방산안보연구
Journal of Defense Industrial Security

방산안보연구, 2권 1호, 2026년 3월, 59-80면.
Journal of Defense Industrial Security, Volume 2, Issue 1, March 2026, Pages 59-80.
https://doi.org/10.64227/jdis.2026.2.1.4

Print ISSN 3091-5546
Online ISSN 3091-5554

원저

공급망 악성코드 위협 대응을 위한 NIST 800-161 기반 보안통제 설계와 딥러닝 탐지 모델의 정량적 평가

권창호[1], 신삼범[2], 전서인[3]

[1]제2작전사령부 서기관
[2]명지대학교 대학원 보안경영공학과 교수
[3]제2작전사령부 부이사관

교신저자: 신삼범 (ktma3419@hanmail.net)

요약

최근 공급망 공격은 공급자 및 다단계 하위 공급망으로 확산되며, 소프트웨어 구성요소를 통한 악성코드 유입이 핵심 위험으로 부각되고 있다. 본 연구는 NIST SP 800-161 Rev.1의 C-SCRM 프로세스를 기반으로 공급망 자산(코드·빌드·업데이트·서드파티 라이브러리 등)에 대한 악성코드 위험을 정량화할 수 있는 통제 연계 구조를 제시하고, 정적 바이너리 표현을 활용한 딥러닝 탐지 모델(이미지 CNN, 시퀀스 1D CNN-GLU)을 설계·평가하였다. 악성 580개/정상 580개 데이터셋에서 Stratified 5-fold 교차검증을 수행한 결과, 1D CNN-GLU는 Accuracy 0.9526±0.0143로 이미지 CNN(0.8647±0.0089) 대비 우수한 성능과 안정성을 보였다. 또한 탐지 결과를 FARM 단계의 의사결정 산출물(위험점수, 통제 적용/예외 승인 기록 등)과 연계하는 계층형(시그니처+딥러닝) 운영모델을 제안한다. 다만 소규모 시나리오 기반 데이터에 한정되므로, 일반화된 절대 성능보다는 모델 간 상대 비교와 운영 연계 가능성에 초점을 둔다. 95% 신뢰구간은 교차검증 fold 결과에 대한 부트스트랩 재표본추출로 산출하였다.

핵심어

공급망 보안, C-SCRM, NIST SP 800-161, 악성코드 탐지, 딥러닝

차례

1. 서론
2. 공급망 악성코드 위협 및 관련 연구 동향
3. NIST 800-161 기반 공급망 사이버보안 위험모델 및 관리요소 도출
4. 공급망 관점 딥러닝 기반 악성코드 탐지 모델 설계
5. 실험 방법
6. 실험 결과
7. 논의/시사점
8. 결론

Open Access

접수일: 2026년 1월 6일
수정일: 2026년 1월 24일
게재승인일: 2026년 1월 31일
출판일: 2026년 3월 31일

Original Article

NIST SP 800-161-Aligned Security Control Design and Quantitative Evaluation of a Deep-Learning-Based Malware Detection Model for Supply-Chain Threats

Changho Kweon[1], Sambum Shin[2], Seoin Jeon[3]

[1]2nd Operations Command, Secretary
[2]Professor, Dept. of Security Management Engineering, Graduate School, Myongji University
[3]2nd Operations Command, Senior Deputy Director

Corresponding Author: Sambum Shin (ktma3419@hanmail.net)

ABSTRACT

Recent supply-chain attacks increasingly propagate across suppliers and multi-tier subcontracting networks, making malware infiltration through software components a critical risk. This study, grounded in the NIST SP 800-161 Rev.1 C-SCRM process, presents a control-linked structure to quantify malware risk across supply-chain assets and designs and evaluates deep-learning-based detection models using static binary representations. Using a dataset of 580 malicious and 580 benign binaries, we conducted stratified 5-fold cross-validation. The 1D CNN-GLU achieved higher and more stable performance than the image-based CNN, with an accuracy of 0.9526 ± 0.0143 versus 0.8647 ± 0.0089. We further propose a layered operational model that links detection outcomes to FARM-stage decision artifacts. Because the evaluation is limited to a small, scenario-based dataset, the study emphasizes relative model comparisons and operational linkage rather than fully generalizable absolute performance. The 95% confidence intervals were estimated via bootstrap resampling of cross-validation fold results.

KEYWORDS

Supply chain security, C-SCRM, NIST SP 800-161, Malware detection, Deep learning

1. 서론

최근 공급망 공격(Supply Chain Attack)의 빈도와 정교함에 따른 위험성이 증가함에 따라, 보안 위협의 범위는 조직 내부 자산을 넘어 외부 공급자, 서비스 제공자, 그리고 다단계 하위 공급망으로 확산되고 있다. SolarWinds 공급망 공격 사건과 Log4j 취약점 사태는 소프트웨어 및 서비스 공급망에서 발생한 취약점이 전 세계 다수 조직으로 전파될 수 있음을 보여주며, 공급망 보안이 더 이상 선택이 아닌 핵심 보안 과제임을 명확히 입증하였다.

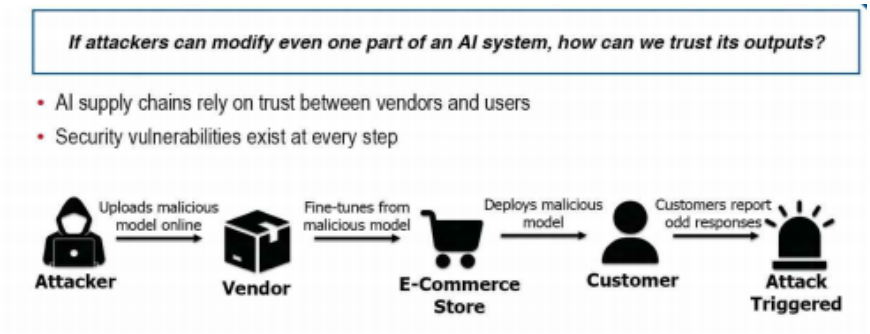


그림 1. 공급망 공격의 위험성

그림 1에서 보여주듯이 이러한 위협 환경 변화에 대응하여 미국 국립표준기술연구소(National Institute of Standards and Technology, NIST)는 조직의 공급망 보안 리스크 관리(Cybersecurity Supply Chain Risk Management, C-SCRM)를 체계적으로 지원하기 위해 SP 800-161 Rev.1을 발간하고, 공급망 전주기에 걸친 관리·통제 방안을 제시한다. 그러나 기존 연구와 표준은 주로 조직 차원의 관리적·절차적 통제에 초점을 두고 있어, 공급망 경로를 통해 유입되는 악성코드 위협을 정량적으로 분석하고 첨단 분석 기법을 C-SCRM 관점에서 통합적으로 활용하는 방안에 대해서는 상대적으로 논의가 부족하다. 본 연구는 NIST SP 800-161 Rev.1 기반 C-SCRM 맥락에서 통제 매핑을 정리하고, 이미지-CNN과 시퀀스-1D CNN-GLU의 탐지 성능을 비교하여 운영 프레임에 통합하는 방법을 제시한다.[1]

1.1. 연구 방법

본 연구는 문헌·표준 분석을 통해 공급망 악성코드 위험요소를 도출하고, 이를 NIST SP 800-161 Rev.1의 FARM 프로세스 산출물과 매핑한 뒤, 정적 바이트 표현 기반 딥러닝 모델 2종의 성능 지표를 산출하여, 탐지 결과가 통제 의사결정에 반영되는 운영 절차를 제안한다.

첫째, 국내·외 문헌조사와 표준 분석을 통해 공급망 공격 및 소프트웨어 공급망 보안과 관련된 선행연구, NIST SP 800-161 Rev.1에서 제시하는 공급망 보안 리스크 관리체계, 그리고 딥러닝 기반 악성코드 탐지 연구 동향을 종합적으로 검토하였다. 이를 통해 공급망 내 악성코드 관련 위험요소와 기존 연구의 개선점을 도출하였다.[2,3]

둘째, NIST SP 800-161 Rev.1의 통제 항목과 공급망 공격 사례 분석을 바탕으로, 악성코드 관련 사이버보안 위험을 구조화하고 분류 체계를 설계하였다. 이를 통해 공급망 보안 리스크 관리 관점에서 악성코드 위협을 설명할 수 있는 표준 기반 위험모델을 제안한다.

셋째, 도출된 위험 요소와 공격 시나리오를 반영하여 딥러닝 기반 악성코드 탐지 모델을 설계하였다. 공개 악성코드·정상코드 데이터셋 및 공급망 공격과 관련된 샘플을 수집·전처리하고, 특징

추출 방법과 딥러닝 모델 구조를 정의한 후, 학습·검증·테스트 절차를 통해 탐지 성능을 측정하였다. 이 과정에서 기존 일반 탐지 모델과의 비교를 통해 공급망 맥락을 반영한 모델의 성능 특성과 차별성을 분석하였다.

넷째, 제안한 위험모델과 딥러닝 탐지 모델의 실험 결과를 통합하여, 공급망 보안 리스크 관리 체계 내에서 악성코드 탐지 기능을 어떻게 배치·운영할 수 있는지에 관한 통합 프레임워크를 설계하였다. 마지막으로, 위험도 감소 및 통제 수준 향상 등 정량 지표를 활용하여 제안 프레임워크의 효과와 실무적 시사점을 도출하였다.

이와 같이 본 연구는 문헌연구 및 표준 분석에 기반한 방법과, 딥러닝 모델 구축·실험을 통한 정량적 분석 방법을 병행하는 혼합 연구 방법을 사용하였다.

1.2. 논문 구성

본 논문은 공급망 보안 리스크 관리와 딥러닝 기반 악성코드 탐지 모델을 통합적으로 분석하기 위해 다음과 같은 구조로 구성된다.

제1장에서는 연구의 배경과 필요성을 제시하고, 공급망 공격 환경 변화에 따른 문제의식을 바탕으로 연구 목적과 연구 질문을 정의한다. 또한 연구의 범위, 방법론, 그리고 본 연구의 학문적·실무적 기여를 설명한다.

