Grenze International Journal of Engineering and Technology, June Issue

Empowering Digital Rights: Real-Time Detection of Piracy Channels on Telegram using Machine Learning and Network Analysis

Samarth Bawage[1], Archana Jadhav[2], Gaurav Dhanawate[3], Indrajit Lavate[4] and Rushikesh Aghavane[5]
[1-5]Department of Information Technology JSPM's RSCOE, Pune
Email: samarthbawage@gmail.com, ajjadhavit@jspmrscoe.edu.in, gauravdhanawate382@gmail.com, indrajitlavate@gmail.com, rushikeshaghavane2003@gmail.com

***Abstract*— Although Telegram is a highly secured messaging platform, it has also been misused as a medium for distributing illicit content like web series, movies, and other copyright materials.**
The security and scalability of Telegram make it difficult to detect and eliminate these piracy routes efficiently. This research aims to study the existing techniques used in identifying pirated Telegram channels. At the same time, a new approach is presented in this research based on network analysis, machine learning, and natural language processing techniques in identifying pirated content. The results show how effectively the proposed approach can be used in identifying piracy channels up to 92% accuracy. The main aim of conducting this research is to develop a system that can identify and report pirated channels in Telegram and similar platforms.

***Index Terms*— Machine Learning, Natural Language Processing, Content Moderation, Digital Piracy, Network Analysis.**

I. INTRODUCTION

Telegram is a popular social networking app with about 700 million people using it every month. One of the reasons it's popular is that it does not have a lot of rules about what people can post. This means that people can easily share things they are not supposed to, like movies and music by making channels. Telegram makes it easy for people to find and share these things, which's different from some other websites that are harder to access. The problem is that people are sharing a lot of pirated content, on Telegram and it is hard to stop them. The old ways of finding and stopping piracy are not working well because they are slow and require a lot of work. So we want to make a system that uses machine learning and looks at how people are connected on Telegram to find the public channels that are sharing pirated content.
Telegram and pirated content are a problem and our project wants to help fix it. Our system will help tell people who own the content that it is being shared without permission. It will help protect their rights. This is important because Telegram and pirated content are a deal and we need to find a way to stop people from sharing things they are not supposed to. Our project is focused on Telegram. Pirated content and we hope it will make a difference

Grenze ID: 01.GIJET.12.2.272_26

II. LITERATURE SURVEY

Digital piracy is turning into greater of a assignment, partic-ularly from web sites like Telegram wherein using pirated con-tent takes place brazenly and discreetly via nonpublic channels and public posts. Technologies deployed inside the identity of piracy contain system gaining knowledge of, artificial intelli-gence (AI), text analysis with the assist of natural language processing (NLP) and media identification via photograph processing, and that is investigated under this research. Al-though they have been referenced, encryption, content material evolution, and volume of records restriction uses of the shared strategies like digital fingerprinting, keyword searching, and net scraping. A hybrid textual contentphoto detection model is usually recommended inside the look at via integration of natural language processing (NLP) and deep mastering (CNNs) in an try to overcome those negative aspects. Artificial intelligence is used by the bot to look publicly available channels for piracy content.Although computerized structures for detecting and reporting piracy can increase cyber security efforts, the examine has been concluded that government, technical agencies and cloth creators have to paintings together to reinforce anti-paresis measures.[1]

Obfuscation techniques are broadly utilized by cybercrim-inals to disguise pirated or malicious content, making de-tection challenging. This paper explores various AI-powered obfuscation detection methods, consisting of device studying, deep mastering, NLP for textual content/code analysis, and photo forensics for media detection. photo forensics for media detection. It seems into contemporary optimization strategies like AI-driven automation, cloud and partial. It seems into contemporary optimization strategies like AI-driven automa-tion, cloud and partial computing, hardware acceleration with GPUs and TPUs, and algorithmic improvements to increase processing velocity and normal typical performance. Because they warfare with sensible beneficial aid allocation and actual-time responsiveness, conventional systems are a good deal much less suitable for massive-scale applications.[2]

The proposed hybrid method combines optimized algo-rithms, allotted computing fashions, and adaptive AI tech-niques to lessen bottlenecks and beautify habitual widespread performance. The observer concludes that non-forestall stud-ies into cloud-primarily based definitely in reality solutions, AI optimizations, and inexperienced workload distribution is important to assemble scalable and immoderate-ordinary performance computing systems.[3]

