

基于小样本数据的 YBCO 薄膜生长工艺参数初步探索 与机器学习分析

王鑫^{1) #} 张济全^{1) #} 李陈志航^{2) #} 任津生^{2) †} 王宇彤¹⁾ 张玉婷¹⁾ 肖
俊成¹⁾ 杨立红¹⁾ 江星宇^{1) †} 袁洁¹⁾

1) (中国科学院物理研究所, 超导国家重点实验室, 北京 100190)

2) (启元实验室, 智能部组件研究中心, 北京 100095)

摘要

YBa₂Cu₃O_{7-δ} (YBCO) 高温超导薄膜的制备受多项工艺参数的共同影响, 然而实际的薄膜生长实验难以快速积累大规模数据。本文基于小样本实验数据, 探究了神经网络模型在优化脉冲激光沉积 (PLD) 制备 YBCO 薄膜过程中的指导作用。数据涵盖衬底种类、氧压、生长时间等十个关键工艺参数, 以 YBCO 的 (005) 峰位及摇摆曲线半高全宽作为薄膜结构特性的表征指标。通过构建多层感知机 (MLP) 模型, 并引入 Bootstrap 重采样与早停机制以缓解有限样本下的过拟合风险。在此基础上, 结合皮尔逊相关性检验, 对工艺参数与结构特征间的复杂关联进行初步探索与可视化分析, 从有限数据中提取关键物理模式。本工作为小样本条件下的超导薄膜实验提供了可迁移的分析框架, 也为后续的数据采集和更深入的“工艺—结构—性能”关系研究奠定了基础。

关键词: YBCO 薄膜, 小样本学习, 数据效率, 机器学习, 工艺优化

PACS: 02.10.Yn, 33.15.Vb, 68.55.-a, 07.05.Mh

基金: 国家重点基础研究发展计划 (批准号: 2022YFA1603903, 2021YFA0718700)、国家自然科学基金 (批准号: U23A6015, 60573172)、中国科学院基础研究青年团队 (批准号: 2022YSBR-048)、中国博士后基金 (批准号: 2023TQ0365, 2024M753464, GZB20230826)、启元创新基金 (批准号: S20230201023) 资助的课题。

同等贡献作者.

† 通信作者.E-mail: xingyujiang@iphy.ac.cn; renjinsheng@qi Yuanlab.com

第一作者 .E-mail: wangxin_@iphy.ac.cn; zhang_jiquan@iphy.ac.cn;
li_chenzhihang@bit.edu.cn

1. 引言

超导薄膜在超导机理探索、超导器件研发中有着非常重要的作用。YBa₂Cu₃O_{7-δ} (YBCO) 高温超导薄膜作为最早被发现且研究最为广泛的铜氧化物超导体之一^[1-3], 其超导转变温度 (T_c) 高于液氮温区, 在强电 (如超导电缆、强磁体、限流器) 和弱电 (如微波滤波器、约瑟夫森结、超导量子干涉仪) 领域均展现出巨大的应用潜力^[4-10]。因此, 实现高质量 YBCO 薄膜的可控制备, 是推动其从实验室走向实际应用的核心前提。

然而, YBCO 薄膜的制备是一个多参数耦合的复杂过程, 高质量薄膜的制备仍面临严峻挑战。以最常用的脉冲激光沉积 (PLD) 技术为例, 薄膜的超导性能、结晶质量和表面形貌受到衬底种类、沉积温度、氧压、激光能量、生长时间、退火流程等数十个工艺参数的显著影响^[11, 12]。这些参数之间存在复杂的非线性耦合关系, 例如, 提高沉积温度有助于改善结晶性, 但可能导致表面粗糙度恶化; 增加氧压可促进氧掺杂以提升 T_c , 但过高的气压又会削弱沉积粒子的动能, 诱发杂相生成。传统的“试错法”优化策略不仅周期长、成本高昂, 而且严重依赖操作人员的经验, 难以系统性地探索整个参数空间并锁定全局最优的工艺窗口, 成为制约 YBCO 薄膜性能进一步提升与应用推广的主要瓶颈。

为解决上述问题, 将机器学习技术应用于薄膜生长工艺优化, 已成为薄膜材料领域一种极具前景的研究方法^[13-23]。贝叶斯优化、神经网络、进化算法等机器学习方法, 已在氮化钛^[14]、碳化硅^[19]等材料体系中得到验证, 其可通过构建“工艺—结构—性能”高维非线性映射模型, 高效搜索最优工艺条件。这种数据驱动策略可利用代理模型替代昂贵的物理实验与计算仿真, 从而在有限的实验次数内快速收敛至全局最优解, 显著缩短研发周期并降低开发成本。因此, 开展机器学习辅助的 YBCO 薄膜生长参数优化研究, 对于突破其性能瓶颈、实现可控制备、并加速其在能源、医疗与信息技术等领域的商业化应用具有至关重要的作用。

