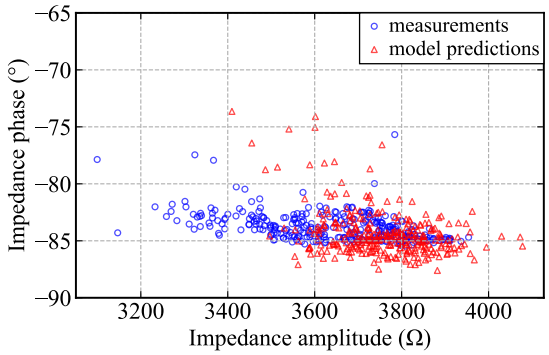
where x_{max} and x_{min} denote the maximum and minimum phases, y_{max} and y_{min} denote the maximum and minimum amplitudes, and d_{max} is the maximum possible distance in the amplitude-phase plane. A similarity S_{EMD} that approaches 1 indicates a smaller discrepancy and thus closer alignment of the two distributions. The six impedance distribution comparisons shown in Fig. 10, from (a) to (f), have point cloud EMD similarities of 0.954, 0.937, 0.97, 0.92, 0.981, and 0.919, respectively.

To complement EMD, the symmetric nearest neighbor root mean square error (RMSE) is also reported to capture local geometric alignment between the predicted and measured point clouds. While EMD quantifies the global distribution shift, RMSE measures pointwise similarity by matching each predicted point to its nearest measured point, and vice versa, averaging the squared distances in both directions before taking the square root. Consistent with the EMD normalization in (11)-(12), RMSE is mapped to $[0, 1]$ as

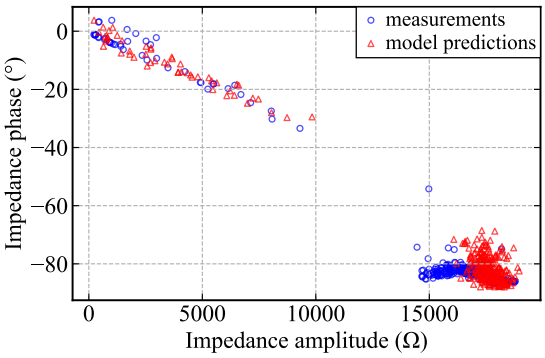
$$S_{RMSE} = 1 - \frac{RMSE}{d_{max}} \tag{13}$$

where a larger S_{RMSE} indicates closer alignment. For the six operating conditions in Fig. 10, the corresponding S_RMSE values are 0.973, 0.992, 0.980, 0.976, 0.994, and 0.993, confirming strong local agreement between predictions and measurements.

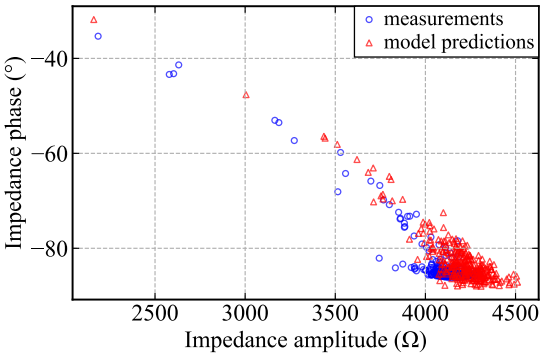
The proposed model not only accurately captures these features and flexibly predicts the coupled relationship between phase and amplitude, but also maintains high accuracy in predictions under diverse conditions, thereby demonstrating its effectiveness for bearing impedance analysis and breakdown risk assessment.