제2장에서는 공급망 공격 및 소프트웨어 공급망 보안에 대한 이론적 배경과 선행연구를 검토한다. 특히 NIST SP 800-161 Rev.1 기반 C-SCRM 체계와 관련 표준 및 기존 연구의 한계를 분석하여 본 연구의 차별성을 도출한다.

제3장에서는 NIST SP 800-161 Rev.1을 기반으로 공급망 사이버보안 위험을 구조화하고, 악성코드 유입 관점에서의 위험요소를 정의한다. 이를 바탕으로 공급망 단계별·위협 유형별 위험 분류 체계와 위험모델을 설계하고, 기존 프레임워크와의 비교를 통해 제안 모델의 특징을 분석한다.

제4장에서는 제안된 위험모델을 기반으로 공급망 환경에 적용 가능한 딥러닝 기반 악성코드 탐지 모델을 설계한다. 데이터셋 구축, 전처리 및 특징 추출 방법, 모델 구조(CNN, 1D CNN-GLU), 학습 및 검증 절차를 구체적으로 기술한다.

제5장에서는 실험 환경 및 평가 방법을 제시한다. 데이터 구성, 학습 설정, 평가 지표(Accuracy, Precision, Recall, F1 등), 교차검증 방식 및 통계적 분석 방법을 설명한다.

제6장에서는 제안한 탐지 모델의 성능을 정량적으로 평가한다. 모델 간 성능 비교, 오탐 및 미탐 분석, 공급망 공격 시나리오 기반 탐지 효과를 분석하고, 운영 관점에서의 적용 가능성을 논의한다.

제7장에서는 NIST SP 800-161 Rev.1 기반 C-SCRM 프로세스와 딥러닝 탐지 모델을 통합한 공급망 보안 리스크 관리 프레임워크를 제안한다. 특히 FARM 프로세스와의 연계를 통해 탐지 결과를 위험평가 및 의사결정 산출물로 활용하는 구조를 제시하고, 정책적·실무적 적용 방안을 논의한다.

제8장에서는 연구 결과를 종합하여 결론을 제시하고, 본 연구의 학문적 의의와 실무적 시사점을 정리한다. 또한 연구의 한계와 향후 연구 방향을 제안한다.

2. 공급망 악성코드 위협 및 관련 연구 동향

2.1. 공급망 사이버보안 위험의 개념과 범위

공급망 사이버보안 위험은 조직이 제공·운영하는 제품과 서비스가 의존하고 있는 공급망 전 과

정에서, 사이버 위협으로 인해 정보자산과 업무 프로세스의 기밀성, 무결성, 가용성이 침해될 가능성을 의미한다. 즉, 생산·개발·유통·운영 등 제품·서비스 생애주기에 참여하는 다양한 공급자, 서비스 제공자, 하위 공급망 참여자와 그들이 보유·운영하는 정보시스템, 소프트웨어, 하드웨어, 데이터, 네트워크 등이 사이버 공격의 매개가 되거나 공격 표면으로 활용될 위험을 포괄하는 개념이다. 본 연구에서는 이러한 관점에서, 공급망 사이버보안 위험을 "조직 외부의 공급망 자산 및 참여자와의 연계를 통해 조직 내부로 전달·증폭될 수 있는 사이버보안 위협의 잠재적 손실"로 정의한다. 전통적인 정보보안은 주로 조직 내부 자산 중심의 보호에 초점을 두어, 물리적 경계나 네트워크 경계를 기준으로 내부 시스템, 서버, 단말, 애플리케이션, 데이터베이스 등 조직이 직접 소유·통제하는 자산을 보호 대상으로 설정해 왔다. 이와 같은 내부 자산 중심 보안은 방화벽, 침입차단시스템, 엔드포인트 보안 솔루션 등 경계 기반 통제를 통해 외부와 내부를 구분하고, 내부 자산을 "신뢰 영역(trusted zone)"으로 간주하는 보안 모델에 기반한다. 이러한 접근에서는 외부 공급자나 서비스 제공자는 통상적으로 계약과 인증, 절차적 점검 등을 통해 신뢰가 확보되었다고 전제되며, 그들의 내부 보안 상태나 소프트웨어 개발·운영 과정에서의 보안 수준은 조직의 기술적 통제 범위 밖에 머무르는 경우가 많다.

반면 공급망 보안 관점에서의 사이버보안 위험은 보호 대상을 조직 내부 자산에 한정하지 않고, 조직의 제품·서비스 제공에 실질적으로 관여하는 제품 구성요소, IT·OT 서비스, 1차·2차·3차 공급자, 클라우드 및 SaaS 제공자, 오픈소스 소프트웨어 및 서드파티 라이브러리, 외주 개발 및 유지보수 업체, 물류·운영 파트너 등 공급망 전반으로 확장한다. 이때 공급망 참여자들 간에는 계약, 조달, 기술적 연동, 데이터 공유, 운영 위탁 등 다양한 형태의 의존 관계가 형성되며, 특정 공급자의 취약점이나 보안 사고가 연쇄적으로 다른 조직으로 전파될 수 있다.

특히 공급망 구조는 다단계 하위 공급망을 포함하는 계층적 구조를 가지는 경우가 많아, 조직이 직접 계약 관계를 맺고 있는 1차 공급자뿐 아니라, 1차 공급자가 다시 의존하는 2차·3차 공급자의 보안 수준과 운영 관행도 간접적으로 조직의 위험 수준에 영향을 미친다. 이와 같이 공급망 사이버보안 위험은 단일 조직의 내부 시스템 범위를 넘어, 제품·서비스·공급자·하위 공급망 전반에 걸쳐 분산된 자산과 프로세스, 이해관계자 네트워크에서 발생하는 연쇄적·전이적(transitive) 위험이라는 특징을 가진다. 따라서 공급망 보안은 내부 자산 중심의 경계 기반 보안을 보완·확장하여, 공급망 전체 생태계를 대상으로 한 통합적인 위험 식별, 평가, 통제, 모니터링 체계를 요구한다.[4]

이러한 관점에서 볼 때, 공급망 사이버보안 위험은 단순히 개별 시스템의 기술적 취약점 문제가 아니라, 조직 경계 밖에 존재하는 다양한 공급자와 하위 공급망 참여자의 보안 수준, 개발·조달·운영 프로세스, 계약 및 관리·감독 체계, 그리고 이들 사이의 상호 연계 구조까지 포괄하는 거버넌스·관리·기술이 결합된 복합적 위험 영역으로 이해할 수 있다. 본 연구는 NIST SP 800-161 Rev.1에서 제시하는 공급망 보안 리스크 관리 체계를 토대로 이러한 공급망 사이버보안 위험을 구조화하고, 단계 및 위협 유형별 분류 체계를 제안하고자 한다.

2.2. 공급망 사이버보안 위험의 주요 발생 요인

공급망 사이버보안 위험은 제품과 서비스 자체뿐만 아니라, 이를 제공·운영하는 공급자 및 다단계 하위 공급망 전반에서 발생한다. 특히 NIST SP 800-161 Rev.1은 공급망 전주기에 걸쳐 다양한 이해관계자와 자산이 상호 연계되면서, 여러 유형의 보안 위협 요인이 복합적으로 작용할 수 있음을 강조한다. 대표 요인은 악의적 기능 삽입, 위조·품질 미달, 개발·제조·통합 결함, 공급망 복잡성이다.

2.2.1. 의도적인 악의적 기능(Malicious Functionality)의 삽입

악의적 기능이 삽입된 보안 위협 요소는 시스템이나 소프트웨어에 고의적으로 포함된 악성 코드 또는 보안 취약 기능을 의미한다. 이는 개발, 통합, 유통·배포 등 공급망의 다양한 단계에서 은밀하게 추가될 수 있으며, 정상 소프트웨어나 업데이트 패키지, 서드파티 라이브러리, 펌웨어 등에 백도어, 정보 유출 기능, 원격 제어 기능 등을 내장시키는 형태로 나타난다. 이러한 악의적 기능은 공급망 상의 신뢰 관계를 악용하여 최종 사용자 조직 내부로 전파되기 때문에, 전통적인 경계 기반 보안 통제로는 탐지·차단이 어려운 특징을 가진다.

표 1. 의도적으로 삽입된 보안 위협 요소 사례

사건명	방식	삽입 위치	탐지 시점	피해 범위
SolarWinds	공급망 공격	소프트웨어 업데이트	수개월 후	글로벌 기업·정부
XZ Utils	오픈소스 내부자 공격	압축 라이브러리	출시 직전	리눅스 시스템 전반
Juniper	암호화 백도어	방화벽 펌웨어	수년 후	정부기관·기업
CCleaner	빌드	시스템 침해	설치 ·수개월 후	수백만 사용자

2.2.2. 위조 또는 품질 미달로 인한 보안 취약성 요소

위조품 또는 저품질 부품은 단순히 기능 저하의 문제가 아니라, 직접적인 보안 취약점을 유발하거나 공격자가 시스템을 침해하는 경로로 악용될 수 있다. 이는 특히 하드웨어기반 시스템이나 임베디드 시스템, 공공/국방 시스템에서 치명적이다.

표 2. 위조 또는 품질 미달로 인한 보안 취약성 요소 사례

사례	부품 유형	위협요소	보안 영향
Cisco 위조 장비	네트워크 장비	백도어, 취약 펌웨어	시스템 침입, 데이터 유출
중국산 반도체	IC 칩	품질 미달, 백도어 가능	무기 시스템 오작동, 침해 가능성
Supermicro	메인보드	스파이 칩 삽입	원격 감시 및 명령 주입
위조 TPM	보안 칩	신뢰 실패	암호화 무력화, Secure Boot 우회

2.2.3. 부적절한 개발·제조·통합 과정에서 발생하는 보안 결함은 주로 보안 고려 없이 설계

개발/조립된 시스템에서 발생하며, 이는 기능에는 이상이 없어 보이나 실제 사용 중에 치명적인 보안 취약점을 유발한다. 이와 같은 문제는 관리 부실, 검증 부족, 개발 프로세스 미비 등에서 비롯되며, 특히 보안이 요구되는 시스템에서는 심각한 결과를 초래한다.

