基于可解释 LightGBM 的电动汽车充电站入侵检测方法*

姚沁怡，龙莆均，陈世伦
（重庆科技大学 数理科学学院，重庆 401331）

摘 要：在电动汽车充电站（EVCS）网络安全问题日益严峻的背景下，传统入侵检测方法存在诸多不足，机器学习和深度学习虽有成效但存在“黑箱”问题。提出一种基于可解释轻量级梯度提升机（LightGBM）的 EVCS 入侵检测框架。利用 SHAP 进行特征选择，经模拟退火算术优化算法（SAOA）对 LightGBM 超参数调优，集成 SHAP、LOCO、CEM、PFI、LIME 和 ALE 等多种可解释性人工智能（XAI）技术。在 CICEVSE2024 和 Edge-IIoTset 数据集上的实验结果显示，模型检测准确率分别达 97.53% 和 88.89%，*F*1 分数分别为 98.01% 和 88.98%，且可解释性强，能为安全运维提供清晰依据。该研究为提升 EVCS 网络安全提供了高效、可解释的解决方案，具有重要的理论与实践意义。

关键词：电动汽车充电站；入侵检测；轻量级梯度提升机；可解释人工智能
中图分类号：TN915；TP309　　**文献标识码：**A　　**DOI：**10.19358/j.issn.2097-1788.2025.05.001
引用格式：姚沁怡，龙莆均，陈世伦. 基于可解释 LightGBM 的电动汽车充电站入侵检测方法［J］. 网络安全与数据治理，2025，44(5)：1-9，16.

Intrusion detection method for electric vehicle charging station based on interpretable lightweight gradient boosting machine

Yao Qinyi，Long Pujun，Chen Shilun
（School of Mathematics and Big Data，Chongqing University of Science and Technology，Chongqing 401331，China）

Abstract：Against the backdrop of increasingly severe cybersecurity challenges in Electric Vehicle Charging Stations（EVCS），traditional intrusion detection methods exhibit multiple limitations，while machine learning and deep learning approaches，despite their effectiveness，suffer from "black-box" issues. This paper proposes an interpretable Lightweight Gradient Boosting Machine（LightGBM）-based intrusion detection framework for EVCS. The framework employs SHAP for feature selection and utilizes a Simulated Annealing Arithmetic Optimization Algorithm（SAOA）to optimize LightGBM hyperparameters，while integrating multiple Explainable Artificial Intelligence（XAI）techniques including SHAP，LOCO，CEM，PFI，LIME，and ALE. Experimental results on the CICEVSE2024 and Edge-IIoTset datasets demonstrate that the model achieves detection accuracies of 97.53% and 88.89%，with *F*1-scores of 98.01% and 88.98% respectively，while maintaining strong interpretability to provide clear decision-making basis for security operations. This research offers an efficient and interpretable solution for enhancing EVCS cybersecurity，with significant theoretical and practical implications.

Key words：electric vehicle charging stations；intrusion detection；LightGBM；XAI

0 引言

随着电动汽车产业的迅猛发展，电动汽车充电站（Electric Vehicle Charging Stations，EVCS）作为连接电网与终端用户的关键基础设施，正不断向智能化、网络化方向演进。EVCS 不仅集成了通信、控制、计费等多功能模块，还普遍采用 OCPP、IEC 61850、IEC 15118 等标准协议以实现设备间的互联互通。然而，这种开放性和标准化所带来的便利，也暴露出严峻的网络安全隐患，如远程操控、恶意数据注入、服务中断及隐私泄露等问题。传统的基于签名匹配的入侵检测系统（如 Snort、Suricata）难以识别未知攻击，且维护成本高、响应延迟大，难

*基金项目：重庆市教委科学技术研究项目（KJQN202101536）；重庆科技大学硕士研究生创新计划项目（YKJCX2321110）

以满足 EVCS 实时、高可靠的安全需求。因此，研究更具智能性、自适应性与前瞻性的入侵检测方法成为保障 EVCS 网络安全的核心课题。

近年来，基于机器学习与深度学习的入侵检测系统（Intrusion Detection System，IDS）因其对异常行为的建模能力，在 EVCS 安全防护中取得显著成果。例如，Akanda 等[1]利用逻辑回归与随机森林建模 EVCS 网络数据，有效识别静态攻击行为；Makhmudov 等[2]结合自适应随机森林与数据漂移检测算法 ADWIN，增强了模型的流式响应能力；Tulsiani 等[3]系统评估多种分类器的性能，验证了机器学习方法在鲁棒性与泛化性方面的优越性。在深度学习方面，Kilichev 等[4]基于 Edge-IIoTset 构建 CNN、LSTM 与 GRU 融合的 NIDS 架构，实现对多类攻击的高效识别；Basnet 等[5]设计了 LSTM 模型用于 5G 架构下的 DDoS 与 FDI 攻击检测；Miskin 等[6]在 CICIDS2017 数据集上实现了 LSTM 准确率 99.98%；Almadhor 等[7]引入迁移学习策略，有效缓解了不同数据分布下模型性能波动的问题。

尽管深度学习模型表现出强大的检测性能，但其高度非线性结构也带来了“黑箱”问题，导致决策过程难以理解和信任，尤其在安全关键系统中限制了其部署落地。具体而言，这类模型在特征空间中往往缺乏显式的逻辑关联解释，用户难以追溯模型是如何做出判断的。此外，模型性能提升通常伴随着结构复杂度增加，进一步加剧了解释性与性能之间的权衡矛盾[8-9]。在实际应用中，运营人员不仅需要知道“是否发生攻击”，还需了解“为何发生”“哪些特征导致异常”，以支持策略制定和系统防御优化。因此，提升模型的可解释性成为深度学习 IDS 发展的关键方向。

