零信任架构中的持续内部威胁检测：行为分析

Gaurav Sarraf 独
立研究人员

摘要——内部威胁是网络安全领域中尤为棘手的问题，尤其是在移除了隐式信任的零信任架构（ZTA）中。尽管基本原则是永不信任，始终验证，攻击者仍可能利用合法凭证，冒充普通用户的活动。对此，结合机器学习（ML）的行为分析有助于持续监控用户活动并识别异常。这个初步框架使用 CERT内部威胁数据集进行数据清洗、归一化，并通过合成少数类过采样技术（SMOTE）实现类别平衡。此外，该框架还采用主成分分析（PCA）进行降维。研究使用了多个基准模型，包括支持向量机（SVM）、人工神经网络（ANN）和贝叶斯网络（Bayes Net），以开发和评估 AdaBoost 分类器。与 SVM（90.1%）、ANN（94.7%）和 Bayes Net（94.9）相比，AdaBoost 的性能更高，准确率（ACC）为 98.0%，精确率（PRE）为 98.3%，召回率（REC）为 98.0%，F1 分数（F1）表现也更优。受试者工作特征（ROC）研究进一步证实了其优势，曲线下面积（AUC）达到 0.98。这些结果证明，基于 AdaBoost 的行为分析能够有效且可靠地强化零信任环境中的持续内部威胁检测。

关键词—内部威胁、零信任架构、行为分析、人工智能、机器学习、持续监控、网络安全

I. 引言

“网络安全”一词描述了用于保护存储在计算机网络中的私密和敏感信息，以及连接这些网络的系统的措施。人们普遍认为，对网络风险进行实时控制比事后控制更有效。内部威胁已成为目前最难检测的威胁之一，它们由拥有系统合法访问权限的个人发起 [1]。内部威胁分为三类：蓄意造成损害的恶意用户；因不注意相关政策而在不知情的情况下危害安全的无意识用户；以及利用其授权凭证访问所存储敏感信息的外部行为者 [2]。受信任的内部人员可能会泄露机密信息、破坏安全机制、窃取和蓄意破坏，从而造成声誉、财务和运营损失。

即使在零信任安全（ZTS）环境下，也可能存在内部威胁，因为攻击者可以使用合法凭证、模拟正常用户行为，或规避薄弱的控制措施。基于身份和基于情境的访问控制是零信任安全（ZTS）的主要支柱。然而，大规模实施 ZTS 也面临自身的挑战，其中包括在高安全性与可用性之间取得平衡、遗留系统集成、访问数据处理，以及在不造成系统瓶颈的情况下保持系统效率 [3], 。这些问题凸显出这样一个事实：ZTS 能够显著降低风险，但内部威胁仍是一个需要持续监控并得到适当检测的重大问题 [4][5]。

身份与访问管理（IAM）通过识别身份并实施基于角色的访问来支持 ZTS，但传统的身份与访问管理系统在识别细微的行为异常方面仍不够有效 [6]。在这种情况下，人工智能和机器学习是提升安全性的两个关键方面。人工智能和机器学习可以通过分析行为数据（即登录时间、设备使用情况、网络交互、访问日志等）来检测恶意或危险的内部人员行为，并识别异常 [7]。机器学习模型能够适应不断变化的用户行为，提高检测率并减少假阳性。机器学习模型同样能够匹配不断演变的用户行为，并优化检测率和假阳性率。机器学习在安全实施中发挥着至关重要的作用 [8]。人工智能和机器学习能够识别行为模式中的异常，例如人员登录时间、设备使用情况、网络使用情况和访问模式的异常，这些异常可能表明存在恶意或高风险的内部人员活动 [9]。然而，机器学习还具有一些优势：它可以自动识别复杂模式，随着行为的变化而不断演进，减少误警，并且能够轻松应用于大规模数据集。由于具备这些特征，机器学习对于不断演进的零信任环境中的内部威胁识别非常有益 [10]。

将人工智能和机器学习纳入 ZTS，使 IAM 成为动态的风险敏感型系统，而非固定的身份验证工具。行为分析使组织能够持续跟踪用户活动、识别异常行为，并提前应对内部威胁 [11]。这一策略总体上增强了安全性，身份应作为主要边界；同时，它也符合零信任原则，即即使处于受信任的环境中，也不能默认信任任何用户或设备。

A. 研究动机与贡献

本研究旨在应对组织中日益增多的内部威胁：受信任人员滥用其访问权限，危及安全。在零信任架构中，没有任何实体天然值得信任，因此，持续监控并检测内部威胁对于保护敏感信息和系统至关重要。传统检测方案往往面临数据集不平衡、行为信息维度高以及内部人员行为隐蔽等问题。本研究采用结合机器学习的行为分析来应对这些挑战，从而提高检测准确率和可靠性。通过揭示用户活动中的隐藏模式，所提出的框架能够支持主动防御机制，并增强零信任环境的韧性。本研究的主要发现如下：