표 3. 부적절한 개발·제조·통합 과정으로 인한 보안 결함 사례

사례	결함 유형	원인	결과
Stuxnet	통합 보안 부재	오프라인 시스템 통제 미흡	제어 시스템 파괴
항공기 해킹	네트워크 분리 실패	시스템 경계 설계 오류	항법 시스템 침해
POS 단말기	펌웨어 무결성 미검증	제조/배포 관리 미흡	고객 카드 정보 유출
Mirai IoT	하드코딩 비밀번호	개발단계 보안 미반영	대규모 DDoS 공격
Jeep 해킹	ECU 통합 보안 부재	통신 경계 미설정	차량 원격 조작 가능

2.2.4. 공급망의 복잡성은 사이버 보안의 취약점을 은밀하고 치명적으로 확대

책임소재 불명확으로 사고 대응 지연, 취약한 하위 공급업체로 인한 공격의 우회 경로 제공, 지역별 보안 기준 상이로 일관된 보안 관리 불가, 하위 공급자의 공급망이 다시 공격받는 다단계 위험이 발생할 수 있다.

표 4. 공급망 복잡성으로 인한 보안 위험 증폭 사례

사례	주요 복잡성 요소	결과
SolarWinds	다단계 개발, 글로벌 고객사	국가 기반 해킹 확산
NotPetya	현지 규제 따른 로컬 소프트웨어	글로벌 기업 마비
ASUS	OEM 제조/업데이트 체계	정식 채널 통한 악성코드 유포
Huawei	법·규제 환경 차이	통신 인프라 신뢰 불일치
Log4j	수직적·수평적 오픈소스 의존	전 세계적 보안 패닉

2.3. 위험의 특성

공급망 위험은 기술적 결함을 넘어 운영 중단, 재정 손실, 평판 악화로 연결된다. 특히 하위 공급자·외부 서비스의 보안 수준이 낮을수록 전체 시스템 신뢰성이 저하된다. 위험은 위협·취약성·발생 가능성·영향도의 함수로 평가되며, 환경 변화에 따라 동적으로 변한다.

또한 공통 구성요소(SBOM 상 공유 라이브러리)로 인한 상관·동시다발 위험, 특정 공급자 집중으로 인한 집중 리스크, 한 곳의 취약점이 다수 조직에 미치는 연쇄·전이 리스크가 두드러진다.

3. NIST 800-161 기반 공급망 사이버보안 위험모델 및 관리요소 도출

3.1. 정의·분류 체계와 가시성(Visibility) 확보

효율적 위험 관리의 출발점은 가시성(Visibility) 확보이다. 가시성은 공급망 전반에서 '누가· 무엇을·언제·어디서·어떻게' 제공하는지를 주기적으로 파악하는 상태를 의미하며, 재고 수준을 넘어 보안·품질·규제 준수까지 포함한다. 가시성이 부족하면 잠재 위협을 인지하지 못한 채 노출된다. 가시성 확보 후 단계별(조달·개발·생산·운송·운영) 위험평가와 위협 시나리오 기반 분석을 수행한다.

그림 2. 출처: ETRI 오픈소스 거버넌스 대응체계

3.2. 거버넌스 통합

공급망 보안 위험은 공급망 전반에 걸쳐 의사결정, 책임, 규제 준수, 위험 관리, 윤리, 지속가능성 등 다양한 거버넌스 요소들을 통합적으로 관리 연계되어야 한다.

표 5. 공급망 거버넌스 통합의 핵심 구성요소

구성 요소	설명
정책 및 규정 통합	글로벌, 지역, 산업별 규제 및 윤리 기준을 공급망 전반에 걸쳐 통합 적용
리스크 관리	공급망 전반의 리스크(예: ESG, 지연, 품질, 사이버 보안 등)에 대한 공동 대응체계 수립
IT 및 데이터 통합	ERP, SCM, CRM 등 시스템 간 데이터 연계 및 실시간 정보 공유
역할과 책임 구조화	각 참여자의 권한, 책임, 의사결정 권한 등을 명확히 설정
지속 가능성 통합	ESG, 탄소배출, 인권, 공정무역 등의 기준을 공급망 전체에 반영
성과 모니터링 및 평가	KPI를 통한 전체 공급망의 성과 측정 및 개선관리 체계 구축

우선, ISO 28000, ISO 37000, OECD 공급망 지침 등 국제 표준에 기반한 거버넌스 프레임워크를 도입하고, 공급업체에 대한 평가 및 인증 제도, 자체 심사 기준, 제3자 인증 연계 체계를 마련할 필요가 있다. 또한 거버넌스 위원회 또는 전담 조직을 구성하여 공급망 정책, 리스크 관리, 규제 대응을 총괄하고, 조직 간 협업과 파트너십을 강화해야 한다. 아울러 공급업체와의 지속적인 커뮤니케이션, 공동 교육, 목표 설정 등을 통해 공급망 전반의 보안 수준을 체계적으로 향상시키는 노력이 요구된다.

3.3. 이해관계자 협력

내부(보안·조달·법무·개발/운영)와 외부(공급자·SI·서비스 제공자) 협력이 필수다. 의도적 위협은 탐지가 어려우므로, 사전적 통제(계약 보안요구·감사권·취약점 통지·코드/아티팩트 검증)와 사후/상시 모니터링을 결합한다.

수준	이해관계자	역할	활동
1. 전사	경영진[26]	C-SCRM 활동에 대한 경영진의 감독을 확립	• 전사적 C-SCRM 전략 정의 • 거버넌스 구조 및 운영 모델 수립 • 기업의 위험 구성, 위험관리 방식의 기조 설정 • 대체적인 구현 계획, 정책, 목적, 목표를 정의 • 전사적 수준의 C-SCRM 의사 결정 수행 • C-SCRM PMO(프로젝트 관리조직) 구성
2. 프로세스	비즈니스 관리자[27]	기업의 미션과 비즈니스 프로세스 측면에서 공급망의 사이버보안 위험을 평가, 대응, 모니터링	• 비즈니스 프로세스별 전략 개발 • 정책, 절차, 지침, 제약사항 개발 • 신규 IT 프로젝트에서 보안취약점 감축 • C-SCRM 구현 계획 개발 • 기업의 위험관리 체계를 비즈니스 프로세스에 맞게 조정 (예. 위험 허용 범위 설정) • 비즈니스 프로세스 내 위험관리 • C-SCRM에 관해 레벨1에 보고, 레벨3의 보고에 대한 조치
3. 운영	시스템 관리자[28]	개별 시스템 및 업무에 C-SCRM을 적용하고 운영 및 보고	• C-SCRM 계획 개발 • C-SCRM 정책과 요구사항 구현 • 레벨1과 2에서 제공한 제약사항 준수 • 개별 시스템의 상황에 맞게 C-SCRM을 조정하고 SDLC에 적용 • C-SCRM에 관해 레벨2에 보고

그림 3. 이해관계자의 역할에 따른 주요 C-SCRM 활동

3.4. Frame-Assess-Respond-Monitor(FARM) 프로세스

NIST 800-161은 반복적 FARM 프로세스로 공급망 위험을 관리한다.

표 6. FARM 기반 C-SCRM 운영에서 딥러닝 탐지와 산출물 연계(저자 재구성)

단계	공급망 단계	입력	탐지의 역할	산출물/증적 (모니터링 환류)
Frame	조달·계약·개발· 배포 전주기	자산/공급자/구성요소 식별 목록, 중요도	고위험 영역 정의 스캔/검증할지 범위 설정	스캔 정책, 위험기준서, 책임체계
Assess	배포 전/업데이트 수용 전	코드/파일/패키지, 해시 서명, SBOM 등	1차(시그니처) 통과 후 2차(딥러닝)로 악성확률	구성요소별 위험점수표, 판정 로그
Respond	계약/검수/배포/ 운영 의사결정	Assess 결과(위험 점수, FN/FP 고려한 정책)	위험점수에 전략 선택 지원	통제 적용 내역, 예외승인 기록
Monitor	배포 전·후 정기/상시	정기 스캔, 업데이트 이벤트, 신규 IOC	정기 스캔/실시간 모니터링으로 신규 위협 탐지 → 결과를 정책/리스트에 피드백	정기 리포트, 추세(오탐/미탐), 개선 이력

조달 및 계약 단계에서 보안 요구사항을 명확히 규정하고, 이를 하위 공급망까지 확산시키는 것이 중요하다. 또한 전담 책임자 지정, 교육·인식 제고 활동 등을 통해 조직 전반의 대응 역량을 강화해야 한다. Assess 단계의 '판정 로그'는 (1) 대상 구성요소 식별자, (2) 스캔 시각, (3) 모델 버전, (4) 적용 임계값(τ), (5) 악성 확률 p 및 위험점수, (6) 최종 판정(차단/검토/허용)을 최소 필드로 포함하도록 정의한다. Respond 단계의 '예외 승인 기록'은 (1) 대상 구성요소 식별자, (2) 승인 사유, (3) 승인자, (4) 유효기간(만료일), (5) 보완통제(예: 추가 분석, 모니터링 강화 등)를 최소 필드로 포함하도록 정의한다.

4. 공급망 관점 딥러닝 기반 악성코드 탐지 모델 설계

4.1. 문제정의 / 위협모델

본 연구의 목표는 공급망 구성요소(코드, 빌드 산출물, 업데이트 패키지, 서드파티 라이브러리) 단위에서 정적 분석만으로 악성 여부를 판정하고, 그 결과를 NIST SP 800-161 Rev.1의 C-SCRM(FARM) 흐름에 연결 가능한 '통제 의사결정 산출물(위험점수/판정로그/예외승인기록)'로 제공하는 것이다. 위협모델은 "공급망 유입 단계에서 악성 바이너리가 포함되거나 변조된 산출물이 배포되는 상황"을 가정하며, 운영 상의 최우선 목표는 미탐(FN) 최소화 또는 기대비용 최소화 등 정책 목적에 따라 달라질 수 있다.

