引用格式：高新成, 朱城枫. 基于 TCN 与 ChebyKAN 融合网络的恶意软件多分类模型 [J]. 电子科技大学学报, 2026, 55(2): 215-223.
GAO X C, ZHU C F. Malware multi-classification model for based on TCN and ChebyKAN fusion network[J]. Journal of University of Electronic Science and Technology of China, 2026, 55(2): 215-223.

基于 TCN 与 ChebyKAN 融合网络的恶意软件多分类模型

高新成[1, 2*]，朱城枫[2]

（1. 东北石油大学现代教育技术中心，大庆 163318；2. 东北石油大学计算机与信息技术学院，大庆 163318）

摘要：针对基于 API 调用序列的传统恶意软件检测方法存在长距离时序依赖捕捉不足、忽略特征间高阶非线性关系等相关问题，提出了一种融合时序卷积网络（TCN）与 Chebyshev-Kolmogorov-Arnold network（ChebyKAN）的恶意软件多分类模型（TCN-SE-ChebyKAN）。首先，基于 TCN 模块对 API 调用序列提取特征，利用因果卷积和膨胀卷积，突破传统循环神经网络在长距离时序依赖建模中的局限，精准捕捉恶意软件多阶段行为的时序关联；其次，引入 SE 模块构建通道注意力机制，动态优化通道权重，解决关键判别性特征被冗余信息掩盖的问题；最后，通过切比雪夫多项式改进 KAN 模块（ChebyKAN），利用其全局逼近特性增强特征间高阶非线性关系的建模能力，克服原始 KAN 中 B 样条函数局部性强的缺陷。实验结果表明，该模型在 Mal-API-2019 数据集上 AUC 值达 92.53%，精确率、召回率、F1 值等指标均有显著提升。

关键词：恶意软件；API 调用序列；SE 模块；ChebyKAN；时序卷积网络
中图分类号：TP309 **文献标志码：**A DOI: 10.12178/1001-0548.2025099

Malware multi-classification model for based on TCN and ChebyKAN fusion network

GAO Xincheng[1, 2*] and ZHU Chengfeng[2]

(1. Center for Modern Educational Technology, Northeast Petroleum University, Daqing 163318, China;
2. School of Computer and Information Technology, Northeast Petroleum University, Daqing 163318, China)

Abstract: The traditional malware detection methods based on API (application programming interface) call sequences fail to sufficiently capture the long-term temporal dependencies and neglect the high-order nonlinear relationships among features. To address these issues, this paper proposes a multi-classification model for malware (TCN-SE-ChebyKAN) that integrates a temporal convolutional network (TCN) and Chebyshev-Kolmogorov-Arnold network (ChebyKAN). First, the TCN module is employed to extract features from API call sequences. By leveraging causal convolutions and dilated convolutions, the model captures temporal characteristics and long-range dependencies representing malware behavior, thereby obtaining more comprehensive behavioral information. Next, a squeeze-and-excitation (SE) network module is introduced to construct a channel attention mechanism. Through dynamic adjustment of channel weights, the model enhances its ability to capture discriminative features. Finally, the KAN (Kolmogorov-Arnold network) module is utilized to model complex relationships among features. By improving the KAN module with Chebyshev polynomials, the model strengthens its capability to model high-order nonlinear relationships among features, boosting overall detection performance. Experimental results demonstrate that the proposed model achieves an AUC value of 92.53% on the Mal-API-2019 data set, with significant improvements in other detection metrics.

Key words: malicious software; API call sequence; SE module; ChebyKAN; temporal convolutional network

近年来，互联网技术飞速发展，恶意软件数量与种类巨增，给网络安全带来巨大挑战。在企业网络环境中，恶意软件通常借助伪装系统补丁、捆绑应用程序等隐蔽手段渗透至设备，利用系统安全漏洞实施敏感数据窃取、系统配置篡改等恶意行为，严重危害企业网络安全。一旦业务系统被恶意软件

收稿日期：2025−05−12
基金项目：国家自然科学基金（61702093）；中国高校产学研创新基金（2021ITA02011）；黑龙江省教育科学规划课题 （GJB1425352）
作者简介：高新成，博士，教授， 主要从事网络安全、网络管理和大数据应用等方面的研究。
*通信作者 E-mail: gxc@nepu.edu.cn

入侵，企业安全运营中心（SOC）的威胁狩猎工作将受到直接干扰，关键数据泄露或业务停摆等事件可能引发重大经济损失，并对企业声誉造成负面影响。

面对如此严峻的恶意软件威胁，检测技术的有效性至关重要。然而，恶意软件的静态分析方法在面对混淆代码和变种样本时，存在代码特征提取效果不佳的问题[1]。而动态分析方法能够通过监控API 调用序列有效捕获行为特征[2]，在应对复杂恶意软件样本时更具优势。所以，本文基于动态分析展开研究。

传统基于签名特征匹配和启发式规则的检测方法[3]，处理已知恶意软件时效率较高，但面对新型恶意软件，在检测准确率和误报率方面表现欠佳。随着机器学习与深度学习技术的发展，恶意软件检测领域迎来了革命性变革。尤其是深度学习凭借强大的特征自动提取能力与复杂模式建模优势，近年来在恶意软件检测领域得到广泛应用。文献 [2] 通过构建包含卷积和递归网络层的模型架构，深度挖掘 API 调用序列中的特征信息，成功实现对恶意软件的精准分类。文献 [4] 不仅发布了具有重要研究价值的 Mal-API-2019 数据集，还采用单层 LSTM 模型进行检测任务，在该数据集上验证了模型的有效性，为后续相关研究提供了标准数据支撑。文献 [5] 利用 LSTM 和 GRU 两种循环神经网络模型进行实验，充分发挥二者在处理时序数据方面的优势，有效捕捉 API 调用序列中的时间依赖关系，在恶意软件检测任务中取得了良好效果。文献 [6] 提出的 RTF 模型，创新性地应用于恶意软件多分类任务，凭借其独特的网络结构和训练策略，实现了高效的多类别区分，进一步拓展了恶意软件检测模型的应用边界。此外，文献 [7] 结合 Word2Vec 词嵌入技术与 1D-CNN 网络，通过将 API 调用序列转化为语义向量，并利用卷积神经网络的局部特征提取能力，精准捕捉 API 调用前后的依赖关系，显著提升了检测模型的性能。文献 [8] 提出 API 2 Vec++，基于图嵌入和 BERT 预训练技术，将 API 调用关系建模为图结构，并借助预训练模型强大的语义理解能力，实现恶意软件的准确多分类。文献 [9] 提出的 CAFTrans 框架，重点关注特征相关性与多模态融合，通过引入注意力机制和多源数据融合策略，为恶意软件检测领域提供了全新的研究思路与技术方向。文献 [10] 提出 TuneDroid 方法，通过代码图像化与贝叶斯优化动态调整 CNN 模型，检测效果显著优于传统模型。

