

Improved Piracy Site Detection Technique using Search Engine

Eui-Jin Kim¹, Deuk-Hun Kim² and Jin Kwak^{3*}

¹ ISAA Lab., Department of Cyber Security
Ajou University, Suwon, Republic of Korea
[e-mail: dmlwls0403@ajou.ac.kr]

² ISAA Lab., Institute for Information and Communication
Ajou University, Suwon, Republic of Korea
[e-mail: dhkim.isaa@gmail.com]

^{3*} Department of Cyber Security
Ajou University, Suwon, Republic of Korea
[e-mail: security@ajou.ac.kr]

*Corresponding author: Jin Kwak

*Received March 16, 2022; revised May 20, 2022; accepted June 16, 2022;
published July 31, 2022*

Abstract

With the increase in copyright content exports to overseas markets due to the recent globalization of the Korean culture, the added value of the Korean digital content market is increasing at a significant rate. As such, as the size of the copyright market increases, different piracy sites have emerged that generate profits by illegally distributing works without the permission of the copyright holders, resulting in direct and indirect damage to these copyright holders. The existing copyright detection methods used in public institutions for solving this problem are limited, while the piracy sites are ever-changing. Methods are being continuously developed to achieve better detection results. To this end, it is possible to detect the latest infringement site domain by detecting the infringement site domain that is constantly changed through the search engine. This paper proposes an improved piracy site detection method using a search engine to prevent the damage caused by piracy sites.

Keywords: Copyright Detection, Search Engine, Piracy Site, Copyright Infringement

1. Introduction

Due to the recent globalization of the Korean culture, copyright contents are being exported to overseas markets, and the overseas copyright market is continuously increasing in size [1]. With this, various piracy sites have appeared, and these sites generate profits by illegally distributing works of the legal sites without the permission of the copyright holder. The illegal distribution of such copyrighted works on piracy sites causes indirect as well as direct damage to the copyright holder [2], which can cause a downturn in the copyright industry.

Several studies are underway to prevent the damage caused by piracy sites, including one [3,4] that suggests a method for detecting and determining whether copyright infringement has occurred based on the characteristics of the copyright infringement sites [5,6]. However, additional research is needed to clearly detect the infringement because web page composition is similar for both the piracy site and the normal site, and as piracy sites that do not have the characteristics of a copyright infringement site appear. In addition, although public institutions are making efforts to block piracy sites to prevent copyright infringement, the detection speed is slow compared to the speed at which copyright infringement sites are created, and the URL of the copyright infringement sites is continuously maintained during the process of blocking. Further, it must be noted that limitations such as continuously changing URLs continues to pose a challenge [7].

Therefore, to detect intelligent piracy sites that bypass existing detection techniques, this paper uses a search engine in real time to crawl a content list comprising names and types of content to construct a dataset. In addition, a technique for determining whether a site suspected of copyright infringement is a piracy site is presented based on the existence of common features between piracy sites and legally operating sites.

In Chapter 2, the characteristics and limitations of the existing techniques for copyright detection of piracy sites were analyzed. In addition, the F1 score and confusion matrix were analyzed. In Chapter 3, the dataset built using suspected copyright infringement sites crawled in real time was analyzed based on the content list and search engine. In Chapter 4, an advanced copy right detection method using a proposed search engine is explained. In Chapter 5, the detection results of the proposed method are analyzed, and finally, Chapter 6, the conclusions are presented.

2. Related Work

In this chapter, F1 score, confusion matrix, the features and limitations of the infringement site detection techniques based on feature are analyzed.

2.1 F1 Score and Confusion Matrix

The F1 score is a statistical measure of accuracy that applies a confusion matrix configured as shown in **Table 1** [8,9]. The confusion matrix is used to derive precision and recall based on true positive (TP), true negative (TN), false positive (FP), and false negative (FN) predictions. A TP reflects a correctly predicted truth, and an FP reflects an incorrectly predicted truth. A TN reflects a correctly predicted falsehood, and an FN reflects an incorrectly predicted falsehood.