为此，可解释人工智能（Explainable Artificial Intelligence，XAI）方法被逐步引入 EVCS 入侵检测研究中。相关研究如 Khan 等[10]探讨 XAI 在提升 IDS 可信度方面的理论基础；Arreche 等[11]提出基于 XAI 的特征选择框架，增强关键变量识别能力；Attique 等[12]将联邦学习与 SHAP 结合，构建兼具隐私保护与可解释性的 IDS；Mohanty 等[13]则结合 GAN 合成数据与 LightGBM + SHAP 实现电动汽车负载预测中的特征贡献分析；Rahman 等[14]基于 CNN-LSTM 检测架构，引入 SHAP 解释获得了高准确率与良好可理解性。

尽管当前 XAI 方法已在部分研究中展现出前景，但仍面临若干挑战：其一，多数研究仅应用单一解释方法，缺乏对多种 XAI 方法的对比与协同机制的深入探讨；其二，特征选择、模型调参与解释方法之间尚未形成统一的集成框架，降低了解释稳定性与部署可行性；其三，现有研究多以解释性为附加指标，未系统评估解释质量对模型性能及安全运维决策的反馈价值。

针对上述问题，本文提出一种基于轻量级梯度提升机（Light Gradient Boosting Machine，LightGBM）的可解释 EVCS 入侵检测框架，通过 SAOA 算法优化模型超参数，并集成六种 XAI 技术（SHAP、LOCO、PFI、CEM、LIME、ALE）对模型预测机制和特征重要性进行多维度解释。同时，利用混淆矩阵、特征重要性对比分析等指标，展现模型在多攻击场景下的实用性与可部署性，为 EVCS 网络安全提供可信且透明的智能防护方案。

1　EVCS 入侵检测模型

本文首先对所选择的数据集进行了预处理，确保其质量和模型稳定性。基于轻量级梯度提升机构建 EVCS 入侵检测模型，采用模拟退火算术优化算法进行超参数寻优，以提升检测性能和泛化能力。为增强模型可解释性，综合运用 SHAP、LOCO、CEM、PFI、LIME 和 ALE 等多种 XAI 技术，深入分析特征重要性、决策逻辑及局部特征贡献。这种方法不仅提高模型可信度，还为 EVCS 网络安全提供透明高效的防护策略，支持安全运维决策。入侵检测技术框架如图 1 所示。

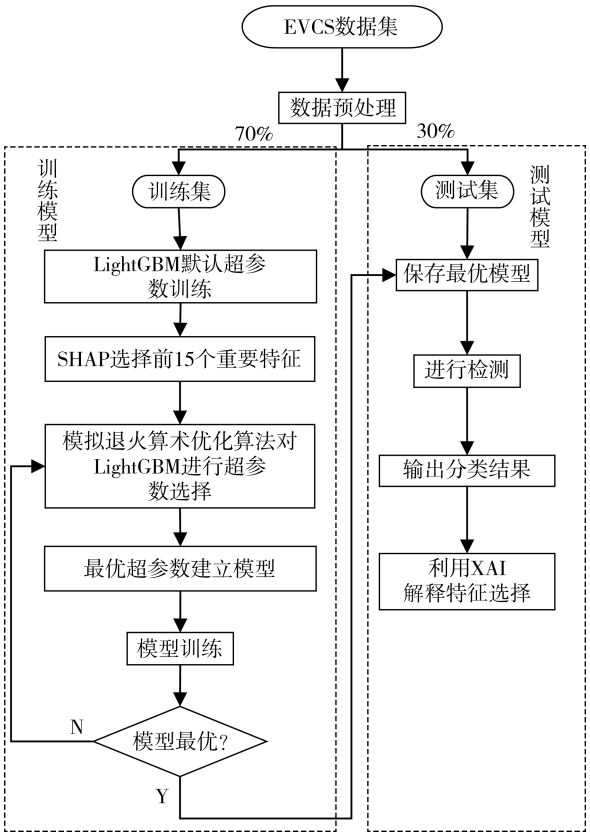


图 1　入侵检测框架图

1.1 LightGBM

LightGBM 是一种基于梯度提升框架的高效决策树算法，用于大规模数据处理优化。其预测函数形式如下：

$$f(x) = \sum_{q=1}^{Q} \alpha_q T_q(x) \tag{1}$$

其中，$f(x)$ 表示最终的预测函数，Q 表示弱学习器的数量，α_q 表示第 q 个弱学习器的权重，x 是训练样本，$T_q(x)$ 表示第 q 个弱学习器的预测结果。

在确定模型的损失函数和训练数据后，Boosting 算法的训练过程就转变为一个优化问题，目标是最小化损失函数。目标函数如下：

$$\arg\min \sum_{h=1}^{H} L(y_h, f(x_h)) \tag{2}$$

其中，H 是样本的数量，h 是样本索引，$f(x_h)$ 是与第 h 个样本对应的目标值，$L(y_h, f(x_h))$ 是第 h 个样本的损失函数值。

作为一种基于梯度下降算法的提升树模型，GBM 在每次添加一个新子模型后，所选损失函数会不断朝具有下一个最高信息含量的变量的梯度方向减少，如式（3）所示：

$$L(F_j(x), Y) < L(F_{j-1}(x), Y) \tag{3}$$

其中，$L(F_j(x), Y)$ 和 $L(F_{j-1}(x), Y)$ 分别是第 j 次和第 $j-1$ 次迭代的损失函数值，$F_j(x)$ 和 $F_{j-1}(x)$ 分别是第 j 次和第 $j-1$ 次样本对应的目标值，Y 是样本的真实目标值。

相较于传统 GBM，LightGBM 采用基于叶子优先的树结构增长策略与基于直方图的特征分裂方法，显著提升了计算效率并降低内存消耗，有效缓解过拟合问题，尤其适用于高维、大规模数据场景。

1.2 模拟退火算术优化算法选择 LightGBM 超参数

尽管 LightGBM 在各领域表现出色，但其超参数优化仍是一个挑战，传统网格搜索方法存在一定的缺陷。为提升优化效率和适应性，本文引入基于模拟退火的算术优化算法（SAOA），SAOA 结合了算术优化算法（AOA）、模拟退火（SA）及改进搜索策略，以均方根误差（RMSE）最小化为优化目标，寻找最优参数组合，从而提升模型性能。SAOA 通过 AOA 提供的全局探索与局部开发机制，实现搜索广度与精度的平衡，同时借助 SA 的温度控制策略，提高算法跳出局部最优的能力。此外，混沌映射与动态逆向学习进一步优化了搜索空间分布，增强了算法的稳定性和收敛速度。通过这些改进，SAOA 显著提升了 LightGBM 的超参数优化效率，降低了计算成本，使其更适用于 EVCS 入侵检测等复杂应用场景。

