# 基于 VAE-CWGAN 和特征统计重要性融合的网络入侵检测方法

刘涛涛[1]，付钰[1]，王坤[1,2]，段雪源[1,3]

（1. 海军工程大学信息安全系，湖北 武汉 430033；
2. 信阳职业技术学院数学与信息工程学院，河南 信阳 464000；
3. 信阳师范大学计算机与信息技术学院，河南 信阳 464000）

**摘 要**：针对传统入侵检测方法受限于数据集类不平衡以及所选特征代表性不强等问题，提出一种基于 VAE-CWGAN 和特征统计重要性融合的检测方法。首先，为提升数据质量对数据集进行预处理；其次，搭建 VAE-CWGAN 模型生成新样本以解决数据集类不平衡问题，使分类模型不再偏向于多数类；再次，使用标准差、中值均值差对特征进行排序，并融合其统计重要性来进行特征选择旨在获得代表性更强的特征，从而使模型更好地学习数据信息；最后，通过一维卷积神经网络对特征选择后的混合数据集进行分类。实验结果表明，所提方法在 NSL-KDD、UNSW-NB15 和 CIC-IDS-2017 数据集上都表现出较好的性能优势，准确率分别为 98.95%、96.24% 和 99.92%，有效提升了入侵检测性能。

**关键词**：入侵检测；网络流量；类不平衡；特征选择；统计重要性融合

**中图分类号**：TP391

**文献标志码**：A

**DOI:** 10.11959/j.issn.1000−436x.2024013

# Network intrusion detection method based on VAE-CWGAN and fusion of statistical importance of feature

LIU Taotao[1], FU Yu[1], WANG Kun[1,2], DUAN Xueyuan[1,3]

1. Department of Information Security, Naval University of Engineering, Wuhan 430033, China
2. School of Mathematics and Information Engineering, Xinyang Vocational and Technical College, Xinyang 464000, China
3. College of Computer and Information Technology, Xinyang Normal University, Xinyang 464000, China

**Abstract:** Considering the problems of traditional intrusion detection methods limited by the class imbalance of datasets and the poor representation of selected features, a detection method based on VAE-CWGAN and fusion of statistical importance of features was proposed. Firstly, data preprocessing was conducted to enhance data quality. Secondly, a VAE-CWGAN model was constructed to generate new samples, addressing the problem of imbalanced datasets, ensuring that the classification model no longer biased towards the majority class. Next, standard deviation, difference of median and mean were used to rank the features and fusion their statistical importance for feature selection, aiming to obtain more representative features, which made the model can better learn data information. Finally, the mixed data set after feature selection was classified through a one-dimensional convolutional neural network. Experimental results show that the proposed method demonstrates good performance advantages on three datasets, namely NSL-KDD, UNSW-NB15, and CIC-IDS-2017. The accuracy rates are 98.95%, 96.24%, and 99.92%, respectively, effectively improving the performance of intrusion detection.

**Keywords:** intrusion detection, network traffic, class imbalance, feature selection, fusion of statistical importance

**收稿日期**：2023−08−24；**修回日期**：2023−11−07

**通信作者**：付钰，fuyu0219@163.com

**基金项目**：国家重点研发计划基金资助项目（No.2018YFB0804104）

**Foundation Item:** The National Key Research and Development Program of China (No.2018YFB0804104)

## 0　引言

随着科技的不断发展，电子设备数量呈几何级增长，网络环境也更加复杂多变，导致网络安全问题频出，如隐私泄露、恶意攻击等，给国民经济造成了巨大的损失。因此，如何有效地避免网络攻击是当前亟待解决的问题。

为解决上述问题，通常构建入侵检测系统（IDS, intrusion detection system）来提升网络空间安全。该系统采用主动防御的方法，可有效检测入侵并及时做出响应[1]，已成为防范网络攻击的重要手段之一。IDS 通常可以分为基于主机的 IDS 以及基于网络的 IDS，前者主要通过监视日志信息进行检测，后者则是分析网络流量来判定是否存在入侵行为[2]。基于机器学习（ML, machine learning）的 IDS 应用较广泛且取得了良好的检测性能，如贝叶斯[3]、支持向量机（SVM, support vector machine）[4]等，但随着网络攻击手段不断进步，ML 难以提取深层特征的缺陷愈发明显，因此仅依靠 ML 的 IDS 已难以检测出入侵行为，亟待寻找新的检测技术。近年来，深度学习（DL, deep learning）因其强大的特征提取能力被应用于各个领域，且已逐渐成为 IDS 中的重要组成部分[5]。不过基于 DL 的 IDS 仍然具有较高的假阳率，其主要原因有两点：数据集不平衡以及所选特征差异性不显著。

在实际网络活动中，正常流量占据大多数，异常流量的数量较少，这种类不平衡数据集导致分类模型在检测过程中会更关注大多数类别，从而忽略少数类攻击或直接将其淹没。但对于 IDS 来说，检测少数类攻击是至关重要的，若其被误检测为正常流量，将给用户设备造成更大的损失。同时，大多入侵数据集都包含较高的特征维度，其中可能存在冗余特征或不相关特征使正常流量与异常流量之间的差异不明显[6]，故容易影响模型对正常流量和异常流量进行判定，导致误报率较高。

为解决上述问题，本文提出一种结合变分自编码器（VAE, variational autoencoder）、条件 Wasserstein 生成对抗网络（CWGAN, conditional Wasserstein generative adversarial network）与特征统计重要性融合的网络入侵检测方法。首先，通过 VAE-CWGAN 对原始数据集中的少数类攻击进行数据生成以平衡数据集；然后，提出一种基于过滤方法的特征选择技术，该技术利用标准差、均值中值差来构建新数据集；最后，使用一维卷积神经网络（1DCNN, one dimension convolutional neural network）对新数据集进行分类。本文工作的主要贡献如下。

1) 针对 IDS 中类不平衡问题，本文提出一种基于 VAE-CWGAN 的样本生成方法，将 VAE 作为生成器，通过 CGAN 来调整网络参数，并在损失函数中添加 Wasserstein 距离及梯度惩罚项以防止梯度弥散无法收敛。

2) 本文设计了一种过滤式的特征选择技术，利用标准差、均值中值差对特征重要性进行统计排序并融合。该技术可视为一种启发式技术，依据特征重要性来推导出新特征集以增强 1DCNN 的检测性能。

3) 本文在 NSL-KDD[7]、UNSW-NB15[8]以及 CIC-IDS-2017[9]这 3 个典型流量数据集上进行实验，通过与其他方法在多个评估指标上的对比可知，所提方法解决了数据不平衡、特征差异不显著等问题，具有较高的检测率（DR），可有效防范网络入侵。

## 1　相关研究工作

### 1.1　机器学习技术

在过去的几十年里，许多机器学习方法被应用于网络入侵检测中。其中，文献[10]提出一种多个支持向量机相结合的入侵检测系统，每一个 SVM 只负责检测特定攻击类型，可准确地检测出攻击，保证了设备的安全；文献[11]利用决策树构建流量分类器，实现了较高的识别精度；文献[12]通过构造随机森林（RF, random forest）实现入侵检测并在 NSL-KDD 数据集上进行实验，取得较好的效果；文献[13]提出一种隐藏朴素贝叶斯模型，可处理高维度、高数据流的入侵检测问题，在提升准确率的同时降低了资源需求。然而，随着网络攻击的多样化，以机器学习技术为代表的浅层学习已经难以满足当前网络流量的需求。因此，利用新型的学习算法实现入侵检测不失为一条有效的研究路线。

### 1.2　深度学习技术

近年来，基于深度学习的入侵检测技术开始崭露头角并逐渐成为主流。文献[14]利用小波变换和栈式自编码器学习样本的多尺度信息，并通过集成学习检测其性能，该方法在 NSL-KDD 数据集上对异常流量能达到 94.61%的精确率。文献[15]首先设计了一种可将流量特征转换为波形或声音特征的

入侵检测系统，然后利用卷积神经网络（CNN）、长短期记忆（LSTM, long short term memory）网络、深度信念网络（DBN, deep belief network）等深度学习方法检测异常行为，该技术在 NSL-KDD 和 CIC-IDS-2017 数据集上分别实现了 84.82%和 99.41%的检测率，体现了该系统的巨大潜力。文献[16]利用密集连接 CNN 进行入侵检测，并通过改进后的混合函数缩小类内距离及放大类间距离，从而更好地识别 KDD CUP 99 数据集中的攻击类型。文献[17]利用门控循环单元学习二元细菌觅食算法所选择的最优特征子集，而且采用 Nesterov 加速自适应矩估计优化超参数，该方法具备一定的新颖性和检测性能。文献[18]提出一种通用的特定自编码器（AE, auto encoder）深度学习架构，其中，通用 AE 学习网络入侵中的共同特征，而特定 AE 则仅学习该领域的特定特征，所取得的效果令人满意。文献[19]构建了一种基于多尺度残差的异常流量检测方法，利用多个不同的残差网络学习重构误差中不同尺度的信息，在 NSL-KDD 和 UNSW-NB15 数据集上对异常流量分别实现了 94.43%和 90.12%的精确率。然而，上述方法未能解决数据集中类不平衡问题，数据集中正常样本远多于异常样本，并且各类攻击在异常样本中所占的比例也相差较大，而深度学习模型需要通过大量数据训练才能取得更高的精度，因此某些攻击不容易被 DL 模型所学习，可能将其判定为其他类型攻击或正常行为，对网络构成较大威胁，所以需要解决数据集中的类不平衡问题以提升检测性能。

