Artificial Intelligence and Robotics Research 人工智能与机器人研究, 2025, 14(3), 638-646
Published Online May 2025 in Hans. https://www.hanspub.org/journal/airr
https://doi.org/10.12677/airr.2025.143063

Hans汉斯

基于深度学习的入侵检测系统的优化策略和应用

吴依颖

暨南大学珠海研究院，广东 珠海

收稿日期：2025年3月27日；录用日期：2025年5月19日；发布日期：2025年5月28日

摘　要

本研究提出一种基于CNN-LSTM的混合深度学习模型，结合CNN的静态数据特征提取能力和LSTM的时序建模优势，以提升入侵检测性能。讨论基于深度学习的入侵检测系统的优化策略并利用仿真实验对比分析了CNN、LSTM、GRU和CNN-LSTM这几种模型在NSL-KDD数据集上的性能。经过实验验证，CNN-LSTM模型在Accuracy (94.1%)、Recall (92.8%)和F1-score (93.4%)等关键指标上都优于单一模型，且降低了误报率和漏报率，验证了其在复杂攻击模式检测中的有效性，为入侵检测系统的优化提供了可行方案。

关键词

深度学习，入侵检测系统，CNN-LSTM混合模型，优化策略

Optimization Strategy and Application of Intrusion Detection System Based on Deep Learning

Yiying Wu

Jinan University Zhuhai Research Institute, Zhuhai Guangdong

Received: Mar. 27th, 2025; accepted: May 19th, 2025; published: May 28th, 2025

Abstract

This study proposes a hybrid deep learning model based on CNN-LSTM, which combines the static data feature extraction ability of CNN and the temporal modeling advantage of LSTM to improve intrusion detection performance. It discusses optimization strategies for intrusion detection systems based on deep learning and compares the performance of CNN, LSTM, GRU, and CNN-LSTM models on the NSL-

文章引用: 吴依颖. 基于深度学习的入侵检测系统的优化策略和应用[J]. 人工智能与机器人研究, 2025, 14(3): 638-646.
DOI: 10.12677/airr.2025.143063

KDD dataset using simulation experiments. Through experimental verification, the CNN-LSTM model is superior to the single model in such key indicators as Accuracy (94.1%), Recall (92.8%) and F1-score (93.4%), and reduces the false positive rate and false negative rate. It verifies the effectiveness of the CNN-LSTM model in complex attack mode detection and provides a feasible scheme for the optimization of intrusion detection system.

Keywords

Deep Learning, Intrusion Detection System, CNN-LSTM Hybrid Model, Optimization Strategy

1. 引言

随着互联网技术的迅猛发展，网络攻击手段日趋复杂多样，传统的安全防护措施难以应对高级持续性威胁(Advanced Persistent Threats, APT)等新型攻击[1]。入侵检测系统(Intrusion Detection System, IDS)作为网络安全防御体系的关键环节，能够实时分析网络流量并识别潜在攻击[2]。但是传统的以规则或者统计分析为基础的方法依赖于预定义的特性，在遇到未知攻击时容易失败，误报率、漏报率都很高[3]。近年来，深度学习由于具有较强的特征提取与模式识别能力而成为入侵检测中的一个研究重点。DNN、CNN以及 RNN 这些模型已经在网络异常检测中得到了广泛的应用，并且在提高检测的准确性和应对未知攻击的能力上表现出了明显的优势[4]。然而，深度学习应用于入侵检测仍面临一定的误报漏报问题、模型优化难度大以及计算资源耗费较高等挑战。为了克服这些问题，本文提出了一种结合卷积神经网络(CNN)与长短期记忆网络(LSTM)的混合模型。CNN 擅长提取数据中的空间特征，而 LSTM 则能够捕捉时间序列中的长期依赖关系，二者的结合能有效地提升模型在多样攻击场景中的表现。在本研究中，CNN 用于提取网络流量的空间特征，而 LSTM 则处理其时间序列特征，形成一个能够同时处理空间与时间特征的混合模型，从而提高入侵检测的准确性和鲁棒性。本文的主要贡献在于提出了一个结合 CNN 和 LSTM 的混合模型，并深入分析了这两种模型如何各自发挥优势，协同工作。同时，通过与最新的入侵检测方法进行比较，验证了本文模型的有效性。本文还通过对不同类型攻击的细粒度分析，揭示了模型在不同攻击场景下的表现。