本文在模型训练过程中采用 10 折交叉验证以确保参数优化的稳健性，并使用实验调整的 SAOA 参数组合，以使 LightGBM 模型在数据集上达到最佳性能。表 1 显示了本研究中使用的 SAOA 关键参数及其参考水平。

表 1 SAOA 参数及其参考水平

参数	含义	参数水平 1	参数水平 2	参数水平 3
N_{pop}	种群大小	30	40	50
A	MOA 控制参数	0.2	0.25	0.3
B	MOA 控制参数	0.8	0.85	0.9
Maxit	迭代次数	100	150	200
T_1	模拟退火初始温度	1 000	1 200	1 500
C	冷却温度	0.92	0.95	0.97

1.3 XAI

为提升 LightGBM 模型在 EVCS 入侵检测任务中的透明度和可解释性，本文采用了多种 XAI 方法，并按照全局和局部进行分类。

全局解释方法主要用于分析整个模型的决策过程，量化不同特征对模型整体预测结果的影响。本文采用 SHAP、LOCO、CEM 和 PFI 作为全局解释工具。

（1）SHAP

SHAP（Shapley Additive Explanations）基于合作博弈论的 Shapley 值来衡量特征对模型预测结果的贡献。

Shapley 值的计算公式如下：

$$\varphi_j = \sum_{S \subseteq F \setminus \{j\}} \frac{|S|!\,(|F|-|S|-1)!}{|F|!} [f(S \cup \{j\}) - f(S)] \tag{4}$$

其中，φ_j 为特征 j 的 Shapley 值；S 为不包含特征 j 的特征子集；F 为所有特征的集合；$f(S)$ 为使用特征子集 S 训练的模型的预测值；组合项 $\frac{|S|!\,(|F|-|S|-1)!}{|F|!}$ 作为权重，保证所有特征组合的公平性。

本研究采用 SHAP 来评估各特征对 LightGBM 预测结果的贡献，并绘制全局特征重要性图。

（2）LOCO

LOCO（Leave-One-Covariate-Out）是一种基于特征移除的解释方法，该方法的主要思想是单独评估每个特征对预测的影响。其计算过程如下：

$$I_j = L(f(X)) - L(f(X_{-j})) \tag{5}$$

其中，I_j 表示特征 j 的重要性；$L(f(X))$ 为原始模型的损失函数（如 RMSE 或 AUC）；$L(f(X_{-j}))$ 为移除特征 j 后的模型损失。

如果去除某个特征后，模型损失显著增加，则说明

该特征在预测中起着关键作用。LOCO 方法计算成本较低，适用于模型的全局特征重要性分析。

（3）CEM

CEM（Contrastive Explanation Method）是一种通过对比实例揭示模型决策依据的解释方法。其核心在于寻找最小扰动 δ，使得模型预测类别发生变化，从而识别对模型决策最关键的特征：

$$\min_{\delta} \|\delta\| \quad \text{s. t.} f(X+\delta) \neq f(X) \tag{6}$$

其中，X 为原始样本；δ 为最优扰动向量；$f(X+\delta)$ 代表模型对扰动样本的预测。CEM 通过这种方式分析决策边界，强调“为什么不是其他类别”，增强模型可解释性。

（4）PFI

PFI（Permutation Feature Importance）通过随机打乱特征值并破坏特征与目标之间的相关性，监测 AI 模型性能（通常是准确性）的变化来评估每个特征的重要性。其计算公式如下：

$$I_j = L(f(X)) - L(f(X_{\pi_j})) \tag{7}$$

其中，X_{π_j} 为对特征 j 进行随机排列后的数据集；$L(f(X))$ 为原始数据的模型损失；$L(f(X_{\pi_j}))$ 为随机打乱特征 j 后的模型损失。

本文采用 LIME 和 ALE 作为局部解释工具。

（1）LIME

LIME（Local Interpretable Model-agnostic Explanations）通过在特定样本附近生成扰动数据，并训练一个简单的可解释模型（如线性回归），以近似模型的局部决策边界。LIME 采用如下优化目标：

$$\min_{g \in G} L(f, g, \pi_x) + \Omega(g) \tag{8}$$

其中，f 为黑箱模型；g 为局部可解释模型；π_x 为局部样本的权重函数；$\Omega(g)$ 为模型复杂度正则项。

LIME 可以生成样本级别的特征贡献图，帮助用户理解 LightGBM 在特定输入上的预测逻辑。

（2）ALE

ALE（Accumulated Local Effects）是一种计算变量局部平均影响的方法，相比 SHAP 和 PFI，ALE 避免了特征共线性问题。其计算公式如下：

$$\text{ALE}_j(x) = \int_{x_{\min}}^{x} \mathbb{E}[f(x_j) \mid x_j = t] \mathrm{d}t \tag{9}$$

其中，$\text{ALE}_j(x)$ 代表特征 j 在样本 x 上的累积影响；$\int_{x_{\min}}^{x} \mathbb{E}[f(x_j) \mid x_j = t]$ 代表特征 j 在不同取值上的局部期望贡献。

ALE 适用于衡量连续变量的非线性影响，能更直观地展示模型对不同特征值范围的依赖关系。

2 实验结果与分析

2.1 实验环境

本文所有实验均在 Windows11 操作系统上进行，配备 AMD Ryzen 77840 Hw/Radeon™ 780M Graphics 处理器（3.80 GHz），32.0 GB 内存，64 位操作系统，为实验提供充足的计算资源。实验采用 Python3.9 进行实现，并基于 Scikit-Learn（Sklearn）等相关库进行算法的训练与优化。

2.2 数据集与预处理

本文实验所用数据集为加拿大网络安全研究所提供的电动汽车充电器攻击数据集 CICEVSE2024 和 Ferrag 等人所提供的网络安全数据集 Edge-IIoTset。选择这两种数据集能够为本文模型评估提供更强的泛化能力。