- 使用了广为人知的 CERT内部威胁数据集，该数据集是识别内部威胁研究的标准数据集。• 进行了数据清洗，以去除噪声、不一致性和异常值。• 使用 SMOTE（即合成少数类过采样技术）来纠正类别不平衡并增加少数类样本的占比。• 采用归一化处理，使特征值标准化，确保各变量的一致性。• 采用主成分分析（PCA）进行降维，在减少冗余的同时捕获关键行为特征。• 开发了基于 AdaBoost 的分类模型，以提升对恶意内部人员行为的检测能力。

- 通过使用 F1 分数、召回率、精确率和准确率以及精确率作为指标进行测试，验证了模型的有效性和可靠性。

B. 论文的依据与创新之处

本研究的依据在于，内部威胁日益复杂，即使在零信任安全模型中也会利用信任边界。传统的身份与访问管理系统无法检测细微的行为偏差，因此亟需先进的解决方案。本研究的创新之处在于将行为分析与稳健的预处理相结合——采用 SMOTE 解决类别不平衡问题，采用主成分分析进行降维，并采用 AdaBoost 实现自适应分类——并将其应用于 CERT内部威胁数据集。不同于传统方法，该框架不仅提高了检测准确率，还确保了可扩展性和适应性，为零信任架构中的持续内部人员威胁监控提供了一种实用的人工智能驱动解决方案。

C. 论文结构

本文结构如下：第二节讨论检测内部威胁的既有研究。第三节详细介绍所提方法中的数据准备、特征工程和模型构建。第四节展示 AdaBoost 分类器的实验结果和性能评估。第五节讨论研究局限，并概述提升内部威胁检测效果的未来方向。第六节总结关键发现和研究贡献，从而结束全文。

II. 文献综述

近期研究指出，内部威胁会利用特权访问权限，因此难以检测。深度学习方法、行为分析和集成机器学习技术已被证明有助于建模用户行为，从而能够在零信任架构中及早识别恶意内部人员。

Tamanna（2020）探讨了内部威胁这一概念，指出这类威胁源于组织内部受信任的个人，会带来安全威胁。由于内部人员了解敏感信息和系统弱点，这类威胁很难检测。该研究提出了一种检测异常用户行为的方法，采用深度神经网络技术，尤其是卷积神经网络与长短期记忆网络的组合。该模型在公开数据集（CMU CERT r4.2版本）上实施，在检测内部风险方面优于传统机器学习方法，受试者工作特征值为0.914。[12]

Le、Zincir-Heywood 和 Heywood（2020年）研究了识别恶意内部威胁的难点，例如数据不平衡和行为变化。针对以用户为中心的内部威胁识别，他们提出了一种能够识别内部人员及有害行为的机器学习系统。即使使用稀疏数据进行训练，测试结果也表明，该系统识别恶意内部人员的准确率可达 85 %，假阳性率仅为 0.78%。值得注意的是，该系统能够在 14 分钟内检测到恶意行为。全面的报告增强了对内部威胁调查的分析洞见。[13]

Singh、Mehtre 和 Sangeetha（2019年）的论文探讨了组织内部日益严峻的内部攻击威胁。论文强调通过分析人类行为来检测有害行为，并批评了当前的安全解决方案，因为它们大多只关注外部威胁。作者采用结合卷积神经网络与 MSLSTM 的集成混合算法，提出了一种用户行为画像方法，可检测时间序列数据中的异常模式。研究结果表明，MSLSTM 比传统模型更有效，在训练集和测试集上的曲线下面积分别为 0.9042 和 0.9047，表明利用公开数据集成功识别内部威胁 [14]

Hu 等人（2019年）的论文讨论了内部攻击者如何利用未授权访问轻易进入内联网系统。论文提出了一种新颖的用户认证系统，利用深度学习和鼠标生物行为特征来实现持续身份验证并阻止内部攻击。通过使用包含 10 名用户的开源数据集进行实验，研究表明，该方法能够在约 7 秒内完成身份验证。假阴性率为 2.28，错误接受率为 2.94，说明该方法效率较高 [15]

Hall 等人（2018年）的论文介绍了信息保护领域当前面临的内部威胁问题，并指出社区需求与解决方案之间仍存在尚未解决的差距。该研究使用 CERT r4.2数据集，采用机器学习分类器来预测恶意内部人员场景，具体而言，是指员工在离开组织前将敏感信息上传至维基解密的行为。研究提出了一种元分类器，其预测性能优于单个模型，并阐述了一种基于组织日志数据生成用户初步摘要的方法。为了提高精确度，研究采用了提升法等方法。根据混淆矩阵和 ROC 曲线下面积，该元分类器的准确率达到 96.2% [16]

Gamachchi 和 Boztas（2017年）说明了不诚实人员所带来的严重危害：这些人可能伪装成可靠的信息来源，同时对国家的财政与安全，以及公共和私人组织的声誉造成严重破坏。为应对内部威胁，必须深入调查用户、系统和网络层面，因为恶意操作具有不可预测且不断变化的特点。本文提出了一种内部威胁检测系统，利用属性图聚类方法和离群点排序系统，识别组织内的可疑用户。实证结果表明，该框架的受试者工作特征曲线值为 0.7648，证明其表现良好，并具备出色的检测能力 [17]