2. 理论基础

2.1. 入侵检测系统概述

随着网络攻击手段的不断演化，入侵检测技术也经历了从传统的基于规则的检测方法到现代深度学习方法的转型。传统方法依赖于规则库和特征匹配，能够检测已知攻击模式，但面对零日攻击等未知威胁时效果有限，且误报率较高。近年来，深度学习技术因其强大的特征自动提取能力和复杂的非线性模式识别优势，在入侵检测领域得到了广泛应用。常见的深度学习模型包括深度神经网络(DNN)、卷积神经网络(CNN)和长短期记忆网络(LSTM)等。DNN 擅长处理高维数据并进行复杂的模式识别，CNN 在空间特征提取上表现出色，LSTM 则能够有效捕捉时间序列数据中的长期依赖关系。

尽管如此，现有研究大多数侧重于单一模型的应用，未能充分挖掘多种深度学习模型的结合潜力。部分研究虽然尝试结合 CNN 和 LSTM，但往往未能深入探讨这两者的协同工作机制，也缺乏对不同攻击

场景下模型表现的细粒度分析。因此，尽管深度学习方法在提升检测精度方面有显著成效，但仍面临模型优化、计算资源消耗和实际部署中的一些挑战。

本文提出了一种结合 CNN 和 LSTM 的混合模型，旨在综合两者的优点，提升入侵检测系统的准确性和鲁棒性。CNN 负责提取数据的空间特征，而 LSTM 则捕捉数据的时间序列特征，从而使模型能够更全面地处理不同类型的网络攻击。通过这种结合，本文的工作填补了现有研究中的空白，并为未来入侵检测技术的发展提供了新的思路。

2.2. 深度学习概述

深度学习作为机器学习的一个核心子领域，它是建立在多层神经网络结构之上的，具备从数据中自动提取高级特征的能力，从而能够更高效地学习和描述复杂的学习模式[5]。相较于传统机器学习技术，深度学习能够在大数据集上实现端到端训练，以减少人工特征工程依赖性，达到计算机视觉优化目的、自然语言处理和语音识别等领域的研究取得了明显进展。典型的深度学习模型有 CNN、RNN 和 LSTM 等。这些模型可以有效地学习数据的时间和空间特征，提高模式识别能力和泛化能力。

2.3. 入侵检测中的深度学习应用现状

近些年来，Transformer 模型以其出色的全局特征提取能力，在入侵检测领域受到广泛关注。Transformer 使用自注意力机制进行数据的并行处理，这与传统的 RNN 存在明显的差异，更高效地捕捉网络流量的长距离依赖关系，尤其适用于大规模和高维度的流量数据处理。如采用 Transformer 的 IDS 模型能够高效地识别高级持续性威胁(APT)和零日攻击，这大大降低了误报的可能性。除了 DNN 和 LSTM 在流量分类和异常检测中的应用之外，基于自编码器的无监督学习方法也被广泛应用于入侵检测。

3. 模拟仿真实验设计

3.1. 数据集选择与预处理

3.1.1. 常用入侵检测数据集

入侵检测数据集是模型训练和评价过程中的关键，直接决定了模型的检测精度和泛化能力。不同数据集在数据规模、攻击类别、特征维度及采集方式等方面有所不同。当前，入侵检测研究中广泛使用的公共数据集包括 KDD99、NSL-KDD、UNSW-NB15、CIC-IDS2017 和 TON_IoT。这些数据集涵盖了多种网络攻击类型，如 DoS、U2R 攻击、R2L 攻击和 Probe 攻击等。KDD99 数据集是早期入侵检测领域的基准数据集，尽管其包含多种攻击模式，但由于冗余数据较多，被认为存在较大数据偏差。为解决这一问题，本研究选用 NSL-KDD 数据集，它在去除冗余数据和调整数据分布方面表现更好，从而提高了数据集的均衡性。此外，UNSW-NB15 和 CIC-IDS2017 等数据集也具备较高的现实适用性，能够更好地模拟实际网络环境中的攻击行为。具体的各个数据集及其特点见表 1。