### 1.3 类平衡技术

类不平衡作为 IDS 中常见的问题已经严重影响系统性能，当前大多研究方法采用重采样技术进行数据生成以解决该问题。文献[20]采用自适应合成（ADASYN, adaptive synthetic）采样和重复编辑最近邻（RENN, repeated edited nearest neighbor）相结合的混合采样方法解决正负样本不平衡的问题。文献[21]将基于聚类的合成少数过采样技术（SMOTE, synthetic minority oversampling technique）与 K-Means 的欠采样技术相结合，既能避免 SMOTE 所带来的过大成本代价，又能防止随机欠采样丢失重要信息，可有效平衡数据。文献[22]使用 K 近邻欠采样方法和 GAN 的变体模型平衡样本，有助于模型更好地学习样本分布。但是，重采样技术可能会出现信息丢失、生成样本拟合度低等问题。而随着生成模型的进步，利用其生成新数据也成为一种新思路。常用的生成模型主要为变分自编码器[23]及生成对抗网络[24]。文献[25]使用改进的条件 VAE 来学习流量数据内在分布以生成指定攻击样本，增加了样本的多样性，从而提升少数类攻击的检测率。文献[26-27]则运用 CWGAN 为少数类攻击生成新样本以缓解类不平衡问题，并通过仿真实验验证了所提方法的有效性。因此，本文在上述文献的基础上，将 VAE 与 CWGAN 相结合来为指定的少数类攻击生成样本，既解决了 VAE 所生成样本模糊的问题，又对样本提前学习降低了 CWGAN 的训练成本，同时加入梯度惩罚项防止数据在训练过程中不收敛。

### 1.4 特征选择技术

通过上述方法缓解了类不平衡问题后，IDS 的检测能力得到提升，不过由于数据集中的特征较多，检测性能易被不相关的特征所影响，因此，研究特征选择技术从而筛选出最能反映数据集中类间差异的特征是非常必要的。

对于入侵检测而言，特征选择技术大体上分为 3 种：过滤法、嵌入法和包装法。过滤法根据离散度和相关性对每个特征进行评分，然后通过所设阈值筛选特征；嵌入法采用机器学习得到每个特征的权重系数，然后进行排序筛选；包装法则基于目标函数，一次选择或排除几个特征。文献[28]提出一种基于包装法的特征选择技术，通过墨鱼算法（CFA, cuttlefish algorithm）进行特征选择，实验表明，对于具有 15 个特征的 KDD CUP 99 数据集而言，该特征选择技术实现了 91.50%的 DR 以及 3.37%的虚警率。文献[29]将二元灰狼优化算法（BGWOA, binary grey wolf optimization algorithm）应用于特征选择中，从结果可知，具有 21 个特征的 NSL-KDD 数据集的 DR 为 98.47%，精确率为 98.60%，在 CIC-IDS-2017 数据集上的 DR 为 99.47%，精确率为 99.48%。

文献[30]采用了基于嵌入法的特征选择技术，利用 RF 计算每个特征的特征重要性并按重要性值对特征进行排序，所提方法在 UNSW-NB15 数据集上进行实验，将其特征缩减为 11 个并取得了不错的效果；文献[31]将 RF 与 SVM 相结合以计算出特征重要性值，在 KDD CUP 99 数据集上通过 14 个特征实现了 93%的 DR 以及 3%的虚警率。

文献[32]提出一种基于信息增益（IG, information gain）的特征过滤技术，该技术将

UNSW-NB15 数据集中的特征缩减为 22 个，实现了 84.83%的准确率；文献[33]利用互信息（MI, mutual information）特征选择技术来筛选特征，并且通过最小二乘支持向量机对缩减后的数据集特征进行学习分类，实验结果表明，该方法在 KDD CUP 99 数据集上的 DR 为 99.46%，在 NSL-KDD 数据集上的 DR 为 98.76%。

综上可知，上述方法已在异常流量数据集上进行验证且均取得了较好的检测性能，现有方法对比如表 1 所示。然而，嵌入法对异常值较敏感，易陷入局部最优；包装法作为搜索式的特征选择技术，计算成本过高。因此，本文采用过滤法进行特征选择，首先通过标准差、均值中值差对每个特征进行排序，其次将特征所对应的 2 个排序值相加，最后根据相加结果由大到小逐个将特征添加到 1DCNN 模型中，直至新添加特征的检测性能不如前一个特征。

**表 1　　现有方法对比**

| 方法 | 特征选择技术 | 数据集 | 检测性能 |
|---|---|---|---|
| 文献[28] | CFA | KDD CUP 99 | DR=91.50% |
| 文献[29] | BGWOA | NSL-KDD,<br>CIC-IDS-2017 | DR=98.47%<br>DR=99.47% |
| 文献[30] | RF | UNSW-NB15 | KNN 准确率为 71.01%<br>DT 准确率为 74.22%<br>BME 准确率为 74.64%<br>XGBoost 准确率为 71.43%<br>RF 准确率为 74.87% |
| 文献[31] | RF-SVM | KDD CUP 99 | DR=93.00% |
| 文献[32] | IG | UNSW-NB15 | 准确率为 84.83% |
| 文献[33] | MI | KDD CUP 99<br>NSL-KDD | DR=99.46%<br>DR=98.76% |

## 2　相关理论基础

### 2.1　VAE

VAE 是在自编码器的基础上引入变分思想的生成学习模型，由 Kingma 等[23]于 2013 年提出，结构如图 1 所示。

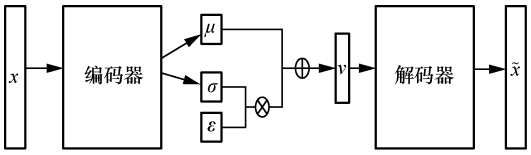



图 1　VAE 结构

图 1 中，$x$ 为真实数据；$\mu$ 为均值；$\sigma$ 为标准差；$\boldsymbol{\varepsilon}$ 为随机向量，符合标准正态分布；$v$ 为隐变量；$\tilde{x}$ 为重构数据。VAE 利用编码器将数据压缩为隐变量，再通过重参数化解决梯度无法求偏导的问题，使隐变量的分布尽量拟合独立的已知分布[34]，最后使用解码器对隐变量进行映射重构以生成数据。VAE 的损失函数定义为

$$L_{\mathrm{VAE}} = -\mathrm{E}_{q(z|x)}[\log p(x|z)] + D_{\mathrm{KL}}[q(z|x)\,\|\,p(z)] \tag{1}$$

其中，第一项为重构误差，$p(x|z)$ 表示解码过程；第二项中 $D_{\mathrm{KL}}$ 为 KL（Kullback-Leibler）散度，可直接测量先验分布 $p(z)$ 与近似后验分布 $q(z|x)$ 之间的相似程度。

### 2.2　GAN

GAN 由生成器和判别器组成，其中，生成器通过学习真实数据的内在分布，用随机噪声生成虚假数据；判别器则是对输入数据是否为虚假数据进行判断。生成器的目的是让生成数据更真实，使判别器对输入数据是否为虚假数据难以预测；判别器的目的是让自身判断更准确。GAN 通过这种对抗方式使虚假数据更好地拟合真实数据分布，其结构如图 2 所示。

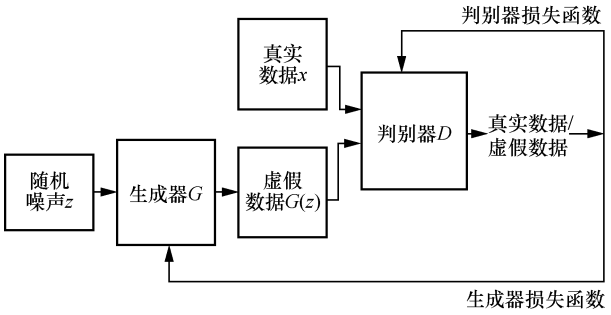



图 2　GAN 结构

由图 2 可知，GAN 中的生成器和判别器通过损失来调整参数，其损失函数为

$$\min_G \max_D V(G,D) = \min_G \max_D \mathrm{E}_{x\sim p_r}[\log D(x)] + \mathrm{E}_{z\sim p_z}[\log(1-D(G(z)))] \tag{2}$$

