斯普林格计算机科学简报

系列编辑

斯坦·兹多尼克，布朗大学，普罗维登斯，RI，美国沙西·谢卡尔，明尼苏达大学，明尼阿波利斯，MN，美国吴信东，佛蒙特大学，伯灵顿，VT，美国拉克米·C·贾因，南澳大利亚大学，阿德莱德，SA，澳大利亚大卫·帕杜亚，伊利诺伊大学厄巴纳-香槟分校，厄巴纳，IL，美国徐民·谢尔曼·申，滑铁卢大学，滑铁卢，ON，加拿大博尔科·弗尔特，佛罗里达大西洋大学，博卡拉顿，FL，美国V·S·苏布拉马尼安，马里兰大学，学院公园，MD，美国马歇尔·赫伯特，卡内基梅隆大学，匹兹堡，PA，美国池克树，东京大学，东京，日本布鲁诺·西西利亚诺，那不勒斯费德里科二世大学，那不勒斯，意大利苏希尔·贾乔迪亚，乔治梅森大学，费尔法克斯，VA，美国牛顿·李，教育研究和奖学金研究所，洛杉矶，CA，美国

斯普林格简报提供了对各个领域尖端研究和实际应用的简明总结。 该系列包含了50至125页的紧凑卷册，涵盖了从专业到学术的各种内容。

典型的主题可能包括：

- 最新分析技术的及时报告
- 作为期刊文章中新研究结果和背景文献综述之间的桥梁

- 热门或新兴主题的快照
- 深入的案例研究或临床示例
- 介绍学生必须理解的核心概念，以便做出独立贡献

简报允许作者以最少的时间投入来展示他们的想法，并让读者吸收这些想法。 简报将作为斯普林格的电子书系列的一部分出版，全球数百万用户可使用。 此外，简报还可供个人购买纸质版和电子版。 简报的特点是快速、全球范围内的电子传播、标准出版合同、易于使用的稿件准备和格式指南，以及加快的出版进度。 我们的目标是在接受后8-12周内出版。 本系列考虑出版征稿和非征稿的稿件。

索引：该系列已被Scopus、Ei-Compendex和zbMATH索引

马尔万·奥马尔

机器学习用于网络安全

创新的深度学习解决方案

Springer

马尔万·奥马尔
ITM和网络安全系
伊利诺伊理工学院
芝加哥，伊利诺伊州，美国

ISSN 2191-5768 ISSN 2191-5776（电子版）
斯普林格计算机科学简报
ISBN 978-3-031-15892-6 ISBN 978-3-031-15893-3（电子书）
https://doi.org/10.1007/978-3-031-15893-3

这本斯普林格印记由注册公司斯普林格自然瑞士股份有限公司出版
注册公司地址为：瑞士钻石街11号，6330 Cham

首先，我想把这本书献给我美妙的妻子玛哈，感谢她激励我进行这个书籍项目，并在整个旅程中支持我。

没有她坚定的支持，这个项目不会见到光明！其次，这本书也献给了我的了不起的孩子们：塔拉和亚当，他们总是激励我追求不可能。

第三，我将这项工作献给我关心和可爱的父母：戈齐和达哈尔，他们总是鼓励我做到最好，追求我的教育梦想。

最后，这本书也献给我可爱的兄弟们：费萨尔、马津、马赫尔、哈齐姆和西南，他们始终相信我，并在我的教育旅程中支持我。

目录

第一章
应用机器学习（ML）解决网络安全威胁

摘要：随着网络安全威胁在规模、频率和影响力上呈指数级增长，传统的威胁检测系统已经被证明是不足够的。 这促使人们开始使用机器学习（以下简称ML）来解决这个问题。
但随着组织越来越多地使用智能网络安全技术，这些基于ML的数字安全系统的整体效力和效益分析成为一个日益受到学术界关注的课题。 本研究旨在通过展示ML基于数据分析技术在网络安全领域的各种问题领域中的应用，扩展和补充这一日益增长的文献库。 为了实现这一目标，采用了对现有学术文献进行快速证据评估（REA）的方法。 目的是呈现ML被应用于帮助解决网络安全威胁挑战的各种方式的快照。

关键词：机器学习安全·深度学习算法·网络安全人工智能·数据分析与网络安全·网络攻击·安全威胁

1.1 引言

网络安全威胁对个人、组织甚至政府造成的损害，无论是立即的还是长期的，都可能是巨大且极其昂贵的。 例如，在2021年，根据可获得的最新信息，研究表明，网络安全攻击导致至少47%的受影响组织遭受严重的工作中断或生产停工。 个人可识别信息（以下简称PII）的丢失是另一个重大影响，影响了近46%的组织（图1.1）。总体而言，这些中断的财务影响是巨大的，并且在最近的过去中有显著增加。 例如，在2015年至2020年期间，向互联网犯罪投诉中心（IC3）报告的与网络犯罪相关的损失从估计的10亿美元增加到42亿美元以上[9]。 除了对个人和组织的影响外，网络安全漏洞还威胁着金融基础设施的根基，并且可能

机器学习在网络安全中，斯普林格计算机科学简报，https://doi.org/10.1007/978-3-031-15893-3_1

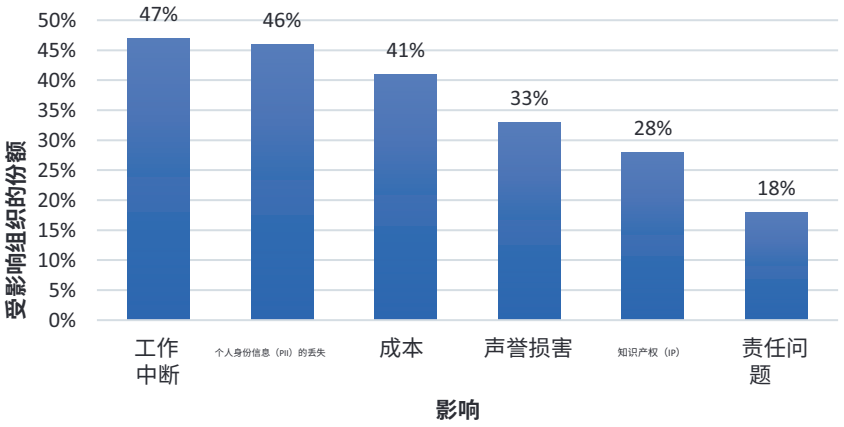


图**1.1** 网络安全攻击对全球组织的影响。 该图描述了202**1**年全球组织遭受网络安全攻击的主要影响。 除了工作中断外，网络攻击还可能导致客户个人身份信息（PII）的丢失（46%），为解决问题而额外支付外部服务的成本（4**1**%），知识产权（IP）的盗窃（28%），持续生产力受损（22%）和业务停工（**1**5%）。 （来源：Sava [**1**6]）

如果被对抗性代理人（包括国家行为者和恐怖组织）部署，并针对交通和能源系统等关键基础设施，网络威胁甚至可能对国家经济和安全构成重大威胁[3]。 因此，缓解网络威胁一直是当今最紧迫的问题之一。

但是，随着网络安全威胁在规模、频率和影响力上呈指数级增长，基于传统的威胁检测系统已被证明是不足够的[14]。 这促使使用机器学习（以下简称ML）来帮助解决这个问题[3, 7]。 但是，随着组织越来越多地使用智能网络安全技术，这些基于ML的数字安全系统的整体功效和效益分析仍然是一个日益受到学术界关注的课题。 本研究旨在通过展示ML基于数据分析技术在网络安全各个问题领域的应用，扩大并增加这一日益增长的文献库。

为了实现这一目标，采用了现有学术文献的快速证据评估（REA）方法来对主题进行概述。 目的是展示ML被应用于帮助解决网络安全威胁挑战的各种方式的快照。

1.2 方法论

市场数据显示，数字安全威胁从恶意软件和病毒攻击到更复杂的网络攻击形式（如分布式拒绝服务攻击和高级持续性威胁）在规模、频率和整体影响方面呈指数增长。 研究表明，包括Rupp [14]的开创性发现在内，其中一个主要原因是网络空间的规模不断增大。

这是一个极大地扩展了可用威胁面（电子商务、物联网、远程办公和自带设备等）的现象。 因此，当代个人和职业生活的方方面面都变得天然易受网络安全风险的影响。 此外，网络犯罪分子变得更加复杂、协调和资源充足，甚至包括国家。 加上“网络犯罪即服务”模式、攻击者采用的DevOps以及几乎所有计算服务的云化等因素，这些因素以及其他一些因素（如加密货币的泛滥）使得网络攻击者不仅能够积累预算和数据，还能够进行研发，创建更加强大和有影响力的攻击模型，具有更高的数量、多样性和速度[14]。 对此，人们开发和使用智能网络安全系统，如机器学习。为了了解机器学习在这方面的应用，采用了REA方法论的方法。

REA是一种快速回顾现有文献的方法，提供关于特定主题的基于证据的解决方案或信息。 这种技术现在由政府社会研究(GSR)网站提出和推进，作为评估关于特定问题已知的内容的方法，通过系统地回顾和批判性评估现有研究，在多个领域，包括计算机科学[6]中找到了广泛的应用。

除了其缺点，如范围和内容的不足，REA提供了一种严格和明确的方法，通过对多个文献进行评估，以回答感兴趣的研究问题(RQ) [6]。 在我们的案例中，RQ是机器学习应用在解决网络安全威胁方面的角色是什么。 因此，来自ScienceDirect、Wiley Online Library、Google Scholar和Elsevier等电子数据库的相关文献对于评估机器学习在网络安全中的潜在应用至关重要。

1.2.1 文献综述

近年来，机器学习和更广泛的人工智能(以下简称AI)领域的兴趣和整体进展显著增加，新的应用在多个领域被积极追求[13]。 与此同时，世界所依赖的数字通信技术也带来了众多安全问题：网络攻击不仅数量上升，而且频率和影响规模也在增加，引起了对网络系统易受攻击的关注，以及提高其安全性的重要性和整体需求[13]。

例如，在过去的5年中，IC3每年至少收到55.2万个网络安全投诉（图1.2）。

这些投诉涵盖了影响全球互联网用户的各种问题。 在这个快速发展的环境中，决策者、研究人员和安全从业人员对于机器学习在增强网络安全方面的潜在应用存在重大和合理的关注[13]。 已经有几项研究探讨了这个问题。

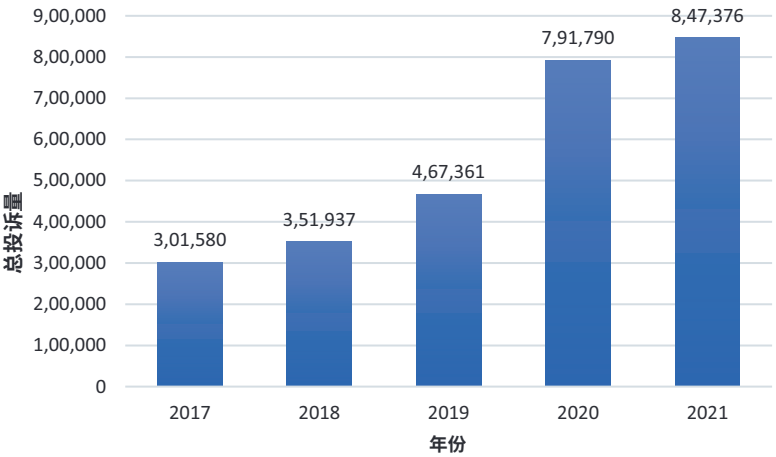


图**1.2** 过去5年的投诉和损失。 (来源：联邦调查局[**F**BI] [4])

Handa等人[8]对机器学习在网络安全中的一些应用进行了文献综述。 他们还研究了对机器学习使用的对抗威胁。 该综述侧重于应用领域，包括（1）电力系统安全，（2）工业控制系统（ICS）按区域划分的网络攻击检测，（3）SCADA系统入侵检测，（4）车载自组织网络（VANETs）入侵检测，以及（5）恶意软件分析。 电力系统安全涉及保护电力系统免受针对压力系统的有意攻击，以造成级联停电。 在这里，基于决策的机器学习算法可以增强网络攻击面，并通过分析诸如电压幅值、电流幅值和角度差异等预测因素来检测这些系统中的压力条件。 机器学习还可以帮助确定在电力系统中部署所需的相量测量单元（PMUs）（用于估计电流或电压的幅值和相位角的设备）的理想数量以及最关键的位置。

在检测ICS上的网络攻击时，机器学习算法可以为保护关键基础设施创建自动入侵检测系统（IDS）。 ICS中最容易受到攻击的组件是可编程逻辑控制器（PLC）- 一个配备有操作系统用于控制机器操作的小型计算机[8]。 如果蠕虫攻击PLC，可能会影响机器的工作和运行。 在这里，基于机器学习的IDS可以通过训练算法来检测数据中的异常模式，从而检测隐藏的网络攻击。

第三，机器学习可以检测用于监控电网和水传输系统等关键基础设施的SCADA系统上的持续入侵。
基于机器学习的入侵检测系统可以识别攻击并发出警报，有助于采取补救措施。 支持向量机是一个机器学习算法的例子，用于区分不同的情况。

在正常和恶意流量之间。 基于机器学习的入侵检测系统也可以用于车载自组织网络（VANETs）——一种将车辆与路边基站连接起来提供有价值的安全和交通信息的无线网络[1]。 这些网络容易受到窃听和干扰等攻击。 在这里，分布式的基于机器学习的入侵检测系统可以帮助节点检测网络上的异常或恶意行为，通过允许节点共享威胁分类数据。

最后，机器学习可以用于恶意软件分析，特别是那些在每次感染后不断改变结构或代码的多态和元-变态恶意软件。 令人担忧的是零日恶意软件带来的威胁——攻击者可以在软件开发者发现和解决之前利用新发现的漏洞。

[10]. 这种攻击无法通过基于签名的方法解决。 在这里，机器学习算法可以检测已知的恶意软件并提供信息来帮助检测新的恶意软件。 基于机器学习的恶意软件分析分为两个阶段：（1）训练阶段，在这个阶段算法提取关于良性和恶意数据的关键信息，并用它创建一个预测模型；（2）测试阶段，在这个阶段创建的预测模型评估未知数据，以确定它们是否对系统构成威胁（见图1.1）。在检测第二代恶意软件中常用的机器学习技术包括决策树、神经网络、深度学习和数据挖掘。

Ford和Siraj [5]研究了机器学习在网络安全中的应用。
这些应用包括网络入侵检测、钓鱼检测、智能电表能源使用的个人化配置、社交网络中的垃圾邮件检测、密码学、人机交互验证和基于击键的身份验证。 在钓鱼检测中，可以使用基于机器学习的分类器，如逻辑回归、神经网络、随机森林、支持向量机和贝叶斯回归树。 其中，逻辑回归具有最高的精确率。 在网络入侵检测中，基于机器学习的检测系统更可取，因为它们可以适应新的和未知的攻击。 例如，这些系统可以将支持向量机（SVM）与误用检测方法相结合，以提高异常识别的准确性。