CICEVSE2024 数据集基于真实电动汽车充电器测试平台和标准通信协议（ISO15118，OCPP）构建，涵盖电动汽车充电器、车辆与后端系统的网络流量数据。通过监测协议交互，数据集可精准捕获正常与异常行为，为本文 EVCS 的入侵检测研究提供了可靠的支持。为提高数据质量并优化机器学习任务，该数据集经过一系列预处理。原始数据以 CSV 文件形式存储在两个目录中，每个文件对应不同场景或攻击类型。数据处理流程包括合并数据源、特征降维以及去除冗余信息。最终预处理后的数据集的样本分类如表 2 所示。

表 2 CICEVSE2024 数据集分类标签的样本数量

分类	攻击类型	数量
0	SYN_Flood	259 481
1	SynonymousIP_Flood	256 730
2	TCP_Flood	256 315
3	PSHACK_Flood	195 952
4	SYN_Stealth_Scan	77 278
5	TCP_Port_Scan	64 455
6	Service_Version_Detection	46 334
7	Vulnerability_Scan	38 023
8	UDP_Flood	32 475
9	OS_Fingerprinting	26 080
10	Aggressive_Scan	21 762
11	Slowloris_Scan	2 340
12	Benign	82
13	ICMP_Flood	32
14	ICMP_Fragmentation	14

Edge-IIoTset 数据集基于 IoT/IIoT 测试平台构建，涵盖网络流量特征、系统资源特征和攻击特征，涉及 DoS/DDoS、MITM、恶意软件等多种攻击类型。数据预处理包

括数据清理、特征选择、数据增强和标签转换。预处理后的数据集的样本分类如表3所示。

表3 Edge-IIoTset数据集分类标签的样本数量

分类	攻击类型	数量
0	Normal	24 101
1	DDoS_UDP	14 498
2	DDoS_ICMP	13 096
3	DDoS_HTTP	10 495
4	SQL_injection	10 282
5	DDoS_TCP	10 247
6	Uploading	10 214
7	Vulnerability_scanner	10 062
8	Password	9 972
9	Backdoor	9 865
10	Ransomware	9 689
11	XSS	9 543
12	Port_Scanning	8 921
13	Fingerprinting	853
14	MITM	358

2.3 评估指标

在多分类任务中，评估模型性能至关重要，常用指标包括准确率（Accuracy）、精确率（Precision）、召回率（Recall）、*F*1分数（*F*1-score）以及混淆矩阵。本文重点评估EVCS网络安全领域的多分类模型，采用宏平均后的准确率、精确率、召回率、*F*1分数等指标。表4是各指标的定义及计算公式。

表4 分类指标和公式

指标	公式
准确率	$\text{Accuracy}=\frac{TP+TN}{TP+TN+FP+FN}$
精确率	$\text{Precision}=\frac{TP}{TP+FP}$
召回率	$\text{Recall}=\frac{TP}{TP+FN}$
*F*1分数	$F1=2\times\frac{\text{Precision}\times\text{Recall}}{\text{Precision}+\text{Recall}}$

混淆矩阵通过展示真实类别与预测类别之间的对应关系，直观反映模型在各类样本上的识别能力与错误分布。

2.4 模型参数设置

LightGBM模型的超参数对其性能具有重要影响，因此选择合适的优化方法至关重要。鉴于元启发式算法在复杂优化问题中的优越性，本文引入了一种高效的优化策略SAOA，以优化LightGBM的超参数配置，从而提升模型性能。考虑到LightGBM具有众多超参数，本文针对四个关键超参数进行优化，以构建更精准的入侵检测模型。具体优化的超参数信息如表5所示，其余参数采用默认值。

表5 LightGBM超参数选择范围

超参数	取值范围	作用	含义
n_estimators	[10，100]	控制训练速率	迭代值
Learning rate	[0.15，0.35]	提高准确率	模型训练学习率
num_leaves	[5，100]	提高准确率	叶子节点数
max_depth	[3，11]	防止过拟合	树的最大深度

在SAOA优化过程中，采用LightGBM在测试集上的均方根误差（RMSE）作为适应度函数，以评估不同超参数配置的优劣。通过迭代优化，SAOA逐步调整LightGBM的关键超参数，以寻找到最优配置，从而提升模型的预测性能。最终，获得的最佳超参数组合如表6所示，其中迭代值（n_estimators）设定为15，学习率（Learning rate）取0.18，叶子数量（num_leaves）设为20，最大深度（max_depth）设为6。

表6 LightGBM最佳超参数

超参数	最佳超参数值
n_estimators	15
Learning rate	0.18
num_leaves	20
max_depth	6

2.5 SHAP特征选择

本研究将SHAP运用到EVCS入侵检测的特征选择流程中，通过量化特征的全局贡献度，从多维特征空间精准筛选出20个关键特征。根据不同的数据集分析显示，tcp.ack等协议层特征在Edge-IIoTset数据集中稳定呈现高SHAP值，而src2dst_stddev_piat_ms等时序特征在CICEVSE2024数据集中同样表现出显著影响力。这种相似但不同的场景区分性验证了SHAP方法的有效性和鲁棒性。随后选取其中前15个关键特征构建LightGBM模型来进行后续实验，在保持低复杂度的同时，对复合攻击模式的识别精度提升显著，展现了特征选择与模型效率间的协同优化效应。

2.6 结果性能分析

由表7和表8可知，LightGBM在Edge-IIoTset和CICEVSE2024数据集上的性能均随着特征选择和SAOA优

化显著提升。在 CICEVSE2024 数据集上，模型整体性能优于 Edge-IIoTset 数据集，表明该数据集更具结构化特征，优化效果更为显著。而在 Edge-IIoTset 数据集上，尽管 SAOA 提升了性能，但模型仍面临更大的挑战，说明数据复杂度和多样性对入侵检测的影响较大。