4.2. 핵심 난제

정적 바이너리 기반 탐지는 난독화/패킹 등으로 특징이 왜곡될 수 있고, 실행 행위 기반 단서가 제한되며, 운영 환경에서는 오탐(FP)·미탐(FN)의 비용이 비대칭이라는 한계를 가진다. 특히 공급망 보안에서는 미탐이 실제 침해로 이어질 수 있어 '재현율 중심 운영'이 요구될 수 있으나, 반대로 과도한 오탐은 배포 지연과 업무 중단 비용을 유발한다. 따라서 모델 설계는 정적 표현의 정보 손실을 최소화하면서, 운영 목적에 맞는 임계값/비용 민감 조정이 가능하도록 구성되어야 한다.

4.3. 통찰

바이너리는 1차원 바이트 시퀀스로 표현되며, 이 시퀀스는 국소 패턴(헤더/섹션/바이트 반복)과 장거리 의존성(멀리 떨어진 구간의 연관) 모두가 존재한다. 이에 본 연구는 바이트를 2D로 재배열

하여 국소 공간 패턴을 학습하는 이미지 기반 CNN과, 원 시퀀스의 순서를 유지하면서 게이팅으로 중요한 특징을 선택하는 1D CNN-GLU를 대비시켜, 동일 데이터/동일 지표 체계 하에서 “표현 방식의 차이”가 탐지 성능을 비교한다.[5,6]

4.4. 모델 구조

본 연구에서는 기본적인 byte2pixel 아이디어를 유지하되 이미지로 변환하지 않고 1차원 시퀀스 형태로 직접 처리하는 시퀀스 기반 접근을 추가로 설계하였다. 이후 임베딩 시퀀스를 입력으로 하는 1차원 합성곱 신경망(1D CNN)을 적용하였다. 시퀀스 기반 접근은 2차원 이미지로의 재배열을 거치지 않으므로, width·height를 임의로 결정하는 과정에서 발생하는 공간적 왜곡을 피하고, 바이트 순서 정보와 구조적 패턴을 1차원에서 그대로 보존한 상태로 학습할 수 있다는 장점이 있다.[7]

4.5. GLU 이론적 근거

GLU(Gated Linear Unit)는 선형 변환 결과에 시그모이드 게이트를 곱해 정보 흐름을 조절하는 구조로, 불필요한 특징을 억제하고 중요한 특징을 선택적으로 통과시키는 효과가 있다.[8] 정적 바이너리의 경우 의미 있는 바이트 패턴이 전체 시퀀스 중 일부 구간에 희소하게 존재할 수 있어, 게이팅은 노이즈를 줄이고 유효 신호를 강조하는 데 기여할 수 있다. 본 연구는 동일한 1D-CNN 골격에서 GLU를 적용함으로써 “게이팅의 유효성”을 실험적으로 확인할 수 있도록 결과에서 구성요소 단위 분석을 추가한다.[9,10]

4.6. 모델 선택 근거

본 연구가 CNN과 1D CNN-GLU를 선택한 이유는 두 모델이 서로 다른 표현 가정을 갖고, 정적 탐지에서 널리 사용되는 합성곱 기반 접근이면서도, 운영 적용을 위한 가능성(혼동행렬 기반 검증) 등 실무 관점 비교가 가능하기 때문이다. 반면 트랜스포머류 모델은 장거리 의존성에는 유리하나 데이터 규모/학습 비용 요구가 커 본 연구의 소규모 시나리오 데이터셋에서는 비교의 공정성을 해칠 수 있어, 향후 연구로 확장 가능성을 논의한다.

4.7. 악성코드 정적 분석 방법

악성코드는 정적·동적·하이브리드 기법으로 분석한다. 정적은 빠르고 안전하나 패킹에 취약하고, 동적은 은닉 행위 포착에 유리하나 환경 우회 가능·비용 제약이 있다. 본 연구는 정적 바이트열을 주 입력으로 사용하였으며, PE 헤더/메타 정보는 해석 근거 수준으로만 활용하였다.

표 7. 시그니처 및 이상탐지(정적분석)

구분	시그니처 탐지	이상탐지
탐지 대상	이미 알려진 공격	알려지지 않은 새로운 공격
탐지 정확도	매우 높음(오탐 적음)	새 공격 탐지 가능하지만 오탐 발생
동작 방식	Rule / Hash / 패턴 매칭	통계 / 머신러닝 / 특징 벡터 기반
장점	빠르고 정확	신종 공격 대응
한계	신종 공격 취약·우회 가능	데이터 품질·설명가능성 과제

표 7에서는 정적 분석의 시그니처 탐지와 이상탐지를 비교하였으며 공급망 맥락에서는 기지 위협과 신종/변종 위협이 공존하므로 시그니처 탐지와 이상 탐지를 결합한 계층형 운영이 요구된다. 딥러닝 기반 탐지를 2차 계층으로 두어 정적분석의 시그니처의 한계를 보완한다.[11]

5. 실험 방법

5.1. 데이터셋

본 연구의 악성 샘플은 MalwareBazaar 정책을 준수하여 수집·검증하였고 해시 기반으로 중복을 제거하였다. 정상 샘플은 PE 실행파일로 범위를 제한하고, Authenticode 서명 검증(유효 서명 여부)과 공식 배포처 확인을 모두 통과한 파일만 포함하였다. 모든 원시 데이터는 정적 읽기만 수행하고 실행은 배제하였다.[12] 수집된 악성코드 샘플은 공개 저장소의 특성상 다양한 패밀리로 구성되어 있으며, 특정 패밀리에 편중되지 않도록 중복 해시 제거 및 시점 기반 샘플링을 통해 분포를 관리하였다.

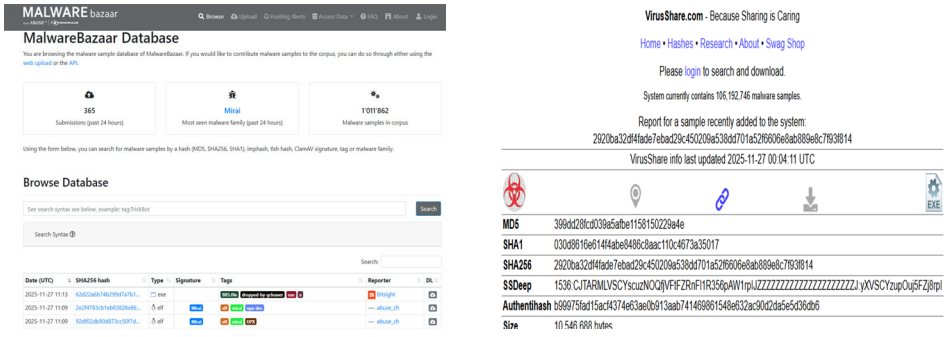


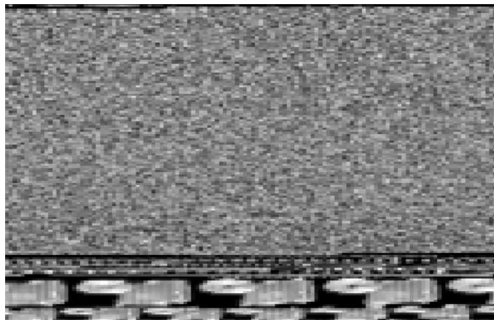
그림 4. 대중에게 공개된 악성코드 샘플

그림 4에서 제시하는 MalwareBazaar 등 대중 공개된 악성코드 데이터셋 그리고 오피스 및 윈도우 머신에서 추출한 .dll 및 .exe 정상 파일을 바탕으로 악성코드 580, 정상 580개로 총 1,160개이다. 입력 표현은 (1) 바이너리를 byte2pixel 방식으로 2차원 배열로 재배열한 뒤 256×256으로 정규화하여 CNN에 입력하는 이미지 경로, (2) 바이너리를 0-255 범위의 바이트열로 변환해 길이 max_len(= 4096)으로 정규화(초과분 절단, 부족분 후행 0-패딩)한 뒤 임베딩-1D CNN-GLU에 입력하는 시퀀스 경로로 구성하였다. 시퀀스 데이터는 (X, y) 형태로 malware_seq_dataset.npz에 저장하였으며(악성 = 1, 정상 = 0), 학습은 Binary Cross-Entropy 손실(BCEWithLogitsLoss; BCE + Sigmoid)과 Adam 옵티마이저(학습률 1e-3)를 사용하였다. 배치 크기 32, 최대 10 epoch, 조기 종료 (patience = 3, min_delta = 0.0)를 적용하였고, 재현성 확보를 위해 전역 시드(SEED = 42) 및 cudnn 결정론 옵션을 고정하였다. 실험은 Ubuntu 20.04 기반 워크스테이션(NVIDIA RTX 2070, Intel Xeon E5-2680 2.70GHz, RAM 56GB) 환경에서 수행하였다.

