

Article

Application of AI-Driven Methods in Quenching Distortion Control of Mesh Belt Furnaces

Xusheng Li ^{1,2,3}, Yixiao Sun ⁴ , Xiaohu Deng ⁵ , Jiangang Wang ⁶ , Yang Ju ³, Mingzhou Wang ², Lingchu Wang ², Hao Chen ² and Dongying Ju ^{2,*}

¹ School of Xinchang, Zhejiang Industry Polytechnic College, Shaoxing 312000, China; lixushengt@sina.com

² Zhejiang XCC Group Co., Ltd., Shaoxing, 312500, China; wangmingzhou@xcc-zxz.com (M.W.); wanglingchu@xcc-zxz.com (L.W.); chenhao@xcc-zxz.com (H.C.)

³ School of Mechanical Engineering, Zhejiang University, Hangzhou 310030, China; yang.ju@zju.edu.cn

⁴ School of Electronic and Information Engineering, University of Science and Technology Liaoning, Anshan 114051, China; robinsunyixiao@163.com

⁵ School of Mechanical Engineering, Tianjin University of Technology and Education, Tianjin 300222, China; dengxh@tute.edu.cn

⁶ College of Material Science and Engineering, Hebei University of Science and Technology, Shijiazhuang 050018, China; wm094212@163.com

* Correspondence: dyjusitec@sina.com

Abstract

Bearing rings are thin-walled components prone to distortion during quenching. Achieving high-precision distortion control for bearing rings remains a critical challenge in high-precision bearing manufacturing. This paper proposes an AI-driven method for distortion control during bearing ring quenching in mesh-belt furnaces. The primary objective is to identify the optimal reverse motor frequency within the furnace. An experimental database is established using distortion results obtained at different reverse motor frequencies. This database serves as training data for deep learning. Subsequently, a large language model (LLM) employs few-shot learning to optimize and predict the reverse motor frequency influencing bearing quenching distortion. Results demonstrate that the LLM method achieves significantly higher prediction accuracy than traditional machine learning approaches. The optimization outcomes validate the effectiveness of generative AI in optimizing the reverse motor frequency of mesh-belt furnaces and controlling distortion during bearing ring quenching.

Keywords: bearing; AI-driven technology; measurement; distortion; reverse motor frequency



Academic Editors: Cheng Zeng and Menghao Yang

Received: 21 December 2025

Revised: 27 January 2026

Accepted: 12 February 2026

Published: 22 February 2026

Copyright: © 2026 by the authors.

Licensee MDPI, Basel, Switzerland.

This article is an open access article distributed under the terms and conditions of the [Creative Commons Attribution \(CC BY\)](https://creativecommons.org/licenses/by/4.0/) license.

1. Introduction

Bearings are ubiquitous and indispensable components across numerous fields including aerospace, precision machine tools, robotics, automobiles, and high-speed rail. They must possess properties such as fatigue resistance, wear resistance, high strength, and extended service life. This necessitates that the heat treatment process, serving as the near-net-shape finishing for bearings, not only achieves fatigue resistance and high strength but also minimizes distortion of the bearing rings during heat treatment. The goal is to maximize the grindable allowance of the rings post-heat treatment, thereby reducing costs while effectively safeguarding and enhancing the overall service performance of the bearings [1–8]. However, as thin-walled annular components, bearing inner and outer rings must be heated above the austenitizing temperature during heat treatment and then immersed in quenching oil, where complex nucleate boiling, vapor film attachment, and

film boiling phenomena occur. These phenomena cause differences in bearing cooling rates, leading to non-uniform phase transformation structures, hardness, and distortion across different regions [9–12]. Therefore, how to predict and control bearing distortion, internal metallic structure changes, and hardness has always been an important research topic in bearing production.

In industrial practice, mesh belt furnaces are widely employed for carburizing and quenching the inner and outer rings of bearings. This automated equipment, optimized for high-efficiency mass production, is often operated with oil as the cooling medium. By precisely controlling furnace temperature, heating duration, and cooling parameters, a stable martensitic microstructure can be obtained, thereby enhancing wear resistance and fatigue strength. Nevertheless, in large-scale production, the uniformity of oil quenching critically affects post-quench distortion. During continuous feeding, components are heated to complete austenitization in the furnace's heating zone before being transferred into the quenching oil. However, stacking during the cooling stage can compromise uniform cooling. To mitigate this issue, an additional secondary mesh belt is installed in the oil tank to ensure that each end face of the bearing ring sequentially contacts the secondary belt and the lifting belt. The residence time of the rings on the secondary belt is crucial—either excessive or insufficient durations fail to achieve the intended effect. The process is illustrated in Figure 1. From a quality control perspective, adjusting the reverse motor frequency of the secondary belt has been identified as an effective approach to minimize distortion, making it a significant factor in large-scale bearing production. However, given the variability in operational conditions across different batches—due to production equipment and tooling constraints—experimentally validating the optimal frequency for each scenario is highly challenging.

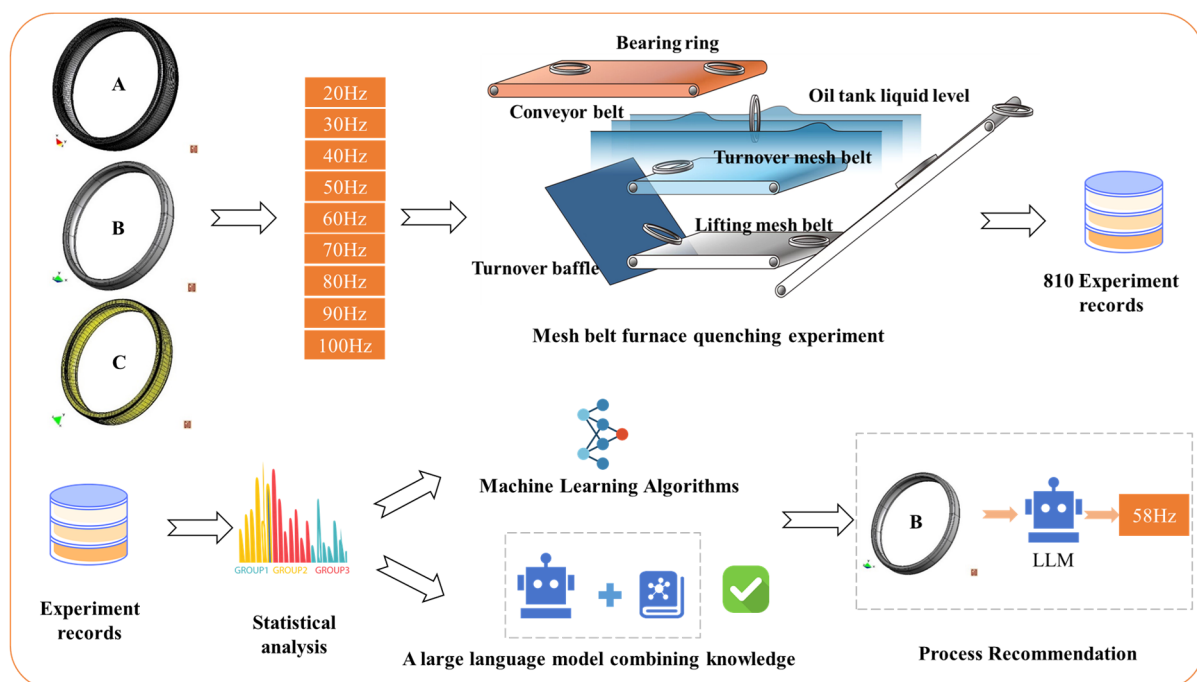


Figure 1. Research roadmap.