表 7　Edge-IIoTset 数据集的性能指标结果（%）

指标	LightGBM	特征选择后 LightGBM	SAOA 优化后 LightGBM
Accuracy	23. 33	74. 77	88. 89
Precision	27. 37	77. 93	91. 20
Recall	23. 21	74. 77	88. 81
*F*1-score	23. 72	72. 23	88. 98

表 8　CICEVSE2024 数据集的性能指标结果（%）

指标	LightGBM	特征选择后 LightGBM	SAOA 优化后 LightGBM
Accuracy	52. 63	83. 73	97. 53
Precision	59. 67	85. 77	98. 59
Recall	52. 63	83. 73	97. 53
*F*1-score	52. 53	85. 15	98. 01

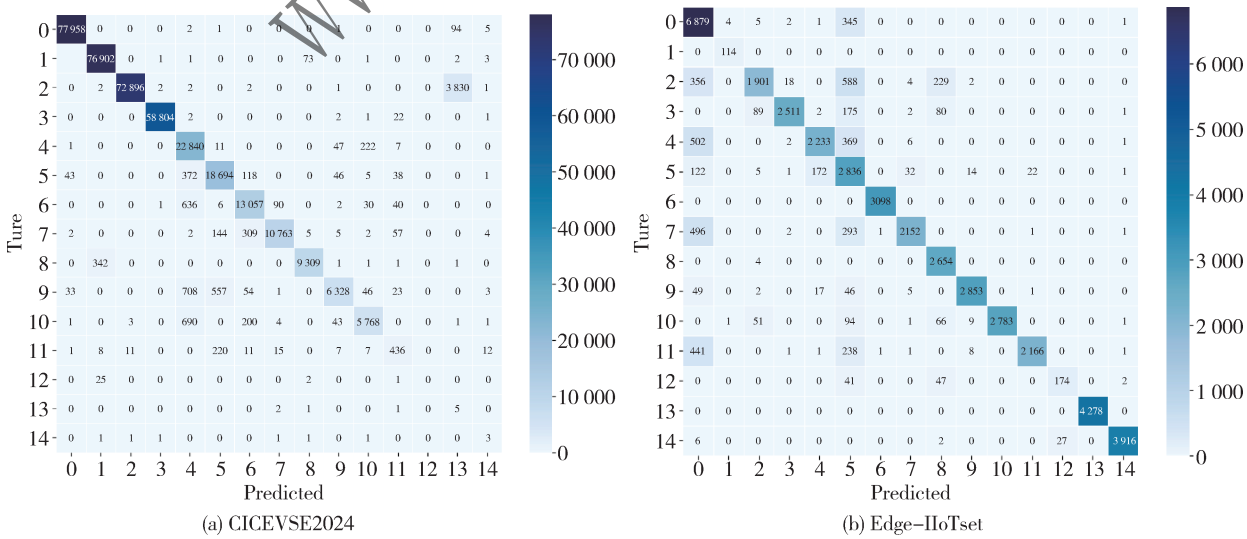
从图 2 混淆矩阵可以看出，经过特征选择与超参数寻优后的 LightGBM 在 CICEVSE2024 和 Edge-IIoTset 数据集上的分类性能均表现良好。对于 CICEVSE2024 数据集，如图 2（a）所示，模型在大多数类别上实现了较高的正确分类率，但仍存在部分误分类情况，例如 OS_ Fingerprinting（类别 9）被错误分类为类别 5（DDoS_ ICMP），这表明 OS 指纹识别攻击的流量模式更容易与 DDoS_ ICMP 攻击混淆。此外，Aggressive_ Scan（类别 10）主要被误分类为了类别 5（DDoS_ ICMP）和类别 6（Vulnerability _Scan），可能是由于它们都涉及大量扫描或请求行为，从而导致模型的误分类。

而在 Edge-IIoTset 数据集上，如图 2（b）所示，类别 1（DDoS_ UDP）、6（Uploading）和 13（Fingerprinting）均能实现 100% 的分类准确率。然而，部分类别仍存在较高的误分类率，其中，分类效果最差的是类别 5（DDoS _TCP），被较多地误分类为类别 2（DDoS_ ICMP）和类别 4（SQL_ injection），可能是因为两者的高流量特性相似和某些 SQL 注入攻击与 TCP 连接密切相关，导致模型无法精准区分。

2. 7　可解释性分析

本文使用的 LightGBM 模型作为典型的黑盒模型，其复杂结构虽然带来较高预测性能，却缺乏可解释性。因此，本文结合全局与局部层面采用多种 XAI 方法对模型决策进行解释。

在全局解释方面，本文引入 LOCO、CEM 和 PFI 三种特征重要性分析方法，分别从特征移除、最小扰动和置换影响角度评估特征对模型性能的贡献。在 Edge-IIoTset 数据集中（图 3、图 4、图 5 左），tcp. ack 和 tcp. seq 等 tcp 相关特征的重要性较高，尤其是 tcp. ack 在 LOCO 方法中的得分超过 0. 7，显示其对模型预测影响显著。相比之下，tcp. flags 等特征重要性较低。在数据集 CICEVSE2024 中（图 3、图 4、图 5 右），src2dst_ duration_ ms 和 src2dst _syn_ packets 等特征的重要性最高，表明其在特定网络环境下具有较强的判别能力。



