IJCRT.ORG ISSN : 2320-2882

INTERNATIONAL JOURNAL OF CREATIVE RESEARCH THOUGHTS (IJCRT)

An International Open Access, Peer-reviewed, Refereed Journal

GuardianEye360 – Sensitive Content Monitoring Extension

Triloki Singh, Kadali Lakshmi Kanishka, Nidhi Singh, Kavvampally Manasa

Department of Computer Science and Engineering, Lovely Professional University, Punjab, India.

Pooja Sharma

Assistant Professor, Department of Computer Science and Engineering, Lovely Professional University, Punjab, India.

Abstract: The use of the internet and search engines are increasing exponentially in this world. Almost In every field including Education, HealthCare, Retail Marketing, business, and many more. There will be many unwanted contents being exposed to the internet users while browsing covering informative material to obscene content. This project introduces GuardianEye360 which is a sensitive image detection system using a combination of TensorFlow and CNN (Convolutional Neural Networks), Cascades, JavaScript. These features are then used to load the content into DOM (Document Object Model). Further the CNN with TensorFlow framework detects the images on the website and blurs the obscene content that is found on the loaded website. This creative browser extension is made to include a trained model that can effectively filter NSFW (Not-Safe-For-Work) content. The extension, which divides discovered images into either NSFW or SFW (Safe-For-Work) domains. The issue of being exposed to sensitive content online is addressed by GuardianEye360. This system operates to protect user experiences while they browse, giving users authority to conceal or hide anything that is too sensitive in the real time. Giving users total control over how they connect online, creates a more secure and safe online environment.

Keywords: User-permission, storage, URLS, image filtering, Convolutional Neural Network, TensorFlow , Obscene image detection, Document Object Model, Not Safe For Work, blur image.

I. Introduction

As the popularity of the internet is rapidly increasing, allowing user to access the online content easily, the risk of exposure to sensitive content is also increasing [1]. The convenience and ease of use social platforms and search engines, more number of people across the world are relaying on the web content, which is further leading to a situation in which users are exposed to sensitive or obscene content even when they indeed aim to search something else [2]. Expressive themes, pictures, and material are connected to many aspects of our lives due to the extensive adoption of an economy that rapidly accepted an excess of sexualization [3]. In addition to the need to address such content for educational purposes, it's important to understand that giving inappropriate attention to sexual content can have very negative effects. Growing interest in the content shows that being exposed to sexually explicit content may be an aspect in the growth of reserved behavior [4]. It is important to distinguish between sexual and educational content since a large portion of the global population, including adults and children spends longer periods of time online consuming various media and data. This is particularly crucial when it comes to protecting the mental health of young people [5]. The effects of being exposed to filthy and pornographic [6] content, is in fact clearly unrelatable to children and teens, and it may have long-lasting effects on their mental health[7].

It can be uncomfortable and embarrassing to come across obscene images by accident, whether at work or at home. When sexual content is accidentally discovered while casually reading a website, it can lead to conflict among colleagues and risk trust and professional relationships. One of these other scenarios is browsing an website at home, and some explicit content shows up on the website. It can be distressing, particularly if children [8] are around. When a child is unintentionally exposed to inappropriate information, it can be upsetting and challenging to manage, especially if they are too young to fully understand or analyse it. One of the causes of this condition is online ads [9]. This discourages users from visiting websites, which negatively impacts their entire surfing experience.

The creation of GuardianEye360, a browser extension provided with a pre-trained model for blurring pornographic images, signifies a crucial advancement in enhancing internet safety and shielding users from inappropriate content. Leveraging a diverse dataset of explicit material, this model ensures robust detection performance. To coherent integrate the extension into popular web browsers like Chrome or Firefox, developers have implemented user-friendly interfaces and real-time detection techniques [2]. The features of the extension include real-time picture analysis when websites are loaded into the browser, along with effective blur or filter techniques when inappropriate information is detected. While providing privacy protections protect sensitive data, user controls that provide modification of sensitivity levels and blacklist particular websites improve accessibility and convenience.