然而，当前基于 API 调用序列的动态分析方法存在计算复杂度高和忽略长距离依赖关系等不足。针对这些问题，本文重点研究 TCN 网络和 ChebyKAN 网络，并将它们有机整合，提出基于 TCN 与 ChebyKAN 融合网络的恶意软件多分类模型（TCN-SE-ChebyKAN）。利用 TCN 的因果卷积和膨胀卷积建模长距离时序依赖关系，引入 SE 注意力机制融合提取多通道特征，并通过 ChebyKAN 的切比雪夫多项式变换进行高阶非线性组合，以此替换传统的全连接层，为恶意软件检测提供有效的解决方案。

1 TCN-SE-ChebyKAN 多分类模型

本文重点围绕恶意软件行为特征的动态性与复杂性展开研究，设计 TCN-SE-ChebyKAN 多分类模型，主要包括数据预处理和恶意软件分类两大模块，整体处理流程如图 1 所示。

数据预处理模块：1）对原始数据集开展清洗、去重操作，将 API 序列长度设置为 150，保证数据的质量和一致性，再借助 SeqGAN 平衡数据分布，为后续分析筑牢数据根基；2）进行词嵌入表征。运用 Word2Vec 对 API 调用序列进行语义向量表征，嵌入层维度设置为 100 维，把离散的符号转换为富含语义的数值表达，以便模型更好地理解和处理数据；3）处理好的数据输入到恶意软件分类模块，进行模型训练与分类。

TCN-SE-ChebyKAN 分类模块：1）利用 TCN 强大的长距离依赖建模能力，深度挖掘 API 调用序列中的时间依赖关系。在恶意软件的行为模式里，API 调用存在复杂的时间顺序和依赖关系。如包含文件加密、网络数据传输、权限提升等操作的 API 调用序列，TCN 可精准捕捉这些关键信息；2）引入 SE 模块，自适应地校准特征通道的权重，增强关键特征的显著性表达，让模型更聚焦于对分类任务有判别性的特征；3）ChebyKAN 网络利用多项式激活函数完成非线性特征降维与分类决策，对从 TCN 和 SE 模块提取的高维复杂特征进行有效降维，并映射到 8 个不同的恶意软件类别空间。通过这些操作，模型能够有效提取 API 调用序列中高维且复杂的深层次特征，精准揭示恶意软件的内在行为模式，完成恶意软件多分类任务。