表I总结了近期关于混合深度学习、集成机器学习和基于内部威胁数据的行为分析的研究，这些研究展现出较高的检测准确率，并克服了可扩展性、实时实施等挑战。

表I. 内部威胁检测技术的最新研究

参考文献	方法	数据集	结果	挑战	未来工作
Tamanna（2020年）	使用混合深度学习 长短期记忆网络 + 卷积神经网络进行异常用户行为检测	卡内基梅隆大学CERT r4.2 (12GB)	ROC = 0.914；优于传统的 ML 方法	High 计算大规模数据的成本数据；在实际实时部署	探索可扩展深度学习架构及其集成方式到零信任环境中的持续监控环境
Le, Zincir-Heywood & Heywood（2020年）	基于机器学习的 以用户为中心检测 with 多层级数据粒度	专有数据集，其中包含真实内部人员场景	检测到 85% 恶意内部人员，假阳性率为 0.78%；最早可在.....之后 14 分钟恶意行为	不平衡 数据，有限的真实标签，处理 行为漂移	开发自适应机器学习模型以应对概念 漂移 and 扩展至流式实时系统
Singh、Mehtre & Sangeetha（2019年）	集成混合机器学习：多状态长短期记忆网络 + 卷积神经网络，用于时空 异常检测	公开内部威胁数据集	AUC = 0.9042（训练集），0.9047（测试集）	计算复杂度； 仍然受限于手工设计特征	将集成学习与自监督 学习 for 持续内部威胁检测
Hu et al.（2019年）	持续 user 身份验证 via 鼠标生物行为特征与深度学习	开源数据集 (10 用户)	每次身份验证 7s；误接受率 = 2.94%，误拒绝率 = 2.28%	有限 to 小型数据集； not 可推广	应用于多模态行为生物识别（键盘、击键、网络活动）在零信任环境中
Hall et al.（2018年）	元分类器 集成（提升 ML 模型）用于检测数据外泄	CERT r4.2	准确率 = 96.2%，曲线下面积 = 0.988	聚焦于单一内部威胁场景；泛化能力有限	扩展至多场景检测并整合 with 零信任安全策略
Gamachchi 和 Boztas（2017年）	属性图聚类 + 用于检测的异常值排序可疑用户	企业网络日志	曲线下面积 = 0.7648	高维异质的 数据，可扩展性问题	改进 基于图的 分析借助深度图神经网络（图神经网络）用于内部人员检测

III. 方法

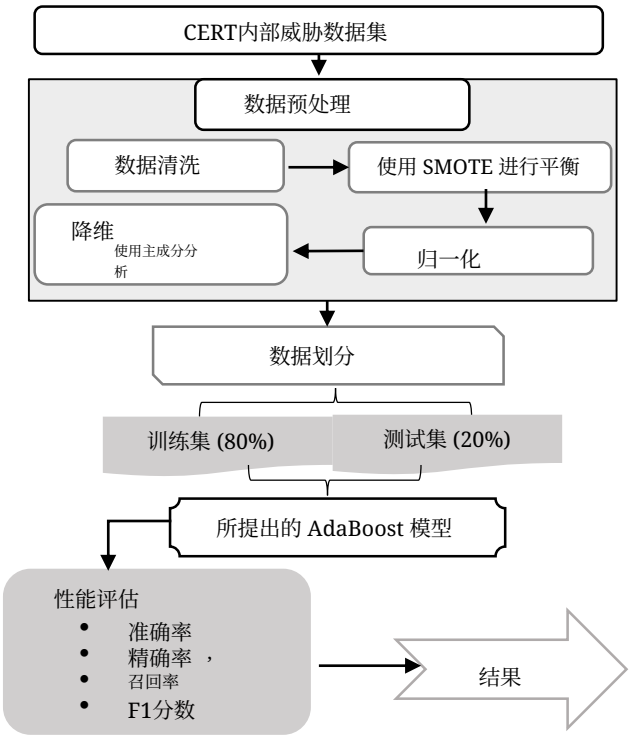

图 1. 使用 AdaBoost 分类器进行内部威胁检测的拟议方法

本研究提出了一种行为分析系统，可供零信任架构使用CERT内部威胁数据集持续检测内部威胁。该方法首先进行有组织的预处理，包括通过数据清洗去除噪声和不一致性。为解决类别不平衡问题，采用SMOTE，并通过归一化将特征值调整至相同水平。随后使用主成分分析降低维度，在减少冗余的同时保留关键模式。预处理后，数据进一步划分为训练集（80%）和测试集（20%）。基于训练数据构建AdaBoost分类模型，并结合多个弱学习器，以提升检测的有效性和稳健性。使用包括准确率、精确率、召回率和F1 分数在内的性能指标，利用测试数据对训练好的模型进行测试。结果表明，AdaBoost能够检测内部威胁。如图1所示，该过程是连续进行的。

