第 52 卷 第 7 期：3437-3447 高电压技术 Vol.52, No.7: 3437-3447
2026 年 7 月 31 日 High Voltage Engineering July 31, 2026
DOI: 10.13336/j.1003-6520.hve.20250619

基于光谱通道拼接与多级注意力融合的绝缘纸聚合度定量评估技术

李　含[1]，孙伟哲[1]，秦少瑞[2]，陈希源[1]，张冠军[1]，李　元[1]
（1. 西安交通大学电气工程学院，西安 710049；2. 国网安徽省电力有限公司电力科学研究院，合肥 230601）

摘　要：近红外光谱技术是一种电力变压器绝缘纸聚合度无损定量评估的新方法。为了提升该技术对现场复杂类型纸样测试的准确性与泛化能力，提出基于光谱微分通道拼接与多级注意力融合的深度学习框架。首先构建了含 6 类典型纸样、300 个不同老化程度样本的光谱数据库；融合一阶与二阶微分预处理通道并通过基线漂移与随机噪声实现数据增强；模型在通道域引入 SENet 自适应提取通道特征权重，在波长域采用 Transformer 编码器捕捉光谱长程非线性关联。结果表明，多级注意力融合建模可以有效建立光谱序列间的长程关联，注意力集中于 C—H 及 C═O 两种基团的特征吸收区。模型在测试集上均方根误差(RMSE)为 45.6，平均相对误差(MAPE)为 5.01%，在 7 种对比模型中表现最优，现场实验中聚合度预测相对误差均低于 5%，可实现复杂场景下绝缘纸老化状态的高精度快速评估。

关键词：绝缘纸；聚合度；近红外光谱；深度学习；注意力机制

Quantitative Assessment of Degree of Polymerization of Insulating Paper by Spectral Channel Concatenation and Multi-level Attention Fusion

LI Han[1], SUN Weizhe[1], QIN Shaorui[2], CHEN Xiyuan[1], ZHANG Guanjun[1], LI Yuan[1]
(1. School of Electrical Engineering, Xi'an Jiaotong University, Xi'an 710049, China;
2. State Grid Anhui Electric Power Research Institute, Hefei 230601, China)

Abstract: Near-infrared spectroscopy (NIRS) is an emerging non-destructive method for quantitatively assessing the degree of polymerization (DP) of transformer insulating paper. To enhance the accuracy and generalization ability of this technique under complex on-site conditions, a deep learning framework based on spectral differential channel concatenation and multi-level attention fusion is proposed. A spectral database consisting of 300 samples with various aging states from six typical paper types is constructed. The first- and second-order differential preprocessing channels are integrated, and data augmentation is performed using baseline drift simulation and random noise injection. A SENet-based module is introduced in the channel domain to adaptively extract the weight of each preprocessing feature, while a Transformer encoder is used in the wavelength domain to capture long-range nonlinear dependencies in the spectral sequences. Results show that the multi-level attention fusion model effectively captures long-range dependencies across spectra and focuses attention on the characteristic absorption regions of C—H and C═O functional groups. The model achieves an RMSE of 45.6 and a MAPE of 5.01% on the test set, outperforming six other benchmark models. In field experiments, the relative error of DP prediction is consistently below 5%, therefore, the proposed method can be used to provide accurate and rapid assessment of insulating paper aging in complex conditions.

Key words: insulating paper; degree of polymerization; near-infrared spectroscopy; deep learning; attention

0　引言

全球能源格局加速转型背景下，电力变压器正朝着高电压、大容量方向发展，这对其绝缘系统稳定性提出了更高的要求[1]。电力变压器的主绝缘系统由油纸复合绝缘组成，其中，绝缘纸/纸板组成的固体绝缘系统承担着电气隔离、机械支撑等重要作用，直接决定了变压器的服役寿命[2]。绝缘纸聚合度(degree of polymerization, DP)是国家和行业标准中唯一直接表征绝缘纸老化程度的指标[3-4]。在绝缘纸原材料热压成型、设备绝缘装配到服役期老化的

基金资助项目：国家自然科学基金(52477159)；陕西省重点研发计划(2025PT-ZCK-14)。
Project supported by National Natural Science Foundation of China (52477159), Key Research and Development Program of Shaanxi Province (2025PT-ZCK-14).

全生命周期中，绝缘纸的 DP 会从 1 200 左右逐渐不可逆衰减至 150 左右[5]。因此，定期监测绝缘纸 DP 对保障变压器安全运行至关重要，而传统的实验室化学测试方法[3]具有破坏性大、测试周期长等局限，难以满足现场快速诊断需求，亟需开发快速无损的 DP 评估技术。

近红外光谱(near infrared spectrum，NIRS)技术具有操作便捷、分析快速、测试无损的突出优势，是有望替代传统化学测试手段实现绝缘纸老化状态无损评估的新型解决方案[6]。2001 年，Ali 等发现绝缘纸近红外光谱中羰基特征吸收带与其老化天数间存在关联性，初步证明了该技术是表征绝缘纸老化状态的有效方法[7]。Baird 构建了绝缘纸近红外光谱-聚合度(NIRS-DP)关联数据库并利用主成分分析和偏最小二乘回归建立 DP 定量评估模型，但受限于纸样种类少、样本库规模小等问题，模型现场测试不准确，无法满足应用需求[6]。

变压器内部绝缘纸包括电缆纸、皱纹纸以及绝缘纸板等多种类型，不同电压等级变压器使用的纸类型号也有差异，建立面向多种纸样的 DP 预测模型可增强技术现场应用的准确性。文献[8]首次建立了包含电缆纸、微皱纸与皱纹纸的混合纸样库，并对比了偏最小二乘回归与神经网络的建模表现，发现神经网络在混合纸样的 DP 预测中表现出更高的准确性。随着研究深入，主成分分析结合径向基神经网络、高斯过程回归[9]等方法被用于进一步提升模型性能。然而，随着纸样类型和样本量的扩增，数据复杂度显著增加，NIRS-DP 间关联更加难以挖掘，现有模型在现场应用时暴露出泛化能力不足的缺陷[10]。这本质上体现了传统特征工程结合浅层机器学习的策略在解析高维数据和复杂关联模式时的固有局限性。

深度学习凭借深层网络架构与参数可扩展性，在特征抽象与模式泛化方面展现出显著优势[11-13]。在基于深度学习的近红外光谱领域研究中，1D-CNN 是最常见的架构形式。文献[14]提出一种基于一维卷积神经网络的分类模型 Teanet，在近红外光谱分类任务的表现优于支持向量回归、随机森林等机器学习算法；文献[15]提出 Inception-ResNet 混合架构，通过 Inception 模块多尺度卷积核提升了光谱序列关系的解析能力；文献[16]论证了光谱预处理与数据增强策略对 CNN 模型收敛速度、稳定性的积极作用。然而，受限于卷积运算的局部性与卷积核大小限制，1D-CNN 的模型无法捕捉光谱序列的长程依赖关系，特征间的关联无法被有效提取[17]。此外，1D-CNN 的卷积操作对高频噪声过度敏感[18]，与近红外光谱数据特点的适配性不佳。

Transformer 架构凭借独特的自注意力机制展现出强大的长序列数据深度处理能力[19-21]，其数据处理方式更适应近红外光谱数据多重共线性、长程相关性的特点。文献[21]直接移植标准 Transformer 架构，在回归预测任务中的效果优于 1D-CNN。文献[23]构建了 CNN-Transformer 混合模型，通过局部特征提取与长程信息关联协同，实现了土壤组分的高精度预测。上述研究均基于大量光谱数据集实现，而绝缘纸全寿命周期样本积累难度大，用于模型训练的数据量不能满足 Transformer 庞大参数空间的需求[24]。同时，绝缘纸光谱采集还受到样本含水量、温度、绝缘油等复杂基质的耦合干扰，NIRS-DP 的关联模式更难以提取。因此，需综合利用预处理、数据增强等策略，并对 Transformer 架构开展针对性优化，实现在低数据资源、强光谱干扰场景的有效应用。