In recent years, with the rapid development of artificial intelligence (AI) technology, data-driven methods have been increasingly introduced into the modeling and process parameter optimization of complex heat treatment processes. Existing research indicates that machine learning and deep learning methods have certain advantages in establishing mapping relationships between process parameters and material properties or distortion

responses. For instance, Kusano et al. [13] predicted the tensile properties of Ti-6Al-4V alloy after additive manufacturing and heat treatment using multiple linear regression combined with quantitative microstructural features. Hernandez et al. [14] compared the applicability of multiple regression and random forest models in predicting the stabilization time of a heat treatment furnace. Zhu et al. [15] and Wang et al. [16] employed XGBoost and support vector machine (SVM) models, respectively, to model steel hardenability and gear heat treatment process control. Such studies have validated the feasibility of data-driven models in predicting heat treatment processes, yet their predictive performance typically relies on a relatively ample sample size. Building upon this foundation, some studies have further integrated machine learning models with optimization algorithms or numerical simulation methods to achieve multi-objective process parameter optimization. Nandal et al. [17] utilized artificial intelligence methods to explore the design of non-isothermal aging processes for Ni–Al alloys, demonstrating the application prospects of AI in complex heat treatment path design. For example, Chintakindi et al. [18] combined principal component analysis, machine learning models, and the particle swarm optimization algorithm to perform multi-objective optimization of the annealing process parameters for Monel 400 alloy. Xia et al. [19] proposed a process optimization framework coupling finite element analysis with a neural network, applied to the optimal design of heat treatment parameters for bearing rings. Oh and Ki [20] utilized deep neural networks to predict the hardness distribution of tool steel during laser heat treatment. Jia et al. [21] proposed a deep learning-based model for predicting post-quench hardness and carburized layer depth, which was then applied to process parameter recommendation. Wang et al. [22] constructed a neural network-based multi-objective optimization model that simultaneously considered hardness, distortion, and total helix deviation, significantly reducing the total helix deviation in gear heat treatment. While such methods demonstrate strong capability in process search, they typically involve complex model construction, high parameter tuning costs, and often struggle to maintain stability when experimental samples are limited.

The potential of large language models (LLMs) in engineering design and manufacturing decision-making is also gaining attention. Sun et al. [23] proposed a large language model-based method for quenching process design, achieving end-to-end intelligent modeling of the process flow. However, existing LLM-related research has predominantly focused on process path generation or decision support at the workflow level. Their capability for quantitative prediction of continuous process parameter–performance relationships and stable interpolation under small-sample conditions still requires further investigation.

In summary, while existing research has achieved significant progress in modeling heat treatment processes and optimizing process parameters, current methods still face challenges of insufficient prediction stability and limited engineering applicability in industrial scenarios characterized by limited experimental samples, highly nonlinear frequency–distortion relationships, and the need to explicitly incorporate empirical process rules. To address these challenges, this paper proposes a prior-guided regression and process parameter recommendation method based on a pre-trained large language model. By uniformly encoding geometric parameters, process conditions, and domain knowledge into structured prompts, the method achieves stable prediction of bearing quenching distortion trends and optimization of the reverse motor frequency under small-sample conditions, without the need for training an additional predictive model.

In this study, targeting the secondary quenching process in a mesh-belt continuous furnace, experiments were designed to quantitatively analyze the process, and a method for distortion prediction and process parameter optimization under small-sample conditions was proposed. First, this paper establishes an experimental database comprising measurements from 810 thin-walled bearing outer rings, covering three typical geometric types

and nine discrete reverse motor frequencies (20–100 Hz). The difference in diameter values between the upper and lower end faces and the oval distortion are adopted as the primary evaluation metrics. Given the sparse and highly nonlinear characteristics of the frequency–distortion relationship, the predictive stability of conventional regression methods under small-sample conditions is analyzed. Innovatively, a pretrained large language model (LLM) is introduced as a prior-constrained regressor. By encoding geometric parameters, frequency information, and historical distortion statistics into structured prompts, along with embedding key process rules, stable predictions of the distortion trend within untested frequency intervals are achieved.

Building upon this foundation, a frequency optimization workflow for process recommendation is further established. Under the premise of satisfying quality constraints, the optimal reverse motor frequency for bearing rings of different geometric types is determined via fine-step search. Supplementary experimental results verify the engineering applicability of this method in terms of reducing testing costs, improving parameter selection efficiency, and supporting decision-making for scaled production.

2. Research Methods

Thin-walled bearing rings are highly prone to distortion after heat treatment, which can be generally classified into two types: the difference in the average diameters between the upper and lower end faces of the bearing, and elliptical distortion. Research has shown that adjusting the reverse motor frequency can effectively reduce bearing distortion. Based on this, the study intends to use experimental research and AI-driven technology to explore the effect of the reverse motor frequency during the quenching process of bearing rings in the mesh belt furnace. First, the experiments were designed and the experimental results were analyzed. A database was established based on the experimental results, and the few-shot learning method combining machine learning and large language models (LLMs) was adopted to predict the distortion of bearings during the quenching process. The experimental results show that LLMs are significantly superior to traditional machine learning methods in terms of prediction accuracy. On this basis, to further explore the application potential of LLMs in intelligent process optimization, LLMs were tested on process recommendation tasks, and experimental verification was conducted accordingly. The main workflow is shown in Figure 1.

2.1. Experimental Material and Models

The bearing material used was high-carbon chromium bearing steel GCr15, and its chemical composition is listed in Table 1.

Table 1. Chemical composition of GCr15 steel.

Compositions	C	Si	Mn	P	S	Ni	Cr	Ti	Fe
Content	0.97	0.26	0.34	0.018	0.002	0.04	1.45	0.0019	Bal.

For this experiment, three types of thin-walled bearing rings (hereafter referred to as A, B, and C) were selected as experimental samples, and the dimensions of the three types are shown in Table 2.

Table 2. Dimensions of the three types.

Parameter	Height	Outer Diameter	Inner Diameter	Wall Thickness
A-Size(mm)	27.25	110.20	95.50	7.35
B-Size(mm)	14.90	115.15	103.90	5.625
C-size(mm)	22.3	125.35	110.6	7.375

2.1.1. Experimental Methods

Based on practical experience, the main factors affecting distortion during actual heat treatment production include the stacking of the workpieces, heating time and temperature, cooling medium, lateral wandering of the mesh belt in the quenching oil, stirring rate of the quenching oil, and the reverse motor frequency of the mesh belt in the quenching tank. This study first excludes the factors among the aforementioned six that are convenient to control and analyze, with a focus on investigating the effect of the mesh belt's reverse motor frequency and its distortion mechanism on thin-walled bearing rings during the continuous mesh belt furnace quenching process. Simultaneously, aiming to minimize quenching distortion in the mesh belt furnace, a method for calculating the optimal reverse motor frequency of the bearing rings on the mesh belt is proposed.

The system comprises a UM010080 drive unit, a quenching medium (model KR468G-6) with a cooling rate of 76 °C/s, and a quenching tank equipped with a temperature-control thermocouple, an oil-tank stirrer motor, a reverse mesh belt, a filter screen, a circulation pump, a sealing device, and an oil-tank hoist, where the reverse motor frequency ranges from 20 to 100 Hz. Based on mechanistic understanding and practical considerations, certain factors were controlled or excluded as follows:

1. The stacking of the workpieces: Using inclined stacking instead of horizontal placement, on the one hand, increases the furnace loading capacity and ensures production output; on the other hand, the use of stacking for arrangement changes the product from surface contact with the mesh belt to point contact, and the increased gap between the bearing and the mesh belt also results in more uniform heating.
2. Heating time and temperature: When a bearing is rapidly cooled from a high temperature, significant thermal stress may be generated. Therefore, the temperature in the rear heating zone of the continuous furnace is appropriately reduced, thereby improving product dimensional stability without affecting microstructure or hardness.
3. Mesh belt wandering in the quenching oil and oil stirring: These measures aim to increase the flow velocity around the bearing to shorten the vapor film duration; however, the resulting oil flow tends to make the distortion of thin-walled bearing rings more unstable.

Empirical observations indicate that axial flow and agitation increase the cooling rate, but may also lead to more uneven cooling, which in turn results in non-uniform expansion during phase transformation of the microstructure. The temperature corresponding to a moderate reverse frequency coincides with the martensite start (Ms) temperature range. Initiating the reversal process within this range promotes more uniform cooling between the upper and lower surfaces. Accordingly, during the experiments, interference from mesh belt wandering and stirring rate in the quenching oil was excluded to investigate the influence of the mesh belt's reverse motor frequency in the quenching oil on both the difference in average diameter between the upper and lower surfaces and the elliptical distortion of bearing rings after heat treatment.

