engineering proceedings

MDPI

Proceeding Paper

Toward the Analytical Assessment of Implicit Harmful Visual Content Using Convolutional Neural Networks †

Shendry Rosero Vásquez [1,2,*] and Julio Ricardo Proaño Orellana [1]

[1] Carrera de Computación, Grupo de Investigación IDEIAGEOCA, Universidad Politécnica Salesiana (UPS), Campus Sur, Quito 170146, Ecuador; jproanoo@ups.edu.ec
[2] Carrera de Tecnologías de la Información, Facultad de Sistemas y Telecomunicaciones (FACSISTEL), Universidad Estatal Península de Santa Elena (UPSE), La Libertad 240204, Ecuador
* Correspondence: sroserov@est.ups.edu.ec
† Presented at the Sixth International Conference on Information Systems and Software Technologies (ICI2ST 2026), Quito, Ecuador, 23–25 September 2026.

Abstract

Implicitly harmful visual content represents a growing challenge in digital environments, particularly when offensive narratives are conveyed through stylized images, cartoons, or other indirect visual representations. Convolutional neural networks (CNNs) offer a promising approach for analyzing visual patterns associated with such content. This study presents an experimental framework based on transfer learning and cosine similarity to assess implicit harmful visual content under controlled conditions. Three pretrained CNN architectures, namely ResNet50, VGG19, and EfficientNetB0, were evaluated using accuracy, recall, F1-score, and confusion matrices. Rather than introducing a novel architecture, this work investigates the suitability of established CNN models combined with similarity-based feature representations as analytical instruments for visual content assessment. Cosine similarity was employed to examine relationships among deep feature embeddings and support the identification of subtle visual patterns. Experimental results demonstrated consistent performance across all evaluated architectures, with ResNet50 and VGG19 achieving the highest classification performance (Accuracy = 0.95; F1-score = 0.95), while EfficientNetB0 obtained competitive results with lower computational requirements. The proposed framework establishes a reproducible methodological basis for future studies involving visually sensitive content, semantic ambiguity, and the assessment of potentially offensive imagery generated by artificial intelligence systems, including text-to-image models.

Keywords: implicit harmful visual content; convolutional neural networks; transfer learning; cosine similarity; visual content assessment; AI-generated imagery

Check for updates

Academic Editor: Leonardo Valdivieso

Published: 18 September 2026

Copyright: © 2026 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license.

1. Introduction

Digital platforms have significantly transformed the way individuals communicate, share information, and express opinions through visual media. Memes, cartoons, stylized illustrations, and other forms of visual content have evolved beyond entertainment, becoming influential mechanisms for shaping public opinion, reinforcing narratives, and generating social interaction in online environments [1,2]. However, the rapid dissemination and anonymity provided by social networks have also facilitated the propagation of harmful visual representations, including cyberbullying-related content, hate-oriented symbolism, and offensive visual narratives targeting both public and private individuals [3].

In recent years, multimodal visual content combining textual and graphical elements has become increasingly relevant in digital communication. Although such content often

appears humorous or symbolic, previous studies have shown that it may indirectly convey harmful messages, reinforce harassment behaviors, or contribute to the normalization of offensive narratives in online communities [4]. From an analytical perspective, this scenario introduces important challenges for artificial intelligence systems, particularly when harmful intent is expressed through implicit visual patterns rather than explicit textual indicators. Consequently, there is a growing need for reproducible analytical frameworks capable of characterizing visually sensitive content under controlled experimental conditions.

To address related problems, several studies have explored deep learning approaches [5] for the classification of offensive multimodal content. For example, architectures combining VGG16 and BiLSTM layers have demonstrated improvements in F1-score compared with traditional classification approaches [4]. Similarly, deeper convolutional architectures such as ResNet152 and DenseNet201 have reported competitive performance in cyberbullying-related classification scenarios, although results vary depending on the complexity and ambiguity of the analyzed visual content [6]. These findings suggest that convolutional neural networks (CNNs) provide suitable representational capabilities for analyzing visually implicit patterns associated with harmful digital content.

Following this line of research, the present study proposes an experimental analytical framework based on three widely adopted CNN architectures: ResNet50, VGG19, and EfficientNetB0. Rather than introducing a novel neural architecture, this work adopts a comparative and methodological perspective focused on evaluating the suitability of established convolutional models as analytical instruments for the assessment of implicit harmful visual content. ResNet50 enables deeper representation learning through residual connections, VGG19 provides architectural simplicity with strong feature extraction capabilities, and EfficientNetB0 offers an efficient balance between accuracy and computational cost through compound scaling strategies [7].