其中，$x$ 表示真实数据，$z$ 为随机噪声，$p_r$ 和 $p_z$ 分别代表真实数据和虚假数据的分布，$G(z)$ 为生成的虚假数据，$D(\cdot)$ 为判别器所给出的分数。

对于入侵检测数据集而言，只需要对几种少数类攻击进行数据生成，而 GAN 并不能生成指定类样本，因此本文通过使用条件生成对抗网络（CGAN, conditional generative adversarial network），将类信息引入 GAN 中作为生成器的输入，同时在鉴别器中接收样本所对应的类信息，以此生成指定类样本。CGAN 结构如图 3 所示。

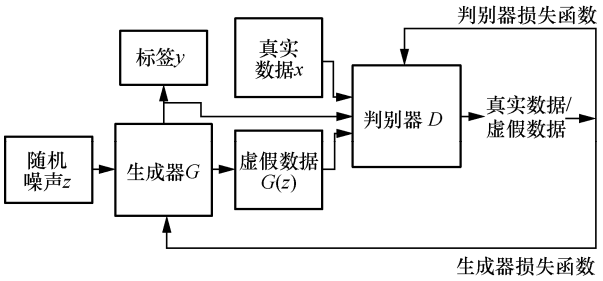



图 3 CGAN 结构

CGAN 仅在 GAN 的基础上添加了类信息，因此其损失函数为

$$\min_G \max_D V(G,D) = \min_G \max_D \mathrm{E}_{x\sim p_r}[\log D(x\mid y)] + \mathrm{E}_{z\sim p_z}[\log(1-D(G(z\mid y)))] \tag{3}$$

但是 GAN 和 CGAN 的损失函数为 JS（Jensen-Shannon）散度，该函数在训练过程中可能会出现模式崩溃及梯度消失等问题。为解决该问题，本文采用 Wasserstein 距离来代替 JS 散度，Wasserstein 距离可度量 2 个样本分布之间的距离，即

$$W(P_r,P_g) = \inf_{\gamma\sim\prod(P_r,P_g)} \mathrm{E}_{(x,y)\sim\gamma}[\|x-y\|] \tag{4}$$

其中，$P_r$ 和 $P_g$ 表示 2 个样本分布，$\prod(P_r,P_g)$ 表示 $P_r$ 和 $P_g$ 联合分布的集合。此时 GAN 被称为 WGAN[35]，其目标函数为

$$V(G,D) = \max_{D\in 1-\mathrm{Lipschitz}} \{\mathrm{E}_{x\sim P_r}[D(x)] - \mathrm{E}_{x\sim P_g}[D(x)]\} \tag{5}$$

其中，$D\in 1-\mathrm{Lipschitz}$ 为约束函数，表示满足 $\|f(x_1)-f(x_2)\|\leqslant\|x_1-x_2\|$，即输出的变化不大于输入的变化。

本文采用的 WGAN-GP[36]在 WGAN 的基础上添加一个梯度惩罚项强制满足 Lipschitz 约束，目标函数为

$$V(G,D) = \max_{D\in 1-\mathrm{Lipschitz}} \{\mathrm{E}_{x\sim P_r}[D(x)] - \mathrm{E}_{x\sim P_g}[D(x)] - \lambda \mathrm{E}_{x\sim P_{\mathrm{penaty}}}[\nabla_x D(x)-1]^2\} \tag{6}$$

其中，$x\sim P_{\mathrm{penaty}}$ 表示 $x$ 为 $P_r$ 与 $P_g$ 分布上两点连线上的点，$\lambda$ 为所设参数。

### 2.3 特征选择

特征选择技术作为网络入侵检测中重要组成部分，能为后续模型的分类识别做出重要贡献。因此本文提出一种基于特征统计重要性的过滤式特征选择技术，从标准差、均值中值差两方面筛选特征，如图 4 所示。


图 4 所提特征选择技术

标准差是一种统计度量工具，用来反映数据集的离散程度。在入侵检测数据集中，通过计算每列特征的标准差可以判断出其离散程度，标准差高表明该特征取值范围大，在所有样本上的差异更显著；而标准差低则说明该特征取值范围小，样本在该特征的差异不大可被舍去。同理，均值和中值同样是统计工具，用来描述数据分布，通过计算均值中值差的绝对值可反映特征值的偏差情况，值越大表示样本的类间差异越明显。因此，利用标准差、均值中值差来进行特征选择有助于提升攻击样本的检测准确率。

通过标准差、均值中值差对特征由大到小进行排序并对其进行排名；然后将每列特征所对应的 2 个排名值相加，并根据相加结果再次由大到小进行排序；最后根据排名值将特征从高到低依次输入模型，直到当前输入特征集精度不再优于前一个特征集精度，此时前一个特征集中的特征为所选择的特征，后续可由所选择的特征构建数据集对检测模型进行训练。

## 3 IDS 框架

### 3.1 所提模型

本文提出的 VAE-CWGAN 模型包含以下三部分：编码器、生成器及判别器，其中生成器也是 VAE 中的解码器，具体结构如图 5 所示。

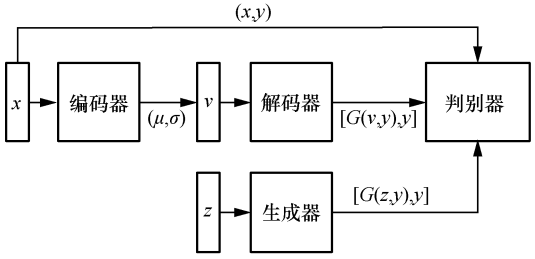



图 5 VAE-CWGAN 模型结构

图 5 中，$x$ 为真实数据，$y$ 为标签，$v$ 为隐变量，$z$ 为随机噪声，判别器分别对原始真实数据、重构数据以及生成数据进行判断以更新模型参数。利用该模型解决数据集类不平衡问题的具体

步骤如下。

**Step1** 通过编码器对原始数据进行编码得到隐变量。

**Step2** 利用解码器或生成器对带有少数类攻击标签的隐变量或噪声进行训练以生成新样本。

**Step3** 固定解码器和生成器，将所有样本输入判别器中进行训练。

**Step4** 固定判别器，通过判别器得到的误差对解码器和生成器进行训练。

**Step5** 重复上述步骤，直到判别器的损失值为 0.5，此时新生成数据与原始数据较接近。

### 3.2　整体架构

本文所提 IDS 框架包含以下四部分：数据预处理模块、数据生成模块、特征选择模块及攻击检测模块，如图 6 所示。

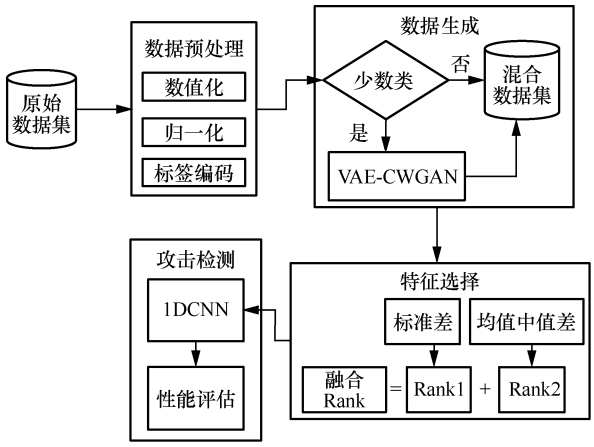



图 6　所提 IDS 框架

由图 6 可知，本文首先对原始数据进行预处理，其次通过 VAE-CWGAN 模型生成少数类攻击数据以构建混合数据集，然后融合统计特征重要性进行特征选择，最后使用 1DCNN 模块对经过特征选择的混合数据集进行检测并评估所提方法性能。

### 3.3　数据集说明

NSL-KDD 数据集通过删除大量冗余数据解决了长期困扰 KDD CUP99 数据集的问题，是入侵检测领域中最经典的数据集之一。NSL-KDD 数据集包含 4 个文件，分别为 KDDTrain+.txt、KDDTest+.txt、KDDTrain-21.txt 以及 KDDTest-21.txt。为了能够更好地对所提方法进行训练，本文选择 KDDTrain+.txt 和 KDDTest+.txt 作为模型的训练集和测试集，具体如表 2 所示。

**表 2　NSL-KDD 流量样本分布**

| 类别 | 训练样本数/个 | 测试样本数/个 |
|---|---|---|
| Normal | 67 343 | 9 711 |
| DoS | 45 927 | 7 460 |
| Probe | 11 656 | 2 421 |
| R2L | 995 | 2 885 |
| U2R | 52 | 67 |
| 合计 | 125 973 | 22 544 |