2.1.2. Heat Treatment Process

After machining, bearing rings were loaded onto the furnace's mesh belt in the specified stacking configuration. The belt conveyed the rings through the heating zone, where they were austenitized at 830 °C and held for 40 min, followed by oil quenching. After quenching, the rings underwent washing and then low-temperature tempering at 170 °C for 120 min, as shown in Figure 2.

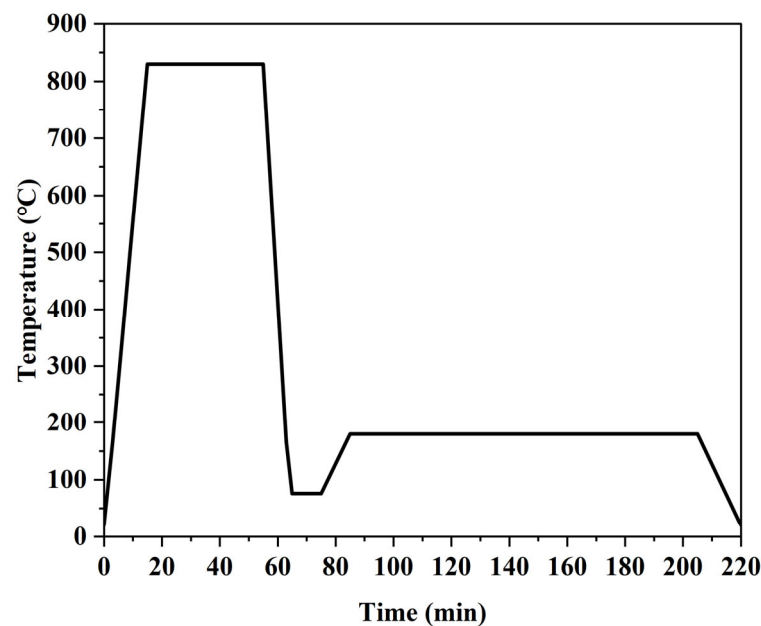


Figure 2. Heat treatment process flow.

2.1.3. Experimental Design

For each of the three bearing types (A, B, and C), quenching experiments were conducted at nine different reverse motor frequencies of the secondary mesh belt (20, 30, 40, 50, 60, 70, 80, 90, and 100 Hz). A total of 810 bearing rings were processed in the experiments, divided into 27 groups according to different frequencies and product types. The data for each group were evenly distributed to ensure representativeness and stability in statistical analysis.

The rings were inclined-stacked on the mesh belt, with 15 pieces per column and two columns per batch, for a total of 30 pieces. After passing through the heating zone, the bearings fell onto the secondary mesh belt in the quenching oil tank. With the movement of the secondary mesh belt, the bearings were flipped by a reversing baffle, then dropped onto the lower mesh belt. The lower belt carried the bearings out of the oil tank and into the next process step.

2.2. Machine Learning and LLM Method

2.2.1. Conventional Machine-Learning Models

In this study, the geometric attributes of the thin-walled bearing rings and the process parameters form the input vector x .

$$x = [h, D_{\text{out}}, D_{\text{in}}, w, f]^T \in \mathbb{R}^5 \quad (1)$$

where h denotes the ring height, D_{out} and D_{in} represent the outer and inner diameters, respectively, w is the wall thickness, and f refers to the secondary mesh-belt reverse motor frequency. The output consists of two distortion indices measured after quenching.

$$y = [T, R]^T \in \mathbb{R}^2 \quad (2)$$

where T denotes the difference in diameter values between the upper and lower end faces, and R denotes the oval distortion. The experimentally constructed sample space is given by $\mathcal{D}$.

$$\mathcal{D} = (x_i, y_i)_{i=1}^N \quad (3)$$

The objective of the distortion-prediction task is to learn a mapping from the dataset $\mathcal{D}$, as formulated in Equation (4).

$$\hat{y} = \phi(x) \quad (4)$$

Furthermore, for each ring geometry $g \in A, B, C$, the recommendation task aims to determine an optimal turnover-belt frequency f^* under the process constraints. Let $\mu_T(g, f)$ and $\mu_R(g, f)$ denote the difference in diameter values between the upper and lower end faces and the oval distortion obtained at geometry g and frequency f , respectively. Let $P_T(g, f)$ and $P_R(g, f)$ represent the corresponding pass rates under these metrics. The process optimization problem can therefore be formulated as the optimization task shown in Equation (5).

$$\begin{aligned} \min_{f \in \mathcal{F}} J_g(f) &= \mu_T(g, f) + \mu_R(g, f), \\ \text{s.t. } P_T(g, f) &\geq 0.9, P_R(g, f) \geq 0.9 \end{aligned} \quad (5)$$

Here, $\mathcal{F}$ denotes the admissible frequency set. The formulation implies that, under the requirement that the qualification rates for both the difference in diameter values between the upper and lower end faces and the oval distortion are not lower than 90%, the objective is to minimize the magnitudes of the two distortion metrics, thereby yielding the recommended frequency f^* with the smallest overall distortion.

In conventional machine learning, the prediction task in Equation (5) can be reformulated as an empirical risk minimization problem shown in Equation (6), which seeks a function $\phi \in \mathcal{H}$ that best fits the available experimental samples.

$$\min_{\phi \in \mathcal{H}} \frac{1}{N} \sum_{i=1}^N \text{Loss}(f(x_i), y_i) + \lambda \mathcal{R}(f) \quad (6)$$

Here, $\mathcal{H}$ denotes the hypothesis space, $\text{Loss}(\cdot)$ represents the loss function, and $\mathcal{R}$ is a regularization term. In the present study, the frequency variable contains only nine discrete levels, and merely three geometric types are available, resulting in a limited sample size. Under these conditions, the feasible functions satisfying Equation (5) are often non-unique, and the interpolated curves exhibit substantial variability between training points, leading to instability in the frequency–distortion relationship as well as in the optimal solution to the optimization problem in Equation (4).

2.2.2. LLM Prior for Small-Sample Process Regression

In the quenching process of the mesh-belt continuous furnace investigated in this study, the available data consist of only three geometries of thin-walled bearing rings, nine discrete reverse motor frequencies, and 27 groups of statistical distortion measurements derived from 810 parts. Both experimental observations and prior domain knowledge indicate that the frequency–distortion relationship is highly nonlinear. Under such a small-sample regime, conventional regression models tend to produce multiple interpolated curves that fit the limited samples equally well yet differ markedly in shape, resulting in predictions that are highly sensitive to model choice and hyperparameters. Moreover, these models struggle to explicitly incorporate process knowledge.

Motivated by these limitations, this work introduces a pretrained large language model (LLM) to serve as a prior-guided regressor. Geometric dimensions, frequencies, and historical distortion data are converted into structured textual representations, from which a small subset of representative samples is retrieved based on similarity. These samples, together with natural-language descriptions of process rules, are provided as prompts to the LLM. Under the dual constraints of few-shot prompting and embedded physical priors, the LLM yields more stable predictions that better reflect the underlying process

mechanisms, particularly for estimating the relationships between frequency and both the difference in diameter values between the upper and lower end faces and the oval distortion within sparse frequency intervals.

The use of a pretrained large language model in this work can be viewed as constructing a conditional probability distribution for the task, as formulated in Equation (7).

$$p_{\theta}(y | x, C(\cdot)) \quad (7)$$

where θ denotes the parameters learned during large-scale pretraining, and C represents the task instructions and domain knowledge injected through prompt engineering. In this study, θ remains fixed during inference and is not updated or fine-tuned on the present dataset.