These technological solutions, leveraging heuristic of acquire a knowledge of models for content analysis, provide robust parental control mechanisms. Through the deployment of obscenity blockers and adaptive filtering techniques, users can enforce tailored content restrictions, encouraging a safer online ecosystem [3]. The imperative to curate age-appropriate content serves as an essential fundamental reducing the countless risks inherent in internet access.

II. RELATED WORK

In their work, P. Taneja, D. Singh, and T. Rajora [2] developed a browser extension that tracks surfing behaviour and identifies Obscene information in order to improve online safety. The JavaScript-based web extension records user activities and sends information to a central server that uses a machine learning model to identify inappropriate content. Data pre-processing ensures consistency and quality, including frame-by-frame analysis for videos and gifs, resizing images, and converting to the RGB colour standard. Information is exchanged with internet security companies and kept in a MongoDB database that is arranged according to the type of content. The ConvNeXt model demonstrated a high degree of accuracy in classifying NSFW photographs, demonstrating the significance of preprocessing and dataset size. These findings support ConvNeXt's effectiveness in NSFW material detection ,shaping future image categorization research.

The problem of automatically identifying offensive or obscene content in videos is discussed in the study by Samal, S., Nayak, R., Jena, S., & Balabantaray, B. K [3]. It presents the Obscenity Detection Transformer (ODT), a revolutionary deep-learning transformer-based system that emphasizes the use of specific information to improve detection accuracy. With the help of vision transformer and long short-term memory layers, the model can extract useful features from video frames. With accuracies of 99.6% and 98.8%, respectively, extensive tests on the Pornography-2k and Pornography-800 datasets show better performance than CNN-based models. Preprocessing includes capturing pertinent contextual information by annotating brief video segments and turning videos into frames. By employing GELU activation functions and multi-head attention mechanisms, the model efficiently handles positional embeddings and enhances classification accuracy. Better detection performance is also achieved by improving temporal dependency modelling with the incorporation of LSTM layers.

D. C. Moreira, E. Torres Pereira and M . Alvarez [4] introduced a dataset of 376K photos for the purpose of detecting pornography, the research tackles the issue of limited data and subjective classification. Pictures from Reddit were taken, showing a variety of scenes from everyday life to porn. Using specific criteria, a strict definition of pornography was established. Furthermore, a standardization approach for the outputs of NSFW picture moderation APIs was put forth, allowing for cross-service comparability. Additionally, a CNN model based on convolutional networks using the PEDA 376K dataset is presented in the article. This model makes use of an effective method for hyperparameter selection. All things considered, the work advances the field of pornography detection research and establishes the framework for next studies in this area.

NOBLE SAJI MATHEWS and SRIDHAR CHIMALAKONDA [5] , created Detox Browser, a Chrome extension, filters Google search results and provides profanity detection and content warnings across websites. Users can alter behaviour and sensitivity. It uses preset patterns to classify HTML nodes with links, eliminating Wikipedia results. It observes changes in search page content with a mutation observer and performs sentiment analysis with the AFINN lexicon. It is implemented online for efficiency using a Multinomial Naive Bayes Classifier and Natural Language Processing. Users can change the sensitivity, override default settings, and add topics to their blacklist. They can also choose to remove or blur information that contains blacklisted keywords and get alerts when they visit websites.

Obscenity detection is improved by the suggested Ensemble learning with Attention-based Yv3 paired with CFC3 loss (EAYv3-CFC3) approach [6]. It makes use of YOLOv3 (Yv3) as the foundational network and integrates CBAM and sandglass blocks into an ensemble backend feature extractor. The feature map loss is handled by the CFC3 loss, which combines the C3 and CFLoss functions. EAYv3-CFC3 outperforms sequential models with 98.85% accuracy, according to performance analysis. The accuracy enhancement justifies a minor increase in computational complexity. AGOI dataset evaluation reveals a 3.75% accuracy gain over clean pictures. When compared to the most advanced techniques, EAYv3-CFC3 processes data faster (0.038 seconds) while maintaining greater accuracy.