Mathew [11] 研究了机器学习在解决网络威胁实现概率方面的应用现状。通过文献综述，该研究确定了机器学习在网络安全中的四个主要应用领域：（1）垃圾邮件检测，（2）恶意软件分析，（3）入侵检测系统（IDS），以及（4）安卓（移动）恶意软件检测。

Das和Morris [2]研究了如何在网络分析中使用机器学习来检测入侵、分类流量和电子邮件过滤。 研究结果确定了网络安全中常见的机器学习技术，包括贝叶斯网络、决策树、聚类、人工神经网络（ANN）、遗传算法和编程、归纳学习以及隐马尔可夫模型（HMM）。 例如，贝叶斯网络可以用于检测异常和已知的攻击模式和签名。 对一些样本贝叶斯网络模型进行的初步测试显示，探测和拒绝服务（DoS）攻击的检测率分别为99%和89%，性能率为89%。 机器学习决策树可以通过允许并行评估和与攻击签名进行比较来提高流量处理速度。 聚类算法可以用于实时检测已知的

攻击签名。 这些算法的样本模型还展示了在检测未知（零日）攻击方面的70-80%的准确率[2]。 总体而言，该研究推荐了三个网络安全领域的机器学习算法：（1）误用检测（例如决策树和遗传算法），（2）异常检测（即聚类算法），以及（3）入侵检测系统。

Salloum等人[15]对机器学习和深度学习技术在检测网络入侵中的文献调查进行了回顾。 与Das和Morris [2]类似，该研究确定了决策树、逻辑回归、朴素贝叶斯和随机森林等常见的机器学习算法用于检测网络入侵。

深度学习——一种新的机器学习分支，在检测基于流的异常方面更加有效。 与机器学习一起使用的常见网络安全数据集包括：（1）ISOT（信息安全与对象技术）数据集，包括1675424个总流量，可以检测异常流量；（2）HTTP CSIC 2010数据集用于Web攻击检测，包括约6000个正常请求和25000个异常请求；（3）CTU-13数据集包括13种不同的恶意软件签名，用于检测僵尸网络流量；（4）UNSW-NB15数据集包括一个小时的匿名跟踪数据，用于DDoS攻击和蠕虫、侦察、DoS、后门和模糊器等9种主要攻击的检测。 因此，机器学习及其独特的数据集可以增强现代网络的网络安全性。

最后，Musser和Garriott [12]评估了机器学习在检测和拦截网络安全攻击方面的潜力，其速度比传统方法更高。
使用四阶段的网络安全模型-预防、检测、响应和恢复以及主动防御-研究分析了最新机器学习创新的影响。
分析得出了四个发现。 首先，机器学习在检测和分类潜在攻击方面更准确。 其次，机器学习可以使几个网络安全任务部分或完全自动化，例如漏洞发现和攻击干扰。
第三，机器学习可以从以前未充分利用的防御策略中提供增量网络安全进展。 第四，不幸的是，机器学习可能改变威胁环境，并使某些策略对攻击者更具吸引力。
传统上，机器学习已用于三个任务-垃圾邮件、入侵和恶意软件检测[12]。 基于机器学习的垃圾邮件检测通过从邮件头中提取更多技术词汇或短语（例如服务器信息和IP地址）提供更好的垃圾邮件分类器。 更复杂的算法可以跟踪用户过去的电子邮件交易，并通过使用深度学习模型标记异常联系人或评估品牌邮件的真实性。 在入侵检测中，机器学习可以用于自动化某些基于误用的检测（通过学习不同类型攻击的显著特征来识别攻击），从而消除了人工创建触发警报的规则列表的需要。 相反，异常行为检测会标记与正常或基线操作不同的任何行为。 在恶意软件检测中，机器学习可以帮助扫描和与恶意软件签名列表进行文件内容匹配。 新的机器学习架构，如深度学习、生成对抗网络（GANs）、强化学习和大规模自然语言模型正在在网络安全的四个阶段中得到采用（参见图1.3、1.4、1.5和1.6）。

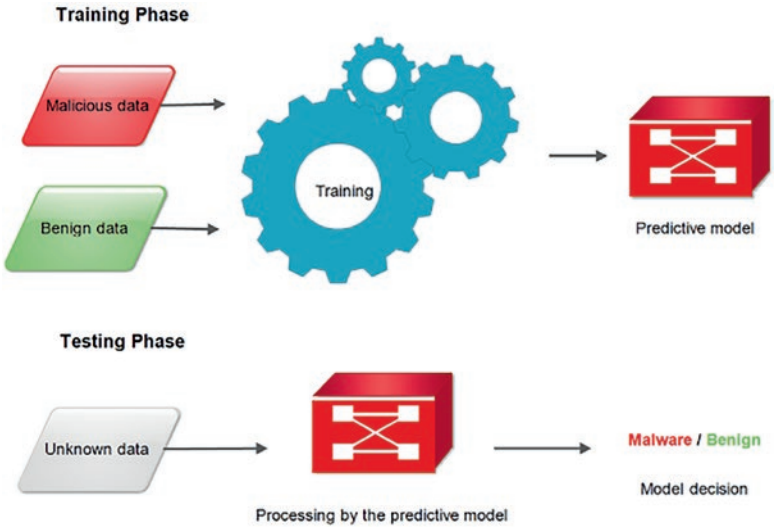


图**1.3** 基于机器学习的恶意软件分析。（来源：Handa等人[8]）

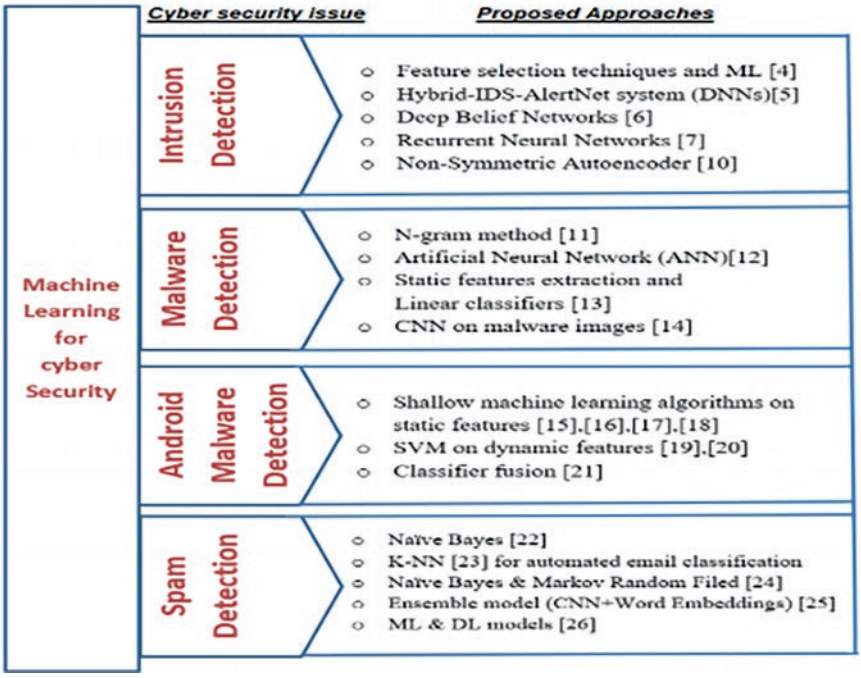
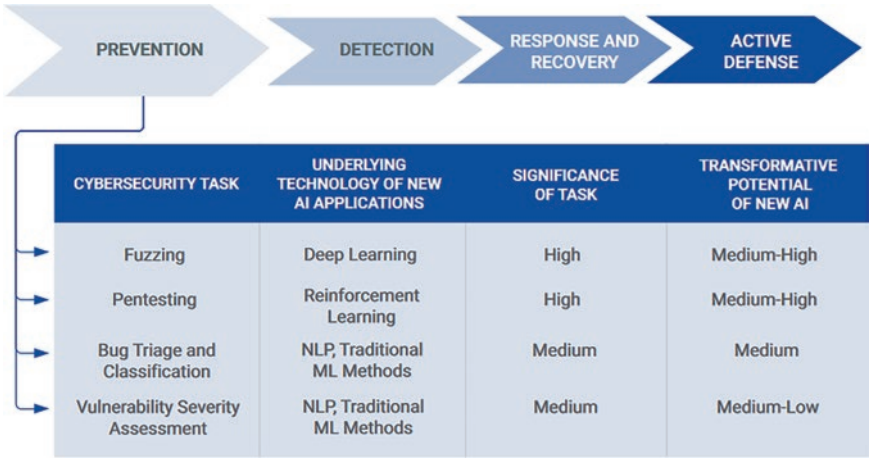
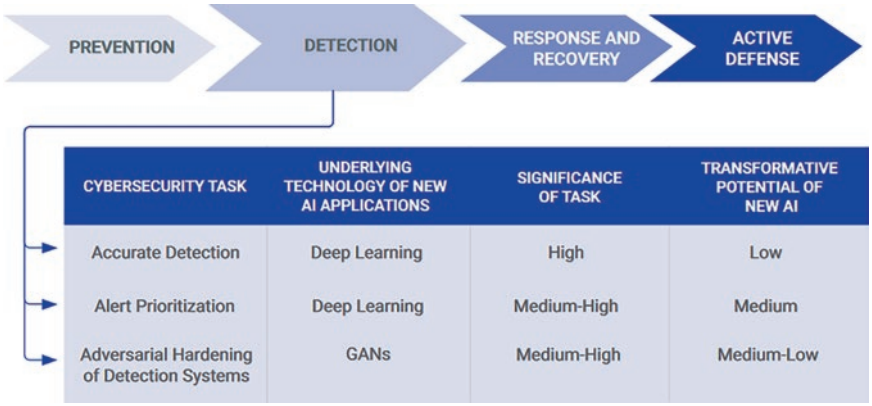


图**1.4** 解决特定网络安全问题的主要机器学习算法。（来源：Mathew[**11**]）



CYBERSECURITY TASK	UNDERLYING TECHNOLOGY OF NEW AI APPLICATIONS	SIGNIFICANCE OF TASK	TRANSFORMATIVE POTENTIAL OF NEW AI
Fuzzing	Deep Learning	High	Medium-High
Pentesting	Reinforcement Learning	High	Medium-High
Bug Triage and Classification	NLP, Traditional ML Methods	Medium	Medium
Vulnerability Severity Assessment	NLP, Traditional ML Methods	Medium	Medium-Low

图**1.5** 预防的新机器学习应用。（来源：Musser和Garriott[12]）

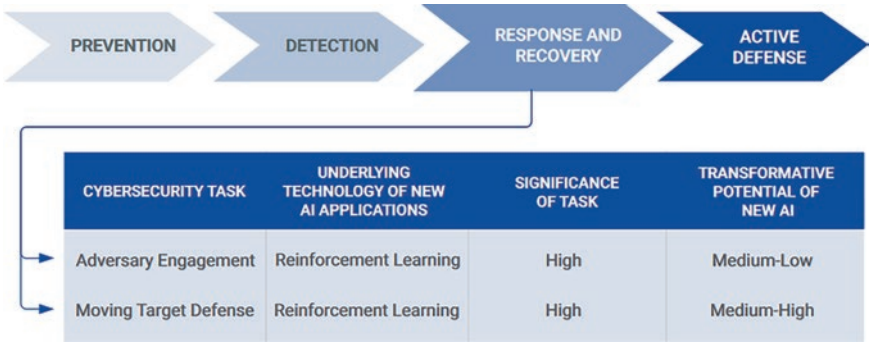


CYBERSECURITY TASK	UNDERLYING TECHNOLOGY OF NEW AI APPLICATIONS	SIGNIFICANCE OF TASK	TRANSFORMATIVE POTENTIAL OF NEW AI
Accurate Detection	Deep Learning	High	Low
Alert Prioritization	Deep Learning	Medium-High	Medium
Adversarial Hardening of Detection Systems	GANs	Medium-High	Medium-Low

图**1.6** 检测的新机器学习应用。（来源：Musser和Garriott[12]）

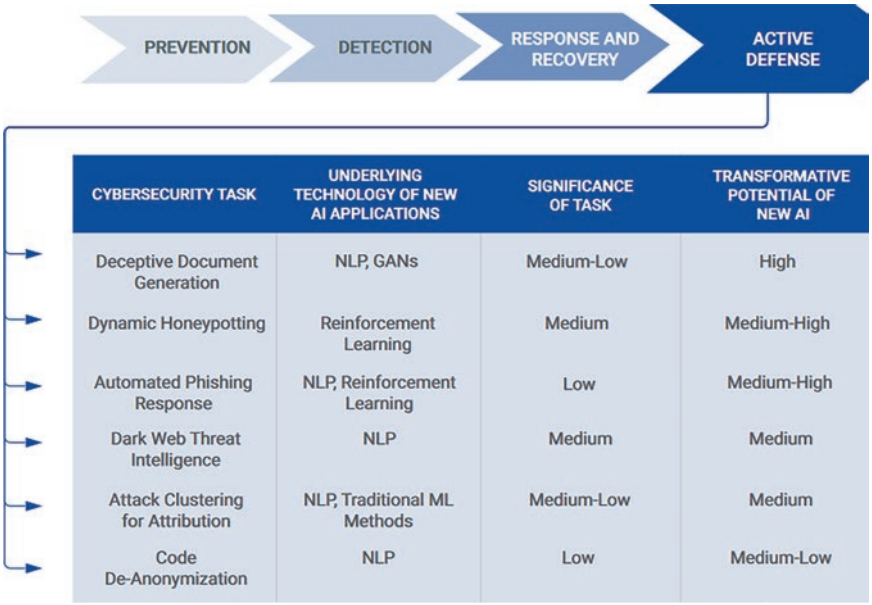
预防涉及搜索和修补漏洞。例如，模糊测试器在代码中搜索漏洞，而渗透测试（穿透测试）在网络中搜索公开已知的漏洞以及不安全的配置。对于运行多个代码的大型组织来说，渗透测试可能既昂贵又耗时，因此使用基于强化学习的AI代理来更有策略地进行此类测试。错误分类和严重性评估涉及使用机器学习来识别最关键的漏洞。

检测是深度学习和新的机器学习模型可能产生变革性影响的阶段。然而，这样的深刻突破尚未发生，因为许多公司仍然使用更简单的检测模型。如今，深度学习的最小使用涉及提供可以帮助更好地优先处理威胁的复杂分析。准确的分析意味着调查时较少的虚假警报。