5.2. 전처리/피처화

이미지 기반 CNN의 경우 바이너리를 고정 길이로 정규화한 후 2D 배열로 변환하였다. 1D CNN-GLU는 동일한 정규화 규칙으로 바이트 시퀀스를 구성하여 입력으로 사용하였다. 전처리

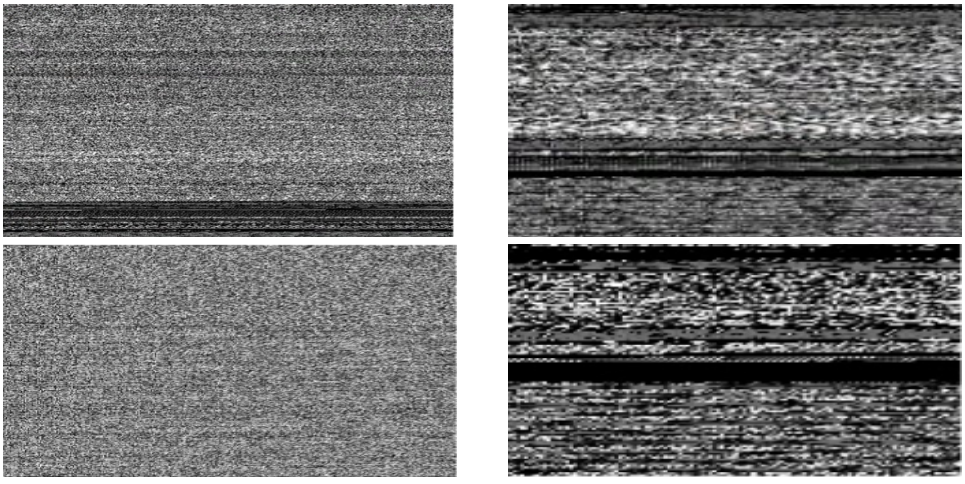
단계는 학습/평가 전 구간에 동일하게 적용되며, 재현성을 위해 알고리즘과 파라미터를 명시한다.[13]



MS-DOS 2.0 Compatible EXE Header
Unused
OEM Identifier OEM Information Offset to PE Header
MS-DOS 2.0 Stub Program and Relocation Table
Unused
PE Header (Aligned on 8-byte boundary)
Section Headers
Import Pages Import information Export Information Base relocations Resourcce information

TABLE I
TYPICAL PORTABLE EXE FILE LAYOUT

그림 5. 악성코드 바이너리의 이미지 변환 개요 및 Windows PE 파일 구조(저자 재구성)



악성파일 정상파일

그림 6. 이미지: width=⌊√size⌋로 2D 재배열 후 256×256으로 리사이즈(잔여(rem) 바이트는 제외)

그림 5, 그림 6에서와 같이 Windows PE(Portable Executable) 형식의 실행파일을 대상으로 정적 바이트 열을 특징 입력으로 사용하였다. 전체 바이너리를 그대로 사용할 경우 시퀀스 길이 증가로 과적합 및 메모리 부담이 커질 수 있으므로, 이미지 경로에서는 byte2pixel 방식으로 바이트를 2차원으로 재배열한 뒤 256×256으로 정규화하여 CNN에 입력하였고, 시퀀스 경로에서는 바이트열을 max_len(=4096)으로 절단/0-패딩하여 임베딩-1D CNN-GLU에 입력하였다. PE 헤더 및 메타 정보는 보조적 해석 근거로만 활용하였다.

5.3. 학습 방법, 하이퍼파라미터 및 구현 세부 사항

입력 소스는 ./data/raw/malware 및 ./data/raw/normal 디렉터리다. 각 바이너리 파일을 바이트열(0–255, uint8)로 파싱하여 길이 max_len의 고정 시퀀스로 정규화한 뒤, (X, y) 배열로 묶어 malware_seq_dataset.npz에 저장하였다. 길이 > max_len은 후미 절단, 길이 < max_len은 후행 0 패딩을 적용한다(라벨: 악성 = 1, 정상 = 0). 모델은 1D CNN-GLU(시퀀스 경로)와 선택적 메타 특징 경로(MLP) 결합 구조를 사용하였다.

손실은 Binary Cross-Entropy, 옵티마이저는 Adam(β1 = 0.9, β2 = 0.999), 초기 학습률 1e-3, Cosine decay 스케줄러를 적용하였다 배치 32, 최대 10 epoch 적용했다. 프레임워크는 PyTorch 2.x, CUDA 12.x. 재현성을 위해 난수 시드 고정했다. 최적 모델 선택 기준(F1, 조기종료 기준), 소규모 데이터셋 특성을 고려해, 테스트셋 기준 정확도(Acc)와 재현율(Recall)은 이항 비율 기반 Wilson 95% 신뢰구간으로 함께 제시하였다.

본 데이터셋은 공급망 공격 시나리오를 모사하기 위한 실험용 샘플로 활용되었으며, Stratified 5-fold 교차검증 및 시나리오 기반 실험 데이터에 한정되므로, 절대 성능의 일반화보다는 모델 간 상대 비교와 운영 연계에 초점을 두었다. 평가는 Stratified 5-fold 교차검증으로 수행하였다. 각 fold에서 학습은 동일한 하이퍼파라미터를 사용하며(seed 고정 포함), 최종 성능은 5개 fold의 평균으로 보고한다. 옵티마이저, 학습률, 배치 크기, 에폭, 조기 종료 등의 설정은 부록을 통해 요약한다.

5.4. 평가 지표 & 통계(5-fold, 평균±표준편차, CI)

지표는 Accuracy/Precision/Recall/F1을 사용하며, fold별 혼동행렬(TP/FP/TN/FN)을 함께 제시한다. 통계치는 5개 fold의 평균±표준편차로 보고하고, 평균에 대한 95% 신뢰구간은 부트스트랩으로 산출하였다.

6. 실험 결과

6.1. 성능 비교(시그니처 기반 악성코드 탐지 모델과 이미지 기반 모델)

시그니처 기반 악성코드 탐지 기법은 알려진 악성코드의 바이트 패턴, 해시, 특정 코드 조각과 같은 고유한 시그니처를 데이터베이스로 관리하고, 검사 대상 파일이 이 시그니처와 일치하는지를 비교함으로써 악성 여부를 판단한다. 이러한 방식은 알려진 악성코드에 대해서는 오탐(false positive)이 적고, 탐지 결과를 해석하기 쉽다는 장점이 있으나, 시그니처가 등록되지 않은 신규 악성코드나 경미하게 변형된 변종(variant)에 대해서는 탐지율이 급격히 저하된다는 구조적 한계를 가진다. 난독화, 패킹, 코드 다형성(polymorphism) 기법이 널리 사용되는 現 환경에서는 시그니처 기반 탐지만으로는 제로데이 공격 및 공급망을 통한 변종 악성코드 유입에 효과적으로 대응하기 어렵다.

이에 비해 이미지 기반 악성코드 탐지 기법은 실행 파일의 전체 바이트 시퀀스나 특정 구간을 픽셀 값으로 매핑하여 2차원 그레이스케일 또는 컬러 이미지로 변환한 뒤, 합성곱 신경망(CNN)과 같은 딥러닝 모델로 분류하는 방식이다. 이러한 접근은 시그니처와 같이 명시적인 특징 정의 없이도 모델이 자동으로 공간적 패턴을 학습할 수 있어, 기존 시그니처에 존재하지 않는 새로운 악성코드 패밀리나 변종에 대해서도 일정 수준의 일반화 성능을 확보할 수 있다는 장점이 있다. 여러 선행 연구에서 CNN 기반 이미지 분류 모델이 공개 악성코드 벤치마크 데이터셋에서 기존 시그니처 기반 또는 전통 머신러닝 기법과 동등하거나 그 이상 수준의 탐지율을 보였으며, 특히 패킹·난독화된 샘플에 대해서도 경쟁력 있는 성능을 보고하고 있다. 다만 이미지 기반 모델은 대규모 학습 데이터와 연산 자원이 필요하며, 모델 구조와 하이퍼파라미터 설정에 따라 성능 편차가 큰 점, 그리고 탐지 결과에 대한 해석 가능성이 낮은 점 등 현실 적용 시 고려해야 할 제약도 존재한다. 반대로 시그니처 기반 탐지는 경량·고속이라는 장점이 있어, 여전히 1차 필터링이나 이미 알려진 위협에 대한 신속한 탐지에는 유효한 수단이다. 따라서 두 접근법은 상호 대체 관계라기보다, 알려진 악성코드에 대한 시그니처 기반 탐지와 미지·변종 악성코드 및 공급망 공격 시나리오에 대한 이미지 기반 딥

러닝 탐지를 병행하는 계층적·하이브리드 탐지 구조로 통합할 때 가장 큰 효과를 발휘할 수 있다. 본 연구에서 수행한 시퀀스 1D CNN-GLU와 이미지 CNN의 성능 비교 실험은 이러한 보완적 관계를 정량적으로 검증하고, 공급망 보안 관점에서 어떤 유형의 위협에 어떤 탐지 기법을 우선 배치할 것인지에 대한 근거를 제공한다는 점에서 의의를 가진다.

본 연구는 악성 580개/정상 580개(총 1,160개)로 구성된 데이터셋을 사용하였다. 이미지 기반 모델의 경우 바이너리 파일을 byte2pixel 방식으로 PNG로 변환하여 학습 데이터로 사용하였다. 성능의 분산을 완화하고 일반화 성능을 안정적으로 추정하기 위해 Stratified 5-fold 교차검증을 수행하였다. 각 fold에서 테스트셋은 232개(악성 116/정상 116)로 구성되며, 나머지 928개를 학습/검증에 사용하였다. 학습 데이터 내부에서 stratified 방식으로 검증셋(10%, 93개)을 분리하고, 검증 정확도 기준으로 최적 에폭 모델(early selection)을 선택한 뒤 해당 모델을 테스트셋에 적용하였다.