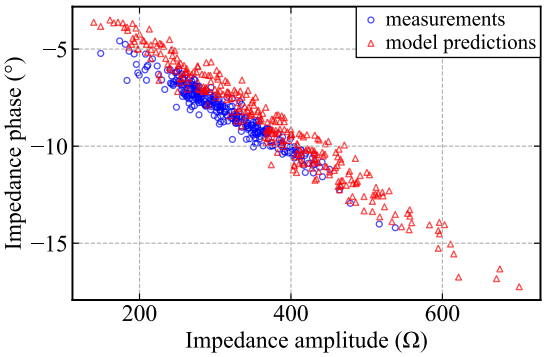
(a)



(b)



(c)



(d)

TABLE II

COMPARISON OF REPRESENTATIVE MODELLING APPROACHES FOR BEARING VOLTAGE/IMPEDANCE PREDICTION

Method	Output form	Phase-amplitude transition modelled	Probabilistic / distribution output	Applicable operating range
Electromagnetic FEA + distributed-parameter equivalent circuit [12]	Time-domain transient and steady-state bearing voltage	No (amplitude only)	No	Various design and material scenarios
Equivalent circuit + variable breakdown resistance [15]	Breakdown process voltage/current waveforms	Yes (capacitive → resistive)	No	High-frequency excitation breakdown tests
Coupled mechanical-electrical analytical model [16]	Impedance (resistance + capacitance)	Yes (breakdown switch model)	No	Considers speed, temperature, load
Single steel ball experiment + thermo-elastic load distribution [17]	Capacitance value	No	No	Static and thermal effects
Data-driven regression (machine learning) [26]	Impedance value	Yes	No	Given speed, temperature, and load
Proposed (Chain-decomposition probabilistic model)	Joint amplitude-phase impedance distribution	Yes (phase-dependent amplitude evolution)	Yes (full probability density)	Cross-RPM, voltage, and frequency conditions

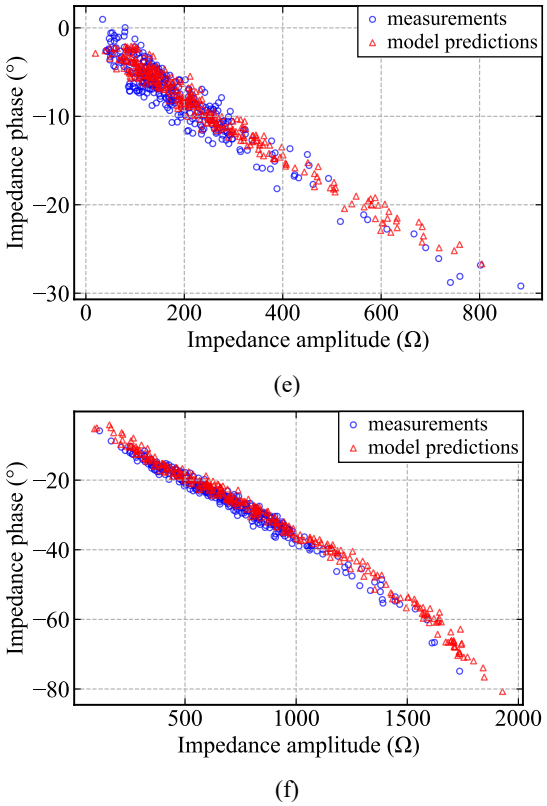


Fig. 10. Predicted impedance distribution versus measured data under six operating conditions. (a) 400 r/min, 2 V, 500 KHz. (b) 200 r/min, 2 V, 100 kHz. (c) 800 r/min, 2 V, 400 KHz. (d) 1200 r/min, 3 V, 800 kHz. (e) 200 r/min, 4 V, 900 KHz. (f) 800 r/min, 4 V, 900 kHz.

D. Comparison with Existing Works

To enhance clarity in reviewing prior studies, Table II summarizes representative modelling approaches for bearing voltage or impedance prediction. The comparison highlights key aspects such as modelling type, output form, capability to capture the capacitive-to-resistive transition, probabilistic treatment, and applicable operating range. It is worth noting that, except for [26], which employs a purely data-driven regression, all other referenced works are fundamentally based on analytical, physics-derived modelling, from equivalent circuit formulations to coupled mechanical-electrical impedance calculations, typically yielding point estimates rather than full probabilistic descriptions. Each has its own strengths: for example, Ref [12] achieves high-fidelity transient voltage simulation via FEA; Ref [15] models the breakdown process with variable resistance; Ref [16] offers physically interpretable predictions across speed, temperature, and load; Ref [17] provides precise capacitance evaluation using controlled single-ball experiments; and [26] reduces analytical-experimental discrepancies using machine learning. In contrast, the present work integrates physical interpretability with a chain-decomposition probabilistic framework, enabling joint amplitude-phase modelling, explicit representation of stochastic variability, and generalization across a wide range of operating conditions.