To complement CNN-based feature extraction, cosine similarity [8–11] was employed as a similarity-based analytical metric for evaluating relationships between image representations in multidimensional embedding spaces [12]. Since feature vectors extracted from deep neural networks encode high-level visual representations, cosine similarity becomes particularly relevant in scenarios where directional similarity between representations is more informative than absolute magnitude differences. Unlike Euclidean distance, cosine similarity emphasizes shared representational patterns and exhibits robustness against scale variations [13], making it suitable for the analysis of visually implicit or semantically ambiguous content [14].

This paper describes the complete experimental pipeline, including data acquisition through web scraping, image preprocessing, feature extraction using convolutional neural networks, and similarity-based analytical assessment using cosine similarity. Experimental results obtained from ResNet50, VGG19, and EfficientNetB0 are comparatively analyzed to evaluate their suitability for characterizing implicitly harmful visual representations. In this sense, the present study constitutes a foundational methodological phase within a broader PhD research agenda focused on the evaluation of visually sensitive and potentially offensive images generated by text-to-image (T2I) systems, where assessing semantic ambiguity and visual offensiveness becomes a central research objective.

The main contributions of this work are summarized as follows:

- The development of a reproducible analytical framework for the assessment of implicitly harmful visual content using convolutional neural networks and similarity-based representations.
- A comparative experimental evaluation of ResNet50, VGG19, and EfficientNetB0 under controlled conditions for analyzing visually sensitive content patterns.

https://doi.org/10.3390/engproc2026153006

- The incorporation of cosine similarity as an analytical metric for evaluating relationships between deep feature representations beyond conventional categorical classification.
- The establishment of a methodological foundation for future research focused on the assessment of visually offensive content generated by text-to-image (T2I) systems.

2. Materials and Methods

This study presents an experimental analytical framework for the assessment of implicitly harmful visual content under controlled experimental conditions. The proposed methodology comprises four consecutive phases: data collection through web scraping, image preprocessing and normalization, convolutional neural network training, and similarity-based classification using cosine similarity. The framework was designed to evaluate the suitability of convolutional neural networks as analytical instruments for characterizing visually implicit harmful patterns. Each stage is described below.

The first phase consisted of data collection through web scraping techniques using the Selenium library in Python 3.11.5 to extract images from X (formerly Twitter). Approximately 300 stylized images associated with potentially harmful or cyberbullying-related content and 300 non-harmful stylized images were obtained and stored in separate directories to facilitate subsequent processing.

During the image preprocessing and normalization phase, images that did not meet the established selection criteria, such as ambiguous context or difficult interpretation, were removed. This process resulted in a balanced dataset comprising 235 potentially harmful images and 235 non-harmful images. All images were standardized to a resolution of 800 × 600 pixels to ensure visual consistency, facilitate manual inspection, and support dataset organization. This resolution was selected to balance visual quality and computational efficiency. During CNN processing, the images were internally adapted to the input dimensions required by each pre-trained architecture, following the standard preprocessing procedures associated with transfer-learning implementations. No architectural modifications were introduced to the original CNN models.

For CNN model training, three convolutional neural network architectures were selected: ResNet50, VGG19, and EfficientNetB0. All architectures were evaluated under homogeneous experimental conditions using baseline transfer-learning configurations based on pre-trained CNN models. The classification layer consisted of two output neurons with a softmax activation function, representing the cyberbullying-related and non-harmful classes. Model training was performed using the Adam optimizer with a learning rate of 0.001, a batch size of 32 images, categorical cross-entropy as the loss function, and the models were trained for 20 epochs. The same hyperparameter configuration was maintained for ResNet50, VGG19, and EfficientNetB0 to ensure a fair comparison. No architecture-specific hyperparameter optimization was performed, since the primary objective was to compare the representational capabilities of the pre-trained CNN architectures while maintaining identical experimental conditions across all evaluated architectures. These models were trained and evaluated using standard classification metrics, including accuracy, recall, F1-score, and confusion matrices, to assess their effectiveness in distinguishing between cyberbullying-related and non-harmful cartoons.

Subsequently, a cosine similarity experiment was designed for the classification stage. An interactive interface was developed to allow the user to select a trained model and upload an image from the test set. Cosine similarity was employed as a matching metric between the feature representation of a new image and those of previously classified images, supporting category assignment based on representational similarity.

The overall experimental workflow is summarized in Algorithm 1.

