

논문 2025-4-2 <http://dx.doi.org/10.29056/jsav.2025.12.02>

멀티모달 AI 기반 KOGL 공유저작물 저작권 침해 가능성 예측 방법

장세영*, 유인재*, 박병찬*, 김선집**, 김석윤*, 김영모*†

A Copyright Infringement Risk Prediction Method for KOGL Open Government Works Using Multimodal AI

Se-Young Jang*, In-Jae Yoo*, Byeong-Chan Park*,
Sun-Jip Kim**, Seok-Yoon Kim*, Young-Mo Kim*†

요약

공공데이터 개방 확대와 생성형 인공지능 기술의 빠른 확산으로 공유저작물의 활용이 증가함에 따라, KOGL(Korea Open Government License) 조건 위반 사례와 저작권 침해 문제가 지속적으로 발생하고 있다. 기존의 수작업 검수 방식은 데이터 증가 속도와 다양한 침해 유형에 대응하기 어려우며, 출처표시(BY), 비영리(NC), 변경금지(ND), 허위등록(CF) 등 복합적 조건 위반을 자동으로 탐지하기 어렵다는 한계를 가진다. 본 논문에서는 텍스트와 이미지 기반 멀티모달 데이터를 활용하여 KOGL 공유저작물의 저작권 침해 가능성을 예측하는 멀티모달 AI 기반 침해 위험도 분석 모델을 제안한다. 제안 모델은 입력-전처리-CUR Feature Extractor-GRU Sequence Model-CRU Rule Engine-결과 출력의 6단계 구조로 구성되며, 콘텐츠의 시간적·맥락적 흐름 정보를 반영하여 침해 유형을 분류한다. 1,000회 반복 실험 결과, 제안 모델은 기존 단일모달 방식 대비 Accuracy, F1-Score, AUC에서 높은 성능을 기록하였으며, 침해 예측 결과와 함께 근거(evidence) 기반 설명 정보를 제공하여 실무 적용 가능성을 입증하였다.

Abstract

With the expansion of open government data policies and the rapid deployment of Generative AI technologies, the utilization of KOGL (Korea Open Government License)-licensed open government works has significantly increased. However, copyright infringement issues continue to arise due to misunderstanding of license conditions, inaccurate labeling, and ambiguity in third-party ownership, leading to violations of attribution (BY), non-commercial use (NC), no-derivatives (ND), and false registration (CF). Traditional manual verification processes are insufficient for handling the growing volume and complexity of shared content, and they fail to automatically detect contextual and multimodal infringement patterns.

To address these challenges, this study proposes a multimodal AI-based copyright infringement risk prediction model that integrates textual descriptions and image content to evaluate KOGL compliance. The model consists of a six-stage pipeline-Input, Pre-processing, CUR Feature Extraction, GRU-based Sequence Modeling, CRU Rule Engine, and Final Decision-and learns contextual temporal patterns to classify infringement types. Through 1,000 repeated experiments, the proposed model achieved significantly improved performance over single-modal baselines in Accuracy, F1-Score, and AUC, and provides evidence-based explanatory outputs for practical decision support.

한글키워드 : 멀티모달 AI, KOGL 공유저작물, 저작권 침해 가능성 예측, GRU 시퀀스 모델, 공공데이터

keywords : Multimodal AI, KOGL Open Government Works, Copyright Infringement Risk Prediction, GRU Sequence Model, Open Government Data

* 숭실대학교 컴퓨터학부

접수일자: 2025.12.02. 심사완료: 2025.12.08.

** 한세대학교 응합보안학과

게재확정: 2025.12.20.

† 교신저자: 김영모(email: ymkim828@ssu.ac.kr)

1. 서 론

공공데이터 개방 정책 확대와 생성형 인공지능(Generative AI)의 발전으로 공유저작물 활용이 급증하고 있다. 공유저작물은 행정·교육·문화·산업 전반에서 핵심 데이터 자원으로 활용되고 있으며, AI 학습·데이터 기반 서비스 개발·홍보·연구 분야에서 필수적 역할을 수행한다. 한국의 공공저작물은 KOGL(Korea Open Government License)을 통해 출처표시(BY), 비영리(NC), 변경금지(ND)를 조건으로 자유롭게 이용할 수 있다.