Beyond this qualitative comparison, quantitative benchmarks are also conducted on two classic regression methods (random forest and gradient boosting) following the configuration in [26], which, to the authors' best knowledge, is the only machine learning study on bearing impedance prediction so far. These models were trained using the same input variables (rotational speed, voltage amplitude, and frequency) to predict the average impedance amplitude.

Since these baseline models cannot produce impedance distributions or capture phase-dependent variations, the comparison was restricted to five operating conditions in predominantly capacitive states and five in predominantly resistive states. For each case, the percentage deviation between the predicted and measured average impedance amplitude was evaluated. As shown in Table III, the proposed model achieves

the highest accuracy in capacitive states (mean deviation 1.01 %), outperforming both baselines. In resistive states, random forest achieves the lowest deviation (1.61 %), while the proposed model outperforms gradient boosting. Importantly, unlike the baselines, the proposed method also models the complete impedance distribution and can capture phase-dependent variations.

TABLE III
COMPARISON WITH OTHER MACHINE LEARNING METHODS

Methods	Mean deviation (%)	
	capacitive states	resistive states
Proposed	1.01	3.07
Random forest	1.27	1.61
Gradient boosting	1.21	3.81

VI. CONCLUSION

A data-driven framework has been presented for modeling the bearing impedance distribution, taking into account both phase and amplitude across varying shaft speeds, voltage amplitudes, and excitation frequencies. A two-part procedure, involving the phase model and the conditional amplitude model, is introduced and validated using experimental data. Following the probabilistic chain rule, the two-dimensional point-cloud distribution is factorized into two one-dimensional distributions, each modeled with an individual MLP, which greatly reduces the difficulty of predicting the full point-cloud distribution directly. The primary findings can be summarized as follows:

1) Phase distribution: abrupt changes or multimodal trends in phase are successfully captured by the bin-based neural network and KDE, effectively distinguishing between capacitive and resistive regimes.
2) Amplitude modeling: adopting a Gaussian assumption within local phase intervals provides an efficient trade-off between simplicity and representational flexibility.
3) Breakdown identification: There is a complex, nonlinear relationship between shifts in the ball bearing impedance and the input conditions. Using this machine learning based model, the probability of failure under various operating conditions can be estimated.

In addition to its current form, the proposed probabilistic framework allows seamless expansion to account for more diverse operating or environmental factors (e.g., temperature, lubrication states) by augmenting the input condition vector. The factorized form $p(x|c)p(y|x,c)$ remains valid when c includes additional variables, ensuring mathematical consistency while preserving the interpretability of the model. Although such variables were not included in the current experiments, the modular design offers a scalable pathway for future improvements. This structure also enables adaptation to other bearing types via minor fine-tuning or by integrating basic bearing descriptors into the input space. Ultimately, combining impedance modeling with extended physical insights may yield more comprehensive bearing current predictions and contribute to improved reliability of inverter-fed electrical machines.

REFERENCES

[1] C. Guo *et al.*, "Investigation of Bearing Current and Degradation in Wide Bandgap Inverter Fed Electric Machines," in *2024 International Symposium on Power Electronics, Electrical Drives, Automation and Motion (SPEEDAM)*, Napoli, Italy: IEEE, Jun. 2024, pp. 1223–1229. doi: 10.1109/SPEEDAM61530.2024.10609178.

[2] S. Chen, T. A. Lipo, and D. Fitzgerald, "Source of induction motor bearing currents caused by PWM inverters," *IEEE Trans. On energy Conversion*, vol. 11, no. 1, pp. 25–32, Mar. 1996, doi: 10.1109/60.486572.

