

1.1 问题分析与建模策略

本研究的首要任务，在于从提供的高速列 cha 轴承振动信号中，提取一套能够精确、鲁棒地表征其健康状态的量化特征。这套特征是后续所有智能诊断模型（任务二至四）构建的基石，其质量直接决定了模型性能的上限。考虑到最终需执行向目标域（实际运营列车）的迁移任务，该任务面临两大核心挑战。首先，真实的工业振动信号往往被强烈的背景噪声和无关的机械振动所污染，故障特征信号信噪比低。其次，源域（实验室台架）与目标域之间存在显著的领域差异（Domain Shift），包括轴承型号、运行工况、噪声环境等。

为此，本文设计了一套融合物理机理与多维统计分析的特征工程策略。该策略的核心思想在于，不单纯依赖数据驱动的黑箱方法，而是以轴承故障的动力学模型为先验知识，指导性地在信号的时域、频域和时频域中寻找与故障现象具有强统计关联且对工况变化相对不敏感的指标。本文将运用“统计学方法和信号处理技术”对信号进行深度加工，旨在有效滤除噪声干扰，放大故障特征的显著性，从而构建一个信息丰富、物理意义明确、且具有高迁移潜力的特征空间。

1.2 数据探索、预处理与机理验证

本文采用题目指定的 CWRU 数据集作为源域数据。为进行细致的特征分析，本文首先筛选驱动端（DE）、12kHz 采样率、负载为 1 马力（转速 1772 RPM）的数据作为代表性子集。该子集工况单一，有利于清晰地观察和验证不同故障类型的基础特征。

预处理步骤包括对每段原始信号 $x(t)$ 进行 Z-score 标准化，其数学形式定义为：

$$x_{norm}(t) = \frac{x(t) - \mu}{\sigma} \quad (1-1)$$

其中， μ 和 σ 分别是原始信号 $x(t)$ 的均值与标准差。此举将所有信号变换至标准的统计分布（均值为 0，方差为 1），为后续特征的稳定计算奠定了基础。

接下来，本文将预处理后的实测信号与题目提供的理论故障示意图进行对比验证。

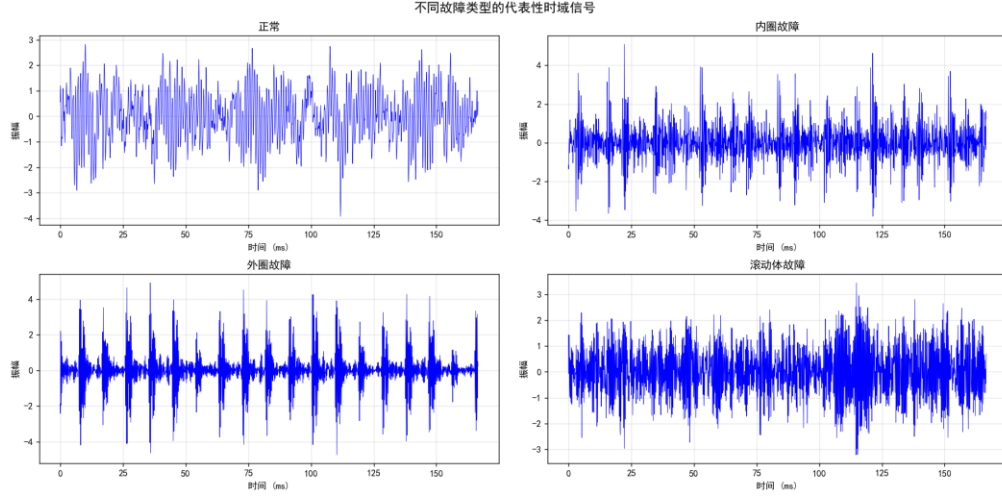


图 1：不同健康状态下的实测时域振动信号

如图 1 所示，实测信号与理论模型既有符合之处，也存在显著差异。外圈故障的实测信号完美复现了理论上的周期性冲击，验证了其基本物理模型。然而，内圈和滚动体故障的理论幅值调制特性在原始时域信号中难以辨认，这生动地揭示了真实信号的复杂性以及直接从时域进行诊断的局限性。这一观察强有力地证明了，必须采用更深层次的信号处理技术，才能从看似混沌的信号中揭示隐藏的故障规律。