UNSW-NB15 数据集是通过 IXIA Perfect Storm 工具创建而成的，该数据集是正常流量和 9 类异常攻击流量的混合体，更符合当前的网络环境。同时该数据集的特征由 Argus 和 Bro-IDS 工具所提取，具体如表 3 所示。

**表 3　UNSW-NB15 流量样本分布**

| 类别 | 训练样本数/个 | 测试样本数/个 |
|---|---|---|
| Normal | 56 000 | 37 000 |
| DoS | 12 264 | 4 089 |
| Fuzzers | 18 184 | 6 062 |
| Analysis | 2 000 | 677 |
| Backdoor | 1 746 | 583 |
| Exploits | 33 393 | 11 132 |
| Generic | 40 000 | 18 871 |
| Reconnaissance | 10 491 | 3 496 |
| Shellcode | 1 133 | 378 |
| Worms | 130 | 44 |
| 合计 | 175 341 | 82 332 |

CIC-IDS-2017 数据集是通过嗅探实时网络数据包创建而成的，该数据集是当前最大和最新的入侵检测数据集之一，包含了大量的 DoS 攻击数据。同时该数据集的特征由 CICFlowMeter 工具所提取。受实验条件所限，本文只选取 CIC-IDS-2017 中的周三数据集进行实验，而且由于周三数据集中 Heartbleed 攻击样本数只有 11 个，太过于稀疏，因此将其删除且数据集中仍存在较为严重的类不平衡问题，具体如表 4 所示。

**表 4　CIC-IDS-2017 流量样本分布**

| 类别 | 训练样本数/个 | 测试样本数/个 |
|---|---|---|
| Benign | 351 719 | 87 964 |
| DoS Hulk | 183 998 | 46 126 |
| DoS GoldenEye | 8 307 | 1 986 |
| DoS Slowloris | 4 648 | 1 148 |
| DoS Slowhttptest | 4 444 | 1 055 |
| 合计 | 553 116 | 138 279 |

### 3.4 数据预处理

为提高数据集中数据质量，本文在数据集输入模型之前对数据进行预处理，以提升模型的收敛速度，具体步骤如下。

1) 数值化。在上述数据集中存在一些非数值特征，例如 NSL-KDD 数据集中的 protocol_type、flag 及 service，模型并不能对其进行学习，因此需要将这些特征数据转换成数值特征，本文通过标签编码函数将其转换为数字，同时对数据标签类别进行数值转换。

2) 归一化。数据集中某些特征维度的取值范围较广泛，对模型的学习贡献并不相同，因此为减小特征维度之间的差异，保证检测结果的准确性，本文采用 min-max 函数对每列特征进行归一化，将其取值范围限制在[0,1]，可保证数据的有效性。

## 4 实验与分析

本节将重点阐述所提方法在 3 个入侵检测数据集上进行的仿真实验，实验环境如下：64 位的 Windows11 操作系统、16 GB RAM 的 AMD Ryzen 7 6800H CPU 以及 NVIDIA GeForce RTX 3050Ti GPU，使用 Python3.8 进行编程。

### 4.1 评估指标

为了能更直观地体现所提方法的性能优势，本文使用准确率（Acc）、精确率（Pre）、召回率（Rec）、误报率（FAR）及 F1 分数 5 个指标对模型性能进行评估，上述指标计算式为

$$\mathrm{Acc}=\frac{\mathrm{TP}+\mathrm{TN}}{\mathrm{TP}+\mathrm{FP}+\mathrm{TN}+\mathrm{FN}} \tag{7}$$

$$\mathrm{Pre}=\frac{\mathrm{TP}}{\mathrm{TP}+\mathrm{FP}} \tag{8}$$

$$\mathrm{Rec}=\frac{\mathrm{TP}}{\mathrm{TP}+\mathrm{FN}} \tag{9}$$

$$\mathrm{FAR}=\frac{\mathrm{FP}}{\mathrm{FP}+\mathrm{TN}} \tag{10}$$

$$\mathrm{Fl}=\frac{2\mathrm{Pre}\times\mathrm{Rec}}{\mathrm{Pre}+\mathrm{Rec}} \tag{11}$$

其中，TP、TN、FP、FN 分别代表真阳性、真阴性、假阳性及假阴性，即正确分类的攻击样本数量、正确分类的正常样本数量、错误分类的攻击样本数量、错误分类的正常样本数量。

### 4.2 参数设置

为防止过拟合以及提高模型的泛化能力，本文在训练和测试过程中选用十折交叉验证，同时采用 Adam 作为优化器，交叉熵函数作为损失函数，选取 ReLU 作为激活函数防止模型不收敛及梯度消失等问题，具体如表 5 所示。

**表 5 模型参数设置**

| 参数 | 参数值 | |
|---|---|---|
| | VAE-CWGAN | 1DCNN |
| 轮 | 200 | 10 |
| 批大小 | 256 | 512 |
| 学习率 | 0.000 1 | 0.001 |
| 优化器 | Adam | Adam |
| 激活函数 | ReLU | ReLU |
| 卷积核大小 | 3×3 | 3×3 |

### 4.3 VAE-CWGAN 模型训练

为解决流量数据集中的类不平衡问题，本文搭建了基于 VAE-CWGAN 的混合模型对少数类进行样本生成，以改善分类模型过于关注多数类的缺点。以 NSL-KDD 数据集为例，对 Probe、R2L 以及 U2R 三类攻击进行样本生成，判别器在 NSL-KDD 数据集上训练的损失函数如图 7 所示。

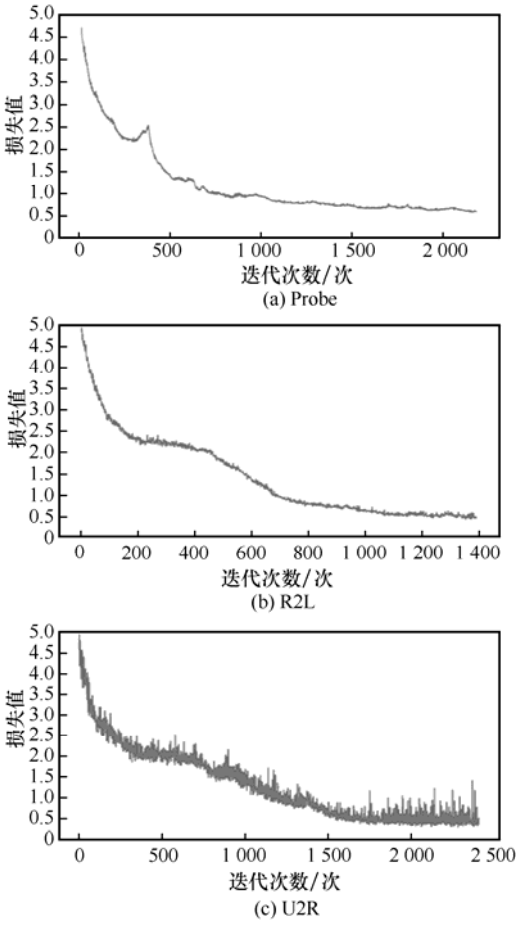



图 7 判别器在 NSL-KDD 数据集上训练的损失函数

由图 7 可知，所提类平衡模型在 NSL-KDD 数据集中的 3 个少数攻击类上经过多次迭代后均能达到纳什均衡，说明生成器生成的样本较理想可靠，和原始样本相似且难以区分。其中，Probe 和 R2L 的损失函数曲线较光滑，同时收敛速度较快，而 U2R 虽然也能收敛至稳定的理想状态附近，但其损失函数曲线存在一定程度的振荡，这是因为 U2R 在数据集中只有 52 个，使 VAE-CWGAN 模型较难学习到其内在分布，故曲线振荡难以收敛。

### 4.4　类平衡方法对比

本节基于 2 种实验背景来验证所提平衡方法的可行性：1) 二分类，用来判断样本属于正常类还是攻击类；2) 多分类，用来判断样本属于正常样本还是具体某一类攻击样本（NSL-KDD 有 5 类，UNSW-NB15 有 10 类，CIC-IDS-2017 有 15 类）。具体实验结果对比如下。