III. METHODOLOGY

This project includes an extension that helps users hide NSFW information in browsers while browsing websites. It provides dynamic restriction settings for those interested in such content, allowing for customization. JavaScript, TensorFlow [10], HTML, and CSS were used during development.

A machine learning technique is used in the system to identify possibly unsuitable content as shown in the Fig 1. Data transfer is quick and efficient, with a streamlined process [7]. The design blends the extension's real-time monitoring skills with the model's [11] prediction powers. This ensures that each extracted component is properly classified before server analysis. Blurring [12] is one type of content control approach that may be used to address offensive content[13]. This solution prioritizes user privacy while providing a safer browsing experience through rigorous data collection and processing.

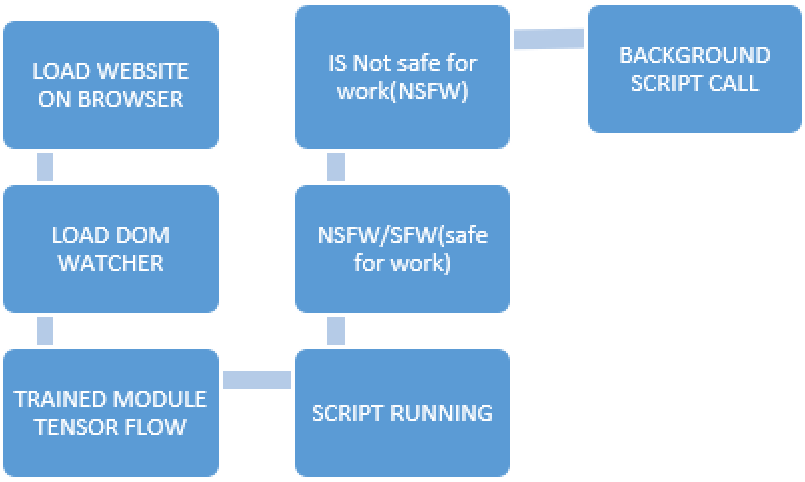


Fig 1(Extension linkage description)

LOAD DOM WATCHER:

Browser extensions are software utilities that intensely improve the web browsing experience. They accomplish this by dynamically interacting with web pages, which allows them to change with and adapt to changes instantly. Extensions must be able to monitor changes inside the Document Object Model (DOM), the blueprint that specifies the structure and content of a webpage. This is where DOM watchers come into play.

Consider a DOM watcher as a specialised observer for a browser extension. It constantly watches the DOM and logs any changes that occur [14]. When a change occurs, the DOM watcher activates, executing a specified function designed explicitly for the extension. This enables the extension to respond appropriately to the specific change it has noticed.

TENSOR FLOW

TensorFlow excels in creating models to train on massive amounts of data, thus interpreting it excellent building extensions for browsers involving machine learning operates that involve sentiment detection along with picture recognition as show in the Fig 2. This procedure occurs outside of the extension itself. Following training, the model [15] can be quickly translated into a smaller format suited for use in browsers. Libraries such as TensorFlow.js allow for the integration of this pre-trained model into the extension, giving it intelligent capabilities.