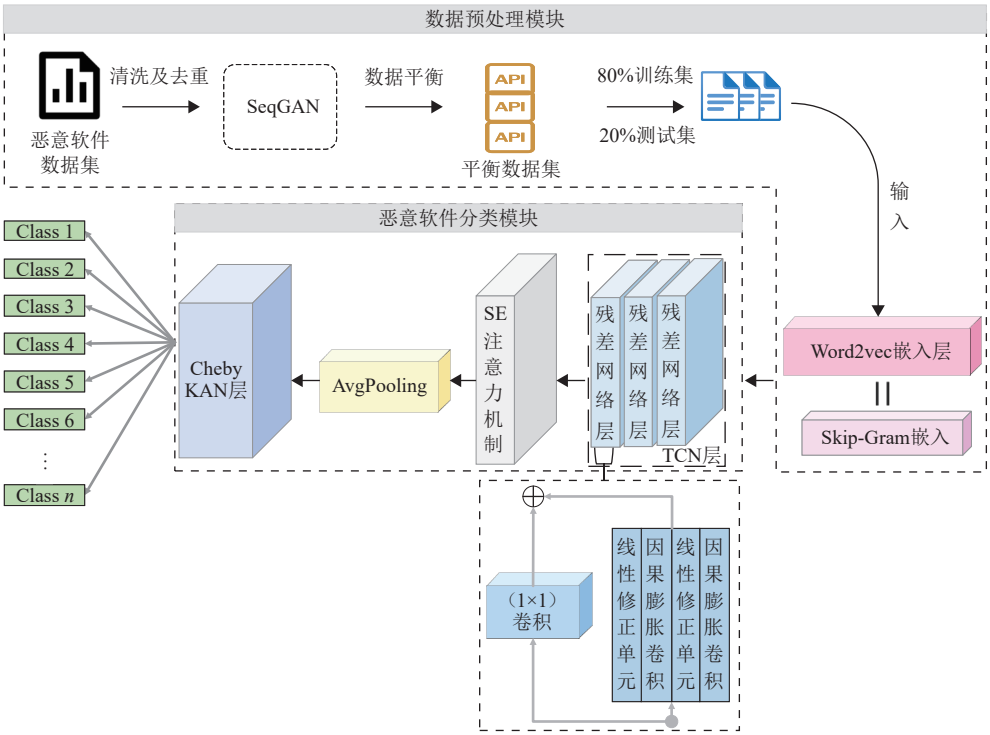


图 1　基于 TCN-SE-ChebyKAN 的恶意软件分类流程图

1.1　词嵌入表征

在恶意软件行为分析领域，API 调用序列非结构化、离散化的特性，极大地增加了语义信息提取的难度。嵌入表征作为自然语言处理领域中的一种关键技术，是文本处理的基础环节，通过将非结构化的文本数据转化为结构化的向量形式[11]，为后续自然语言处理任务提供有效输入，其核心思想是利用训练机制构建词语与实数向量的映射关系，使向量空间能够反映词语的语义关联特性。针对传统方法将 API 调用作为孤立符号处理，导致难以捕捉调用间的语义关联与行为逻辑、无法有效表征序列深层信息的问题，本研究采用 Word2vec 算法中的 Skip-Gram 方法对 API 调用序列进行词嵌入表征。通过将每个 API 调用视为单词，利用 Skip-Gram 模型以目标 API 调用预测上下文的方式[12]，在无监督学习过程中，将 API 调用序列转化为对应的稠密向量。这种处理方式能够有效捕捉 API 调用间的强关联的上下文信息，精准挖掘其语义特征，实现对 API 函数的高效编码。考虑到 API 调用序列非结构化、离散化的特性，设置嵌入维度为 100 维，既能为其语义关联与行为逻辑提供充足的表征空间，有效捕捉调用间的深层语义与上下文关联，又能在语义表征能力与计算复杂度间达成平衡，避免低维度信息压缩不足或高维度参数冗余的问题，为后续模型训练提供可靠的数据基础。模型结构如图 2 所示。

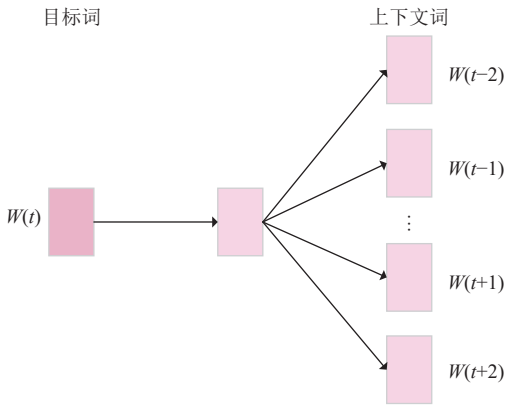


图 2　Skip-Gram 模型

1.2　时序卷积网络

API 调用序列作为反映恶意软件行为的核心数据，其蕴含的时序依赖关系与动态语义特征呈现复杂非线性分布。针对传统时序模型在处理 API 调用序列时存在的梯度传播不稳定、长距离依赖捕捉能力弱、模型记忆范围难以控制等问题，本研究引入时序卷积网络（temporal convolutional networks, TCN）对 API 调用序列进行深层特征提取。

通过构建因果膨胀卷积（dilated causal convolution）层，确保模型在学习过程中仅利用历史时刻的 API 调用信息推断当前行为模式，避免未

来数据的非因果干扰；同时借助膨胀卷积的间隔采样机制，在不显著增加计算复杂度的前提下扩大模型感受野[13]，使其能够识别不同时间段 API 调用间的协作关系，准确把握调用顺序的变化规律。残差模块的引入则为梯度传播开辟额外的路径，在保障网络梯度稳定传导的同时，支撑模型对 API 调用序列中复杂时序特征与语义关联的层级化特征建模。作为一种时序卷积网络模型，TCN 能够有效提取序列数据的深层特征，同时保留较大的感受野，以此更好地控制模型的记忆范围，进而提升时序序列分类的准确性。TCN 因果膨胀卷积和网络结构图分别如图 3、图 4 所示。

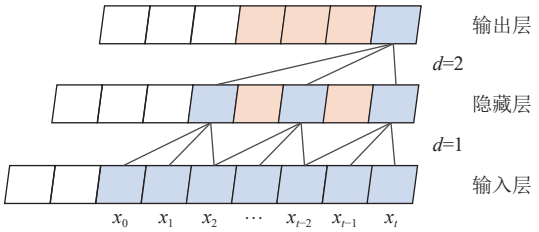


图 3 因果膨胀卷积架构

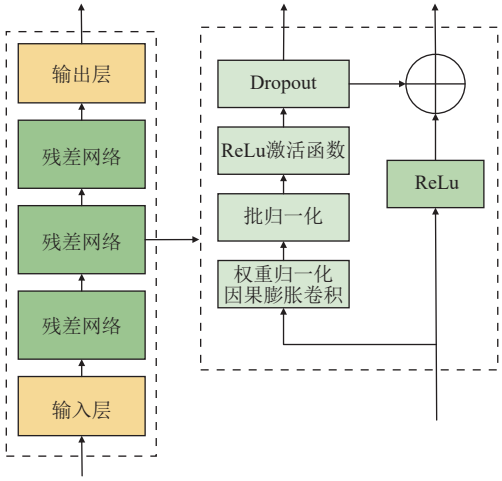


图 4 TCN 时序卷积网络架构

1.2.1 因果卷积

针对传统卷积神经网络在处理时序数据时，存在过去信息泄露从而干扰未来预测的问题，TCN 引入因果卷积确保时序数据的因果关系。因果卷积采用单向结构，通过限制模型只依赖当前和过去的时序步信息，隔离未来数据，避免信息泄露。这种机制为序列数据处理提供了有效方法，能更准确地捕捉数据的时间依赖关系