Table 1. Confusion matrix format

		Truth Condition	
		True	False
Predicted Condition	True	TP	FP
	False	FN	TN

$$Accuracy = \frac{TP+TN}{TP+FN+FP+TN} \quad (1)$$

$$Precision = \frac{TP}{TP+FP} \quad (2)$$

$$Recall = \frac{TP}{TP+FN} \quad (3)$$

$$F1\ Score = 2 \times \frac{Precision \times Recall}{Precision + Recall} \quad (4)$$

Precision is derived using equation (2), and recall is derived using equation (3). The F1 score is derived as the harmonious average of equations (2) and (3), and from this, the accuracy of the proposed technique can be measured. Precision refers to the ratio of TPs to all TPs and all FPs, and recall refers to the ratio of TPs to all TPs and all FNs; hence, they complement each other and tell a more complete story of accuracy. Accuracy (equation (1)) refers to the ratio of all TPs and TNs to all TPs, TNs, FPs, and FNs, which better indicates the full-spread accuracy of the tool. Accuracy is then used for comparative analysis with existing studies [3].

2.2 Piracy Site Detection Techniques

This section explains existing techniques for detecting piracy sites based on their characteristic features.

2.2.1 Feature Analysis of Detection Techniques for Piracy Sites

The 'Feature Analysis of Detection Techniques for Piracy Sites' was proposed by Choi et al., and as the title suggests, this method analyzes the characteristics of piracy sites using the proposed detection method.

In this paper, the types of digital content infringed on piracy sites as well as the size and method of infringement were analyzed. The torrent, video streaming, and webtoon sites that are representative of the copyright infringement sites were also studied. In addition, an algorithm for the piracy site was proposed. The detection process proposed in this study can be described as follows.

First, after extracting the advertisement of the input site, the characteristics of the advertisement banner are analyzed, and if illegality is included in the analyzed characteristics, the content menu of the input site is analyzed by the piracy site type. If the characteristics of the infringing site for each analyzed type exists in the input site, the input site is determined as a piracy site. A confusion matrix-based accuracy was used to evaluate the performance of the copy right detection method. Based on the values obtained for the performance indicators for 114 sites that consist of 57 piracy sites and 57 normal sites to 1:1 ratio, an accuracy of approximately 93% was obtained, as reported in Table 2 [3].

Table 2. Results of Choi's Research

Category	TP	FN	FP	TN	Accuracy	Precision	Recall	F1 Score
Torrent	23	0	2	21	0.956521	0.92	1	0.958333
Video Streaming	13	1	1	13	0.928571	0.928571	0.928571	0.928571
Webtoon	20	0	4	16	0.9	0.833333	1	0.909091
Total	56	1	7	50	0.929824	0.888888	0.982456	0.933333

Using proposed detection method, an experiment using Python programming language was conducted; however, there was no specific mention of the detection method and the module used. In addition, there were limitations in terms of the dataset preparation process and the proposed technique was also not verified.

2.2.2 Intelligent Piracy Site Detection Technique with High Accuracy

The method of 'Intelligent Piracy Site Detection Technique with High Accuracy,' which was proposed by Kim et al., detects and determines whether or not a violation exists based on the characteristics of the piracy site.

In this method, to detect a site suspected of copyright infringement, it is first determined whether the site entered in Step 1 has already been detected as an infringing site by a public institution in advance, and whether or not a similar feature exists in the normal site. In Step 2, after checking whether the business registration number exists only in the normal site, it is determined whether it is a normal site. In addition, the keywords characterize torrent, video streaming, and webtoon infringement sites are listed, and whether or not an infringement has occurred is determined using keyword detection. Finally, in Step 3, after taking a screenshot of the main page of the entered site, it is determined using the Google Vision API whether an advertisement banner on the main page is available. A confusion matrix-based F1 score was used to evaluate the performance of the copyright detection method and the performance indicators for 314 sites that consist of 157 piracy sites and 157 normal sites to 1:1 ratio, were evaluated. Consequently, the F1 score showed an accuracy of approximately 92.9% or more, as presented in [Table 3 \[4\]](#).

Table 3. Results of Kim's Research

Category	TP	FN	FP	TN	Accuracy	Precision	Recall	F1 Score
Torrent	33	1	2	32	0.955882	0.942857	0.970588	0.956522
Video Streaming	64	7	5	66	0.915493	0.927536	0.901408	0.914286
Webtoon	47	5	2	50	0.932692	0.959184	0.903846	0.930693
Total	144	13	9	138	0.927632	0.941176	0.917197	0.929032

However, there were limitations with regard to the configuration of the dataset as well as the method used for calculating the confusion matrix for the infringing sites. Further, a detailed description of the aspects mentioned above was not presented. Therefore, since the existing studies that have presented methods for detecting piracy sites do not mention the dataset configuration details, they cannot be applied to every practical scenario, given the ever-changing nature of piracy sites.