A. 数据收集

CERT 内部威胁数据库是一个大型数据库，旨在检测和分析组织环境中的内部威胁。该数据库包含 693,649 条记录和 830 个特征，涵盖了广泛的用户活动、行为和系统交互，这些都可能是潜在内部人员恶意活动的指标。目标变量是内部人员，用于判断给定记录是否与内部威胁事件相关。以下提供了多种可深入了解该数据集的可视化图表：

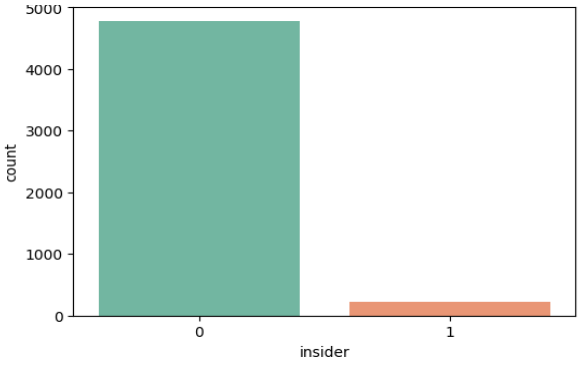

图 2. 内部人员与非内部人员威胁的类别分布

图 2 展示了两类观测数据的数量，横轴标注为内部人员。类别 0（可能是非内部人员）的样本数量远多于另一类，约为 4,800。相比之下，标记为 1（代表内部人员）的部分数量要少得多，约为 200 至 300。如此显著的差异表明数据存在严重的不平衡，非内部人员观测数据的数量远远超过内部人员观测数据。这种分布是高度不平衡的分类问题的典型特征，在数据分析或机器学习任务中可能需要进行特殊处理。

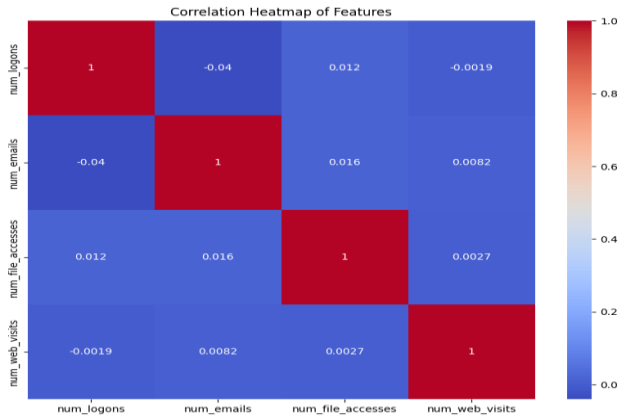

图 3. 数据集相关性热力图

图 3“特征相关性热力图”基于 CERT内部威胁数据，对四项特征进行了可视化：num_logons、num_emails、num_file_accesses 和 num_web_visits。对角线显示相关系数为 1，即完全相关，因为每项特征都与自身相关。所有非对角线数值都接近于零，表明任意两项不同特征之间的线性关系都非常弱或可以忽略。例如，num_logons 和 num_emails 的相关系数仅为 - 0.04。这意味着这些特征大多相互独立，而这正是机器学习模型非常理想的一项属性。

B. 数据预处理

数据预处理是处理原始数据的重要阶段，旨在为机器学习做好准备，使其能够被处理。该阶段负责清理、转换和整理数据，使其在计算上更加高效。适当的预处理对于确保数据集准确、一致且组织良好至关重要，从而提高数据集在后续模型中的可靠性和可用性。以下步骤概述了预处理过程：

1) 数据清洗

预处理所需的预处理工作是数据清洗，它能确保在运行机器学习算法 [18] 之前，数据集具备良好的质量、可靠性和实用性。本研究对 CERT内部威胁数据集进行了数据清洗检查。尽管数据集中未发现缺失值或异常值，但整体流程如下：

- 检查了 830 个特征中是否存在空值或缺失值，未发现任何缺失值。• 对所有数值变量和分类变量进行了异常值验证，未发现严重异常值。• 检查了数据集中是否存在重复记录，未发现此类记录。• 检查了各项特征（时间戳、用户ID、活动日志等）中的数据一致性，数据集中的数据格式和值均一致。

C. 使用 SMOTE 进行平衡处理

机器学习研究人员使用 SMOTE 解决社会阶层不平等的问题。SMOTE 通过为少数类创建合成实例，致力于实现类别分布更加均衡这一主要目标。采样过程如下：首先，选取一个少数类实例。选定实例后，算法会找出它的 K近邻，其中 k=3 在本研究中为 [19]。然后，从已识别出的邻居列表中随机选择一个邻居。最后，将第一个实例的特征与随机选出的邻居的特征相结合，从而创建一个合成实例。这种做法有效地引入了合成数据。经过 SMOTE 处理后，Insider 线程 0 和 1 的类别数量被调整为 6,92,342。