Algorithm 1: Analytical Pipeline for Implicit Harmful Visual Content Assessment

```
Collect and filter images via web scraping using Selenium;
Resize images to 800 × 600 pixels;
Initialize the pre-trained ResNet50, VGG19, and EfficientNetB0 architectures;
Configure two output neurons with softmax activation;
Train the CNN models using Adam (LR = 0.001), a batch size of 32, categorical
  cross-entropy, and 20 epochs;
foreach CNN model (ResNet50, VGG19, EfficientNetB0) do
    Extract deep feature representations from the images;
    Evaluate the model using classification metrics;
Select the trained model and input image via the interface;
Compute cosine similarity between the input image and the labeled images;
Assign the classification label (cyberbullying-related or non-harmful);
```

The overall structure of the proposed experimental framework is illustrated in Figure 1.

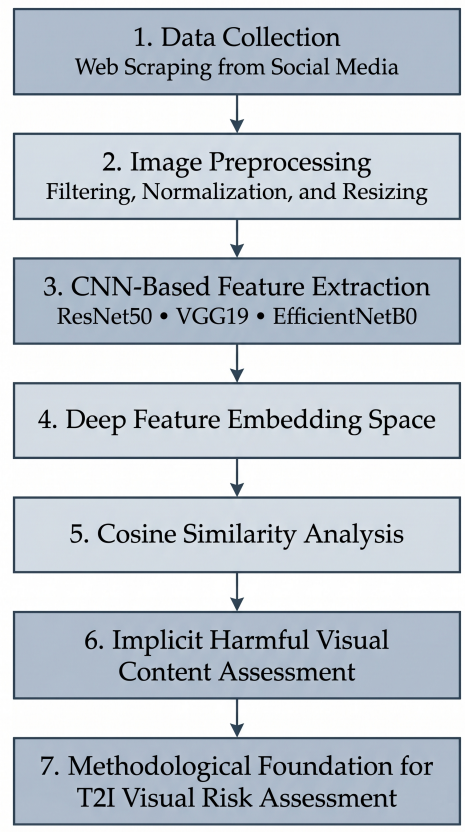


Figure 1. Experimental framework for implicit harmful visual content assessment using convolutional neural networks and cosine similarity.

Cosine similarity was used to compare the feature vector of a new image $\mathbf{I}_{\text{new}}$ with those of the classified images $\mathbf{I}_{\text{trained}}$, and is computed as

$$\cos(\theta) = \frac{\mathbf{I}_{\text{new}} \cdot \mathbf{I}_{\text{trained}}}{\|\mathbf{I}_{\text{new}}\| \|\mathbf{I}_{\text{trained}}\|} \quad (1)$$

where $\cos(\theta) \in [-1, 1]$ represents the degree of similarity between feature vectors. Values close to 1 indicate highly similar image representations, whereas lower values suggest weaker representational correspondence.

For cosine similarity-based classification, similarity scores were computed between the input image and all images in the training set. The class associated with the highest similarity score was subsequently assigned to the input image. To facilitate reproducibility and data management, image labels and file paths were stored in a CSV file and loaded during the classification process.

The dataset was divided into 80% for training and 20% for testing. During training, network weights were optimized according to the selected loss function, enabling a reliable evaluation of the proposed analytical framework.

Experimental Design and Reproducibility

The proposed framework was developed under controlled experimental conditions to ensure consistency and reproducibility across all evaluated CNN architectures. The dataset was balanced between the two target classes to minimize classification bias during both training and evaluation. Furthermore, all images were processed using the same preprocessing pipeline, including filtering, normalization, and resizing procedures, thereby ensuring homogeneous input conditions for feature extraction and classification.

To maintain methodological consistency, an identical training and testing partition strategy was applied to ResNet50, VGG19, and EfficientNetB0. Likewise, evaluation metrics, including accuracy, recall, F1-score, confusion matrices, and cosine similarity analysis, were computed under equivalent experimental settings for all architectures. This design enables a comparative assessment focused on the representational behavior of deep feature embeddings rather than solely on classification performance.

From a methodological perspective, the framework was intentionally structured as a reproducible analytical environment for studying visually implicit harmful content under controlled conditions. In this sense, the proposed approach constitutes a foundational experimental phase for future investigations involving AI-generated visual content and text-to-image (T2I) systems.

3. Results

The performance of the ResNet50, VGG19, and EfficientNetB0 architectures was evaluated using a dataset composed of potentially harmful and non-harmful stylized images. Standard evaluation metrics, including accuracy, recall, F1-score, and confusion matrices, were employed to assess the analytical capability of each model for distinguishing visually implicit harmful content across both target categories.