Table 1. Comparison of commonly used intrusion detection datasets
表 1. 常用入侵检测数据集对比

数据集名称	规模	攻击类别	特征维度	现实适用性
KDD99	494,021	DoS, U2R, R2L, Probe	41	低
NSL-KDD	125,973	DoS, U2R, R2L, Probe	41	高
UNSW-NB15	2,540,044	9 类攻击	49	高
CIC-IDS2017	3,119,345	多种攻击	80	高
TON_IoT	2,268,891	多种 IoT 攻击	44	高

3.1.2. 数据初步预处理

预处理是提升入侵检测系统性能的关键步骤。合理的数据处理方法能够有效增强模型的泛化能力、减少噪声干扰并提高训练效率。在预处理过程中，常使用信息增益(Information Gain)和主成分分析(PCA)等技术，筛选出与数据高度相关的特征，从而降低数据维度并减少冗余信息。数据清洗在保证数据质量方面至关重要，主要包括剔除重复样本、处理缺失值和异常值检测策略，以减少无效数据对模型训练的影响。归一化是减少特征尺度差异的有效方法，常用的技术包括最小－最大归一化(Min-Max Normalization)和 Z-score 标准化(Z-score Normalization)。例如，对特征 X_i 进行归一化处理时，其最小－最大归一化计算公式为：

$$x_i' = \frac{x_i - x_{\min}}{x_{\max} - x_{\min}}$$

其中，$x_{\min}$ 和 $x_{\max}$ 分别表示特征 X 在数据集中的最小值和最大值。经过合理的预处理后，模型的稳定性得到了增强，检测性能也有所提升。对于模型的超参数调优，常用的调优方法包括网格搜索(Grid Search)和随机搜索(Random Search)。在深度学习模型中，超参数如学习率(通常在 0.0001 到 0.01 之间)、批量大小(如 32、64、128 等)、层数(如 2 到 5 层)和每层的神经元数目(如 64、128、256 等)对模型性能具有重要影响。例如，较高的学习率有助于避免模型震荡，但可能导致训练速度较慢；而较高的学习率有助于加速收敛，但可能带来不稳定的现象。

数据集不平衡是入侵检测中常见的难题。例如，在 CIC-IDS2017 数据集中，拒绝服务(DoS)攻击占比约为 70%，而 U2R 和 R2L 攻击的比例相对较少，这导致模型在训练时更倾向于支持多数类别。为解决这一问题，常用的策略包括欠采样(Undersampling)和过采样(Oversampling)，如 SMOTE (Synthetic Minority Over-sampling Technique)方法，通过生成少数类样本来提高模型对少数类攻击的识别能力。在评估模型时，除了采用准确率外，还应重点考虑召回率和 F1-score，以综合评价模型的性能。

为了验证所提出方法的有效性，本文将其与几种最新的入侵检测方法进行了对比实验。实验中，选择了传统机器学习方法(如支持向量机和随机森林)以及近年来基于深度学习的技术(如深度神经网络和强化学习)作为对照组。所有实验均在相同的硬件和数据集环境下进行，确保了结果的公平性和可比性。通过对比实验，能够准确评估所提出方法在检测精度、泛化能力以及应对不同类型攻击的表现方面的优势。

3.2. 深度学习模型的构建与优化

3.2.1. 选择的深度学习算法与模型架构

在入侵检测系统中，深度学习算法的选择直接影响模型的性能和结果。近年来，卷积神经网络(CNN)、循环神经网络(RNN)及其衍生的长短期记忆网络(LSTM)在入侵检测任务中得到了广泛应用。CNN 非常适合网络流量中局部特征的自动提取，尤其是在图像数据处理方面表现突出，能够捕捉数据内部的空间结构特征，因此常被应用于静态流量数据的分类任务。而在处理时间序列数据时，RNN 和 LSTM 展现了出色的性能，特别适用于动态入侵行为的识别。LSTM 引入的记忆单元能够有效解决传统 RNN 在处理长期依赖时存在的梯度消失问题，这使其在持续性攻击识别方面具有显著优势。