3. Dataset for the Proposed Technique

This chapter describes the datasets employed in this proposed technique. Two datasets were used. The first is the content list that is composed of the content name and content type as presented in **Table 4**.

Table 4. Examples from the content list

Number	Content Name	Content Type
1	Video's name 1	Video
...
10	Video's name 10	Video
11	Video's name 1	Torrent
...
20	Video's name 10	Torrent
21	Webtoon's name 1	Webtoon
...
30	Webtoon's name 10	Webtoon

The content list is divided based on the content name and content type, and the content types include webtoons, videos, and torrents. The video type is used to find video streaming sites that infringe video copyrights, and the torrent type to find torrent sites that infringe video copyrights. Likewise, the webtoon type is used to find illegal webtoon sites that infringe webtoon copyrights by illegally posting webtoons. This content list contains 'search keywords' that can be used for crawling the search engine. Note that each 'search keyword' is a combination of the 'Content Name' and 'Content Type;' for example, 'review.'

The second dataset is created by crawling the site URLs from the search engine's search results into the Input Site List. Among the input sites crawled using the search keywords, it is determined whether the actual copyright has been violated, and as presented in **Table 5**, 39 torrent sites, 86 video streaming sites, and 25 webtoon sites were derived. In this paper, to evaluate the performance of the proposed method, the normal site and the site that infringes the copyright are in a 1:1 ratio. A confusion matrix is used to evaluate the performance of the proposed technique.

Table 5. Number of input sites

Category	Number of Suspected Piracy Sites	Number of Normal Sites	Total
Torrent	39	39	78
Video Streaming	86	86	172
Webtoon	25	25	50
Total	150	150	300

4. Improved Piracy Site Detection Technique Using the Proposed Search Engine

This chapter describes the improved method for the copy right detection of piracy sites using the proposed search engine.

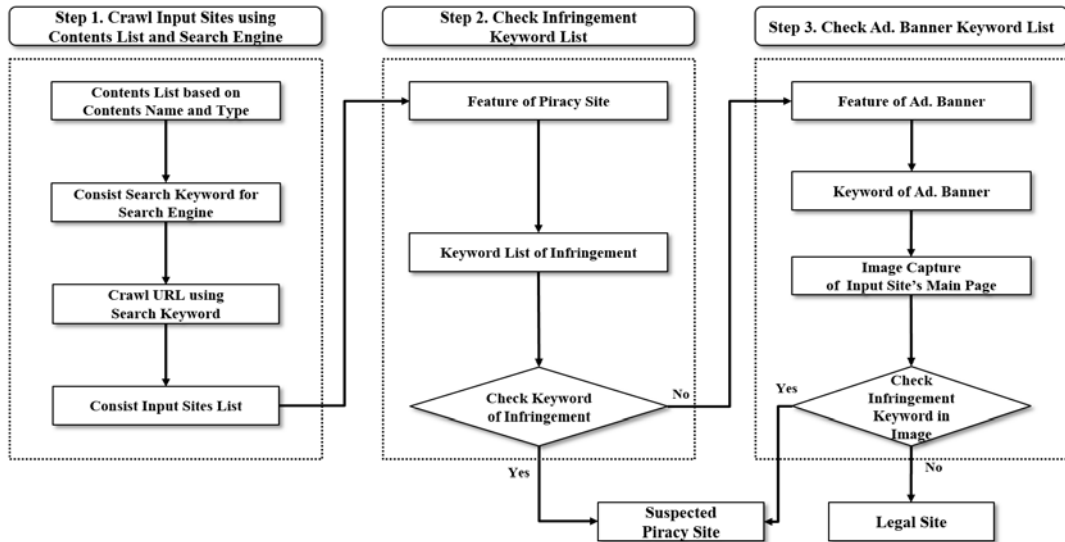


Fig. 1. Improved piracy site detection technique based on the proposed search engine