CYBERSECURITY TASK	UNDERLYING TECHNOLOGY OF NEW AI APPLICATIONS	SIGNIFICANCE OF TASK	TRANSFORMATIVE POTENTIAL OF NEW AI
Adversary Engagement	Reinforcement Learning	High	Medium-Low
Moving Target Defense	Reinforcement Learning	High	Medium-High

图**1.7** 响应和恢复的新机器学习应用。（来源：Musser和Garriott [12]）



CYBERSECURITY TASK	UNDERLYING TECHNOLOGY OF NEW AI APPLICATIONS	SIGNIFICANCE OF TASK	TRANSFORMATIVE POTENTIAL OF NEW AI
Deceptive Document Generation	NLP, GANs	Medium-Low	High
Dynamic Honeypotting	Reinforcement Learning	Medium	Medium-High
Automated Phishing Response	NLP, Reinforcement Learning	Low	Medium-High
Dark Web Threat Intelligence	NLP	Medium	Medium
Attack Clustering for Attribution	NLP, Traditional ML Methods	Medium-Low	Medium
Code De-Anonymization	NLP	Low	Medium-Low

图**1.8** 主动防御的新机器学习应用。（来源：Musser和Garriott [12]）

此外，生成对抗网络可以用于加固机器学习系统，以预测潜在攻击（图1.7）。

人工智能和机器学习可以通过两种方式支持响应和恢复：（1）准确分类正在进行的攻击并选择适当的响应策略，以及（2）自动化决策，例如施加用户限制以限制感染并将受损机器与网络隔离（图1.8）。

主动防御涉及应对新威胁的策略。防御的三个领域适用于机器学习：（1）欺骗，（2）威胁情报和（3）归因

[12]。欺骗是最简单的防御机制，涉及使用机器学习生成逼真的文档、文件或活动配置文件来误导攻击者。

威胁情报涉及收集有关潜在对手的信息，以预测攻击并创建更强大的防御措施。 机器学习可以通过对暗网用户进行聚类或文本挖掘来提供帮助。 归因涉及将攻击追溯到特定的对手。

在这里，机器学习模型可以通过聚类信息来识别具有相似属性或描述的攻击，这些属性或描述与先前的攻击相似。

1.3 结论

互联网连接设备、系统和网络的广泛使用加剧了对网络安全策略的研究。 尽管传统的网络安全方法（如杀毒软件/反恶意软件程序、加密、防火墙和入侵检测系统）仍然提供对已知攻击模式和签名的基线防御，但它们越来越无法应对不断出现的复杂网络威胁。 随着信息和通信技术的进步，攻击者用于攻击计算机网络的方法和技术也在不断发展。 这一趋势的一个危险后果是零日攻击的风险加大，这对互联网连接的工业控制系统和关键基础设施构成严重威胁，从而威胁到社会稳定和国家安全。 因此，人工智能和机器学习在预防、检测、响应和防御策略方面可以发挥关键作用，特别是对于具有广泛数据能力的公司。 此外，机器学习可以实现许多常规网络安全任务的自动化，这些任务占据管理员大部分时间，或者增加新的信息流以提高管理员的情境意识。

然而，当前自动化的范围受到专有业务操作的限制，使得网络安全系统的标准化变得困难。 此外，自动化需要明确的目标，而计算机系统和网络的不同目标使得情况更加复杂。

影响当代网络安全的另一个主要挑战是难以跟上产生的大量安全警报。 未来，机器学习模型的进一步发展预计将解决这些挑战，尽管创新可能是渐进的，正如过去的发展所证明的那样。 这些进步也无疑将增强新一代网络安全专家的培训。

参考文献

1. Chadha, R. D. (2015). 车载自组织网络（VANETs）：一项综述。国际计算机与通信工程创新研究杂志，*3*（*3*），*2339–2346*。
https://www.rroij.com/open- access/vehicular- ad- hoc- network- vanets- a- review.pdf

2. Das, R., & Morris, T. H. (2017). 机器学习和网络安全。2017年国际计算机、电气和通信工程会议（*ICCECE*）。htt*p*s://doi.org/10.1109/iccece.2017.8526232
3. Dua, S., & Du, X. (2016).数据挖掘和机器学习在网络安全中。CRC出版社。
4. 联邦调查局（FBI）。 （2021）。2021年互联网犯罪报告。 htt*p*s://www.ic3.gov/Media/PDF/AnnualReport/2021_IC3Report.pdf
5. Ford, V., & Siraj, A. (2014). 机器学习在网络安全中的应用。*I*S*C*A第*27*届国际工业和工程应用计算机会议（*CAINE*-2014）。 https://vford.me/papers/Ford%20Siraj%20Machine%20Learning%20in%20Cyber%20Security%20final%20manuscript.pdf
6. Grant, M. J., & Booth, A. (2009). 一种评论的分类学：对14种评论类型和相关方法的分析。健康信息与图书馆杂志，26 (2)，91–108。 https://doi.org/10.1111/j.1471- 1842.2009.00848.x
7. Halder, S., & Ozdemir, S. (2018).亲身体验网络安全的机器学习：利用*Python*生态系统使您的机器智能，保护您的系统。*Packt Publishing*。
8. Handa, A., Sharma, A., & Shukla, S. K. (2019). 机器学习在网络安全中的应用：一项综述。*WIREs*。数据挖掘与知识发现，*9* (4)。 h*t*t*p*s*://doi.org/*10.1002/*wid*m.1306
9. Johnson, J. (2021, March 18). 网络犯罪：*IC3* 2020年报告的损失。Statista. 从https://www.statista.com/statistics/267132/中检索到的2022年5月4日的美国网络犯罪造成的总损害
10. 卡巴斯基（2021年6月17日）。什么是零日攻击？—定义和解释。www.kaspersky.*c*om。从*htt*ps://www.kaspersky.*co*m/reso*u*rce-cen*t*er/*defi*n*itio*ns/*z*ero-*d*ay-e*x*p*loit*中检索到的2022年5月4日
11. Mathew, A. (2021). 机器学习在网络安全威胁中的应用。*SSRN*电子期刊。https://d*o*i.*o*rg/10.2139/ssrn.3769194
12. Musser, M., & Garriott, A. (2021).机器学习和网络安全：炒作与现实。安全与新兴技术中心（CSET）。 https://cset.georgetown.edu/wp-content/uploads/Machine-Learning-and-Cybersecurity.pdf
13. 美国国家科学院（NAS）。 （2020）。人工智能对网络安全的影响：研讨会记录。 美国国家科学院出版社。
14. Rupp, M.（2022年3月4日）。 机器学习和网络安全：简介。VMRay。 2022年5月4日从https://www.vmray.com/cyber-security-blog/machine-learning-and-cyber-security-an-introduction/检索
15. Salloum, S. A., Alshurideh, M., Elnagar, A., & Shaalan, K.（2020）。 机器学习和深度学习技术在网络安全中的应用：综述。智能系统和计算的进展，50-57。https://doi.org/10.1007/978-3-030-44289-7_5
16. Sava, J. A.（2022年2月14日）。2021年对企业的网络安全攻击影响。Statista. 2022年5月4日从https://www.statista.com/statistics/1255679/检索cyber- security- impact- on- businesses/

第2章
使用优化的卷积神经网络进行恶意软件检测的新方法

摘要 网络犯罪在近年来已成为一个价值数十亿美元的行业。 大多数网络犯罪/网络攻击都涉及部署某种类型的恶意软件。
恶意软件恶意地针对每个行业、每个部门、每个企业，甚至个人，已经展示出其能够使整个商业组织离线并造成数十亿美元的重大财务损失的能力。

恶意软件作者在攻击策略和复杂性方面不断发展，并开发出难以检测并能在后台休眠相当长时间以逃避安全控制的恶意软件。 鉴于上述论点，传统的恶意软件检测方法已不再有效。 因此，深度学习模型已成为检测和分类恶意软件的新兴趋势。 本文提出了一种新的卷积深度学习神经网络，以高精度准确有效地检测恶意软件。 本文与文献中的大多数其他论文不同之处在于，它采用了一种专家数据科学方法，通过从头开始开发卷积神经网络来建立性能模型的基线，然后探索和实施改进模型，最后评估最终模型的性能。 基线模型最初达到了98%的准确率，但在增加CNN模型的深度后，其准确率达到了99.183%，超过了文献中的大多数CNN模型。 最后，为了进一步巩固这个CNN模型的有效性，我们使用改进的模型对我们数据集中的新恶意软件样本进行预测。

关键词卷积神经网络·深度学习·恶意软件检测·图像特征·恶意软件可视化·恶意软件数据集·恶意软件分类

2.1 引言

信息技术的使用对现代生活是一种福音，因为它使我们在生活和工作方面达到了新的高度，但它也给我们的生活增加了相当多的漏洞和威胁。 一个无害的行为，比如浏览恶意网站或打开电子邮件附件，可能会给现代企业的运营带来混乱和破坏

不经常更新系统或无意中安装恶意软件可能会完全暴露计算机系统的漏洞和风险，使其易受网络攻击的威胁。 网络犯罪活动近年来飙升，黑客成功地通过恶意软件扰乱了某个行业的关键运营，将整个企业组织作为人质。 通过恶意软件的帮助，网络犯罪分子成功地劫持了一个部门或行业的关键运营，使整个企业组织陷入困境。

勒索软件是一种恶意软件，越来越多地被网络攻击者用来将目标计算机系统扣为人质，直到赎金要求得到满足。 最早的勒索软件事件之一发生在1989年，当时国际艾滋病大会的参会者收到了感染恶意软盘，失去了对文件的访问权限。 然后，他们被指示支付189美元到巴拿马的一个特定邮政信箱，以恢复对文件的访问权限[2]。 如今，勒索软件攻击的趋势显著增长，网络攻击者通过针对具有重要敏感信息或财务资源的特定组织进行高价值目标攻击。 这些组织对于像美国这样的国家的经济至关重要。如果这些目标实体的系统被扣为人质，将影响这些关键部门的日常运营，因此，这些实体更有可能支付赎金以恢复正常运营[2]。 2021年对美国最大的精炼产品管道Colonial Pipeline的勒索软件攻击就是这种高价值目标攻击的一个例子。 因此，网络犯罪在近年来已经成为一个数十亿美元的产业。 大多数网络犯罪/网络攻击都涉及部署某种类型的恶意软件。 随着杀毒技术变得更加强大并演变为反恶意软件软件，恶意软件创建者也提出了更复杂和强大的变种，难以检测，并可以在后台休眠而不引起安全控制的怀疑[3]。

在过去几年中，野外检测到的恶意软件样本数量一直在持续增长。 根据麦咖菲实验室的研究，2020年第四季度检测到了超过1,224,628个恶意软件威胁，其中包括7899个独特的新哈希值。 这就需要加强对恶意软件检测和预防的努力，特别是在网络黑客每天都在开发新的恶意软件变种[4]的情况下。

恶意软件分类是确定恶意软件名称、家族或类型以及恶意软件行为和签名的关键步骤，然后才能采取必要的行动，如删除和隔离。 对于恶意软件分类，主要有两种方法，包括基于签名的方法和基于行为的方法。 尽管传统上更多地使用基于签名的分类方法，因为它精确且快速，但它无法检测到通过混淆技术[5]（如打包、加密、元变形和多态性[6]）生成的恶意软件变种。 基于行为的分类可以解决基于签名的分类所面临的问题，因为所有恶意软件变种的行为几乎相似。 然而，检索与恶意软件行为相关的数据是耗时的，因为它应该在恶意软件激活期间收集。

因此，一种基于图像处理的恶意软件分类的新方法在最近变得流行起来[7]。 它使分类器能够通过检查恶意软件的图像纹理来检测和分类恶意软件。

与传统的基于签名或行为的恶意软件检测分类不同，这种方法实际上不依赖于研究恶意软件的签名和行为[7]，因此克服了传统恶意软件分类方法的一些弱点。

传统上，恶意软件的检测和分类涉及恶意URL或文件的行为和目的的监控过程。 最常用的三种恶意软件分析方法包括静态分析、动态分析或两者之间的混合分析。静态分析主要通过检查应用程序的清单和反汇编代码来获取特征。

[8]。 动态分析监视应用程序在执行过程中的行为，而混合分析则在应用程序安装之前和执行过程中监视应用程序。 在这三种分析方法中，混合分析是最强大的，因为单独的静态和动态分析无法检测到最复杂的恶意软件变种的威胁[8]。徐等人在他们的研究中发现，如果将混合分析与深度学习技术相结合，混合分析可以变得更加强大。 使用深度学习模型可以显著提高恶意软件检测和分类的性能，准确率达到95-99% [9]。

应用机器学习，特别是深度学习模型，来检测恶意软件已经存在了几年，恶意软件可视化已成为近年来网络安全研究人员热门的研究课题。 不同的传统机器学习方法，包括K最近邻算法、向量机、随机森林、决策树和朴素贝叶斯，已被用于检测和分类已知的恶意软件[2]。 通常以可执行文件（PE）格式（.EXE文件）出现的恶意软件样本是由一系列位构成的。

每个恶意软件二进制都由一串零和一组成。 这些零和一可以转换或表示为8位向量。 然后，这些8位向量被组织成二维矩阵，形成灰度图像，即一个二维矩阵（对应图像的高度和宽度）。 因此，每个灰度像素由一个取值范围为0到255的值表示[10]。

Nataraj等人首次提出了使用图像处理方法对恶意软件进行分类和检测，首先将恶意软件二进制文件转换为灰度图像[2]。 将恶意软件转换为灰度图像的前提是从图像处理的角度来观察恶意软件。 基于图像的恶意软件分类和检测旨在通过研究恶意软件图像的纹理来检测和分类恶意软件二进制文件的存在，这很容易从收集到的二进制恶意软件中转换出来。 这种恶意软件检测方法很容易适用于深度学习模型，例如广泛用于图像识别的卷积神经网络（CNN）。 这种方法的优点是，与其他恶意软件检测方法不同，它甚至可以检测到恶意软件代码的任何小改动，更好的是，它甚至可以检测到打包和混淆的恶意软件。 这是可能的，因为当恶意软件作者对其代码进行更改或打包其二进制文件甚至对其进行混淆时，纹理将出现在不同的位置上。

代表恶意软件的图像[11]。 当我们将恶意软件样本转换为图像时，我们可以应用深度学习算法来寻找恶意软件家族之间的视觉模式或相似之处，因为恶意软件作者经常重复使用代码来创建新的变种。 众所周知，恶意软件不再是编写而是组装[12]。