#### 4.4.1　二分类方法对比

为了验证模型所生成的样本对攻击检测的有效性，现将本文方法与传统的采样方法以及深度生成模型方法进行对比，具体结果如表 6 所示，其中加粗字体表示最佳性能。

如表 6 所示，本文方法在 3 个数据集上都取得了较好的性能，尤其是在精确率和召回率这对矛盾的指标上，本文方法实现了很好的兼顾。对于 NSL-KDD 数据集来说，本文方法在 5 个指标上均取得了最优的结果，其中，准确率、精确率、召回率以及 F1 分数均为 98.95%，误报率仅为 1.04%，皆远胜于其他方法，尤其是在召回率上相较于其他类平衡方法提升了 0.1%～22.4%。这主要是因为 ROS 只是简单地复制原始数据，容易造成过拟合问题。BorderlineSMOTE 是对 SMOTE 的改进，通过使用边界上的样本来克服类不平衡问题，不过其合成样本较相似，仍存在过拟合问题。ADASYN 也是基于 K 近邻产生数据，只是其对不同的少数类赋予不同的权重，但是易受异常值的影响。ENN 和 RENN 则是通过随机删除多数类中的样本以实现类平衡，不过某些重要的样本信息也会因此丢失。这些传统的采样方法缺乏对原始数据更深层次的学习，无法对含噪声的数据进行判别，故生成样本时产生了大量含噪声的数据，进而使模型的分类效果下降。

对于 UNSW-NB15 数据集而言，本文方法在各项性能指标上基本最优，在召回率上提升了 7.93%～14.10%，具备较好的检测性能。其原因在于传统的采样方法只是从样本局域邻域的角度出

**表 6**　　**本文方法与传统的采样方法以及深度生成模型方法对比结果**

| 数据集 | 方法 | Acc | Pre | Rec | F1 | FAR |
|---|---|---|---|---|---|---|
| NSL-KDD | ROS[37] | 82.17% | 90.65% | 76.55% | 83.07% | 9.95% |
| | BorderlineSMOTE[37] | 84.20% | 91.69% | 78.96% | 84.91% | 8.87% |
| | RENN[20] | 98.02% | 98.06% | 98.02% | 98.03% | — |
| | ADASYN[20] | 98.35% | 98.83% | 98.35% | 98.57% | — |
| | ENN[20] | 98.85% | 98.87% | 98.85% | 98.85% | — |
| | 本文方法 | **98.95%** | **98.95%** | **98.95%** | **98.95%** | **1.04%** |
| UNSW-NB15 | ROS[21] | 87.31% | **99.04%** | 82.14% | 89.80% | — |
| | SMOTE[21] | 87.22% | 98.23% | 82.70% | 89.80% | — |
| | ADASYN[21] | 90.47% | 97.44% | 88.31% | 92.65% | — |
| | WGAN[21] | 88.82% | 98.90% | 84.51% | 91.14% | — |
| | CGAN[21] | 88.92% | 98.64% | 84.89% | 91.25% | — |
| | 本文方法 | **96.24%** | 96.23% | **96.24%** | **96.24%** | **4.74%** |
| CIC-IDS-2017 | ROS[22] | **99.94%** | 99.74% | 99.97% | 99.85% | 0.07% |
| | SMOTE[22] | 99.93% | 99.69% | **99.98%** | 99.84% | 0.08% |
| | ADASYN[22] | 99.92% | 99.65% | 99.96% | 99.80% | 0.09% |
| | RUS+SMOTE[22] | 99.92% | 99.65% | 99.96% | 99.80% | 0.09% |
| | K-Means+SMOTE[22] | 99.93% | 99.69% | 99.97% | 99.83% | 0.08% |
| | 本文方法 | 99.92% | **99.92%** | 99.92% | **99.92%** | **0.07%** |

发，没有考虑样本整体分布。WGAN、CGAN 以及本文方法通过对原始样本的内在分布进行学习，生成的样本质量较高，但 WGAN 与 CGAN 未引入梯度惩罚项及类信息向量，梯度消失和模型易崩溃的问题仍存在，故两者的性能稍差。同时本文方法又引入 VAE 可提前学习到输入的空间分布，弥补了 GAN 可能生成无意义样本的缺点，同时 GAN 也改善了 VAE 多样性不强的弊端，因此在 UNSW-NB15 数据集上本文方法的性能最优。同理，在 CIC-IDS-2017 数据集上，虽然将传统的采样方法进行结合，避免了原始样本信息的丢失，但其本质上仍只是对样本的浅层学习，分类效果难以得到提升。而本文方法即使在某些指标上并没有达到最优，不过其与最优结果相差甚小，仅在百分位上存在差异，检测性能基本一致。本文方法在 3 个数据集的二分类混淆矩阵如图 8 所示。从图 8 可以看出，大部分样本都集中在矩阵的对角线上，说明模型整体性能较好，也验证了本文方法更适合解决类不平衡问题，可对网络攻击流量实现有效的检测。

### 4.4.2 多分类方法对比

面对不同的攻击方式，所采用的防御措施也不尽相同，而单纯的二分类检测并不能识别异常流量属于哪一类具体攻击，因此有必要进行多分类实验，从而对异常流量进行细分。为使本文方法更加客观合理，将本文方法与其他多分类平衡方法进行对比，结果如表 7 所示。

如表 7 所示，本文方法在 NSL-KDD 数据集上的准确率、精确率、召回率、F1 分数均为 98.91%，误报率仅为 1.07%，在 5 个指标上均实现了最优。相较于其他方法，本文方法在准确率上提升了

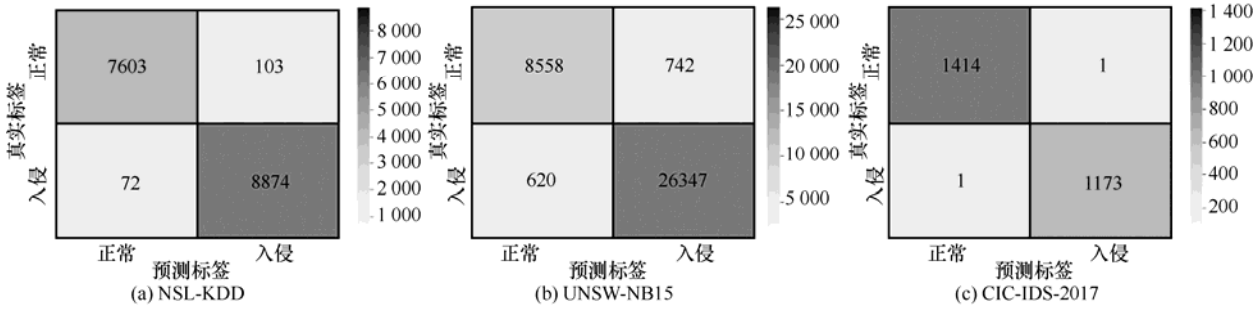



图 8 二分类混淆矩阵

**表 7 本文方法与其他多分类平衡方法对比结果**

| 数据集 | 方法 | Acc | Pre | Rec | F1 | FAR |
|---|---|---|---|---|---|---|
| NSL-KDD | ROS[25] | 78.26% | 92.34% | 67.41% | 77.93% | 7.39% |
| | SMOTE[25] | 81.16% | 96.42% | 69.48% | 80.76% | 3.41% |
| | ADASYN[25] | 80.10% | 96.16% | 67.74% | 79.49% | 3.57% |
| | CVAE[25] | 85.97% | 97.39% | 77.43% | 86.27% | 2.74% |
| | CWGAN[26] | 90.34% | 96.74% | 85.92% | 91.01% | 3.83% |
| | 本文方法 | **98.91%** | **98.91%** | **98.91%** | **98.91%** | **1.07%** |
| UNSW-NB15 | ROS[26] | 81.70% | 77.32% | 94.49% | 85.05% | 33.96% |
| | SMOTE[26] | 82.44% | 78.05% | **94.77%** | 85.60% | 32.65% |
| | ADASYN[26] | 82.15% | 77.76% | 94.65% | 85.38% | 33.16% |
| | WGAN[27] | 81.49% | 84.71% | 82.51% | 83.60% | — |
| | CWGAN[27] | 85.59% | 86.11% | 85.57% | 85.84% | — |
| | 本文方法 | **87.58%** | **89.22%** | 87.58% | **88.39%** | **8.85%** |
| CIC-IDS-2017 | ROS[22] | 99.76% | 99.81% | 99.76% | 99.77% | — |
| | SMOTE[22] | 99.76% | **99.83%** | 99.76% | 99.77% | — |
| | ADASYN[22] | 99.76% | **99.83%** | 99.76% | 99.77% | — |
| | RUS+SMOTE[38] | 99.72% | 99.81% | 99.72% | 99.75% | — |
| | K-Means+SMOTE[38] | 99.70% | 99.81% | 99.70% | 99.74% | — |
| | 本文方法 | **99.79%** | 99.79% | **99.79%** | **99.79%** | **0.19%** |