Unlike directly training a regression network in the numerical feature space, this work leverages a pretrained LLM through a combination of few-shot prompting and physics-based constraints. Specifically, the experimental dataset $\mathcal{D}$ and the physical rules governing mesh-belt quenching are encoded as natural-language prompts and provided to the model as conditioning information, as formulated in Equation (8).

$$C_{phys}(x) = \{D, \text{domain rules}\} \quad (8)$$

Under the given input x and the physical context $C_{phys}(x)$, the LLM-based joint prediction of (T, R) can be reformulated as a probabilistic modeling task, as formulated in Equation (9).

$$\hat{y} = \text{LLM}_{\theta}(x) = \operatorname{argmax}_y p_{\theta}(y | x, C_{phys}(x)) \quad (9)$$

From a function-space perspective, the pretrained LLM together with the few-shot prompts defines a prior-constrained hypothesis set, as formulated in Equation (10).

$$\mathcal{H}_{\mathcal{LLM}} = \text{LLM}_{\theta}(\cdot; C_{phys}(x)) \quad (10)$$

Here, $\text{LLM}_{\theta}(\cdot; C_{phys}(x))$ denotes the family of functions that map the input x to the output y , induced by the pretrained language model under the fixed textual context $C_{phys}(x)$, where $\cdot$ represents the model input and θ is the frozen parameter set of the LLM. The context $C_{phys}(x)$ is constructed from a limited number of experimental samples together with the physics-based domain rules relevant to the quenching process. Under the same set of 27 experiments, the admissible function family $\mathcal{H}_{\mathcal{LLM}}$ is substantially smaller than the hypothesis space $\mathcal{H}$ of conventional regression models, which is typically spanned by flexible families such as polynomials or deep neural networks, and whose flexibility often leads to multiple disparate interpolants that all fit the sparse data. As a result, functions violating the physical mechanisms underlying the frequency–distortion relationship are implicitly excluded. On top of this, few-shot examples further guide the model toward function families that remain consistent with local experimental trends, yielding more stable and mechanically plausible interpolation across the unexplored frequency range.

After obtaining the predictions from the LLM, the process parameter optimization described in Section 2.2.1 can be decomposed into two sequential steps; the first is to identify a feasible frequency interval, and the second is to determine the optimal frequency within that interval. First, within the admissible frequency range $\mathcal{F}$, a set of candidate frequencies is constructed, as formulated in Equation (11).

$$\{f_m\}_{m=2}^M \subset \mathcal{F} \quad (11)$$

Using a step size of 2 Hz while including the nine experimentally tested frequencies, an input vector is constructed for each candidate frequency f_m based on the geometric parameters of the bearing x_m , as formulated in Equation (12).

$$x_m = [h, D_{out}, D_{in}, w, f_m]^T \quad (12)$$

This process yields estimates of the predicted mean values of the difference in diameter values between the upper and lower end faces and the oval distortion, denoted as $\hat{\mu}_T(g, f)$ and $\hat{\mu}_R(g, f)$, as well as the corresponding qualification probabilities $\hat{P}_T(g, f)$ and $\hat{P}_R(g, f)$. These estimates are then used to determine the optimal frequency based on the LLM-derived predictions.

From a probabilistic perspective, the above procedure for identifying the optimal frequency can be interpreted as decision-making based on the posterior distribution induced by the LLM. By treating the distortion outcome predicted by the LLM at a given frequency f as a random variable $Y_f \sim p_\theta(y|x_f, C_{phys})$, the optimal frequency f^* can be formally defined as the solution that minimizes the expected distortion risk subject to the prescribed probabilistic constraints, as formulated in Equation (13).

$$\begin{aligned} f^* = \underset{f \in f_m}{\operatorname{argmin}} \quad & (\hat{\mu}_T(g, f) + \hat{\mu}_R(g, f)) \\ \text{s.t.} \quad & \hat{P}_T(g, f) \geq 0.9, \hat{P}_R(g, f) \geq 0.9 \end{aligned} \quad (13)$$

where $\hat{\mu}(g, f) = E_{y \sim p_\theta(\cdot|x, C_{phys})}[y]$ denotes the mathematical expectation under the posterior distribution constructed by the LLM, characterizing the average distortion level at frequency f for geometry g . Likewise, $\hat{P}_R(g, f) = \int_\Omega p_\theta(y|x, C_{phys}) dy$ represents the confidence probability that the generated outcome falls within the acceptable region Ω .

2.2.3. Data Preprocessing

In this study, the pretrained large language model is used in a frozen manner, and no model training or parameter optimization is involved. The experimental data are processed differently for the distortion prediction task versus the process recommendation task.

For the distortion prediction task, a leave-one-out data partition strategy is applied. Specifically, for each prediction case, 26 experimental data points are used as the input context for the LLM, while the remaining one data point is withheld and used to evaluate the prediction performance.

For the process recommendation task, all available experimental data are provided to the LLM to support trend analysis and frequency selection. Based on the recommended frequencies, additional quenching experiments are conducted for validation. These experimental results are not included in the input data and are used solely to evaluate the effectiveness of the recommended process parameters.

This section addresses the distortion prediction and frequency selection problems in the quenching process of the mesh-belt continuous furnace, formulating them as a regression and optimization task under small-sample conditions. We first formalize the objective function and process constraints within a conventional machine-learning framework and analyze the limitations of interpolation-based regression when only sparse experimental samples are available. Building upon this, we introduce a pretrained large language model as a prior-guided regressor, in which few-shot prompting and embedded process knowledge are used to restrict the effective function space. This enables the model to produce more stable distortion predictions under data scarcity and provides data-driven support for selecting the optimal reverse motor frequency. For reproducibility, the exact prompt

templates for both prediction and recommendation tasks, numerical feature representations, and inference settings are provided in Supplementary Materials.

3. Results and Discussions

3.1. Experimental Results

During the experiments, bearing components underwent identical heating and quenching processes. The diameter difference between the upper and lower surfaces and the elliptical distortion were measured and recorded to assess the influence of varying reverse motor frequencies on distortion characteristics. Since distortion of individual components in production environments may be affected by random factors, mean values for each experimental group were calculated to improve the stability of the experimental data (as shown in Table 3), thereby characterizing the overall distortion trend under each condition.

Table 3. Diameter difference between upper and lower surfaces and elliptical distortion of three products at different frequencies.

Distortion Model		20	30	40	50	60	70	80	90	100
R (mm)	A	0.868	0.783	0.847	0.778	1.022	1.010	0.975	0.895	0.943
T (mm)	A	1.398	1.167	0.987	1.018	0.878	1.040	0.901	0.828	1.277
R (mm)	B	1.322	1.288	1.160	1.145	1.538	1.392	1.390	1.570	1.513
T (mm)	B	0.365	0.358	0.297	0.375	0.315	0.362	0.31	0.343	0.46
R (mm)	C	0.937	0.780	0.888	1.138	1.017	1.003	0.90	1.272	1.210
T (mm)	C	0.563	0.473	0.402	0.468	0.433	0.407	0.473	0.468	0.660

The experimental data indicate that the reverse motor frequency affects both the diameter difference between the upper and lower surfaces and the elliptical distortion of the bearings. Under different frequency conditions, the distortion characteristics of the three bearing types exhibited certain fluctuations. In certain frequency ranges, the magnitude of distortion was relatively low, suggesting that an appropriate reverse motor frequency helps to reduce overall distortion. Compared to low or high frequencies, a moderate frequency range (40–60 Hz) showed a trend of reduced distortion for some products, as shown in Figure 3, indicating that proper control of the reversing process can optimize cooling uniformity during quenching, thereby reducing distortion. However, the optimal reverse motor frequency varied among the bearing products, showing certain differences: Type A exhibited smaller distortion under specific frequencies but greater fluctuation at higher frequencies. Type B elliptical distortion was more sensitive to frequency changes, resulting in larger distortion at some frequencies. Type C demonstrated generally low distortion overall but still displayed a degree of frequency dependence, as shown in Figure 4

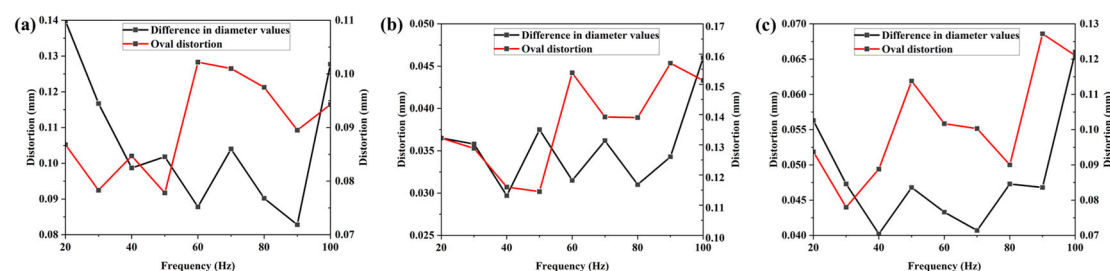


Figure 3. Distortion Results of Three Components at Different Frequencies. (a) Distortion Results of Type A; (b) Distortion Results of Type B; (c) Distortion Results of Type C.