2.1.1 研究的必要性

如前所述，机器学习（ML），尤其是深度学习模型，已广泛应用于解决许多网络安全挑战，包括恶意软件检测。 在过去几年中，已经提出了许多深度学习算法来解决恶意软件分类和检测的问题。 这些深度学习模型依赖于从恶意软件的静态和动态分析中提取重要特征的过程，这被称为“特征工程”。实质上，特征工程允许研究人员从恶意软件的静态和动态分析中选择各种特征。 与特定类别的恶意软件对应的特征被用来训练深度学习模型，以在恶意软件和干净软件之间创建一个分离平面[13]。 尽管之前的研究工作[1,9,11-14]在使用各种深度学习算法进行更高效的恶意软件检测技术方面取得了重大进展，但主要问题是大多数这样的工作开发了一个算法，将其应用于少数数据集，达到了可接受的准确性水平，但并没有努力优化他们的深度学习模型。 这主要是因为大多数网络安全研究人员并非数据科学家，因此他们缺乏将最佳实践应用于优化他们的学习模型的专业知识。 我们坚信这在文献中造成了一个空白，即缺乏具有优化特性的深度学习模型。 我们认为，即使我们开发了一种新的学习算法，我们也应该将模型提升到一个新的水平，并对其进行优化，使其更高效并实现更高的准确性水平结果。 本研究的目的是通过提出一种新的卷积深度学习神经网络来准确有效地检测恶意软件，并具有高精度，以弥补文献中的这一空白。 与文献中大多数其他论文不同的是，本文采用了专业的数据科学方法，首先从头开始开发一个卷积神经网络来建立性能模型的基线，然后从基线模型中探索和实施改进模型，最后评估最终模型的性能。

2.1.2 研究的主要贡献

为了填补文献中的空白，本文提出了一种新的卷积深度学习网络架构用于恶意软件检测。 这个新的卷积神经网络基于专家数据科学方法，从头开始开发一个卷积神经网络，然后使用数据科学的最佳实践和方法来优化模型，最终实现卓越的性能。

检测准确率。 总体而言，我们的研究工作的主要贡献包括以下几点：

1. 我们提出了一种新的卷积深度学习神经网络，能够准确有效地检测恶意软件，并具有高精度。
2. 我们采用专家数据科学方法，从头开始开发了一个卷积神经网络，首先建立了一个性能模型的基准，并从基准模型中探索和实现改进模型。
3. 我们评估了最终模型的性能。 基准模型最初达到了98%的准确率，但在增加了卷积神经网络模型的深度后，其准确率达到了99.183%。
4. 我们的新颖深度神经网络（DNN）模型在文献中超过了大多数CNN模型。
5. 最后，为了进一步巩固这个CNN模型的有效性和准确性，我们使用改进后的模型对我们数据集中的新恶意软件样本进行预测。

本文的其余部分组织如下。 第2.2节回顾了传统上用于恶意软件分类模型的各种方法，考虑了各种方法。在第2.3节中，介绍了深度学习架构，以便深入了解这个研究背景。 第2.3节介绍了在这项研究中用于深度学习的实现架构以及用于评估分类器性能的统计指标。 第2.4节描述了方法和数据集。第2.5节讨论了使用我们提出的CNN模型进行恶意软件分类的实验研究和所得结果。 第2.6节介绍了我们的结果与以前工作的比较。

最后，第2.7节提供了本研究的结论和未来工作。

2.2 相关工作

为了突出我们工作的重要性，我们调查了其他研究人员在恶意软件检测和分类方面应用机器学习和深度学习技术的情况。Nataraj等人[1]被认为是在恶意软件检测方面使用机器学习的先驱。 他们使用了一种机器学习技术，k-NN，用于基于图像的恶意软件分类。 他们使用GIST描述符从输入图像中提取特征。 他们使用了非常流行但相对较小的Malimg数据集，其中包含来自25个不同恶意软件家族的9339个恶意软件二进制文件。 在他们提出的模型中，他们实现了令人印象深刻的98%的准确率。

他们的新方法的一个缺点是GIST描述符被认为是耗时且过于复杂的。

在另一项相关研究中，Yajamanam等人[15]采用了略有不同的方法，使用特征工程技术作为影响恶意软件分类器准确性的重要因素。 因此，他们从GIST的320个特征中选择了60个特征进行训练。 不幸的是，他们的模型只达到了

92%的准确率，因为减少特征可能对他们的技术的有效性和准确性产生负面影响。

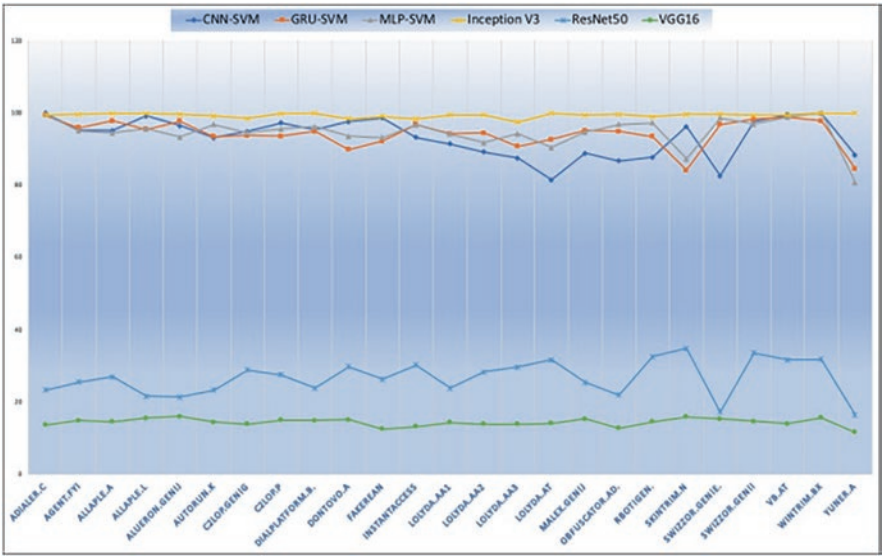
[16]中的作者使用了不同的描述符来提取预训练深度学习模型所需的特征。 他们使用了ImageNet作为数据集，其中包含了120万个恶意软件二进制文件的1000个类别。 他们认为数据集越大，模型的效果和准确性就越好。这是基于一个前提，即任何给定的深度学习模型只有训练数据好，它才能表现出色。 令他们惊讶的是，提出的深度学习模型只达到了92%的准确率，考虑到他们使用的庞大数据集和文献中其他相关模型的比较，这并不令人印象深刻。

Jiawei等人采用了一种轻量级方法来检测分布式拒绝服务恶意软件[17]。研究人员将恶意软件二进制文件转换为灰度图像，然后将其输入到经过微调的卷积神经网络中。机器学习分类器用于进一步对恶意软件二进制文件进行分类。 然而，由于签名匹配系统包含每个恶意软件样本的详细信息，其数据库庞大，因此对于资源有限的物联网设备来说效率不高。 一旦训练完成，只需要少量的训练数据就可以通过机器学习进行恶意软件分类，因此研究人员在恶意软件检测中使用了一个小型的两层浅层卷积神经网络。 研究人员在良性和恶意恶意软件的分类上实现了94%的平均准确率。

与以往文献中使用图像描述符的其他研究不同，Quan Le等人[18]没有使用描述符；相反，他们使用输入图像训练了他们的深度学习模型。 在他们的研究中，他们将原始输入图像转换为一维固定大小的向量，然后将其输入到他们的卷积神经网络模型中。
尽管他们的恶意软件分类器在98%的准确率方面取得了令人印象深刻的成绩，但必须注意，将图像转换为一维的固定大小向量可能会对图像质量产生负面影响，可能导致信息丢失。

在类似的思路中，论文[19]的作者使用原始的恶意软件图像来训练卷积神经网络进行恶意软件分类。 然而，他们的模型存在缺陷，因为他们操纵数据并以一种提供更高准确率的方式进行平衡。

Bensaud等人[2]使用了六个深度学习模型来检测和分类恶意软件图像，然后将这些模型的性能进行了比较。 研究人员在Malimg数据集上运行的六个模型包括Inception V3、VGG16-Net、ResNet50、CNN-SVM、MLP-SVM和GRU-SVM。由于这些图像的输入层需要3、224、224的形状来表示图像的红色、绿色和蓝色(RGB)通道，因此研究人员无法使用灰度图像与VGG16-Net和ResNet50这两个模型。而灰度图像的输入层需要1、224、224的形状。 下图展示了这六个模型的预测准确率：



正如我们所看到的，VGG16-Net和ResNet50在与其他模型相比表现较差，因为这两个模型的架构是为了识别RGB格式的彩色图像而设计的。 因此，当它们在灰度图像上进行测试时，这两个模型的准确性最低。 Inception V3模型实现了最高的准确率，达到了99.24%。

CNN-SVM、GRU-SVM和MLP-SVM的准确率也很高，分别达到了93.22%、94.17%和94.55%。 但是ResNet 50和VGG16的准确率分别为26.66%和14.31%。

Lad和Adamuthe [20]提出了一种结合了CNN和混合CNN+SVM模型的恶意软件图像检测和分类方法。 研究人员没有使用softmax作为激活函数。 他们使用SVM根据CNN模型提取的特征进行恶意软件分类任务。 所提出的模型通过全连接层生成了一个256个神经元的向量，作为SVM的输入。该模型的准确率达到了98.03%，优于其他CNN模型的准确率，如Xception（97.56%）、InceptionV3（97.22%）和VGG16（96.96%）[20]。

从上述研究中可以清楚地看出，在图像处理中并不需要使用图像描述符（例如GIST或SIFT [21]）来提取特征；相反，我们可以简单地利用恶意软件图像的灰度像素值，并将这些值用作训练机器学习模型的特征。 这里的想法是减少不必要的复杂性，以及构建高效深度学习恶意软件分类模型所需的时间。 但需要注意的是，为了使这个想法发挥最佳效果，我们需要确保恶意软件输入图像的尺寸大约为32*32，这样我们才不会给恶意软件分类器带来负担。

受到上述挑战的启发，本文研究了输入恶意软件图像的维度和学习的影响

技术，即卷积神经网络（CNN），应用于基于图像的恶意软件分类的性能。实验在Malimg数据集上进行。我们相信，即使我们开发了一种新的学习算法，我们也应该将模型提升到一个新的水平，并对其进行优化，使其更高效并获得更高的准确性水平结果。为了实现这一目标，我们旨在通过提出一种新的卷积深度学习神经网络来准确有效地检测恶意软件，并具有高精度。

2.3 系统架构

本节介绍了卷积神经网络（CNN）作为一种深度学习模型，用于分类和检测基于图像的恶意软件。如图2.1所示，我们将恶意可执行文件转换为灰度图像，然后将它们分组为恶意软件家族，如僵尸网络、银行木马、后门、蠕虫等。完成这一步骤后，我们将进入基于图像的恶意软件分类器，即我们的CNN模型。

恶意软件二进制文件的PE格式是具有扩展名的程序，包括.bin、.exe和.dll[2]。 PE文件是根据其组成部分进行识别的，这些组成部分被称为.tex、.rdata、.data和.rsrc。.tex是表示代码段并包含程序指令的第一个组件。

.rdata是包含只读数据的部分，而.data包含可修改的数据。 .rsrc是PE文件的最后一个组件，代表恶意软件使用的资源[2]。 将二进制文件转换为灰度图像的一种简单方法是将恶意软件二进制文件的字节序列（8位组）视为灰度图像的像素值（由8位编码）[主要论文参考]。 下图展示了以文本模式组成的灰度图像中恶意软件二进制文件的各个部分[2]。 基于文本模式，可以对恶意软件进行分类。

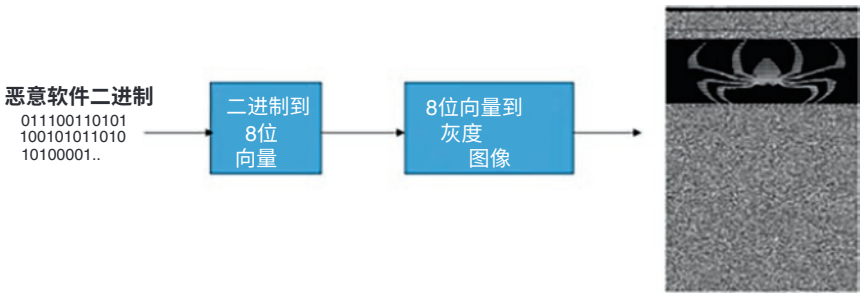
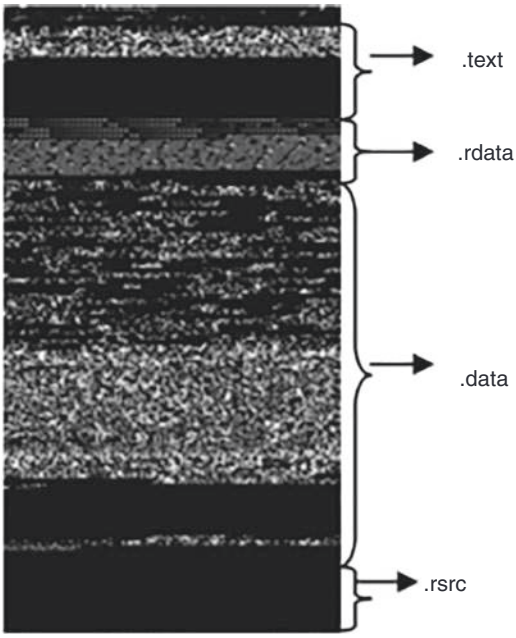


图**2.1** 将恶意软件转换为图像



通过图像可视化恶意软件二进制，因为这样隐藏在图像中的模式变得更加明显和可见。 如前所述，每个灰度像素由一个8位向量表示，其值的范围可以从 00000000(0) 到11111111 (255). 每个8位向量由一个数字表示，并可以转换为恶意软件图像中的像素，如下所示：

尽管基于恶意软件的图像的高度可能因恶意软件可执行文件的大小而异，但其宽度通常是固定的32、64和128个像素。
由于图像的宽度通常固定为32、64和128个像素[1,6,19]。

因此，不同的恶意软件二进制生成不同的恶意软件图像，这些图像具有不同的形状，如图2.2a-c所示，分别对应于Alueron.gen!J、Dialplatform.B和Swizzor.gen!E三个恶意软件家族。

使用基于图像的恶意软件分类模型的优势在于同一恶意软件家族的变体在纹理上非常相似，如图2.3所示。Dontovo A家族的三个变体与原始恶意软件非常相似。
这三个变体是从Malimg数据集[6]中的431个变体中随机选择的。 纹理上的这种相似性将使CNN模型能够根据图像纹理的相似性有效地对恶意软件进行分类到相应的恶意软件家族。