受限于仅有的近百组完整的“工艺—表征”数据, 本研究面临典型的小样本

挑战。在此条件下，直接构建兼具高精度与强泛化能力的预测模型极为困难。因此，本研究聚焦 YBCO 薄膜的小样本工艺优化。基于现有有限实验数据，利用机器学习工具挖掘工艺参数与薄膜结构表征之间的关联规律，识别关键影响因素，为后续定向实验设计提供可参考的假设与优先级建议。本研究的目标不仅是预测结构参数，更在于创建适用于凝聚态实验的小样本数据分析方法，为数据匮乏场景下的材料工艺研究提供技术支撑。

2. 研究方法

2.1 薄膜制备与表征

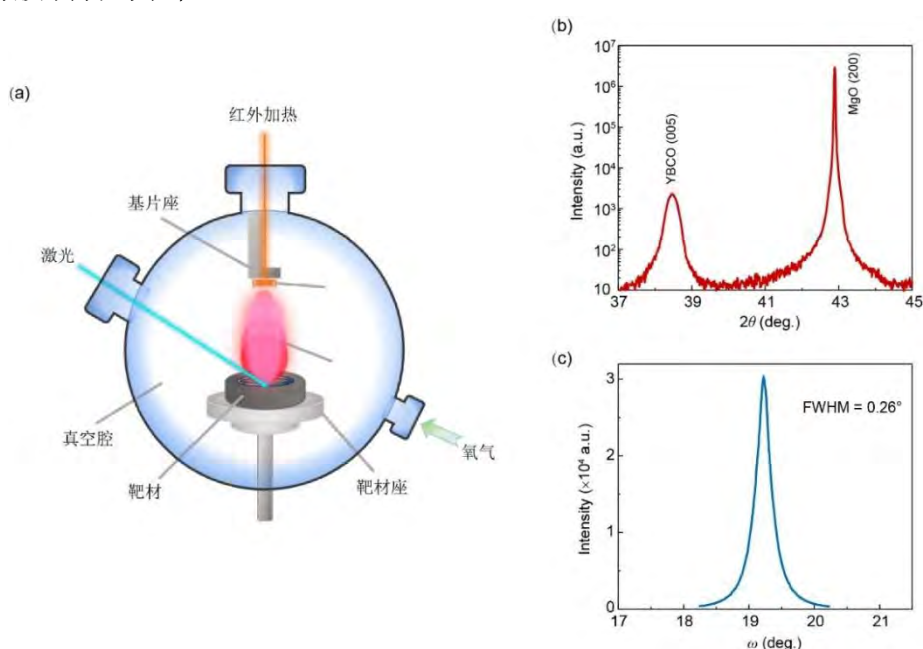


图 1 (a) PLD 原理图；(b) XRD $\theta - 2\theta$ 扫描；(c) YBCO (005) 峰的摇摆曲线。

Fig.1. (a) PLD schematic diagram; (b) XRD $\theta - 2\theta$ scan;

(c) Rocking curve of the (005) peak of YBCO.

YBCO 薄膜采用 PLD 技术生长^[24]，如图 1 (a) 所示。生长薄膜所用的衬底包括 SrTiO_3 (100) (STO) 和 MgO (100) (MO) 两种单晶衬底，使用银胶和压片两种方法安置在两种不锈钢基片座上。Beamtech 的 Nimma-900 型脉冲固体激光器提供了脉冲激光光源，其波长为 354.8 nm。松山湖材料实验室实用超导薄膜团队提供了高致密度的 YBCO 靶材。薄膜生长前，设备真空度达到 10^{-6} 至 10^{-7} torr。激光能量密度和光斑大小由激光器输出电压和聚焦位置共同决定。生长过程中，我们使用红外激光将基片座加热至 700-900 °C，在腔体中充入 0.2-0.55 torr 的高

纯氧气，并使用了不同的靶基距（靶材—衬底间距）、激光频率和生长时间，得到了不同厚度的 YBCO 薄膜。生长过程结束后对薄膜进行退火处理：在腔体中充入 1 个大气压的高纯氧气，后逐步冷却至室温，以期得到最佳超导转变温度。

薄膜制备完成后利用 X 射线衍射仪对其进行基本的结构表征。图 1 (b)和 (c) 分别展示了薄膜的 $\theta - 2\theta$ 扫描图谱以及 (005) 峰的摇摆曲线。YBCO (005) 峰的位置信息可用于计算薄膜的 c 轴晶格常数，其反映了薄膜的应力状态和氧含量。摇摆曲线的半高全宽 (full width at half maximum (FWHM)) 是表征晶粒取向一致性与晶格畸变的关键指标，其值越小表明薄膜结晶质量越高。

2.2 数据预处理

我们收集了大量薄膜样品的生长参数以及结构参数 (XRD 表征数据)，表 1 以其中 6 个样品为例，展示了我们所使用的生长参数（包括衬底种类、衬底大小、激光焦距、靶基距、输出电压、激光频率、氧压、生长时间、基片温度、退火时间）和结构参数（YBCO (005) 峰位和摇摆曲线半高全宽）。考虑到衬底峰偏移引起的系统误差，我们首先对所有的 $\theta - 2\theta$ 数据进行了衬底峰校正，即将 STO (200) 或 MO (200) 峰移动到统一位置，再对 YBCO (005) 峰做高斯拟合，提取出峰位信息。