Step 1. Crawl Input Sites using Content List and Search Engine

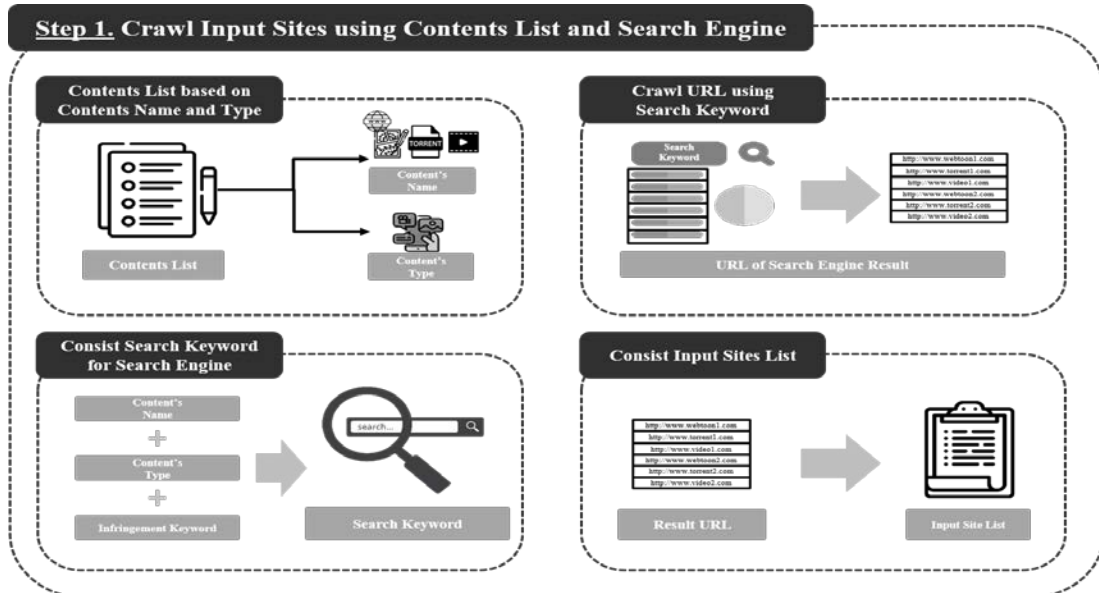


Fig. 2. Step 1 process of our proposed technique

Table 6. Psuedo code for crawling input sites using the content list and search engine

Function for check URL using white list:	
1	contents_list ← It is predefined content list dataset in chapter 3
2	contents_name, contents_type ← Component of content list
3	infringement_keyword ← It contains infringing keywords such as ‘free view’ and ‘replay’
4	search_keyword ← It consists ‘contents_name’, ‘contents_type’ and ‘infringement_keyword’
5	driver.get(search_keyword) : It searches search_keyword on search engine
6	str_html = soup.select(‘specific tag’) : The URL is searched by selecting a ‘specific tag’ of the search engine
7	input_sites_list.csv ← It saved str_html in csv

This process involves finding sites that may have undergone copyright infringement based on the contents of the content list. It involves searching for a search keyword in the search engine and then composing a list of URLs. In this manner, as mentioned in Chapter 3, the input site list is constructed in the search engine using the information in the content list and the infringing keywords.

First, for constructing a content list, search keywords are created by adding infringing keywords such as ‘review’ and ‘free view’ to the content names and content types existing in the content list. Using the composed URLs from the search engine as the input sites, the page sources for the input sites are dynamically retrieved using the Python module, “Selenium.” After crawling a URL located in a specific tag of the page source of the search engine, it is saved in the “input_sites_list.csv” format and listed. The listed input file is of the format presented in **Table 7**. The csv file is then delivered to Step 2.

Table 7. Examples of input_sites_list.csv

Number	Type	URL
1	Torrent	http://www.torrent1.com
2	Video	http://www.video1.com
3	Webtoon	http://www.webtoon1.com
...
N