如前所述，机器学习模型可以使用灰度恶意软件图像进行训练。 准确地说，灰度像素值可以作为输入图像的特征，而不是从图像描述符中提取的特征（图2.4）。

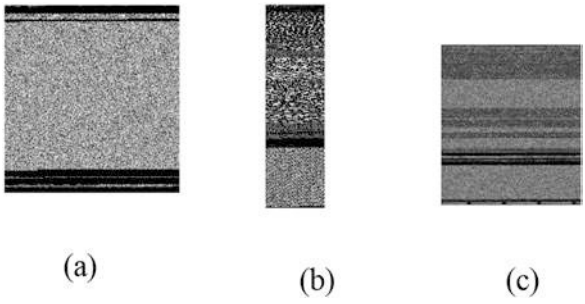


图**2.2** 恶意软件图像（**a**）Alueron.**g**en!J；（**b**）**Dia**lpl**a**tform.**B**；和（**c**）Sw**i**zzor.**g**en!E

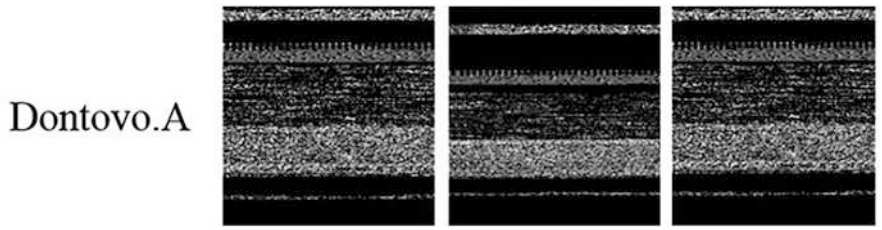


图**2.3** Dontovo家族的三个变种

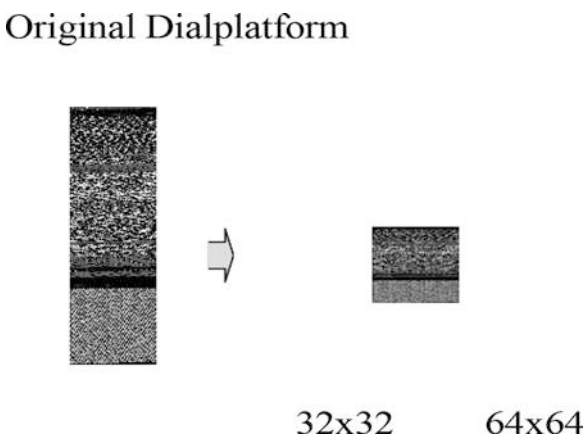
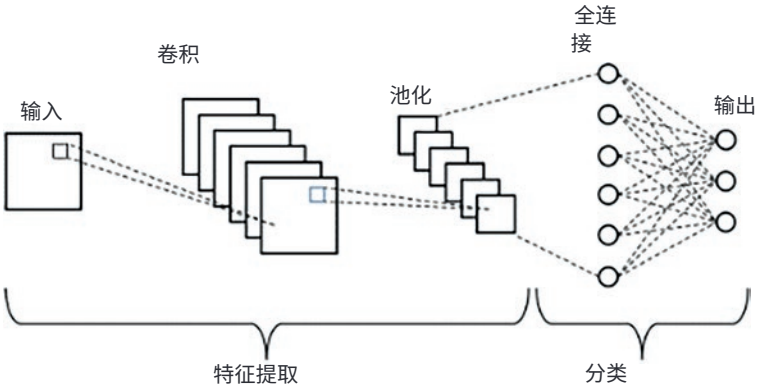


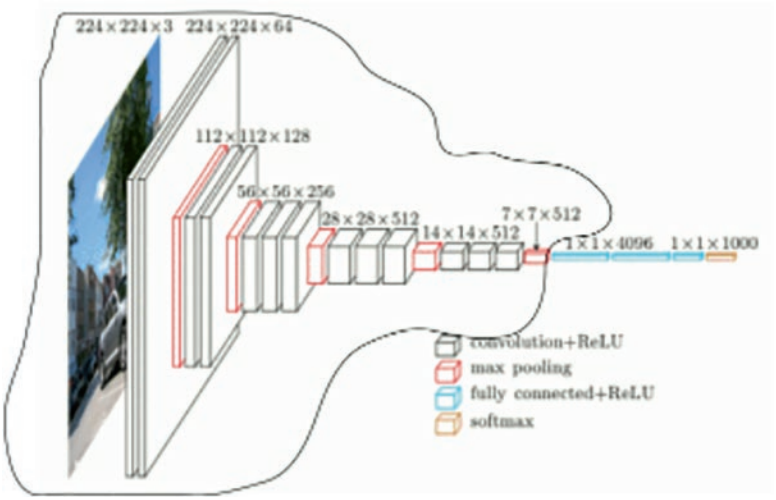
图**2.4** **Di**alplatform恶意软件在归一化后的图像

CNN是神经网络的一个子类，可以有效地从图像中分类和识别特定特征，因此在视觉图像分析中广泛使用CNN。 CNN的应用范围可能涉及图像或视频的识别、图像分类、自然语言处理[14]、医学图像分析和计算机视觉[9]。 CNN有两个主要功能，包括从图像中提取特征和对图像进行分类[22]。下图说明了CNN架构的两个功能：



CNN的两个功能[22]

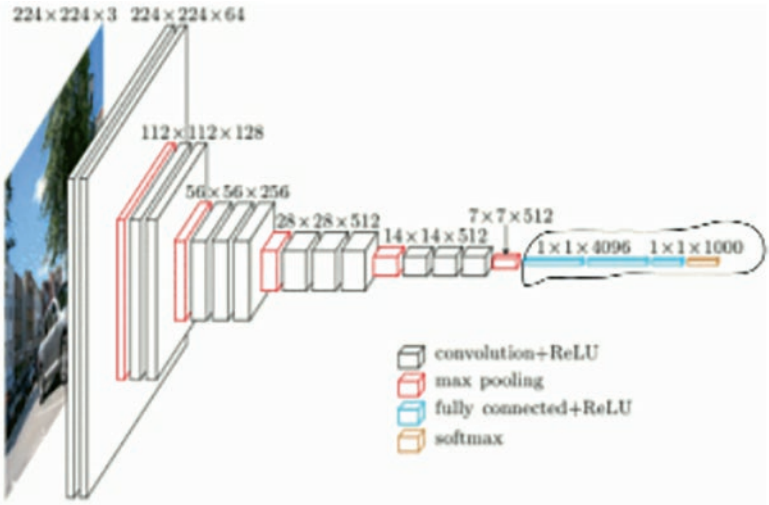
CNN的架构主要由两个模块组成[23]。 第一个模块通过卷积滤波操作将模板与图像匹配，作为特征提取器。 第一层负责使用多个卷积核对图像进行滤波，并生成特征图，然后对其进行调整或归一化。 这个滤波图像并生成特征图，然后进行归一化和调整的过程会重复多次[24]。 最后从最后一个特征图中得到的值最终用于连接成一个向量。 第一个模块的输出和第二个模块的输入仅由该向量定义。 下图[24]展示了第一个模块，用黑色圈出来：



在所有神经网络用于分类之后，第二个模块的功能开始发挥作用。 第二个模块的输入因子值通过多个激活函数和线性组合进行转换，生成一个新的向量作为输出。 最后的向量与类别数量相同。

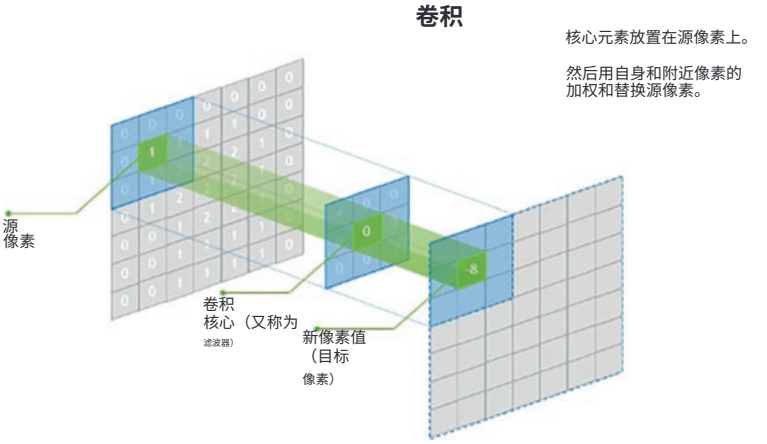
例如，元素i表示图像属于类别i的概率。

每个元素的值介于0和1之间，所有元素的总和为1 [24]。 这些概率的计算是通过第二个模块的最后一层完成的，该层使用二元分类的逻辑函数和多类别分类的softmax函数作为激活函数。 与普通神经网络一样，梯度反向传播确定层的参数。 例如，在训练阶段，最小化交叉熵，但这些参数在CNN的情况下特指图像特征[24]。 下图中用黑色圈出的是第二个模块：



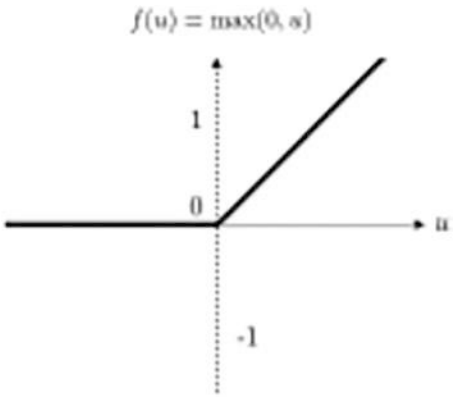
卷积神经网络（CNN）有四种类型的层，包括卷积层、池化层、全连接层[25]和ReLU修正层[24]。

卷积层是CNN的第一层。该层的目的是检测输入图像中的多个特征的存在。
[26]。 该层通过卷积滤波器来实现这一目的。 卷积层的原理是在图像上拖动代表特征的窗口，然后估计扫描图像的每个部分与特征本身之间的卷积乘积。 然后将特征视为滤波器。 卷积层接收多个图像作为输入，并计算每个图像与每个滤波器的卷积。 滤波器准确地表示我们希望在图像中看到的特征。 对于每对图像和滤波器，得到一个特征图，表示图像中特征的位置[24]。 下图演示了卷积过程，其中核的中心元素位于源像素上方，然后用自身和附近像素的加权和替换该像素：



池化层的定位在两个卷积层之间。 池化层接收多个特征图，然后对每个特征图执行池化操作。 池化操作通过减小图像的尺寸而不影响其重要特征来实现。 池化层通过将图像切分为规则大小的单元，并保持每个单元内的最大值来实现这一目标。 最常见的单元大小为2×2或3×3个像素单元[24]。 这些单元之间保持2个像素的间隔。 通过减少网络内的参数和计算量，池化层通过避免过度学习来提高网络的效率。

ReLU修正层代表修正线性单元，指的是通过应用ReLU(x) = max (0,x) [24]公式实现的真实非线性函数。 其可视化表示如下：



ReLU修正层将所有作为输入接收到的负值替换为零。 该层执行激活函数。

CNN的最后一层是全连接层，它通过应用线性组合和激活函数对接收到的输入向量生成一个新的输出向量。 该层将输入图像作为网络的输入进行图像分类，然后生成一个大小为N的向量，其中N表示图像类别的数量。 向量的每个元素表示输入图像属于某个类别的概率。 该层通过将每个输入元素与权重相乘，然后求和并应用激活函数来计算概率。 全连接层确定了电子邮件中特征的位置与其类别之间的关系。 该层的输入表是前一层的输出。 因此，输入表对应于特定特征的特征图。 输入表的高值指示了图像中特征的位置（表2.1）。因此，整体CNN架构如下图所示（图2.5）：

表**2.1** Malimg数据集家族[11]

	家族名称	变种
1	Allaple.L	1591
2	Allaple.A	2949
3	Yuner.A	800
4	Lolyda.AA 1	213
5	Lolyda.AA 2	184
6	Lolyda.AA 3	123
7	C2Lop.P	146
8	C2Lop.gen!G	200
9	Instantaccess	431
10	Swizzor.gen!I	132
11	Swizzor.gen!E	128
12	VB.AT	408
13	Fakerean	381
14	Alueron.gen!J	198
15	Malex.gen!J	136
16	Lolyda.AT	159
17	Adialer.C	125
18	Wintrim.BX	97
19	Dialplatform.B	177
20	Dontovo.A	162
21	Obfuscator.AD	142
22	Agent.FYI	116
23	Autorun.K	106
24	Rbot!gen	158
25	Skintrim.N	80
	总计	**9339**

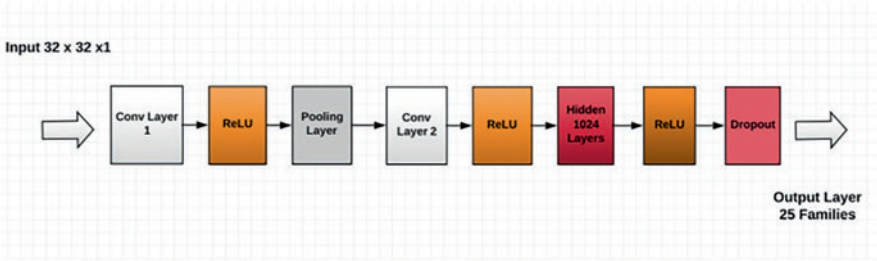


图**2.5** 整体CNN模型架构

2.4 方法和数据集

Malimg数据集在过去几年的许多研究项目和实验中被广泛使用，因为它非常适合于良好的深度学习卷积神经网络。 这项研究的独特之处在于研究人员决定从头开始开发一个新的CNN模型，而不是回顾已经表现良好的模型的文献。 这里的另一个独特之处在于，与文献中的大多数研究不同，研究人员不仅仅开发了一个模型并呈现结果；相反，他们超越了正常的期望，从头开始开发了一个CNN模型，开发了一个强大的基准模型性能评估，并探索了基准模型的扩展，以提高模型的学习能力，最后，研究人员将开发一个最终的CNN模型，评估最终的模型，并将其用于对新的恶意软件进行预测。

Malimg数据集非常适合我们的CNN模型，因为它具有足够的训练和测试数据集。为了了解我们的CNN模型的性能和学习曲线，在给定的训练运行中，我们可以进一步将训练数据集分为训练集和验证集。 我们使用传统的概念，将数据集的70%用于训练，剩余的30%用于测试[27]。

A. 开发基准模型

为了开发用于恶意软件分类和检测任务的基准CNN模型，我们将首先开发测试工具的基础设施，这将使我们能够在Malimg数据集上评估我们的模型。 这一步也很重要，因为它将建立模型的基准线，评估它，并最终改进模型。 开发我们的测试工具的基础设施包括以下步骤：加载Malimg数据集，准备数据集，定义CNN模型，评估模型，最后利用模型对Malimg数据集中的新恶意软件样本进行预测（保留测试）。