在进行模型训练前先对原始数据进行一致性检查与异常值识别。对于连续变量，计算第一/第三四分位数 $Q1$ 、 $Q3$ ，并以 $IQR = Q3 - Q1$ 构建经验正常范围 $[Q1 - 1.5IQR, Q3 + 1.5IQR]$ ，超出范围的记录标记为潜在离群点；同时结合箱线图与 Z-score 统计进行复核。鉴于样本量有限，离群点处理以“统计—核验—决策”为原则：对明显由记录错误或测量异常导致的样本予以剔除；对处于工艺窗口边界但经实验确认有效的样本予以保留，以覆盖更广参数空间并减少过度数据清洗带来的信息损失。

在数据集划分上，由于样本数量较少，我们从原始数据集中随机抽取了 5 条数据样本作为独立的测试集用于模型性能评估，剩余数据作为训练集。在模型训练前先对数据进行预处理，具体包括对衬底种类 (STO 和 MO) 做独热编码，对连续特征采用 StandardScaler 标准化处理（减去均值并除以标准差），以消除不同物理量纲与数值范围差异对模型训练的影响。为了确保模型评估的独立性和可信度，只有训练集的数据参与模型训练，从而避免测试集信息的泄露。

表 1 部分生长参数与 XRD 表征数据。

Table 1. A portion of growth parameters and XRD characterization data.

样品编号	生长参数									结构参数	
	衬底种类	衬底大小	激光焦距	靶基距	输出电压	激光频率	氧压	生长时间	基片温度	峰位	半高全宽
1	STO	5	-12	50	780	1	0.45	300	775	38.54	0.16
2	STO	5	-12	50	780	1	0.45	480	770	38.56	0.19
3	STO	5	-12	55	690	1	0.45	300	740	38.62	0.30
4	MO	10	-18	70	730	1	0.35	1020	870	38.47	0.58
5	MO	10	-18	70	730	1	0.35	1020	895	38.48	0.26
6	MO	10	-15	55	690	1	0.20	600	690	38.49	0.43

物理量单位：衬底大小 (mm)；激光焦距 (mm)；靶基距 (mm)；输出电压 (V)；激光频率 (Hz)；氧压 (torr)；生长时间 (s)；基片温度 (°C)；峰位 (deg.)；半高全宽 (deg.)。

2.3 神经网络模型搭建

模型选择阶段，本文对比评估了 3 种典型的机器学习方法，包括决策树 (Decision Tree)^[25]、支持向量机 (SVM)^[26]和多层感知机 (MLP)^[27-29]。决策树是一类常见的监督学习模型，通过递归划分特征空间实现回归预测，模型结构直观、训练效率较高。但单一决策树在小样本下易过拟合，难以充分刻画多参数耦合的复杂非线性关系。支持向量机在小样本数据集上有较好的泛化能力，适用于非线性回归拟合，但在处理高维特征时训练时间较长，且对超参数选择敏感。相比之下，多层感知机作为一种基础的前馈神经网络，在捕捉复杂非线性关联与高维输入的任务中具有更强的表达能力与更优的拟合效果，因此更适用于本研究的工艺—结构关系建模。

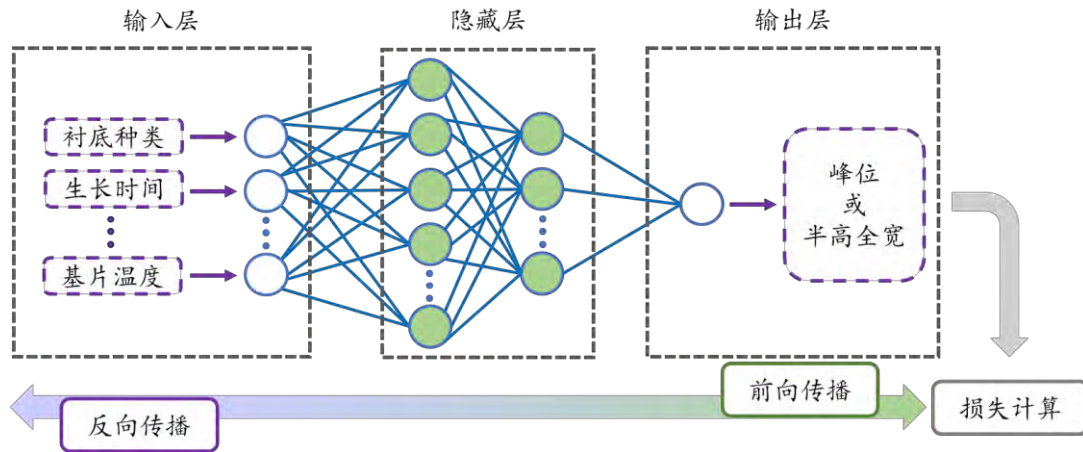


图 2 MLP 模型示意图。

Fig.2. Schematic diagram of the MLP model.