D. 归一化

使用距离度量或神经网络的分类技术（包括最近邻和聚类算法）能够从归一化中显著受益，因为归一化可以加快模型训练阶段 [20]。最常用的归一化方法包括最小-最大、Z分数和小数定标。本研究采用最小-最大归一化，其一般公式如下（1）：

$$x_{new} = \frac{x - \min(x)}{\max(x) - \min(x)} \quad (1)$$

E. 使用主成分分析进行降维

降维算法用于解决高维数据带来的问题。这类数据不仅会增加计算复杂度和实时性要求，还会妨碍故障诊断性能，因为许多维度无关紧要，并可能掩盖噪声数据中的簇 [21]。主成分分析可以将相关特征转换为规模更小、彼此不相关的主成分集合，从而降低数据集的维度。这样既能加快训练过程，又能在尽量减少数据方差损失的同时提升模型的可解释性。根据公式（2），应首先计算协方差矩阵 Σ，以获得用于降维投影的矩阵。

$$\Sigma = \frac{1}{m}\sum_{i=1}^{m}\left(X^{(i)}\right)\left(X^{(i)}\right)^{T} \quad (2)$$

F. 数据划分

我们将数据集划分为训练集和测试集，以便全面测试模型。数据划分如下：80% 用于训练，20% 用于测试。通过分层抽样保留了各子集内类别的分布。模型构建和交叉验证均使用训练集完成。使用测试集，我们能够了解最终模型在新数据上的表现。

G. 所提出的 AdaBoost 模型

自适应提升（AdaBoost）[22], 最初由 Freund 和 Schapiro 于 1996 年提出。将弱分类器转化为强分类器是 AdaBoost 的任务。它要解决的是一个二元分类问题。AdaBoost 主要用于构建提升算法。AdaBoost 使用一棵小型决策树。树生成后，会立即评估其在每个训练实例上的表现。由于难以预测哪些训练数据会受益于增加权重，因此必须先决定应给予树的下一级多少关注。为了可靠地预测异常观测值，AdaBoost 会迭代训练数据样本，同时调整分类器权重 [23]。为了完善分类器，需要在许多加权训练样本上对其进行交互式训练。每轮训练中，它都力求通过降低训练误差，尽可能贴合这些样本。AdaBoost 中的每个弱学习器都会根据其在加权训练数据上的表现获得相应的贡献（权重），其计算方式如公式（3）所示：

$$\alpha_t = \frac{1}{2}\ln\left(\frac{1-\epsilon_t}{\epsilon_t}\right) \quad (3)$$

使得错误率较低的学习器对最终决策的贡献更大。任何弱学习器的错误权重如公式（4）所示：

$$\epsilon_t = \sum_{i=1}^{N} w_i^{(t)} \mathbb{I}(h_t(x_i) \neq y_i) \quad (4)$$

其中，误分类样本对误差的贡献更大，因此下一次迭代会着重处理更难的样本。最后，强分类器通过聚合所有弱学习器构建而成，如公式（5）所示：

$$F(x) = sign\left(\sum_{t=1}^{T} \alpha_t h_t(x)\right) \quad (5)$$

提升模型是借助其他弱分类器的力量所获得的模型。

H. 评估指标

机器学习模型的性能、效能和可靠性通过评估指标进行评估。这些指标对于确保模型稳健且具有泛化能力，以及评估其模式识别和预测能力，具有重要价值。本研究所用的指标详细说明了模型的性能。

1) 混淆矩阵 O

ne 一种测试分类器区分不同类别准确率的方法是使用混淆矩阵[24]。通过查看 TP 和 TN 数值，我们可以了解分类器何时正确分类数据；而查看 FP 和 FN 数值，则可以了解它何时分类错误。

• TP（真正例）：具有正向预测值且实际值为正的大量数据。• FP（假正例）：可用于正向和负向预测的信息量。• FN（假负例）：实际值为正的数据，以及实际值为负但包含预测值的数据。• TN（真负例）：预期值和实际值均为负的大量数据。

以下指标用于评估模型的性能：

准确率：用百分比表示准确率。它表示对实例进行预测的准确程度 [25]。其计算公式见公式（6）：

$$Accuracy = \frac{TP+TN}{TP+FP+TN+FN} \quad (6)$$

精确率：准确率的一种衡量指标，是正确答案占总决策数的比例。其计算方法是将 TP 总数除以 FP 总数与 TP 总数之和。见公式（7）。

$$Precision = \frac{TP}{TP+FP} \quad (7)$$

召回率：正类真正率占正类评估总数的比例即为召回率。其计算方法见公式（8）：

$$Recall = \frac{TP}{TP+FN} \quad (8)$$

F1分数：可使用 F 值衡量准确率。公式（9）将其表示为两个相互竞争概念之间的平衡：敏感度和精确率：

$$F1\ Score = \frac{2\ (Precision \times Recall)}{Precision \times Recall} \quad (9)$$