因果卷积通过下一层 t 时刻的值和之前 $x_1x_2\cdots x_t$ 的值来计算上一层 t 时刻的值 y_t，使得 y_t 接近于实际值，计算方法为：

$$P(x)=\prod_{t=1}^{T}P(x_t|x_1,x_2\cdots,x_{t-1}) \tag{1}$$

式中，$P(x)$ 表示函数的概率密度函数；T 表示延迟的时序。

1.2.2 膨胀卷积

TCN 通过引入膨胀卷积，逐步增大感受野，确保能够捕捉时序数据中的长距离依赖信息，同时克服传统神经网络需要采用逐层叠加多个卷积层的方式增强时序序列建模能力的局限性，其与普通卷积在输入数据的采样方式上存在明显差异。普通卷积在操作时，会将每个输入数据纳入计算，而膨胀卷积在计算中采用间隔式采样机制，通过设定膨胀系数来灵活调整采样间隔，从而实现跳跃采样。感受野的计算公式为：

$$\mathrm{RF}=(K-1)d+1 \tag{2}$$

式中，RF 表示感受野大小；K 为卷积核大小；d 为空洞系数。TCN 增大感受野的方式主要有两种：一是增大膨胀系数，二是选择更大的卷积核尺寸。在膨胀卷积中，膨胀系数按指数增长，从而显著扩展每一层的感受野范围。通过这种方式，膨胀卷积能够让每一层覆盖更广的范围，以较少的网络层数实现更大的感受野。随着膨胀系数的增加，感受野能够跨越更大范围，从局部到全局捕捉和提取更深层次的信息，从而提升模型的表达能力[14]。此外，这种设计使网络更加轻量化，计算效率更高。本文模型中因果膨胀卷积共设置 3 层，每一层的空洞数 d 分别设为 1、2、4，旨在以更深的模型层数和更多的空洞数来获取更大的感受野。

1.2.3 残差模块

在 TCN 里，残差模块把模型所提取到的特征与初始特征进行结合，防止在特征提取过程中遗失关键信息，从而增强模型的稳定性以及对数据的表征能力。残差模块中每个模块包含两个膨胀卷积层和权值归一化层，以 ReLU 作为激活函数实现非线性变换[15]。在正则化设计中，TCN 在膨胀卷积操作后在各残差块中引入 dropout 层，通过随机失活神经元的机制抑制过拟合，提升模型泛化能力。针对输入与输出特征在通道维度可能存在的差异，采用 1×1 卷积层对通道数进行适配，确保残差连接时特征张量的维度一致性，从而保障跨层信息融合的有效性。残差不仅有效缓解梯度消失和梯度爆炸问题，还能加速深层网络的训练收敛过程，提升模型性能。残差模块的具体结构如图 5 所示。

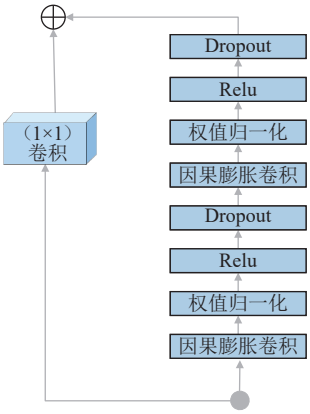


图 5　残差模块结构

1.3　压缩和激励模块

针对 API 调用序列经过 TCN 提取时序特征后，判别性特征易受冗余通道噪声干扰的问题，本文引入压缩和激励模块 SE（squeeze-and-excitation）[16]，该模块通过全局平均池化压缩通道全局特征，利用两层全连接网络学习通道间的非线性依赖关系。最终通过 Sigmoid 函数生成通道注意力权重矩阵，对 TCN 输出的多尺度特征进行通道特征重加权。该机制会增强与恶意行为强相关的 API 调用模式，同时抑制非关键通道的干扰，有效提升模型对判别性特征的捕捉能力。SE 模块结构如图 6 所示。

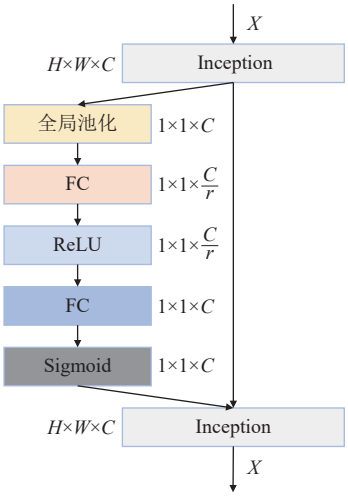


图 6　SE 模块结构图

1.3.1　全局特征压缩

首先对 TCN 输出的特征图$F \in \boldsymbol{R}^{L\times C}$进行全局平均池化（global average pooling），将每个通道的空间维度（序列长度 L）压缩为通道描述符向量$z \in \boldsymbol{R}^{C}$：

$$z_c = \frac{1}{HW}\sum_{i=1}^{H}\sum_{j=1}^{W} F(i,c) \quad c = 1,2,\cdots,C \tag{3}$$

式中，$F(i,c)$表示特征图在第i个时间步、第c个通道的特征值；z_c为第c个通道的全局统计特征。

1.3.2　通道依赖建模

通过两层全连接网络（FC）对通道描述符$\boldsymbol{z}$进行非线性变换，学习通道间的依赖关系，计算公式为：

$$\boldsymbol{f}_1 = \mathrm{ReLU}(\boldsymbol{W}_1 \cdot \boldsymbol{z} + \boldsymbol{b}_1) \tag{4}$$

$$\boldsymbol{s} = \sigma(\boldsymbol{W}_2 \cdot \boldsymbol{f}_1 + \boldsymbol{b}_2) \tag{5}$$

式中，式 (4) 表示第一层全连接层降维，$\boldsymbol{W}_1 \in \boldsymbol{R}^{\frac{C}{r}\times C}$，$r$ 为降维比例（本研究取 r=16）；$\boldsymbol{b}_1$为偏置项。式 (5) 表示第二层全连接层升维，$\boldsymbol{W}_2 \in \boldsymbol{R}^{C\times\frac{C}{r}}$；$\boldsymbol{b}_2$为偏置项；$\sigma$为 Sigmoid 激活函数。最终生成通道注意力权重向量$\boldsymbol{s} \in \boldsymbol{R}^{C}$。