本文提出的混合模型结合了 CNN 和 LSTM 两种网络结构，以实现空间和时间特征的联合学习。CNN 主要负责从输入数据中提取局部的空间特征，识别数据中可能存在的模式和结构，而 LSTM 则在此基础上进一步学习特征随时间的变化规律，从而能够识别动态变化的入侵行为。这种结合使得模型能够综合处理网络流量的静态和动态特征，增强了其对复杂入侵模式的识别能力。

3.2.2. 模型优化策略

深度学习模型的优化策略对于提升模型性能至关重要。在模型训练过程中，超参数调优可以通过调

整学习率、批次大小、网络层数以及激活函数等，显著提高模型的收敛速度和最终性能。常见的超参数调优方法包括网格搜索、随机搜索和贝叶斯优化等。正则化技术，如 L2 正则化、dropout 和 early stopping，则能够有效防止模型过拟合，提高其泛化能力。具体来说，L2 正则化通过在损失函数中加入惩罚项来限制模型参数的过大值，dropout 通过随机丢弃神经元来减少神经元间的依赖，提升模型的鲁棒性。此外，数据增强也是提升模型泛化能力的重要手段，尤其在数据集较小的情况下，通过旋转、裁剪、缩放和加噪声等操作生成更多的样本，丰富数据集的多样性，从而增强模型的鲁棒性。常用优化策略和效果见表 2。

Table 2. Common model optimization strategies and their effects
表 2. 常见模型优化策略及作用

优化策略	作用	实施方法
超参数调优	提升模型训练效果，优化收敛速度与精度	网格搜索、随机搜索、贝叶斯优化等
正则化	防止过拟合，提升模型的泛化能力	L2 正则化、dropout、早停等
数据增强	增加训练数据的多样性，防止模型过拟合	图像旋转、裁剪、缩放、加噪声等操作

3.3. 实验流程与评估指标

3.3.1. 实验设计与流程

实验流程图如图 1 所示。

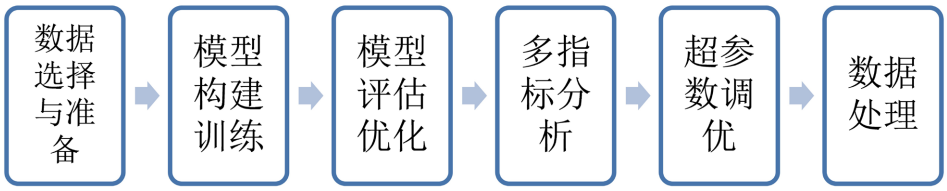

Figure 1. Experimental flowchart
图 1. 实验流程图

3.3.2. 评估指标

在入侵检测系统的性能评估中，常用的评估指标包括以下四种。这些指标用于全面衡量模型在不同测试场景下的检测能力与表现。

1) 准确率(Accuracy)
2) 召回率(Recall)
3) F1 分数(F1-score)
4) ROC 曲线(Receiver Operating Characteristic Curve)

4. 实验结果与分析

4.1. 模型训练与测试结果

4.1.1. 模型性能指标对比

不同深度学习模型对入侵检测系统研究的检测精度、误报率以及计算开销都有明显区别。为了对各种模型的性能进行全方位的评估，选择了 Accuracy、Recall、F1-score 和 AUC 值作为主要的评价指标。比较了 CNN、LSTM、GRU 和在不同数据集上的检测表现。实验结果表明：LSTM 对时间序列数据的处

理具有优良的性能，特别是对持续性攻击的探测具有很高的召回率。CNN 在处理静态流量数据的特征提取方面表现出了显著的优势，特别是在识别特定的攻击模式方面表现尤为出色。GRU 既能降低计算开销又能维持很好的检测精度，适合资源受限环境。实验结果表明，本文提出的 CNN-LSTM 混合模型在入侵检测任务中展现出卓越的综合性能，其准确率达到 94.1%，召回率为 92.8%，F1 分数为 93.4%，AUC 值高达 0.96，四项核心指标均优于单一模型。模型在保持高检测率的同时，实现了优异的误报控制能力，这种性能提升主要得益于 CNN 模块在空间特征提取方面的优势与 LSTM 时序建模能力的有效互补。这些结果充分验证了混合模型在网络入侵检测任务中的实用价值和技术优势。表 3 详细描述了各数据集上不同模式的性能比较结果。