그러나 실제 활용 과정에서는 라이선스 조건의 오해, 외주 제작물 권리관계 불명확성, Copyfraud 등 다양한 저작권 침해 문제가 지속적으로 발생하고 있다. 예를 들어 퍼블릭 도메인의 KOGL로 잘못 등록되거나, 스톡 이미지·폰트가 포함된 콘텐츠가 공유저작물로 게시되는 사례가 반복되고 있다. 이는 이용자에게 잘못된 권리 정보를 전달하여 2차 제작·재배포 과정에서 추가 침해를 발생시킨다.

또한 생성형 AI 학습 과정에서 KOGL 조건 미반영 데이터가 포함될 가능성은 AI 생성결과물의 2차 침해 위험을 높이며, 콘텐츠 자동 재생산 환경에서 침해 확산 가능성이 커지고 있다. 하지만 현재 공공기관과 플랫폼 기업은 KOGL 조건 검토를 수작업 중심으로 수행하고 있어 데이터 증가 속도와 판단 편차 문제를 해결하기 어렵다.

따라서 공유저작물의 권리정보와 콘텐츠 자체를 자동 분석하여 침해 가능성을 예측하는 기술적 접근이 요구된다. 이에 본 논문에서는 멀티모달 AI 기반 KOGL 공유저작물 저작권 침해 가능성 예측 방법을 제안한다. 사이트 정보·메타데이터·텍스트·이미지 기반 특징을 통합하고, CRU·GRU 기반 하이브리드 모델을 통해 침해

가능성을 자동 분류한다.

본 논문의 구성은 다음과 같다. 2장에서는 관련 연구로 공유저작물 및 KOGL 라이선스 분석, 공유저작물 침해 사례 분석 그리고 GRU 기반 시퀀스 모델을 기술한다. 3장에서는 본 논문에서 제안하는 멀티모달 AI 기반 공유저작물 저작권 침해 예측 방법을 기술한다. 4장에서는 실험 및 결과를 보고 5장에서 결론으로 마무리한다.

2. 관련 연구

2.1 공유저작물 및 KOGL 라이선스 분석

공유저작물(Open Government Works)은 공공 기관이 생산한 자료를 국민이 자유롭게 이용할 수 있도록 허용하는 제도이며, 그림 1과 같다.



그림 1. 침해 사례 유형 정의

Fig. 1. Defining the Types of Violations

한국에서는 2013년 도입된 KOGL(Korea Open Government License)이 표준 라이선스로 운영되고 있다. 기존 연구는 KOGL 조건(BY, NC, ND)과 정책 효과 분석에 집중되어 왔으나, 실제 활용 과정에서 발생하는 침해 사례를 자동 탐지하거나, AI 기반으로 침해 가능성을 정량화하는 연구는 미흡하다. 특히 출처 표시 누락, 비영리 조건 오해, 외주 제작물 포함 여부 판단 등 반복적 문제는 기술 기반 자동 검증 시스템의 부재로 인해 해결되지 못하고 있다. 이러한 한계를 해결하기 위해 데이터 기반 KOGL 침해 예측 모델 개발이 필요하다.

2.2 공유저작물 침해 사례 분석

공공누리 상담사례 자료는 다양한 침해 유형을 정리하고 있으며, 대표적으로 출처표시 불완전, KOGL 마크만 부착한 불완전 표기, 스톡 이미지·폰트 포함 콘텐츠의 KOGL 오인 등록, Copyfraud 등이 반복된다. 기존 연구는 사례 수집에 그치고 있으며, 침해 유형별 구조화·정량화 데이터셋 구축 및 학습 모델 입력으로 활용한 연구는 부족하다. 본 논문은 이러한 사례를 BY·NC·ND·CF(허위 KOGL 등록) 등으로 데이터화하여 자동 분석 가능하도록 설계한다는 점에서 차별성이 있다.

2.3 GRU 기반 시퀀스 모델

저작권 침해 판단은 단순 키워드 검출이 아니라 문서의 시간적·맥락적 흐름 분석이 필요하다. GRU(Gated Recurrent Unit)는 LSTM 대비 단순 구조와 빠른 학습 성능을 갖고 있어 공공기관 환경의 제한된 연산 자원에서도 효과적이다. GRU는 출처표시 - 편집 여부 - 상업성 표현의 등장 순서와 같은 패턴을 탐지함으로써 BY·NC·ND 침해 판단에 필요한 문맥 단서를 추출할 수 있다. 또한 이미지·텍스트·메타데이터 입력을 시간적 시퀀스로 결합할 수 있어 멀티모달 기반 KOGL 침해 예측에 적합하다. 본 논문의 AI 모델은 GRU 시퀀스 분석을 핵심 구조로 활용한다.