1.3.3　特征通道重加权

将通道注意力权重$\boldsymbol{s}$与原始特征图 F 进行逐通道相乘，得到加权后的特征图:

$$F'(i,c) = s_c F(i,c) \tag{6}$$

式中，s_c表示第c个通道的重要性权重。该操作增强了与恶意行为强相关的通道特征（如文件操作、网络通信相关 API 的特征通道），抑制了冗余通道的干扰。

1.4　Chebyshev-KAN（ChebyKAN）

经 TCN 和 SE 模块处理后的数据，仍存在复杂模式挖掘不足与训练过程中收敛稳定性欠佳等问题。鉴于此，本文引入 ChebyKAN 网络对其进行优化，解决深层模式挖掘的局限性，增强训练过程的收敛稳健性。

1.4.1　Kolmogorov-Arnold 网络

传统 KAN（Kolmogorov-Arnold network）[17]是一种新型神经网络架构，受 Kolmogorov-Arnold 表示定理的启发，通过重新定义权重参数与激活函数的关系，打破了传统多层感知机（multi-layer perceptron, MLP）的设计框架。KAN 将可学习的激活函数放置在网络边缘，用自适应的单变量样条函数替代传统的权重参数。KAN 的网络结构如图 7 所示。

KAN 网络的核心主要由外部函数和内部函数共同构成。外部函数负责捕获全局特征，而内部函数则通过样条函数的形式对局部细节进行精细拟合。KAN 网络一般形式可表示为：

$$\mathrm{KAN}(x)=f(x_1,x_2,\cdots,x_n)=\sum_{q=1}^{2n+1}\Phi_q\left(\sum_{p=1}^{n}\phi_{q,p}(x_p)\right) \quad (7)$$

式中，$\phi_{q,p}$是内部单变量函数，定义域为 [0,1]，值域为实数集 $\boldsymbol{R}$，负责处理输入 x 的第 p 个分量 x_p，第 q 个外部函数的求和提供输入项；Φ_q是外部函数，对内部函数$\phi_{p,q}$的输出求和结果做映射，定义域和值域均为 $\boldsymbol{R}$。

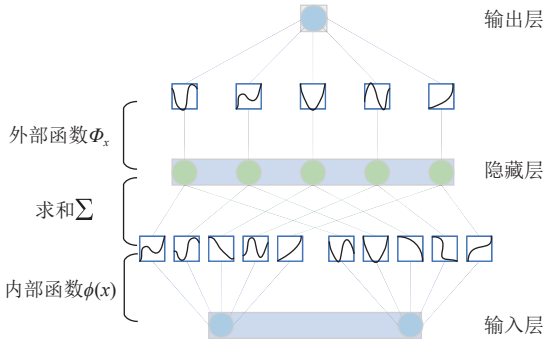


图 7 KAN 网络结构图

式 (7) 内部函数形成了一个 n 输入和（$2n$+1）输出的 KAN 层，输出函数形成了一个（$2n$+1）输入和 n 输出的 KAN 层，因此 KAN(x) 本质上由两个 KAN 层组成。为使 KAN 更易于优化，在 KAN 原文中提出了一种残差激活策略，通过引入基函数$b(x)$，使可训练激活函数$\phi(x)$是基函数$b(x)$和样条函数spline(x)的和[18]：

$$\phi(x)=w_b b(x)+w_s\mathrm{spline}(x) \quad (8)$$

式中，w_b和w_s实质上是冗余的，它可以被吸收到$b(x)$和样条函数spline(x)中。

$b(x)$ 设置为 SiLU 函数：

$$b(x)=\mathrm{SiLU}(x)=\frac{x}{1+\mathrm{e}^{-x}} \quad (9)$$

spline(x) 参数化为 B 样条曲线的线性组合：

$$\mathrm{spline}(x)=\sum_{i=i}^{n}c_i B_i(x) \quad (10)$$

式中，c_i表示可训练的参数；spline(x)是一组一维函数组合；$B_i(x)$是预先定义的 Base-样条基函数。在训练过程中样条参数c_i持续优化，以调整样条形状，从而拟合训练数据。

1.4.2 基于 Chebyshev 多项式的改进设计

传统 KAN 采用 B 样条函数作为基函数，其局部紧支撑特性（仅在小区间内非零）导致对全局模式的捕捉能力不足。ChebyKAN 通过替换为 Chebyshev 多项式，利用其全局正交性（在区间 [−1,1] 全域非零且满足正交性）[19]，相比于 B 样条函数，能使用少量系数即可表示全局模式，实现对长距离依赖特征的直接拟合：

$$\mathrm{Cheby}(x)=\sum_{k=0}^{m}a_k T_k(x) \quad (11)$$

式中，a_k表示可训练的参数；T_k是 k 阶 Chebyshv 多项式；m 是多项式阶数。

在误差上界推导公式中，设目标函数为 $f(x)$，ChebyKAN 与 KAN 的逼近误差满足式 (12) 和式 (13)。

ChebyKAN 误差：

$$\|f-C_m\|_{L^2}\leqslant C_1 m^{-r} \quad (12)$$

KAN 误差:

$$\|f-S_n\|_{L^2}\leqslant C_2 n^{-r/2} \quad (13)$$

式中，m 为多项式阶数；n 为节点数；r 为函数光滑度；C_1、C_2为常数，由于 B 样条需要更多参数拟合全局模式，当参数规模 m=n 时，ChebyKAN 的误差衰减速率 ($O(m^{-r})$)显著优于 B-SplineKAN $\left(O\left(n^{-r/2}\right)\right)$，当 r>0 时，即：

$$\frac{\|f-C_m\|_{L^2}}{\|f-S_n\|_{L^2}}\leqslant\frac{C_1}{C_2}n^{-r/2}<1 \quad (14)$$

相较于传统 KAN，ChebyKAN 通过全局正交基函数的设计，解决了传统 KAN 在复杂模式挖掘中的局限性。理论推导表明，其全局逼近误差上界随参数规模的增大衰减速率更快，且在收敛稳健性上具有显著优势，波动更小，有效降低了陷入局部最优解的风险，在复杂任务中表现出更优异的性能。