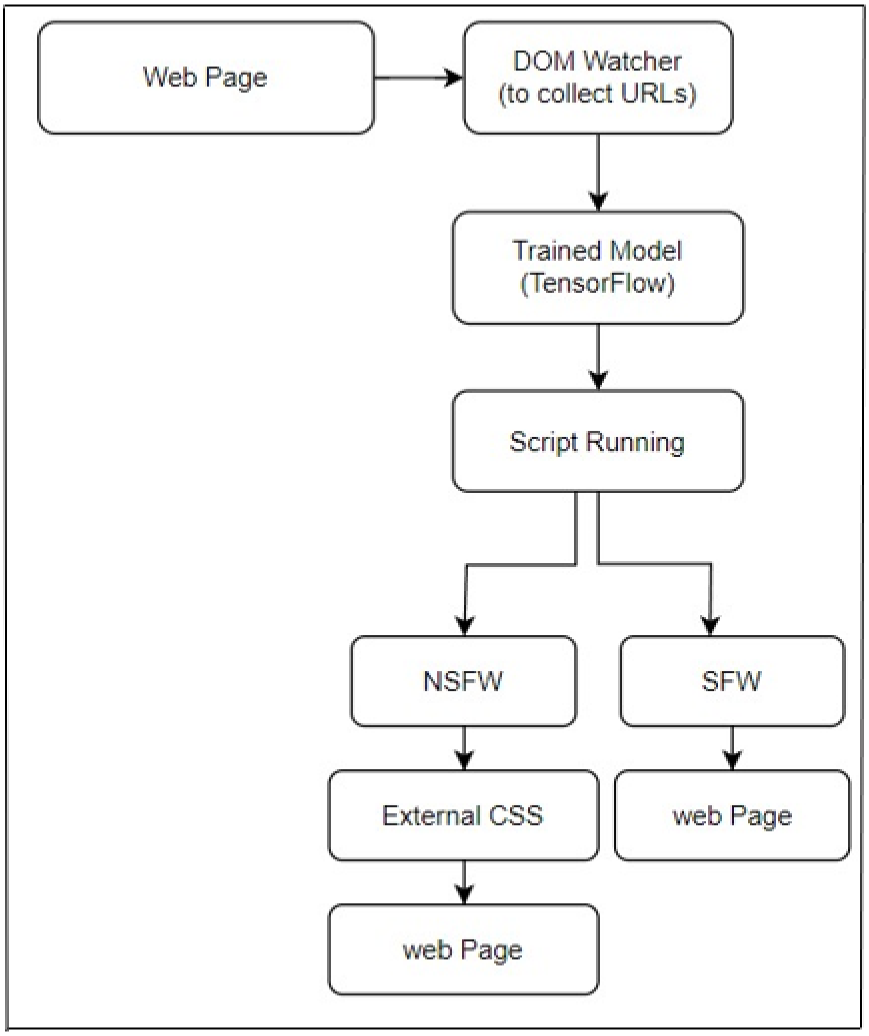


Fig 2(Experimental architecture for validation of system)

TensorFlow is a useful tool for testing and developing. It enables immediate replication of the machine-learning model's behaviour in the environment of the browser extension. This procedure improves the extension's logic, ensuring that it is consistent with the model's predictions. After attaining the required [10] functionality, TensorFlow.js could be integrated into the extension to deploy the pre-trained model in real-world use cases.

SCRIPT OPERATING IN BACKGROUND

Browser extensions can interpret this vision into reality through the implementation of background scripts to dynamically filter and hide undesired images. These background scripts run in the background, continually monitoring the screen for unwanted images. They can be customised with a variety of filtering options, allowing users to customise the experience to their preference.

In contrast to content scripts associated with unique webpages, the background scripts run in the background continually. This allows them to look through each webpage that loads and find images that fit the filter. Images from blocked domains [16] or those with image alt tags including search terms fall under this category. Furthermore, inappropriate materials could be detected through background scripts using picture recognition methods that depend on visual attributes. This can be especially helpful in locating and eliminating inappropriate or unnecessary pictures that might escape keyword-based filters.

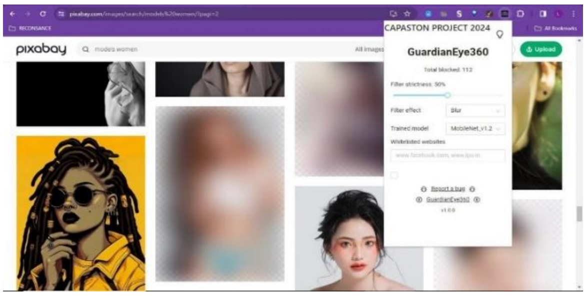
This comprehensive filtering method guarantees that no undesired image escapes identification, promoting a more efficient and customised surfing experience. Consumers can experience a clutter-and distraction-free online experience, which improves their pleasure level when browsing.