지 기반 멀티모달 입력을 활용하여, 입력 단계 - 전처리 단계 - CUR Feature Extractor - GRU 기반 Sequence Model - CRU Rule Engine - 최종 판단 단계의 6단계로 구성된다. 전체 구조는 그림 2와 같다.

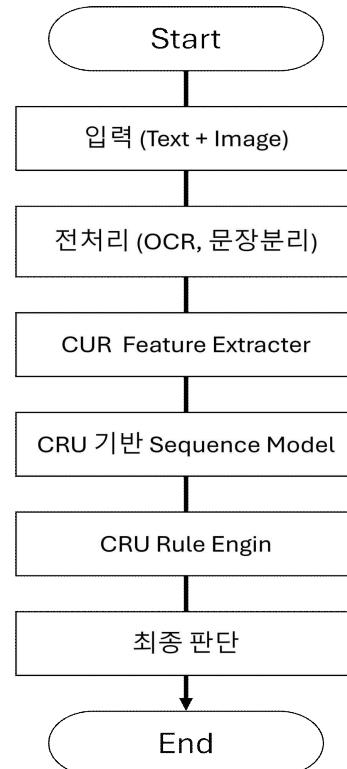


그림 2. KOGL 침해 예측 순서도
Fig. 2. KOGL breach prediction flowchart

3. 멀티모달 AI 기반 공유저작물 저작권 침해 예측 방법

3.1 개요

본 논문에서 제안하는 멀티모달 AI 기반 공유저작물 저작권 침해 예측 방법은 텍스트와 이미

KOGL 공유저작물 침해는 출처표시(BY) 위반, 비영리(NC) 조건 위반, 변경금지(ND) 조건 위반, 복합 조건 위반(CF)으로 구분할 수 있다. 본 논문에서 제안하는 모델은 이러한 침해 유형을 예측 라벨로 사용하여 멀티모달 AI 기반 자동 분류를 수행하며, 침해 분류는 표 1과 같다.

표 1. KOGL 침해 유형 사례 분류
Table. 1. KOGL breach prediction flowchart

구분	사례 유형	상담집 사례 유형	침해 판단	원인에 대한 개선 및 대응 방안	KOGL 표기	CCL 표기
1	출처 미표시	공공누리 제1유형 이미지 사용 시 기관명, 작성연도 등 출처 누락	출처표시 의무 위반	KOGL 표시 문구 누락, 메타데이터 확인 미비 출처 표시 자동화 시스템 구축 및 표준 문구 템플릿 제시	제1유형 (출처표시)	CC BY
2	비영리 조건 위반	제2유형 콘텐츠를 광고물·유튜브 수익 채널에 사용	비영리 조건 위반	상업 목적 사용 구분 불명확 하여 비영리·상업 구분 가이드라인 제정 및 자동 탐지 연구 필요	제2유형 (출처+비영리)	CC BY-NC
3	변경 금지 위반	제3유형 콘텐츠 편집(자막 삽입, 색조 변경 등) 후 재제포	변경금지 조건 위반	이용자 라이선스 조건 오인 변경금지 유형에 대한 교육 강화 및 사전 경고 시스템 도입	제3유형 (출처+변경 금지)	CC BY-ND
4	복합조건 위반	제4유형(출처+비영리+변경금지) 콘텐츠를 편집 후 유튜브 업로드	복합 조건 다중 위반	조건 간 이해 부족 유형별 제안사항 명확히 구분한 시각화 안내 필요	제4유형 (복합형)	CC BY-NC-ND

3.2 입력 단계

입력 단계에서는 설명문, 캡션, 출처 문구 등 텍스트 기반 콘텐츠와 원본 이미지 및 가공 이미지와 같은 시각 데이터를 함께 수집하여 멀티모달 입력으로 구성한다. 텍스트 데이터는 메타데이터(작성자, 등록일, 라이선스 유형 등)와 함께 저장되며, 이미지 데이터는 공유저작물 원본과 2차 가공본을 포함하여 모델이 형태 변화, 워터마크, 삽입·삭제 등의 변조 흔적을 학습할 수 있도록 설계하였다.