2 实验与分析

2.1 数据集及处理

Mal-API-2019[4] 是恶意软件动态行为分析领域广泛认可的基准数据集，其通过沙箱环境完整捕获 Windows 恶意软件 API 调用序列，以动态行为日志的形式精确表征恶意代码的运行时特征。基于上述优势，本研究采用该数据集进行实验分析。数据采用标准化 CSV 文件格式存储，共包含 7 107 条恶意软件样本，数据集中标注的恶意软件分为 8 类，具体分布如表 1 所示。

在模型对数据样本进行训练前，需要对数据集进行预处理操作。首先，依据空格分隔符从数据中提取 API 调用序列；其次，对所有 API 名称进行

标准化处理，统一转换为小写形式并移除特殊字符，以消除格式不一致带来的噪声干扰；最后，在严格保持 API 调用顺序的前提下，对每条序列执行去重操作，确保每个 API 在单条序列中仅出现一次。其中，API 调用序列最大长度不超过 150，因此设置长度为 150，既能完整覆盖所有样本的有效行为序列，又能通过统一长度保证数据输入的一致性，从而有效减少冗余信息对后续模型训练的负面影响。

表 1　Mal-API-2019 数据集

类型	数量
Worms	1 001
Virus	1 001
Trojans	1 001
Downloaders	1 001
Backdoors	1 001
Droppers	891
Spyware	832
Adware	379

由于数据集中各类别样本数量差异显著，直接用于模型训练会导致模型向多数类样本偏移。参考文献 [20] 使用 CycleGAN 解决不平衡数据的思路，本研究引入 SeqGAN（序列生成对抗网络）[21]，通过强化学习生成语义合理的 API 序列，扩充少数类样本，完成数据集的类别平衡操作。

经 SeqGAN 平衡后的数据集，各分类样本数量均统一为 1 001 条。随后，将数据集按 8∶2 的比例划分训练集和测试集，为后续模型训练与性能评估奠定基础。

2.2　性能评价指标

本文采用精确率（Precision）、召回率（Recall）、F1 值（F1-score）和 AUC（曲线下面积）值 4 个指标进行评价，这些评价指标已在相关研究中得到广泛应用。

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}} \tag{15}$$

$$\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}} \tag{16}$$

$$\text{F1} = 2\frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \tag{17}$$

$$\text{AUC} = \frac{\sum_{\text{ins}_i \in \text{positiveclass}} \text{rank}_{\text{ins}_i} - \dfrac{M(M+1)}{2}}{MN} \tag{18}$$

式中，TP 表示真阳性；FN 表示假阴性；FP 表示假阳性；TN 表示真阴性；$\text{rank}_{\text{ins}_i}$ 表示第 i 条样本的序号（概率得分从小到大排序，第 rank 个位置）；M,N 分别表示正样本的个数和负样本的个数；$\sum_{\text{ins}_i \in \text{positiveclass}}$ 为正样本序号的相加。

2.3　超参数对比实验

为充分考量 SE 压缩率 r 与 ChebyKAN 的阶数 m 对模型性能的影响，本节通过控制变量法进行实验。对 r 进行实验时，将阶数 m 统一设置为 4，对阶数 m 进行实验时，将压缩率 r 统一设置为 16。SE 模块压缩率 r 与 Chebyshev 阶数 m 的对比结果分别如图 8、图 9 所示。

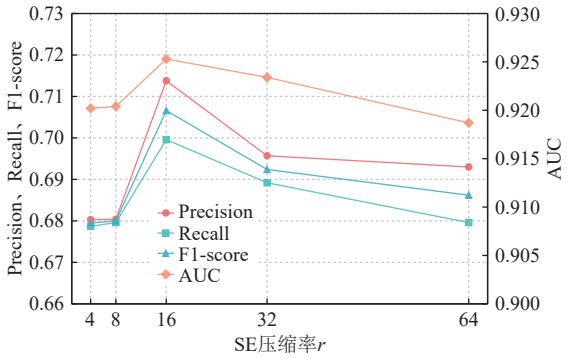


图 8　SE 中不同压缩率 r 的对比结果图

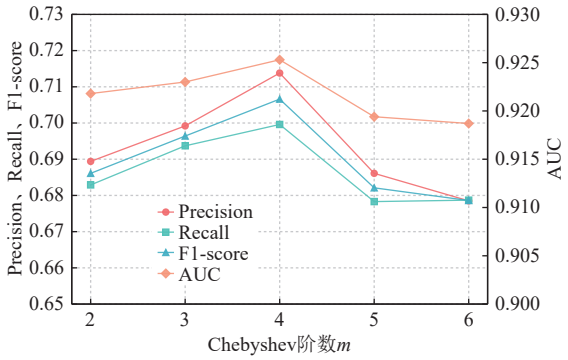


图 9　Chebyshev 中不同阶数 m 的对比结果

SE 模块的压缩率 r 与 ChebyKAN 中 Chebyshev 阶数 m 是影响模型性能的关键超参数。r 直接调控通道注意力机制效能，r 过小时，通道维度压缩不足，易引发过拟合且无法筛选关键特征；r 过大则过度压缩信息，导致注意力权重计算失效。实验表明，r=16 时通道信息提炼与压缩平衡，模型性能最佳。m 决定基函数多项式复杂度与特征映射能力，m 过小难以挖掘高阶语义信息，过大则增加计算负担、易过拟合。实验数据显示，m 对模型性能影响呈先升后降趋势，m=4 时基函数复杂度适配特征学习需求，Precision、Recall 等指标达最优，二

者共同为后续消融实验奠定参数基准。

2.4 消融实验

为了验证 SE 模块和 ChebyKAN 分别对本文模型的影响，本节进行了消融实验。在实验过程中记录了准确率、F1 分数和 AUC 分数等指标，实验结果如表 2 所示。