(a) CICEVSE2024　　(b) Edge-IIoTset

图 2　混淆矩阵图

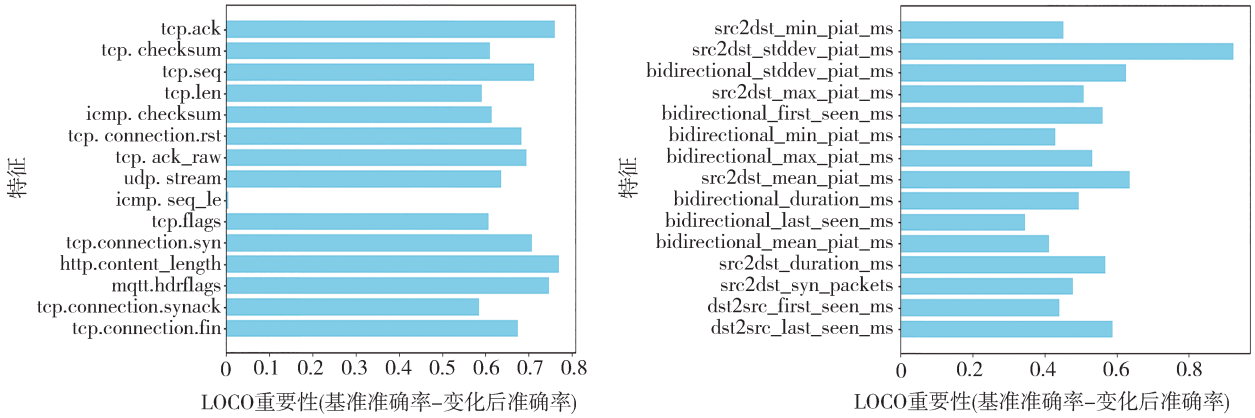


图 3　LOCO 特征图

图 4　CEM 特征图

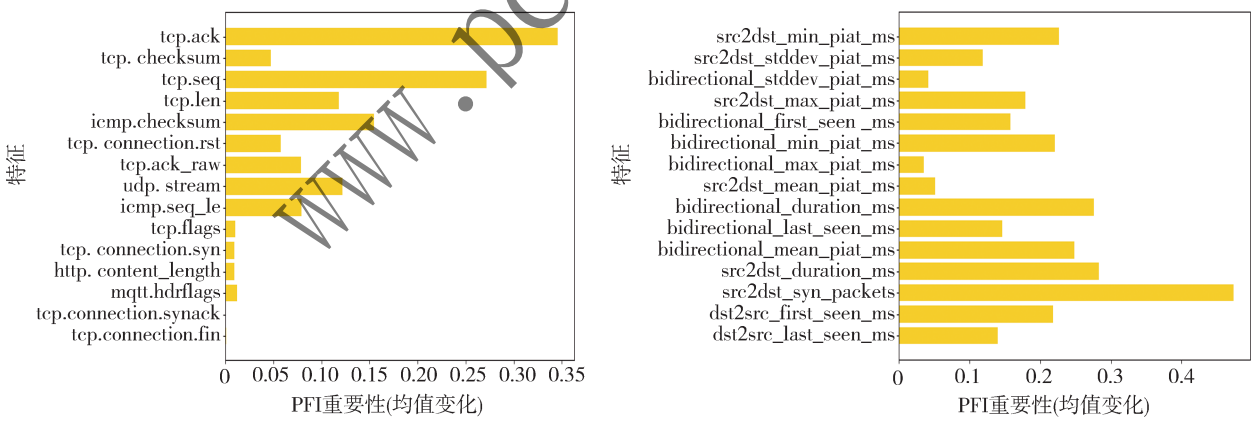


图 5　PFI 特征图

鉴于不同可解释性方法在评估特征重要性时存在关注维度与计算机制上的差异，可能导致解释结果在排序和权重上的不一致性，因此本文采用方法互补性原则，即不同方法从特征剔除（LOCO）、扰动响应（CEM）与随机置换（PFI）三个维度提供多角度的解释结果，能够在特征层面形成互补性视角，可更全面地揭示模型内部的决策机制，有效增强模型在入侵检测任务中的可解释性与可信度。为此，本文对各方法所选 Top-k 关键特征进行交集分析，通过重叠度度量提升特征选择的一致性与可信度；其次，为了验证结果稳健性，利用 Spearman 等级相关系数对各方法的重要性排序进行一致性评估，该方式能够很好地解释模型结果在不同样本划分下的稳定性与泛化能力。

在局部解释层面，本文采用 LIME 和 ALE 来选择任意一个样本进行深入剖析。LIME 解释显示，如图 6 所示，在 Edge-IIoTset 数据集中，其中一个样本被高概率预

测为 Normal 类，tcp. checksum = 1. 42 对预测为 MITM 类具有显著正向影响，而 tcp. connection. syn = －0. 36 则抑制该预测。在 CICEVSE2024 数据集中，如图 7 所示，模型准确识别出了样本的攻击类型，其中的特征如何影响分类结果在规则集中也清晰可见。

ALE 方法进一步揭示了特征值变化对预测概率的平均影响趋势，如图 8 与图 9 所示。例如，Edge-IIoTset 中 tcp. checksum 值与 tcp. ack_raw 值增大时，分别对 Vulnerability_scanner 类预测概率产生正向与负向影响。而在数据集 CICEVSE2024 中，src2dst_max_piat_ms 值和 bidirectional_max_piat_ms 值的增加均对 Slowloris_Scan 预测有明显的提升。