针对上述问题，本文提出一种基于微分通道拼接与多级注意力融合的深度学习框架，旨在构建面向现场应用的混合纸样 DP 定量评估模型。主要创新思路在于：(1)融合多光谱微分通道加权和基线漂移−噪声耦合数据增强策略，在有限数据下通过特征增强和数据增强使模型性能达到最优；(2)融合通道注意力机制(channel attention, CAT)[25]与纯编码器 Transformer(encoder-only transformer, EOT)[26]的多级注意力机制深度架构 CAT-EOT，实现了跨通道与跨波段特征的联合建模。

本文提出了基于 NIRS 定量评估绝缘纸 DP 的技术路线，建立典型类别绝缘纸的 NIRS-DP 数据库，提出预处理联合数据增强的数据集优化方法，并解构 CAT-EOT 架构的原理和特点，进而开展 NIRS-DP 的关联建模，最后从实验室检测与现场实验多方面分析验证了模型的先进性。

1 近红外光谱 DP 定量评估方法

近红外光谱定量评估需要构建光谱特征向量与 DP 之间的映射关系。假设一组包含 n 个样品的绝缘纸近红外光谱数据集 $\boldsymbol{X}=\{\boldsymbol{x}_1,\cdots,\boldsymbol{x}_n\}\in\mathbb{R}^d$，其中，$\boldsymbol{x}_i=[v_1,\cdots,v_d]$为样品 i 的 d 维近红外光谱数据，v_j 为第 j 个波长点处对应的吸光度。该光谱数据集所对

应的 DP 数据为 $\boldsymbol{y}=[y_1,\cdots,y_n]$。光谱与 DP 的目标映射关系可表述为：

$$y=f(\boldsymbol{x}_i;\theta)+\varepsilon,\ i\in\mathbb{N}_n \tag{1}$$

式中：f 为映射函数；θ 为模型参数空间；ε 为误差项。映射函数能否充分反映光谱与 DP 的关系取决于函数空间与数据分布的同构性，以及数据集中的纸样种类代表性和样本量。

对此，本文对模型训练提出两方面要求：(1)样本需要覆盖典型绝缘纸类型和全寿命劣化状态；(2)所建立的 NIRS-DP 映射函数具备高维非线性表达能力。本研究按照数据库建设—增强建模—优化验证的技术路线开展绝缘纸 DP 定量评估模型研究，如图 1 所示。利用实验室加速热老化实验，建立包含电缆纸、绝缘纸板及皱纹纸 3 种类型、6 种型号绝缘纸的全老化周期近红外光谱数据库，使用 CAT-EOT 多级注意力架构解析 NIRS-DP 的复杂映射关系，建立高精度模型，并通过实验室测试数据集和现场实验对模型进行充分优化与验证。

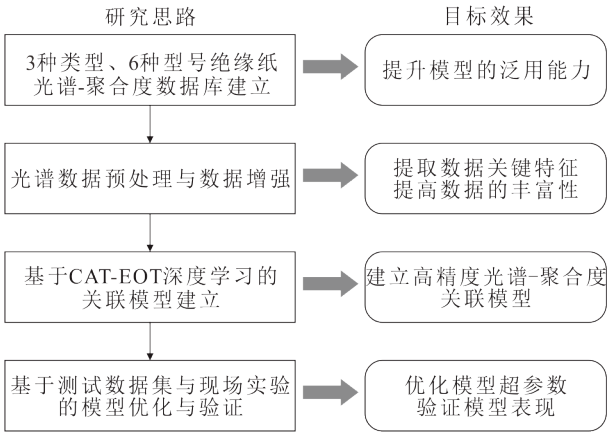


图 1　近红外光谱分析技术实现思路

Fig.1　Implementation Strategy of NIR Spectroscopy

2　绝缘纸 NIRS-DP 数据库建设

建立代表性绝缘纸 NIRS-DP 数据库是保障模型准确性和泛化能力的关键，数据库建立流程如图 2 所示，包括样品预处理与真空浸渍、加速热老化实验、绝缘纸 DP 测试(黏度法)与绝缘纸 NIRS 采集 4 个阶段。样本涵盖 35~500 kV 变压器典型纸样，如图 3 所示。纸样种类为 BZZ-75、BZZ-125、GDL、22HCC、JW50 以及 B31A，每种纸样制备 50 个标准样品，在真空干燥箱中与 KI25X 绝缘油真空浸渍，并在 130 ℃、133 Pa 条件下进行模拟加速热老化，老化周期 100 d，取样间隔为 2 d。

DP 测试依据 DL/T 984—2018 标准实施双平行测试，数据有效性原则为两次测试偏差阈值≤2.5%。测试结果如图 4 所示，所有纸样的 DP 随老化时间均呈指数衰减，3 种匝间绝缘纸 DP 的下降速率基本一致。由于热改性工艺，22 HCC 丹尼松纸在老化周期内的 DP 下降速率远低于普通绝缘纸和纸板。

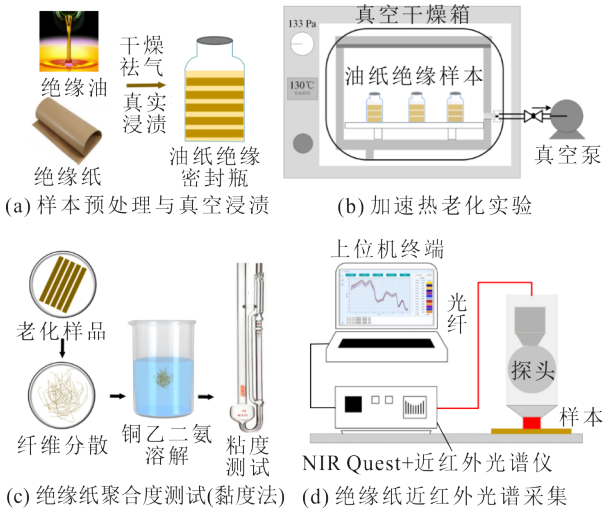


图 2　绝缘纸光谱-DP 数据库建立的流程

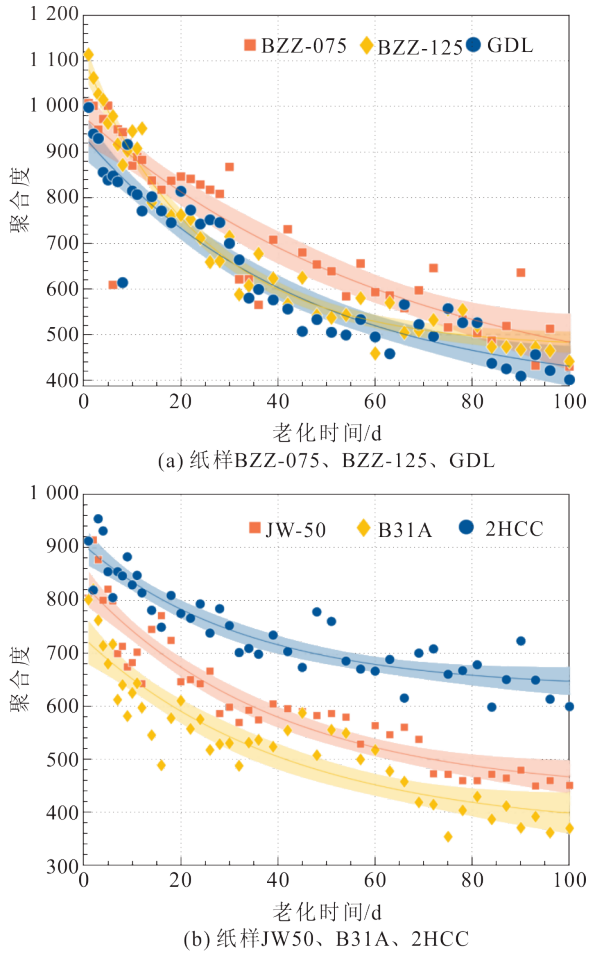
Fig.2　Establishment of the insulating paper spectrum-DP database



图 3　所选的 6 种典型绝缘纸

Fig.3　Six kinds of typical insulating paper selected

光谱采集使用 NIR Quest+2.2 近红外光谱仪，波长范围 900~2 200 nm，像元数 512，波长分辨率 5 nm，积分时间 80 ms，扫描次数 12 次。在每个样本的不同位置取 3 个采集点以消除绝缘油分布不均匀的影响。最终采集得到的 6 类纸样的代表性光谱及其吸收带归属如图 5 所示，可以看到，各样品的光谱曲线在整体形态上高度一致，但原始光谱数据与 DP 之间并不体现直接相关性，表明样本光谱同时受到样本类型、含水量、温度波动及绝缘油光学吸收特性等非老化关联变量的耦合干扰。