8.57%～20.65%，在精确率上提升了 1.52%～6.57%，在召回率上提升了 12.99%～31.50%，在 F1 分数上提升了 7.90%～20.98%，在误报率上降低了 1.67%～5.69%。对于 UNSW-NB15 数据集而言，本文方法在除召回率外的 4 个指标上都为最优，主要是因为在原始的 UNSW-NB15 数据集中 DOS 类和 Exploits 类、Normal 类与 Fuzzers 类存在部分样本特征完全一致的情况，这使模型无法对其进行判别，使召回率偏低。不过其余指标都得到了较大提升，尤其是将误报率大幅降低，这说明本文方法在对流量进行判别时较为稳定，可有效保护设备安全。对于 CIC-IDS-2017 数据集而言，本文方法在性能指标方面基本上都为最优，仅在精确率上相差 0.04%，而在准确率上最大提升了 0.09%、在召回率上最大提升了 0.09%、在 F1 分数上最大提升了 0.05%，这种幅度的提升对于接近 100% 的性能指标来说已殊为不易。本文方法在 3 个数据集上的多分类混淆矩阵如图 9 所示。

通过类平衡方法在 3 个数据集上的二分类和多分类结果对比，体现了本文所提类平衡方法的有效性，为解决数据集类不平衡问题提供了一条可行的思路。

### 4.5　特征选择方法对比

为了验证本文所提特征选择方法的优势，本节采用卡方检验法、随机森林、互信息法、递归消除法以及遗传算法这 5 种特征选择方法对本文数据集进行特征筛选，并在本文分类模型上开展仿真实验。本文方法与其他特征选择方法对比结果如表 8 所示。

由表 8 可知，与现有的典型特征选择方法相比，本文方法具有更好的检测性能。在 NSL-KDD 数据集上准确率提升了 0.67%～3.35%；在 UNSW-NB15 数据集上准确率提升了 2.05%～5.15%；在 CIC-IDS-2017 数据集上准确率提升了 0.34%～5.72%。精确率与召回率体现了模型对入侵问题的关联程度以及敏感程度，在 3 个数据集中本文方法在这 2 个指标上皆为最优，最大可提升 5%。同时，本文方法在兼顾精确率和召回率的度量指标 F1 分数上分别为 98.91%、88.39%以及 99.79%，相比于其他方法都取得了一定程度的提高，此外，本文方法误报率也大幅下降。主要原因在于卡方检验为独立性假设，而入侵数据集中存在复杂的非线性关系，故卡方检验筛所选出的特征可能只是局部相关，未能考虑到特征整体的相关性；同理，互信息法和递归消除法也忽略了特征之间的相关性；随机森林法为集成模型，在处理高维数据时可能会过拟合，导致所选择的特征泛化性不强；遗传算法容易陷入局部最优且难以处理多目标问题，所筛选特征可能不会取得令人满意的效果。本文所提出的启发式特征选择方法作为实现底层分类问题的有效合理媒介，具备一定的科学性和可行性[6]。其通过融合标准差和均值中值差的统计重要性来选择具备高辨识性和高类间差异性的特征，同时也考虑到了特征之间的整体相关性，因此可有效提升模型的检测性能。

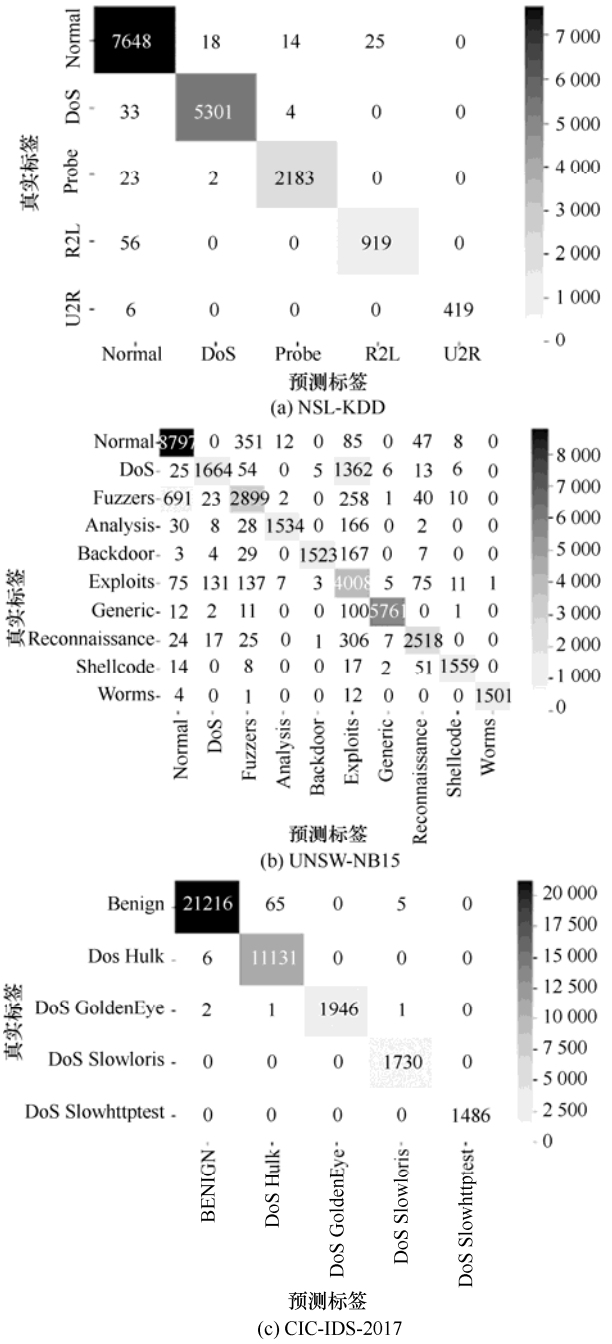



图 9　多分类混淆矩阵

表 8 本文方法与其他特征选择方法对比结果

| 数据集 | 方法 | Acc | Pre | Rec | F1 | FAR | Size |
|---|---|---|---|---|---|---|---|
| NSL-KDD | 卡方检验法 | 95.56% | 95.59% | 95.56% | 95.44% | 4.46% | 13% |
| | 随机森林法 | 97.84% | 97.86% | 97.84% | 97.81% | 1.62% | 17% |
| | 互信息法 | 96.49% | 96.54% | 96.49% | 96.50% | 3.96% | 15% |
| | 递归消除法 | 98.07% | 98.06% | 98.07% | 98.06% | 2.57% | 20% |
| | 遗传算法 | 98.24% | 98.24% | 98.24% | 98.24% | 2.54% | 24% |
| | 本文方法 | **98.91%** | **98.91%** | **98.91%** | **98.91%** | **1.07%** | 22% |
| UNSW-NB15 | 卡方检验法 | 82.43% | 85.20% | 82.43% | 82.52% | 11.11% | 13% |
| | 随机森林法 | 84.90% | 86.30% | 84.90% | 85.04% | 11.03% | 19% |
| | 互信息法 | 84.18% | 86.11% | 84.18% | 84.30% | 11.97% | 25% |
| | 递归消除法 | 85.51% | 87.35% | 85.51% | 85.68% | 10.78% | 21% |
| | 遗传算法 | 85.53% | 87.27% | 85.53% | 85.92% | 11.57% | 32% |
| | 本文方法 | **87.58%** | **89.22%** | **87.58%** | **88.39%** | **8.85%** | 21% |
| CIC-IDS-2017 | 卡方检验法 | 94.52% | 95.25% | 94.52% | 94.60% | 3.97% | 13% |
| | 随机森林法 | 99.30% | 99.31% | 99.30% | 99.31% | 0.48% | 36% |
| | 互信息法 | 99.31% | 99.31% | 99.31% | 99.31% | 0.56% | 35% |
| | 递归消除法 | 99.37% | 99.37% | 99.37% | 99.37% | 0.40% | 20% |
| | 遗传算法 | 99.45% | 99.46% | 99.45% | 99.46% | 0.40% | 38% |
| | 本文方法 | **99.79%** | **99.79%** | **99.79%** | **99.79%** | **0.19%** | 41% |

### 4.6 与现有二分类方法对比

为了检验本文所提 VAE-CWGAN 对异常流量的检测性能，将其与 3 种二分类方法进行比较，分别为 MFC[14]、Tad-GAN[39]以及 MSRC[19]，具体如表 9 所示。

表 9 本文方法与二分类方法对比结果

| 模型 | 数据集 | Pre | Rec | F1 |
|---|---|---|---|---|
| MFC | NSL-KDD | 94.61% | 95.06% | 94.83% |
| | UNSW-NB15 | 89.03% | 90.68% | 89.85% |
| Tad-GAN | NSL-KDD | 83.43% | 88.71% | 85.99% |
| | UNSW-NB15 | 86.81% | 89.91% | 88.33% |
| MSRC | NSL-KDD | 94.43% | 94.92% | 94.67% |
| | UNSW-NB15 | 90.12% | 91.79% | 90.95% |
| VAE-CWGAN | NSL-KDD | **98.95%** | **98.95%** | **98.95%** |
| | UNSW-NB15 | **96.23%** | **96.24%** | **96.24%** |