표 8. 5-fold 혼동행렬 및 성능(시퀀스 기반 1D CNN-GLU, τ=0.5)

Fold	TP	FP	TN	FN	Acc	Prec	Rec	F1	AUC
1	111	9	107	5	0.9397	0.9250	0.9569	0.9407	0.9854
2	111	10	106	5	0.9353	0.9174	0.9569	0.9367	0.9890
3	111	5	111	5	0.9569	0.9569	0.9569	0.9569	0.9949
4	111	3	113	5	0.9655	0.9737	0.9569	0.9652	0.9956
5	114	6	110	2	0.9655	0.9500	0.9828	0.9661	0.9934

표 9. 5-fold 혼동행렬 및 성능(이미지 기반 CNN, τ=0.5)

Fold	TP	FP	TN	FN	Acc	Prec	Rec	F1	AUC
1	102	16	100	14	0.8707	0.8644	0.8793	0.8718	0.9567
2	106	20	96	10	0.8707	0.8413	0.9138	0.8760	0.9564
3	111	26	90	5	0.8664	0.8102	0.9569	0.8775	0.9569
4	93	12	104	23	0.8491	0.8857	0.8017	0.8416	0.9569
5	113	28	88	3	0.8664	0.8014	0.9741	0.8794	0.9605

표 8은 시퀀스기반, 표 9는 이미지기반 5-fold 혼동행렬 및 성능을 비교하였다. 이진 분류 성능 평가는 테스트 데이터에서의 혼동행렬 원소(TP, FP, TN, FN)에 기반하여 산출하였다. TP는 악성(1)을 악성으로 예측한 수, FN은 악성(1)을 정상(0)으로 예측한 수, FP는 정상(0)을 악성(1)으로 예측한 수, TN은 정상(0)을 정상(0)으로 예측한 수로 정의한다. 평가지표는 Accuracy = (TP + TN) / N, Precision = TP / (TP + FP), Recall = TP / (TP + FN)로 계산하였다. 모델 출력은 sigmoid (logit) 확률로 변환하였고, 임계값 τ = 0.5에서 y_pred = 1(y_score ≥ τ)로 이진 예측을 생성하였다. AUC는 예측 점수(y_score)에 기반하여 계산하였다.

표 10. 5-fold

구 분	시퀀스 기반 1D CNN-GLU		이미지 기반 CNN	
Metric	Mean ± SD	95% CI (mean, bootstrap)	Mean ± SD	95% CI (mean, bootstrap)
Accuracy	0.9526 ± 0.0143	[0.9414, 0.9638]	0.8647 ± 0.0089	[0.8569, 0.8698]
Precision	0.9446 ± 0.0232	[0.9269, 0.9622]	0.8406 ± 0.0356	[0.8129, 0.8683]
Recall	0.9621 ± 0.0116	[0.9569, 0.9724]	0.9052 ± 0.0687	[0.8483, 0.9552]
F1	0.9531 ± 0.0137	[0.9423, 0.9639]	0.8693 ± 0.0157	[0.8548, 0.8779]
AUC	0.9917 ± 0.0043	[0.9882, 0.9949]	0.9575 ± 0.0017	[0.9566, 0.9590]

표 10은 5-fold 교차검증에서 두 모델의 평균 성능을 비교한 결과이다. 전반적으로 1D CNN-GLU는 재현율/PR 관점에서 강점을 보이며, 이미지 기반 CNN은 특정 조건에서 정밀도 측면의 장점을 보였다. 이는 표현 방식과 게이팅 유무가 오탐·미탐 트레이드오프에 영향을 줄 수 있음을 시사한다. 또한 fold 결과(5개)에 대해 재표본추출(bootstrap, 10,000회)로 평균 성능의 95% 신뢰구간(95% CI)을 추정하였다.

표 11. pooled 혼동행렬 및 Wilson 95% CI

구 분		시퀀스 기반 1D CNN-GLU	이미지 기반 CNN
Pooled confusion (N=1160)		TP=558, FP=33, TN=547, FN=22	TP=525, FP=102, TN=478, FN=55
Wilson 95% CI	Accuracy CI	[0.9388, 0.9634]	[0.8438, 0.8831]
	Precision CI	[0.9226, 0.9600]	[0.8064, 0.8641]
	Recall CI	[0.9432, 0.9748]	[0.8786, 0.9264]

표 11에서는 Stratified 5-fold 교차검증 결과, 추가로 모든 fold의 예측 결과를 합산한 pooled 혼동행렬을 구성하고, pooled 비율(Accuracy/Precision/Recall)에 대해 Wilson 95% 신뢰구간을 함께 제시하였다.

시퀀스 기반 1D CNN-GLU 모델은 Accuracy 0.9526 ± 0.0143(95% CI [0.9414, 0.9638]), Recall 0.9621 ± 0.0116(95% CI [0.9569, 0.9724])로 높은 분류 성능을 보였으며, pooled 혼동행렬 기준 Wilson 95% CI에서도 Accuracy [0.9388, 0.9634], Recall [0.9432, 0.9748]로 안정적인 성능 범위를 확인하였다. 반면 이미지 기반 CNN은 Accuracy 0.8647 ± 0.0089 (95% CI [0.8569, 0.8698])로 상대적으로 낮았고, 특히 Precision의 변동(±0.0356)과 fold별 FP 증가가 관찰되어 오탐 비용이 큰 운영 환경에서는 임계값/정책 기반 보정이 필요함을 시사한다. 결과적으로 본 데이터셋에서는 시퀀스 기반(정적 바이트열) 모델이 이미지 기반 모델 대비 전반적 성능과 안정성 측면에서 우수하였다.

종합하면, NIST SP 800-161에서 제시하는 공급망 위험 기반 접근(Risk-based C-SCRM)의 관점에서, 시그니처 기반 탐지를 1차 필터로 유지하되, 제안한 시퀀스 및 이미지 기반 딥러닝 탐지 모델을 2차 심층 분석 계층으로 통합하는 하이브리드 구조가 공급망 보안 수준을 실질적으로 향상시키는 현실적인 방안으로 판단된다.

6.2. Ablation(통찰/구성요소 검증)와 유의성

통찰(게이팅의 효과)을 검증하기 위해 1D CNN-GLU에서 게이트를 제거한 1D CNN을 구성하

였으며 두 모델 간 성능 차이에 대해 fold 기반 부트스트랩 95% 신뢰구간을 산출하였다. 신뢰구간이 제한적으로 겹치는 지표에 대해서는 운영상 의미 있는 차이로 해석할 수 있으며, 엄밀한 유의성 검정(p-value) 비교는 향후 더 큰 데이터셋에서의 검증으로 확장할 예정이다.

7. 논의/시사점

본 장에서는 (i) NIST 800-161 통제 산출물 관점에서 본 탐지 결과의 활용 방식(위험점수/판정 로그/예외승인기록), (ii) 비용 구조에 따른 임계값 조정의 실무적 의미, (iii) 데이터 규모·외부 데이터셋 확장 및 최신 모델과의 직접 비교의 한계를 정리한다. 또한 이전 장에서 도출한 공급망 사이버보안 위험모델과 딥러닝 기반 악성코드 탐지 모델을 바탕으로, 공급망 보안 강화를 위한 통합 프레임워크를 제안하고 그 시사점을 논의한다. 먼저 NIST SP 800-161 Rev.1 기반 공급망 위험관리 체계와 본 연구에서 설계한 딥러닝 탐지 모델을 어떻게 연계할 수 있는지 통합 구조를 제시하고, 이어서 정책적·제도적 시사점, 조직 차원의 적용 방안 및 운영 절차, 산업별·조직규모별 적용 시 고려사항을 제시함으로써 연구 결과의 실무적 활용 가능성을 구체화하고자 한다.

7.1. NIST 800-161 기반 공급망 위험관리와 딥러닝 탐지 모델의 통합 프레임워크 제안

본 연구에서 제안하는 통합 프레임워크는 NIST SP 800-161 Rev.1에서 제시하는 C-SCRM (Cybersecurity Supply Chain Risk Management) 프로세스를 기반으로, 공급망 단계별 위험 식별·평가·통제·모니터링 과정에 딥러닝 기반 악성코드 탐지 기능을 결합한 구조를 갖는다. NIST SP 800-161 Rev.1은 조직이 공급망 전 생애주기에서 사이버보안 위험을 관리하기 위해 '거버넌스 수립–위험 식별–위험 평가–위험 대응–모니터링 및 개선'의 반복적 관리 사이클을 수행할 것을 요구한다. 본 연구는 이 관리 사이클 상에서 소프트웨어 및 서비스 공급망 관련 자산(코드, 빌드 아티팩트, 업데이트 패키지, 서드파티 라이브러리 등)을 대상으로, 딥러닝 기반 악성코드 탐지 모델을 정량적 평가 도구로 활용하는 구조를 제안한다.

구체적으로, 제안 프레임워크는 첫째, 공급망 자산 및 이해관계자를 식별하는 단계에서 본 연구의 위험 분류 체계를 적용하여 악성코드 관련 고위험 영역을 도출한다. 둘째, 이러한 고위험 영역에 대해 딥러닝 기반 악성코드 탐지 모델을 활용하여 코드·파일·패키지 수준에서의 악성 여부를 정량적으로 평가함으로써, 전통적인 문서 점검이나 설문 기반 평가를 보완한다. 셋째, 탐지 결과를 위험 평가 단계의 입력으로 반영하여, 각 공급자·서비스·제품 구성요소에 대한 위험도 점수 및 우선순위를 산정하고, 이에 따라 공급망 통제 전략을 결정한다. 넷째, 운영 단계에서는 배포 전·후 정기적 스캔과 실시간 모니터링을 통해 새로운 악성코드 위협을 탐지하고, 탐지 결과를 피드백하여 관리 체계를 지속적으로 개선한다.

이와 같은 통합 프레임워크를 통해 딥러닝 기반 악성코드 탐지 모델은 단순한 기술적 보안 솔루션이 아니라, 공급망 위험관리 프로세스에 내재화된 정량 평가 수단으로 기능하게 된다. 이는 공급망 보안이 관리적·정책적 통제에 치우쳐 있는 기존 C-SCRM 접근을, 데이터 기반·모델 기반 위험관리로 확장하는 역할을 한다. 또한 제안 프레임워크는 조직이 한정된 보안 자원 하에서 어떤 공급망 단계에 탐지 역량을 우선 배치해야 하는지, 탐지 결과를 어떻게 계약·감사·통제 활동과 연계해야 하는지에 대한 의사결정 구조를 제공한다는 점에서 실무적 활용 가치가 크다.