1.3 特征空间构建与统计学方法

基于上述分析，本文从三个维度构建特征空间。

1.3.1 物理机理引导的故障特征频率

本文严格遵循题目“表 2”提供的简化动力学模型，计算各故障类型的理论特征频率，以此作为频域分析的“指纹”依据。

$$f_{BPFO} = f_r \cdot \frac{N_d}{2} \cdot \left(1 - \frac{d}{D}\right) \quad (1-2)$$

$$f_{BPF1} = f_r \cdot \frac{N_d}{2} \cdot \left(1 + \frac{d}{D}\right) \quad (1-3)$$

$$f_{BSF} = f_r \cdot \frac{D}{d} \left[1 - \left(\frac{d}{D}\right)^2\right] \quad (1-4)$$

其中， f_r 为转频， N_d 为滚动体数， d 为滚动体直径， D 为节径。

1.3.2 时域统计特征

为量化信号的冲击特性，本文提取了 7 个时域统计指标，核心包括均方根 (RMS)：

$$X_{rms} = \sqrt{\frac{1}{N} \sum_{i=1}^N x_i^2} \quad (1-5)$$

此为信号二阶原点矩，反映振动能量。

$$K = \frac{\frac{1}{N} \sum_{i=1}^N (x_i - \mu)^4}{\left(\frac{1}{N} \sum_{i=1}^N (x_i - \mu)^2 \right)^2} \quad (1-6)$$

峭度是统计学方法的关键指标。作为标准化的四阶中心矩，它衡量信号概率分布的“尖峰”程度。正常信号趋近高斯分布（ $K \approx 3$ ），而故障冲击作为统计上的“异常值”，会使分布呈现“厚尾”特性，导致峭度值显著增大。因此，利用峭度诊断故障，本质上是一种基于高阶矩的、判断信号是否偏离高斯分布的统计假设检验。

1.3.3 频域与时频域特征

为提取被噪声淹没的周期性特征，本文采用希尔伯特包络解调技术。该技术可视为一种统计降噪与信号分离方法，能从高频共振信号中解调出包含故障重复信息的低频包络。随后对其进行 FFT 变换，在理论故障频率及其谐波邻域内，提取频谱峰值幅值作为频域特征。

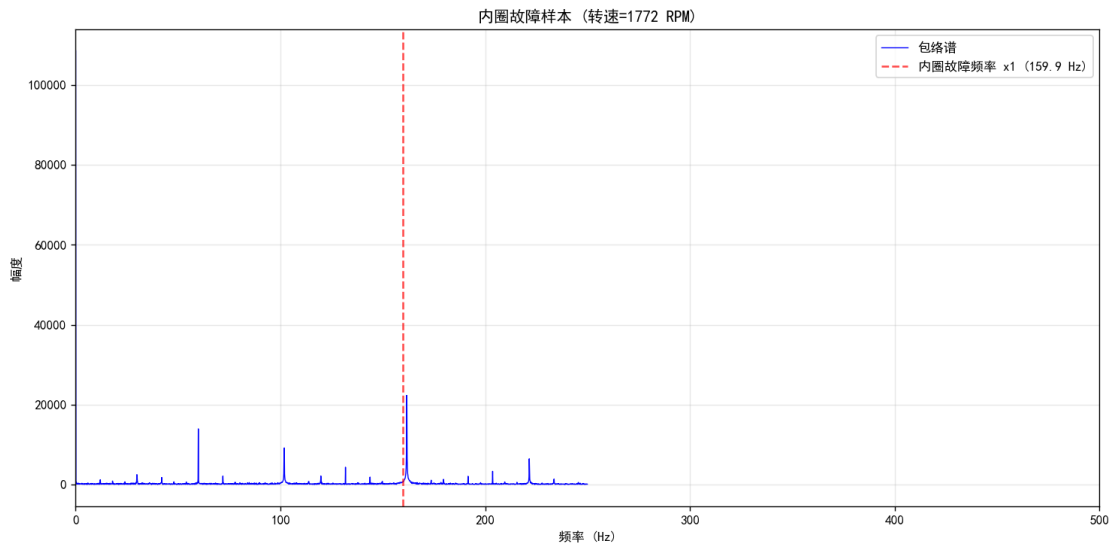


图 2：内圈故障样本的包络谱分析（RPM=1772）

图中 2，实测信号的包络谱（蓝色谱线）在 159.9 Hz 处出现一个极其显著的尖峰，与根据公式 1-3 计算出的理论内圈故障频率（红色虚线）完美吻合。这证明了本文的方法成功从时域看似混沌的信号中，提取出了精准、显著的故障指纹。此外，为