Table 1 summarizes the comparative performance obtained for each CNN architecture. The results indicate that ResNet50 and VGG19 demonstrate consistent representational capability for distinguishing visually implicit harmful patterns, achieving accuracy, recall, and F1-score values of 0.95. Among the evaluated architectures, ResNet50 exhibited the most consistent analytical behavior across both target categories, reinforcing its suitability for representing harmful visual patterns. Although EfficientNetB0 presented a slightly lower performance (0.94), it remained effective while offering improved computational efficiency, making it suitable for scenarios involving constrained computational resources.

In addition to conventional evaluation metrics, cosine similarity was employed as a similarity-based analytical criterion operating over deep feature representations extracted by the CNN architectures. For a given input image, its feature vector was compared against vectors from the labeled dataset, assigning the category associated with the highest similarity score. This approach emphasizes the orientation of feature vectors within the embedding space, providing robustness when distinguishing visually complex or semantically implicit patterns.

https://doi.org/10.3390/engproc2026153006

Table 1. Comparative analytical performance of CNN architectures for implicit harmful visual content assessment.

Model	Actual	Cyberbullying	Normal	Acc.	Rec.	F1
ResNet50	Cyberbullying Normal	222 9	13 226	0.95	0.95	0.95
VGG19	Cyberbullying Normal	220 10	15 225	0.95	0.95	0.95
EfficientNetB0	Cyberbullying Normal	219 11	16 224	0.94	0.94	0.94

Overall, ResNet50 and VGG19 demonstrated a slight analytical advantage over EfficientNetB0 in terms of representational consistency and classification stability. Nevertheless, EfficientNetB0 achieved competitive results while maintaining lower computational requirements, reinforcing the suitability of similarity-based CNN analytical frameworks for the assessment of implicitly harmful visual content under controlled experimental conditions.

Comparative Visual Analysis

The experimental phase evaluated the analytical behavior of the ResNet50, VGG19, and EfficientNetB0 architectures using stylized images reserved exclusively for evaluation purposes, including samples obtained from external online sources. The assessment focused on accuracy, recall, and F1-score metrics, as well as comparative analysis aimed at examining the representational capability of each architecture for distinguishing visually implicit harmful patterns across the two target categories.

Figure 2 presents a comparative bar chart of accuracy, recall, and F1-score across the evaluated CNN architectures, providing an overall performance overview of the proposed analytical framework. ResNet50 and VGG19 demonstrate stronger representational consistency relative to EfficientNetB0, particularly in terms of accuracy and F1-score, suggesting that deeper architectures provide improved capability for modeling visually implicit and semantically complex patterns.

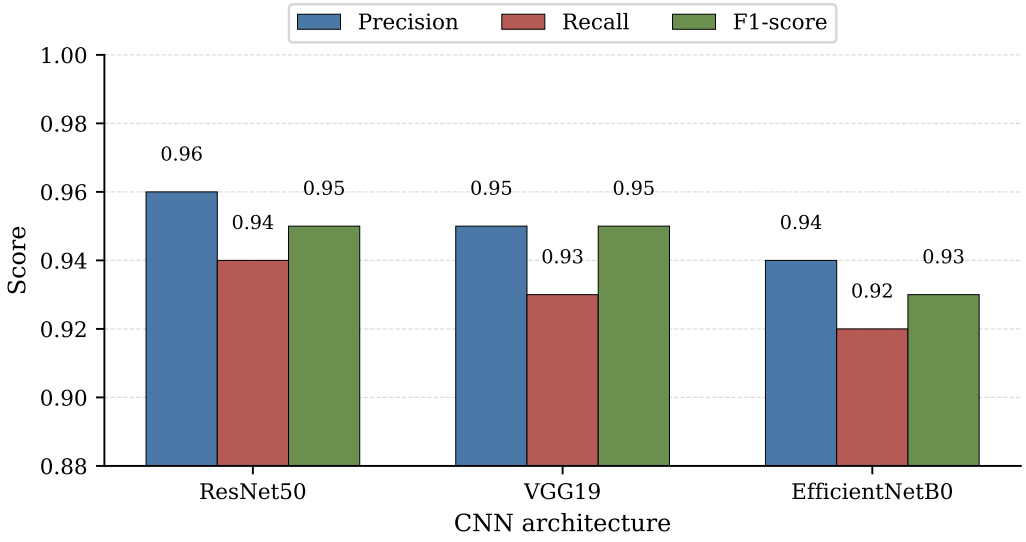


Figure 2. Comparative performance overview across CNN architectures using accuracy, recall, and F1-score metrics.

Complementarily, Figure 3 provides a heatmap-based representation of metric consistency across the evaluated architectures. This visualization highlights the stability of

ResNet50 and VGG19 across all assessed metrics while also illustrating the competitive behavior of EfficientNetB0 under reduced computational complexity. Together, both visualizations reinforce the suitability of similarity-based convolutional representations for analytical assessment tasks involving visually implicit harmful content.