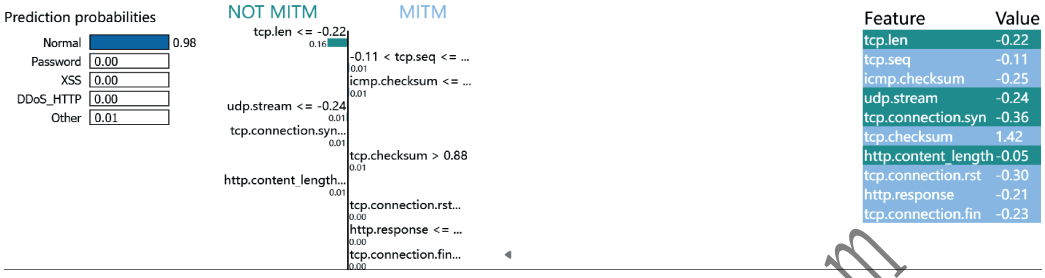


图 6　Edge－IIoTset 数据集的 LIME 局部解释图

图 7　CICEVSE2024 数据集的 LIME 局部解释图

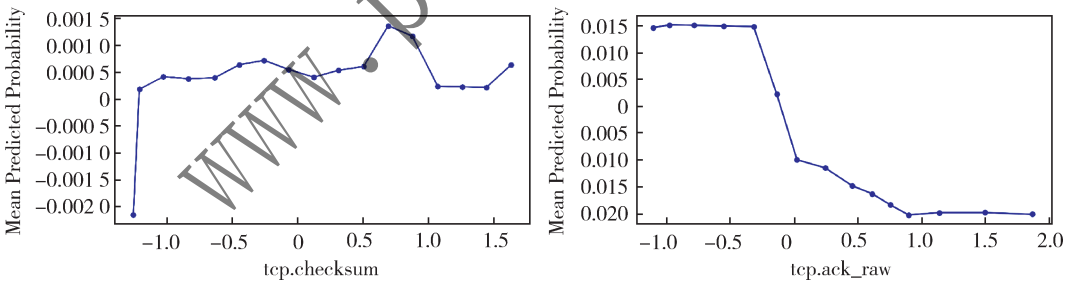


图 8　Edge-IIoTset 数据集的 ALE 局部解释图

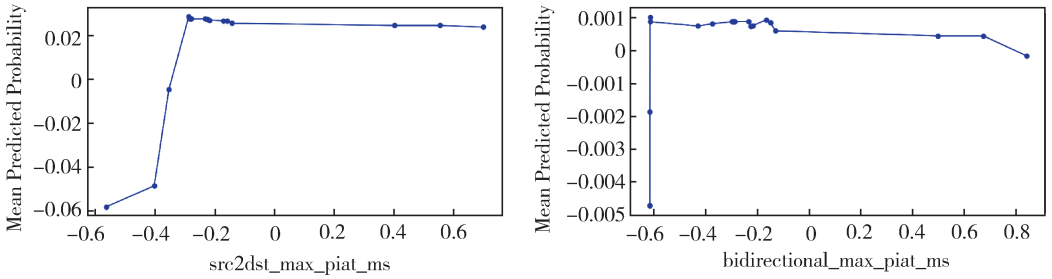


图 9　CICEVSE2024 数据集的 ALE 局部解释图

LIME 和 ALE 提供了样本级别的局部可解释性，揭示特征如何具体影响分类概率，有助于理解模型在关键攻击样本上的预测逻辑，提高入侵检测系统的透明性和可信度，以确保系统的网络安全。

3 讨论

本研究基于 CICEVSE2024 和 Edge-IIoTset 数据集，采用 LightGBM 结合多种可解释性方法，对 EVCS 入侵检测方法进行深入的研究。如表 9 所示，相比近两年的方法，本文模型在保持较高检测准确率的同时，显著提升了决策过程的透明度和可解释性，有助于安全运维人员理解模型判定依据。此外，本研究在特征选择与超参数优化方面进行了改进，使模型在不同数据集上的适应性更强，为提升 EVCS 网络安全提供了一种高效、可解释的解决方案。

表 9 与现有方法比较

方法	数据集	模型	可解释性	准确率/%
文献［15］	CICEVSE2024	RF	无	93.74
文献［16］		TCN	无	93.00
文献［12］		CNN-LSTM	有	97.15
本文		LightGBM	有	97.53
文献［17］	Edge-IIoTset	RNN	无	74.40
文献［18］		FI-SEL	无	90.16
本文		LightGBM	有	88.89

4 结论

本文提出了一种融合 LightGBM 与多种可解释性技术的电动汽车充电站入侵检测框架，通过 SHAP 方法筛选关键特征，结合 SAOA 优化 LightGBM 超参数，有效提升了模型性能与计算效率。在 CICEVSE2024 和 Edge-IIoTset 数据集上，该模型分别取得了 97.53% 和 88.89% 的准确率，*F*1 分数分别为 98.01% 和 88.98%。此外，集成 LOCO、CEM 和 PFI 全局解释方法与 LIME 和 ALE 局部解释方法，深入剖析模型的特征依赖与预测机制，显著提升了模型透明度与可信度。该框架兼具高精度与强可解释性，为充电站威胁识别与安全运维提供了有力支持。未来可拓展至实时预警系统与跨场景应用，进一步增强其在复杂网络环境下的适应性与实用性。

参考文献

［1］AKANDA M R K, LIMA J R Q P S D O, HOLMES A A, et al. Safeguarding the future of mobility: cybersecurity issues and solutions for infrastructure associated with electric vehicle charging［J］. arXiv preprint, arXiv: 2502.00035, 2025.

［2］MAKHMUDOV F, KILICHEV D, GIYOSOV U, et al. Online machine learning for intrusion detection in electric vehicle charging systems［J］. Mathematics, 2025, 13 (5): 712.

［3］TULSIANI J, NAYAK K, LAKRA P P, et al. Privacy preserving scheme for EV charging station using machine learning based intrusion detection system［C］//2024 Second International Conference on Intelligent Cyber Physical Systems and Internet of Things (ICoICI). IEEE, 2024: 545－552.

［4］KILICHEV D, TURIMOV D, KIM W. Next-generation intrusion detection for IoT EVCS: integrating CNN, ISTM, and GRU models［J］. Mathematics, 2024, 12 (4): 571.

［5］BASNET M, ALI M H. Exploring cybersecurity issues in 5G enabled electric vehicle charging station with deep learning［J］. IET Generation, Transmission & Distribution, 2021, 15 (24): 3435－3449.

［6］MISKIN S V, CHANDARAGI P, WALI U V. Intrusion detection system for electric vehicle charging station［C］//2023 3rd International Conference on Mobile Networks and Wireless Communications (ICMNWC). IEEE, 2023: 1－7.

［7］ALMADHOR A, ALSUBAI S, BOUAZZI I, et al. Transfer learning for securing electric vehicle charging infrastructure from cyber-physical attacks［J］. Scientific Reports, 2025, 15 (1): 9331.

［8］ARRIETA A B, DÍAZ-RODRÍGUEZ N, DEL SER J, et al. Explainable artificial intelligence (XAI): concepts, taxon-omies, opportunities and challenges toward responsible AI［J］. Information Fusion, 2020, 58: 82－115.

［9］DOSHI-VELEZ F, KIM B. Towards a rigorous science of interpretable machine learning［J］. arXiv preprint, arXiv: 1702.08608, 2017.

