Journal of Current Research in Engineering and Science
Bi-Annual Online Journal (ISSN: 2581 - 611X)
VOLUME:8, ISSUE:2, AUGUST 2025, PAPER NUMBER:33

Hawkeye360- Ensuring Content Safety with Precision

Nidhi Singh[1], Kadali Lakshmi Kanishka[2], Manasa Kavvampally[3], Triloki Singh[4], Pooja Sharma[5],

[1,2,3,4] Students, and[5] Faculty

Dept. of Computer Science Engineering,
Lovely Professional University, Punjab, India.
nidhisinghbgr@gmail.comm

Abstract: The world's use of search engines and the internet is growing at an exponential rate. In all industries, including business, retail marketing, healthcare, education, and more. While perusing the internet, people will come across a variety of undesired materials, from vulgar to educational. With the help of TensorFlow, CNN (Convolutional Neural Networks), cascades, and JavaScript, this project presents Hawkeye360, a sensitive image detecting system. The material is subsequently loaded into the Document Object Model (DOM) using these characteristics. Additionally, the CNN using the TensorFlow framework blurs the pornographic information on the loaded webpage after detecting the images on it. Not Safe-For-Work (NSFW) content can filter with the use of a trained model, which is built into this inventive browser extension. The extension that classifies found photos into Safe-For-Work (SFW) or NSFW domains. Online exposure to sensitive content is a problem that Hawkeye360 addresses. This feature helps to protect user experiences when browsing by enabling users to quickly mask or cover up anything that is excessively sensitive. You can create a more secure and safe online environment by giving people total control over their connections to the internet.

Keywords: Obscene picture identification, TensorFlow, Convolutional Neural Network, storage, URLS, image filtering, user permission, and Document Object Model Not Safe for Work; imagine blurred.

1. INTRODUCTION:

The risk of coming into contact with sensitive content is rising along with the internet's rapid rise in popularity and ease of access [1]. Due to social media's and search engines' convenience and usability, more people worldwide are relying on online content, which is further putting users in situations where they are exposed to offensive or sensitive material even when they are actually trying to find something else [2]. Because of the widespread adoption of an economy that quickly accepted an excess of sexualization, expressive themes, images, and material are tied to many facets of our lives [3]. Giving inappropriate attention to sexual content can have very harmful impacts, which is another reason why it's necessary to address it for educational purposes. A growing body of research suggests that exposure to sexually explicit content may have a role in the development of reserved behaviour [4]. Given that a significant section of the world's population, including adults and children, spends extended lengths of time online ingesting various media and data, it is imperative to distinguish between sexual and instructional information. This is especially important in terms of safeguarding youth mental health [5]. It is obvious that children and teens cannot relate to the implications of being exposed to dirty and pornographic [6] information, and it may have long-term effects on their mental health [7]. Whether at work or at home, accidentally coming across offensive photos can be awkward and embarrassing. Unintentionally coming across sexual information while browsing a website might cause tension among coworkers and ruin relationships based on trust. One of these other situations is when you're at home viewing a website and it has some explicit stuff. It can be upsetting, especially if kids are present [8]. It can be distressing and difficult to handle when a youngster is unintentionally exposed to unsuitable information, particularly if they are too young to completely comprehend or process it. Online advertisements are one of the conditions causes [9]. This makes people less likely to visit the websites, which has an adverse effect on their entire browsing experience. The creation of Hawkeye360, which gives users a strong tool to protect themselves against objectionable content, marks a substantial improvement

in internet safety. This browser extension uses a pre-trained model that guarantees strong detection performance, exploiting a wide dataset containing explicit material. User-friendly interfaces and real-time detection techniques [2] enable its easy integration into well-known web browsers such as Chrome and Firefox. This allows for quick picture analysis while webpages load and the application of efficient blur or filter techniques upon unsuitable content identification. One of Hawkeye360's important advantages is its customizable settings, which improve accessibility and convenience by letting users ban particular websites and change sensitivity levels. Additionally, this extension puts user privacy first by enforcing privacy safeguards to protect sensitive data. Hawkeye360 functions as a powerful parental control tool due to these technical advancements, empowering users to impose customized content limitations and promote a safer online environment—especially for young people. The extension's dedication to selecting age-appropriate content and reducing the risks associated with uncontrolled internet access is demonstrated by its capacity to implement obscenity blockers and adaptive filtering approaches. Users may confidently browse the digital environment using Hawkeye360, knowing they are fully protected from coming across explicit material. Innovations like Hawkeye360, which enable users to retain control over their online experience and guarantee a safer and more secure internet for everyone, are essential to the ongoing evolution of technology.