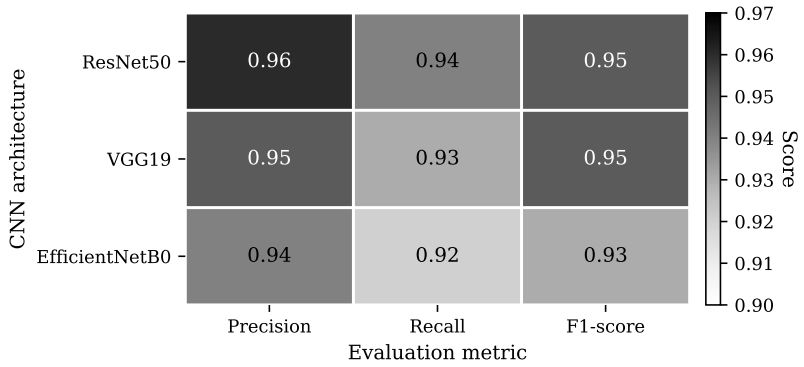


Figure 3. Heatmap representation of metric consistency across evaluated CNN architectures.

The comparative confusion matrices presented in Figure 4 further reinforce the analytical consistency of the evaluated architectures. ResNet50 exhibited the most stable discrimination behavior across the two target categories, with minimal representational ambiguity and reduced misclassification rates. Similarly, VGG19 demonstrated strong representational stability, maintaining classification behavior comparable to that of ResNet50 across visually implicit harmful patterns.

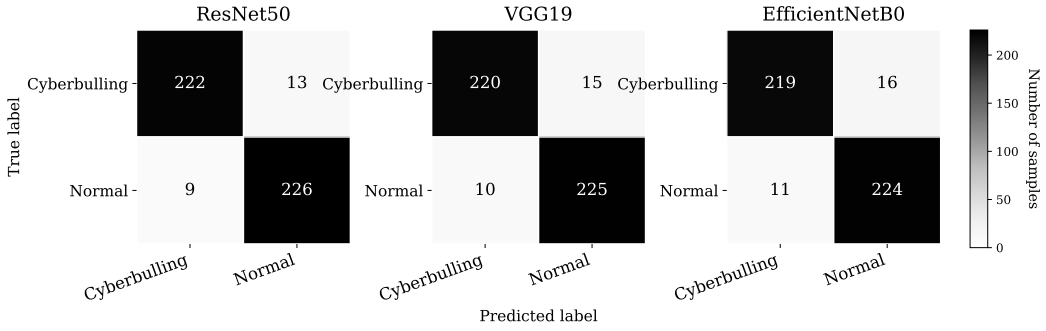


Figure 4. Comparative confusion matrices illustrating representational consistency across the evaluated CNN architectures.

Although EfficientNetB0 presented slightly lower performance levels, the confusion matrix analysis indicates that it maintained effective analytical capability when processing visually complex images. Considering its reduced computational requirements, this architecture remains suitable for analytical environments involving constrained resources or real-time processing scenarios.

Overall, the comparative analysis indicates that ResNet50 and VGG19 provide higher representational consistency than EfficientNetB0 under the evaluated experimental conditions. Nevertheless, all assessed architectures demonstrated the suitability of combining convolutional neural networks with cosine similarity for the analytical assessment of implicitly harmful visual content. These findings reinforce the viability of similarity-based deep feature representations as methodological instruments for future investigations involving visually sensitive content and AI-generated imagery.

https://doi.org/10.3390/engproc2026153006

4. Discussion

The experimental results obtained using the ResNet50, VGG19, and EfficientNetB0 architectures indicate that cosine similarity constitutes an effective analytical metric for distinguishing visually implicit harmful patterns from non-harmful visual representations. Beyond its quantitative performance, cosine similarity proved particularly suitable for scenarios involving semantically nuanced or visually complex content, as it emphasizes the orientation of deep feature representations rather than their absolute magnitude. This characteristic becomes especially relevant when analyzing stylized or indirectly encoded visual content, where harmful semantic cues may emerge through subtle representational variations rather than explicit visual elements.

Previous research supports the use of similarity-based approaches in image analysis tasks. For instance, Park and Kim [15] demonstrated that incorporating similarity measures between feature maps in convolutional neural networks can reduce computational complexity while maintaining competitive performance. Similarly, An et al. [16] employed cosine similarity loss functions in convolutional architectures for complex image classification, achieving superior performance compared with traditional cross-entropy approaches. These findings are conceptually consistent with the present study and reinforce the suitability of similarity-based representations for enhancing the analytical capability of convolutional feature embeddings, particularly when subtle or semantically implicit visual differences are involved.