多层感知机 (MLP) 是一种前馈神经网络, 由输入层、隐藏层和输出层三个部分组成, 每一层的节点或神经元之间通过权重连接, 并通过反向传播算法调节节点权重, 使得参数沿着损失函数梯度相反的方向进行更新, 从而实现模型预测中的最小化误差。具体步骤包括由输入层依次传递到不同隐藏层、输出层的前向传播, 每个神经元通过加权求和以及激活函数处理后进行输出, 隐藏层中的每个神经元分别作为上一层的输出与下一层的输入, 由此逐级传递直至输出最终模型。前向传播完成后, 采用损失函数计算模型在训练集上的损失, 本研究为回归任务, 选用均方根误差 (RMSE) 量化预测值与真实值的偏差。随后将误差进行反向传播, 同时优化更新模型参数进行新一轮的训练。整个模型的训练需要不断重复这一过程直至模型收敛。采用决定系数 (R^2) 和均方根误差 (RMSE) 评价模型预测值与真实值之间的拟合程度, R^2 越接近 1, RMSE 越小, 表明拟合程度越高。

如图 2 所示, 模型以薄膜生长参数作为输入层, 以待预测的结构参数作为输出层。在模型训练时首先输入汇总好的数据信息, 执行由输入层到隐藏层到输出层的前向传播, 在输出预测信息后, 进行损失计算和反向传播直至模型收敛。考虑到本实验数据集样本量较少, 为降低单一划分训练集随机划分带来的偏差, 我们在训练过程中引入了 Bootstrap 方法^[30]。Bootstrap 是一种重采样技术, 它通过从原始训练集随机抽取数据 (允许重复抽取), 生成多个子训练集用于模型训练与评估。这一方法能有效增加训练数据的多样性, 缓解小样本数据带来的偏差, 从而提升模型的泛化能力。我们通过对原始数据集进行有放回的随机重采样, 构

建多个训练子集进行模型训练与评估，并采用早停、L2 权重衰减等正则化策略，有效缓解了小样本数据可能导致的过拟合问题，保证了模型评估的可靠性。

为了定量分析各生长参数与结构参数之间的关系，本文对预处理后的数据进行了相关性分析。分别以 YBCO (005) 峰位和摇摆曲线半高全宽为目标量，将所有样本的工艺参数与结构参数组成数据矩阵，通过计算每个生长参数与目标量之间的皮尔逊 (Pearson) 相关系数 r 来衡量变量之间线性相关程度和方向，实现关键影响因素的快速筛选。相关系数 $r \in [-1, 1]$ ，绝对值越大表示线性相关性越强，正值/负值分别对应正相关/负相关。

3. 结果与讨论

3.1 MLP 模型性能评估

图 3 (a, b) 分别展示了模型在训练集上对 YBCO (005) 峰位和摇摆曲线半高全宽的预测性能。对于峰位预测 (图 3 (a))，模型实现了 $R^2 = 0.97$ 、 $RMSE = 0.01$ 的优异拟合效果，表明神经网络能够有效学习工艺参数与晶体结构之间的复杂非线性关系；在半高全宽预测任务中 (图 3 (b))， $R^2 = 0.87$ 、 $RMSE = 0.05$ ，虽略低于峰位预测精度，但仍体现出良好的建模能力，说明模型可准确捕捉薄膜结晶质量的变化趋势。

为了对模型的泛化能力进行检验，我们使用独立于训练集的 5 组新实验数据进行测试，结果如图 3 (c, d) 所示。在峰位预测中 (图 3 (c))， $R^2 = 0.96$ ， $RMSE = 0.01$ ，预测值与真实值高度吻合，未出现系统性偏差。在半高全宽预测中 (图 3 (d))， $R^2 = 0.79$ 、 $RMSE = 0.06$ ，性能相对峰位有所下降，但整体趋势保持一致，所有测试点均落在合理误差范围内。测试集的预测性能与训练集相当，说明模型具有良好的泛化能力。

值得进一步讨论的是，摇摆曲线半高全宽的预测精度明显低于 (005) 峰位。这一现象在物理上是合理的：(005) 峰位主要反映 YBCO 薄膜的 c 轴晶格常数，受氧含量和宏观应力状态主导，其与衬底种类、氧压等关键参数的关系相对明确；而半高全宽表征的是晶粒取向分布的弥散程度，即薄膜的结晶质量，其形成过程涉及原子的表面扩散、成核竞争、界面能调控等多个动力学环节，对激光能量密度、靶基距、生长时间乃至衬底表面状态等参数的依赖更为敏感且高度非线性。

此外，微小的工艺波动（如羽辉稳定性、局部温差）更容易在半高全宽上体现为显著变化，导致其数据噪声水平更高、规律性更弱。因此，半高全宽的高精度预测对模型的表达能力和数据质量提出了更高要求，这也解释了为何其预测性能略逊于峰位。