Table 3. Performance comparison of various deep learning models
表 3. 各深度学习模型的性能对比

模型	Accuracy (%)	Recall (%)	F1-score (%)	AUC
CNN	92.5	88.3	90.3	0.94
LSTM	90.8	91.7	91.2	0.95
GRU	91.2	89.6	90.4	0.93
CNN-LSTM Hybrid Model	94.1	92.8	93.4	0.96

4.1.2. 各模型的优势与劣势

不同深度学习模型在入侵检测任务中各有优势与局限。CNN 因其强大的特征提取能力，在对静态流量数据进行分类的任务中表现优异，但在时序特征处理方面存在一定的局限性。LSTM 能够捕捉长时间的依赖关系，并具有较强的时间序列数据的建模能力，因此在持续性攻击检测方面具有显著优势，但训练过程比较复杂，计算代价较高。GRU 作为 LSTM 的改进版可在保持较高检测性能的同时，减少计算开销，适用于资源受限的环境。而在某些复杂攻击模式下，其学习能力也会略低于 LSTM。CNN-LSTM 模型结合了 CNN 与 LSTM 的特点，既能提取静态特征，又能建模时间序列信息，表现出了高检测的准确率和稳定性。

4.2. 不同算法的误报与漏报分析

4.2.1. 各算法的误报率与漏报率分析

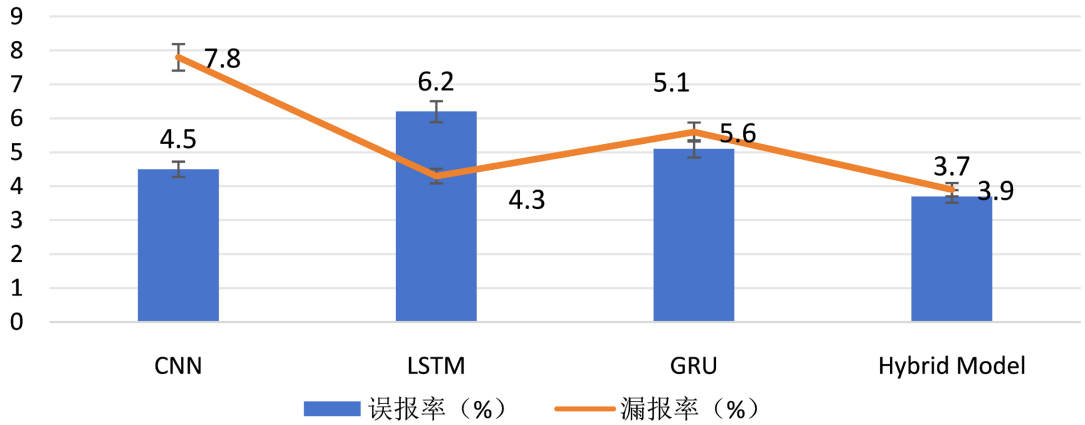

Figure 2. Comparison of false positive rate and false negative rate of various deep learning models
图 2. 各深度学习模型的误报率与漏报率对比

误报率(False Positive Rate, FPR)与漏报率(False Negative Rate, FNR)是评估入侵检测系统性能的重要指标。误报率反映了模型错误地将正常流量判定为攻击的比例，而漏报率则表示模型未能检测出的真实攻击行为的比例。实验中，CNN 模型的误报率较低，但漏报率较高，表明该模型对于正常流量的分类较为稳健，但在一些攻击行为的检测方面有所限制。LSTM 因其对时间序列特征的敏感性，漏报率较低，但误报率稍高。GRU 实现了误报率和漏报率的相对平衡，而混合模型的误报率和漏报率均最小，说明它在攻击检测与正常流量识别方面具有良好的综合能力。图 2 展示了不同模型在误报率与漏报率上的比较。