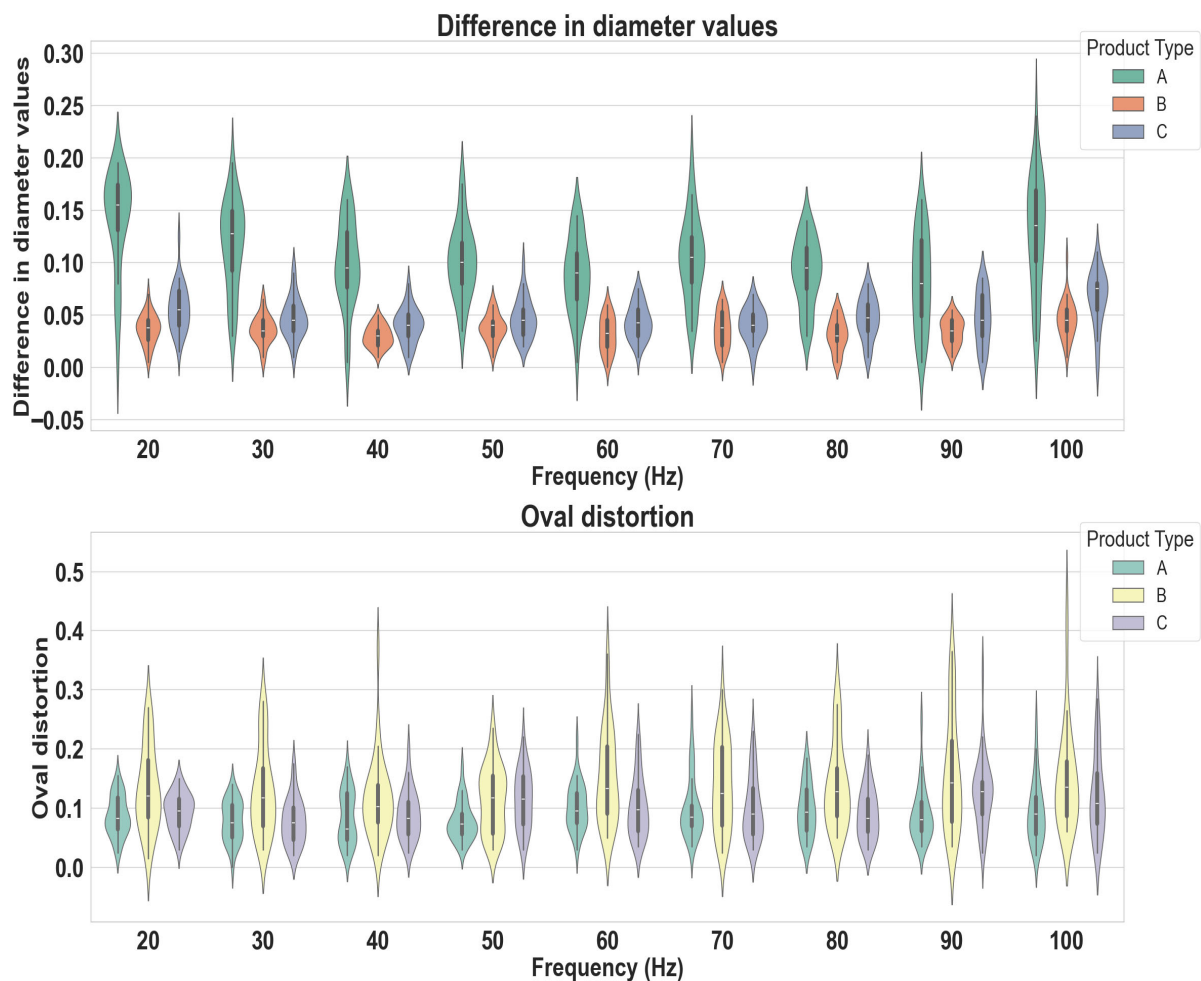


Figure 4. Distribution of the diameter difference between upper and lower surfaces and elliptical distortion of three components at different frequencies.

In summary, the findings highlight that selecting an appropriate reverse motor frequency is critical for controlling heat treatment distortion. Proper reverse motor parameters can improve cooling uniformity during quenching, thereby reducing distortion and enhancing product quality. Further data analysis could enable predictive modeling to recommend optimal reverse motor frequencies for different bearing geometries, providing valuable support for process optimization in heat treatment.

To explore the relationship between the mean diameter difference in the upper and lower surfaces and oval distortion after bearing heat treatment, a standardized deviation analysis method was employed, and a visualization analysis was performed on the distortion characteristics under different reverse motor frequencies. Specifically, measurements of mean diameter difference and oval distortion were obtained for three bearing types (A, B and C) under nine secondary mesh belt reverse motor frequencies, then standardized to eliminate the effect of scale differences across products and frequency conditions. Standardized taper deviation was plotted on the x-axis and standardized ovality on the y-axis, and scatter distribution plots were constructed for different reverse motor frequencies to visually illustrate the relationship between them, as shown in Figure 5.