如表 9 所示，本文方法在 2 个数据集上都取得了最佳的检测性能，其次是 MSRC 和 MFC，两者性能相差不大，而 Tad-GAN 则表现最差，这主要是因为 Tad-GAN 虽然是双向结构，能更好地捕捉数据的内在分布，但其对数据的依赖性较强，而数据集中存在较多重复数据，因此其性能不如其他 3 种方法。而 MSRC 和 MFC 通过小波变换可提取样本的时频信息进行学习，兼顾了流量的原始信息和细节信息，不过这 2 种方法在进行小波变换时损失了部分数据信息，且未能解决数据集样本不平衡及所选特征差异不大等问题，而 UNSW-NB15 数据集又较为冗余复杂，因此这 2 种方法在 NSL-KDD 数据集上的指标比 Tad-GAN 高 11%左右，在 UNSW-NB15 数据集上只是略优于 Tad-GAN，而本文所提 VAE-CWGAN 不仅解决了类不平衡问题，而且所选特征差异性较高，故本文方法在各个指标上都远优于另外 3 种方法。综上，本文方法与其他二分类检测方法相比具备更强的稳定性和泛化性，有效提升了检测性能。

### 4.7 与现有多分类研究方法对比

通过上述类平衡方法与特征选择技术的对比实验可知，本文方法取得了不错的效果。为了进一步验证本文所提基于 VAE-CWGAN 和特征统计重要性融合的入侵检测方法的可行性，将本文方法与近几年所提方法进行比较，具体结果如表 10～表 12 所示。

**表 10　NSL-KDD 数据集上对比结果**

| 方法 | Acc | Pre | Rec | F1 |
|---|---|---|---|---|
| KD-TCNN[29] | 98.44% | 98.60% | 98.47% | 98.57% |
| CSK-CNN[22] | **99.14%** | 94.03% | 98.70% | 96.31% |
| ABC-BWO-CONV-LSTM[40] | 98.67% | 97.48% | **100%** | 98.73% |
| Bagging-GBM[41] | 91.57% | 98.67% | 86.18% | 91.50% |
| 本文方法 | 98.91% | **98.91%** | 98.91% | **98.91%** |

**表 11　UNSW-NB15 数据集上对比结果**

| 方法 | Acc | Pre | Rec | F1 |
|---|---|---|---|---|
| GMM-WGAN[42] | **87.70%** | 88.46% | 87.70% | 85.44% |
| MCNN-DFS[43] | 80.51% | 81.00% | 81.00% | 81.00% |
| CNN-BiLSTM[44] | 77.16% | 82.63% | 79.91% | 81.25% |
| FCWGAN-BiLSTM[27] | 85.59% | 86.11% | 85.57% | 85.84% |
| 本文方法 | 87.58% | **89.22%** | **87.58%** | **88.39%** |

**表 12　CIC-IDS-2017 数据集上对比结果**

| 方法 | Acc | Pre | Rec | F1 |
|---|---|---|---|---|
| TACGAN[21] | 95.86% | 96.85% | 94.79% | 95.81% |
| SVM-GAC[45] | **99.93%** | 98.73% | 99.61% | 98.91% |
| KD-TCNN[29] | 99.44% | 99.48% | 99.47% | 99.46% |
| ADASYN-RENN[36] | 99.65% | 99.63% | 99.65% | 99.64% |
| 本文方法 | 99.79% | **99.79%** | **99.79%** | **99.79%** |

由表 10 可知，本文方法在准确率上略低于 CSK-CNN，这是因为该方法在进行多分类之前已实现了二分类，避免了正常样本的干扰，不过本文方法在其他指标上优于 CSK-CNN，同理在召回率上仅低于 ABC-BWO-CONV-LSTM，不过其 100%的召回率在实际网络中难以实现。F1 分数作为最为重要的评价指标，兼顾了召回率和精确率，本文方法的 F1 分数为 98.91%，比 KD-TCNN、CSK-CNN、ABC-BWO-CONV-LSTM 及 Bagging-GBM 分别高了 0.34%、2.6%、0.18%和 7.41%，说明本文方法能有效检测到攻击流量并正确分类。如表 11 所示，在准确率方面，本文方法仅比 GMM-WGAN 低 0.12%，但比 MCNN-DFS、CNN-BiLSTM 和 FCWGAN-BiLSTM 分别高 7.07%、10.42%和 1.99%，而在 F1 分数上比这 4 种方法分别高了 2.95%、7.39%、7.14%和 2.55%。综上，虽然该数据集包含了大量的重复数据，但本文方法在每个指标上基本可获得最优的结果。由表 12 可知，各种方法都取得了较好的结果，这是因为 CIC-IDS-2017 数据集含有大量数据且较为简单，在这种情况下，本文方法在精确率上仅比 SVM-GAC 低 0.14%，但比 TACGAN、KD-TCNN 和 ADASYN-RENN 分别高了 3.93%、0.35%和 0.14%。而且在 F1 分数上比这 4 种方法分别高了 3.98%、0.88%、0.33%和 0.15%。通过与现有方法在 3 个数据集上的对比，说明了本文方法在检测攻击流量时的有效性，从而可在一定程度上保护网络设备，使其免受威胁。

## 5　结束语

本文提出了一种基于 VAE-CWGAN 和标准差、中值均值差的统计重要性融合的入侵检测方法。该方法不仅较好地解决了流量数据集类不平衡问题，而且所选择的特征子集也具有高辨别性和高类间差异性。采用 NSL-KDD、UNSW-NB15 及 CIC-IDS-2017 数据集对本文方法进行二分类和多分类性能评估，并从准确率、精确率、召回率、F1 分数和误报率 5 个指标上验证了本文方法的优势，同时在与其他方法进行对比时体现了本文方法具备较优的检测性能。不过本文仅在 3 个流量数据集上进行实验，未来将考虑从样本更为真实、场景更为广泛的流量数据集出发以提升模型的普适性，同时设计轻量级模型挖掘特征以提升检测效率。

## 参考文献：


[1] HE J X, WANG X D, SONG Y F, et al. Network intrusion detection based on conditional Wasserstein variational autoencoder with generative adversarial network and one-dimensional convolutional neural networks[J]. Applied Intelligence, 2023, 53(10): 12416-12436.

[2] 王伟. 基于深度学习的网络流量分类及异常检测方法研究[D]. 合肥: 中国科学技术大学, 2018.
WANG W. Deep learning for network traffic classification and anomaly detection[D]. Hefei: University of Science and Technology of China, 2018.

[3] PANDA M, PATRA M R. Network intrusion detection using naive bayes[J]. International Journal of Computer Science and Network Security, 2007, 7(12): 258-263.

[4] MEHEDI H M A, NASSER M, PAL B, et al. Support vector machine and random forest modeling for intrusion detection system (IDS)[J]. Journal of Intelligent Learning Systems and Applications, 2014, 6(1): 45-52.

[5] YAN J Q, JIN D, LEE C W, et al. A comparative study of off-line deep learning based network intrusion detection[C]//Proceedings of the 2018 Tenth International Conference on Ubiquitous and Future Networks (ICUFN). Piscataway: IEEE Press, 2018: 299-304.

[6] THAKKAR A, LOHIYA R. Fusion of statistical importance for feature selection in deep neural network-based intrusion detection system[J]. Information Fusion, 2023, 90: 353-363.

[7] DHANABAL L, SHANTHARAJAH S P. A study on NSL-KDD dataset for intrusion detection system based on classification algorithms[J]. International Journal of Advanced Research in Computer and Communication Engineering, 2015, 4(6): 446-452.

[8] MOUSTAFA N, SLAY J. UNSW-NB15: a comprehensive data set for network intrusion detection systems (UNSW-NB15 network data set)[C]//Proceedings of the 2015 Military Communications and Information Systems Conference (MilCIS). Piscataway: IEEE Press, 2015: 1-6.

[9] SHARAFALDIN I, HABIBI L A, GHORBANI A A. Toward generating a new intrusion detection dataset and intrusion traffic characterization[C]//Proceedings of the 4th International Conference on Information Systems Security and Privacy. Setúbal: SciTePress, 2018: 108-116.

[10] VIJAYANAND R, DEVARAJ D, KANNAPIRAN B. Support vector machine based intrusion detection system with reduced input features for advanced metering infrastructure of smart grid[C]//Proceedings of the 2017 4th International Conference on Advanced Computing and Communication Systems (ICACCS). Piscataway: IEEE Press, 2017: 1-7.

[11] TONG D, QU Y R, PRASANNA V K. Accelerating decision tree based traffic classification on FPGA and multicore platforms[J]. IEEE Transactions on Parallel and Distributed Systems, 2017, 28(11): 3046-3059.