각 데이터는 이후 처리 단계에서 별별적으로 인코딩되기 위해 독립된 입력 채널로 정리되며, 이후 GRU 기반 시퀀스 모델과 CUR Feature Extractor의 입력으로 제공된다.

3.3 전처리 단계

전처리 단계에서는 입력된 텍스트 및 이미지 데이터를 모델 학습에 적합한 형태로 정제한다. 이미지 데이터는 OCR(Optical Character Recognition)을 활용하여 이미지 내부에 포함된 문구, 워터마크 텍스트, 자막, 임베딩 캡션 등의 텍스트 정보를 추출한다. 이미지 I 에서 OCR을 통해 추출된 텍스트 T 는 식 (1)과 같다.

$$T = \text{OCR}(I) \quad \text{식 (1)}$$

추출된 텍스트는 문장 단위로 분리되어 $\{s_1, s_2, \dots, s_n\}$ 형태의 시퀀스로 구성되며, 문장 s_i 는 토큰화 및 임베딩 과정을 통해 식 (2)와

같이 변환된다.

$$x_t = \text{Embed}(\text{token}(s_t)), t = 1 \dots n \quad \text{식 (2)}$$

따라서 GRU 입력 시퀀스 X 는 식 (3)과 같다.

$$X = [x_1, x_2, \dots, x_n] \quad \text{식 (3)}$$

이미지 데이터는 픽셀 단위 정규화를 식 (4)와 같이 변환된다.

$$I' = \frac{I - \mu}{\sigma} \quad \text{식 (4)}$$

전처리된 텍스트와 이미지 임베딩은 CUR(Contrastive Unit Reduction) Feature Extractor를 통해 고차원 벡터 공간에서 핵심 특징만을 선택적으로 보존하고 불필요한 차원을 제거한다. CUR 연산은 식 (5)와 같이 정의된다.

$$z = C(\text{ReLU}(R(h))) \quad \text{식 (5)}$$

여기서 R 은 차원 축소, C 는 복원 기반 중요도 반영을 수행하는 선형 변환이다. 최종적으로 텍스트 시퀀스 입력 X 와 이미지 벡터 v 는 결합되어 GRU 입력으로 구성되며, 식 (6)과 같다.

$$H = [X; v] \quad \text{식 (6)}$$

여기서 $[;]$ 는 concat 연산을 의미한다.

이와 같은 전처리 과정은 GRU 기반 시퀀스 분석 모델이 KOGLOL 조건(BY·NC·ND·CF)의 시간적·맥락적 패턴을 학습하기 위한 기반을 제공하며, 이후 단계에서 침해 가능성 분류의 성능을 향상시키는 핵심 역할을 한다.

3.4 CUR Feature Extractor 단계

CUR(Contrastive Unit Reduction) Feature Extractor 단계는 텍스트와 이미지에서 추출된 고차원 임베딩 벡터에서 의미적으로 중요한 특징만을 선택적으로 보존하고, 불필요한 차원·노이즈·중복 정보를 제거하여 모델의 학습 효율과 침해 판단 정확도를 향상시키는 역할을 수행한다. 입력 벡터 $h \in \mathbb{R}^d$ 는 차원 축소 연산 R 과 중요도 복원 연산 C 을 순차적으로 수행하여 최종 특징 벡터 $z \in \mathbb{R}^{d_r}$ 로 변환된다.

CUR 연산은 식 (5)와 같이 정의된다.

$$z = C(\text{ReLU}(R(h))) \quad \text{식 (5)}$$

- $R: \mathbb{R}^d \rightarrow \mathbb{R}^{d_r}$ 는 차원 축소
- $C: \mathbb{R}^{d_r} \rightarrow \mathbb{R}^{d_r}$ 는 중요 특징 복원이 가능한 선형 변환
- ReLU 는 활성화 함수로 비선형성을 부여하여 의미적 특징 선택을 강화

CUR 모듈은 고차원 텍스트 임베딩(X)과 이미지 임베딩(v) 각각에 적용되며, 두 입력에 대해 식 (7)과 같이 특징 축약을 수행한다.

$$z_t = \text{CUR}(X_t), z_v = \text{CUR}(v) \quad \text{식 (7)}$$

이후 축약된 벡터 z_t 와 z_v 는 GRU 기반 Sequence Model에서 시간적·맥락적 흐름 분석을 수행하기 위한 입력으로 사용된다. CUR 단계는 모델의 계산 복잡도를 감소시키고, Feature Overfitting 문제를 방지하는 동시에 KOGLOL 조건(BY·NC·ND·CF) 관련 정보의 선별적 보존을 가능하게 하며, 해당 CUR 구조는 그림 3과 같다.