B. 加载数据集

我们对数据集有一些了解。 所有图像都具有相同的32×32像素的正方形尺寸，并且图像是灰度的。 因此，我们可以加载图像并重新调整数据数组以具有单色通道。 下面的代码片段演示了如何实现上述任务：

```
# 加载数据集
(trainX, trainY), (testX, testY) = malmig.load_data()
# 重新调整数据集以具有单个通道
trainX = trainX.reshape((trainX.shape[0],  32, 32, 1))
testX = testX.reshape((testX.shape[0], 32, 32, 1))
```

C. 定义我们提出的CNN模型

现在，我们需要为恶意软件检测和分类定义一个基准卷积神经网络模型。 该模型有两个主要方面：特征提取前端由卷积和池化层组成，分类器后端将进行预测以确定二进制样本是否为恶意软件。我们将使用随机梯度下降优化器的保守设置，学习率为0.01，动量为0.9。我们将优化分类交叉熵损失函数，适用于多类别分类，并监控分类准确度指标[28]。

我们的模型可以使用以下代码来定义：

```
# 定义卷积神经网络模型
deffl 定义模型():
 模型 = Sequential()
   模型.add(Conv2D(32,     (3,   3),   activation='relu',   kernel_
initializer='he_uniform', input_shape=(28,  28, 1)))
 模型.add(MaxPooling2D((2, 2)))
 模型.add(Flatten())
模型.
add(Dense(100,
activation='relu', kernel_initializer='he_uniform'))
 模型.add(Dense(10, activation='softmax'))
 # 编译模型
 opt = SGD(lr=0.01, momentum=0.9)
  model.compile(optimizer=opt,  loss='categorical_crossentropy',
metrics=['accuracy'])
 return model
```

D. 评估我们的CNN模型

一旦我们完成了定义恶意软件检测的CNN模型的任务，我们现在需要对其进行评估。我们将使用五折交叉验证来评估模型。 *k*的值= 5将确保我们获得基准线

重复评估，并且还将确保相对较短的运行时间。 因此，k= 5意味着我们的训练数据集将被分成五个测试集[28]。

此外，为了确保我们的模型在上述五个折叠中包含相同的训练和测试数据集，我们将在开始分割过程之前对训练数据集进行洗牌[28]。 这个过程将确保我们进行“苹果对苹果”的比较（公平的模型比较），可以说是如此。 训练我们的CNN基线模型的批次大小将为32个恶意软件样本，训练时使用10个时期。 这个设置将使我们能够估计我们的CNN基线模型的性能，并跟踪每次运行的结果历史和每个折叠的恶意软件分类准确性。

以下代码演示了如何实现上述任务：

```
# 使用k折交叉验证评估模型
deffl evaluate_model(dataX, dataY, n_ffolds=5):
 scores, histories = list(), list()
 # 准备交叉验证
 kffold = KFold(n_ffolds, shuffffle=True, random_state=1)
 # 枚举拆分
 fflor train_ix, test_ix in kffold.split(dataX):
 # 定义模型
 model = deffine_model()
 # 选择用于训练和测试的行
trainX, trainY, testX, testY = dataX[train_ix], dataY[train_ix],
dataX[test_ix], dataY[test_ix]
 # 拟合模型
history = model.ffit(trainX, trainY, epochs=10, batch_size=32, vali -
dation_data=(testX, testY), verbose=0)
 # 评估模型
 , acc = model.evaluate(testX, testY, verbose=0)
 print('> %.3f' % (acc * 100.0))
 # 存储分数
 scores.append(acc)
 histories.append(history)
 return scores, histories
```

2.5 实证结果

在评估我们的模型之后，下一个逻辑步骤是显示和呈现结果。
我们将重点关注CNN模型的学习行为，然后估计其
性能。 正如深度学习社区中广为人知的，过拟合
和欠拟合是几乎所有深度学习模型都面临的两个主要问题

模型。 为此，为了确保我们的CNN模型既不过拟合也不欠拟合，我们创建了一条线图来显示和报告我们的模型在测试数据集和训练数据集上在每个五折交叉验证期间的性能。 线图将提供关于我们的模型是否对恶意软件数据集[19]过度拟合或欠拟合的见解。

只需要一个包含两个子图的图形，就可以了解过拟合和欠拟合问题，一个子图用于损失，一个子图用于我们的CNN模型的准确性[29]。 我们的模型在训练数据集上的性能使用蓝色线显示，而橙色线将显示我们的CNN模型在保留测试数据集上的性能。 下面的代码片段将实现这个任务：

```
# 总结模型性能
deffl summarize_perfflormance(scores):
 # 打印总结
 print('准确率：平均值=%.3ffl 标准差=%.3ffl, n=%d' % (mean(scores)*100,
std(scores)*100, len(scores)))
 # 结果的箱线图
 pyplot.boxplot(scores)
 pyplot.show()
```

我们可以看到模型达到完美技能的情况。这些是很好的结果。

```
>98.467
>98.683
>98.642
>98.850
>98.583
```

上面的数字令人印象深刻，它们清楚地显示了我们的模型评估的进展。

接下来，显示了一个诊断图，可以了解模型在每个折叠中的学习行为（图2.6）。

在这种情况下，我们可以看到模型通常能够很好地拟合，训练和测试学习曲线趋于收敛。 没有明显的过拟合或欠拟合的迹象。

接下来，计算模型性能的摘要。 我们可以看到，在这种情况下，模型的估计技能约为98.6%，这是合理的。

最后，创建一个箱线图来总结准确性分数的分布。

2.5.1 改进基准模型

有许多方法可以改进基线模型。

我们可以探索CNN模型配置的各个方面，这些方面很可能会带来改进，即所谓的低悬果实。 第一种是改变学习算法，第二种是增加模型的深度[28]（图2.7）。

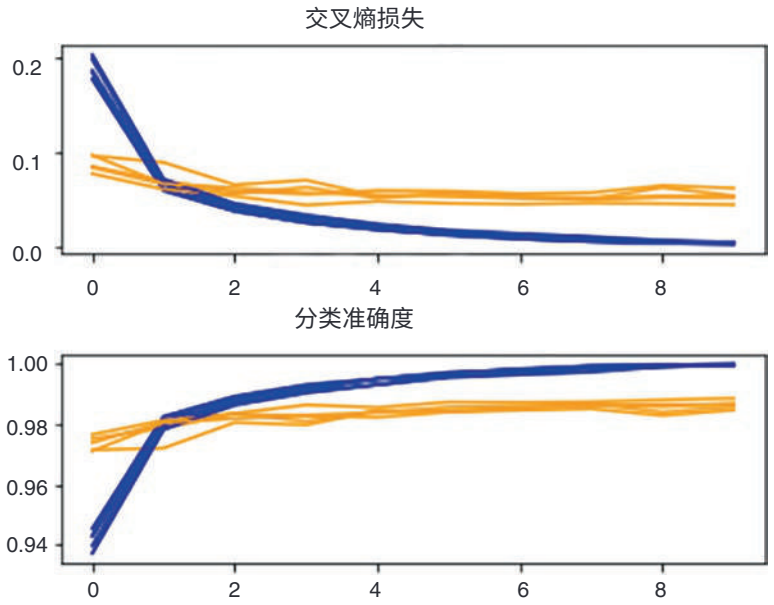


图**2.6** 基线模型在k折交叉验证期间的损失和准确率学习曲线

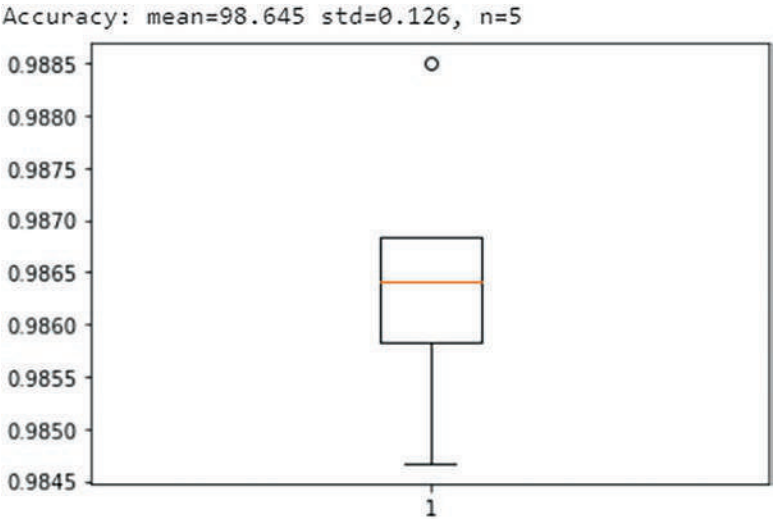


图**2.7**基于k折交叉验证评估的基准模型准确性得分的箱线图

现在我们可以探索一些改进我们的CNN模型的方法。
我们可以改变模型配置来探索对基准模型的改进。 两种常见的方法包括改变模型的特征提取部分的容量或改变分类器部分的容量或函数[28]。

我们可以增加模型的特征提取器部分的深度，按照VGG的模式添加更多的卷积和池化层，使用相同大小的滤波器，同时增加滤波器的数量。 在这种情况下，我们将添加一个具有64个滤波器的双卷积层，然后是另一个最大池化层[28]。

运行示例报告交叉验证过程的每个折叠的模型性能如下所示：

>98.775
>98.683
>98.967
>99.183
>99.008

上面每个折叠的得分可能表明相对于基准模型有一些改进。

2.5.2 完善我们的模型并进行预测

尽管继续改进我们的模型似乎很有趣，但在这一点上，我们将选择我们的CNN模型的最终配置。 我们的模型的最终版本将是更深的模型。 我们通过将其拟合到整个训练Malimg训练数据集上，然后加载模型并进行评估来完成我们的模型。

表**2.2** 提出的DL-CNN与其他现有深度学习方法的比较

年份	研究人员	方法	技术	准确率 (%)
2011	Nataraj等人	GIST	机器学习 98	
2017	S. Yue	卷积神经网络	深度学习	97.32
2017	Makandar和Patrot	Gabor小波-kNN	机器学习 89.11	
2018	Yajamanam等人	GIST+kNN+SVM	机器学习 97	
2018	Cui, Xue等人	GIST+SVM [29]	深度学习	92.20
2018	Cui, Xue等人	GIST+kNN	深度学习	91.90
2018	Cui, Xue等人	GLCM+SVM	深度学习	93.20
2018	Cui, Xue等人	GLCM+kNN	深度学习	92.50
2018	Cui, Xue等人	IDA+DRBA	深度学习	94.50
2019	Cui, Du等人	卷积神经网络，NSGA-II	深度学习	97.6
2020	Mallet	卷积神经网络，Keras	深度学习	95.15
2020	Vasan等人	IMCFN，彩色图像	深度学习	98.82
2021	Moussas和Andreatos	图像和文件特征，ANN两级	ANN 99.13	
2021	Omar	卷积神经网络，Keras	深度学习	99.18

它。 我们将在留存测试Malimg数据集上评估我们的模型性能和准确性。 这将使我们对我们的模型在真实恶意软件数据集上的实际准确性有更深入的了解（表2.2）。

对测试数据集上的模型分类准确率进行计算并打印。 在这种情况下，我们可以看到该模型的准确率为99.180%，或者略低于1%，这个结果还是不错的，标准差约为半个百分点（例如，99%的分数）。

>99.180

2.6 与之前工作的结果比较

为了确保公平比较（苹果与苹果比较），我们提供了使用相同数据集对恶意软件分类和检测的深度学习CNN模型与其他深度学习算法性能的比较。 我们提出的卷积深度学习模型的最终版本在准确率上达到了99.18，而Mallet [30]只有95.15。这清楚地表明我们的模型在文献中表现优于其他模型。 有趣的是，Mallet [30]几乎使用了完全相同的深度学习架构，包括CNN和Keras科学环境。 此外，我们的模型超过了Cui等人提出的深度学习模型[19]，其准确率为97.6。有趣的是，我们的模型甚至超过了Moussas和Andreatos [31]在完全相同数据集上的最新工作，其准确率达到了99.13。

2.7 结论

在本文中，我们提出、开发和展示了一个用于恶意软件检测的深度学习卷积神经网络模型。 我们的模型采用了一种独特的方法来进行恶意软件检测，即从头开始开发了一个深度学习卷积神经网络，开发了一个基准模型，开发了一个对基准模型进行鲁棒性性能评估的方法，并探索了对基准模型的扩展，以提高模型的学习能力，最后，我们开发了一个最终的卷积神经网络模型，评估了该模型，并将其用于对新的恶意软件进行预测。

我们使用了流行的Malimg数据集，其中包含9339个样本，属于25个恶意软件家族。 与现有的基于深度学习的恶意软件检测模型相比，我们的模型表现出色，明显优于这些模型。

作为未来的工作，我们相信可以使用更大的数据集，如BIG 2015数据集和新的数据集，进一步测试和评估我们的模型的性能。

从Kaggle等存储库中提供的数据集。另一个未来的方向是将我们模型的准确性得分与其他类似的深度学习模型（如k-NN模型等）的准确性得分进行比较。

参考文献

1. Ye, Y., Li, T., Adjeroh, D., & Iyengar, S. S. (2017). 使用数据挖掘技术进行恶意软件检测的调查。A*CM*计算调查（*CSUR*），*5*0（3），文章编号41。
2. Bensaoud, A., 等 (2020).使用卷积神经网络模型对恶意软件图像进行分类. 科罗拉多大学科罗拉多斯普林斯分校计算机科学系[在线]. 可用*h*ttps://ar*x*i*v*.or*g*/pd*f*/2010.16108.pd*f*

3. Kaspersky. (2019).Kaspersky安全公报2019[在线]. 可用h*tt*ps://se*cu*re*list*. com/kaspersky- security- bulletin- threat- predictions- for- 2019/88878/
4. Cybersecurity Ventures. (2018). [在线]. 可用 https://cybersecurityventures. com/- cybercrime- damages- 6- trillion- by- 2021/
5. Souri, A., & Hosseini, R. (201*8*). 基于数据挖掘技术的恶意软件检测方法的最新调查。人本计算与信息科学，*8*(3)，1–22。

6. Nataraj, L., Karthikeyan, S., Jacob, G., & Manjunath, B. (2011). 恶意软件图像：可视化和自动分类。 在第*8*届国际网络安全可视化研讨会上，匹兹堡，宾夕法尼亚州，美国。

7. Lee, C., et al. (2020). 使用机器学习评估基于图像的恶意软件分类。 在计算集体智能进展中，第*12*届国际会议，ICCCI 2020，越南岘港，2020年*11*月*30*日至*12*月*3*日，论文集。htt*p*s://doi.org/10.1007/978- 3- 030- 63119- 2_11