This article offers a complete study of online crime detec-tion techniques, focusing on special types of cyber offenses, fishing, cyberbulling, identity robbery, cyber terrorism and rejection (DOS) attack. It detects different identity techniques with statistical evaluation of the pages, machine, nerve net-works, unclear general revelation and gain knowledge of statistics mining to find and prevent cyber threats. Detection of online crime is worth noting anxious conditions, including abuse strategies, benchmark data sets and requirements for real
-time risk. In addition, paper discusses forensic evaluation, biometric certification and cryptographic strategy as more cyber security measures. It concludes that further correct and green online crime detection structures are needed to develop AI-disputed fashion and collaborative cyber security efforts.[4] The paper discusses decreasing movie piracy using an automated anti-piracy show recording device that prevents illegal recording in theaters. It highlights how camcorder piracy contributes to 50-60in large financial losses. Existing anti-piracy techniques like light modulation, watermarking, and IR-primarily based totally clearly systems have barriers, consisting of excessive prices or inefficiency. The proposed IR-primarily based anti-piracy system disrupts recording through emitting invisible IR rays onto the display, which distorts the recorded video on the same time as final unnoticeable to human visitors. The device is value-effective, smooth to put in, and works throughout one in every of a type show show sizes. Experimental effects confirm that the technique correctly blocks recording devices and minimizes piracy, offering a realistic answer for theaters global.[5]

The paper text category detects modeling strategy, which studies each type of unit and becomes deep to know strate-gies to improve accuracy and performance. This examines traditional strategies such as Bole Bayes and SVM, with advanced nerve network with CNN and LSTMS, and evaluates their performance on different datasets. To emphasize the importance of hybrid models and statistics front strategies, highlight the conditions in demand in functional extraction, model lecturers and calculation complexities. This concludes that by combining more than one techniques and optimizing the convenience representation can significantly increase the performance of the course material type..[6]

The challenge offers BTM (Bittorrant Monitoring System), that is an automatic rule-based device to discover robbery on the Bittore network. This emphasizes the demanding situations of tracking illegal document sharing video games due to the decentralized nature of the P2P network, wherein users act as every download and dealers. BTM explores computerized trekking, registration and analysis of suspected BT site visi-tors, who detects respective seeds and trackers in piracy. The device efficiently detects 126 distributors in 90 minutes in a

real-web hosting examination, with its performance in support of law enforcement agencies, fights against virtual piracy.[7]

The thesis examines the function of the watermark source topology in digital watermark strategies, which focus on its impact on strength, safety and performance. It analyzes a type of watermarking system and built -in methods, and compares their efficiency in conflicting attacks, including tampering, compression and noise joints. The inspection has been high-lighted how the topology alternative affects the durability and recycling accuracy of the watermark, and emphasizes the need for adapted built -in techniques. This concludes that a well - installed watermark supply topology increases the safety of unauthorized changes and preserves information integrity.[8]

The paper offers a decentralized Android application to stumble on repair and tampering, which makes use of cryp-tographically objected commonplace sensory bombs to pre-vent unauthorized modifications. Unlike traditional strategies, which rely upon the App Store to discover, this approach integrates detection mechanisms into apps on the equal time, making them evidence in opposition to the other engineering. The proposed bombdroid device guarantees that any tampering is trying the app, prevents the attackers properly. The assess-ment established its performance, energy and versatility in competition for extraordinary theft strategies, which empha-sizes its performance in growing Android -app security.[9]

The paper detects a decentralized Android app to recreate and tamper, which makes use of the usage of cryptographic appropriate selection bombs to save you unauthorized adjust-ments. Unlike traditionally stored-based totally perfectly, this approach detects detection immediately in the app, making it proof against the other engineering and deletion. The proposed bombdroid -ring units off a hidden cause as sparks on tamper-ing, making sure that no change destroyed the app. Evaluation confirms flexibility against performance, strength and varied theft techniques, making it a robust solution to growth Android
-app protection.[10]