[3] I. Husain *et al.*, "Electric Drive Technology Trends, Challenges, and Opportunities for Future Electric Vehicles," *Proc. IEEE*, vol. 109, no. 6, pp. 1039–1059, Jun. 2021, doi: 10.1109/JPROC.2020.3046112.

[4] Shaotang Chen and T. A. Lipo, "Bearing currents and shaft voltages of an induction motor under hard- and soft-switching inverter excitation," *IEEE Trans. on Ind. Applicat.*, vol. 34, no. 5, pp. 1042–1048, Oct. 1998, doi: 10.1109/28.720444.

[5] A. L. Julian, G. Oriti, and T. A. Lipo, "Elimination of common-mode voltage in three-phase sinusoidal power converters," *IEEE Trans. Power Electron.*, vol. 14, no. 5, pp. 982–989, Sep. 1999, doi: 10.1109/63.788504.

[6] D. De Gaetano, W. Zhu, X. Sun, X. Chen, A. Griffo, and G. W. Jewell, "Experimental Ball Bearing Impedance Analysis Under Different Speed and Electrical Conditions," *IEEE Trans. Dielect. Electr. Insul.*, vol. 30, no. 3, pp. 1312–1321, Jun. 2023, doi: 10.1109/TDEI.2023.3271958.

[7] D. Zheng *et al.*, "The Effect of Bearing Impedance on Online Condition Monitoring for Rotor Winding Insulation of Doubly-Fed Induction Generator," *IEEE Trans. Power Electron.*, vol. 39, no. 9, pp. 11697–11707, Sep. 2024, doi: 10.1109/TPEL.2024.3416324.

[8] P. Han, G. Heins, D. Patterson, M. Thiele, and D. M. Ionel, "Combined Numerical and Experimental Determination of Ball Bearing Capacitances for Bearing Current Prediction," in *2020 IEEE Energy Conversion Congress and Exposition (ECCE)*, Detroit, MI, USA: IEEE, Oct. 2020, pp. 5590–5594. doi: 10.1109/ECCE44975.2020.9235700.

[9] O. Magdun, Y. Gemeinder, and A. Binder, "Investigation of influence of bearing load and bearing temperature on EDM bearing currents," in *2010 IEEE Energy Conversion Congress and Exposition*, Atlanta, GA: IEEE, Sep. 2010, pp. 2733–2738. doi: 10.1109/ECCE.2010.5618061.

[10] J. Park, T. R. Wellawatta, S.-J. Choi, and J. Hur, "Mitigation Method of the Shaft Voltage According to Parasitic Capacitance of the PMSM," *IEEE Trans. on Ind. Applicat.*, vol. 53, no. 5, pp. 4441–4449, Sep. 2017, doi: 10.1109/TIA.2017.2717378.

[11] J.-H. Im, Y.-H. Choo, S.-H. Lee, J. Hur, and J.-H. Heo, "Modeling Equivalent Circuit of Shaft Voltage for EV Motor with Ceramic Bearings," in *2024 IEEE Energy Conversion Congress and Exposition (ECCE)*, Phoenix, AZ, USA: IEEE, Oct. 2024, pp. 5394–5397. doi: 10.1109/ECCE55643.2024.10861564.

[12] P. Han, G. Heins, D. Patterson, M. Thiele, and D. M. Ionel, "Modeling of Bearing Voltage in Electric Machines Based on Electromagnetic FEA and Measured Bearing Capacitance," *IEEE Trans. on Ind. Applicat.*, vol. 57, no. 5, pp. 4765–4775, Sep. 2021, doi: 10.1109/TIA.2021.3092700.

[13] P. Han *et al.*, "On the Modeling of Bearing Voltage and Current in PWM Converter-fed Electric Machines Using Electromagnetic Finite Element Analysis," in *2021 IEEE Energy Conversion Congress and Exposition (ECCE)*, Vancouver, BC, Canada: IEEE, Oct. 2021, pp. 4606–4610. doi: 10.1109/ECCE47101.2021.9595362.