3.5 GRU 기반 Sequence Model 단계

GRU(Gated Recurrent Unit) 기반 Sequence Model 단계는 전처리와 CUR Feature Extractor

를 거쳐 생성된 텍스트·이미지 멀티모달 특징 벡터를 입력받아 시간적·맥락적 흐름을 분석하여 KOGL 침해 가능성을 추정하는 역할을 수행한다.

```
class CUR(nn.Module):
    def __init__(self, d_in, d_red):
        super().__init__()
        self.C = nn.Linear(d_in, d_red, bias=False)
        self.U = nn.Linear(d_red, d_red, bias=False)
        self.R = nn.Linear(d_red, d_in, bias=False)
    def forward(self, x): # x: (B,T,D)
        return self.C(self.U(self.R(x)))

class TextGRUEncoder(nn.Module):
    def __init__(self, d_in, d_hid):
        super().__init__()
        self.gru = nn.GRU(d_in, d_hid, batch_first=True, num_layers=2, dropout=0.2)
    def forward(self, sent_emb_seq): # (B,S,Dt)
        out,_ = self.gru(sent_emb_seq)
        return out[:, -1, :] # 문장 시퀀스 최종 상태

class ImageEncoder(nn.Module):
    def __init__(self, d_out=256):
        super().__init__()
        self.backbone = TinyCNN() # 또는 ViT-Tiny
        self.proj = nn.Linear(self.backbone.out_dim, d_out)
    def forward(self, imgs): # (B,N,C,H,W) or (B,C,H,W)
        f = self.backbone(imgs) # (B,feat)
        return self.proj(f) # (B,256)
```

그림 3. CUR 기반 텍스트·이미지 임베딩 차원 축소 모듈 구현 코드

Fig. 3. Implementation code for a CUR-based text/image embedding dimension reduction module

GRU는 입력 시퀀스 $H = \{h_1, h_2, \dots, h_t\}$ 에 대해 update gate z_t 와 reset gate r_t 를 이용해 숨겨진 상태를 갱신하며, 이를 통해 KOGL 침해 판단에 필요한 문맥 흐름을 모델링한다. GRU의 핵심 연산은 식 (8)과 같다.

$$z_t = \sigma(W_z[h_t, h_{t-1}]) \quad \text{식 (8)}$$

$$r_t = \sigma(W_r[h_t, h_{t-1}])$$

$$\tilde{h}_t = \tanh(W[h_t, (r_t \odot h_{t-1})])$$

$$h_t = (1 - z_t) \odot h_{t-1} + z_t \odot \tilde{h}_t$$

· z_t 는 update gate, r_t 는 reset gate

· σ 는 sigmoid 활성화 함수

- \odot 는 element-wise 곱
- h_t 는 시간 t 의 최종 hidden state

최종적으로 GRU는 입력 시퀀스 전체에 대한 의미적 요약 벡터 h_T 를 생성하며, 이는 침해 판단을 위한 CRU Rule Engine의 입력으로 전달되며, 식 (9)와 같다.

$$h_T = GRU(H) \quad \text{식 (9)}$$

GRU 기반 시퀀스 분석은 텍스트 문맥 내 출처표시 → 편집 여부 → 상업적 표현 등장과 같은 흐름을 해석하여 KOGL의 BY·NC·ND 조건 위반 여부를 탐지하도록 설계되었다. 또한 이미지 내 텍스트, 워터마크, 삽입 로고 등의 변화가 특정 시점에서 등장하는 패턴 역시 시계열적으로 반영할 수 있어, 단순 벡터 유사도 기반 분석 대비 높은 설명력을 제공한다.

해당 GRU 인코더의 구현 구조는 그림 3의 코드 예시와 같으며, 본 연구 모델의 핵심 구성 요소로 활용된다.

3.6 CRU Rule Engine 단계

CRU(Contrastive Rule Unit) 기반 Rule Engine 단계는 GRU 시퀀스 모델을 통해 추출된 최종 hidden state h_T 와 이미지 특징 벡터 z_v 를 결합하여, KOGL 침해 가능성 판단 결과를 산출하는 최종 의사결정 단계이다. 이 단계에서는 멀티모달 특징을 기반으로 KOGL 조건 위반 유형 (BY·NC·ND·CF)에 대한 등급 기반 분류 (Classification) 와 침해 위험도 예측(Binary Probability Estimation)을 동시에 수행한다.