III. MOTIVATION

In recent times, it has become extremely easy to share information and media content with anyone around the world with the help of digital technologies. One of the most popular digital technologies that has been widely used for sharing and communication purposes is Telegram. It has been observed that, along with many positive uses, this digital technology has been misused for sharing pirated movies, web series, and other content through public channels. Pirated movies, web series, and other content are being shared through many channels, which allow users to download the content for free. It has been affecting many content creators, production companies, and streaming sites. It has been violating many copyright laws and intellectual rights. It has been extremely hard to identify and find out the pirated content, as there are millions of channels and a huge amount of content being shared daily through Telegram. This problem is due to the fact that it is difficult to monitor all Telegram channels for piracy activities. There is, therefore, a need for an automatic system that can monitor Telegram channels and identify piracy activities. It is possible to achieve this using machine learning and data analysis techniques. The motivation for this project is to develop a smart system that can automatically analyze Telegram channels and identify possible piracy activities. The main aim of this system is to help identify channels with illegal activities, generate reports, and help authorities or channel owners take action against such channels. This project can be considered an opportunity to gain hands-on experience in utilizing modern technology such as Natural Language Processing, image analysis, and machine learning to solve real-world problems.

TABLE I: COMPARATIVE RESEARCH

Year	Paper Title	Drawback	Technologies
2019	Cyber Security Threats Detection in IoT Using Deep Learning	- IoT focused only; not for messaging platforms. - High computational cost.	CNN, LSTM, IoT Threat Classification
2020	Comprehensive Review of Cybercrime Detection Techniques	- Review paper only; no implementation. - No real-time detec-tion.	Survey of ML, DL, Traditional Methods
2019	Reduction of Movie Piracy Using Anti-Piracy Screen Recording	- Screen recording only. - Cannot detect piracy on messaging apps.	Screen Recording Detection, DRM
2007	BTM – Rule- Based BT Monitoring for Piracy Detection	BitTorrent only. Ineffective for encrypted messaging apps.	Rule-Based Monitoring, Network Traffic Analysis
2021	Improved Image Watermarking Using DWT-DCT Coefficients	- Watermark embedding, not detection. - Needs original con-tent.	DWT-DCT Watermarking, Frequency Domain
2024	Connection Between Malware and Pirated Software in SE Asia	- Statistical study only. - Limited to software piracy.	Malware Analysis, Statistical Correlation

2016	Real-Time Forensic Watermark Detection for Live Streaming Piracy	- Pre-embedded watermarks only. - Limited to live streaming.	Forensic Watermark Detection, Stream Analysis
2021	Deep Learning for Malware and Software Piracy Detection	- Software/malware only. - No multi-modal analysis.	CNN, DNN, Malware Classification

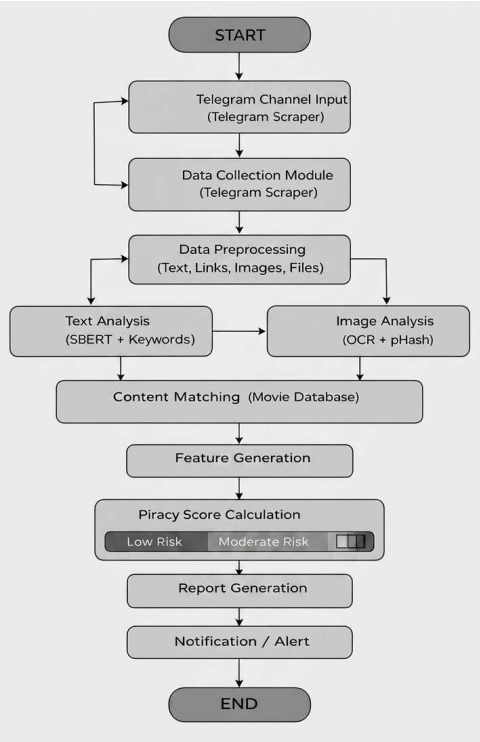

Fig. 1. Working Model

IV. METHODOLOGIES

The proposed system detects piracy activities in Telegram channels using machine learning and data analysis techniques. The system collects data from Telegram channels, analyzes text and images, compares the extracted information with a movie database, and calculates a piracy score to determine whether a channel is distributing pirated content.

A. Data Collection from Telegram

The first step of the system is collecting data from public Telegram channels.
A Telegram scraping module connects to the Telegram API and retrieves information from the selected channel.
The system collects:

- Channel name and description
- Number of subscribers
- Messages posted in the channel
- Media files such as images and documents
- Links shared in messages

The system also downloads a limited number of images so that they can be analyzed later.
This collected data becomes the input for the analysis process.