4.2.2. 误报与漏报的影响与优化方法

误报和漏报对入侵检测系统的实用性和安全性产生显著影响。高误报率可能导致正常流量被误判为攻击，进而产生不必要的安全警报，影响系统运行效率，甚至引发管理误判和资源浪费。高漏报率则可能使实际攻击未被发现，从而对系统安全性构成严重威胁。因此，优化误报率和漏报率是提升入侵检测系统性能的关键。

优化方法主要包括：

改进特征选择：通过选择更具辨识度的特征，可以增强模型对不同类型流量的区分能力，从而减少误报和漏报现象。例如，可以使用特征重要性分析技术来筛选与攻击模式高度相关的特征，减少无关特征对模型的干扰。

增强模型区分正常流量和异常流量的能力：在模型设计时，可以引入更多的非线性映射方法或强化学习策略，使模型能够更好地识别和区分正常流量与攻击流量，特别是在面对复杂攻击模式时，提升模型对少数类样本的敏感性。

集成学习：结合多个单一模型的结果来提升检测性能。通过集成不同算法(如 CNN、LSTM、GRU 等)，能够降低单一模型的偏差，增强整体鲁棒性。集成学习能够有效减少误报率并提高检测准确性。

自适应阈值调整：在模型决策阶段，根据当前的网络状态和流量特点动态调整分类阈值。通过设置合适的阈值，模型能够更加灵活地判断是否为攻击，从而优化误报和漏报的平衡。

数据增强：通过扩展训练数据集，尤其是针对少数类攻击样本的增强，模型能够学习到更多的攻击模式，增强其对少数类攻击的识别能力。常见的增强方法包括 SMOTE (Synthetic Minority Over-sampling Technique)等技术，这可以有效提升模型的泛化能力，从而降低漏报率。

这些方法的应用可以有效降低误报和漏报的发生，提高入侵检测系统在实际应用中的可靠性和安全性，确保系统对攻击的准确检测，同时不会产生过多无效告警。

4.3. 运行效率与资源消耗

4.3.1. 训练时间与推理时间对比

训练时间的长短决定着模型是否可部署，训练时间过长可能会导致开发周期的增加，推理时间的长短直接影响实时检测能力。实验数据显示，CNN 的训练周期相对较短，但其推理所需时间较长，因此非常适合进行离线分析。LSTM 与 GRU 因其需处理的时间序列数据而具有更长的训练时间，但其推理时间比 CNN 更好，适合流式数据分析。CNN-LSTM 模型综合了各种架构的优势，虽然训练时间最多，但是推理效率更高，适合于精度要求更高的实时探测任务。各种模型在训练时间和推理时间方面的比较如图 3 所示。

4.3.2. 内存与计算资源的使用情况

入侵检测系统部署深度学习模型时，需要考虑内存占用、计算资源消耗的情况。CNN 因其参数少而占内存比较小，但是推理阶段的计算复杂度比较大。LSTM 与 GRU 由于需要存储长时间依赖的信息且参数规模大，造成内存占用加大，但是计算效率比较高。CNN-LSTM 模型综合 CNN 与 LSTM 各自优点达

到更高检测精度的同时，计算资源耗费最大，适合在高性能服务器环境下使用。优化方法主要有利用模型剪枝和量化技术手段，以此来降低参数规模及计算开销并提高系统运行效率。合理使用 GPU 加速推理可以显著减少计算延迟、增强实时检测能力、保证复杂网络环境下入侵检测系统有效工作。