8. 徐，L.等（2016年）。 HADM：用于检测恶意软件的混合分析。*S*A*I*智能系统会议*2016*年，*21*（22），1037-1048。
9. Farabet，C.，Couprie，C.，Najman，L.和LeCun，Y.（2013年）。 学习层次特征用于场景标记。*IEEE*模式分析和机器智能交易，*35*（8），1915-1929。

10. Douze，M.等（2009年）。 评估GIST描述符用于网络规模图像搜索。在*ACM*国际图像和视频检索会议论文集，第*19*篇，希腊。

11. Ahmadi，M.，Ulyanov，D.，Semenov，S.，Trofimov，M.和Giacinto，G.（2016年）。 新颖特征提取，选择和融合以实现有效的恶意软件家族分类。 在第6届A*C*M数据和应用安全与隐私会议上的论文集，美国路易斯安那州。12. 韩，K. S.，林，J.
H.，康，B.和任，E. G.（2015年）。 使用可视化图像和熵图进行恶意软件分析。国际信息安全杂志，1*4*（1），1-1*4*。
13. Vinayakumar, R., Alazab, M., Soman, K., Poornachandran, P., & Venkatraman, S. (2019). 使用深度学习的强大智能恶意软件检测。*IEEE Access, 7*, 46717–46738.
14. Kalchbrenner, N., Grefenstette, E., & Blunsom, P. (2014). 用于建模句子的卷积神经网络。在《计算语言学协会第*52*届年会论文集》中，美国马里兰州（第*655*-665页）。

15. Bhodia, N., Prajapati, P., Troia, F. D., & Stamp, M. (2015). 基于图像的恶意软件分类的迁移学习。 在《第5届国际信息系统安全与隐私会议论文集》中（第719-726页）。

16. Alex, T. 使用机器学习进行恶意软件检测 [在线]. 可用 *htt*ps://*githu*b.com/*tu*ff96/*Ma*l*war*e-*detection-using-Machin*e-Le*arning*

17. Su, J., Danilo Vasconcellos, V., Prasad, S., Daniele, S., Feng, Y., & Sakurai, K. (2018). 基于图像识别的物联网恶意软件轻量级分类. 在*2018*年I*EEE*第*42*届年度计算机软件和应用会议*(COMPSAC)*中，东京 *(pp. 664–669)*.
18. Le, Q., Boydell, O., Mac Namee, B., & Scanlon, M. (2018). 深度学习在非领域专家的恶意软件分类中的应用.数字调查, *26*(1), 5118–5126.
19. Cui, Z., et al. (2018). 基于深度学习的恶意代码变体检测.*IEEE*工业信息学报, 1*4*(7), 3187–3196.
20. Lad, S. S., & Adamuthe, A. C. (2020). 恶意软件分类与改进的卷积神经网络模型。国际计算机网络与信息安全杂志，*1*2(6)，30-43。https://doi.org/*1*0.58*1*5/ijcnis.2020.06.03

21. Tareen, S. A. K., & Saleem, Z. (2018). SIFT、SURF、KAZE、AKAZE、ORB和BRISK的比较分析。在国际计算、数学和工程技术会议（i*C*o*M*ET 2018）中，巴基斯坦苏库尔。
22. Gurucharan, M. K. (2020).基本的*CNN*架构：解释卷积神经网络的*5*层。 可用h*ttps://www*.u*pg*rad.c*o*m/b*l*og/ba*si*c-cnn-arch*itec*ture*/#:~:text=*Th*ere%20ar*e*%20*t*hr*ee*%20*types*%20*of*,*C*NN%20arch*itec*ture%20will%20be%20formed

23. Agarwal, B., et al. (2020).深度学习技术在生物医学和健康信息学中的应用. 学术的.
24. Smeda, K. (2019).理解*CNN*的架构. T*o*war*d*s Data S*cienc*e [在线]. 可用https://towardsdatascience.com/understand- the- architecture- of- cnn- 90a25e244c72
5. Albelwi, S., & Mahmood, A. (2017). 一个设计深度卷积神经网络架构的框架熵，19(6)，文章242.
26. El-Baz, A. S., & Suri, J. (2021).神经工程技术用于自闭症谱系障碍. 学术的.
27. 美国参议院。 (2022).美国公司勒索软件攻击案例研究。员工报告[在线]。 可用*htt*p*s://w*ww.*hs*gac.*s*enate.*go*v/*imo*/media/doc/Americas%20Data%20Held%20Hostage.pdf

28. 机器学习的精要。 (2021).如何从头开始开发卷积神经网络。 从htt*p*s://*m*achine*l*earning*m*astery.*com*/h*ow*-t*o*-*de*v*elop*-a-generati*v*e-a*d*versaria*l*-net*w*ork-*f*or-a- 1- *dim*ensi*o*na*l*- *functi*on - *f*rom- scratch- in- keras/检索
29. Rieck, K., Trinius, P., Willems, C., & Holz, T. (2011). 使用机器学习自动分析恶意软件行为。计算机安全杂志 (*JCS*), 19(4), 639–668.
30. Madhuanand, L., Nex, F., Yang, M., Paparoditis, N., Mallet, C., Lafarge, F., et al. (2020). 深度学习用于无人机图像的单目深度估计。*ISPRS* Annals of the *Photogrammetry, Remote Sensing and Spatial Information Sciences, 2*, 451–458.
31. Moussas, V., & Andreatos, A. (2021). 基于代码可视化和两级分类的恶意软件检测。信息，12(3)，118。

第3章
使用本地离群因子技术进行恶意软件异常检测

摘要 恶意软件异常检测是一个重要的研究领域，因为新的恶意软件变种继续对商业组织造成严重破坏。 在这项研究中，我们提出了一种基于本地离群因子算法的新技术，用于检测异常的恶意软件行为。 我们在真实数据集上经验性地验证了我们的技术的性能和有效性。 这是一种高效的恶意软件检测技术，因为为此目的训练的模型基于无监督学习。

该模型在异常情况下进行训练，即进程中的异常行为，使其显著有效。

关键词 本地离群因子 · 离群点检测 · 异常检测 · 恶意软件检测 · 恶意软件数据集 · 数据集验证

3.1 引言

世界正在以迅猛的速度发展。 最先进的技术被应用于涉及技术使用的每个领域。 科学和技术领域的广泛扩展和进步归功于通信和信息系统的使用。 随着信息技术应用在过去几十年中的显著扩展，依赖通信和信息技术应用的用户数量也显著增加，显然在安全和隐私方面引起了各种关注。

显然，IT领域在过去几十年中呈指数增长，潜在的安全威胁也增加了。 网络安全正在采用新的创新方法来应对网络犯罪分子采取的侵入性方法。 这些进展的应用和实施可以看作是数字取证和监视、恶意软件和僵尸网络、入侵检测和预防系统等领域的成功。[1] 全球范围内的网络犯罪分子不断推导出新的方法和方式来突破高度复杂系统的安全和完整性，以此作为挑战。

M. Omar，机器学习在网络安全中，斯普林格计算机科学简报，https://doi.org/10.1007/978-3-031-15893-3_3

为了解决这些问题，使用最新技术如机器学习、人工智能、联邦学习、密码学等，开发了各种入侵检测和预防系统。人工智能和机器学习在决策支持、人类行为分析、推荐、过程控制等方面发挥了重要作用。机器学习的众多应用之一是恶意软件检测。

针对威胁向量和潜在威胁的识别算法，利用机器学习进行恶意软件检测有许多不同的方法[2]。 机器学习技术能够适应变化的环境，解决复杂的问题和难题。 机器学习和人工智能在网络安全领域的应用解决了钓鱼检测、网络和应用程序入侵检测和预防系统、密码学、垃圾邮件检测、拒绝服务和分布式拒绝服务预防等问题[3]。

3.1.1 恶意软件检测

恶意软件是对计算机系统造成损害的最持久和最重要的安全威胁之一，这是由于缺乏网络安全实践所导致的。 恶意软件检测是在主机系统上检测恶意文件或程序的过程，这些文件或程序可能对系统的组件造成潜在的损害[4]。 恶意软件的发展已经取得了长足的进步，发展出了高级持续性威胁（APT）攻击、动态链接库（DLL）攻击等。这些攻击非常难以检测，甚至更难以缓解[5]。

恶意软件根据攻击向量、传递方式、生命周期、威胁向量等因素被分为不同的类别。一些常见的恶意软件类型包括木马、Rootkit、僵尸网络、蠕虫、拒绝服务等。由于每天都有成千上万种不同变种的恶意软件被开发和发布，安全专业人员开发了各种工具和软件，如CWSandbox、ANUBIS、Norman Sandbox等，来检测和缓解这些威胁[4]。

3.1.2 入侵检测系统

正如先前提到的，网络犯罪分子设计了成千上万种新方法来利用软件和硬件层面的漏洞。 为了确定威胁和攻击向量，安全专业人员开发了不同的工具来检测和减轻此类攻击。 入侵检测系统是一种软件系统，用于检测信息系统或网络上的恶意活动的发生。

级别。 这些入侵检测系统专门开发用于维护数据和信息存储在计算机系统上的完整性、机密性和可用性[6]。 入侵检测系统通过采用不同的技术限制未经授权的访问和攻击者的提升权限。 由于其重要性，入侵检测系统在每个组织中都具有重要意义。 以下原因证明了在当前技术时代使用和实施入侵检测系统的必要性：

- 通过增加风险和威胁发现来预防问题行为。
- 检测到未被标准安全控制所减轻的漏洞和攻击。
- 检测和缓解“门把手摇动”和网络攻击。
- 现有威胁和漏洞的文档和报告。
- 质量和安全控制用于管理和安全设计
  实施。
- 对现有和潜在未来威胁的即兴诊断和分析
  以及恢复和漏洞分析[6]。

入侵检测系统根据用户的使用和要求而有不同类型，如基于间隔、实时、基于网络、基于主机、基于应用程序等。每个入侵检测系统根据保护标准和漏洞实施不同的策略，涵盖攻击向量的不同方面。 本研究重点介绍了网络入侵检测系统的使用和实施。

3.1.3 基于网络的入侵检测系统

网络入侵检测系统是商业公司中最常用的入侵检测系统之一。 网络入侵检测系统监视网络流量和检查数据包，以在它们通过网络到达系统之前检测异常行为和分析不同的攻击。 一个网络入侵检测系统可以监听多个主机的网络段，从而为整个网络提供实时威胁分析（图3.1）。

在更常见的方法中，NIDS由两个网络接口组成；一个接口以混杂模式用于监听网络信息传输，而另一个接口用于报告和控制。 将流量作为输入传递给NIDS有许多不同的方法；随着交换技术的发展，端口镜像技术为NIDS提供了流量。例如，思科采用了交换式端口分析器的方法。 根据功能和使用要求的不同，NIDS可以是基于异常的或基于签名的[8]。

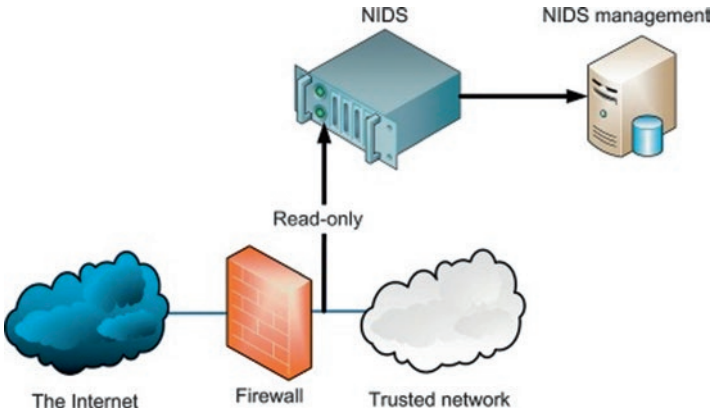


图**3.1** 网络IDS架构[7]

3.1.4 网络入侵检测系统的优势

• 大型跨多个节点的网络可以通过少量的IDS轻松监控。

• 将基于网络的IDS纳入已建立的网络中只需要很少的努力，并且对网络的影响非常小。 NIDS是被动系统，实施时只监听网络流量，不会对网络的正常功能和操作造成任何干扰。

• NIDS的实施和部署可以在不同范围的攻击下显著增强安全性和保护性。 甚至可以使NIDS的实施对攻击者不可见。

正如先前提到的，通过实施和应用先进的机器学习技术，可以检测和缓解具有各种变量的不同攻击。 本研究的重点是了解和实施一种称为“局部离群因子”的机器学习方法，以缓解基于异常的攻击。

3.2 相关工作

最近，各种研究人员开发了异常检测方法。 Kim等人采用局部离群因子来确定端点环境中的陌生入侵。Kim等人采用的一种方法。

[9] 使用两个模型来检测异常行为，这些行为与正常情况有所偏离。异常检测是通过

事件日志的分析。这将生成异常分数。根据事件发生的频率，确定攻击类型。通过使用局部离群因子和自编码器，该模型被设计为高效地检测异常。根据攻击者提出的不同攻击场景，利用攻击配置文件的分析用于检测异常。Kim等人的研究[9]表明，已经进行了各种研究，以通过使用带标签的数据（如拒绝服务（DOS）攻击，用户对根（U2R）攻击，远程探测攻击和远程到本地（R2L）攻击）来检测基于监督学习的攻击行为。与上述研究相比，本研究实施了一种模型，用于确定行为与正常情况下的行为偏差。所提出的模型使用LOF和自编码器进行异常分数的计算，该分数表示从现有日志和新生成日志中收集的数据点（图3.2）之间的数据点。

Kotu和Deshpande [10]提出了各种在数据集中检测异常的方法。 研究人员详细解释了检测数据集中异常的不同方法。 在提出的不同方法中，突出了使用统计方法进行异常检测和使用数据挖掘进行异常检测的两种方法。 通过采用统计分布模型，识别与模型不符的数据点。 根据数据点在标准差曲线上的位置来检测异常值。 通过应用数据挖掘方法进行异常检测，各种参数和变量突出显示数据中的异常值。 这些变量包括距离、密度、分布、聚类和分类技术。 Kotu和Deshpande [10]在所进行的研究中详细解释了每个变量及其工作原理。