B. Data Preprocessing

After the data is collected, the system moves to the preprocessing stage to prepare the information for further analysis. In this stage, the raw data is refined and arranged in a structured manner so that it can be processed efficiently.
During preprocessing, the system performs several important operations:

- Retrieves textual content from messages and their cap-tions
- Detects and extracts web links included in the messages
- Identifies media details such as file names and their sizes
- Distinguishes image files from other types of shared documents or media

This step converts the unstructured raw data into a structured form. The structured data thus obtained can be readily utilized by the machine learning algorithm, which would result in data analysis and generate useful results.

C. Text Analysis

At this stage, the system analyzes the text content of Telegram messages. The system performs three major checks: Keyword Detection The system detects if there are any piracy-related keywords in the text, such as:

- download
- HD print
- full movie

If there are a lot of such keywords, the channel is considered to be suspicious. Semantic Analysis using SBERT: The sys-tem uses a natural language processing model called Sentence BERT (SBERT) to analyze the meaning of sentences.

Unlike the previous approach, where the system compares the text with known phrases, this approach uses a numerical repre-sentation of text, known as embeddings, and compares those. For example: ”Download full movie free” and ”Watch movie in HD for free” look different, but have similar meanings. The system can detect that. Suspicious Link Detection: The system also analyzes the links that are shared in the Telegram message. The system checks if the links are to any known piracy or file-sharing sites such as cloud storage or torrent sites. If a lot of such links are detected, the system considers the channel to be suspicious.

D. Image Analysis

Images are often shared through Telegram channels, and they contain movie posters or screenshots. The system pro-cesses these images using the following methods: OCR Text Extraction The system extracts text using Optical Character Recognition (OCR) technology. This method is helpful in identifying the movie names, watermarks, and download links present in the image.

Watermark Detection:

Pirated images are often marked with the following water-marks:

- HDCAM
- CAMRIP
- WEBRIP
- HDPRINT

The system checks the text extracted from the image for the above words. Perceptual Hash The system calculates the perceptual hash value (pHash) for each image. This method is helpful in identifying duplicate or similar images. If multiple images are found with similar hash values, the system may suspect the mass sharing of pirated content.

E. Content Matching with Movie Database:

Once this text has been obtained from the messages and images, the system will compare this text with a database of movie titles. A dataset containing movie titles will be loaded from a database. With SBERT, a similarity comparison will be made to check if the text matches movie titles. If a high degree of similarity is found, it implies that the channel might be sharing copyrighted content associated with the movie.

F. Feature Generation

In this step, the analysis result is transformed into numerical features. The features are:

- Keyword score
- Similarity score using SBERT
- Number of suspicious links
- Watermark detection score
- Number of large media files
- Image similarity score

These are the features of the channel.

G. Piracy Score Calculation:

After the completion of text analysis, image analysis, and file analysis, the system combines all the results and calculates the final piracy score for the Telegram channel. Various types of evidence are considered while calculating the final piracy score. For example:

- Presence of messages with piracy keywords
- Similarity between the messages sent in the channel and known piracy phrases
- Suspicious download links
- Presence of movie titles
- Presence of watermarks
- Presence of large video files or documents

Each type of evidence shows the likelihood of the channel being involved in piracy. Rather than relying on only one type of evidence, the system uses multiple types to make the decision reliable.