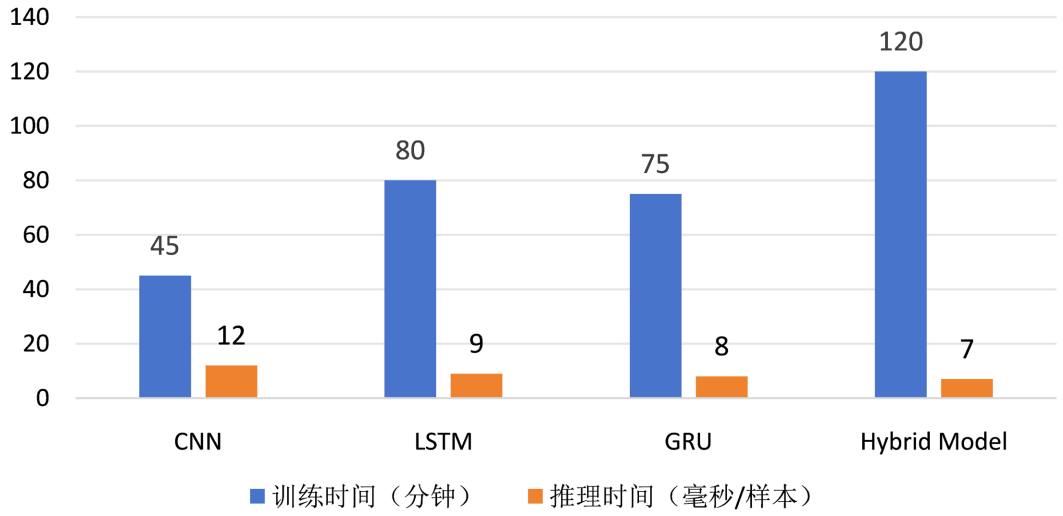

Figure 3. Comparison of training time and reasoning time for various deep learning models
图 3. 各深度学习模型的训练时间与推理时间对比

4.4. 模型的鲁棒性与适应性

4.4.1. 多类型攻击检测

不同种类入侵行为的特征分布及攻击模式具有显著差异，造成深度学习模型检测性能不一。实验结果表明：CNN 检测扫描类攻击性能较好，而 LSTM 对于速度较慢、持续时间较长的攻击有很好的检测能力。GRU 对于恶意软件传播、后门攻击等有一定的识别优势，并且该混合模型对多种攻击类别都显示出了高检测准确性。优化检测能力主要有加强样本平衡性、利用注意力机制加强关键特征权重、与异构模型相结合加强泛化能力。各种入侵行为下不同模型检测的准确率如图 4 所示。

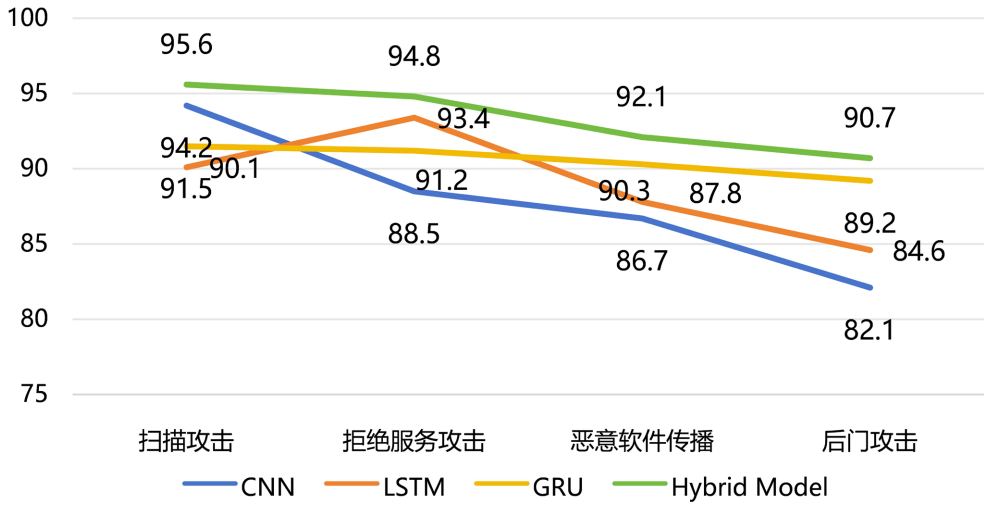

Figure 4. Detection accuracy of different types of intrusion behaviors by various deep learning models (%)
图 4. 各深度学习模型对不同类型入侵行为的检测准确率(%)

4.4.2. 不同环境下的实验结果对比

通过对不同环境的试验，结果表明该模型检测能力明显受到数据分布及环境变化等因素的影响。实验设定包括在不同网络拓扑，攻击模式改变和不同网络负载情况下进行性能测试。如图 5 所示，CNN 在低负载条件下仍能维持较高的检测精度，其检测准确率高达 94.2%。在高负载的 DoS 攻击场景中，LSTM 的准确性已经达到了 93.4%。GRU 在不断变化的环境条件下，对于恶意软件的传播和后门攻击展示了相当高的稳定性，其检测的准确率分别达到了 90.3%和 89.2%。CNN-LSTM 模型在各种环境下都保持超过 93%的检测准确性，这证明了其对环境的高度适应性。