这一全面的评估流程能够详细评估模型性能的多个方面，从而支持识别并选择最合适、最有效的模型。

IV. 结果分析与讨论

该方法强调在 ZTA 中持续进行行为分析，以实现内部威胁检测。为评估 AdaBoost模型的有效性，研究人员使用一个全面的内部威胁数据集开展实验，该数据集模拟了零信任环境中的用户活动。实验采用一台配备 Intel Xeon W-2295 CPU @ 3.0 GHz 和 32 GB 随机存取存储器的工作站，并使用 Python、scikit-learn 及相关机器学习库实现模型。表 II 展示了 AdaBoost模型用于检测持续性内部威胁的结果。模型的准确率高达 98.0%，表明其在检测潜在内部威胁方面具有出色的整体预测能力。AdaBoost模型的精确率为 98.3、召回率为 98.0，表现均衡，能够准确地对恶意行为进行分类，同时将假阳性和假阴性的数量控制在较低水平。F1 分数达到 98.0%，也进一步表明该模型能够应对序列行为安全数据的复杂性。这些发现说明，AdaBoost模型适用于零信任安全模型中的持续监控和威胁缓解。

表 II. 零信任架构中用于内部威胁检测的 AdaBoost模型性能

矩阵	AdaBoost
准确率	98.0
精确率	98.3
召回率	98.0
F1 分数	98.0

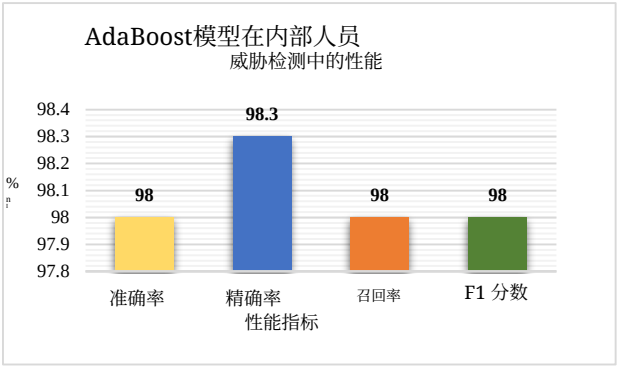

图 4. AdaBoost模型用于内部威胁检测的性能

表 II 和图 4 展示了 AdaBoost 模型在分类任务中的表现。该模型的准确率高达 98.0%，表明它能正确预测绝大多数实例。精确率最高，为 98.3%，体现了其在尽量减少假阳性的同时识别正类的强大能力。召回率和 F1 分数均为 98.0%，表明模型能够稳健地检测实际正类，并在精确率与召回率之间实现平衡。图中以 97.8% 至 98.4% 的百分比刻度展示了这些指标，凸显了 AdaBoost 在内部威胁检测方面始终如一的高性能和可靠性。

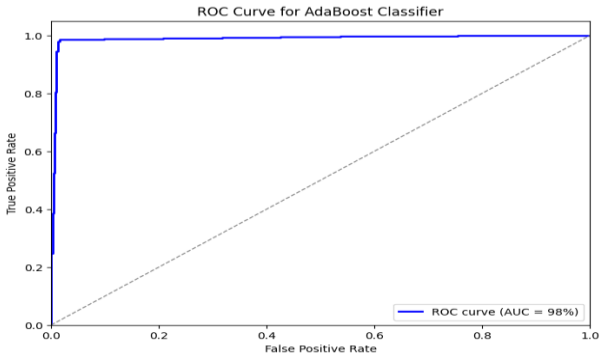

图 5. AdaBoost模型的受试者工作特征曲线

AdaBoost 分类器的 ROC 曲线如图 5 所示。该图展示了两个变量之间的关系：敏感度（或真正率）和特异度（或假阳性率

) 除以 1。模型的蓝色性能曲线非常接近图表的左上角，表明其表现出色。如文中明确指出的，其曲线下面积（AUC）高达 98%（或 0.98）。虚线对角线表示曲线下面积为 0.50 的随机分类器。曲线陡峭上升且接近理想点 (0,1)，这清楚地表明 AdaBoost模型能够出色地区分这两个群体。

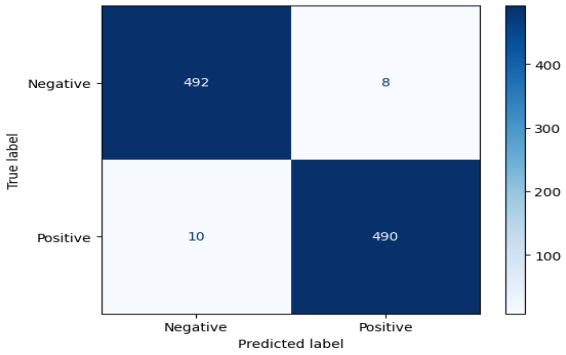

图6。AdaBoost模型的混淆矩阵

图 6 中的蓝色热力图展示了一个二元分类模型的混淆矩阵。矩阵的纵轴显示真实标签，横轴显示预测标签。两个标签均分为正类或负类。左上角单元格中有 492 个真负例（被准确识别为负类的实例）和 8 个假阳性（被错误预测为正类的负类实例）。另一方面，左下角显示 10 个错误的负类预测（正类案例）；右下角则显示 490 个正确分类的正类示例（真正例）。