2. RELATED WORK:

To increase online safety, P. Taneja, D. Singh, and T. Rajora [2] created a browser plugin that monitors browsing habits and detects offensive content. The JavaScript-powered browser extension logs user activity and transmits data to a central server that recognizes problematic material using a machine learning model. Pre-processing of the data, such as resizing photographs, converting to the RGB colour standard, and doing frame-by frame analysis for movies and gifs, guarantees consistency and quality. Data is stored in a MongoDB database that is organized by content type and shared with internet security businesses. The ConvNeXt model showed how important preprocessing and dataset size are by accurately categorizing NSFW photos with a high degree of accuracy. These results validate ConvNeXt's efficacy in NSFW material identification and will influence future studies on picture classification.

The paper by Samal, S., Nayak, R., Jena, S., & Balabantaray, B. K. [3] discusses the issue of automatically detecting objectionable or obscene content in videos. It introduces the novel deep-learning transformer-based Obscenity Detection Transformer (ODT) [8], which stresses the utilization of particular information to increase detection accuracy. The large short-term memory layers and vision transformer enable the model to extract meaningful characteristics from video frames. Extensive tests on the Pornography-2k and Pornography-800 datasets demonstrate greater performance than CNN-based models, with accuracies of 99.6% and 98.8%. Preprocessing involves converting films into frames and annotating short video portions to obtain relevant contextual information. The model effectively manages positional embeddings and improves classification accuracy through the use of multi-head attention mechanisms and GELU activation functions. Using LSTM layers to improve temporal dependency modelling also results in better detection performance.

In order to detect pornography, D. C. Moreira, E. Torres Pereira, and M. Alvarez [4] presented a dataset of 376K images. The study addresses the problem of incomplete data and arbitrary classification. Images from Reddit were collected, showcasing a range of scenarios from regular life to explicit content. A stringent definition of pornography was developed using particular standards. Moreover, a standardization strategy was proposed for the NSFW photo moderation API results, enabling cross service comparability. The paper also presents a CNN model that uses the PEDA 376K dataset and is based on convolutional networks. This model uses a successful hyperparameter selection technique. Given the circumstances, the work creates the groundwork for future research in this area and advances the field of pornography detection investigations.

Detox Browser is a Chrome addon developed by NOBLE SAJI MATHEWS and SRIDHAR CHIMALAKONDA [5] that filters Google search results and offers content warnings and profanity detection on many websites. Users can change their sensitivity and behaviour. It filters Wikipedia results by classifying HTML nodes with links using preset patterns. It uses an AFINN lexicon for sentiment

analysis and a mutation observer to track changes in the content of search results pages. For efficiency, it is implemented online with Natural Language Processing and a Multinomial Naive Bayes Classifier. Users have the ability to add topics to their block list, modify the sensitivity, and override default settings. Additionally, they can decide to ignore or obfuscate content that includes block-listed keywords and receive notifications whenever they visit websites.

The proposed Ensemble training with Attention-based Yv3 paired with CFC3 loss (EAYv3-CFC3) technique improves obscenity detection [6]. YOLOv3 (Yv3) is used as the base network, while CBAM and Sandglass blocks are integrated to create an ensemble backend feature extractor. The CFC3 loss, which combines the C3 and CFLoss functions, manages the feature map loss. Performance research shows that EAYv3- CFC3 outperforms sequential models with 98.85% accuracy. A slight increase in computing complexity is justified by the accuracy improvement. An assessment of the AGOI dataset shows a 3.75% improvement in accuracy over clean images. EAYv3-CFC3 analyses data more quickly (0.038 seconds) and more accurately than the most sophisticated methods.