表 2 消融实验

模型	Precision	Recall	F1-score	AUC
TCN	0.662 8	0.659 6	0.658 1	0.903 7
TCN+SE	0.672 5	0.675 4	0.672 2	0.909 9
TCN+ChebyKAN	0.680 3	0.680 4	0.678 7	0.914 4
TCN+SE+KAN	0.678 3	0.678 3	0.675 5	0.913 2
TCN+SE+ChebyKAN	0.713 8	0.699 6	**0.706 6**	**0.925 3**

实验结果表明，SE 模块的引入使模型性能得到显著提升，基于通道注意力机制，有效增强了关键特征的提取，精确率从基线模型 TCN 的 0.662 8 提升至 0.672 5，提高了 0.97%。ChebyKAN 模块通过 Chebyshev 多项式实现更高效的非线性变换，相较于传统 KAN 架构，在各项指标上均表现更优，单独引入后精确率提升至 0.680 3。

值得注意的是，当 TCN、SE 和 ChebyKAN 这 3 个模块组合时，模型在多项评估指标上均达到最优表现，精确率达到 0.713 8，相比基线模型提升 5.10%；召回率为 0.699 6，F1 分数为 0.706 6，AUC 值达 0.925 3，这充分验证了各模块之间显著的协同增效作用，证明了本文模型设计的有效性和合理性。

2.5 模型对比实验

将本文模型与近几年在恶意软件分类领域中表现良好的几种模型进行对比实验。这些方法均使用本文中经过平衡的数据集且在同一实验环境下进行实验，相关指标如表 3 所示。

表 3 模型的分类结果

模型	Precision	Recall	F1-score	AUC
LSTM[5]	0.614 8	0.612 4	0.613 5	0.861 2
GRU[5]	0.620 5	0.618 8	0.619 6	0.873 6
1D-CNN[7]	0.661 3	0.656 7	0.655 8	0.903 0
RTF Model[6]	0.641 2	0.638 6	0.639 9	0.887 5
Extratrees[22]	0.634 6	0.634 2	0.634 5	0.879 3
CAFTrans[9]	0.672 7	0.675 3	0.674 0	0.910 2
TCN+SE+ChebyKAN	**0.713 8**	**0.699 6**	0.706 6	0.925 3

实验结果表明，TCN+SE+ChebyKAN 模型在 Precision 和 Recall 上均有显著提升。相较于 LSTM、GRU 和 1D-CNN 这 3 类基础模型，该模型的精确率分别提升 9.90%、9.33% 和 5.25%，召回率分别提升 8.72%、8.08% 和 4.29%。

在与前沿模型对比中，在精确率方面，较 RTF 模型提升 7.26%，较 Extratrees 提升 7.92%，较 CAFTrans 模型提升 4.11%。在召回率方面，较 RTF 模型提升 6.10%，较 Extratrees 方法提升 6.54%，较 CAFTrans 模型提升 2.43%。这一系统性优势进一步证明了新型模块组合在处理复杂数据时的优势。

已有研究表明，在多分类场景下，传统 Precision、Recall 等指标难以全面衡量模型性能[23]。因此，本研究选取 AUC 和 F1 分数作为核心评估指标。采用这两个指标既契合多分类任务特性，也便于与同类研究进行横向对比，确保评估结果的科学性与客观性[6]。具体实验结果如图 10 所示。在关键指标 F1 分数与 AUC 上，TCN+SE+ChebyKAN 模型同样表现突出。相比近期表现优异的 CAFTrans 模型，F1 分数从 67.40% 提升至 69.25%，AUC 值由 91.02% 增至 92.53%，分别实现 2.74% 和 1.66% 的增长，表明该模型在综合性能与泛化能力上更具竞争力，有效验证了模块协同设计的有效性。

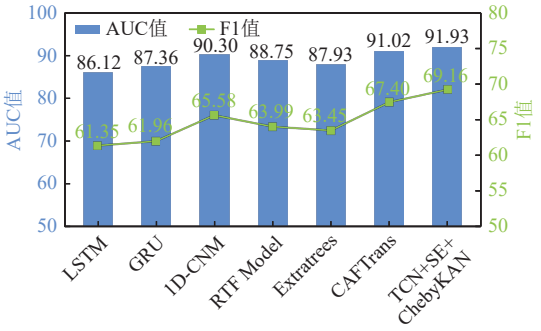


图 10 对比线-柱图

3 结束语

本文提出的融合时序卷积网络（TCN）与 ChebyKAN 的恶意软件多分类模型（TCN-SE-ChebyKAN）有效解决了恶意软件多分类任务中，对长距离时序依赖捕捉不足、忽略特征间高阶非线性关系等问题。首先，该模型通过 TCN 的因果卷积与膨胀卷积，精准提取 API 调用序列中的时序特征，有效建模长距离依赖关系，全面获取恶意软件行为信息；其次，融合 SE 模块，实现对关键通道的动态提取，显著增强特征的判别能力，进一步丰富特征表达；然后，借助 ChebyKAN 模块的切比雪夫多项式变换，对特征间的高阶非线性关系进行有效建模，突破了传统神经网络在表达能力上的局

限。实验结果表明，在基于 API 序列的恶意软件多分类任务中，本文模型的各项指标显著高于其他模型，充分验证了该模型在特征处理方面的有效性，为恶意软件多分类问题提供了可靠的解决方案。

尽管本研究在特征建模与分类性能上取得一定突破，但由于模型融合了 TCN 的因果膨胀卷积、SE 模块的通道注意力机制及 ChebyKAN 的高阶多项式变换等复杂结构，相比 GRU、LSTM 等轻量化模型，在时间性能上存在一定劣势，未来仍存在优化空间。后续研究可基于现有模型架构，构建多维度特征协同机制，通过解析特征与恶意行为的动态关联性，同时探索模型压缩与计算效率优化策略，在保持性能优势的前提下平衡时间开销，进一步提升对未知威胁的检测泛化能力。

参 考 文 献

[1] ZELINKA I, AMER E. An ensemble-based malware detection model using minimum feature set[J]. Mendel, 2019, 25(2): 1-10.