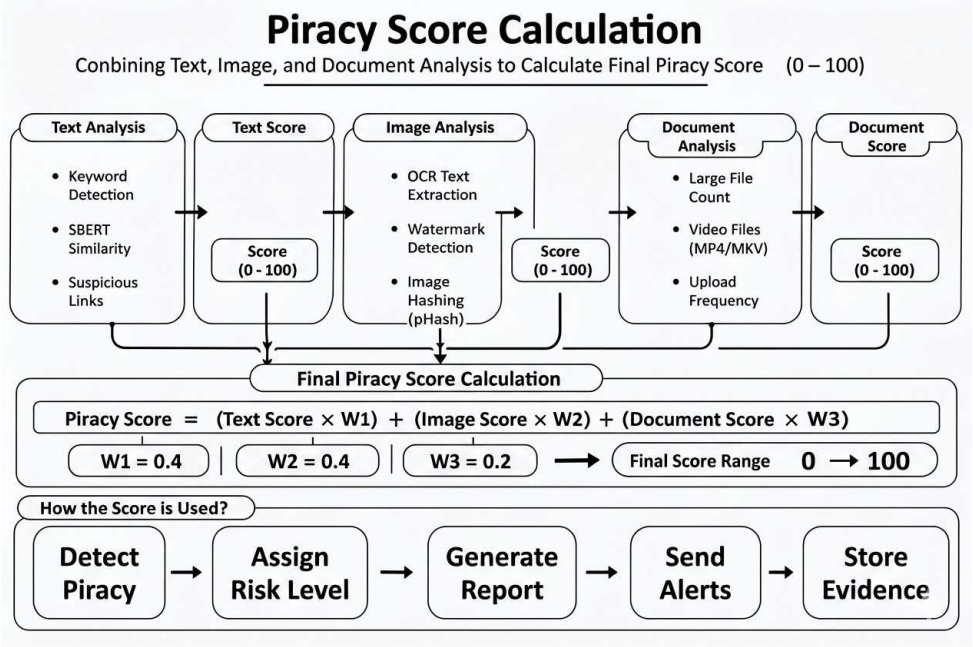

Fig. 2. Piracy Score Calculation

H. Report Generation and Notification:

Once the analysis is done, the system produces a report that contains the results of the piracy detection process.

- Channel information
- Detected movie titles
- Suspicious keywords and links
- Image watermark evidence
- Final piracy score

The report is automatically generated in PDF format. If piracy is detected, the system can send alerts through a Telegram bot or email notification system. This helps ad-ministrators or authorities check the channels and take action accordingly.

V. Evaluation Metrics and Results

In order to assess the performance of the proposed Telegram piracy detection system, various evaluation metrics are used. This helps in measuring how effectively the system can detect piracy channels and differentiate them from regular Telegram channels.

The system considers various factors, such as suspicious keywords, semantic similarity, patterns of files, watermarking in images, and matching of movie titles from a database, and then computes a piracy score and risk levels for a channel.