3. METHODOLOGY:

The goal of the project is to create a browser extension that will let users hide NSFW (Not Safe for Work) information when they visit the internet. With the use of this extension's dynamic restriction settings, users may customize their browsing to suit their tastes for particular types of information. Utilizing JavaScript, TensorFlow, HTML, and CSS during the development process allowed for the building of a reliable and approachable solution.

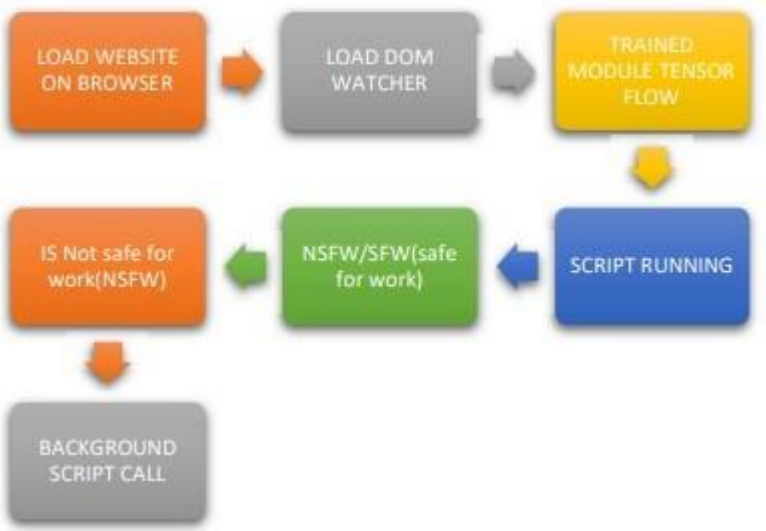

Fig 1(Description of the extension linkage)

The system's use of machine learning algorithms to detect potentially inappropriate content is a crucial component. This is made possible by an optimized process that guarantees rapid and effective data transfer. The machine learning model's prediction capabilities and the extension's real-time monitoring features are smoothly integrated by the design. The approach improves identification and handling content by carefully classifying each extracted component prior to server analysis. Blurring, grey scaling and hiding are some of the content control strategies that the system uses to address objectionable content. This method offers a safer browsing experience without compromising user privacy. By means of comprehensive data gathering and analysis, the system seeks for and provides users with efficient methods to safeguard themselves from coming across unsuitable content on the internet. The initiative seeks to enhance the security and usability of the internet by integrating the latest technologies with user-focused design concepts.

A. LOAD DOM WATCHER:

Browser extensions are useful software tools that significantly enhance online surfing experience. They are able to instantaneously adjust to changes by means of dynamic interaction with web sites. The Document Object Model (DOM), the blueprint that describes the layout and content of a webpage, must be able to be monitored by extensions. This is the application of DOM watchers. Think of a DOM watcher as a specialized observer for an addon for a browser. It continuously monitors the DOM and records any alterations that take place [14]. Upon detecting a modification, the DOM watcher initiates, carrying out a designated function specifically created for the extension. This allows the extension to react to the particular change it has detected in an appropriate manner.

B. TENSOR FLOW

As seen in Fig. 2, TensorFlow is a valuable tool for developing machine learning extensions for browsers that entail sentiment detection and photo identification. It excels at generating models that can be trained on enormous quantities of data. This process takes place apart from the extension. The model [15] may be immediately converted into a smaller version appropriate for browser use after training. The addon gains intelligence through the incorporation of this pretrained model made possible by libraries like TensorFlow.js. When it comes to testing and creating machine learning models, TensorFlow is quite helpful, especially when it comes to browser extensions. TensorFlow makes it easier to enhance the logic of the extension iteratively and guarantees that it remains consistent with the model's predictions by enabling instantaneous replication of the model's behaviour within the extension environment. When the required functionality is attained, TensorFlow.js integration allows the pre-trained model to be deployed in real-world circumstances, enabling the extension to efficiently detect and manage NSFW material in real-time while preserving browser compatibility. This method makes the most of the extension's capacity to provide visitors a safer online experience by utilizing machine learning to analyses dynamic information.