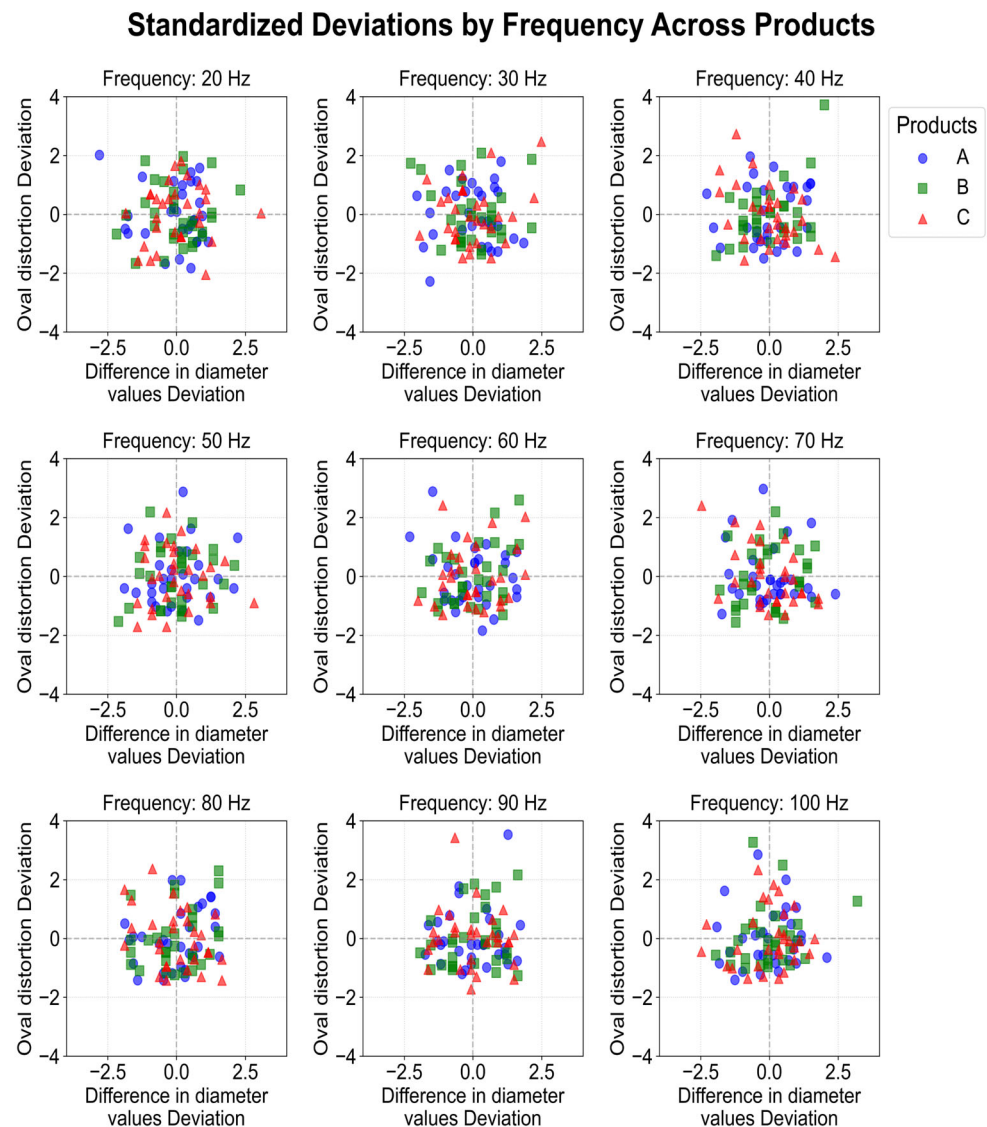


Figure 5. Visualization of distortion characteristics.

The analysis showed that, under all frequency conditions, the standardized deviations exhibited the characteristics of random distributions, with no evident linear or nonlinear correlation. While slight variations existed among products at different frequencies, the overall data distribution in two-dimensional space was uniform, lacking clear trends or clustering. Furthermore, under certain frequency conditions, individual data points deviated from the central region, but this deviation did not follow a consistent pattern across all products, further indicating no statistically significant relationship between the mean diameter difference and oval distortion.

Based on the above analysis, it can be inferred that the factors influencing the mean diameter difference and oval distortion may be largely independent, with the reverse motor frequency affecting them through different mechanisms, or their formation being dominated by other process parameters. Consequently, to further optimize distortion control in bearing heat treatment, it is necessary to establish independent models for the mean diameter difference and oval distortion, respectively, incorporating more process parameters for in-depth analysis to identify the key variables governing distortion characteristics.

3.2. AI-Driven Method and Accuracy Analysis

In this study, a database established through experimental results was used to develop a machine learning prediction model for the average diameter difference between the upper and lower surfaces (T) and the oval distortion (R) of bearing components to investigate the influence of reverse motor frequency and geometric parameters on distortion and further optimize process parameters. The dataset included key dimensional parameters of the bearing parts, taking height, outer diameter, inner diameter, wall thickness, and reverse motor frequency as input features, and the average diameter difference (T) and oval distortion (R) as target variables. Due to differences in the dimensional units of the features, all input variables were standardized to avoid the impact of feature scale on model training. To ensure generalization capability and make full use of the limited experimental data, Leave-One-Out (LOO) cross-validation was employed: each time, one sample was left out as the test set while the remaining samples were used for training. This process was repeated until all samples had been used once as the test set, thereby reducing overfitting risk and improving model robustness.

To comprehensively evaluate the predictive capability of different algorithms, five common regression models were selected for training and comparative analysis: linear regression (LR), random forest regression (RF), gradient boosting-based eXtreme Gradient Boosting (XGB) model, decision tree regression (DT), and support vector regression (SVR). The linear regression model was used to establish a baseline to assess the linear relationships between input and target variables. Decision tree regression can capture nonlinear characteristics in the data but is prone to overfitting. Random forest regression integrates multiple decision trees to reduce the variance of single-tree models and improve generalization. SVR uses kernel methods to find the optimal hyperplane in high-dimensional space, enabling more accurate nonlinear regression fitting. XGB, as an efficient regression method based on gradient boosted decision trees, offers strong generalization and computational efficiency in multiple predictive tasks. In addition, a Multi-Layer Perceptron (MLP) model was employed, which utilizes a multi-layer neural network architecture to capture complex nonlinear relationships and has been widely applied in recent years to process outcome prediction and materials property prediction.

3.2.1. Few-Shot LLM Prediction Method

To further explore the applicability of methods based on data-driven approaches to predicting bearing part distortion, a large language model (LLM) was utilized for the few-shot prediction. A domain-specific prompt was designed for the mesh-belt furnace quenching process to emulate expert reasoning in the field of materials science and engineering. This method combined data retrieval and knowledge reasoning to achieve the prediction of the average diameter difference (T) and oval distortion (R) under specific process conditions. Model performance was evaluated using the coefficient of determination (R^2), mean squared error (MSE), and mean absolute error (MAE); the results are shown in Figure 6.

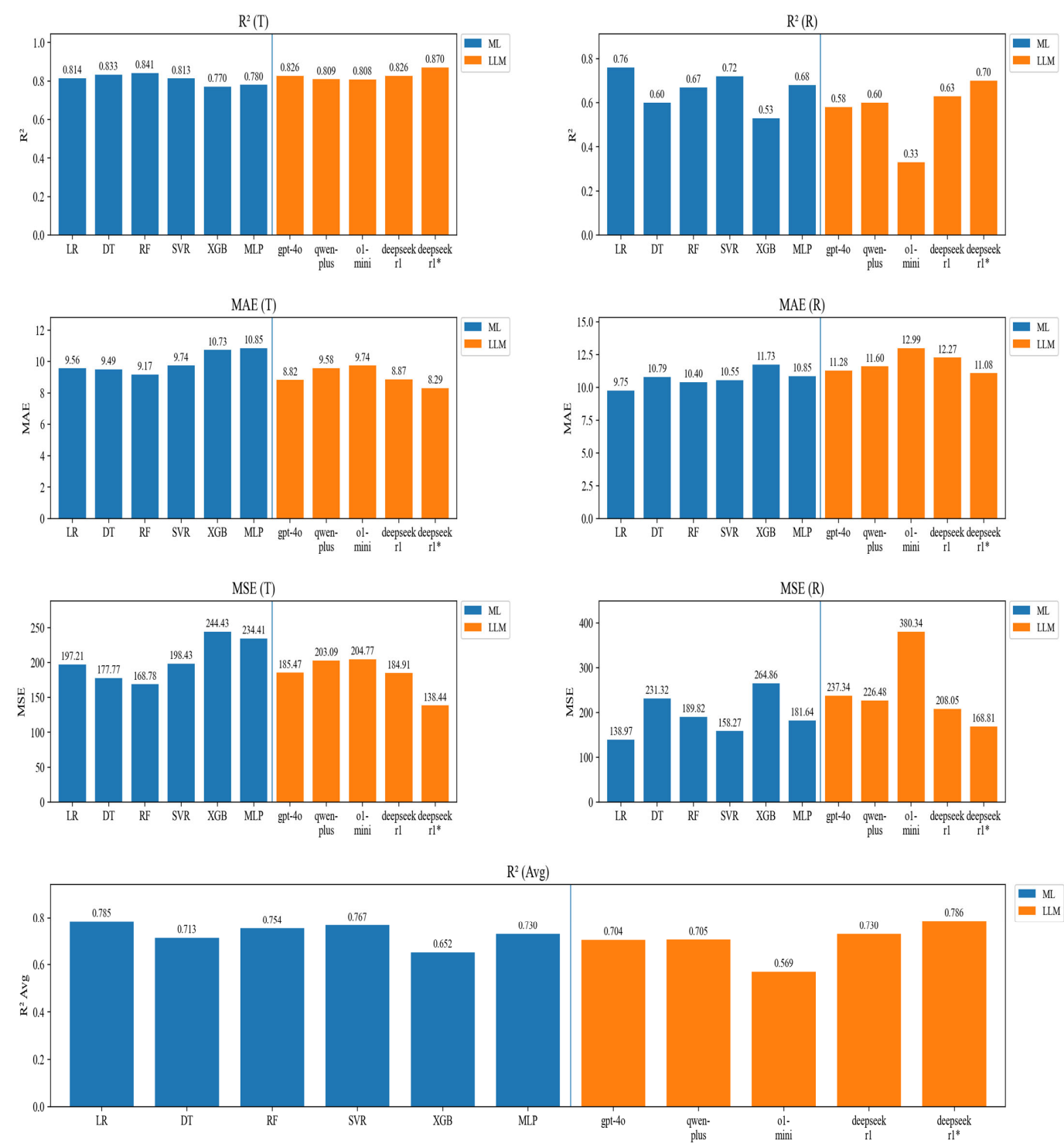


Figure 6. Comparison of predictive performance between machine learning and LLM methods.