A. 比较与讨论

表 III 展示了对零信任基础设施中用于持续内部威胁检测的各种机器学习模型进行基于准确率的评估结果。在区分良性与恶意内部人员行为方面，支持向量机取得了90.10%的准确率，表明其具有中等的预测能力。人工神经网络的准确率达到94.74%，性能略有提升，体现出更强的模式识别能力。同样，贝叶斯网络（BaysNet）达到94.87%的准确率，在检测能力方面实现了相当的平衡。相比之下，AdaBoost模型显著优于其他模型，准确率达到98.0%，证明其在零信任安全环境中的行为分析和持续威胁检测方面具有优越性。

表 III. 用于内部威胁检测的机器学习模型准确率比较

模型	准确率
BaysNet[26]	94.87
支持向量机[27]	90.10
人工神经网络[28]	94.74
AdaBoost	98.0

研究结果表明，与 SVM、人工神经网络和贝叶斯网络等传统模型相比，AdaBoost 提供了更优的内部威胁检测方法。其集成结构能够捕捉行为的细微偏差，而这些偏差可能会被其他模型忽略。对比分析显示，传统分类器虽然擅长识别总体趋势，但无法对内部人员活动中出现的细微异常作出响应。结果表明，应将行为分析与机器学习的自适应方法相结合，应用于零信任系统。通过跟踪用户活动，组织能够提前防范内部威胁，同时增强整体系统安全性和韧性。

V. 结论与未来展望

将行为分析与机器学习相结合，已被证明能有效支持零信任架构中的持续内部威胁检测。数据清洗、SMOTE、归一化和主成分分析等预处理方法确保了高质量的特征表示。对比实验表明，AdaBoost 的表现显著优于传统模型。支持向量机（准确率为 90.1%）、人工神经网络（94.7%）和贝叶斯网络（94.9%）取得了中等成效，而 AdaBoost 的表现更佳，准确率为 98.0%、精确率为 98.3%、召回率为 98.0%，F1 分数为 98.0%。ROC分析确认曲线下面积为 0.98，表明其具有卓越的判别能力和均衡的性能。这些结果凸显了 AdaBoost 在检测细微行为异常、同时减少假阳性和假阴性方面的优势，因此非常适用于动态零信任环境。未来研究可聚焦于实时部署，以验证其在企业级工作负载下的可扩展性。研究范围若超越 CERT数据集，还可纳入击键动态特征、生物特征信号和网络模式等多模态数据，以丰富行为画像。包括图神经网络、强化学习和自监督方法在内的先进学习技术，或可增强系统对不断演变威胁的适应性。融入可解释人工智能也是关键一步，可提高模型输出的透明度，增强分析人员的信任，并提供可付诸行动的洞见。这些方向共同推动内部威胁检测发展为面向零信任安全框架的主动、可扩展且可解释的解决方案。

参考文献

[1] S. F. Ahmed 和 N. A. Hikal，《现代企业网络安全测量与评估方法综述》，《JOIV信息学与可视化国际期刊》，第3卷第3期，第157–176页，2019年8月，doi: 10.30630/joiv.3.3.241。[2] S. Neeli，《NoSQL数据库的重要性：战略业务方法与管理技术》，《高级开发研究期刊》，第10卷，11，2019年1期，页码未标明。[3] J. Modini、M. Vanzomeren、S. Fowler、K. Joiner 和 T. Lynar，《迎接中等强国面临的内部威胁挑战》，载于《国际网络战与安全会议》，2020年。doi：10.34190/ICCWS.20.131。[4] P. Pathak、A. Shrivastava 和 S. Gupta，《容迟网络中各种安全问题综述》，《高级Shell程序期刊》，第2卷第2期，第12–18页，2015年。[5] V. M. L. G. Nerella，《自动化跨平台数据库迁移与高可用性实现》，《土耳其计算机与数学教育期刊》，第9卷第2期，第823–835页，2018年。[6] A. Balasubramanian，《利用主动机器学习方法打击金融领域的洗钱行为》，《国际工程、多学科与物理科学创新研究期刊》，第7卷第2期，第1–15页，2019年。doi：10.5281/zenodo.14508474。[7] M. K. Omopariola，《新兴经济体中的零信任架构部署：尼日利亚案例研究》，《国际计算机应用技术研究期刊》，第5卷第12期，2016年。[8] A. Balasubramanian，《人工智能赋能的需求响应：实现更智能能源管理的框架》，《国际核心工程与管理期刊》，第5卷第6期，第96–110页，2018年。[9] D. S. Berman、A. L. Buczak、J. S. Chavis 和 C. L. Corbett，《网络安全深度学习方法综述》，《Information》，第10卷第4期，2019年。doi：10.3390/info10040122。[10] S. S. S. Neeli，《无服务器数据库：一种经济高效且可扩展的解决方案》，《国际工程、多学科与物理科学创新研究期刊》，第7卷第6期，第7页，2019年。[11] M. Zamani 和 M. Movahedi，《用于入侵检测的机器学习技术》，arXiv预印本 arXiv1312.2177，2013年。[12] Tamanna，《基于LSTM--CNN模型的深度学习内部威胁检测》，都柏林，爱尔兰国立学院，2020年。