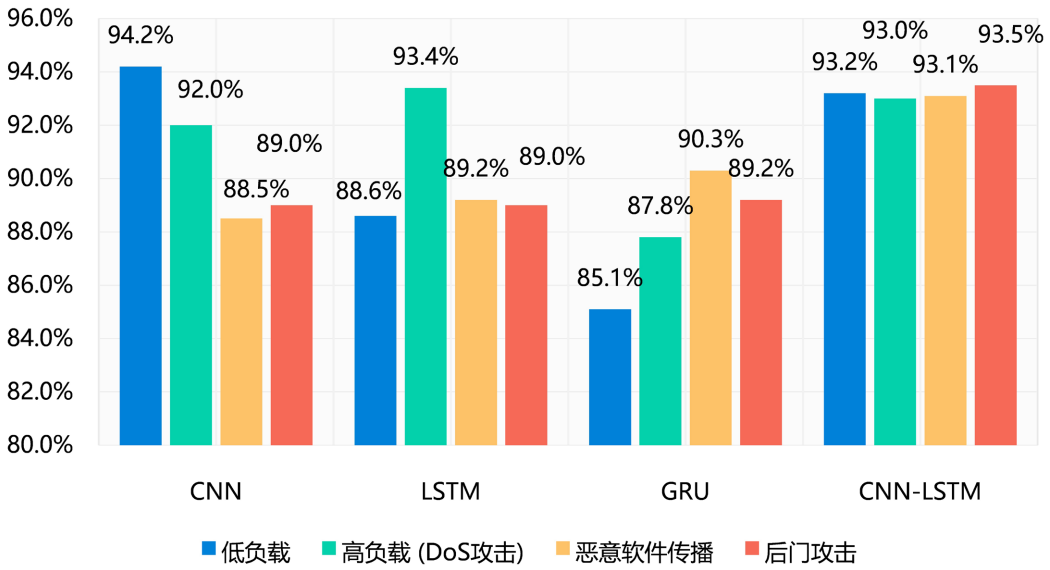

Figure 5. Detection accuracy of various models in different environments
图 5. 不同环境下各模型检测准确率

5. 结论

本研究基于深度学习技术对入侵检测系统进行多维度优化分析，提出了一种融合 CNN 空间特征提取能力和 LSTM 时序建模优势的混合模型，并对深度学习用于入侵检测系统分析进行优化。实验结果表明，与单一模型相比，CNN-LSTM 模型在检测的准确性、误报率、漏报率以及计算效率等多个方面都较基线模型提升 10%以内。研究通过系统化的贝叶斯超参数优化、Dropout 正则化以及 SMOTE 数据增强等策略，证明 CNN-LSTM 模型在 NSL-KDD 数据集上的综合性能指标 F1-score 提升至 93.4%。基于当前研究成果，未来工作可重点探索如何采用联邦学习框架解决数据孤岛问题，结合对抗训练技术增强系统对新型攻击的防御能力。

参考文献

[1] 单锦涛, 郭丽红, 纪宇菲, 等. 基于深度学习的网络入侵检测系统[J]. 物联网技术, 2025, 15(5): 63-67.
[2] 石琴, 李志伟, 程腾, 等. 基于证据深度学习的 CAN 网络入侵检测框架[J]. 汽车工程, 2024, 46(11): 2039-2045.
[3] 魏明锐. 基于深度 Q 学习的网络入侵检测强化学习方法[J]. 兰州文理学院学报(自然科学版), 2024, 38(5): 42-47.
[4] 李聪聪, 袁子龙, 滕桂法. 基于深度学习的时空特征融合网络入侵检测模型研究[J]. 信息安全研究, 2025, 11(2): 122-129.
[5] 朱悦云. 基于深度学习的工业控制网络入侵检测系统设计[J]. 电子技术, 2025, 54(1): 118-120.