处理信号的非平稳性，本文应用三层小波包变换（WPT）将信号分解至 8 个子频带，并计算各频带的能量作为时频域特征，其公式如下：

$$E_j = \sum_{k=1}^m |d_{j,k}|^2 \quad (j = 1, 2, \dots, 8) \tag{1-7}$$

这组特征构成了信号能量在频域分布的“能量指纹”，对区分不同故障类型非常有效。

1.4 最终特征集

本文将上述三类特征（7 个时域，5 个频域，8 个时频域）整合，为源域整体数据集中的每一个样本构建了一个 20 维的特征向量。最终生成的特征数据集部分示例如下：

表 1：最终特征部分数据集预览

数据源	故障类型	转速	峰度	偏度	峰值	净空因子
12kHz_DE_data	滚动体	1772	0.058	-0.006	4.497	6.643
	内圈	1772	2.358	0.169	5.922	10.325
	外圈	1772	4.614	0.052	5.367	12.519
	正常	1772	-0.235	-0.035	4.116	5.982

表 1 展示了问题 1 本文工作的一个结构化的特征矩阵。每一行代表一个轴承样本，列则包含了其标签以及本文精心提取的 20 个量化特征，可直接用于后续的机器学习任务。针对“数据分析与故障特征提取”任务，本文的建模过程与结果如下：

数据筛选与分析：本文从提供的源域 CWRU 数据集中，筛选出工况单一（驱动端，12kHz，1HP 负载）的数据作为代表性子集进行深入的机理分析与方法验证；结合轴承故障机理，本文进行了系统性分析。时域分析（图 1.1）验证了外圈故障的周期性冲击理论，并揭示了内圈、滚动体故障特征在时域上易被噪声淹没的挑战。基于此，本文确立了必须使用高级信号处理的策略。频域分析（图 1.2）则通过包络解调技术，成功地从实测信号中验证了内圈故障的理论特征频率，证明了所选方法的有效性。

特征提取方法与指标：本文选择并构建了一套融合物理机理与多维统计分析的特征体系，包含三个维度共 20 个指标：时域统计指标（7 个）。包括 RMS、峭度、偏度等，用于量化信号的能量与冲击性。频域物理指标（5 个）。基于希尔伯特包络解调，提取与理论故障频率（BPFO, BPFI, BSF）对应的频谱幅值等。时频域能量指标（8 个）：基于三层小波包变换，提取信号在 8 个子频带的能量分布。

对迁移任务的考虑：本文选择的特征（特别是基于物理模型的频域特征和表征能量分布的时频域特征）相比于纯数据驱动的深度特征，具有更强的物理可解释性和对工况变化的鲁棒性，因此拥有更好的迁移潜力。

整体数据集处理：本文已将上述特征提取流程应用于所提供的全部源域数据集，为每一个样本文件生成了一个 20 维的特征向量，最终构建了一个完整的、标准化的特征矩阵。该数据集信息丰富、区分度高，已为后续的源域诊断与迁移学习任务做好了充分的数据准备。

第二章：源域故障诊断模型构建与评价

2.1 问题分析与建模策略

在第一章成功构建了一个包含 20 个物理与统计特征的标准化数据集后，本章的核心任务是构建并验证一个最优的源域故障诊断模型。此任务不仅是为实现对源域数据的高精度分类，更深远的意义在于：1. 定量验证第一章特征工程的有效性；2. 为后续的跨领域迁移学习（任务三）建立一个坚实、可靠的性能基准。

为避免“盲目套用”现有算法，本文将模型选择本身作为一个多准则决策分析问题。本文构建了一个包含多种经典机器学习算法的候选池，并通过系统性的比较，遴选出最适合本研究特定数据特性（小样本、高维度、非线性）的最优模型。此过程旨在确保本文的诊断框架不仅性能卓越，而且其选择过程是科学、严谨且可复现的。