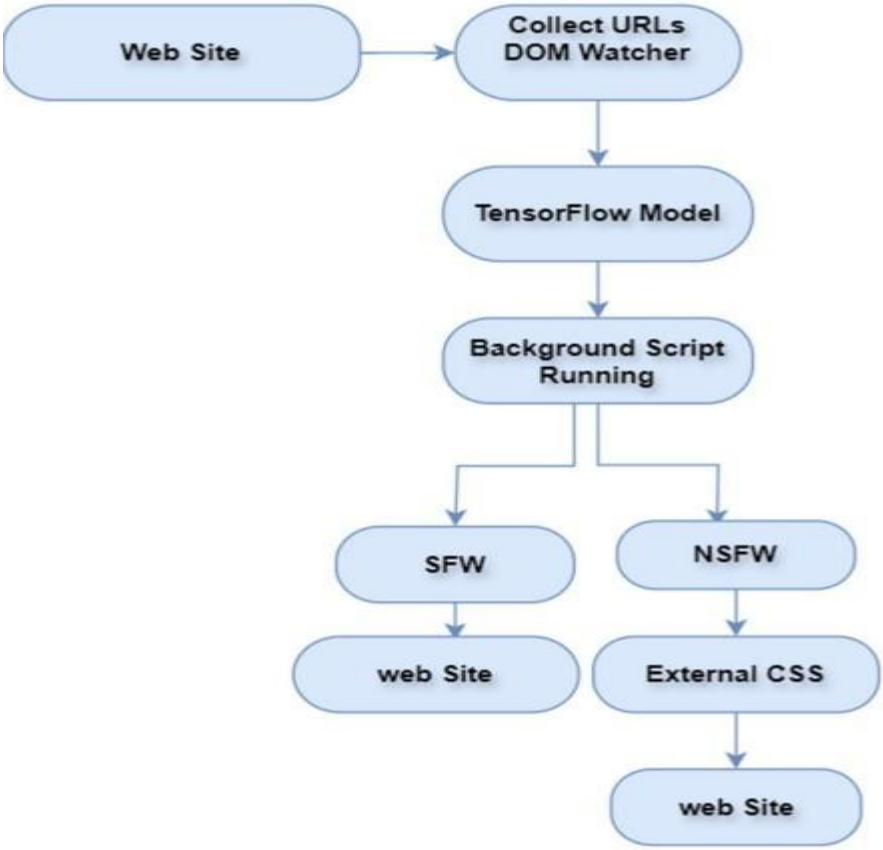

Fig 2 (A validation architecture to check the system)

C. BACKGROUND SCRIPT OPERATION:

With the help of background scripts, browser extensions may turn this vision into reality by automatically filtering and hiding unwanted pictures. While operating in the background, these scripts scan the screen continuously for undesirable visuals. Users may personalize the experience to suit their preferences by customizing them with an array of filtering choices. Unlike content scripts linked to distinct webpages, background scripts continuously operate in the background. They can now go through every webpage that loads and locate pictures that match the filter by doing this. This category includes images from restricted domains [16] and images whose image alt tags contain certain search phrases. In addition, by utilizing picture recognition techniques that rely on visual characteristics, background scripts might be used to identify unsuitable items. Finding and removing unsuitable or excessive images that may escape keyword-based filters can be made easier using this. A more effective and personalized browsing experience is encouraged by this thorough filtering process, which ensures that no unwanted image fails to reach detection. Customers might enjoy a more enjoyable web browsing experience when there are fewer obstacles and distractions.

D. NOT SAFE FOR WORK (NSFW):

Users are shielded from potentially offensive information by NSFW filtering extensions. These extensions employ a variety of methods. Well-known NSFW websites' curated lists are often updated to either hide or prohibit content that is appropriate. To further personalize the experience of browsing for users, use custom filtration rules based on keywords or patterns [3]. The most advanced technique uses machine learning algorithms to analyse photos once they have been trained on an extensive amount of labelled material. This makes it possible to identify potentially inappropriate information with remarkable precision. This deals with emerging trends or weak indications that might be missed by search filters.