(a) 纸样BZZ-075、BZZ-125、GDL

(b) 纸样JW50、B31A、2HCC

图 4 绝缘纸 DP 随老化时间的变化

Fig.4 Variation of DP of insulating paper with aging time

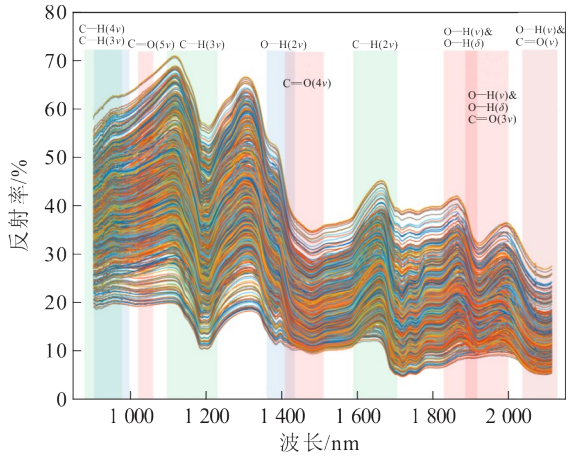


图 5 6 种绝缘纸的近红外光谱及吸收带归属

Fig.5 Near-infrared spectra of six types of insulating papers and absorption band assignments

近红外光谱响应本质上源于分子振转能级跃迁的合频与倍频效应，涉及 C═O、C—H、O—H 等官能团的复杂量子化振动模式。例如，与老化高度相关的羰基在 900~1 050 nm[C═O(5ν)]、1 400~1 500 nm[C═O(4ν)]、1 800~1 900 nm[O—H(ν)&C═O(ν)]、1 900~2 000 nm[C═O(3ν)]以及 2 050~2 110 nm[C—H(ν)&C═O(ν)]多个波段上表现出吸收特征，特征间存在长程多重关联，其谱带展宽与吸收带重叠效应显著增加了特征解析难度[27]。因此，需要构建基于深度学习的多尺度特征挖掘模型，深入解析光谱响应中隐含的老化关联信息。

3 CAT-EOT 模型建立

3.1 模型整体架构

CAT-EOT 模型架构如图 6 所示，包含 3 个重要步骤。(1)采用随机抽样把 NIRS-DP 数据集划分成训练集和测试集，其中训练集将并行进入原始光谱、一阶导及二阶导 3 个微分预处理通道，经随机噪声

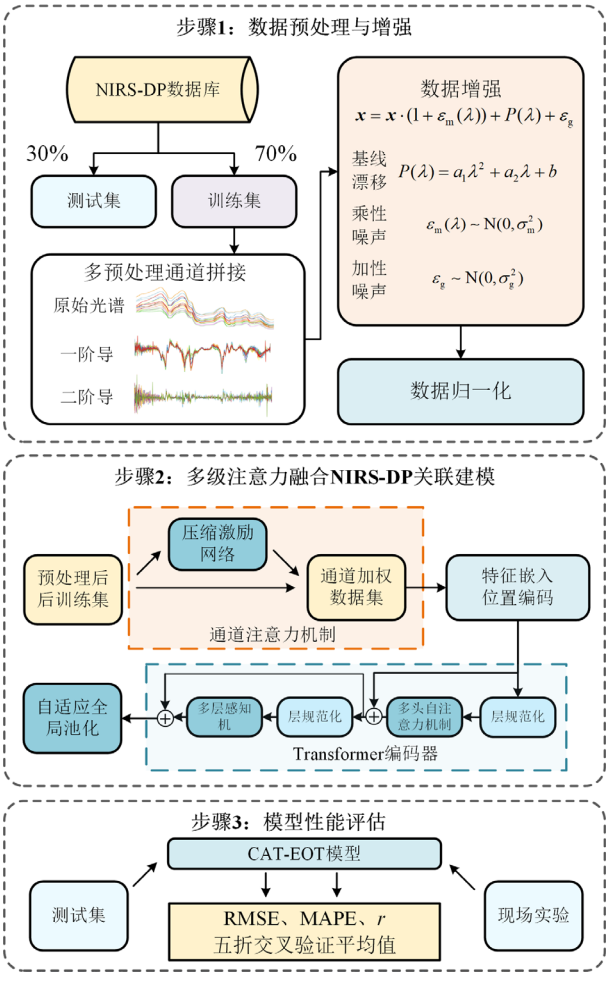


图 6 CAT-EOT 模型整体架构

Fig.6 The architecture of CAT-EOT model

注入与基线漂移模拟实现数据增强，归一化处理后进入模型训练阶段；(2)数据进入通道注意力模块，采用压缩激励网络量化不同预处理模式的特征贡献

度，实现跨通道特征自适应提取；再进入 EOT 核心模块，利用多头自注意力机制捕捉特征间的长程依赖关系，提取 NIRS-DP 间的非线性关联模式，最后通过自适应全局池化层整合 EOT 的输出；(3)应用实验室测试集与现场实验数据对模型性能进行评估。

3.2　数据预处理与样本增强

对 NIRS 微分运算可有效解耦重叠谱峰，突出原始数据背后的关键信息[28]，而直接利用深度学习提取相似的微分光谱特征则需要深层、复杂的网络拓扑结构支持[29]，而捕捉到的相关特征仍是局部的。数据预处理仍是光谱分析中的必要环节。传统预处理策略的选择往往采用穷举式组合筛选，该方法存在计算冗余且受限于先验知识，难以建立普适性的策略组合[30-31]。本研究利用不同微分光谱处理的特点构建互补的光谱信息特征空间，并创新性引入通道注意力机制来调节各微分预处理方法的权重贡献，自动优化预处理策略。对于样本 i 的原始光谱向量 $\boldsymbol{x}_i$，其一阶、二阶微分变换可以表示为：

$$\begin{cases} \boldsymbol{x}_i^{(1)} = \dfrac{\mathrm{d}v_i}{\mathrm{d}\lambda} \\ x_i^{(2)} = \dfrac{\mathrm{d}^2 v_i}{\mathrm{d}^2 \lambda} \end{cases} \tag{2}$$

式中：$\boldsymbol{x}_i^{(1)}$与 $\boldsymbol{x}_i^{(2)}$分别表示样本 i 的一阶与二阶导数，λ 为波长。$\boldsymbol{X}_{\mathrm{comb}}=[\boldsymbol{x}_i; x_i^{(1)}; x_i^{(2)}]$为组合后样品 i 的微分光谱输入。其中，一阶导可以消除基线漂移对羰基等特征吸收带来的影响，二阶导可以进一步放大光谱曲率，分离微弱老化特征[32]。

CAT-EOT 架构的多头自注意力层包含大量参数，使 CAT-EOT 建模要求较高的数据量。本研究设计了基于近红外光谱仪器响应特性的数据增强策略，构建包含多项式基线漂移函数 $P(\lambda)$、乘性噪声 $\varepsilon_{\mathrm{m}}(\lambda)$与加性噪声 ε_{g} 的混合干扰模型：

$$\boldsymbol{x} = \boldsymbol{x}\cdot(1+\varepsilon_{\mathrm{m}}(\lambda)) + P(\lambda) + \varepsilon_{\mathrm{g}} \tag{3}$$

$$P(\lambda) = a_1\lambda^2 + a_2\lambda + b \tag{4}$$

$$\varepsilon_{\mathrm{m}}(\lambda) \sim \mathrm{N}(0, \sigma_{\mathrm{m}}^2) \tag{5}$$

$$\varepsilon_{\mathrm{g}} \sim \mathrm{N}(0, \sigma_{\mathrm{g}}^2) \tag{6}$$

式中：a_1 与 a_2 为随机斜率；b 为随机偏移；σ_{m} 用于控制乘性噪声强度；σ_{g} 用于控制高斯噪声强度；基线漂移函数 $P(\lambda)$模拟光源与环境扰动；乘性噪声 $\varepsilon_{\mathrm{m}}(\lambda)$表征样品随机杂质的光散射效应；加性噪声 ε_{g} 对应光电探测器本底噪声。通过随机选择 a_1、a_2、σ_{m} 与 σ_{g} 即可得到包含混合随机噪声的增强光谱数据，扩展训练样本的物理扰动边界，并满足 CAT-EOT 架构的数据量需求。