NSFW

NSFW filtering extensions act as protectors, defending users from content that can be deemed improper. These expansions use a combination of techniques. Curated lists of well-known NSFW websites are updated often to either obscure or ban content when it matches. Create customised filtration rules are based on keywords or patterns so you can further customise the browsing experience for users [3]. The most sophisticated method analyses photographs by applying machine learning algorithms that have been trained on a large amount of labelled content. This allows for the amazing accuracy of spotting possibly unsuitable content. These addresses growing patterns or subtle indications that could elude keyword filters. When an extension detects a potentially offensive image, it provides several control options. Some allow viewers to make educated viewing decisions through blocking the image entirely, while others blur it for partial visibility or show a warning notice.

This all-encompassing strategy provides a strong barrier against inappropriate interactions on the internet by enabling individuals to design a secure and distraction-free surfing experience. Browser extensions that filter images provide users the ability to create a more organised and distraction-free online experience. These additions go beyond simply banning offensive content. They facilitate personalisation of the browsing experiences by offering a variety of filtering options [8]. Images that may be considered offensive are detected using techniques such as blacklists, image analysis, and user-defined criteria. With the use of extensions, you can hide, obscure, or offers warnings on content that has been marked.

IV. RESULT ANALYSIS



CAPASTON PROJECT 2024

GuardianEye360

Total blocked: 112

Filter strictness: 50%

Filter effect: Blur

Trained model: MobileNet_v1.2

Whitelisted websites: www.facebook.com, www.ipo.in

Report a bug

GuardianEye360

v1.0.0

Fig 3 (Enabling of extension)

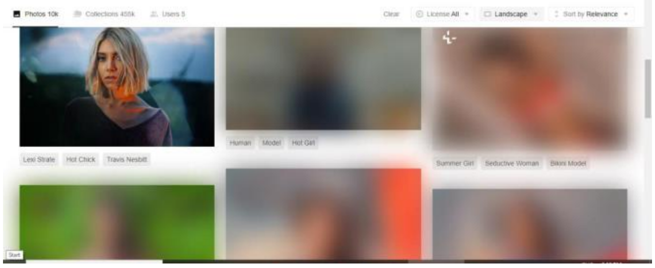


Fig 4 (Response to obscene content vs. non-obscene content)

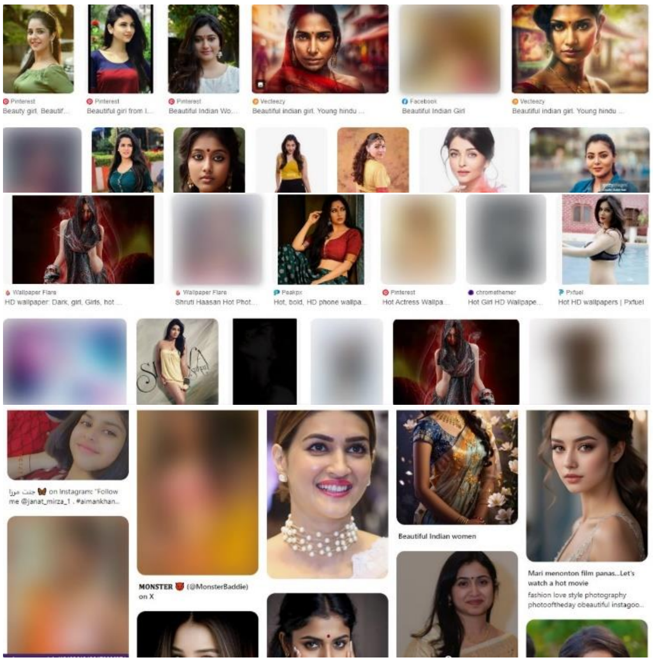

Pinterest
Beauty girl. Beautif...
Pinterest
Beautiful girl from i...
Pinterest
Beautiful Indian Wo...
Vecteezy
Beautiful indian girl. Young hindu ...
Facebook
Beautiful Indian Girl
Vecteezy
Beautiful indian girl. Young hindu ...
Wallpaper Flare
HD wallpaper: Dark, girl, Girls, hot ...
Wallpaper Flare
Shruti Haasan Hot Phot...
Peakpx
Hot, bold, HD phone wallpa...
Pinterest
Hot Actress Wallpa...
Hot Girl HD Wallpape...
Pxfuel
Hot HD wallpapers | Pxfuel
Beautiful Indian women
MONSTER (@MonsterBaddie) on X
Mari menonton film panas...Let's watch a hot movie