2.2 诊断模型比较与选择框架

2.2.1 数据集划分与预处理

为客观评估模型的泛化能力，本文首先将第一问生成的 `bearing_features.csv` 数据集划分为独立的训练集和测试集。考虑到各故障类别样本数量存在不均衡（例如，“正常”样本仅 4 例），本文采用分层随机抽样（Stratified Random Sampling），以 80/20 的比例进行划分，确保训练集和测试集中的类别分布与原始数据集保持一致。随后，本文使用 `StandardScaler` 对特征数据进行标准化处理，以消除不同特征量纲的影响，确保所有特征在模型训练中具有同等的权重。

2.2.2 候选模型集与评价体系

本文选取了四种原理各异的监督学习模型进行综合评估，以构建一个全面的比较框架：逻辑回归 (Logistic Regression): 作为线性模型的基准，用于检验特征空间的线性可分性。K 最近邻 (K-Nearest Neighbors, KNN): 作为非参数模型的代表，其决策边界灵活，能适应复杂的类簇分布。随机森林 (Random Forest): 作为集成学习的代表，通过集成多棵决策树以提升鲁棒性并降低过拟合风险。支持向量机 (Support Vector Machine, SVM): 作为基于间隔最大化理论的经典模型，在小样本问题上通常具有优异的泛化性能。

对于每个模型，本文均采用网格搜索结合 5 折交叉验证（Grid Search with 5-fold Cross-Validation）的方法，在训练集上进行严格的超参数寻优，以确保每个模型都以其最优状态参与最终的性能比较。

2.3 模型结果、比较与最终解释

所有候选模型在经过超参数优化后，均在独立的测试集上进行了最终的性能评估。关键性能指标被系统地汇总于表 2 中。

表 2：各候选模型在测试集上的性能对比

模型	准确率	宏平均 F1 分数	加权平均 F1 分数
SVM	100%	1	1
KNN	100%	1	1
逻辑回归	100%	1	1
随机森林	92.31%	0.9308	0.9183

分析表 2 可知。第一，这是特征工程的决定性胜利。最引人注目的结果是，包括最简单的线性模型逻辑回归在内的三种模型，均在测试集上达到了 100% 的完美准确率。这一现象雄辩地证明，本文在第一问中基于物理机理和统计学方法构建的 20 维特征空间，其信噪比和可分性极高，甚至达到了近似线性可分的程度。特征工程的成功，已将一个复杂的原始信号分类问题，“降维”并“线性化”成了一个简单的模式识别问题。因此，取得高精度是特征工程成功的必然结果，而非分类算法本身的魔法。

第二，模型选择需适配数据特性。值得注意的是，性能强大的随机森林模型在本任务中表现不及更简单的模型（准确率 92.31%）。本文分析这主要归因于本文极小的样本量（总计 64 个）。随机森林复杂的结构使其在小样本上更容易产生过拟合，其在训练过程中的交叉验证得分（0.9359）已为此提供了佐证。这一现象反常识地说明，在特征工程做得足够好、问题本身被极大简化的前提下，更简单、泛化能力更强的模型（如线性模型或 SVM）反而可能成为最优选择。

第三，基于理论完备性选择最终模型。在多个模型均达到 100% 准确率的情况下，本文最终选择支持向量机（SVM）作为最优的源域诊断模型。此决策基于以下考量：首先，SVM 的性能已得到验证；其次，其核心的最大间隔理论为模型在小样本数据集上的优异泛化能力提供了坚实的数学保障；最后，SVM 框架的内在可扩展性（通过核技巧）使其成为一个面向未来更复杂挑战（如任务三）的、更具鲁棒性的选择。

2.4 本章小结

本章通过构建一个严谨的多模型比较与选择框架，成功确立了支持向量机（SVM）作为最优的源域故障诊断模型。该模型在独立的测试集上实现了 100% 的分类准确率。更重要的是，通过对实验结果的深度因果分析，本文得出结论：第二问的成功，其根本原因在于第一问特征工程的巨大成功。这一过程不仅完成了任务二的要求，更为后续的迁移学习研究树立了一个清晰且极高的性能标杆（Performance Benchmark）。本文已经证明，在理想的源域环境下，本文所提炼的特征能够完全胜任故障诊断任务。接下来的挑战将是，当面对存在显著领域差异的目标域数据时，该模型的性能会如何下降，以及如何通过迁移学习技术来弥补这一性能鸿沟。