An extension offers many control choices when it finds a picture that can be objectionable. Some blur the image for partial sight or display a warning message, while still others block the image completely to give viewers the opportunity to make informed viewing decisions. This all-inclusive approach lets users create a safe and distraction-free browsing experience, which acts as a powerful barrier against unwanted interactions on the internet. With the use of browser extensions, users may make their time online more organized and distraction-free. These enhancements go beyond essentially prohibiting objectionable material. By providing a range of filtering choices, like filtering strictness they make it easier to customize surfing experiences [8]. Image analysis, block lists, and user-defined criteria are some of the methods used to identify images that can be declared objectionable. You can obfuscate, conceal, or display cautions on information that has been clearly marked by using extensions.

4. RESULT ANALYSIS

One of Hawkeye360's most notable features is its ability to analyses photos powerfully. Unwanted pictures that appear while you're online can be frustrating and conflict with your ability to browse without interruption. Hawkeye360 is the solution because it prioritizes user demands and a more personalized online encounter. This content filtering technique blurs, greyscale or hides undesirable photos, rather than prohibiting entire web pages [11]. The findings are displayed in Fig. 5,6,7. The capacity to create block lists allows users to maintain customized profiles of webpages that contain objectionable content. The system compares websites to this list. If a match is identified, Hawkeye360 delicately blurs the offensive picture on that one particular website while maintaining unhindered access to the rest of the content.

Fig 3 (Graphical representation of extension having customizable features)

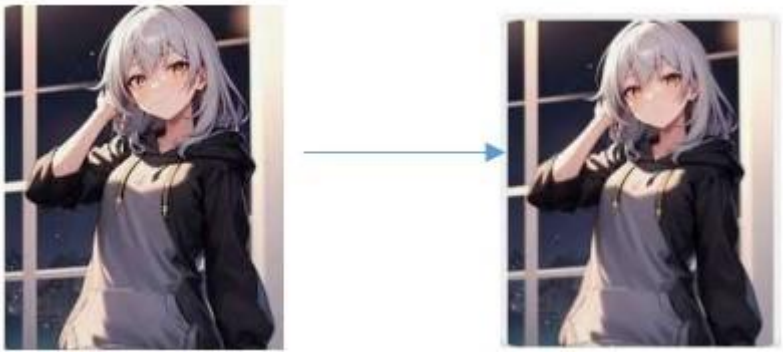

Fig 4 (SFW image)

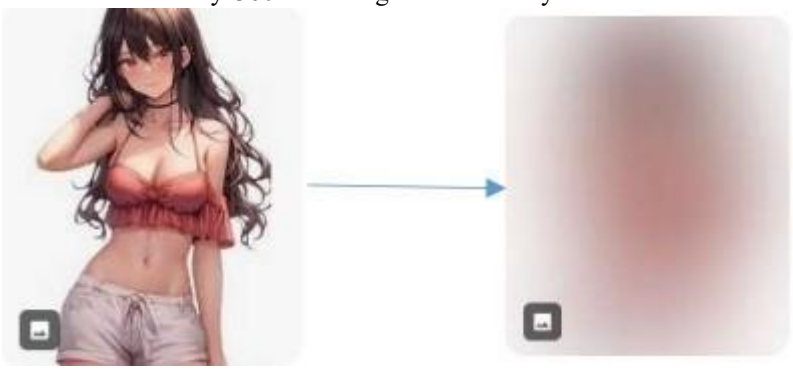

Fig 5 (NSFW image blurred)

Fig 6 (NSFW image Grey Scaled)

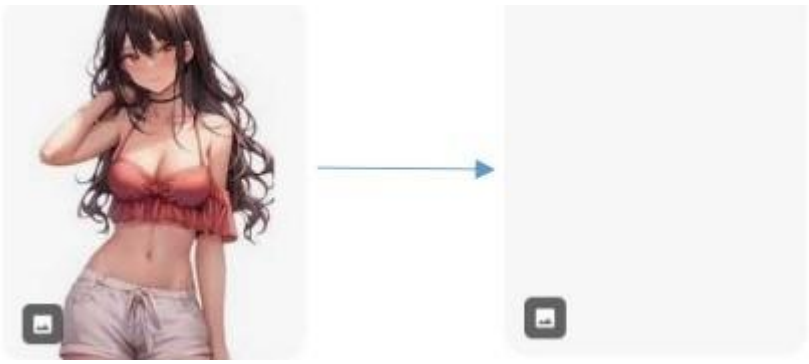

Fig 7 (NSFW image hidden)