Further supporting this perspective, Ahmad et al. [17] demonstrated the effectiveness of cosine similarity variants in hyperspectral image analysis, emphasizing the importance of structured similarity representations for separating visually complex categories. Although developed within a different application domain, these findings align with the methodological principles adopted in this work, suggesting that similarity-based analytical metrics provide robust representational capabilities across diverse image analysis scenarios involving high-dimensional feature spaces.

Beyond classification performance, the assessment of implicitly harmful visual content increasingly intersects with broader research domains including trustworthy artificial intelligence, multimedia security, and robust AI safety frameworks. Recent studies have emphasized that harmful content assessment should not be restricted to localized feature discrimination but should also consider system-level robustness, reliability, and secure deployment scenarios involving AI-generated multimedia content. In particular, Lin et al. [18] highlighted that multimedia systems require systematic security-oriented analyses to identify structural vulnerabilities and improve their robustness, reinforcing the importance of comprehensive evaluation frameworks beyond conventional performance metrics. From this perspective, the present analytical framework contributes toward the development of interpretable and security-oriented methodologies for evaluating semantically sensitive visual representations.

In the comparative evaluation conducted, ResNet50 and VGG19 demonstrated stronger representational consistency than EfficientNetB0, which is consistent with previous studies highlighting the importance of hierarchical and multiscale feature representations in modeling semantically complex visual patterns and supporting higher-level visual abstraction processes [19,20]. Nevertheless, EfficientNetB0 achieved competitive analytical performance while maintaining lower computational requirements, a characteristic highlighted in recent studies involving lightweight convolutional architectures for resource-constrained analytical environments [15]. This trade-off between representational depth and computational efficiency reinforces the importance of selecting architectures according to the analytical objectives, operational constraints, and interpretability requirements of the evaluated scenario.

A comparison between the evaluated CNN architectures and previous studies is presented in Table 2.

Table 2. Comparison of results between the evaluated CNN architectures and previous studies.

Study	Main Metric	Accuracy
An et al. (2020) [16]	Cosine Similarity	96%
Ahmad and Mazzara (2024) [17]	Convolutional Networks	92%
Our study (ResNet50)	CNN + Cosine Similarity	95%
Our study (VGG19)	CNN + Cosine Similarity	95%
Our study (EfficientNetB0)	CNN + Cosine Similarity	94%

Despite the promising results obtained, certain limitations should be acknowledged. The experimental dataset consisted of 470 balanced stylized images collected under controlled conditions, which may limit the direct generalization of the findings to large-scale social media environments characterized by substantially higher visual diversity and continuously evolving harmful representations. Consequently, the representational behavior observed within the cosine similarity embedding space should be interpreted as an exploratory analytical result rather than as definitive evidence of universal generalization. Future studies should validate the proposed framework using larger, heterogeneous, and dynamically evolving datasets to further assess its robustness and external validity. Overall, the findings confirm that combining convolutional neural networks with cosine similarity provides a robust analytical framework for assessing visually implicit harmful content under controlled experimental conditions. Rather than functioning exclusively as a categorical classification mechanism, the proposed approach can be interpreted as a methodological instrument for characterizing semantically sensitive visual representations through deep feature embeddings and similarity-based analysis. In this sense, the framework establishes a foundational methodological basis for subsequent research phases focused on evaluating AI-generated imagery and text-to-image (T2I) systems, where the assessment of visual offensiveness, semantic ambiguity, and policy-relevant risk becomes increasingly relevant. Future work may explore hybrid architectures that balance representational depth and computational efficiency, as well as extend the proposed analytical framework toward heterogeneous image domains and multimodal generative environments.

5. Conclusions

This study presented an experimental analytical framework for the assessment of implicitly harmful visual content using three convolutional neural network architectures, namely ResNet50, VGG19, and EfficientNetB0, in combination with cosine similarity as a similarity-based analytical metric. The primary objective was to evaluate the suitability of convolutional feature representations for distinguishing visually implicit harmful patterns from non-harmful visual content under controlled experimental conditions. The experimental results demonstrated that the proposed framework achieved consistent analytical performance in terms of accuracy, recall, and F1-score, reaching values of up to 95% for the deeper architectures.