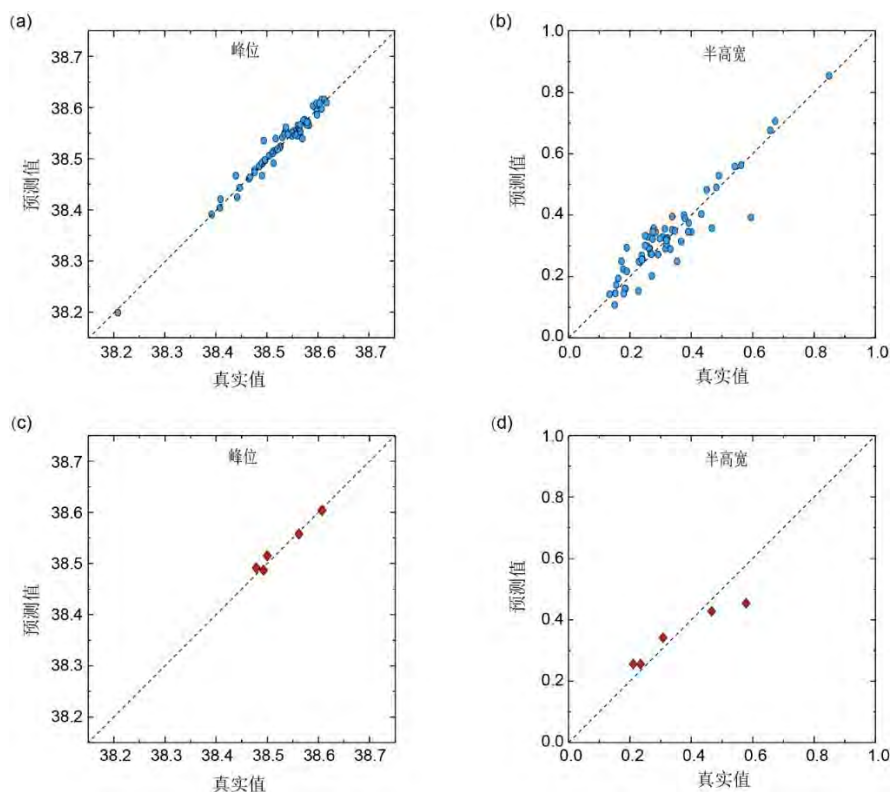


图3 模型训练与测试结果。(a) 峰位与 (b) 半高全宽的训练结果。(c) 峰位与 (d) 半高全宽的测试结果。虚线表示预测值等于真实值。

Fig.3. Training and testing results of MLP model. Training results for (a) peak position and (b) FWHM. Testing results for (c) peak position and (d) FWHM. The dashed line indicates where the predicted value equals the true value.

3.2 工艺参数影响机制分析

图4展示了各生长参数与YBCO(005)峰位和摇摆曲线半高全宽之间的皮尔逊相关系数(r)，直观反映了不同工艺变量对薄膜晶格结构和结晶质量的影响程度。从图中可见，衬底种类是影响峰位和半高全宽的最主要因素。STO衬底与MO衬底呈现出趋势相反的相关性。一方面，STO衬底使(005)峰位向高角度偏移($r = +0.50$)，MO衬底使得峰位向低角度偏移($r = -0.50$)；另一方面，STO衬底显著降低半高全宽($r = -0.42$)，MO衬底则表现出相反趋势($r =$

+0.42)。这些现象均源于 STO、MO 两种衬底与 YBCO 的晶格匹配差异：四方相 YBCO 的面内晶格常数约为 3.86 Å，STO 和 MO 的晶格常数分别为 3.905 Å 和 4.21 Å。YBCO 在 STO 衬底上可实现共格外延生长，其面内晶格被迫与衬底晶格匹配，产生面内拉伸应变。根据泊松效应 (Poisson effect)，材料在某一方向受外力拉伸时，垂直外力的方向将发生收缩。因此，YBCO 的 c 轴受到压缩，导致 (005) 衍射峰向高角度偏移（即 c 轴晶格常数减小），STO 衬底的 c 轴压缩效应已被 TEM 研究证实^[11, 27]，这与本文观察到的正相关性 ($r = +0.50$) 一致。相反，在 MO 衬底上，由于失配度过大，难以维持共格外延，薄膜倾向于形成多晶或非外延结构，局部应力分布不均，整体表现为较弱的面内约束和相对松弛的 c 轴，因而 (005) 峰位向低角度移动 ($r = -0.50$)。此外，大失配还会诱发高密度位错和晶界，破坏晶粒取向一致性，导致摇摆曲线半高全宽显著增大。