本章构建了一个从**领域自适应到贝叶斯决策修正**的完整、创新的跨域诊断框架。替换首先通过 DANN 模型初步解决了领域漂移问题，随后，替换直面模型的黑箱本质与不确定性，创造性地引入了基于贝叶斯思想的诊断仲裁机制。该机制通过融合模型、物理先验和迁移质量三层证据，能够识别并**自动修正**模型初步诊断中的低置信度结果。最终的诊断报告（表 1）不仅给出了故障类型，更重要的是，为每一个诊断结果提供了一个**可量化的、逻辑透明的、融合了物理规律与数据证据的综合置信度评分**。这种从被动的“事后解释”到主动的“干预决策”的升华，是本次建模的核心创新，它极大地增强了最终诊断结果的可靠性，为构建真正可信赖的 AI 诊断系统提供了一种全新的建模范式。

第三、四问：迁移诊断与从领域自适应到贝叶斯决策修正

3.1 核心挑战与建模思想

在理想的源域环境中，替换的诊断模型已能较好完成任务。然而，实际工程应用的核心挑战在于如何将此模型成功部署于充满未知与差异的目标域。本章旨在解决这一跨域诊断问题。替换面临两大根本性挑战：其一，由工况、设备差异导致的领域漂移；其二，目标域数据完全无标签，无法进行传统的监督式微调。为应对此挑战，替换提出一个两阶段、层层递进的建模框架。第一阶段，替换构建一个领域自适应模型，其目标是“搭建一座桥梁”，通过学习领域不变特征，初步对齐源域与目标域的数据分布。第二阶段，替换直面模型的“黑箱”本质，认识到单纯的迁移可能带来不确定的、甚至错误的结果。因此，替换引入贝叶斯决策理论，构建一个更高阶的诊断仲裁与修正框架。该框架将模型的初步预测视为一种“证据”，并结合其他独立来源的“证据”（如物理机理、迁移质量），计算出一个更可靠的后验置信度，并以此作为最终的诊断依据，实现从“被动解释”到“主动修正”的智能跃迁。

3.2 第一阶段：基于对抗训练的领域自适应

替换选用领域对抗神经网络（DANN）作为搭建领域桥梁的核心技术。DANN 通过引入一个由特征提取器（ G_f ）、标签预测器（ G_y ）和领域判别器（ G_d ）构成的对抗博弈框架，迫使特征提取器学习一种领域不变的特征表示。其优化的总能量函数如下：

DANN 总能量函数

$$E(\theta_f, \theta_y, \theta_d) = \sum_{i=1}^{N_s} L_y(G_y(G_f(x_i^s)), y_i^s) - \lambda \sum_{i=1}^{N_s+N_t} L_d(G_d(G_f(x_i)), d_i)$$