[14] M. T. A. Êvo, A. M. Alzamora, I. O. Zaparoli, and H. D. Paula, "Inverter-Induced Bearing Currents: A Thorough Study of the Cause-and-Effect Chains," *IEEE Ind. Appl. Mag.*, vol. 29, no. 3, pp. 57–66, May 2023, doi: 10.1109/MIAS.2022.3214026.

[15] X. Ren, R. Liu, and E. Yang, "Modelling of the Bearing Breakdown Resistance in Bearing Currents Problem of AC Motors," *Energies*, vol. 12, no. 6, p. 1121, Mar. 2019, doi: 10.3390/en12061121.

[16] Y. Gemeinder, M. Schuster, B. Radnai, B. Sauer, and A. Binder, "Calculation and validation of a bearing impedance model for ball bearings and the influence on EDM-currents," in *2014 International Conference on Electrical Machines (ICEM)*, Berlin, Germany: IEEE, Sep. 2014, pp. 1804–1810. doi: 10.1109/ICELMACH.2014.6960428.

[17] S. Puchtler, J. Van Der Kuip, and E. Kirchner, "Analyzing Ball Bearing Capacitance Using Single Steel Ball Bearings," *Tribol Lett*, vol. 71, no. 2, p. 38, Jun. 2023, doi: 10.1007/s11249-023-01706-7.

[18] A. Muetze and A. Binder, "Calculation of Circulating Bearing Currents in Machines of Inverter-Based Drive Systems," *IEEE Trans. Ind. Electron.*, vol. 54, no. 2, pp. 932–938, Apr. 2007, doi: 10.1109/TIE.2007.892001.

[19] D. Busse, J. Erdman, R. J. Kerkman, D. Schlegel, and G. Skibinski, "Bearing Currents and Their Relationship to PWM Drives," *IEEE TRANSACTIONS ON POWER ELECTRONICS*, vol. 12, no. 2, 1997.

[20] O. Magdun, Y. Gemeinder, and A. Binder, "Rotor impedance of the high frequency circulating bearing current path in inverter-fed AC machines," in *2013 IEEE Energy Conversion Congress and Exposition*, Denver, CO, USA: IEEE, Sep. 2013, pp. 3512–3519. doi: 10.1109/ECCE.2013.6647163.

[21] S. Węglarczyk, "Kernel density estimation and its application," *ITM Web Conf.*, vol. 23, p. 00037, 2018, doi: 10.1051/itmconf/20182300037.

[22] P. Welzbacher, S. Puchtler, A. Geipl, and E. Kirchner, "Uncertainty Analysis of a Calculation Model for Electric Bearing Impedance," *Proc. Des. Soc.*, vol. 2, pp. 653–662, May 2022, doi: 10.1017/pds.2022.67.

[23] H. Gupta and S. Das, "Statistical analysis of oil insulation breakdown voltage," in *2017 IEEE International Conference on Industrial Engineering and Engineering Management (IEEM)*, Singapore: IEEE, Dec. 2017, pp. 2044–2048. doi: 10.1109/IEEM.2017.8290251.

[24] R. Turnbull, R. Rahmani, S. Paul, and H. Rahnejat, "Electrotribodynamics of ball bearings in electrical machines," *Tribology International*, vol. 188, p. 108817, Oct. 2023, doi: 10.1016/j.triboint.2023.108817.

[25] S. Marimuthu, T. Mani, T. D. Sudarsanam, S. George, and L. Jeyaseelan, "Preferring Box-Cox transformation, instead of log transformation to convert skewed distribution of outcomes to normal in medical research," *Clinical Epidemiology and Global Health*, vol. 15, p. 101043, May 2022, doi: 10.1016/j.cegh.2022.101043.

[26] E. Kirchner, C. Bienefeld, T. Schirra, and A. Moltschanov, "Predicting the Electrical Impedance of Rolling Bearings Using Machine Learning Methods," *Machines*, vol. 10, no. 2, p. 156, Feb. 2022, doi: 10.3390/machines10020156.