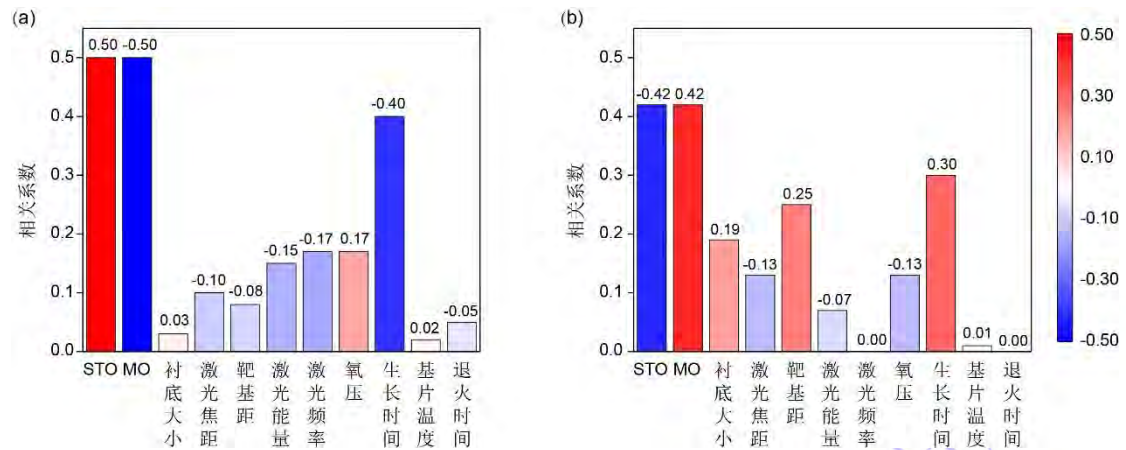


图 4 各生长参数与 (a) YBCO (005) 峰位、(b) 摇摆曲线半高全宽间的相关性。

Fig. 4. Correlation between growth parameters and (a) the peak position of YBCO (005) and (b) FWHM of the rocking curve.

其次，生长时间对峰位和半高全宽均产生显著但方向不同的影响。生长时间与峰位呈负相关 ($r = -0.40$)，说明随着生长时间延长（即膜厚增加），(005) 峰位逐渐向低角度移动，对应 c 轴晶格常数增大，这可能与膜厚增加导致的应变弛豫有关，或受到氧含量分布的影响。同时，生长时间与半高全宽呈正相关 ($r = +0.30$)，说明结晶质量随膜厚增加而下降。这一现象可归因于 YBCO 薄膜典型的“层-岛结合” (Stranski-Krastanov) 生长机制：初始阶段以层状生长为主；随着厚度增加，三维岛状生长逐渐主导，晶粒取向一致性降低，导致半高全宽增大。

除了衬底种类和生长时间外,其余参数对峰位和半高全宽的影响存在明显差异。氧压、激光频率和激光能量对峰位的影响较为显著,但对半高全宽的影响较弱;退火时间对峰位有一定影响,而对半高全宽几乎无影响。这表明这些参数主要通过调控薄膜中的氧含量(进而改变 c 轴晶格常数)来影响峰位。相反,靶基距和衬底尺寸对半高全宽的影响较大,但对峰位影响较小,反映出它们主要通过调节沉积粒子动能、表面扩散能力等生长动力学过程,影响成核密度与晶粒取向排列,从而调控结晶质量。这种参数响应的差异性也解释了为何在机器学习建模中,半高全宽的预测难度高于峰位:前者受多种非线性、动态的表面与界面过程耦合影响,数据噪声更高、规律性更弱;而后者主要由热力学平衡态下的氧掺杂水平和宏观应力状态决定,关系相对明确且稳定。

在本研究的数据集中,基片温度与峰位及半高全宽的相关性均接近于零,表明在当前实验参数窗口(700 - 900°C)内,未观察到基片温度与 XRD 指标的显著线性相关性。这可能是因为当前温度窗口(700 - 900°C)接近 YBCO 最优外延区间^[12],其效应趋于饱和;亦或其影响被其它强耦合参数(如衬底种类、生长时间)所掩盖,后续需在低温区间(<600°C)或高温区间(>950°C)验证温度敏感性。

MLP 是黑箱模型,因其复杂的非线性结构而难以直接解释,高 R^2 不代表揭示了真实物理机制。相关性分析虽能揭示变量间的关联,但只能作为生成科学假设的起点,关键结论需后续单变量实验验证(如固定其他参数,系统调控生长时间)。

4. 结论

本文基于 PLD 技术制备 YBCO 高温超导薄膜的实验数据集,构建了 MLP 神经网络模型,实现了对 YBCO (005) 衍射峰位及摇摆曲线半高全宽的高精度预测。该模型在独立测试集上表现出良好的泛化性能,验证了数据驱动方法在复杂薄膜工艺建模中的可行性。进一步的相关性分析揭示了各生长参数对薄膜晶格结构与结晶质量的不同影响机制:其中,(005) 峰位反映晶格常数,主要受氧含量与宏观应力调控,关系相对明确;而半高全宽表征的结晶质量则涉及复杂的生长动力学过程,对工艺参数的依赖更为敏感且高度非线性。当前模型对峰位

的预测误差 (RMSE = 0.01°) 已接近 XRD 测量精度 ($\pm 0.005^\circ$), 可为 YBCO 薄膜工艺优化提供可靠参考。

本工作为 YBCO 薄膜的智能工艺优化提供了一条高效、可行的技术路径, 也为深入理解“工艺—结构”之间的映射关系提供了量化依据。在此基础上, 未来可通过以下方向进一步拓展: 一方面, 扩大数据规模并拓宽参数空间, 例如引入更多衬底种类、更精细的退火条件及原位监测信息等参数, 进一步提升模型的泛化能力与适用范围; 另一方面, 在现有结构参数预测的基础上, 将模型预测范围延伸至超导转变温度 (T_c)、临界电流密度 (J_c) 等关键电输运性能, 推动工艺优化从“结构可控”迈向“性能导向”的闭环设计, 为高温超导材料从实验室研究走向规模化产业应用提供强有力的技术支撑。