3.3　通道注意力机制

本文所构建的 CAT 结构基于压缩—激励网络(squeeze and excitation network, SENet)，可实现不同微分通道的重要性评价[23]，其基本原理如图 7 所示，首先将预处理后的光谱矩阵 $\boldsymbol{X}_{\mathrm{comb}}$ 重构为 $1\times d\times C$ 形式的 $\boldsymbol{X}'_{\mathrm{comb}}$，其中通道维度 C 对应预处理方法个数，光谱特征数为 d。并行进入全局平均池化(global average pooling, GAP)与全局最大池化(global max pooling, GMP)对各通道光谱向量进行池化操作，即 $F_{\mathrm{sq}}(\cdot)$操作

$$x_{\mathrm{c}}^{\mathrm{avg}} = \frac{1}{d}\sum_{i=1}^{d}\boldsymbol{x}_{\mathrm{c}}(j) \tag{7}$$

$$x_{\mathrm{c}}^{\mathrm{max}} = \max_i \boldsymbol{x}_{\mathrm{c}}(i) \tag{8}$$

式中：$\boldsymbol{x}_{\mathrm{c}}$ 表示通道 c 对应的光谱向量；$x_{\mathrm{c}}^{\mathrm{avg}}$ 为 GAP 得到的全局信息平均值；$x_{\mathrm{c}}^{\mathrm{max}}$ 为 GMP 得到的全局最显著值；$x_{\mathrm{c}}^{\mathrm{avg}}$ 与 $x_{\mathrm{c}}^{\mathrm{max}}$ 的数值组合与不同微分通道的特性有关。随后，在特征激励阶段通过多层感知机结合 Sigmoid 函数(F_{sigmoid})和 ReLU 函数(F_{ReLU})的参数化门控机制实现通道权重 s 的学习：

$$\begin{aligned} s = F_{\mathrm{ex}}(x_{\mathrm{c}}^{\mathrm{avg/max}}, \boldsymbol{W}) = \\ F_{\mathrm{Sigmoid}}(\boldsymbol{W}_2 \cdot F_{\mathrm{ReLU}}(\boldsymbol{W}_1 \cdot x_{\mathrm{c}}^{\mathrm{avg/max}})) \end{aligned} \tag{9}$$

式中：$F_{\mathrm{ex}}(\cdot,\boldsymbol{W})$为激励操作函数；$\boldsymbol{W}_1$ 与 $\boldsymbol{W}_2$ 为感知机参数矩阵；$x_{\mathrm{c}}^{\mathrm{avg/max}}$ 为池化操作后的特征信息。该结构通过非线性变换建立 $x_{\mathrm{c}}^{\mathrm{max}}$ 和 $x_{\mathrm{c}}^{\mathrm{avg}}$ 与通道重要性之间的映射关系。最后，通过 $F_{\mathrm{scale}}(\cdot,\cdot)$实现对原数据 $\boldsymbol{X}'_{\mathrm{comb}}$ 在通道维度的乘积加权，即：

$$F_{\mathrm{scale}}(\boldsymbol{X}'_{\mathrm{comb}}, s) = s\cdot\boldsymbol{X}'_{\mathrm{comb}} \tag{10}$$

通过 GAP 和 GMP 的互补性，可以兼顾全局特征的平均值和显著性，克服单一池化的信息损失。利用注意力权重实现训练时特征空间的动态聚焦，有效提升模型对关键预处理操作的感知能力。

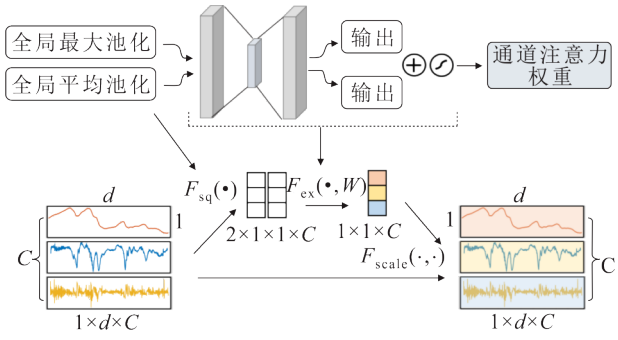


图 7　通道注意力模块的网络结构

Fig.7　Network structure of channel attention module

3.4 Transformer 编码器

本文采用的 EOT 结构基于多头自注意力机制，它能自适应地捕捉光谱数据中不同波长间的长程依赖关系[21]，有效解析近红外光谱中分子振动的耦合效应，如图 8 所示。高维多重相关的光谱信号经过分割与通道一维重构、线性嵌入投影和位置编码后进入编码器。在 EOT 中，光谱信号经层规范化后进入多头自注意力层，经残差连接后输入多层感知机，并输出至下一编码器层。其中，多头自注意力机制通过并行计算多个不同自注意力“头”，关注输入的不同部分或不同特征。对于每个“头”，将输入进行 $\boldsymbol{Q}$、$\boldsymbol{K}$、$\boldsymbol{V}$ 3 个关键的向量变换：

$$\begin{cases}\boldsymbol{Q}=\boldsymbol{x}^{\mathrm{w}}\boldsymbol{W}^{\boldsymbol{Q}}\\ \boldsymbol{K}=\boldsymbol{x}^{\mathrm{w}}\boldsymbol{W}^{\boldsymbol{K}}\\ \boldsymbol{V}=\boldsymbol{x}^{\mathrm{w}}\boldsymbol{W}^{\boldsymbol{V}}\end{cases} \tag{11}$$

式中：$\boldsymbol{Q}$、$\boldsymbol{K}$ 及 $\boldsymbol{V}$ 分别为查询向量、键向量以及值向量；$\boldsymbol{W}^{\boldsymbol{Q}}$、$\boldsymbol{W}^{\boldsymbol{K}}$ 与 $\boldsymbol{W}^{\boldsymbol{V}}$ 是对应的权重矩阵，实现特征空间到注意力子空间的线性变换；$\boldsymbol{x}^{\mathrm{w}}$ 为通道注意力加权后的光谱向量。每个注意力头通过缩放点积运算 F_{softmax} 给出注意力输出为：

$$\boldsymbol{O}=F_{\mathrm{softmax}}\left(\frac{\boldsymbol{Q}\boldsymbol{K}^{T}}{\sqrt{d_{\mathrm{k}}}}\right)\boldsymbol{V} \tag{12}$$

式中：d_{k} 为子空间维度调节因子，用于稳定梯度传播。随后，各注意力头的输出通过拼接算子 F_{Concat} 实现异构特征整合，即：

$$\boldsymbol{M}_{\mathrm{head}}=F_{\mathrm{Concat}}(\boldsymbol{O}_1,\cdots,\boldsymbol{O}_h) \tag{13}$$

式中：h 为注意力头数。该操作通过多头并行计算挖掘光谱互补信息，建立跨序列(光谱波段)的非线性关联。残差连接与层归一化模块协同可有效缓解深度网络训练中的梯度消失问题，保障模型收敛稳定性。最后，所有的编码器输出由自适应全局池化进行全局特征提取，得到最终 CAT-EOT 模型。