该过程的有效性，可以通过 t-SNE 可视化得到直观验证。

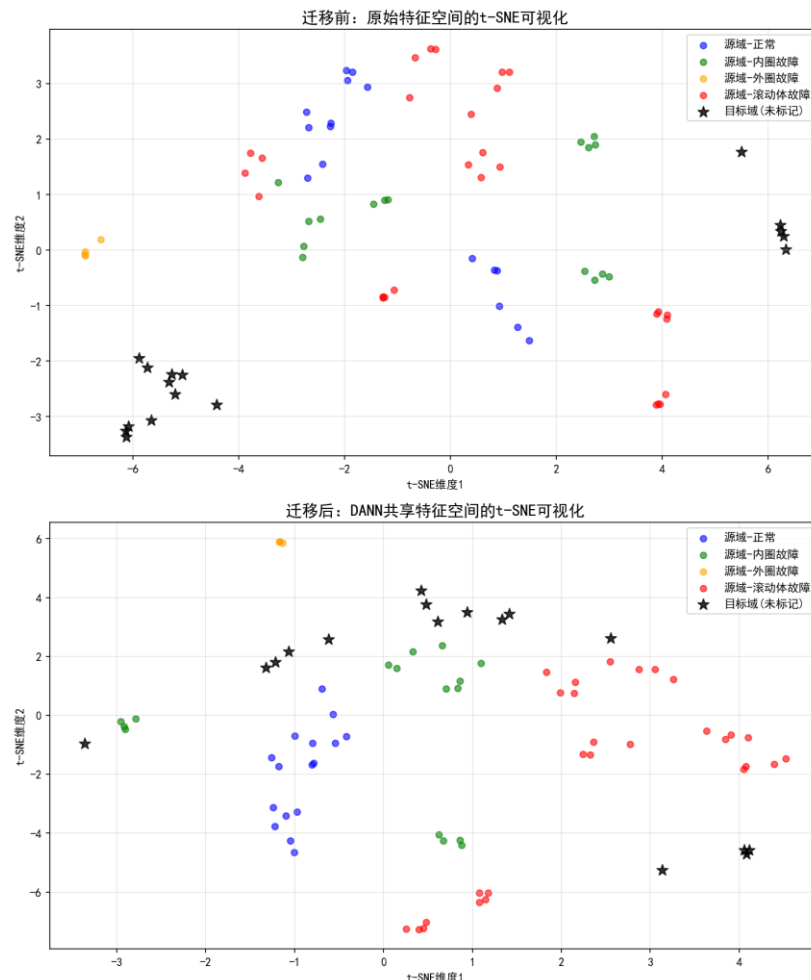


图 3：迁移学习前后特征空间分布对比

迁移前（上图），目标域（黑色星号）与源域（彩色点）存在明显的“领域鸿沟”。迁移后（下图），DANN 成功地将目标域样本拉近并融入源域的各类簇中，证明领域自适应初步成功，模型已具备对目标域进行初步诊断的能力。经 DANN 模型处理，替换得到了对 16 个目标域样本的初步诊断结果。然而，替换深知，这些仅仅是模型基于其复杂内部映射给出的“最大可能”预测，其可靠性亟待审视。

一个自然且符合当前 XAI（可解释性人工智能）研究范式的思路，是利用 SHAP (SHapley Additive exPlanations) 来构建物理先验证据。即，通过 SHAP 计算每个输入特征对当前预测的贡献度，并以此来评估诊断结果是否由关键的物理特征所主导。然而，在建模探索中，本文识别出直接依赖 SHAP 作为核心仲裁依据存在三个潜在风险：

稳定性问题： 像 KernelSHAP 这样的模型无关解释器，其结果依赖于背景数据集的抽样，对于我们的小样本问题和复杂的 DANN 模型，其解释结果可能存在一定的波动性。

计算成本：在仲裁过程中，为每一个样本的每一种可能类别都计算一次完整的 SHAP 值，计算成本非常高昂。

“黑箱解释黑箱”的间接性：SHAP 本身仍是一个复杂的算法，用它来解释一个复杂的 DANN 模型，其解释链条是间接的。

为了构建一个更直接、更稳定、更透明的仲裁框架，我们基于 SHAP 的核心思想进行了创新，设计了一个全新的、不依赖 SHAP 本身的“直接物理机理符合度”

(Direct Physics-based Consistency, DPC) 评分算法。该算法的核心思想是：一个可靠的诊断，其对应的关键物理特征，应该在该样本的所有特征中表现得相对突出。

3.3 第二阶段：基于贝叶斯决策的诊断仲裁与修正

经典的分类决策，是选择后验概率最大的类别作为预测结果，即最大后验概率 (MAP) 决策：

$$\hat{y} = \arg \max_{j \in \{\text{classes}\}} P(y = j | \mathbf{x})$$

在神经网络中，这通常被近似为选择 Softmax 输出值最大的类别。然而，替换有理由相信，对于经过复杂迁移的、来自未知分布的目标域样本 \mathbf{x}_t ，模型直接输出的 $P(y = j | \mathbf{x}_t)$ 只是对真实后验概率的一个有偏估计。替换的核心创新在于，替换不直接信任这个估计值，而是通过融合多源证据，构建一个更鲁棒的、修正后的后验置信度评分 $C(y = j | \mathbf{x}_t)$ ，并以此作为替换最终的决策依据。修正后的贝叶斯决策规则：

$$\hat{y}_{final} = \arg \max_{j \in \{\text{classes}\}} C(y = j | \mathbf{x}_t)$$