7.2. 정책적·제도적 시사점

본 연구의 결과는 국가 및 산업 차원에서 공급망 보안 정책·제도 설계 시 다음과 같은 시사점을

제공한다. 첫째, 공급망 보안 규제와 가이드라인은 단순히 공급자에 대한 인증·문서 제출 요구에 그치지 않고, 악성코드 탐지와 같은 기술적 검증 절차를 C-SCRM 프레임워크에 명시적으로 포함할 필요가 있다. 특히 공공조달, 국가핵심기반시설, 국방·에너지·금융 등 국가·사회 기반이 되는 분야에서는 소프트웨어 및 서비스 공급자에 대해 NIST SP 800-161 Rev.1 수준의 공급망 위험관리 체계를 요구함과 동시에, 악성코드 탐지 등 정량적 검증 지표를 포함한 보안 성숙도 평가를 제도화할 필요가 있다.

둘째, 정책 입안자는 공급망 보안 정책을 설계할 때, 딥러닝 기반 탐지 기술과 같은 첨단 보안 기술의 도입을 단순 권고 수준이 아닌, 위험도 및 중요도에 따른 단계별 의무화·차등 적용 방안과 연계하는 것이 바람직하다. 예를 들어, 고위험군 공급자 또는 중요 시스템에 관계된 공급자는 배포 전 코드 스캔, 주기적 악성코드 검증, 로그 기반 이상 행위 탐지 등 강화된 기술적 검증을 수행하도록 요구할 수 있다. 이러한 요구를 뒷받침하기 위해, 정부 차원에서 공용 악성코드 샘플 저장소 및 벤치마크 데이터셋, 레퍼런스 탐지 모델 등을 제공하여 중소 공급자도 일정 수준 이상의 검증을 수행할 수 있도록 지원하는 정책도 고려할 수 있다.

셋째, 공급망 보안 관련 제도는 국제 표준 및 글로벌 규제 환경과의 정합성을 확보해야 한다. NIST SP 800-161 Rev.1, ISO/IEC 27036, SBOM(Software Bill of Materials) 관련 정책 등과 연계하여, 국내 기업이 해외 조달시장·글로벌 공급망에 참여할 때 중복 규제나 기준 불일치로 인한 부담이 최소화되도록 조정할 필요가 있다. 이를 위해 국내 가이드라인에서는 공급망 위험모델과 딥러닝 탐지 모델과 같은 기술적 검증 요소를 국제 표준에서 요구하는 통제 항목과 매핑하여, 기업이 동일한 보안 활동을 다양한 규제 준수 증빙으로 활용할 수 있도록 하는 것이 유리하다.[14]

7.3. 조직 차원의 적용 방안 및 운영 절차(프로세스) 제안

조직이 본 연구에서 제안한 통합 프레임워크를 실제로 적용하기 위해서는 조직 구조, 역할·책임, 프로세스 측면에서의 구체적인 운영 방안을 수립해야 한다. 우선, 최고정보보안책임자(CISO) 및 공급망 관리 책임 부서가 공동으로 C-SCRM 거버넌스를 수립하고, 공급망 보안 정책, 표준, 절차를 제정해야 한다. 이 과정에서 정보보안 부서, 구매·조달 부서, 개발·운영(Dev/Ops) 부서, 법무·준법 부서 등 관련 조직 간 역할과 책임(R&R)을 명확히 정의하고, 공급망 위험 식별·평가·대응·모니터링에 대한 협업 구조를 마련해야 한다.

운영 절차 측면에서는 제품·서비스 생애주기 상의 주요 단계에 공급망 보안 점검 및 딥러닝 기반 악성코드 검증 절차를 통합하는 것이 중요하다. 예를 들어, 신규 공급자 선정 단계에서는 공급자의 보안 성숙도와 C-SCRM 체계 보유 여부를 평가하고, 계약 단계에서는 소스코드·바이너리 검증, 정기적 보안 테스트, 인시던트 통보 의무 등을 계약 조항에 명시할 수 있다. 개발 및 통합 단계에서는 CI/CD 파이프라인에 딥러닝 기반 악성코드 탐지 도구를 연계하여, 서드파티 라이브러리·오픈소스 컴포넌트·외주 개발 코드에 대한 자동 스캔을 수행하도록 구성한다. 배포 및 운영 단계에서는 업데이트 패키지, 패치, 구성 변경 등에 대해 사전·사후 검증 절차를 적용하고, 로그 등 데이터를 활용한 지속적인 모니터링을 수행한다.

또한, 조직은 딥러닝 탐지 모델의 운영과 관련하여 성능 모니터링, 모델 업데이트, 오탐·미탐 대응 절차를 포함한 MLOps(Machine Learning Operations) 체계를 마련해야 한다. 악성코드 위협 환경은 지속적으로 변화하므로, 새로운 위협 정보와 인시던트 데이터를 반영하여 모델을 주기적으로 재학습하는 절차가 필요하다. 이와 더불어, 탐지 결과에 대한 해석과 대응 의사결정을 지원하기 위해 보안관제센터(SOC) 또는 CSIRT(Computer Security Incident Response Team)와의 연계, 경보 우선순위 설정, 대응 플레이북(playbook) 정의 등이 요구된다. 이러한 프로세스를

통해 조직은 공급망 보안 통제를 일회성 점검이 아닌, 지속적인 위험관리 활동으로 정착시킬 수 있다.

7.4. 산업별/조직규모별 적용 시 고려사항

공급망 보안 통합 프레임워크의 효과적인 적용을 위해서는 산업별 특성과 조직규모를 고려한 차별화된 접근이 필요하다. 먼저, 전력·가스·교통·통신 등 국가기간산업과 금융, 국방, 보건의료와 같이 높은 가용성과 안전성이 요구되는 산업에서는 공급망 보안 위협이 서비스 중단, 물리적 피해, 국민 생명·재산 침해로 직결될 수 있다. 이러한 산업에서는 NIST SP 800-161 Rev.1 수준 이상의 엄격한 C-SCRM 체계와 딥러닝 기반 탐지를 포함한 고도화된 기술 통제를 도입하는 것이 필요하며, 규제기관의 감독·점검과 연계된 의무 수준도 상대적으로 높게 설정될 수 있다. 반면, 상대적으로 위험 영향이 제한적인 산업에서는 동일한 프레임워크를 적용하되, 위험도에 따라 통제 수준과 적용 범위를 조정하는 위험 기반(risk-based) 접근이 적절하다.

조직규모 측면에서는 대기업과 중소기업의 차이를 고려해야 한다. 대규모 조직은 자체적으로 C-SCRM 전담 조직과 SOC, MLOps 팀을 운영할 수 있는 인력·비용·기술 역량을 보유한 경우가 많아, 본 연구에서 제안한 통합 프레임워크를 비교적 직접적으로 구현할 수 있다. 반면, 중소규모 조직과 소규모 공급자는 제한된 자원으로 인해 동일한 수준의 통제와 기술을 자체 구축하기 어렵다. 이 경우, 클라우드 기반 보안 서비스(Security as a Service), 공공·산업 단위의 공동 악성코드 분석 인프라, 정부 또는 산업협회가 제공하는 레퍼런스 탐지 모델 및 가이드라인을 활용하여 간접적으로 프레임워크를 구현하는 방안이 현실적이다.

마지막으로, 글로벌 공급망에 참여하는 조직의 경우, 다양한 국가·지역의 규제와 표준, 고객 요구사항을 동시에 충족해야 하는 과제가 존재한다. 따라서 산업별·지역별로 상이한 규제 요구를 분석하고, NIST SP 800-161 Rev.1 기반 내부 프레임워크를 공통 플랫폼으로 삼아 그 위에 개별 규제 및 고객 요구사항을 매핑하는 전략이 필요하다. 이를 통해 조직은 공급망 보안에 대한 내부 기준을 일관되게 유지하면서도, 각 이해관계자의 요구에 유연하게 대응할 수 있다. 이러한 산업별·조직규모별 고려를 반영할 때, 본 연구에서 제안한 통합 프레임워크는 다양한 환경에서 실질적인 적용 가능성을 가질 수 있을 것이다.

8. 결론

8.1. 딥러닝 기반 악성코드 탐지에 대한 연구

본 연구는 NIST SP 800-161 기반 위험모델과 딥러닝 탐지의 통합을 통해, 공급망 악성코드 위협에 대한 정량적·운영적 대응 방안을 제시했다. 시퀀스 기반 1D CNN-GLU는 높은 판별력과 균형 잡힌 성능을 보였다. 또한 공급망 맥락 정보를 반영한 탐지 접근법이 어떤 차별적 탐지 패턴과 장·단점을 가지는지 확인하였다. 이러한 분석 결과는 공급망 공격 시나리오에서 악성코드를 보다 효과적으로 탐지하기 위해 어떤 특징 정보를 선택·결합해야 하는지, 그리고 딥러닝 모델 구조를 어떻게 설계·조정할 필요가 있는지를 보여줌으로써, 향후 공급망 특화 악성코드 탐지 기법 연구에 기초자료를 제공한다는 점에서 의의를 가진다.

8.2. NIST 800-161 기반 통합 프레임워크에 대한 결론

또한 본 연구는 NIST SP 800-161 Rev.1에서 제시하는 공급망 보안 리스크 관리 체계를 기반

으로, 제안한 위험 분류 체계와 딥러닝 기반 악성코드 탐지 모델을 연계한 통합 프레임워크를 제시하였다. 이를 통해 조직이 공급망 단계별로 어떤 위험 요소를 우선적으로 관리해야 하는지, 어떤 지점에서 딥러닝 탐지 모델을 배치·운영하는 것이 효과적인지, 그리고 탐지 결과를 어떻게 위험평가 및 통제수준 개선 활동과 연계할 수 있는지를 구조적으로 설명하였다. 이와 같은 통합 프레임워크는 공급망 보안을 단순히 기술적 탐지 기능의 도입으로만 바라보는 것이 아니라, NIST 표준 기반의 거버넌스·정책·프로세스와 결합된 전사적 리스크 관리의 관점에서 재구성했다는 점에서 의미가 있다. 궁극적으로 본 연구에서 제시한 통합 관점은 조직이 제한된 자원 하에서 공급망 보안 투자를 우선순위화하고, 관리적·기술적 통제를 유기적으로 결합하는 데 참고할 수 있는 실질적 방향성을 제공한다.