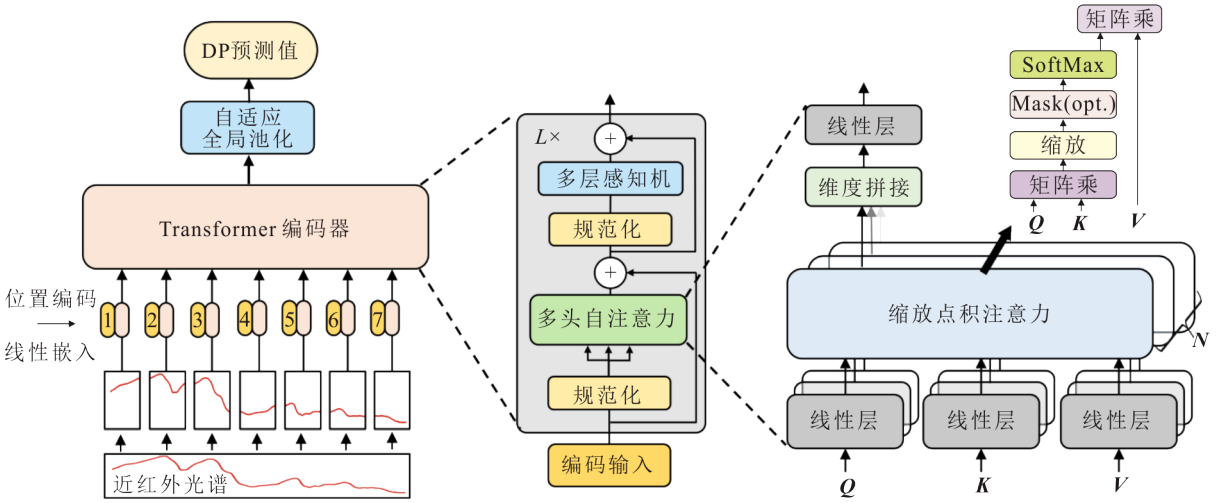


图 8 基于多头自注意力机制的 EOT 架构

Fig.8 EOT architecture based on multi-head self-attention mechanism

3.5 模型训练

本文提出模型的训练基于 Windows10 下的 Python 3.9 平台与 Pytorch 2.5.1 深度学习框架。输入数据结构为(937,512)的光谱数据与(937,1)的 DP 标签数据，训练集比例为 70%，测试集比例为 30%。在数据增强模块中，设置 0~100 等差递增(公差为 10)的数据扩增倍数，以分析模型训练收敛速度与性能变化，基线漂移系数范围为 $a_1\in[-10^{-4},10^{-4}]$、$a_2\in[-10^{-3},10^{-3}]$、$b\in[-10^{-1},10^{-1}]$，乘性噪声与高斯噪声方差符合均匀分布 $\sigma\sim U[0.5,1]$。EOT 的序列数为 64，嵌入维度为 48，编码器层数 L=4，注意力头数 N=4，编码器，前馈层维度设置为 128，参数规模中等。应用 Adam 优化器，初始学习率为 10^{-4}，使用 ReduceLROnPlateau 策略动态调整学习率。

模型性能评估基于 3 个互补性指标：均方根误差(root mean square error，RMSE)，计算方法见式(14)；平均相对误差(mean absolute percentage error, MAPE)，计算方法见式(15)；相关系数 r，计算方法见式(16)。

$$E=\sqrt{\frac{\sum_{i=1}^{n_{\mathrm{t}}}(y_{\mathrm{t}}^{i}-y_{\mathrm{pre}}^{i})^{2}}{n_{\mathrm{t}}}} \tag{14}$$

$$F=\frac{1}{n_{\mathrm{t}}}\sum_{i=1}^{n_{\mathrm{t}}}\left|\frac{y_{\mathrm{t}}^{i}-y_{\mathrm{pre}}^{i}}{y_{\mathrm{t}}^{i}}\right|\times 100\% \tag{15}$$

$$r=\frac{\sum_{i=1}^{n_{\mathrm{t}}}(y_{\mathrm{t}}^{i}-\bar{y}_{\mathrm{t}})(y_{\mathrm{pre}}^{i}-\bar{y}_{\mathrm{pre}})}{\sqrt{\sum_{i=1}^{n_{\mathrm{t}}}(y_{\mathrm{t}}^{i}-\bar{y}_{\mathrm{t}})^{2}}\cdot\sqrt{\sum_{i=1}^{n_{\mathrm{t}}}(y_{\mathrm{pre}}^{i}-\bar{y}_{\mathrm{pre}})^{2}}} \tag{16}$$

式中：y_{t}为测试集 DP 标签值；$\bar{y}_{\mathrm{t}}$为其均值；y_{pre}为模型对应的预测；$\bar{y}_{\mathrm{pre}}$为其均值；n_{t}为测试集样本数量。RMSE 越小，MAPE 越小，r 越接近于 1，说明模型表现越好。本文采用五折交叉验证方法评估模型性能，并分别对每折计算的 RMSE、MAPE 和 r 值取平均值，作为最终性能指标。

4　结果与分析

4.1　预处理与数据增强效果分析

不同数据增强倍数对模型收敛性的影响如图 9 所示。建模实验表明，数据增强倍数的提高可加速 CAT-EOT 模型收敛，但对最终的模型准确度基本不产生影响。因此，综合权衡训练效率与计算成本，本文选择数据增强倍数为 30，该参数配置在保证模型泛化能力的同时提高了计算效率。

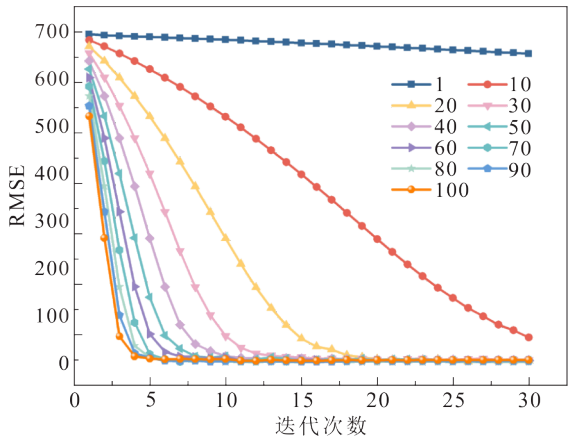


图 9　不同数据增强倍数下模型的收敛特性

Fig.9　Convergence of model training under different data augmentation multiples

预处理方式显著影响着模型的训练结果，本文对比了原始光谱(RAW)、一阶导数(1D)光谱与二阶导数(2D)光谱直接建模，以及维度拼接和通道拼接建模等 6 种预处理方式下的模型性能差异，如图 10 所示。相较 RAW 直接建模(RMSE=74.8)，1D 处理后模型 RMSE 减小至 56.1，而 2D 处理后模型性能无提升。RAW+1D 的维度拼接模型的 RMSE=51.6，说明 RAW 结合 1D 光谱可以提供更多的信息，使模型误差进一步降低。若继续拼接 2D 处理(RAW+1D+2D)则使 RMSE 升高为 76.5，这是由于维度过大导致模型的局部特征提取能力下降，同时引入大量冗余信息使有效信息被稀释。而通道拼接则契合 CAT-EOT 的通道信息映射方式，在保留有效信息的前提下，通过 CAT 提取通道权重，增加模型对重要通道的关注，最终模型的 RMSE 为 45.6，为所有预处理策略最优结果。

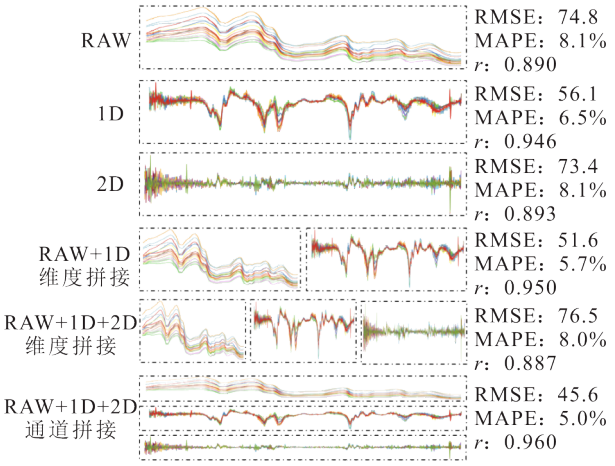


图 10　不同预处理策略下的模型性能

Fig.10　Model performance under different preprocessing strategies

通道注意力机制可自适应计算每个通道的权重，通道权重可以理解为每个预处理方式的贡献。图 11 展示了模型迭代过程中 3 个通道的权重变化。1D 通道的权重逐渐增加，最终稳定在(0.83±0.1)区间，其余两通道的权重逐渐降低。可见，对建模来说 1D 光谱的特征比 RAW 和 2D 光谱更为重要。