Fig 5(Response to obscene content vs. non-obscene content)

GuardianEye360 is powerful analysis of images abilities is an essential part of its functionality Unexpected images that pop up while browsing can be annoying and interrupt with an uninterrupted seamless surfing experience. The answer is GuardianEye360, which puts the needs of the user and a more customised browsing experience first. Instead of banning whole web pages, this content screening technology does this by blurring objectionable images [11]. The results are shown in the Fig 5.

The ability to generate blacklists gives individuals the ability to keep individualised lists of websites that are known to host offensive content. Websites are compared to this list by the system. GuardianEye360 carefully blurs the objectionable image on that one specific website if a match is found, preserving uninterrupted access to the remaining material.

GuardianEye360 provides even more detailed oversight with customised filter rules, going beyond blacklists. These guidelines can be made using pattern or words found in the images. stepping upon a webpage that has pictures on it that include a particular phrase. GuardianEye360 can automatically blur those images while maintaining the remaining portion of the information visible by implementing a blurring rule that targets that phrase.

This focused strategy has various benefits. The ability to customise of the blacklist, filter [14] rules, and the degree of obscuring offers individuals complete control over the system, enabling them to adapt it to their unique requirements and tastes. Compared to website blocking, blurring offers a more covert and privacy-focused option. Users are exposed to less unwanted images while remaining can find the webpage's core content. Furthermore, blurring photos can aid in lowering data usage, which is especially advantageous for customers with data-limited plans.

It is imperative to bear in mind that the efficacy of GuardianEye360 is contingent upon appropriate configuration. To guarantee best performance, blacklists [9] and filter rules must be regularly maintained. Even while GuardianEye360 offers a useful degree of security, using caution when browsing the internet is still necessary for a secure and safe online experience.

V.FUTURE SCOPE

The research paper on GuardianEye360 offers a promising solution for safer web browsing. Currently, GuardianEye360 focuses on image filtering. Consider incorporating functionalities to detect and filter sensitive text content. Explore pre-trained models that can classify NSFW [2] content into subcategories for users to define specific filtering preferences. Allow users to define the level of filtering (e.g., blur, block entirely) for different NSFW categories. Investigate real-time image analysis to detect and filter sensitive content dynamically while browsing. Provide users with clear explanations on how GuardianEye360 identifies and filters content. Consider the possibility of offline image filtering for situations with limited internet access [15]. Address data storage concerns by exploring on-device storage or federated learning techniques to minimize privacy risks.

By exploring these future research areas, GuardianEye360 can become an even more robust and user-friendly tool for creating a safer and more secure online browsing experience.

VI.CONCLUSION

GuardianEye360" is a web extension designed to enhance online browsing safety by automatically identifying and blurring sensitive or obscene content on websites. Upon loading a webpage, the extension's backend processes the content using a pre-trained model, which distinguishes between NSFW (Not Safe For Work) and NON_NSFW [2] content categories. Any identified NSFW content is promptly blurred, while the remaining content is displayed without interruption. What sets "GuardianEye360" apart is its user-friendly interface and customizable features, empowering users with control over their browsing experience. The extension seamlessly integrates into popular web browsers, providing a solution for users to navigate websites without fear of encountering inappropriate material. With the future introduction of the "GuardianEye360" extension, users can confidently explore the internet knowing they have a reliable tool at their disposal to safeguard against encountering offensive or explicit content [1],[2]. This initiative aims to promote a more secure and comfortable online environment for all users, fostering a positive browsing experience devoid of hesitation or concern.

VII. REFERENCE

[1]. N. Gautam and D. K. Vishwakarma, "Obscenity Detection in Videos Through a Sequential ConvNet Pipeline Classifier," in IEEE Transactions on Cognitive and Developmental Systems, vol. 15, no. 1, pp. 310-318, March 2023, doi: 10.1109/TCDS.2022.3158613.