[13] D. C. Le、N. Zincir-Heywood 和 M. I. Heywood，《使用机器学习分析内部威胁检测的数据粒度级别》，IEEE网络与服务管理汇刊，第17卷第1期，第30–44页，2020年。doi：10.1109/TNSM.2020.2967721。[14] M. Singh、B. M. Mehtre 和 S. Sangeetha，《利用集成方法进行用户行为分析以检测内部威胁》，载于 ISBA 2019——第五届IEEE身份、安全与行为分析国际会议，2019年。doi：10.1109/ISBA.2019.8778466。[15] T. Hu、W. Niu、X. Zhang、X. Liu、J. Lu 和 Y. Liu，《基于鼠标动态特征与深度学习的内部威胁检测方法》，《安全与通信网络》，第2019卷，2019年。doi：10.1155/2019/3898951。[16] A. J. Hall、N. Pitropakis、W. J. Buchanan 和 N. Moradpoor，《利用组织数据和异构堆叠分类器预测恶意内部威胁场景》，载于《2018年IEEE大数据国际会议（Big Data）》，2018年，第5034–5039页。doi：10.1109/BigData.2018.8621922。[17] A. Gamachchi 和 S. Boztas，《通过属性图聚类检测内部威胁》，载于《2017年IEEE Trustcom/BigDataSE/ICESS会议》，2017年，第112–119页。doi：10.1109/Trustcom/BigDataSE/ICESS.2017.227。[18] S. S. S. Sindhu、S. Geetha 和 A. Kannan，《基于决策树并采用封装式方法的轻量级入侵检测》，《专家系统与应用》，第39卷第1期，第129–141页，2012年。[19] D. Gonzalez-Cuautle 等，《用于优化僵尸网络和入侵检测系统数据集分类任务的合成少数类过采样技术》，《应用科学》第10卷第3期，第794页，2020年1月。doi：10.3390/app10030794。[20] I. A. Khan、D. Pi、Z. U. Khan、Y. Hussain 和 A. Nawaz，《HML-IDS：一种用于SCADA系统入侵检测的混合多层异常预测方法》，《IEEE开放获取》，第7卷，第89507–89521页，2019年。doi：10.1109/ACCESS.2019.2925838。

[21] U. A. Korat 和 A. Alimohammad，“用于主成分分析的可重构硬件架构”，《电路、系统与信号处理》，第38卷，第5期，第2097–2113页，2019年5月，doi: 10.1007/s00034-018-0953-y。[22] A. B. Abhale 和 S. S. Manivannan，“用于确定无线传感器网络入侵检测异常类型的监督式机器学习分类算法方法”，《光学存储器与神经网络》，第29卷，第3期，第244–256页，2020年7月，doi: 10.3103/S1060992X20030029。[23] B. Mueller，“利用机器学习技术和信息安全事件管理（SIEM）增强网络安全中的隐蔽威胁检测”，2020年。[24] A. Yulianto、P. Sukarno 和 N. A. Suwastika，“ 提升基于 AdaBoost 的入侵检测系统（IDS）在 CIC IDS 2017数据集上的性能”，《物理学杂志：会议系列》，第1192卷，2019年3月，doi: 10.1088/1742-6596/1192/1/012018。[25] S. Zhao、W. Li、T. Zia 和 A. Y. Zomaya，“用于物联网中基于异常的入侵检测的降维模型和分类器”，载于《2017年电气电子工程师学会第15届国际可靠性、自主性与安全计算会议、第15届国际普适智能与计算会议、第3届国际大数据智能与计算及网络科学与技术大会（DASC/PiCom/DataCom/CyberSciTech）》，电气电子工程师学会，2017年11月，第836–843页。doi: 10.1109/DASC- PICom-DataCom-CyberSciTec.2017.141。[26] D. C. Le、S. Khanchi、A. N. Zincir-Heywood 和 M. I. Heywood，“针对内部威胁检测的进化计算方法基准测试”，载于《遗传与进化计算大会论文集》，2018年7月，第1286–1293页。doi: 10.1145/3205455.3205612。[27] O. Almomani，“基于粒子群优化、灰狼优化算法、萤火虫算法和遗传算法的网络入侵检测系统特征选择模型”，《对称性》（巴塞尔），第12卷，第6期，2020年，doi: 10.3390/sym12061046。[28] M. Al-Zewairi、S. Almajali 和 M. Ayyash，“使用浅层和深层人工神经网络分类器检测未知安全攻击”，《电子学》，第9卷，第12期，2020年，doi: 10.3390/electronics9122006。