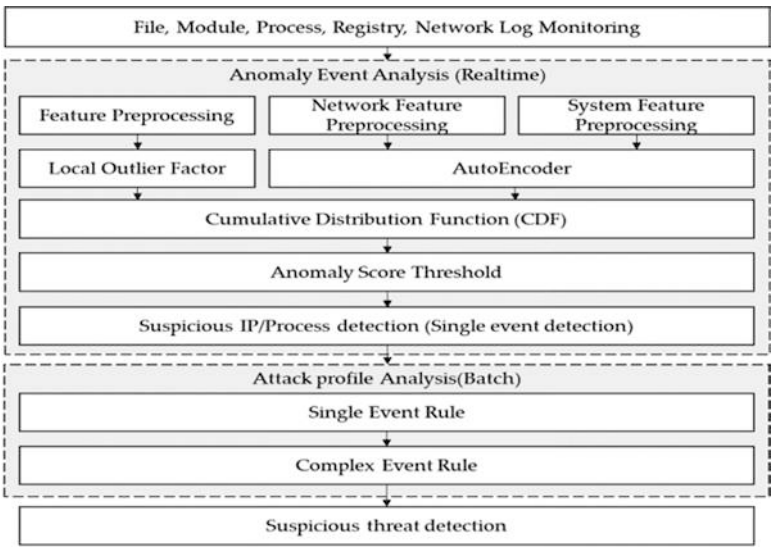


图**3.2** Kim等人提出的模型[9]

Tang等人的研究和研究强调在进行利用或攻击时检测恶意软件。这种策略作为提高检测传递负载准确性的早期威胁检测器。 基于异常的恶意软件检测方法的基本直觉是通过奇怪的非本地脚本或代码块分析程序正常执行的变化。

如果这些扰动被识别和分析，它们可以构成恶意软件检测的基础。 唐等人进行的研究[11]使用Internet Explorer 8和Adobe PDF Reader 9来确定攻击向量，因为这些程序更常被攻击者利用。

3.3 提出的方法论

在网络安全领域引入先进的机器学习技术显著降低了与技术使用相关的风险。 如前所述，各种机器学习技术被应用于对抗网络犯罪者的攻击。 有监督学习和无监督学习的使用为关键数据和信息的保护开辟了新的维度。

为了信息安全，采用了各种机器学习技术，如决策树、相似哈希（局部敏感哈希）、流入流聚类、行为建模等[12]。 用于检测恶意软件的机器学习方法之一是局部离群因子技术，这是本研究中采用和实施的基本方法。

3.3.1 本地离群因子

3.3.1.1 对局部离群因子方法的简要理解

局部离群因子是一种利用无监督学习原理进行异常检测的机器学习技术。 数据集中的离群点被表示为异常分数。 这是通过计算数据点相对于最近数据点的局部密度的偏差来实现的。

3.3.1.2 局部离群因子的基本工作原理

通过估计彼此之间接近的不同数据点之间的距离来确定局部密度。 可以通过将具有低密度的数据点与附近的具有相似密度的数据点进行比较来测量每个数据点的密度。 低密度的数据点被视为离群值[13]。

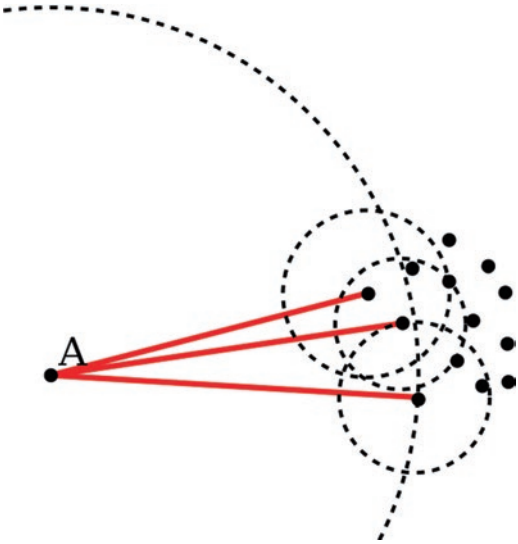


图**3.3** LOF中的相对密度[14]

$$\text{reachability-distance}_k(A,B)=\max\{k\text{-distance}(B),\ d(A,B)\}$$

图**3.4** 两点之间的可达距离和点的k距离[13]

$$\text{lrd}_k(A):=1/\left(\frac{\sum_{B\in N_k(A)}\text{reachability-distance}_k(A,B)}{|N_k(A)|}\right)$$

图**3.5** 本地可达性距离公式 [13]

下图显示了本地异常因子的基本分类（图3.3）。

最初，两点之间的距离是k-距离。 这些数据彼此之间相对接近；因此，确定它们的k个最近邻点之间的距离。 第二个最近的点是初始点的最近邻。这用于测量可达性距离，即两个不同点之间的最大距离和该特定点的k-距离（图3.4）。

3.3.1.3 本地可达性距离

测量与特定点靠近的k个最近点的本地可达性距离，通过计算所有k个最近点的可达性距离的倒数之和来实现。 如果点彼此之间更接近，则距离较低，密度增加（图3.5）。

图**3.6** 局部异常因子公式[13]

$$\mathrm{LOF}_k(A) := \frac{\sum_{B \in N_k(A)} \frac{\mathrm{lrd}_k(B)}{\mathrm{lrd}_k(A)}}{|N_k(A)|} = \frac{\sum_{B \in N_k(A)} \mathrm{lrd}_k(B)}{|N_k(A)| \cdot \mathrm{lrd}_k(A)}$$

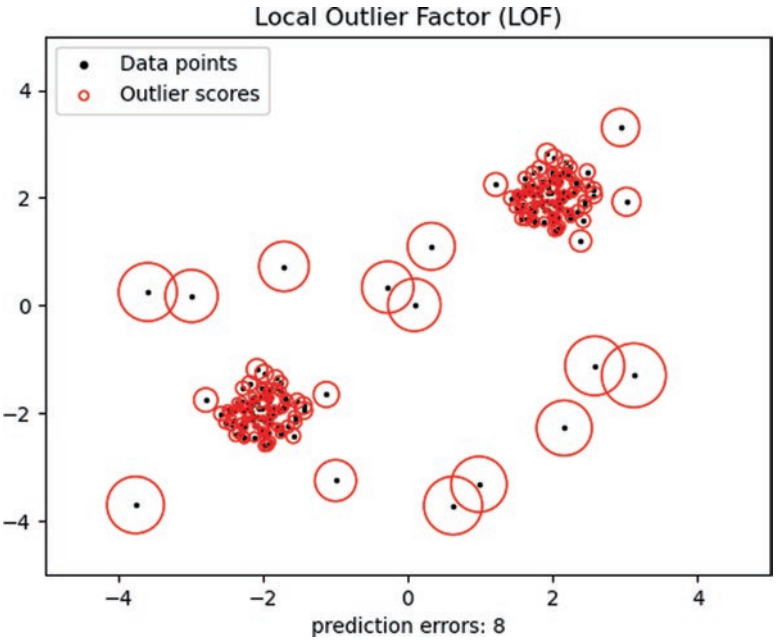


图**3.7** 局部异常因子在数据集上的实现结果[15]

局部异常因子（LOF）通过邻居的平均局部可达距离与该点的局部可达距离的比率来计算（图3.6和图3.7）。

3.4 结果与讨论

我们在广受欢迎的NSL-KDD数据集上进行了实验。 NSL-KDD是KDD'99数据集的新版本（详见表3.1）。这是一个有效的基准数据集，可以帮助研究人员比较不同的入侵检测方法。 在下图中，我们展示了我们的LOF技术在NSL-KDD数据集上的结果（图3.8）。

在下图中，我们展示了我们的LOF技术在NSL-KDD数据集上的预测性能（图3.9）。

Python 代码

下面是我们在KDD数据集上使用基于LOF技术运行实验时的Python代码片段。

表**3.1**NSL-KDD的特征描述

特征名称	描述
持续时间	连接的长度（以秒为单位）
协议类型	协议类型（例如tcp，udp等）
源字节	从源到目的地的数据字节数
目的字节	从目的地到源的数据字节数
服务计数	过去2秒钟内与当前连接具有相同服务的连接数目的主机与源端口相同速率
	连接到相同源端口的连接数

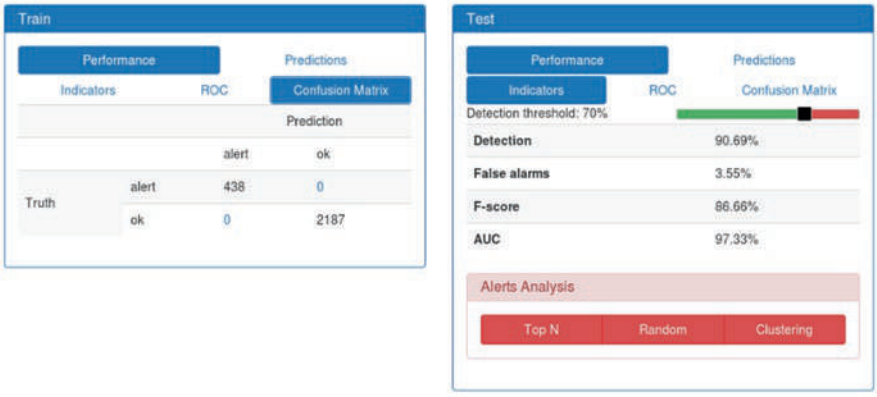


图**3.8**LOF技术在NSL-KDD数据集上的性能

图**3.9**LOF技术在NSL-KDD数据集上的预测性能

```
import numpy as np
from sklearn.metrics import roc_auc_score,  roc_curve

warnings.fiilterwarnings("ignore",   message="numpy.dtype    size
changed")
with warnings.catch_warnings():
 warnings.fiilterwarnings("ignore",category=DeprecationWarning)
warnings.fiilterwarnings("ignore",  message="Using   a   non-tuple
sequence")

import copy
import matplotlib.pyplot as plt
firom mlxtend.evaluate import confiusion_matrix
firom modules import utils, dimension_reduction as dim_red, evalu-
ation as eval, clustering as cluster
import sys

try:
 path = sys.argv[1]
except IndexError:
 is_product = False
else:
 is_product = True

DIMENSION = 30
SEPARATOR = "================================"

# 0. 数据加载
ifi is_product:
 train, ytrain = utils.load_train_data(path,  is_product)
else:
  train,  ytrain  =  utils.load_train_data('./data_in/satan_normal.
csv', is_product)

# 1. 降维
T  = DIMENSION
n  = train.shape[0]
projected = dim_red.pca(train, T, is_product)

# 3. 聚类
predict = cluster.LOF_score(projected)
train["rate"] = predict
train["label"] = ytrain

# 4. 评估
```

```
ifi is_product:
 fior i in train["rate"]:
 print(i)
else:
 fipr, tpr, threshold = roc_curve(ytrain,   train["rate"])
 t = np.arange(0., 5., 0.001)
 utils.plot(1, 1, fipr, tpr, 'b', t, t, 'r')
 print("AUC score : ", roc_auc_score(ytrain,   train["rate"]))
 print("fiinish")
```

3.5 结论

本研究详细介绍了各种新兴技术在网络安全领域的作用。 随着每个互联网用户在现代技术应用中的可靠性显著提高，与广泛使用这些技术相关的风险也随之出现。

尽管技术在过去几十年中显著发展，但网络攻击也呈指数增长。 这些网络攻击威胁到数据的机密性、完整性和可用性，同时破坏了关键基础设施。 为了应对这种侵犯隐私的行为，安全研究人员和专业人士不断设计新的方法和技术来应对安全挑战。 高级机器学习技术，如决策树、局部离群因子、k 最近邻算法、卷积神经网络等，被用来对抗这些攻击。 本文解决了使用局部离群因子检测恶意软件的问题。 这是一种基于无监督学习的高效恶意软件检测技术。 该模型通过异常行为训练，即进程中的异常行为，使其显著有效。

参考文献

1. Anbar, M., Abdullah, N., & Manickam, S. (2019).网络安全的进展. 斯普林格出版社。
2. Nayyar, S. (2021).博客文章：机器学习对网络安全的带来. 福布斯。 2022年5月2日检索自https://www.forbes.com/sites/forbestechcouncil/2021/10/01/what- machine- learning- can- bring- to- cybersecurity/?sh=72e575e01203
3. Ford, V., & Siraj, A. (2014). 机器学习在网络安全中的应用。 在工业和工程领域的第27届国际计算机应用会议论文集中 (Vol. 118). IEEE Xplore.
4. Katzenbeisser, S., Kinder, J., & Veith, H. (2011). 恶意软件检测。 密码学和安全性百科全书, 752–755.https://doi.org/10.1007/978- 1- 4419- 5906- 5_8385.. G*u*ezzaz, A., Benkirane, S., Azro*u*r, M., & Kh*u*rram, S. (2021). 一种可靠的网络入侵检测方法，使用决策树和增强数据质量。 安全和通信网络，2021, 1–8.https://doi.org/10.1155/2021/1230593

6. Bace, R., & Mell, P. (2001). 入侵检测系统。*NIST*，*51*。 从http://cs.uccs.edu/~cchow/pub/ids/*NIST*sp800- 31.pdf检索于2022年*5*月2日
7. Conrad, E., Misenar, S., & Feldman, J. (2017). 领域7。C*ISSP*®第十一小时，145-183。 https://doi.org/10.1016/b978- 0- 12- 811248- 9.00007- 3
8. 入侵检测系统简介。 (2003). 1-38。 https://doi.org/10.1016/b978- 193226669- 6/50021- 5
9. Kim, S., Hwang, C., & Lee, T. (2020). 基于异常的未知入侵检测在终端环境中。电子学，9(6)，1022。https://doi.org/10.3390/electronics906102210. Kotu, V., & Deshpande, B. (2015). 异常检测。预测分析和数据挖掘，329-345。https://doi.org/10.1016/b978- 0- 12- 801460- 8.00011- 2
11. Tang, A., Sethumadhavan, S., & Stolfo, S. (2014). 基于硬件特征的无监督异常基于恶意软件检测攻击、入侵和防御研究，109–129. https://doi.org/10.1007/978- 3- 319- 11379- 1_6
12. 卡巴斯基. (无日期). 卡巴斯基官网. 2022年5月5日检索自https://www.kaspersky.com/enterprise- security/wiki- section/products/machine- learning- in- cybersecurity.
13. GeeksforGeeks. (2020).局部离群因子—GeeksforGeeks. GeeksforGeeks. 2022年5月3日检索自h*tt*ps://www.geeksforgeeks.og/*l*o*cal*- o*utli*er- f*act*or/
14. ajulsam. (无日期).局部离群因子 | 图形实验创建用户指南. Ajulsam.gitbooks.io. 2022年5月3日检索自https://ajulsam.gitbooks.io/graphlab- create- user- guide/content/anomaly_detection/local_outlier_factor.html
15. sckit-learn. (n.d.).使用局部离群因子（*LOF*）进行异常检测. sc*i*k*it-learn*. 从*htt*ps://sc*i*k*it-lea rn.org/st*a*b*le/auto_ex*amples/*nei*g*hb*ors/*plot_l*of_out-l*ier_detection.h*tml*#:~:te*x*t=*局部离群因子（*LOF*）是一种用于检测异常值的算法，它通过比较每个样本的局部密度与其邻居的局部密度来确定异常值。具有比其邻居更低密度的样本被认为是异常值。