8.3. 향후 연구 과제

첫째, 악성코드 탐지 모델의 학습·평가에 활용된 데이터셋 등 본 연구는 총 1,160개 샘플을 분할하여 평가하였다. 향후 연구에서는 보다 다양한 공급망 공격 사례와 실제 현장에서 수집된 로그 등 데이터를 반영하여, 제안 모델의 일반화 성능과 실효성을 검증할 필요가 있다. 제한된 규모의 데이터셋과 공급망 시나리오 중심의 실험 설계를 기반으로 수행되었으며, 이에 따라 성능 분석을 중심으로 평가를 진행하였다. 향후 연구에서는 보다 대규모 데이터셋을 활용하여 탐지 성능에 대한 심층 분석을 수행할 예정이다. 둘째, 본 연구에서 제안한 딥러닝 모델은 성능 향상에 초점을 두었으나, 공급망 보안 의사결정 과정에서 모델의 판단 근거를 설명할 수 있는 설명가능 인공지능(XAI, Explainable AI) 기법과의 결합이 충분히 다루어지지 못하였다. 향후에는 탐지 정확도뿐 아니라 설명 가능성과 해석 가능성을 함께 고려한 모델 설계가 요구된다. 셋째, 본 연구에서 제안한 통합 프레임워크는 개념적·구조적 수준의 설계에 초점을 맞추었으며, 실제 조직 환경에서의 적용 사례와 비용·편익 분석은 제한적으로 다루어졌다. 향후 연구에서는 다양한 산업·조직 규모를 대상으로 시범 적용 사례를 축적하고, 경제적 영향과 운영상의 제약 요인을 정량적으로 분석함으로써 프레임워크의 현실 적합성을 검증해야 한다. 넷째, NIST SP 800-161 Rev.1 외에도 ISO/IEC 27036, SBOM(Software Bill of Materials) 등 공급망 보안과 관련된 다른 국제 표준·제도와의 정합성을 고려한 확장 연구를 통해, 보다 포괄적인 글로벌 공급망 보안 거버넌스 모델을 제시할 수 있을 것이다.

참고문헌

[1] NIST Special Publication 800-161 Rev.1: Cybersecurity Supply Chain Risk Management Practices for Systems and Organizations. pp. 1-15. 2022.

[2] NTIA. The Minimum Elements for a Software Bill of Materials (SBOM). pp. 6-14. 2021.

[3] NTIA. Framing Software Component Transparency: Establishing a Common Software Bill of Materials (SBOM). pp. 6-14. 2021.

[4] 국가정보원, 과학기술정보통신부. SW 공급망 보안 가이드라인 v1.0, 2024.

[5] 송무준, 이종관. 악성코드의 이미지화 방법이 딥러닝 기반의 악성코드 분류 성능에 미치는 영향 분석, 한국군사학논집, 제77권 제1호, pp. 512-517. 2021.

[6] A. Nataraj et al. Malware Images: Visualization and Automatic Classification. Proceedings of the International Conference on Cybercrime Security and Digital Forensics, pp. 9-11. 2011.

[7] 김혜정, 윤은준. 악성코드로부터 빅데이터를 보호하기 위한 이미지 기반의 인공지능 딥러닝 기법. 전자공학회논문지, 제54권 제2호, pp. 76-82. 2017.

[8] 권혜윤. 머신 러닝을 이용한 악성코드의 분류. 한국정보과학회 학술대회. 2017.

[9] 조영복, 딥러닝 기반의 R-CNN을 이용한 악성코드 탐지 기법. 한국디지털콘텐츠학회 논문지, 제19권 제6호, pp. 1177-1183. 2018.

[10] 최선오 등. 딥러닝을 이용한 악성코드탐지 연구동향. 정보보호학회지, 제27권 제3호, pp. 20-26. 2017.

[11] 정성민 등. V-그램: 명령어 기본 블록과 딥러닝 기반의 악성코드 탐지. 정보과학학회논문지, 제46권 제7호, pp. 599-605. 2019.

[12] abuse.ch. MalwareBazaar. https://bazaar.abuse.ch/ (검색일 2025. 12. 21).

[13] 최양서 등. 악성프로그램 탐지를 위한 PE헤더 특성 분석 기술. 융합보안논문지, 제8권 제2호, pp. 63-70. 2008.

[14] 류원옥 등. SW공급망 관리 및 SBOM 동향. 한국정보통신연구원, pp. 6-14. 2023.

부록. 실험 세부 내용

1. 경로 및 폴더 구조

영역	경로	설명
원시 데이터(악성)	data/raw/malware/	정적 읽기만 수행(무실행), 해시/출처 검증
원시 데이터(정상)	data/raw/normal/	서명·출처 검증 파일
시퀀스 NPZ	datasets/malware_seq_dataset.npz	X:(N,max_len), y:(N,)
이미지 출력	data/raw/malware_output/*.png	byte2pixel: width= ⌊√size ⌋ 로 설정한 뒤, rem=(size mod width) 바이트를 제외한 size−rem 바이트만으로 2D 재배열하고 256×256으로 리사이즈
이미지 출력	data/raw/normal_output/*.png	256×256 리사이즈
실험 결과	runs/〈타임스탬프〉/	metrics.csv, report.txt, curves_*.png, preds_*.npz

2. 환경 및 하이퍼파라미터

2.1. 계산 환경

항목	값	비고
OS / Python	Ubuntu 20.04.1(focal Fossa) Python 3.8.3	pip freeze 별도 첨부 가능
GPU	NVIDIA RTX 2070	없으면 CPU로 자동
CPU / RAM	Intel Xeon E5-2680 / 56 GB	
패키지	torch, scikit-learn, pillow, numpy, matplotlib	requirements.txt

2.2. 하이퍼파라미터

항목	값	설명
max_len	4096	시퀀스 길이
embed_dim / channels	128 / 128	1D-CNN + GLU
kernel_size	3	공통
dropout	이미지 0.5 / 시퀀스 0.2	모델별 dropout
weight_decay	0.0	정규화 미적용
optimizer / lr	Adam / 1e-3	학습 설정
batch / epochs / patience	32 / 10 / 3	조기종료 기준
이미지 크기	256×256	byte2pixel 후 리사이즈
임계값(threshold)	0.5	p ≥ 0.5 → 악성(테스트 평가 적용)

본 연구는 기준 임계값 0.5를 사용했으나, 실제 적용 시에는 미탐(FN) 비용을 반영하여 ROC 기반 최적 임계값 또는 비용민감(cost-sensitive) 임계값으로 재설정하는 절차가 필요하다.

3. 전처리 규칙

- 시퀀스: 앞에서부터 max_len까지 절단, 부족분 0-패딩(후행).

- 이미지: byte2pixel 방식으로 width= $\lfloor \sqrt{size} \rfloor$ 를 적용해 rem 바이트를 제외(size-rem만 사용)하여 2D로 재배열한 뒤, 256×256으로 리사이즈하였다.

4. 학습/검증/테스트 구성

- 검증 손실(val_loss)이 3 epoch 연속 개선되지 않으면 학습을 조기 종료하였다(patience = 3, min_delta = 0.0).

Fold별 혼동행렬 및 성능 (1D CNN-GLU, τ = 0.50)

Fold	TP	FP	TN	FN	Acc	Prec	Rec	F1	AUC
1	111	9	107	5	0.9397	0.9250	0.9569	0.9407	0.9854
2	111	10	106	5	0.9353	0.9174	0.9569	0.9367	0.9890
3	111	5	111	5	0.9569	0.9569	0.9569	0.9569	0.9949
4	111	3	113	5	0.9655	0.9737	0.9569	0.9652	0.9956
5	114	6	110	2	0.9655	0.9500	0.9828	0.9661	0.9934

Fold별 혼동행렬 및 성능 (Image CNN, τ = 0.50)

Fold	TP	FP	TN	FN	Acc	Prec	Rec	F1	AUC
1	102	16	100	14	0.8707	0.8644	0.8793	0.8718	0.9567
2	106	20	96	10	0.8707	0.8413	0.9138	0.8760	0.9564
3	111	26	90	5	0.8664	0.8102	0.9569	0.8775	0.9569
4	93	12	104	23	0.8491	0.8857	0.8017	0.8416	0.9569
5	113	28	88	3	0.8664	0.8014	0.9741	0.8794	0.9605

5. 평가 지표·통계 검정 보고 형식

- 지표: Acc, Prec, Rec, F1 등

6. 재현성 통제

- 전역 시드를 SEED=42로 고정하고(cudnn deterministic 설정 포함), 부록 A의 하이퍼파라미터·전처리 규칙과 동일한 설정으로 실험을 수행하였다.

※ 본 연구는 Stratified 5-fold 교차검증 hold-out으로 학습/검증/테스트를 하였다. 양성(positive) 클래스는 '악성코드'로 정의하였으며, 테스트 평가는 임계값(threshold) τ=0.5에서 p≥0.5이면 악성으로 판정하였다. 평가 지표는 Acc, Prec, Rec, F1를 사용하였다. 소규모 데이터셋에서 추정 분산을 완화하기 위해 테스트셋 예측 결과를 부트스트랩으로 재표본화하여 95% 신뢰구간을 산출하였고, 본문에 제시한 성능 값은 부트스트랩 평균(95% CI) 기준으로 보고하였다. 재현성 확보를 위해 전역 시드는 SEED=42로 고정하였다.

※ 지면 제한으로 부록 A의 추가 실험 로그(학습 곡선, 추가 통계/표, 코드 일부)는 생략하였다. 다만 본문 및 부록에 기재된 데이터 분할, 전처리 규칙, 하이퍼파라미터와 동일한 설정으로 실험을 수행하면 동일 조건에서 결과 재현이 가능하다.