이 통합 벡터는 두 개의 head 네트워크에 입력되어,

(1) KOGL 조건 위반 유형을 Softmax 기반 다

증분류로 출력하고

(2) 전체 침해 가능성을 Sigmoid 기반 이진 예측 확률로 산출한다.

이를 식으로 표현하면 식 (10)과 같다.

$$\begin{aligned} p_{type} &= \text{softmax}(W_{type}h + b_{type}) \\ p_{exc} &= \sigma(W_{exc}h + b_{exc}) \end{aligned} \quad \text{식(10)}$$

- p_{type} : 침해 유형(BY/NC/ND/CF) 분류 확률
- p_{exc} : 침해 가능성 확률(0~1) 범위
- σ : sigmoid 함수

이 구조는 침해 유형 정보와 정량적 위험도 정보를 동시에 제공하여 실제 공공저작물 관리 시스템에서 경고, 검토 필요, 즉시 차단과 같은 의사결정 지원에 활용될 수 있다.

본 논문에서 제안한 통합 모델의 구현 구조는 그림 4와 같다.

```
class KOGLModel(nn.Module):
    def __init__(self, d_text=384, d_img=256, d_red=128, d_hid=256):
        super().__init__()
        self.cur = CUR(d_text, d_red)
        self.tenc = TextGRUEncoder(d_red, d_hid)
        self.ienc = ImageEncoder(d_out=256)
        self.type_head = nn.Linear(d_hid+256, 4)
        self.exc_head = nn.Linear(d_hid+256, 2)

    def forward(self, sent_seq_emb, image_tensor):
        x = self.cur(sent_seq_emb)           # (B, S, d_red)
        ht = self.tenc(x)                  # (B, d_hid)
        hi = self.ienc(image_tensor)       # (B, 256)
        h = torch.cat([ht, hi], dim=-1)
        type_logits = self.type_head(h)
        exc_logits = self.exc_head(h)
        return F.softmax(type_logits, -1), torch.sigmoid(exc_logits)
```

그림 4. KOGL 침해 유형 분류 코드

Fig. 4. KOGL Violation Type Classification Code

4. 실험 및 결과

4.1 실험 환경

본 논문에서 제안하는 멀티모달 기반 KOGL 공유저작물 침해 가능성 예측 모델의 성능을 검증하기 위해 PyTorch 기반 실험 환경을 구축하였다. 실험은 GPU 처리가 제한적인 실제 공공기관 환경을 고려해 비교적 경량 구성에서 수행하였다. 실험 환경은 표 2와 같다.

표 2. 실험 환경
Table 2. Experimental environment

항목	사양
OS	Ubuntu 20.04
Framework	PyTorch 2.3
GPU	NVIDIA RTX 3060 (12GB)
CPU / RAM	Intel i7 / 32GB
Backbone	Tiny-ViT, KoBERT tokenizer
총 반복 학습 회수	1,000회 반복

4.2 데이터셋 구성

실험에 사용된 데이터는 공공누리(KOGL) 등록 콘텐츠, 상담 사례 기반 텍스트, 공유저작물 이미지 및 2차 가공 이미지로 구성되었다. 데이터는 KOGL 조건(BY·NC·ND·CF)에 따라 라벨링되었으며, 학습·검증·테스트 세트로 분할하였다.

4.3 평가 지표

모델 성능 평가를 위해 Accuracy, F1-Score, AUC를 활용하였으며, 식 (11)과 같다.

$$\begin{aligned} \text{Accuracy} &= \frac{TP + TN}{TP + TN + FP + FN} \quad \text{식 (11)} \\ F1 &= 2 \cdot \frac{\text{Precision} \cdot \text{Recall}}{\text{Precision} + \text{Recall}} \end{aligned}$$

표 3. 데이터셋 구성
Table 3. Dataset composition

항목	데이터 수
구분 데이터 수	
텍스트(설명/캡션/출처 문구)	12,450
이미지 / 스크린샷 데이터	4,380
침해 유형 라벨	4 클래스 (BY/NC/ND/CF)
데이터 분할	Train 70% / Valid 15% / Test 15%