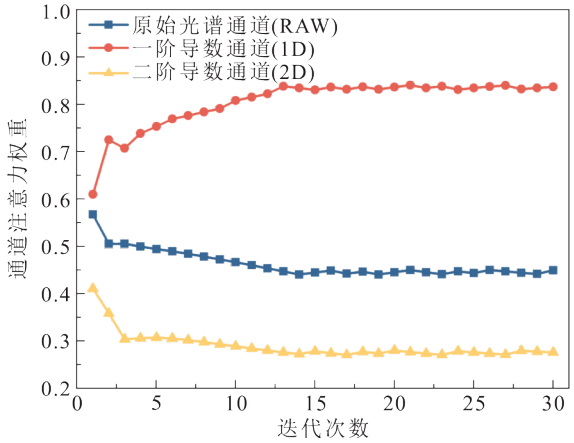


图 11　通道权重随迭代次数的变化

Fig.11　Variation of channel weights as a function of iterations

4.2 注意力权重分析

在EOT训练过程中，每8个波长点为一个序列，共64个序列，通过查询向量($\boldsymbol{Q}$)与键向量($\boldsymbol{K}$)的点积衡量相似度，缩放后经Softmax归一化，得各输入序列对当前输出的重要性分布，如图12所示。发现测试集的平均注意力权重在902、961、1 205、1 460、1 950及2 080 nm呈现显著峰值。通过与图5交叉对比可见，高权重序列对应的波长大多位于C—H及C═O两种基团的特征吸收区[27]。值得注意的是，O—H吸收带并不受到注意力重点关注，这可能是由于绝缘纸样本易吸收环境水分，与自身O—H基团的光谱响应产生叠加，干扰对老化状态的量化。通过注意力权重分析可知，大部分序列权重约为0，说明EOT有效地将宽吸收带的共线性序列特征聚焦在单一序列上，在频带混叠的干扰下，准确提取光谱特征的长程关联，将注意力集中于与老化有关的基团光谱响应上。

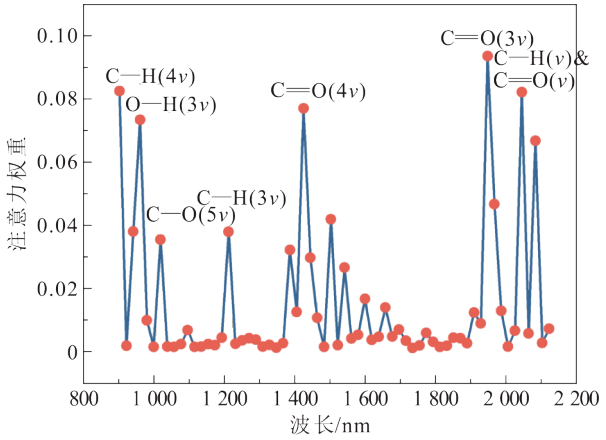


图12 不同序列对应的注意力权重

Fig.12 Attention weights corresponding to different sequences

4.3 不同模型性能对比

为验证CAT-EOT模型的先进性，构建了包含传统机器学习与深度学习的7种对比模型，分别为PLS、GPR、SVR、1D-CNN、Transformer、Vision Transformer以及CAT-EOT。各模型超参数均经过优化，其中，由于PLS、GPR与SVR没有多通道及大规模数据处理能力，这3个模型只优选预处理方式，但不进行数据增强。4种深度学习模型除网络架构不同外均采用相同的预处理方式，最终各模型参数设置如表1所示。对比模型也采用五折交叉验证方式评估其表现。

各模型在测试集上的预测结果如图13所示。结果表明，基于注意力机制的模型在近红外光谱建模中展现出显著优势，CAT-EOT模型预测准确度最高，RMSE=45.6、MAPE=5.01%、r=0.960。相较于其他经典的机器学习方法，RMSE分别降低了20.1、18.7与12.5。对于深度学习模型，1D-CNN模型的精度最差，该框架无法适应近红外光谱数据长程多重相关的特性；CAT-EOT模型准确度高于原始Transformer和VIT模型，说明通道注意力机制可以更有效地提取不同预处理方式中所隐含的关键信息。

表1 各模型的参数设置

Table 1 Parameter setting of each model

方法	设置
PLS	主成分数=20
GPR	径向基核 C=31.6，ε=0.869
SVR	径向基核 C=92 004，ε=15.29
1D-CNN	Inception模块3个并行卷积层 (1)卷积核大小为3，输出通道为4 (2)卷积核大小为5，输出通道为4 (3)最大池化后卷积，输出通道为4 后接2个全连接层
Transformer	设置与3.5节相同，不设置通道注意力模块
VIT	设置与3.5节相似，不设置通道注意力模块
CAT-EOT	设置见3.5节

4.4 现场实验验证

某220 kV主变于2001年12月投入运行，2024年9月例行运检中发现高压套管C相氢气含量超标(1 179 μL/L)，超出注意值150 μL/L，同时中压套管B相低频介损超出注意值10%，怀疑套管内部绝缘出现劣化故障。为进一步查明氢气和介损超标原因，对高压C相套管与中压B相套管进行解体检查。应用近红外光谱分析技术现场快速测试绝缘纸DP，现场测试方式与测试结果如图14所示。每隔10个电容屏采集一次光谱，每次选择3个采集点，并裁取纸样用于黏度法测试。注意，所采集的光谱(图14(b))在900~1 100 nm处存在较多噪声，这与实验室标准测试条件下的结果有较大差异。

为直观展示CAT-EOT的特征提取挖掘能力，采用主成分分析(principal component analysis，PCA)、Isomap及t-SNE对比分析了原始数据与编码器输出数据(前馈网络及全局池化之前)的降维效果，如图15所示。3种降维方法结果均表明，对于原始数据，现场实验结果与训练集数据在降维后具有较大差异。这是因为套管为少油设备，现场套管纸样的含油量与实验室纸样有所差异，且由于长时间暴露，样品一定程度受潮，导致光谱产生偏移。而编码器处理后的现场数据与建模数据在特征空间