5. 致谢

本研究工作得到中国科学院超导专项的大力支持(资助编号:[SCZX-0102])。

参考文献

- [1] Zhao Z X, Chen L Q, Yang Q S, Huang Y Z, Chen G H, Tang R M, Liu G R, Cui C G, Chen L, Wang L Z, Guo S Q, Li S L, Bi J Q [1987 Chin. Sci. Bull. 32 412](#) (in Chinese)
[赵忠贤, 陈立泉, 杨乾声, 黄玉珍, 陈赓华, 唐汝明, 刘贵荣, 崔长庚, 陈烈, 王连忠, 郭树权, 李山林, 毕建清 1987 科学通报 32 412]
- [2] Wu M K, Ashburn J R, Torng C J, Hor P H, Meng R L, Gao L, Huang Z J, Wang Y Q, Chu C W [1987 Phys. Rev. Lett. 58 908](#)
- [3] Dijkkamp D, Venkatesan T, Wu X D, Shaheen S A, Jisrawi N, Min-Lee Y H, McLean W L, Croft M [1987 Appl. Phys. Lett. 51 619](#)
- [4] Iijima Y, Tanabe N, Ikeno Y, Kohno O [1991 Physica C 185-189 1959](#)
- [5] Goyal A, Norton D P, Christe D K, Specht E D, Paranthaman M, Kroeger D M, Budai J D, He Q, List F A, Feenstra R, Kerchner H R, Lee D F, Hatfield E, Martin P M,

- Mathis J, Park C [1996 *Appl. Supercond.* **4** 403](#)
- [6] Antončík F, Jankovský O, Hlásek T, Bartůněk V [2020 *Nanomaterials* **10** 1429](#)
- [7] Borisenko I V, Kotelyanski I M, Shadrin A V, Komissinski P V, Ovsyannikov G A
[2005 *IEEE Trans. Appl. Supercond.* **15** 165](#)
- [8] Cybart S A, Cho E Y, Wong T J, Wehlin B H, Ma M K, Huynh C, Dynes R C [2015
Nat. Nanotechnol. **10** 598](#)
- [9] Greer J A [1992 *J. Vac. Sci. Technol. A* **10** 1821](#)
- [10] Tian Q W, Yuan P S, Yu H Q, Wang S N, Liu X Y, Li L Y, You L X [2025 *Acta Phys.
Sin.* **74** 057401](#) (in Chinese) [田清文, 原蒲升, 余慧勤, 汪书娜, 刘晓宇, 李凌云, 尤立星 2025 物理学报 **74** 057401]
- [11] Chen F C 2023 *Ph. D. Dissertation* (Beijing: University of Chinese Academy of Sciences) (in Chinese) [陈赋聪 2023 博士学位论文 (北京: 中国科学院大学)]
- [12] Lin Z F, Sun W X, Liu T X, Tu S J, Ni Z, Bai X B, Zhao Z Y, Zhang J Q, Chen F C, Hu W, Feng Z P, Yuan J, Jin K [2023 *J. Synth. Cryst.* **52** 1036](#) (in Chinese) [林泽丰, 孙伟轩, 刘天想, 涂思佳, 倪壮, 柏欣博, 赵展艺, 张济全, 陈赋聪, 胡卫, 冯中沛, 袁洁, 金魁 2023 人工晶体学报 **52** 1036]
- [13] Kim H J, Chong M, Rhee T G, Khim Y G, Jung M H, Kim Y M, Jeong H Y, Choi B K, Chang Y J [2023 *Nano Converg.* **10** 10](#)
- [14] Ohkubo I, Hou Z, Lee J N, Aizawa T, Lippmaa M, Chikyow T, Tsuda K, Mori T
[2021 *Mater. Today Phys.* **16** 100296](#)
- [15] Pfeifer S, Pokuri B S S, Du P F, Ganapathysubramanian B [2018 *Mater. Discov.* **11**](#)

- [16] Price C C, Li Y S, Zhou G Y, Younas R, Zeng S S, Scanlon T H, Munro J M, Hinkle C L [2024 Nano Lett. 24 14862](#)
- [17] Provence S R, Thapa S, Paudel R, Truttmann T K, Prakash A, Jalan B, Comes R B [2020 Phys. Rev. Mater. 4 083807](#)
- [18] Shen C, Zhan W K, Xin K Y, Li M Y, Sun Z Y, Cong H, Xu C, Tang J, Wu Z F, Xu B, Wei Z M, Xue C L, Zhao C, Wang Z G [2024 Nat. Commun. 15 2724](#)
- [19] Tang Z R, Zhao S B, Li J, Zuo Y H, Tian J, Tang H Y, Fan J J, Zhang G Q [2024 Case Stud. Therm. Eng. 59 104507](#)
- [20] Wahab H, Jain V, Tyrrell A S, Seas M A, Kotthoff L, Johnson P A [2020 Carbon 167 609](#)
- [21] Yao T S, Tang C Y, Yang M, Zhu K J, Yan D Y, Yi C J, Feng Z L, Lei H C, Li C H, Wang L, Wang L, Shi Y G, Sun Y J, Ding H [2019 Chin. Phys. Lett. 36 068101](#)
- [22] Zhang Q, Tan W, Ning Y Q, Nie G Z, Cai M Q, Wang J N, Zhu H P, Zhao Y Q [2024 Acta Phys. Sin. 73 230201](#) (in Chinese) [张桥, 谭薇, 宁勇祺, 聂国政, 蔡孟秋, 王俊年, 朱慧平, 赵宇清 2024 物理学报 73 230201]
- [23] Han P, Zang H, Yu Y, Zhang F, Zhou N D, Shi Z M, Sun X J, Li D B [2026 Acta Phys. Sin. 75 080710](#) (in Chinese) [韩鹏, 臧行, 于艳, 张凤, 周诺丹, 石芝铭, 孙晓娟, 黎大兵 2026 物理学报 75 080710]
- [24] Zhang D M 2011 *Pulsed Laser Deposition Dynamics* (Beijing: Science Press) (in Chinese) [张端明 2011 脉冲激光沉积动力学原理 (北京: 科学出版社)]
- [25] Breiman L, Friedman J, Olshen R A, Stone C J 1984 *Classification and Regression Trees* (New York: Chapman and Hall/CRC)