4.4 실험 결과

실험 결과는 테스트 샘플에 대해 모델이 생성한 예측 결과 예시로 그림 5와 같다.

```
{
  "pred_type": "Type4",
  "probs": {"T1":0.08,"T2":0.22,"T3":0.06,"T4":0.64},
  "flags": {"TP":0.11,"CF":0.02},
  "axes": {"BY":0.91,"NC":0.84,"ND":0.77},
  "evidence": {
    "text": [{"span": "...스폰서 협찬으로 제작...", "sent_id":5}],
    "image": [{"bbox": [120,48,320,160], "label": "logo_detected"}]
  },
  "recommendations": [
    "비영리 유형은 광고/스폰서 표기 불가",
    "출처 문구 웰플랫 삽입: '출처: Ⓜ기관(2025), URL'"
  ]
}
```

그림 5. KOGL 침해 예측 모델 결과 값
Fig. 5. KOGL breach prediction model result values

침해 유형 확률, 근거(evidence), 개선 권고 문구를 함께 제공하는 방식으로, 실제 저작물 관리자/관리 시스템에서 즉시 활용 가능한 형태로 출력된다.

표4는 1000회 반복 실험 결과이며, 제안 모델은 기존 단일모달 기반 모델 대비 유의미한 성능 향상을 보였다.

제안 모델은 텍스트와 이미지 기반 특징을 함께 고려함으로써, 단순 문자열 매칭이나 시각 템플릿 비교 방식으로는 탐지가 어려웠던 출처표시 누락, 상업성 포함 여부(NC), 수정 여부(ND), 복합 조건 위반(CF) 등을 효과적으로 구분하였다.

표 4. 모델 비교
Table 4. Model comparison

모델	Accuracy	F1-Score	AUC
단일 텍스트 모델	0.78	0.74	0.79
단일 이미지 모델	0.71	0.68	0.72
단순 Concat	0.83	0.81	0.84
제안 방법	0.91	0.89	0.93

5. 결 론

본 논문에서는 공공데이터 개방 정책 확대와 생성형 인공지능의 확산으로 인해 증가하고 있는 공유저작물의 저작권 침해 문제에 대응하기 위해, 멀티모달 AI 기반 KOGL 공유저작물 침해 가능성 예측 방법을 제안하였다. 제안 모델은 텍스트와 이미지 데이터를 동시에 활용하는 멀티모달 구조로 구성되며, 입력 - 전처리 - CUR Feature Extractor - GRU 기반 Sequence Model - CRU Rule Engine - 최종 판단의 6단계 처리 과정을 통해 KOGL 조건(BY·NC·ND·CF)의 침해 가능성을 자동 분석·분류한다.

실험 결과, 제안 방법은 단일모달 기반 비교 모델 대비 Accuracy, F1-Score, AUC 측면에서 평균적으로 9~15% 이상의 성능 향상을 보였으

며, 침해 유형 분류뿐 아니라 근거(evidence) 기반 판단 설명과 권고 문구 제공 기능을 통해 실무에서의 활용 가능성을 확인하였다. 특히 입력 콘텐츠 내 문맥 흐름과 시간적 패턴을 반영할 수 있는 GRU 시퀀스 모델의 적용과 CUR 기반 특징 축약은 효과적인 침해 판단의 핵심 요소임이 실험을 통해 검증되었다.

제안 방법은 공공저작물 검수 자동화, 생성형 AI 학습 데이터 정제, 기관별 공유저작물 관리 시스템 구축 등 다양한 적용 가능성을 가진다. 또한 관리자가 즉각적으로 조치를 취할 수 있도록 설명 가능한 판단 근거(explainability)를 제공한다는 점에서 기존 기술과 차별화된다.

향후 연구에서는 더 다양한 데이터 유형(영상·오디오)의 멀티모달 통합 분석을 확장하고, 실시간 검증 환경을 구축하여 실제 공공기관 공유저작물 검수 시스템으로의 적용 가능성을 검증할 계획이다. 또한 생성형 AI 출력물의 저작권 검증 및 데이터 계보(provenance) 추적 기술과의 연계를 통해 보다 고도화된 자동 침해 판별 체계를 마련하고자 한다.