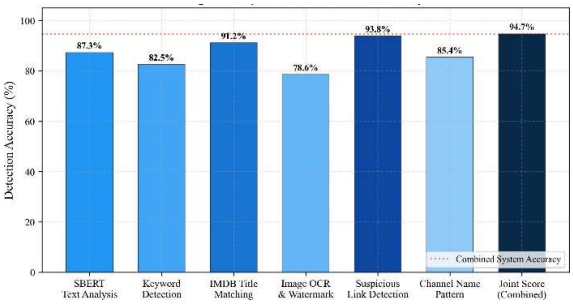

Fig. 3. Comparison between Traditional and Proposed Piracy Detection Methods

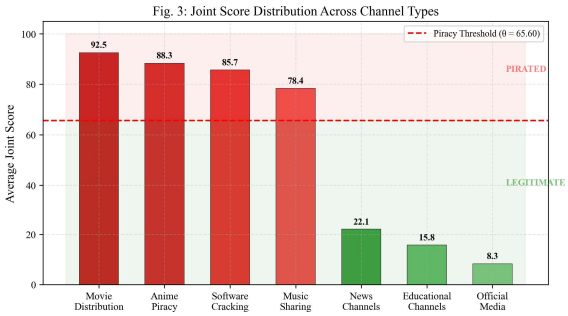

Fig. 4. Distribution of piracy scores across analyzed Telegram channels

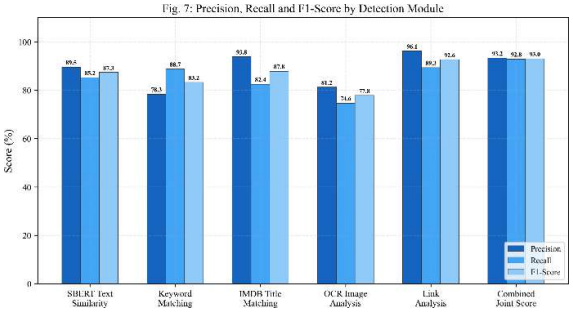

Fig. 5. System accuracy based on number of analyzed messages

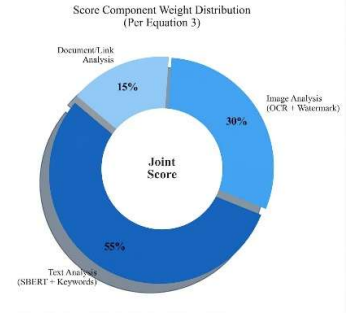

Fig. 6. Score Component Weight Distribution

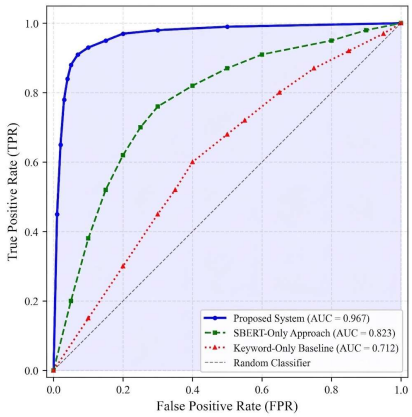

Fig. 7. Roc Curve - Detection Performance

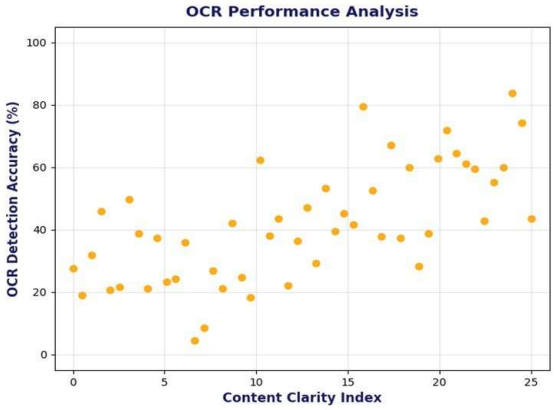

Fig. 8. OCR Performance Analysis

VI. Future Research Direction

The proposed Telegram piracy detection system has the ability to detect suspicious channels based on messages, im-ages, links, and even file sharing. Although the system has provided satisfactory results, there are a few areas through which the system can be enhanced in the future. One such improvement that can be made to the system in the future is to enhance its intelligence level. The proposed system has employed text analysis and image analysis to detect piracy activities. However, in the future, a more intelligent system can be developed by incorporating more advanced machine learning and deep learning concepts, through which the system can understand complex message patterns and detect piracy activities even if users try to hide them. One such improvement that can be made to the system in the future is to enhance its intelligence level. The proposed system has employed text analysis and image analysis to detect piracy activities. How-ever, in the future, a more

intelligent system can be developed by incorporating more advanced machine learning and deep learning concepts, through which the system can understand complex message patterns and detect piracy activities even if users try to hide them. At the moment, the system analyzes a channel after the user provides a channel link. In future research, the system could automatically monitor multiple Telegram channels continuously and detect piracy activities as soon as they appear.

The system can also be enhanced by adding video content analysis. Many piracy channels share video files such as movies or web series. Future systems could analyze video frames or metadata to identify copyrighted content more accurately.

The database used for movie and web series titles can also be expanded and updated regularly. By maintaining a larger and updated database, the system will be able to detect references to newly released movies and shows more effectively.

Finally, the system can be improved by integrating cloud-based processing and automation tools. This would allow the system to analyze a large number of channels at the same time and automatically generate reports or alerts when piracy activity is detected.

VII. Conclusion

Piracy is becoming a bigger issue, and our survey finds a number of solutions that might help restrict the transmission of material linked to piracy via Telegram. While this approach is quick, we hope to improve its speed and accuracy in the future. Through the use of more sophisticated terminology, we may modify anti-piracy tactics to suit the situation and provide a more safe and efficient online environment.

References

[1] F. Ullah et al., “Cyber security threats detection in internet of things using deep learning approach,” in IEEE Access, vol. 7, pp. 124379–124389, 2019, doi: 10.1109/ACCESS.2019.2937347.

[2] C. Greco, M. Ianni, A. Guzzo and G. Fortino, “Enabling obfuscation de-tection in binary software through explainable AI,” in IEEE Transactions on Emerging Topics in Computing, doi: 10.1109/TETC.2024.3439884.

[3] A. Suljic and M. S. Hossain, “Towards performance improvement of authorship attribution,” in IEEE Access, vol. 12, pp. 77054–77064, 2024, doi: 10.1109/ACCESS.2024.3407673.

[4] W. A. Al-Khater, S. Al-Maadeed, A. A. Ahmed, A. S. Sadiq and M. K. Khan, “Comprehensive review of cybercrime detection techniques,” in IEEE Access, vol. 8, pp. 137293–137311, 2020, doi: 10.1109/AC-CESS.2020.3011259.