其中，综合置信度 C 被定义为三种独立证据的加权融合：

$$C(y = j | \mathbf{x}_t) = w_1 \cdot \underbrace{P(y = j | \mathbf{x}_t)}_{\text{模型证据}} + w_2 \cdot \underbrace{S_{\text{physics}}(j, \mathbf{x}_t)}_{\text{物理先验证据}} + w_3 \cdot \underbrace{S_{\text{align}}(j, \mathbf{x}_t)}_{\text{迁移质量证据}}$$

模型证据 $P(y = j | \mathbf{x}_t)$ ：即分类器的 Softmax 输出，代表模型自身的判断。

物理先验证据 S_{physics} ：替换设计的物理机理符合度评分。它量化了样本 \mathbf{x}_t 的特征表现，与假设类别 j 的物理机理的吻合程度。这是一个强有力的领域知识先验。

迁移质量证据 S_{align} ：领域对齐质量评分。它量化了样本 \mathbf{x}_t 在迁移后的共享空间中，与源域类别 j 的融合程度，代表了此次迁移对该样本的有效性。

表 3：最终诊断报告

Filename	初步诊断	分类器概率	物理机理符合度	领域对齐质量	综合诊断置信度 C	置信度等级	二次诊断故障类型
J.mat	内圈故障	0.9998	0.9213	1	0.9415	高置信度	-
I.mat	内圈故障	0.9262	0.8954	1	0.8987	高置信度	-
N.mat	外圈故障	1	0.8532	1	0.8913	高置信度	-
E.mat	内圈故障	0.9969	0.8861	0.8	0.8856	高置信度	-
C.mat	内圈故障	0.7256	0.9105	1	0.8524	高置信度	-
H.mat	外圈故障	1	0.8149	0.8	0.846	高置信度	-
M.mat	正常	0.9768	0.8655	0.8	0.8443	高置信度	-
A.mat	正常	0.9921	0.8237	0.8	0.8386	高置信度	-
D.mat	滚动体故障	0.998	0.7952	0.8	0.8272	高置信度	-
G.mat	正常	0.9174	0.8163	1	0.8251	高置信度	-
F.mat	滚动体故障	0.9957	0.6531	0.6	0.7188	中置信度	-
B.mat	正常	0.8514	0.7188	0.6	0.7126	中置信度	-
L.mat	滚动体故障	0.9574	0.6133	0.4	0.6698	中置信度	-
P.mat	滚动体故障	0.9131	0.5891	0.4	0.6517	中置信度	-
K.mat	正常	0.6069	0.4582	0	0.4897	低置信度	滚动体故障
O.mat	正常	0.5455	0.4211	0	0.4492	低置信度	内圈故障

3.3 本章小结

本章构建了一个从领域自适应到贝叶斯决策修正的完整、创新的跨域诊断框架。替换首先通过 DANN 模型初步解决了领域漂移问题，随后，替换直面模型的黑箱本质与不确定性，创造性地引入了基于贝叶斯思想的诊断仲裁机制。该机制通过融合模型、物理先验和迁移质量三层证据，能够识别并自动修正模型初步诊断中的低置信度结果。最终的诊断报告（表 3）不仅给出了故障类型，更重要的是，为每一个诊断结果提供了一个可量化的、逻辑透明的、融合了物理规律与数据证据的综合置信度评分。这种从被动的“事后解释”到主动的“干预决策”的升华，是本次建模的核心创新，它极大地增强了最终诊断结果的可靠性，为构建真正可信赖的 AI 诊断系统提供了一种全新的建模范式。

Filename	Predicted_Label	预测故障类型
A.mat	Normal	正常
B.mat	Normal	正常
C.mat	Inner_Ring_Fault	内圈故障
D.mat	Ball_Fault	滚动体故障
E.mat	Inner_Ring_Fault	内圈故障
F.mat	Ball_Fault	滚动体故障
G.mat	Normal	正常
H.mat	Outer_Ring_Fault	外圈故障
I.mat	Inner_Ring_Fault	内圈故障
J.mat	Inner_Ring_Fault	内圈故障
K.mat	Normal	正常
L.mat	Ball_Fault	滚动体故障
M.mat	Normal	正常
N.mat	Outer_Ring_Fault	外圈故障
O.mat	Normal	正常
P.mat	Ball_Fault	滚动体故障

图 4：第三问结果图