Step 2. Check Infringement Keyword List

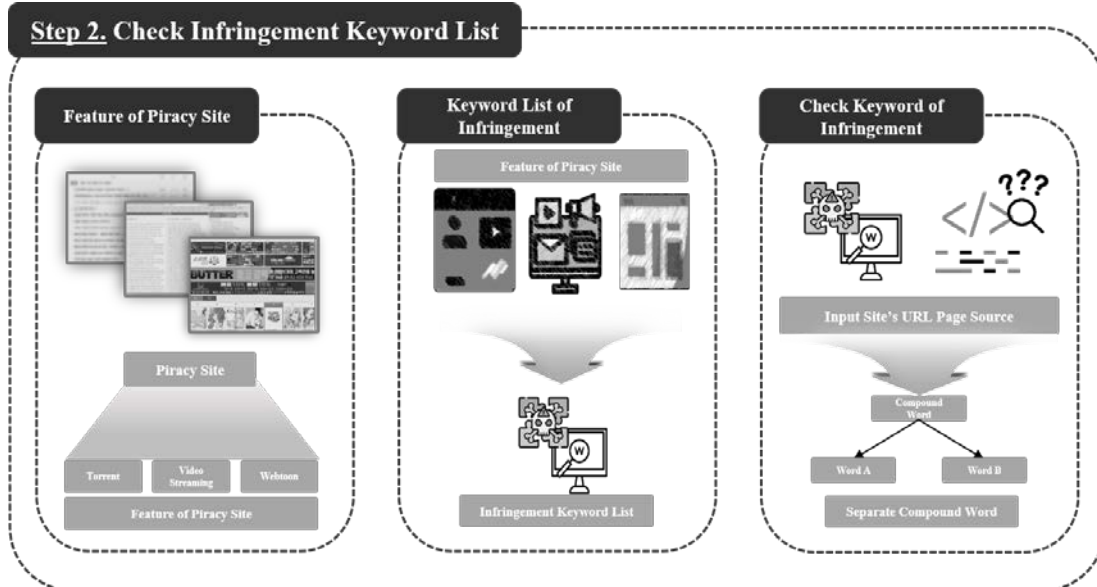


Fig. 3. Step 2 process of our proposed technique

Table 8. Psuedo code for checking the infringement keyword list

Function for inspection of included keyword:	
1	input_sites_list.csv ← It is websites that has been crawled and was delivered from Step 1
2	Keyword_list ← It is keyword list that often used in the piracy site
3	compoundKeyword ← Array to save after separating the 'compound keyword'
	FOR keyword in Keyword_list:
4	IF the keyword is matched using the regular expression, a string containing the keyword is extracted
5	Separate the extracted strings with 'Segment' module
6	After separating the string, if string contains the keywords save in compoundKeyword
7	If keyword in IF compoundKeyword exists, RETURN 0
8	RETURN -1

This process involves determining whether an infringement keyword exists for each infringement site of the input page source. In other words, by extracting the characteristics of the infringement site and checking the existence of an infringement keyword, this process determines whether an infringement has occurred.

First, the characteristics of the piracy site are analyzed by type. Infringement keywords that mainly appear on the infringement site are extracted based on the characteristics of the analyzed piracy site. When removing the compound word containing the extracted infringement keyword, the compound word is separated using the Python module, 'Segment,' and then the existence of the pre-extracted infringement keyword is reaffirmed. If an infringement keyword exists, it is judged as a suspected infringement site, otherwise Step 3 is analyzed. Note that the piracy site changes the existing keywords or generates new keywords to cope with site advancement and blocking. Table 9 presents the infringement keyword list for each type of piracy site.

Table 9. Infringement keyword list by infringement site

Category	Keyword list	Number of Keyword
Torrent	“torrent,” “magnet,” “seed,” “trt,” etc.	15
Video Streaming	“HDVid,” “Evoload,” “UpStream,” “FlashVid,” “Jawcloud,” etc.	10
Webtoon	“BL,” “GL,” etc.	12