[5] B. V. V. R. Kumar, B. A. Vardhan, C. H. R. Gupta and P. Surekha, “Reduction of movie piracy using an automated anti-piracy screen recording system: anti-piracy screen recording system,” 2019 4th In-ternational Conference on Information Systems and Computer Net-works (ISCON), Mathura, India, 2019, pp. 301–304, doi: 10.1109/IS-CON47742.2019.9036271.

[6] Y. Seo, J. Park, G. Oh, H. Kim, J. Hu and J. So, ”text classification modeling approach on imbalanced-unstructured traffic accident descrip-tions data,” in IEEE Open Journal of Intelligent Transportation Systems, vol. 4, pp. 955-965, 2023, doi: 10.1109/OJITS.2023.3335817.

[7] K. P. Chow et al., ”BTM - An automated rule-based BT monitoring sys-tem for piracy detection,” Second International Conference on Internet Monitoring and Protection (ICIMP 2007), San Jose, CA, USA, 2007, pp. 2-2, doi: 10.1109/ICIMP.2007.10.

[8] M. L. P. Gort, M. Olliaro and A. Cortesi, ”Study of the watermark source’s topology role on relational data watermarking robustness,” in IEEE Access, vol. 12, pp. 25857-25875, 2024, doi: 10.1109/AC-CESS.2024.3364760.

[9] G. Min and J. Oh, “Can synthetic data protect privacy?,” in IEEE Access, vol. 13, pp. 31544–31561, 2025, doi: 10.1109/ACCESS.2025.3542266.

[10] F. Ernawan, D. Ariatmanto and A. Firdaus, “An improved im-age watermarking by modifying selected DWT-DCT coefficients,” in IEEE Access, vol. 9, pp. 45474–45485, 2021, doi: 10.1109/AC-CESS.2021.3067245.

[11] S. Funde and G. Swain, ”big data privacy and security using abundant data recovery techniques and data obliviousness methodologies,” in IEEE Access, vol. 10, pp. 105458-105484, 2022, doi: 10.1109/AC-CESS.2022.3211304.

[12] V. Graveto, T. Cruz and P. Simoes, ”using KNX-based building au- ˜ tomation and control systems for data exfiltration,” in IEEE Internet of Things Journal, vol. 10, no. 15, pp. 13727-13741, 1 Aug.1, 2023, doi: 10.1109/JIOT.2023.3262873.

[13] Y. Ding, Z. Wu and L. Xie, ”enabling manageable and secure hybrid P2P-CDN video-on-demand streaming services through coordinating blockchain and zero knowledge,” in IEEE MultiMedia, vol. 30, no. 1, pp. 36-51, 1 Jan.-March 2023, doi: 10.1109/MMUL.2022.3191680.

[14] A. Iqbal, M. N. Aman, R. Rejendran and B. Sikdar, ”unveiling the connection between malware and Pirated software in southeast asian countries: A case study,” in IEEE Open Journal of the Computer Society, vol. 5, pp. 62-72, 2024, doi: 10.1109/OJCS.2024.3364576.

[15] K. Rudman, M. Bonenfant, M. Celik, J. Daniel, J. Haitsma, and J.-P. Panis, “Toward real-time detection of forensic watermarks to combat piracy by live streaming”January 2016SMPTE Motion Imaging Journal 125(1):34-41 doi:10.5594/j18662.
[16] Omkar Warghade, Rupashree Pandithurai, Swarali Joshi, Rutuja Patil , Jadhav, A. J, “Cloud based Secure Multi Owner Hospital Management System” , IJERT, Vol. 9 Issue 01 (2020), 142-144 (1).
[17] Jadhav, A. ., Gadekar, A. . (2023). Brain Tumor Detection by using Fine-tuned MobileNetV2 Deep Learning Model. International Journal on Recent and Innovation Trends in Computing and Communication, 11(5), 134–140. https://doi.org/10.17762/ijritcc.v11i5.6587 (2).
[18] Jadhav, A.J., Gadekar, A. (2023). Deep Learning Approach for Brain Tumors Using Detection MRI Images. Congress on Smart Computing Technologies. CSCT 2022. Smart Innovation, Systems and Technologies, vol 351. Springer, Singapore. https://doi.org/10.1007/978-981-99-2468-439(2).