The results indicate that in process optimization tasks, LLMs combined with domain knowledge in heat treatment can significantly enhance predictive capability. Future work will further explore how to efficiently integrate physical knowledge and experimental data to improve reliability and generalization. The LLM recommendation process is illustrated in Figure 7.

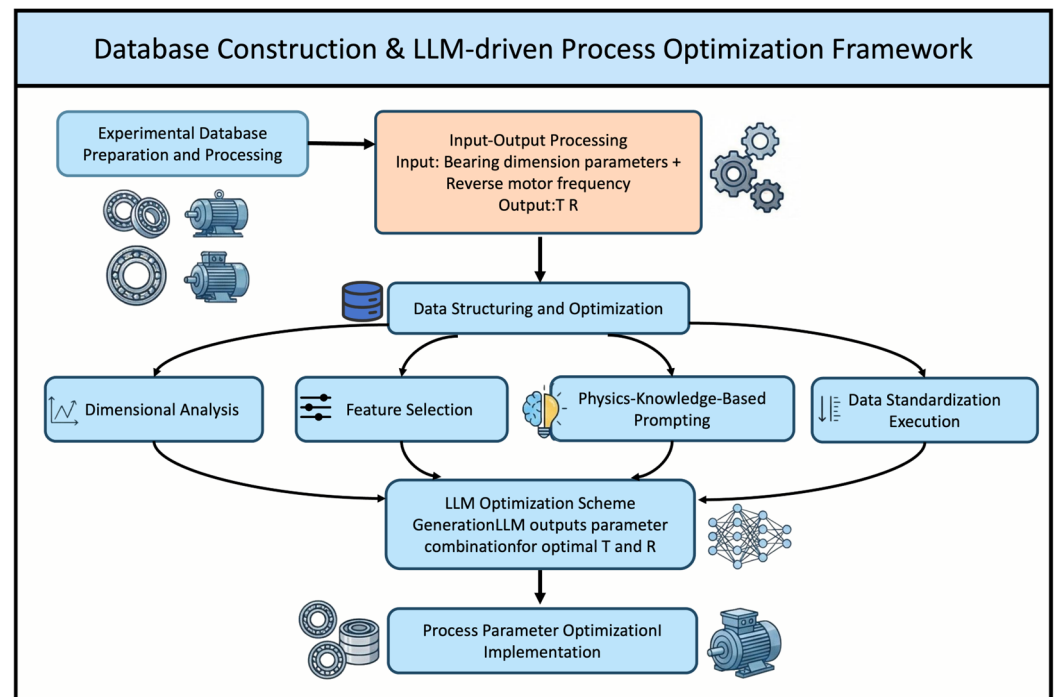


Figure 7. Recommendation process of LLM.

3.2.2. Prediction Results for Three Types of Thin-Walled Bearing Rings

In the preceding study, both machine learning and LLM few-shot methods were applied to distortion prediction tasks in bearing quenching processes. The experimental results show that LLM significantly outperforms traditional machine learning methods in prediction accuracy. Based on this, to further explore the application potential of LLMs in intelligent process optimization, the few-shot method was used to test the LLM on process recommendation tasks, followed by experimental verification of the recommended process schemes to assess their feasibility.

To improve the LLM understanding and reasoning capability for process optimization, a specialized few-shot prompt was constructed for the design of the reversing mesh belt in bearing quenching. This prompt combined data-driven analysis with knowledge reasoning, ensuring that the model could make reasonable process recommendations under limited data conditions. The core design of the prompt revolved around enhancing the model reasoning capacity: first, by assigning the model the role of a domain expert in materials science and engineering, guiding it to integrate data with physical knowledge; second, by emphasizing the influence of the “multi-stage flipping quenching” process on distortion, prompting it to fully consider the mechanism of this process when recommending the optimal reverse motor frequency. In terms of optimization objectives and constraints, the prompt explicitly required that the LLM recommend a reverse motor frequency that minimizes both taper and roundness distortion while maintaining a product qualification rate of no less than 96%. Furthermore, considering that experimental data only covered nine discrete frequency points from 20 Hz to 100 Hz, and that it is impractical to test all possible frequencies in actual production, the prompt instructed the LLM to first identify potential optimal intervals and then perform interpolation within these intervals, requiring it to derive more reasonable process parameters from the limited dataset. During data matching and trend analysis, the LLM was required to match bearings with similar shapes, analyze distortion trends under different reverse motor frequencies, deduce possible optimal frequencies, and make reasonable predictions beyond the existing experimental data points to improve accuracy and feasibility. Accordingly, a specialized few-shot prompt

was developed for process optimization tasks, using the DeepSeek-R1 model for process recommendation testing and experimental validation.

From the experimental results in Table 4, the LLM recommended process achieved certain optimizations in some cases but did not universally outperform experimental data across all bearing models. This indicates that although the LLM can recommend process parameters outside the experimental range, its interpolation and extrapolation capabilities may still have limitations, preventing precise matching to optimal processes in all cases. Overall, the LLM can optimize individual indicators in certain cases, but its recommended frequencies did not universally outperform experimental results for all bearing models. Distortion results from the coupled effects of thermal stress, phase transformation stress, and residual stress—a highly nonlinear physical process. Even for bearings from the same furnace batch, the distortion can vary under identical processing conditions due to subtle differences in material microstructure, which contributes to significant experimental scatter, suggesting further improvements are needed in process parameter optimization.

Table 4. Comparison between the LLM-Recommended process and the original optimal experimental group.

Product	Frequency (Hz)	Mean Ovality (μm)	Ovality Compliance Rate (%)	Ovality SD (μm)	Difference in Mean Diameter (μm)	Compliance Rate of Mean Diameter Difference (%)	Mean Diameter Difference SD (μm)
A	80	97.5	86.7	44.19	90.17	80	31.91
A1	78	88.2	86.7	55.27	100.8	66.67	42.40
B	50	116	60	55.03	37.5	93.33	13.05
B1	58	120	66.67	66.40	32.3	96.67	14.00
C	40	88.83	86.67	40.78	40.17	86.67	16.69
C1	35	98.50	76.67	50.1	46.5	86.67	17

4. Conclusions

This study measured and calculated the oval distortion and the difference between the average values of the upper and lower surfaces of bearings after quenching using a coordinate measuring machine, and established a database based on the calculation results. Through the AI-driven optimization method proposed in this study, based on the experimentally measured results, the optimal reverse motor frequency of the secondary mesh belt for minimizing distortion of bearing rings during reversing was determined. Mesh belt furnaces are used to produce various types of bearings. The method proposed in this study can reduce production costs and lower the experimental costs associated with launching new products.

This study proposes a data-driven method for distortion prediction and reverse motor frequency recommendation in mesh-belt furnace quenching, addressing the engineering challenge of analyzing the frequency-distortion relationship from discrete, small-sample data in the secondary quenching process of a continuous mesh-belt furnace. The method is built upon modeling and comparative analysis using multi-batch post-quenching distortion data, and introduces a pretrained large language model (LLM) to enable frequency–distortion relationship learning and process recommendation under small-sample conditions. Results show that the LLM can effectively integrate historical statistical information and process rules under data scarcity, delivering more stable trend judgments and interpretable recommendations. This work provides a referential technical pathway for nonlinear process optimization tasks characterized by high experimental costs and limited data availability. At the same time, relying solely on language-model outputs may still fall short of simultaneously meeting requirements for prediction accuracy and feasibility under

quality constraints. Future work will explore integrated solutions that combine physical mechanism modeling, numerical optimization, and simulation to enhance robustness and engineering usability under complex operating conditions.