[12] FARNAAZ N, JABBAR M A. Random forest modeling for network intrusion detection system[J]. Procedia Computer Science, 2016, 89: 213-217.

[13] KOC L, MAZZUCHI T A, SARKANI S. A network intrusion detection system based on a hidden naïve Bayes multiclass classifier[J]. Expert Systems with Applications, 2012, 39(18): 13492-13500.

[14] 段雪源, 付钰, 王坤, 等. 基于多尺度特征的网络流量异常检测方法[J]. 通信学报, 2022, 43(10): 65-76.
DUAN X Y, FU Y, WANG K, et al. Network traffic anomaly detection method based on multi-scale characteristic[J]. Journal on Communications, 2022, 43(10): 65-76.

[15] ALDARWBI M Y, LASHKARI A H, GHORBANI A A. The sound of intrusion: a novel network intrusion detection system[J]. Computers and Electrical Engineering, 2022, 104: 108455.

[16] 缪祥华, 单小撤. 基于密集连接卷积神经网络的入侵检测技术研究[J]. 电子与信息学报, 2020, 42(11): 2706-2712.
MIAO X H, SHAN X C. Research on intrusion detection technology based on densely connected convolutional neural networks[J]. Journal of Electronics & Information Technology, 2020, 42(11): 2706-2712.

[17] ALTHOBAITI M M, PRADEEP M K, GUPTA D, et al. An intelligent cognitive computing based intrusion detection for industrial cyber-physical systems[J]. Measurement, 2021, 186: 110145.

[18] THAKUR S, CHAKRABORTY A, DE R, et al. Intrusion detection in cyber-physical systems using a generic and domain specific deep autoencoder model[J]. Computers & Electrical Engineering, 2021, 91: 107044.

[19] DUAN X Y, FU Y, WANG K. Network traffic anomaly detection method based on multi-scale residual classifier[J]. Computer Communications, 2023, 198: 206-216.

[20] CAO B, LI C H, SONG Y F, et al. Network intrusion detection model based on CNN and GRU[J]. Applied Sciences, 2022, 12(9): 4184.

[21] DING H W, CHEN L Y, DONG L, et al. Imbalanced data classification: a KNN and generative adversarial networks-based hybrid approach for intrusion detection[J]. Future Generation Computer Systems, 2022, 131: 240-254.

[22] SONG J M, WANG X J, HE M S, et al. CSK-CNN: network intrusion detection model based on two-layer convolution neural network for handling imbalanced dataset[J]. Information, 2023, 14(2): 130.

[23] KINGMA D P, WELLING M. Auto-encoding variational Bayes[J]. arXiv Preprint, arXiv: 1312.6114, 2013.

[24] GOODFELLOW I J, POUGET-ABADIE J, MIRZA M, et al. Generative adversarial nets[C]//Proceedings of the 27th International Conference on Neural Information Processing Systems. Cambridge: MIT Press, 2014: 2672-2680.

[25] YANG Y Q, ZHENG K F, WU C H, et al. Improving the classification effectiveness of intrusion detection by using improved conditional variational autoencoder and deep neural network[J]. Sensors, 2019, 19(11): 2528.

[26] ZHANG G L, WANG X D, LI R, et al. Network intrusion detection based on conditional Wasserstein generative adversarial network and cost-sensitive stacked autoencoder[J]. IEEE Access, 2020, 8: 190431-190447.

[27] MA Z X, LI J, SONG Y F, et al. Network intrusion detection method based on FCWGAN and BiLSTM[J]. Computational Intelligence and Neuroscience, 2022, 2022: 6591140.

[28] EESA A S, ORMAN Z, BRIFCANI A M A. A novel feature-selection approach based on the cuttlefish optimization algorithm for intrusion detection systems[J]. Expert Systems with Applications, 2015, 42(5): 2670-2679.

[29] WANG Z D, LI Z Y, HE D J, et al. A lightweight approach for network intrusion detection in industrial cyber-physical systems based on knowledge distillation and deep metric learning[J]. Expert Systems with Applications, 2022, 206: 117671.

[30] KHAN N M, MADHAV C N, NEGI A, et al. Analysis on improving the performance of machine learning models using feature selection technique[C]//Proceedings of 18th International Conference on Intelligent Systems Design and Applications. Berlin: Springer, 2020: 69-77.

[31] CHANG Y P, LI W, YANG Z M. Network intrusion detection based on random forest and support vector machine[C]//Proceedings of the 2017 IEEE International Conference on Computational Science and Engineering (CSE) and IEEE International Conference on Embedded and Ubiquitous Computing (EUC). Piscataway: IEEE Press, 2017: 635-638.

[32] KUMAR V, SINHA D, DAS A K, et al. An integrated rule based intrusion detection system: analysis on UNSW-NB15 data set and the real time online dataset[J]. Cluster Computing, 2020, 23(2): 1397-1418.

[33] AMBUSAIDI M A, HE X J, NANDA P, et al. Building an intrusion detection system using a filter-based feature selection algorithm[J]. IEEE Transactions on Computers, 2016, 65(10): 2986-2998.

[34] 段雪源, 付钰, 王坤. 基于 VAE-WGAN 的多维时间序列异常检测方法[J]. 通信学报, 2022, 43(3): 1-13.
DUAN X Y, FU Y, WANG K. Multi-dimensional time series anomaly detection method based on VAE-WGAN[J]. Journal on Communications, 2022, 43(3): 1-13.

[35] ARJOVSKY M, CHINTALA S, BOTTOU L. Wasserstein generative adversarial networks[C]//Proceedings of the 34th International Conference on Machine Learning. New York: ACM Press, 2017: 214-223.

[36] GULRAJANI I, AHMED F, ARJOVSKY M, et al. Improved training

of Wasserstein GANs[J]. arXiv Preprint, arXiv: 1704.00028, 2017.

[37] 尹梓诺, 马海龙, 胡涛. 基于联合注意力机制和一维卷积神经网络-双向长短期记忆网络模型的流量异常检测方法[J]. 电子与信息学报, 2023, 45(10): 3719-3728.
YIN Z N, MA H L, HU T. A traffic anomaly detection method based on the joint model of attention mechanism and one-dimensional convolutional neural network-bidirectional long short term memory[J]. Journal of Electronics & Information Technology, 2023, 45(10): 3719-3728.

[38] ZHANG H P, HUANG L L, WU C Q, et al. An effective convolutional neural network based on SMOTE and Gaussian mixture model for intrusion detection in imbalanced dataset[J]. Computer Networks, 2020, 177: 107315.

[39] GEIGER A, LIU D Y, ALNEGHEIMISH S, et al. TadGAN: time series anomaly detection using generative adversarial networks[C]//Proceedings of the 2020 IEEE International Conference on Big Data (Big Data). Piscataway: IEEE Press, 2020: 33-43.

[40] KANNA P R, SANTHI P. Hybrid intrusion detection using MapReduce based black widow optimized convolutional long short-term memory neural networks[J]. Expert Systems with Applications, 2022, 194: 116545.

[41] LOUK M H L, TAMA B A. Dual-IDS: a bagging-based gradient boosting decision tree model for network anomaly intrusion detection system[J]. Expert Systems with Applications, 2023, 213: 119030.

[42] CUI J Y, ZONG L S, XIE J H, et al. A novel multi-module integrated intrusion detection system for high-dimensional imbalanced data[J]. Applied Intelligence, 2023, 53(1): 272-288.

[43] AL-TURAIKI I, ALTWAIJRY N. A convolutional neural network for improved anomaly-based network intrusion detection[J]. Big Data, 2021, 9(3): 233-252.

[44] JIANG K Y, WANG W Y, WANG A L, et al. Network intrusion detection combined hybrid sampling with deep hierarchical network[J]. IEEE Access, 2020, 8: 32464-32476.

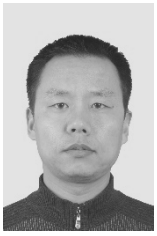
[45] SHUKLA A K. Detection of anomaly intrusion utilizing self-adaptive grasshopper optimization algorithm[J]. Neural Computing and Applications, 2021, 33(13): 7541-7561.

[作者简介]

**刘涛涛**（1996− ），男，江西吉安人，海军工程大学博士生，主要研究方向为人工智能、信息处理、网络安全。

**付钰**（1982− ），女，湖北武汉人，博士，海军工程大学教授、博士生导师，主要研究方向为信息安全、人工智能。

**王坤**（1981− ），女，河南信阳人，海军工程大学博士生，主要研究方向为信息安全、人工智能。



**段雪源**（1981− ），男，河南开封人，海军工程大学博士生，主要研究方向为人工智能、信息处理、网络安全。