Step 3. Check Ad. Banner Keyword List

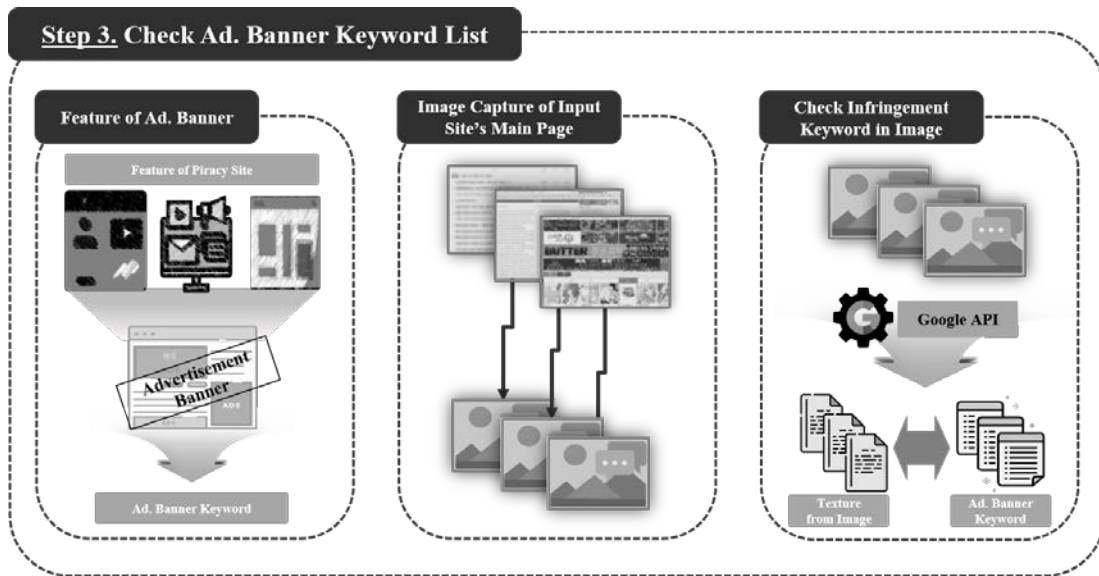


Fig. 4. Step 3 process of our proposed technique

Table 10. Pseudo code for check Ad. Banner keyword list

Function for Check Ad. Banner Keyword List:
1 Ad_keyword ← keywords inside the advertising banner
2 img ← A webpage image which has been extracted using selenium module
3 img_text ← Extract text in img using Google Vision API
4 FOR keyword in Ad_keyword:
5 RETURN 0 if keyword matches img_text via IF regular expression
6 RETURN -1

In this process, the words with a high usage frequency from the piracy site and illegal advertisement banners are extracted, the piracy site keyword list and advertisement banner keywords are generated, and the existence of the keywords are checked. The image is secured by capturing the main page of the entered site using the web driver. The obtained image is used to extract text. Using the OCR function provided by the Google Vision API [9], the text from the input site's main page is extracted from the obtained image. Because it is difficult to detect the information in the advertisement banner image by analyzing the page source for the text in the illegal advertisement banner existing on the piracy site, by extracting the text existing in the website main page image, the information in the page source text present in the

advertisement banner can also be secured. The text extracted from the main page image is compared with a predefined list of illegal advertisement banner keywords. If the advertisement banner keyword exists in the list, it is judged as a site suspected of infringing; if not, the keyword infringing copyright is analyzed. **Table 11** shows a list of keywords that are frequently used in illegal advertisement banners.

Table 11. Keyword list of advertisement banners

Category	Keyword List	Number of Keyword
Advertisement	“Jackpot,” “Casino,” “Toto,” “Proto,” “Porn,” “Unover,” etc.	10

5. Experimental Analysis

In this chapter, a confusion matrix is derived to measure the performance of the proposed method in Chapter 4, and the accuracy of the proposed method is measured with the F1 score through the derived confusion matrix. The reason for measuring the accuracy of the proposed technique using F1 score is that Accuracy can be measured high if the number of categories in the dataset is similar, but F1 score reflects accuracy when the dataset configuration is different [11]. In this paper, F1 score is used as a criterion for measuring the accuracy of the proposed technique because the dataset is different from 39 torrent sites and 86 video streaming sites to 25 webtoon sites. **Table 5** of Chapter 3 shows the configuration of the data set used to measure the proposed technique in this paper, and the conditions and results for accuracy measurement are as follows.

5.1. Experimental environment setup

Table 12 shows the experimental environment for the search engine-based improved infringing site copy right detection method proposed in this paper. Windows 10 Pro 64bit was used for the operating system. 64GB of RAM was used, and Python 3.8.0 version was the programming language. In addition, the Touch VPN 1.0.22 application was used to bypass the piracy site access blocking.

Table 12. Experimental setup

Component	Specification
Operating System	Windows 10 Pro 64bit
RAM	64GB
Programing Language	Python v.3.8.0
VPN	Touch VPN 1.0.22

5.2. Performance Technique Performance Analysis

To measure the accuracy of the proposed method in Chapter 4, this section describes the results of repeating the proposed method three times with 300 datasets.

5.2.1. Analysis for Torrent Sites

The results of repeating the method three times for a total of 78 datasets with 39 suspected torrent infringement sites and 39 normal sites in a 1:1 ratio are as follows. Accuracy was consistent with the true condition at 0.970086, and the precision of predicting suspected piracy sites was 0.974138. Recall, which indicates that it was predicted as a suspected infringement site among actual infringement sites, was 0.965812. From this, the final F1 score was measured at 0.969957, meaning that the overall accuracy of the proposed method with torrent sites was about 97.0%. **Table 13** shows the analysis results for the torrent sites.