Supplementary Materials: The following supporting information can be downloaded at: <https://www.mdpi.com/article/10.3390/pr14040718/s1>.

Author Contributions: Conceptualization, X.L. and D.J.; Methodology, X.L., X.D., J.W. and D.J.; Software, Y.S.; Validation, L.W.; Formal analysis, X.L. and J.W.; Investigation, L.W. and H.C.; Resources, Y.J.; Data curation, X.L., Y.S. and X.D.; Writing—original draft, X.L. and Y.S.; Writing—review & editing, X.L. and D.J.; Visualization, H.C.; Supervision, Y.J.; Project administration, M.W. and D.J.; Funding acquisition, M.W. and D.J. All authors have read and agreed to the published version of the manuscript.

Funding: This study was conducted as an internal project of Zhejiang XCC Group Co., Ltd., and was carried out as part of our daily operations. No external funding sources are applicable.

Data Availability Statement: The original contributions presented in this study are included in the article/Supplementary Materials. Further inquiries can be directed to the corresponding author.

Acknowledgments: Our heartfelt thanks to the Heat Treatment Plant for providing experimental materials and equipment.

Conflicts of Interest: Authors Xusheng Li, Mingzhou Wang, Lingchu Wang, Hao Chen and Dongying Juwere employed by the company Zhejiang XCC Group Co., Ltd. The remaining authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest. The authors declare no conflict of interest.

References

1. Sun, G.; Wang, Q.; Li, Z. Microstructural evolution of high-nickel steel during quenching, lamellarizing, and tempering heat treatment. *J. Mater. Res. Technol.* **2024**, *33*, 8618–8630. [\[CrossRef\]](#)
2. Zhang, X.; Tang, J.; Zhang, X. An optimized hardness model for carburizing-quenching of low carbon alloy steel. *J. Cent. S. Univ.* **2017**, *24*, 9–16. [\[CrossRef\]](#)
3. Giordani, T.; Clarke, T.R.; Kwietniewski, C.E.F.; Aronov, M.A.; Kobasko, N.I.; Totten, G.E. Mechanical and metallurgical evaluation of carburized, conventionally and intensively quenched steels. *J. Mater. Eng. Perform.* **2013**, *22*, 2304–2313. [\[CrossRef\]](#)
4. Su, S.; Song, R.; Chen, C.; Wang, J.; Zhang, Y. The novel process of spheroidizing-critical annealing used to optimize the properties of carburized steel and its effect on hardening mechanism of quenching and tempering. *Mater. Sci. Eng. A* **2019**, *765*, 138322. [\[CrossRef\]](#)
5. Ferguson, B.; Li, Z.; Freborg, A. Modeling heat treatment of steel parts. *Comput. Mater. Sci.* **2005**, *34*, 274–281. [\[CrossRef\]](#)
6. Jiang, B.; Mei, Z.; Zhou, L.; Liu, G.; Wang, Z.; Huang, B.; Liu, Y. High toughness and multiphase microstructure transition product of carburizing steel by a novel heat treatment cooling process. *Mater. Sci. Eng. A* **2016**, *675*, 361–370. [\[CrossRef\]](#)
7. Lingamanaik, S.; Chen, B. The effects of carburising and quenching process on the formation of residual stresses in automotive gears. *Comput. Mater. Sci.* **2012**, *62*, 99–104. [\[CrossRef\]](#)
8. Tewary, U.; Mohapatra, G.; Sahay, S. Distortion mechanisms during carburizing and quenching in a transmission shaft. *J. Mater. Eng. Perform.* **2017**, *26*, 4890–4901. [\[CrossRef\]](#)
9. Atraszkiewicz, R.; Januszewicz, B.; Stachurski, W.; Dybowski, K.; Rzepkowski, A. High pressure gas quenching: Distortion analysis in gears after heat treatment. *Mater. Sci. Eng. A* **2012**, *558*, 550–557. [\[CrossRef\]](#)
10. Samuel, A.; Prabhu, K.N. Residual stress and distortion during quench hardening of steels: A review. *J. Mater. Eng. Perform.* **2022**, *31*, 5161–5188. [\[CrossRef\]](#)
11. Cocaño, P.; Fernández, P.; Muñoz-Calvente, M. Residual stress removal in rings: An analytical model for predicting displacement and stress from machining of bearings. *Results Eng.* **2025**, *26*, 104610. [\[CrossRef\]](#)
12. Lübben, T.; Frerichs, F. Influence of Rewetting Behavior on the Distortion of Bearing Races. *J. Mater. Eng. Perform.* **2013**, *22*, 2323–2330. [\[CrossRef\]](#)
13. Kusano, M.; Miyazaki, S.; Watanabe, M.; Kishimoto, S.; Bulgarevich, D.; Ono, Y.; Yumoto, A. Tensile properties prediction by multiple linear regression analysis for selective laser melted and post heat-treated Ti-6Al-4V with microstructural quantification. *Mater. Sci. Eng. A* **2020**, *787*, 139549. [\[CrossRef\]](#)

14. Hernandez, R.C.; Flores, J.L.M.; Cervera, C.M.; Olivos, P. Comparative Analysis of Factorial Analysis-Multiple Regression and Random Forest for the Prediction of the Stabilization Time in Furnaces in the Heat Treatment Area in a Metalworking Company. *Contemp. Math.* **2025**, *6*, 307–329. [[CrossRef](#)]
15. Zhu, D.; Wang, B.; Zhao, H.; Wu, S.; Li, F.; Huang, S.; Wu, H.; Wang, S.; Zhang, C.; Gao, J.; et al. Enhanced hardenability prediction in 20CrMo special steel via XGBoost model. *J. Iron Steel Res. Int.* **2025**, *32*, 1023–1033. [[CrossRef](#)]
16. Wang, Y.; Zhang, L.; Su, Y.; Liu, H.; Yang, H.; Chen, Y. Heat treatment control technology of high-strength steel gears based on support vector machine. *Sci. Rep.* **2025**, *15*, 7657. [[CrossRef](#)]
17. Nandal, V.; Dieb, S.; Bulgarevich, D.S.; Osada, T.; Koyama, T.; Minamoto, S.; Demura, M. Artificial intelligence inspired design of non-isothermal aging for γ - γ' two-phase, Ni–Al alloys. *Sci. Rep.* **2023**, *13*, 12660. [[CrossRef](#)]
18. Chintakindi, S.; Alsamhan, A.; Abidi, M.H.; Kumar, M. Annealing of monel 400 alloy using principal component analysis, hyper-parameter optimization, machine learning techniques, and multi-objective particle swarm optimization. *Int. J. Comput. Intell. Syst.* **2022**, *15*, 18. [[CrossRef](#)]
19. Xia, T.; Chen, Y.; Song, T.; Cui, P.; Liu, Y.; Zhu, J. Optimization of Heat Treatment Process Parameters for 8Cr4Mo4V Bearing Ring Using FEA-NN- PSO Method. *Met. Mater. Int.* **2025**, *31*, 2776–2796. [[CrossRef](#)]
20. Oh, S.; Ki, H. Deep learning model for predicting hardness distribution in laser heat treatment of AISI H13 tool steel. *Appl. Therm. Eng.* **2019**, *153*, 583–595. [[CrossRef](#)]
21. Jia, H.; Ju, D.; Cao, J. Machine learning based optimization method for vacuum carburizing process and its application. *J. Mater. Inform.* **2023**, *3*, 9. [[CrossRef](#)]
22. Wang, J.; Yang, S.; Li, J.; Ju, D.; Li, X.; He, F.; Li, H.; Chen, Y. Mathematical simulation and experimental verification of carburizing quenching process based on multi-field coupling. *Coatings* **2021**, *11*, 1132. [[CrossRef](#)]
23. Sun, Y.; Li, X.; Liu, C.; Deng, X.; Zhang, W.; Wang, J.; Zhang, Z.; Wen, T.; Song, T.; Ju, D. Development of an intelligent design and simulation aid system for heat treatment processes based on LLM. *Mater. Des.* **2024**, *248*, 113506. [[CrossRef](#)]

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.