Beyond block lists, Hawkeye360 offers even more precise supervision through customized filter rules as shown in fig5,6,7. Patterns or words identified in the photographs can be used to create these recommendations. strolling over a webpage with images that display a specific word. By using the blurring rule that specifically targets that word, Hawkeye360 may automatically blur certain photos while keeping the other part of the content viewable.

Fig 8 (Results on Sexually explicit images Vs Normal images)

This targeted approach offers a number of advantages. People may fully customize the list of undesirable content, filter [14] rules, and degree of concealing to suit their own preferences and needs they can blur, hide and greyscale the images. This gives them total control over the system. obscuring provides a more discrete and privacy-focused solution than website blocking. While still being able to access the primary content of the website, users are subjected to fewer undesirable graphics. Moreover, blurring images can help reduce data use, which is particularly helpful for users of bandwidth-limited plans. It is important to remember that proper setup is necessary for Hawkeye360 to function effectively. Regular maintenance of filter rules and block lists [9] is necessary to ensure optimal performance. Hawkeye360 provides a proficient level of protection, but for a safe and secure online experience, exercise caution whenever you browse the internet.

5. FUTURE SCOPE

In the area of imagine filtering, in particular, the study writing on Hawkeye360 offers a potential path toward improving the safety of web browsing. It is suggested that features for identifying and filtering obscene content be added to further enhance its capabilities. Hawkeye360 would be able to offer complete security with this extension since it would cover both textual and visual NSFW content. Through the examination of pre-trained models that are able to categorize non-sexual images into various categories, users may establish customized filtering choices based on their own requirements. This method gives consumers more control over the surfing experience by allowing them to choose the degree of filtering, from blurring to completely barring various NSFW categories. Additionally, by looking into real-time image analysis methods, Hawkeye360 automatically identifies and filter critical information while people browse the internet. This real-time feature makes sure that users are always shielded from unsuitable content, which increases the extension's overall effectiveness. Taking into account the potential for offline image filtering also reduces worries about restricted internet access. Through the investigation of options like federated learning methods or on-device storage, Hawkeye360 can reduce privacy hazards related to data storage and offer reliable security independent of internet access. Future developments in these fields might make Hawkeye360 an even more powerful and intuitive tool that provides users with a secure and safer internet surfing experience.

6. CONCLUSION

"Hawkeye360" is a groundbreaking web plugin that automatically identifies and blurs offensive or sensitive information that is viewed while browsing the internet, therefore revolutionizing online safety. Not Safe For Work (NSFW) and non-NSFW (NON_NSFW) content categories are quickly distinguished by its backend processing, which is driven by an advanced pre-trained model when a webpage loads. This differentiation guarantees that any information deemed inappropriate for children is quickly hidden, maintaining the excellence of the online experience, and protecting children from potentially harmful content. "Hawkeye360" gives people a dependable way to browse online with self-assurance, aware they are protected by a watchful guardian shielding them from unwanted information by expertly integrating this feature into well-known web browsers.

The focus "Hawkeye360" places on user control and customization is what really makes it stand out. Users have complete control over the extension because of its user-friendly interface and adaptable features, which let them customize (blur, greyscale, hide) their browsing experiences to suit their own demands and comfort levels. With this degree of freedom, people may browse the internet more confidently, knowing that they can build a surfing experience that fits within their own personal boundaries and ideals. Users may anticipate a more trustworthy and secure online environment with the impending launch of
"Hawkeye360," where the anxiety of coming across unsuitable information will be eliminated, promoting a happy and stress-free surfing experience for everyone.

7. 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, "EAYv3CFC3: 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 DeepLearning 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).Ahmad, I., & Kumar, R. (2018). DDoS attack detection using machine learning algorithms. In 2018 9th International Conference on Computing, Communication and Networking Technologies (ICCCNT) (pp. 1-5). IEEE.