[2] KOLOSNJAJI B, ZARRAS A, WEBSTER G, et al. Deep learning for classification of malware system call sequences[C]//AI 2016: Advances in Artificial Intelligence. Cham: Springer International Publishing, 2016: 137-149.

[3] YE Y F, LI T, ADJEROH D, et al. A survey on malware detection using data mining techniques[J]. ACM Computing Surveys, 2018, 50(3): 1-40.

[4] CATAK F O, YAZı A F, ELEZAJ O, et al. Deep learning based Sequential model for malware analysis using Windows exe API Calls[J]. PeerJ Computer Science, 2020, 6: e285.

[5] LI C, ZHENG J J. API call-based malware classification using recurrent neural networks[J]. Journal of Cyber Security and Mobility, 2021, 10(3): 617-640.

[6] DEMIRKıRAN F, AYKUT Ç, UĞUR Ü, et al. An ensemble of pre-trained transformer models for imbalanced multiclass malware classification[J]. Computers & Security, 2022, 121: 102846.

[7] PANDA B, BISOYI S S, PANIGRAHY S. An ensemble approach for imbalanced multiclass malware classification using 1D-CNN[J]. PeerJ Computer Science, 2023, 9: e1677.

[8] CUI L, YIN J N, CUI J C, et al. API2Vec: Boosting API sequence representation for malware detection and classification[J]. IEEE Transactions on Software Engineering, 2024, 50(8): 2142-2162.

[9] QIAN L P, CONG L. Channel features and API frequency-based transformer model for malware identification[J]. Sensors, 2024, 24(2): 580.

[10] BAKıR H. A new method for tuning the CNN pre-trained models as a feature extractor for malware detection[J]. Pattern Analysis and Applications, 2025, 28(1): 26.

[11] 曾骏, 王子威, 于扬, 等. 自然语言处理领域中的词嵌入方法综述[J]. 计算机科学与探索, 2024, 18(1): 24-43.
ZENG J, WANG Z W, YU Y, et al. A survey of word embedding methods in natural language processing[J]. Computer Science and Exploration, 2024, 18(1): 24-43.

[12] 高新成, 张海洋, 朱城枫. 基于图层级异构图注意力网络的 JavaScript 恶意代码检测[J/OL]. 吉林大学学报（工学版）, [2025-05-05]. https://link.cnki.net/doi/10.13229/j.cnki.jdxbgxb.20240962.
GAO X C, ZHANG H Y, ZHU C F. JavaScript malicious code detection based on layer-level heterogeneous graph attention network[J/OL]. Journal of Jilin University (Engineering and Technology Edition), [2025-05-05]. https://link.cnki.net/doi/10.13229/j.cnki.jdxbgxb.20240962.

[13] FAN Y Y, LI C J, YI Q, et al. Classification of field moving targets based on improved TCN network[J]. Computer Engineering, 2021, 47: 106-112.

[14] 李思聪, 王坚, 宋亚飞, 等. 基于 BiTCN-DLP 的恶意代码分类方法[J]. 信息网络安全, 2023, 23(11): 104-117.
LI S C, WANG J, SONG Y F, et al. Malicious code classification method based on BiTCN-DLP[J]. Cyberspace Security, 2023, 23(11): 104-117.

[15] 郭玥秀, 杨伟, 刘琦, 等. 残差网络研究综述[J]. 计算机应用研究, 2020, 37(5): 1292-1297.
GUO Y X, YANG W, LIU Q, et al. A survey of residual network research[J]. Application Research of Computers, 2020, 37(5): 1292-1297.

[16] HU J, SHEN L, SUN G. Squeeze-and-excitation networks[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. New York: IEEE, 2018: 7132-7141.

[17] LIU Z, WANG Y, VAIDYA S, et al. Kan: Kolmogorov-ar-nold networks[EB/OL]. [2024-11-10]. https://arxiv.org/pdf/2404.19756.

[18] 袁立宁, 冯文刚, 刘钊. 基于 Kolmogorov-Arnold 网络的节点分类算法[J]. 计算机科学与探索, 2025, 19(3): 645-656.
YUAN L N, FENG W G, LIU Z. Node classification based on kolmogorov-arnold networks[J]. Journal of Frontiers of Computer Science & Technology, 2025, 19(3): 645-656.

[19] 韩梦如, 郭子磊, 赵运晓, 等. 基于 Chebyshev 点的 B 样条配置法在奇异摄动两点边值问题中的应用研究[J]. 应用数学进展, 2025, 14(5): 262-273.
HAN M R, GUO Z L, ZHAO Y X, et al. Study on application of Chebyshev B-spline collocation method on singularly perturbed two-point boundary value problems[J]. Advances in Applied Mathematics, 2025, 14(5): 262-273.

[20] TEKEREK A. A novel architecture for web-based attack detection using convolutional neural network[J]. Computers & Security, 2021, 100: 102096.

[21] YU L T, ZHANG W N, WANG J, et al. SeqGAN: Sequence generative adversarial nets with policy gradient[J]. Proceedings of the AAAI Conference on Artificial Intelligence, 2017, 31(1): 493.

[22] CANNARILE A, CARRERA F, GALANTUCCI S, et al. A study on malware detection and classification using the analysis of API calls sequences through shallow learning and recurrent neural networks[C]//Proceedings of the ITASEC'22: Italian Conference on Cybersecurity. Rome: [s.n.], 2022.

[23] BRANCO P, TORGO L, RIBEIRO R P. Relevance-based evaluation metrics for multi-class imbalanced domains[M]//Advances in Knowledge Discovery and Data Mining. Cham: Springer International Publishing, 2017: 698-710.

责任编辑　叶　芳