［10］KHAN N, AHMAD K, TAMIMI A A, et al. Explainable AI-based intrusion detection system for Industry 5.0: an overview of the literature, associated challenges, the existing solutions, and potential research directions［J］. arXiv preprint, arXiv: 2408.03335, 2024.

［11］ARRECHE O, GUNTUR T, ABDALLAH M. XAI-based feature selection for improved network intrusion detection systems［J］. arXiv preprint, arXiv: 2410.10050, 2024.

［12］ATTIQUE D, WANG H, WANG P, et al. EX-DFL: an explainable deep federated-based intrusion detection system for Industrial IoT［C］//2024 21st International Joint Conference on Computer Science and Software Engineering (JCSSE). IEEE, 2024: 358－364.

［13］MOHANTY P K, REDDY K H K, PANIGRAHY S K, et al. Leveraging generative & explainable AI for electric vehicle energy towards sustainable, consumer-centric transportation［J］. IEEE Access, 2024, 12: 143721－143732.

（下转第 16 页）

研究，2024（1）：217－219.
[2] 刘鹏远. 数字化驱动，让培训更智能 [J]. 人力资源，2025（3）：80－81.
[3] 韩佼男，张万莉，王春秀，等. 智慧培训平台赋能企业培训管理数字化转型 [J]. 中国管理信息化，2024，27（22）：144－146.
[4] 马春梅. 在线教育平台信息安全防护策略——以 XX 公共培训服务平台为例 [J]. 青岛远洋船员职业学院学报，2023，44（2）：74－76.
[5] 王若晗，向继，管长御，等. 零信任架构的回望与未来发展研究 [J]. 信息安全研究，2024，10（10）：896－902.
[6] WARD R，BEYER B. BeyondCorp：a new approach to enterprise security [J]. the Magazine of USENIX & SAGE，2014，39：6－11.
[7] 诸葛程晨，王群，刘家银，等. 零信任网络综述 [J]. 计算机工程与应用，2022，58（22）：12－29.
[8] 薛人瑞，吴华佳. Hisec 零信任安全解决方案 [C]// 2021 年国家网络安全宣传周“网络安全产业发展论坛”论文集，2021：119－123.
[9] 蔡东赟. 腾讯 iOA 零信任安全技术实践 [J]. 信息安全与通信保密，2020（S1）：98－102.
[10] 李治宇. 基于零信任策略在网络攻防演练中的实践研究 [J]. 信息产业报道，2024（1）：60－62.
[11] 罗栗，黎臻，陈洋. 零信任网络架构与实现技术的研究与思考 [J]. 通信技术，2023，56（4）：509－514.
[12] 莫爵君，陈哲，春增军，等. 基于企业现有安全架构的零信任架构可行性研究 [C]//2023 电力行业信息化年会，2023：226－232.
[13] 徐言海. 零信任安全体系的设计与实现 [J]. 集成电路应用，2024，41（2）：329－331.
[14] 贾万祥，张平华. 零信任架构下的智慧校园安全性实测技术 [J]. 鄂州大学学报，2024，31（1）：99－101.
[15] VASWANI A，SHAZEER N，PARMAR N，et al. Attention is all you need [J]. arXiv：1706. 03762，2017.
[16] 孙振中. 零信任和 SDP 技术框架在新华社移动办公中的应用研究 [C]//中国新闻技术工作者联合会 2024 年学术年会论文集，2024：232－236.
[17] 安宁，许文静，刘珠慧，等. 基于零信任模型的细粒度数据库安全控制方法 [J]. 电子技术应用，2024，50（10）：63－68.
[18] 吕忠亭，朱丹妮，雷世斌，等. 基于零信任体系的数字身份安全平台设计与研究 [J]. 微型电脑应用，2024，40（2）：45－49.

（收稿日期：2025－03－14）

作者简介：

秦文远（1998－），男，硕士，助教，主要研究方向：软件工程、数据安全、深度学习、计算生物学。

安宁（1990－），男，硕士，高级工程师，高级经济师，主要研究方向：软件工程、云计算、人工智能、大数据、信息安全、数字教育等。

（上接第 9 页）

[14] RAHMAN M M，CHAYAN M M H，MEHRIN K，et al. explainable deep learning for cyber attack detection in electric vehicle charging stations [C]//Proceedings of the 11th International Conference on Networking，Systems，and Security，2024：1－7.
[15] BUDDI E D，GHORBANI A A，DADKHAH S，et al. Enhancing ev charging station security using a multi-dimensional dataset：CICEVSE2024 [C]//IFIP Annual Conference on Data and Applications Security and Privacy. Cham：Springer Nature Switzerland，2024：171－190.
[16] BENFARHAT I，GOH V T，SIOW C L，et al. Temporal convolutional network approach to secure open charge point protocol（OCPP）in electric vehicle charging [J]. IEEE Access，2025，13：15272－15289.
[17] SHEN T，DING L，SUN J，et al. Edge computing for IoT security：integrating machine learning with key agreement [C]//2023 3rd International Conference on Consumer Electronics and Computer Engineering（ICCECE）. IEEE，2023：474－483.
[18] ABDULKAREEM S A，FOH C H，CARREZ F，et al. A lightweight SEL for attack detection in IoT/IIoT networks [J]. Journal of Network and Computer Applications，2024，230：103980.

（收稿日期：2025－04－07）

作者简介：

姚沁怡（1999－），女，硕士研究生，主要研究方向：网络安全、工业大数据。

龙莆均（1987－），通信作者，男，博士，讲师，主要研究方向：最优化方法。E-mail：longpujun@ gmail. com。

陈世伦（1998－），男，硕士研究生，主要研究方向：图像处理。

版权声明

凡《网络安全与数据治理》录用的文章，如作者没有关于汇编权、翻译权、印刷权及电子版的复制权、信息网络传播权与发行权等版权的特殊声明，即视作该文章署名作者同意将该文章的汇编权、翻译权、印刷权及电子版的复制权、信息网络传播权与发行权授予本刊，本刊有权授权本刊合作数据库、合作媒体等合作伙伴使用。同时，本刊支付的稿酬已包含上述使用的费用，特此声明。

《网络安全与数据治理》编辑部