While ResNet50 and VGG19 demonstrated stronger representational consistency for modeling semantically complex visual patterns, EfficientNetB0 achieved competitive results with slightly lower performance (94%) while offering important advantages in computational efficiency. This trade-off between representational depth and computational cost becomes particularly relevant in analytical environments involving constrained computational resources or scalable visual assessment scenarios. Furthermore, the use of cosine similarity enabled robust similarity-based discrimination by emphasizing the orientation of

deep feature vectors and facilitating the identification of shared visual patterns within the embedding space.

The findings obtained are consistent with previous studies, such as those reported by Park and Kim [15], who demonstrated that similarity-based representations can improve both efficiency and analytical performance in convolutional architectures. Likewise, the results align with those of An et al. [16], who showed that cosine similarity approaches outperform traditional cross-entropy strategies in complex image analysis tasks. Together, these works reinforce the methodological validity of combining convolutional neural networks with similarity-based analytical metrics for studying visually implicit representations.

Beyond its immediate experimental performance, the contribution of this study lies in validating CNN-based analytical frameworks combined with similarity metrics as methodological instruments for characterizing semantically sensitive and potentially harmful visual content. In this sense, the proposed approach should not be interpreted solely as a final categorical classification mechanism but rather as a foundational methodological contribution that supports subsequent analytical research phases. In particular, the framework establishes a reproducible basis for future investigations focused on evaluating images generated from textual prompts, where assessing visual offensiveness and semantic ambiguity becomes increasingly important.

Future Work

The findings of this study open several research directions, particularly toward the systematic analysis of visually generated content within sensitive digital environments. In particular, the present work constitutes a foundational methodological phase for a subsequent research stage focused on evaluating images generated from textual prompts, where the primary objective is to assess degrees of visual offensiveness, semantic ambiguity, and policy-relevant risk embedded within generated visual content. This future phase will extend the validated analytical framework presented in this paper from curated stylized image datasets toward heterogeneous image domains produced by text-to-image (T2I) generation systems.

One of the principal research directions involves investigating how convolutional neural networks combined with similarity-based analytical metrics can be employed as methodological instruments for characterizing AI-generated imagery under controlled and reproducible experimental conditions. This includes analyzing variations associated with prompt formulations, generation seeds, model configurations, and representational consistency across generated outputs, as well as studying the emergence of visually implicit harmful or borderline content patterns within multimodal generative environments. In this context, extending the framework toward multimodal analysis through the incorporation of textual information associated with image generation represents a natural progression for improving the interpretability and analytical scope of the proposed methodology.

Additional research directions include exploring hybrid architectures that balance the representational capacity of deep convolutional models such as *ResNet50* with the computational efficiency of lightweight architectures such as *EfficientNetB0*. Likewise, future investigations may incorporate adaptive or continual learning mechanisms capable of responding to evolving visual patterns and emerging forms of harmful content in dynamic digital environments. Addressing issues related to dataset bias, fairness, and representational imbalance also remains an important research direction, particularly in the context of generative systems that may amplify latent biases contained within training datasets.

Finally, the proposed framework may be extended through simulation-based analytical evaluations and controlled experimental scenarios designed to approximate real-world content moderation conditions without requiring direct deployment in production systems.

https://doi.org/10.3390/engproc2026153006

Complementary research avenues include the use of generative models for controlled data augmentation, reinforcement learning strategies to support analytical decision-making processes, and the exploration of alternative similarity or distance metrics for characterizing semantically sensitive visual representations. Overall, these future directions position the present study as a foundational methodological contribution within a broader doctoral research agenda focused on understanding and evaluating analytical and generative systems in the context of visually implicit harmful content assessment.

Author Contributions: Conceptualization, S.R.V. and J.R.P.O.; methodology, S.R.V.; software, S.R.V.; validation, S.R.V. and J.R.P.O.; formal analysis, S.R.V.; investigation, S.R.V.; data curation, S.R.V.; writing—original draft preparation, S.R.V.; writing—review and editing, S.R.V. and J.R.P.O.; visualization, S.R.V.; supervision, J.R.P.O. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The data used in this study are available from the corresponding author upon reasonable request.

Acknowledgments: The authors would like to thank the Universidad Politécnica Salesiana and the Universidad Estatal Península de Santa Elena for their academic support during the development of this research. During the preparation of this manuscript, AI-assisted tools were used for language refinement and editorial support. The authors reviewed and edited all generated content and take full responsibility for the final manuscript.