Table 13. Result of torrent site

Number of Repetitions	TP	FN	FP	TN	Accuracy	Precision	Recall	F1 Score
1	38	1	1	38	0.974359	0.974359	0.974359	0.974359
2	38	1	1	38	0.974359	0.974359	0.974359	0.974359
3	37	2	1	38	0.961539	0.973684	0.948718	0.961039
Total	113	4	3	114	0.970086	0.974138	0.965812	0.969957

5.2.2. Analysis for Video Streaming Sites

The results of repeating the method three times on 172 datasets with 86 suspected video streaming piracy sites and 86 normal sites in a 1:1 ratio are as follows. Accuracy was consistent with the true condition at 0.918605, and the precision of predicting suspected piracy sites was 0.888489. Recall, which indicates that it was predicted as a suspected infringement site among actual infringement sites, was 0.957364. From this, the final F1 score was measured at 0.921642, meaning that the overall accuracy of the proposed technique with video streaming sites was about 92.2%. **Table 14** shows the analysis results for the video streaming sites.

Table 14. Result of video streaming site

Number of Repetitions	TP	FN	FP	TN	Accuracy	Precision	Recall	F1 Score
1	82	4	10	76	0.918605	0.891304	0.953488	0.921348
2	83	3	10	76	0.924419	0.892473	0.965116	0.927374
3	82	4	11	75	0.912791	0.881720	0.953488	0.916201
Total	247	11	31	227	0.918605	0.888489	0.957364	0.921642

5.2.3. Analysis for Webtoon Sites

The results of repeating the method three times on 50 datasets with 25 suspected webtoon piracy sites and 25 normal sites in a 1:1 ratio are as follows. Accuracy was consistent with the true condition at 0.973333, and the precision of predicting suspected piracy sites was 0.961039. Recall, which indicates that it was predicted as a suspected infringement site among actual infringement sites, was 0.98667. From this, the final F1 score was measured at 0.973684, meaning that the accuracy of the proposed technique with webtoon sites was about 97.4%. **Table 15** shows the analysis results for the webtoon sites.

Table 15. Result of webtoon site

Number of Repetitions	TP	FN	FP	TN	Accuracy	Precision	Recall	F1 Score
1	24	1	1	24	0.96	0.96	0.96	0.96
2	25	0	1	24	0.98	0.961539	1.0	0.980392
3	25	0	1	24	0.98	0.961539	1.0	0.980392
Total	74	1	3	72	0.973333	0.961039	0.986667	0.973684

5.2.4. Experimental Results

The three-time performance measurement results using the proposed method for the site suspected of copyright infringement are shown in **Table 16**. The final measured accuracy of the proposed technique was 94.1%, precision was 92.1%, recall was 96.4%, and the F1 score was 94.2%. The reason why each try of F1 score is different is we use VPN for crawling. Thus, the web page source may or may not be crawled in time.

Table 16. Accuracy results

Number of Repetitions	TP	FN	FP	TN	Accuracy	Precision	Recall	F1 Score
1	144	6	12	138	0.94	0.923077	0.96	0.941176
2	146	4	12	138	0.946667	0.924051	0.973333	0.948052
3	144	6	13	137	0.936667	0.917197	0.96	0.938111
Total	434	16	37	413	0.941111	0.921444	0.964444	0.942454

The existing piracy site copyright detection method studied by Choi et al.[3] derived an accuracy of 93% using the confusion matrix-based accuracy, but the accuracy of the method proposed in this study is approximately 94.1%, which is relatively better. In addition, the accuracy of the technique proposed by Kim et al[4]. was measured using the F1 score based on the confusion matrix and an accuracy of approximately 92.9% was obtained. However, the F1 score of the technique proposed in this study was approximately 94.2%, showing that the accuracy is relatively excellent, as shown in **Table 17**.

F1 score's Precision is related to low False Positive, and Recall is related to False Negative. For models with low Precision and high Recall, the predicted ratio correctly compared to True Condition is high but not accurate, and for models with high Precision and low Recall, the predicted ratio correctly compared to True Condition is low but accurate. Therefore, when both Precision and Recall have high values, the accuracy of the model may be determined to be accurate [9]. Therefore, it can be seen that the proposed technique of this paper, which has higher Precision and Recall than Choi's Research [3] and Kim's Research [4], is relatively excellent.

Table 17. Comparison with Previous Researches

Research	Accuracy	Precision	Recall	F1 Score
Choi's Research [3]	0.929824	0.888888	0.982456	0.933333
Kim's Research [4]	0.927632	0.941176	0.917197	0.929032
Proposed Technique	0.941111	0.921444	0.964444	0.942454

6. Conclusion

Through the technique proposed in this study, the lack of response to continuously changing piracy sites owing to insufficient dataset configuration, which is a limitation of the piracy site detection method to prevent copyright infringement, is overcome. This paper proposed a search engine-based piracy site detection method and achieved an accuracy of approximately 94.1% for Accuracy and 94.2% for F1 score, which are better than the accuracy of approximately 93% and F1 score of approximately 92.9% of the existing copy right detection method. Through the proposed technique in this paper, it will be possible to block piracy sites and protect the copyright holder's rights by detecting and determining whether or not the piracy site is constantly changed. Through the proposal technique of this paper, it was proved that it was possible to respond to intelligent copyright infringement sites that constantly change, and compared to the existing proposal technique, copyright detection and infringement were determined through improved detection techniques. Through this paper, it will be possible to contribute to the protection of copyright holder's rights by detecting and blocking copyright infringement sites.