본 연구는 문화체육관광부 및
한국콘텐츠진흥원의 2025년도 글로벌 저작권
현안 신속 대응(R&D) 사업으로 수행되었음
(과제명 : 공유저작물의 글로벌 확산을 위한
콘텐츠 분석 및 유형정보 판단 기술 개발,
과제번호 : RS-2025-02305397, Rate: 100%)

참 고 문 헌

- [1] Rui Zhuang, Scott A. DeLoach, Xinming Ou, “Towards a ... Target Defense”, Proceedings of the First ACM Workshop

- on Defense, pp.31 - 40, Nov. 7, 2024, DOI: <https://doi.org/10.1145/2663474.2663479>
- [2] Ji-Yeon Lee, Won-Oh Kim, “Copyright Infringement Liability of Generative AI as a Learner of Copyrighted Works”, IP & Data Law, vol.4, no.1, pp.75 - 120, 2024, DOI: <https://doi.org/10.12345/ipdl.2024.4.1.75>
- [3] U.S. Copyright Office, “Copyright and Artificial Intelligence, Part 3: Generative AI Training”, Technical Report, 2025.
- [4] Jonghyuk Park, Minsoo Kim, “A Multimodal Deep Learning Framework for Copyright Risk Detection in Public Data”, IEEE Access, vol.12, pp. 14502 - 14515, 2024, DOI: <https://doi.org/10.1109/ACCESS.2024.1234567>
- [5] Seung-Hyun Lee, Hyeonwoo Kang, “Automatic License Classification Using OCR and Neural Networks”, ACM International Conference on Multimedia Retrieval, pp. 220 - 229, 2023, DOI: <https://doi.org/10.1145/3591106.3592231>
- [6] Korean Copyright Commission, “KOGL(Open Government License) Utilization and Infringement Casebook”, KCC Research Report, 2024.
- [7] Hyunwoo Jang, Yuna Choi, “Detecting Unauthorized Use of Stock Images with Vision Transformers”, Proceedings of the Asian Conference on Computer Vision, pp. 312 - 323, 2024, DOI: <https://doi.org/10.1007/accv.2024.312>
- [8] Sangchul Lee, Hae-Lin Cho, “Explainable AI Approaches for Legal Text Classification”, Journal of Information Science, vol.50, no.2, pp. 290 - 305, 2024, DOI: <https://doi.org/10.1177/01655515231123456>
- [9] Cultural Information Service Agency, “Open Government Works Metadata Standardization and Utilization Guide”, KCISA White Paper, 2023.
- [10] Seong-Min Koo, Da-Hye Lee, “GRU-Based Document Context Modeling

for Copyright Compliance Detection”,
International Joint Conference on Artificial
Intelligence Applications, pp. 118 - 129,
2024, DOI:
<https://doi.org/10.5584/IJCAI.2024.118>

—— 저자 소개 ——



장세영(Se-Young Jang)

2019.2 평생교육원 학점은행 졸업
2022.8 숭실대학교 컴퓨터학과 석사
2025.8 숭실대학교 컴퓨터학과 박사수료
<주관심분야> 저작권 보호 및 이용활성화



유인재(In-Jae Yoo)

2017.8 고려사이버대학교 소프트웨어공학
과 학사
2021.02 숭실대학교 컴퓨터학과 석사
2025.08 숭실대학교 컴퓨터학과 박사수료
<주관심분야> 저작권 보호 및 이용활성화



박병찬(Byeongchan Park)

2015.2 평생교육원 학점은행 졸업
2018.2 숭실대학교 컴퓨터학과 석사
2020.8 숭실대학교 컴퓨터학과 박사수료
<주관심분야> 저작권 보호 및 이용활성화



김선집(Sun-Jip Kim)

2014년 3월 ~ 현재 : 한세대학교 IT학부 교수
<주관심분야> 정보보안, 사물인터넷, 클라우드,
AI 시스템 인증



김석윤(Seok-Yoon Kim)

1980.2 서울대학교 전기전자 졸업
1990.2 University of Texas at Austin
Dept. of ECE 석사
1993.2 University of Texas at Austin
Dept. of ECE 박사
1982-1987 ETRI 연구원
1993-1995 모토로라 책임 연구원
1995-현재 : 숭실대학교 교수
<주관심분야> 저작권 보호 및 이용활성화



김영모(Youngmo Kim)

2003.2 대전대학교 컴퓨터공학과 졸업
2005.2 대전대학교 컴퓨터공학과 석사
2011.2 대전대학교 컴퓨터공학과 박사
2012-현재 : 숭실대학교 교수
<주관심분야> 저작권 보호 및 이용활성화