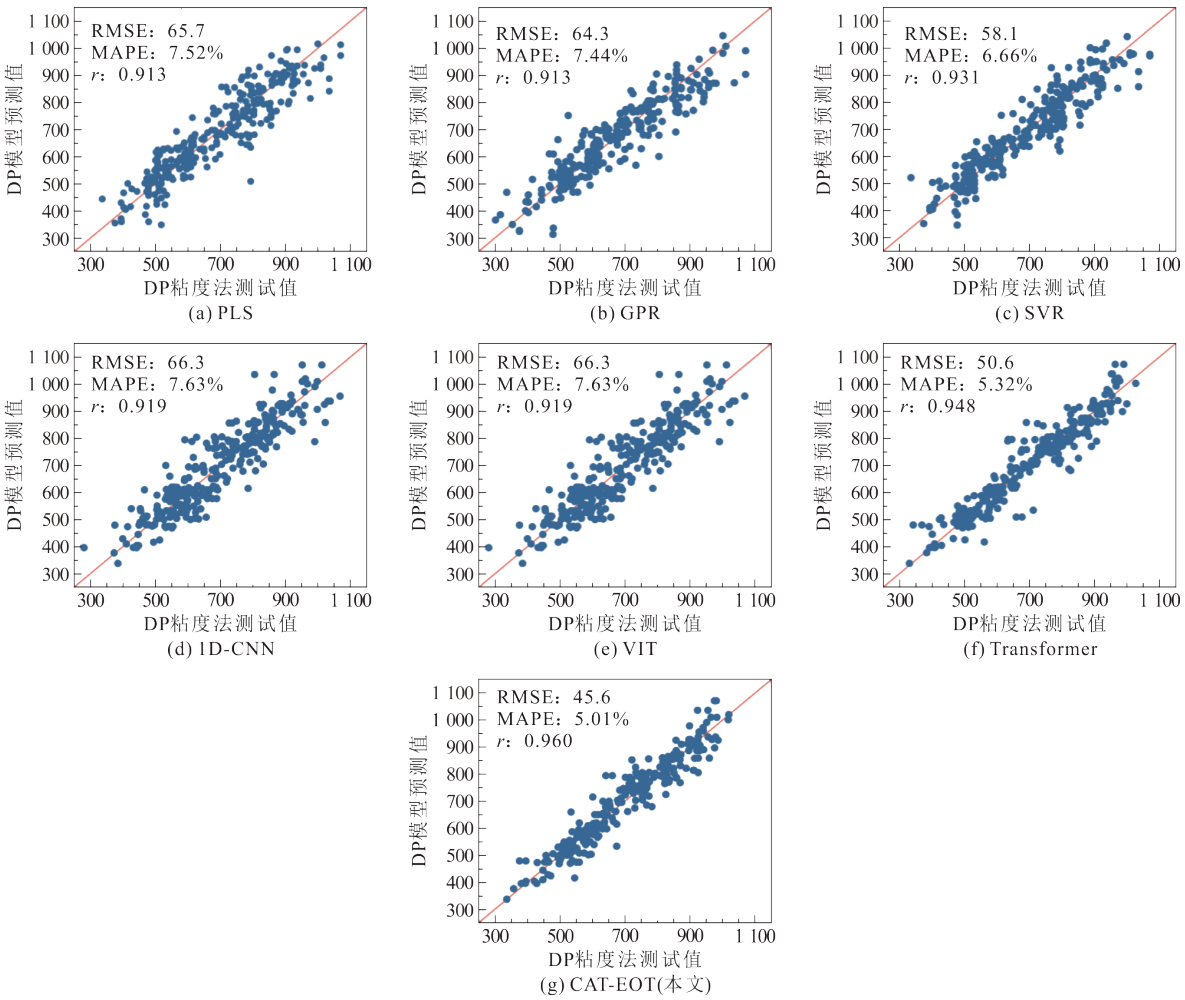


图 13　7 种模型在测试集上的预测性能比较

Fig.13　Comparison of performance of the seven models on the test set

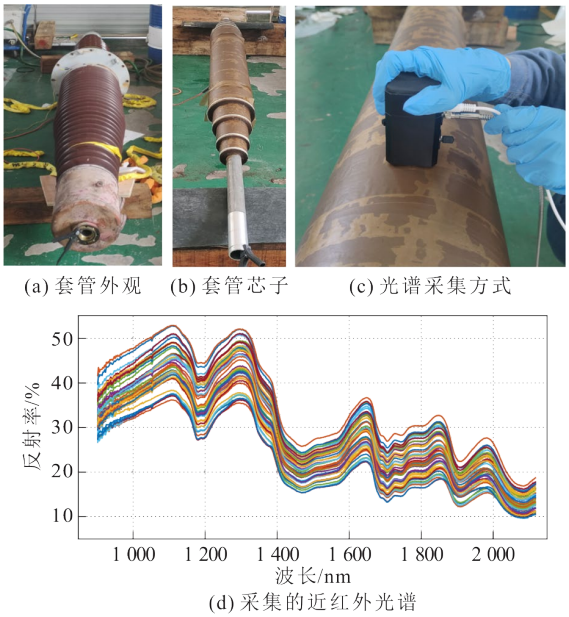


图 14　220 kV 变压器套管绝缘纸近红外光谱现场测试

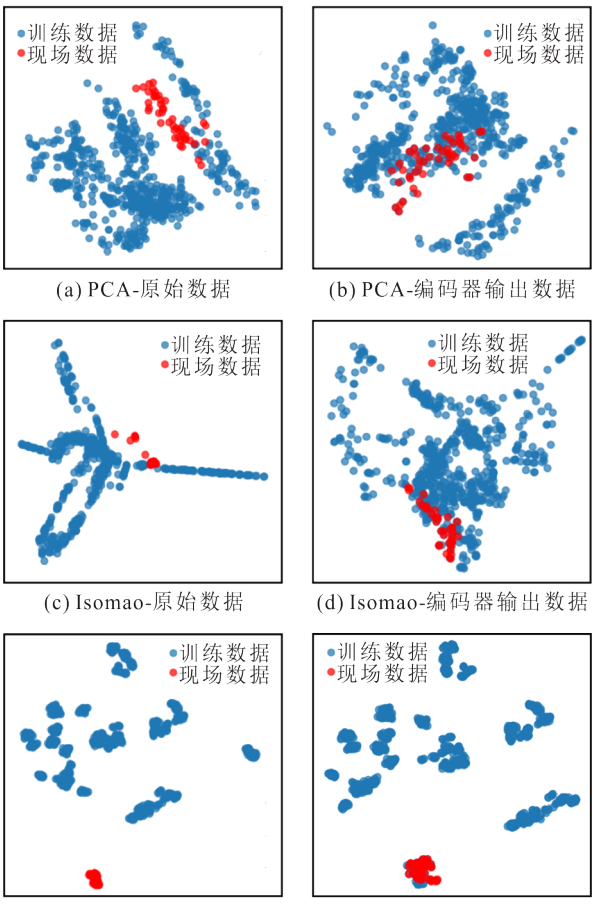
Fig.14　On-site NIR spectral testing of insulating paper in 220 kV transformer bushings

中重合，说明 CAT-EOT 准确挖掘了光谱中与 DP 相关的数据特征，展现出对样品差异及仪器响应差异的鲁棒性修正能力。最终的测试结果如表 2 所示，对 8 个测量点 DP 的预测误差均小于 5%，其中 3 个采样点预测精度高于 1%，验证了 CAT-EOT 模型在实际工程场景中优异的泛化能力。

表 2　现场测试中 CAT-EOT 模型的 DP 预测结果

Table 2　DP prediction of CAT-EOT model in field test

位置	DP 测试值 (黏度法)	DP 预测值 (CAT-EOT)	相对误差/%
110-1	876	869.67	0.72
110-3	1 087	1 081.21	0.53
110-5	1 028	1 001.79	2.54
110-7	847	860.26	1.56
220-1	991	972.21	1.89
220-3	1 008	975.46	3.22
220-5	991	997.74	0.68
220-7	918	942.80	2.70



(a) PCA-原始数据 (b) PCA-编码器输出数据

(c) Isomao-原始数据 (d) Isomao-编码器输出数据

(e) t-SNE-原始数据 (f) t-SNE-编码器输出数据

图 15 原始数据与编码器输出数据的降维分析结果

Fig.15 Dimensionality reduction analysis results of the original data and the encoder output data

5 结论

1）多通道微分预处理可有效解耦重叠光谱特征峰，通道权重分析表明，一阶微分光谱在老化特征表征中具有主导作用(通道权重约 0.83)。通道注意力机制的引入实现了光谱多尺度特征的挖掘与融合，使模型预测的 MAPE 较传统原始光谱直接建模降低 3.1%。

2）CAT-EOT 模型 DP 预测精度相较于经典的机器学习方法(PLS、SVR 与 GPR)提升明显(RMSE 平均降低 20%)，在所有对比模型中表现最优。实验验证表明，编码器处理后的现场数据与训练数据间具有更高的特征空间匹配度，现场测试误差小于 5%。该方法实现了面向复杂纸样数据库的光谱序列长程相关性提取与 NIRS-DP 深度关联模式挖掘。

本文从近红外光谱预处理与建模两方面增强了 DP 评估模型的精度和泛化能力，突破了现有方法的应用局限。未来还应关注基于迁移学习的多油品适应性提升和克服不同光谱仪台间差异的模型传递方法研究。

参考文献 References

[1] 李立涅，罗兵，董旭柱，等. 高电压技术发展现状及展望[J]. 高电压技术，2025，51(8)：3682-3720.
LI Licheng, LUO Bing, DONG Xuzhu, et al. Advances and prospects of high voltage technology[J]. High Voltage Engineering, 2025, 51(8): 3682-3720.

[2] 梁一霖，张明泽，刘骥，等. 复杂运行条件下非均匀老化油纸绝缘典型电场变化特性[J]. 高电压技术，2024，50(4)：1731-1741.
LIANG Yilin, ZHANG Mingze, LIU Ji, et al. Variation characteristics of typical electric field of non-uniform aging oil-paper insulation under complex operating conditions[J]. High Voltage Engineering, 2024, 50(4): 1731-1741.

[3] 国家能源局. 油浸式变压器绝缘老化判断导则：DL/T 984—2018[S]. 北京：中国电力出版社，2018.
National Energy Administration. Guide for the diagnosis of insulation aging in oil-immersed power transformer: DL/T 984—2018[S]. Beijing, China: China Electric Power Press, 2018.

[4] ZHOU L J, ZHANG J, LI Z X, et al. Degree of polymerisation estimation for insulation paper immersed in silicone oil using frequency domain dielectric spectroscopy[J]. High Voltage, 2023, 8(5): 1041-1054.