[26] Cortes C, Vapnik V [1995 *Mach. Learn.* **20** 273](#)

[27] Rosenblatt F [1958 *Psychol. Rev.* **65** 386](#)

[28] Minsky M L, Papert S A 1988 *Perceptrons: An Introduction to Computational Geometry* (Cambridge, MA: MIT Press)

[29] Samek W, Montavon G, Lapuschkin S, Anders C J, Müller K R [2021 *Proc. IEEE* **109** 247](#)

[30] Efron B [1979 *Ann. Stat.* **7** 1](#)

录用稿件，非最终出版稿

Machine Learning Exploration of YBCO Thin Film Growth Parameters under Small Data Constraint

Xin Wang ^{1)♯} Jiquan Zhang ^{1)♯} Chenzhihang Li ^{2)♯} Jinsheng Ren ^{2)†} Yutong Wang ¹⁾

Yuting Zhang ¹⁾ Juncheng Xiao ¹⁾ Lihong Yang ¹⁾ Xingyu Jiang ^{1)†} Jie Yuan ¹⁾

1) (National Laboratory for Superconductivity, Institute of Physics, Chinese Academy of Sciences, Beijing 100190, China)

2) (Qiyuan Lab, Beijing 100095, China)

Abstract

The fabrication of $\text{YBa}_2\text{Cu}_3\text{O}_{7-\delta}$ (YBCO) high-temperature superconducting thin films is governed by multiple process parameters. However, the accumulation of large-scale experimental data in actual thin-film growth processes remains challenging. This study utilizes a small-sample experimental dataset to examine the guiding role of a neural network model in optimizing the pulsed laser deposition (PLD) processes for YBCO thin films. The dataset encompasses ten key process parameters, including substrate type, substrate size, laser focal length, target-substrate distance, output voltage, laser frequency, oxygen pressure, growth time, substrate temperature, and annealing time, which serve as the model inputs. The YBCO (005) peak position and the full width at half maximum (FWHM) of the rocking curve, obtained from X-ray diffraction (XRD), are the structural parameters to be predicted and constitute the model outputs. A multilayer perceptron (MLP) neural network model was constructed, incorporating bootstrap resampling and early stopping mechanisms to mitigate overfitting under small-sample conditions. The model yields reasonable predictions of the peak position and the FWHM, and demonstrates favorable generalization performance on an independent test set, thereby validating the feasibility of data-driven approaches for modeling complex thin-film processes. Notably, the root-mean-square error (RMSE) for the (005) peak position prediction

is as low as 0.01° , approaching the instrumental resolution of XRD measurements ($\pm 0.005^\circ$). On this basis, Pearson correlation analysis is further conducted to perform a preliminary exploration and visual assessment of the complex interrelations between the process parameters and the structural features. Specifically, the peak position, which reflects lattice constant variations, is primarily regulated by oxygen content and macroscopic stress, exhibiting a relatively clear relationship. In contrast, the FWHM, which reflects crystalline quality, is governed by complex growth kinetics and exhibits a more sensitive and highly nonlinear dependence on the process parameters. This work establishes a scalable and transferable analytical framework for superconducting thin-film experiments under small-sample constraints. It paves a feasible pathway for data-driven process optimization in resource-limited circumstances, and facilitates comprehensive in-depth interpretation of process-structure-property correlations.

Keywords: YBCO thin film, small sample, data efficiency, machine learning, growth process optimization _

* Project supported by the National Key Basic Research Program of China (Grant Nos. 2022YFA1603903, 2021YFA0718700), the National Natural Science Foundation of China (Grant Nos. U23A6015, 60573172), CAS Project for Young Scientists in Basic Research (Grant No. 2022YSBR-048), the China Postdoctoral Science Foundation (Grant Nos. 2023TQ0365, 2024M753464, GZB20230826) and Qiyuan Innovation Program (Grant No. S20230201023).

These authors contributed equally.

† Corresponding author. E-mail: xingyujiang@iphy.ac.cn; renjinsheng@qiyuanlab.com
The first author. E-mail: wangxin@iphy.ac.cn; zhang_jiquan@iphy.ac.cn;
li_chenzhihang@bit.edu.cn