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Wang, J.C.; Hong, Y.J.; Sanjaya, E.; Santoso, H.B. Engaging or Silent? Social Media Commenting on Controversial versus Uncontroversial Issues. *Pac. Asia J. Assoc. Inf. Syst.* **2025**, *17*, 135–162.
2. Ismail, O.F. Born-Digital Memes as Archival Discourse: A Linked-Data Analysis of Cultural Sentiment and Polarization. *J. Media* **2025**, *6*, 28. [CrossRef]
3. Fernando, Z.J.; Furqon, A.F. Semiotics and Cyber Victimology in Interpreting the Signs of Digital Victimization. *Int. J. Semiot. Law* **2025**. [CrossRef]
4. Kumari, P.; Singh, R. Multimodal Cyberbullying Meme Detection from Social Media Using Deep Learning Approach. *Int. J. Comput. Sci. Inf. Technol. (IJCSIT)* **2023**, *15*, 31–45. [CrossRef]
5. Almomani, A.; Nahar, K.; Alauthman, M.; Al-Betar, M.A.; Yaseen, Q.; Gupta, B.B. Image Cyberbullying Detection and Recognition Using Transfer Deep Machine Learning. *Int. J. Cogn. Comput. Eng.* **2024**, *5*, 14–26. [CrossRef]
6. Jha, P.; Maity, K.; Jain, R.; Verma, A.; Saha, S.; Bhattacharyya, P. Meme-ingful Analysis: Enhanced Understanding of Cyberbullying in Memes Through Multimodal Explanations. In Proceedings of the 18th Conference of the European Chapter of the Association for Computational Linguistics (Volume 1: Long Papers), St. Julian's, Malta, 17–22 March 2024; pp. 930–943. [CrossRef]
7. Tan, M.; Le, Q. EfficientNet: Rethinking model scaling for convolutional neural networks. *arXiv* **2019**, arXiv:1905.11946.
8. DataStax. What Are Vector Databases? Available online: https://docs.datastax.com/en/astra-db-serverless/get-started/vector-concepts.html (accessed on 1 August 2026).
9. Steck, H.; Ekanadham, C.; Kallus, N. Is Cosine-Similarity of Embeddings Really About Similarity? In Proceedings of the ACM Web Conference 2024, Singapore, 13–17 May 2024. [CrossRef]
10. Singh, K.; Mishra, M.; Singh, E.S. Content-based Recommender System Using Cosine Similarity. *Int. J. Res. Appl. Sci. Eng. Technol.* **2024**, *12*, 2541–2548. [CrossRef]
11. Levy, A.; Shalom, B.R.; Chalamish, M. A Guide to Similarity Measures and Their Data Science Applications. *J. Big Data* **2025**, *12*, 188. [CrossRef]
12. Thapliyal, S.; Kumar, N. Fusion of Heuristics and Cosine Similarity Measures: Introducing HCSTA for Image Segmentation via Multilevel Thresholding. *Int. J. Syst. Assur. Eng. Manag.* **2025**. [CrossRef]

13. Elmore, K.L.; Richman, M.B. Euclidean Distance as a Similarity Metric for Principal Component Analysis. *Mon. Weather Rev.* **2001**, *129*, 540–549. [CrossRef]
14. Pericherla, S.; Ilavarasan, E. Cyberbullying Detection and Classification on Social Media Images Using Convolution Neural Networks and CB-YOLO Model. *Evol. Syst.* **2025**, *16*, 43. [CrossRef]
15. Park, K.; Kim, D.H. Accelerating Image Classification using Feature Map Similarity in Convolutional Neural Networks. *Appl. Sci.* **2019**, *9*, 108. [CrossRef]
16. An, S.; Lee, M. An ensemble of simple convolutional neural network models for MNIST digit recognition. *arXiv* **2020**, arXiv:2008.10400.
17. Ahmad, M.; Mazzara, M. SCSNet: Sharpened Cosine Similarity-Based Neural Network for Hyperspectral Image Classification. *IEEE Geosci. Remote Sens. Lett.* **2024**, *21*, 5503204. [CrossRef]
18. Lin, Y.; Wei, Y.; Chen, D.; Li, Y.; Erkan, U.; Toktas, A.; Gao, S.; Zhang, Y. Cryptanalysis and Improvement of a Video Cryptosystem via Chaos and S-Box. *Acm Trans. Multimed. Comput. Commun. Appl.* **2026**, *22*, 173. [CrossRef]
19. Batisteli, J.P.O.; Guimaraes, S.J.F.; Patrocinio, Z.K.G., Jr. Hierarchical Multiscale Representation in Remote Sensing Scene Classification. *IEEE Geosci. Remote Sens. Lett.* **2025**, *22*, 2504505. [CrossRef]
20. Muratore, P.; Alemi, A.; Zoccolan, D. Unraveling the Complexity of Rat Object Vision Requires a Full Convolutional Network and Beyond. *Patterns* **2025**, *6*, 101149. [CrossRef] [PubMed]

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.