[5] 江长明，郝震，唐云鹏，等. 变压器绝缘故障发展规律及主动防御技术[J]. 高电压技术，2024，50(10)：4338-4348.
JIANG Changming, HAO Zhen, TANG Yunpeng, et al. Insulation fault development laws and active defense technology for transformer[J]. High Voltage Engineering, 2024, 50(10): 4338-4348.

[6] BAIRD P J, HERMAN H, STEVENS G C. On-site analysis of transformer paper insulation using portable spectroscopy for chemometric prediction of aged condition[J]. IEEE Transactions on Dielectrics and Electrical Insulation, 2008, 15(4): 1089-1099.

[7] ALI M, EMSLEY A M, HERMAN H, et al. Spectroscopic studies of the ageing of cellulosic paper[J]. Polymer, 2001, 42(7): 2893-2900.

[8] LI Y, ZHANG Y, ZHANG W B, et al. Comparative study of partial least squares and neural network models of near-infrared spectroscopy for aging condition assessment of insulating paper[J]. Measurement Science and Technology, 2020, 31(4): 045501.

[9] LI Y, ZHANG W B, LI H, et al. Aging assessment of oil-impregnated-paper electrical equipment via near infrared spectroscopy powered by improved PCA-RBF-NN: modelling and field practices[J]. IEEE Transactions on Dielectrics and Electrical Insulation, 2021, 28(6): 2035-2042.

[10] LI Y, LI H, ZHANG W B, et al. Gaussian process regression for quantitative degree of polymerization analysis of oil-paper insulation by EPO-NIRS: a solution to field moisture[J]. IEEE Transactions on Industry Applications, 2023, 59(3): 2750-2759.

[11] 马恒瑞，袁傲添，王波，等. 基于深度学习的负荷预测研究综述与展望[J]. 高电压技术，2025，51(3)：1233-1250.
MA Hengrui, YAUN Aotian, WANG Bo, et al. Review and prospect of load forecasting based on deep learning[J]. High Voltage Engineering, 2025, 51(3): 1233-1250.

[12] MISHRA P, PASSOS D, MARINI F, et al. Deep learning for near-infrared spectral data modelling: hypes and benefits[J]. TrAC Trends in Analytical Chemistry, 2022, 157: 116804.

[13] DEBUS B, PARASTAR H, HARRINGTON P, et al. Deep learning in analytical chemistry[J]. TrAC Trends in Analytical Chemistry, 2021, 145: 116459.

[14] YANG J R, WANG J, LU G D, et al. TeaNet: deep learning on Near-Infrared Spectroscopy (NIR) data for the assurance of tea quality[J]. Computers and Electronics in Agriculture, 2021, 190: 106431.

[15] WANG X, YU L, TIAN S W, et al. End-to-end analysis modeling of vibrational spectroscopy based on deep learning approach[J]. Journal of Chemometrics, 2020, 34(10): e3291.

[16] NG W, MINASNY B, MENDES W D S, et al. The influence of training sample size on the accuracy of deep learning models for the prediction of soil properties with near-infrared spectroscopy data[J]. Soil, 2020, 6(2): 565-578.

[17] SUN Q, ZHAO G R, XIA X Y, et al. Hyperspectral image classification based on multi-Scale convolutional features and multi-attention mechanisms[J]. Remote Sensing, 2024, 16(12): 2185.

[18] KIRANYAZ S, AVCI O, ABDELJABER O, et al. 1D convolutional neural networks and applications: a survey[J]. Mechanical Systems and Signal Processing, 2021, 151: 107398.

[19] VASWANI A, SHAZEER N, PARMAR N, et al. Attention is all you need[C]//Proceedings of the 31st International Conference on Neural Information Processing Systems. Long Beach, USA: Curran Associates Inc., 2017: 6000-6010.

[20] 金涛，王万豪，黄钦瑜，等. 一种利用轻量级多通道注意力融合网络的窃电检测边缘计算方法[J]. 中国电机工程学报，2026，46(6)：2420-2433.
JIN Tao, WANG Wanhao, HUANG Qinyu, et al. Edge computing approach for electricity theft detection utilizing lightweight multi-channel attention fusion network[J]. Proceedings of the CSEE, 2026, 46(6): 2420-2433.

[21] 郑尚坡，刘俊峰，曾君，等. 基于多尺度注意力自适应去噪网络的局部放电模式识别[J]. 高电压技术，2025，51(4)：1958-1968.
ZHENG Shangpo, LIU Junfeng, ZENG Jun, et al. Partial discharge pattern recognition based on multi-scale attention adaptive denoising network[J]. High Voltage Engineering, 2025, 51(4): 1958-1968.

[22] LI L Q, LU F, WANG Z J, et al. A transformer-based model for quantitative analysis of near-infrared spectra[EB/OL]. 2024,[2026-06-22]. https://www.aminer.cn/pub/65ffa61e13fb2c6cf6f24435/a-transformer-based-model-for-quantitative-analysis-of-near-infrared-spectra.

[23] CAO L Y, SUN M, YANG Z C, et al. A novel transformer-CNN approach for predicting soil properties from LUCAS Vis-NIR spectral data[J]. Agronomy, 2024, 14(9): 1998.

[24] 蔡兆龙，张建良，孟宏宇，等. 面向小样本数据的电机转子振动故障诊断研究[J]. 中国电机工程学报，2025，45(6)：2378-2389.
CAI Zhaolong, ZHANG Jianliang, MENG Hongyu, et al. Research on vibration fault diagnosis of motor rotor under small quantity of samples[J]. High Voltage Engineering, 2025, 45(6): 2378-2389.

[25] 孟高军，吴佳凤，张仰飞，等. 基于双重注意力机制与时空门控融合的风电集群功率短期预测[J]. 高电压技术，2026，52(2)：663-673.
MENG Gaojun, WU Jiafeng, ZHANG Yangfei, et al. Short-term wind power prediction for wind farm clusters based on dual attention mechanism and spatio-temporal gated fusion[J]. High Voltage Engineering, 2026, 52(2): 663-673.

[26] SARROUTI M, TAO C, RANDRIAMIHAJA Y M. Comparing encoder-only and encoder-decoder transformers for relation extraction from biomedical texts: an empirical study on ten benchmark datasets[C]//Proceedings of the 21st Workshop on Biomedical Language Processing. Dublin, Ireland: IEEE, 2022: 376-382.

[27] WORKMAN J JR, WEYER L. Practical guide and spectral atlas for interpretive near-infrared spectroscopy[M]. 2nd ed. Boca Raton, USA: CRC Press, 2012: 231-249.

[28] MOKARI A, GUO S X, BOCKLITZ T. Exploring the steps of infrared (IR) spectral analysis: pre-processing, (classical) data modelling, and deep learning[J]. Molecules, 2023, 28(19): 6886.

[29] ZHANG J W, LIU T L, TAO D C. Going deeper, generalizing better: an information-theoretic view for deep learning[J]. IEEE Transactions on Neural Networks and Learning Systems, 2024, 35(11): 16683-16695.

[30] RINNAN Å, BERG F V D, ENGELSEN S B. Review of the most common pre-processing techniques for near-infrared spectra[J]. TrAC Trends in Analytical Chemistry, 2009, 28(10): 1201-1222.

[31] 张朱珊莹，顾瀚文，谢凯文，等. 基于近红外光谱的预处理及组合方法[J]. 激光与光电子学进展，2021，58(16)：1617001.
ZHANG Zhushanying, GU Hanwen, XIE Kaiwen, et al. Pretreatment and combined method based on near infrared spectroscopy[J]. Laser & Optoelectronics Progress, 2021, 58(16): 1617001.

[32] SHAO X, CUI X, WANG M, et al. High order derivative to investigate the complexity of the near infrared spectra of aqueous solutions[J]. Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy, 2019, 213: 83-89.



李　含
1998—，男，博士生
主要从事基于近红外光谱的绝缘材料老化状态无损评估技术研究
E-mail: lihan0412@stu.xjtu.edu.cn

LI Han
Ph.D. candidate



李　元(通信作者)
1984—，男，博士，副教授，博导
主要从事高电压试验技术、电力设备状态感知与智能诊断等研究
E-mail: liyuan8490@xjtu.edu.cn

LI Yuan
Ph.D.
Associate professor
Corresponding author

收稿日期　2025-04-28　在线出版日期　2025-11-17　　编辑　曹昭君