[2]. P. Taneja, D. Singh and T. Rajora, "A Safer Web Experience: Deep Learning-Enhanced Obscene Content Filtering Plugin," 2023 2nd International Conference on Futuristic Technologies (INCOFT), Belagavi, Karnataka, India, 2023, pp. 1-4, doi: 10.1109/INCOFT60753.2023.10425199.

[3]. S. Samal, Y. -D. Zhang, J. M. G. Saez, S. -H. Wang, B. K. Balabantaray and R. Nayak, "EAYv3-CFC3: Ensemble Learning With Attention-Based Yv3 Combined With CFC3 Loss for Obscenity Detection," in IEEE Transactions on Emerging Topics in Computational Intelligence, doi: 10.1109/TETCI.2023.3320553.

[4]. Mathews, Noble & Chimalakonda, Sridhar. (2021). Detox Browser -- Towards Filtering Sensitive Content On the Web.(Detox)

[5]. D. C. Moreira, E. Torres Pereira and M. Alvarez, "PEDA 376K: A Novel Dataset for Deep-Learning Based Porn-Detectors," 2020 International Joint Conference on Neural Networks (IJCNN), Glasgow, UK, 2020, pp. 1-8, doi: 10.1109/IJCNN48605.2020.9206701.

[6]. S. L. Hor et al., "An Evaluation of State-of-the-Art Object Detectors for Pornography Detection," 2021 IEEE International Conference on Signal and Image Processing Applications (ICSIPA), Kuala Terengganu, Malaysia, 2021, pp. 191-196, doi: 10.1109/ICSIPA52582.2021.9576796.

[7]. C. Liambas and A. Manios, "Pornography Image Detection in Digital Forensics," 2023 8th International Conference on Frontiers of Signal Processing (ICFSP), Corfu, Greece, 2023, pp. 88-92, doi: 10.1109/ICFSP59764.2023.10372879.

[8]. T. C. Nagavi and A. D. S., "Detection and Classification of Toxic Content for Social Media Platforms," 2021 4th International Conference on Recent Developments in Control, Automation & Power Engineering (RDCAPE), Noida, India, 2021, pp. 368-373, doi: 10.1109/RDCAPE52977.2021.9633647.

[9]. N. Aldahoul et al., "An Evaluation of Traditional and CNN-Based Feature Descriptors for Cartoon Pornography Detection," in IEEE Access, vol. 9, pp. 39910-39925, 2021, doi: 10.1109/ACCESS.2021.3064392.

[10]. Awad, Abdelrahman Mohamed, et al. "Development of automatic obscene images filtering using deep learning." Advances in Robotics, Automation and Data Analytics: Selected Papers from iCITES 2020. Springer International Publishing, 2021.

[11]. Samal, S., Nayak, R., Jena, S., & Balabantaray, B. K. (2023). Obscene image detection using transfer learning and feature fusion. Multimedia Tools and Applications, 82(19), 28739-28767.

[12]. Bargavi, Manju, Sakshi Dhruva, Tenzin Kunsang, S. Subham Patra, and Tenzin Nyima. "Icensor: Unwanted Image Detection and Censoring." (2023).

[13]. Rautela, K., Sharma, D., Kumar, V., & Kumar, D. (2024). Obscenity detection transformer for detecting inappropriate contents from videos. Multimedia Tools and Applications, 83(4), 10799-10814.

[14]. Al Naffakh, H. A. H., Ghazali, R., El Abbadi, N. K., & Razzaq, A. N. (2021). A review of human skin detection applications based on image processing. Bulletin of Electrical Engineering and Informatics, 10(1), 129-137.

[15]. Mazinani, M. R., & Ahmadi, K. D. (2021). An Adaptive Porn Video Detection Based on Consecutive Frames Using Deep Learning. Rev. d'Intelligence Artif., 35(4), 281-290.

[16]. Phan, D. D., Nguyen, T. T., Nguyen, Q. H., Tran, H. L., Nguyen, K. N. K., & Vu, D. L. (2022). Lspd: A large-scale pornographic dataset for detection and classification. International Journal of Intelligent Engineering and Systems, 15(1).