Acknowledgement

This research project was supported by the Ministry of Culture, Sports and Tourism (MCST) and Korea Copyright Commission in 2021(2019-PF-9500).

References

- [1] Korea Copyright Protection Agency, "English Version of C STORY 2016," *KCOPA Report*, Dec. 2016. [Article \(CrossRef Link\)](#)
- [2] Department of Communications, "ONLINE COPYRIGHT INFRINGEMENT RESEARCH," A *MARKETING RESEARCH REPORT*, Australia, June. 2015. [Article \(CrossRef Link\)](#)
- [3] S.K. Choi and J. Kwak, "Feature Analysis and Detection Techniques for Piracy Sites," *KSII Transactions on Internet and Information Systems*, Vol. 14, No. 5, pp. 2204-2220, May, 2020. [Article \(CrossRef Link\)](#)
- [4] E.J. Kim and J. Kwak, "Intelligent Piracy Site Detection Technique with High Accuracy," *KSII Transactions on Internet and Information Systems*, Vol. 15, No. 1, pp.285-301, Jan. 2021. [Article \(CrossRef Link\)](#)
- [5] Jung-Sik Hwang and Hyun-Gon Kim, "Blockchain-based Copyright Management System Capable of Registering Creative Ideas," *Journal of Internet Computing and Services*, Vol. 20, No. 5, pp. 57-65, Oct. 2019. [Article \(CrossRef Link\)](#)

- [6] Hee-Wan Park, "Design and Implementation of Server-based Resource Obfuscation Techniques for Preventing Copyrights Infringement to Android Contents," *Journal of the Korea Contents Society*, Vol. 16, No. 5, pp.13-20, May. 2016. [Article \(CrossRef Link\)](#)
- [7] Korea Copyright Protection Agency, "2021 Annual Report on Copyright Protection," *KCOPA Report*, p.23, Apr. 2021. [Article \(CrossRef Link\)](#)
- [8] J.H Lee and H.K Lee, "A Study on Korean Emotion Index Using F1_score," *The Journal of Internet Electronic Commerce Research*, Vol. 20, No. 1, pp.131-145, Feb. 2020. [Article \(CrossRef Link\)](#)
- [9] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, and B. Thirion, "Scikit-learn: Machine Learning in Python," *Journal of Machine Learning Research*, vol. 12, pp. 2825-2830, 2011, [Article \(CrossRef Link\)](#)
- [10] R. Girshick, J. Donahue, T. Darrell and J. Malik, "Region-Based Convolutional Networks for Accurate Object Detection and Segmentation," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 38, No. 1, pp. 142-158, 1 Jan. 2016. [Article \(CrossRef Link\)](#)
- [11] DeVries Zachary, Locke Eric, Goda Mohamad, Moravek Dita, Phan Kim, Stratton Alexandra, Kingwell Stephen, K. Wai Eugene and Phan Philippe, "Using a national surgical database to predict complications following posterior lumbar surgery and comparing the area under the curve and F1-score for the assessment of prognostic capability," *The Spine Journal*, Vol.21, no. 7, pp.1135-1142, Feb. 2021. [Article \(CrossRef Link\)](#)



Eui-Jin Kim is a postgraduate student of the master's course at Dept. of Cyber Security in Ajou University, Republic of Korea. He received his B.S. degree from Ajou University, Republic of Korea. His research interests include Copyright protection, Cryptographic protocols, and Network security.



Deuk-Hun Kim is a postdoctoral researcher at the institute for information and communication in Ajou University, Republic of Korea. He received his Ph.D. degree from Ajou University, Republic of Korea. His research interests include Convergence security, Cryptographic protocols, and Network security.



Jin Kwak is a professor at the Dept. of AI Convergence Network and Dept. of Cyber Security in Ajou University, Republic of Korea. He received his Ph.D. degree from SKKU, Republic of Korea. His research interests include Copyright protection, Cryptographic protocols, and Applied security mechanisms for cloud and big data system.