Hindawi
Applied Computational Intelligence and Soft Computing
Volume 2018, Article ID 1463546, 13 pages
https://doi.org/10.1155/2018/1463546

Research Article

Explicit Content Detection System: An Approach towards a Safe and Ethical Environment

Ali Qamar Bhatti,[1] **Muhammad Umer**,[1] **Syed Hasan Adil**,[1] **Mansoor Ebrahim**,[2] **Daniyal Nawaz**,[1] **and Faizan Ahmed**[1]

[1]*Iqra University, Pakistan*
[2]*Sunway University, Malaysia*

Correspondence should be addressed to Syed Hasan Adil; hasan.adil@iqra.edu.pk

Received 7 February 2018; Revised 26 May 2018; Accepted 5 June 2018; Published 4 July 2018

Academic Editor: Miin-Shen Yang

An explicit content detection (ECD) system to detect Not Suitable For Work (NSFW) media (i.e., image/ video) content is proposed. The proposed ECD system is based on residual network (i.e., deep learning model) which returns a probability to indicate the explicitness in media content. The value is further compared with a defined threshold to decide whether the content is explicit or nonexplicit. The proposed system not only differentiates between explicit/nonexplicit contents but also indicates the degree of explicitness in any media content, i.e., high, medium, or low. In addition, the system also identifies the media files with tampered extension and label them as suspicious. The experimental result shows that the proposed model provides an accuracy of ~ 95% when tested on our image and video datasets.

1. Introduction

With the advent of modern technology, information and its accessibility on the Internet has dramatically increased. In addition, people are getting easier access to general information plus the adult information (adult images, videos, and animation) especially youths which is an alarming sign. To prevent youths from accessing such adult contents is one of the major challenges of the modern society. One solution to this is to develop a mechanism that can detect and filter the adult content from the data volume. However, accurately identifying the adult media content from a bundle of information is an important constraint that needs to be considered. The adult media content can be categorized as exposed body contents, detailed erogenous parts contents, and pornographic action [1].

Filtering media with adult contents is imperative to avoid offensive content over the Internet. In literature, different applications to restrict the accessibility of such adult contents on computer exist such as blocking unwanted sites (CyberPatrol, ContentProtect, NetNanny, Family.net, and K9 Web Protection [2]) or identifying explicit content media (SurfRecon, Porn Stick Detection [3]). In addition, many researchers have focused their research on developing explicit content detection mechanism for media contents using different techniques such as identification of skin region, skin detection, YCbCr space color, and HSV color model [1–4].

An explicit content detection (ECD) system is developed and implemented in this research work by using deep learning solution for Not Suitable/Safe for Work (NSFW) media (image, video) contents. The approach is based on image processing, skin tone detector, and pattern recognition techniques. In the first step, YCbCr color space is used to transform the image to classify various objects that are not of interest. Secondly, skin tone detection threshold of the image is calculated to filter various segments existing within the image. Finally the image explicitness probability is estimated to determine whether the image contains explicit content or not.

The key highlights of the proposed ECD system include the following: (i) being an open source system, (ii) being computationally efficient, (iii) being highly robust, and (iv) ease of deployment in multiple modes which include standalone on individual desktop, with proxy server, and dedicated server as content filtering system to detect or restrict the explicit content. The paper is organized as follows. In Section 2, research

work related to explicit content detection is discussed. The overall system architecture is presented in Section 3, followed by a detailed explanation of the proposed model described in Section 4. Section 5 presents and discusses all the simulation results. Finally, the paper is concluded in Section 6.

The main objectives of the research are highlighted as follows:

(a) Development of an explicit content detection (ECD) system

(b) Software application of the ECD system

(i) Checking the selected file for tampered extensions.

(ii) Detecting explicit images and based on degree of explicitness marking them as high, medium, and low.

(iii) Detecting explicit video and based on degree of explicitness marking them as high, medium, and low.

2. Related Work

Many previous works have already focused on developing explicit content detection mechanism for media contents based on different techniques such as identification of skin region, skin detection, YCbCr space color, and HSV color model. The details about these works can be found in [1–10].

An explicit content detection algorithm, which makes use of Support Vector Machine (SVM) for classification, is proposed in [1]. The SVM approach is built on statistics learning theory basis that helps to predict, analyze, tune, and identify explicit content in an image. The proposed model is a three-step process that includes (i) skin filtering (skin is one of the most important features for detecting explicitness in an image, (ii) attributes (skin percentage, pornography-weight, skin area geometric distribution, skin pixels in skin correlation, hair-inside-body, and skin-region-smoothness), and (iii) SVM-prediction (training of SVM model using the discussed six attributes). However, the proposed work has very low prediction accuracy, i.e., the model predicts images in which people wear short clothes (bikinis, shorts) as explicit; that is not true. In addition, the work is not valid for videos.

The work done by [2] employs skin region identification approach. The proposed work adopts a combined approach in which HSV color model is grouped with YIQ and YUV models. The proposed approach initially uses the white balance algorithm to achieve better skin area. Next, the texture model based on grey level comatrix and geometrical human structure is applied to lower the disturbance of background area that is similar to skin area. Finally, the SVM model is used for the transformed image to successfully classify the image as explicit and nonexplicit.

In [3], a survey of various skin modelling and classification approaches using color information in the visual range is conducted. The review focuses on color spaces used for skin modelling and detection and the use of skin-color constancy and dynamic adaptation techniques adopted by various approaches to improve the skin detection performance in ambiguous environmental conditions. In addition, the paper also indicates the various factors under which the skin detection techniques perform well.

In [4], an explicit content detection algorithm, which makes use of skin region detection, is proposed. The proposed work is based on HSV color model rather than RGB color model to detect the skin in the images as RGB color model is exposed to some lighting issues. The HSV model not only improves the lighting issues but also the visibility of the skin tone of the images. The proposed system formulates only the skin pixels as output and based on the skin pixel values the explicitness in the image is detected. The images having larger values of skin tone pixels are referred to as explicit. Further in [5], the author extended the approach proposed in [4] and evaluates its performance on complex dataset.

In contrast to the approach proposed in [5], an explicit content detection model that makes use of YCbCr space color is proposed in [6]. The key objective of the proposed model is to apply the model for forensic analysis or pornographic images detection on storage devices such as hard disk, USB memories, etc. The proposed model estimates the color pixels percentages that are within the images that are susceptible to be a tone skin. In other words, ratio and proportion of the skin content in a given image with respect to the total image is calculated (#skin color pixels / #image pixels in total). Once the skin percentage is calculated it is compared with a certain defined threshold to labeling images as explicit or nonexplicit. If the value is greater than the defined threshold then it is classified as explicit and else nonexplicit. The proposed approach is only applicable for identifying and classifying explicitness/nonexplicitness in images (i.e., not applicable for videos).

A new approach for detecting pornographic images is introduced in [7]. The proposed approach suggests two new features, i.e., Fourier descriptors and signature of boundary of skin region. These two features in combination with other simple traditional features provide decent difference between explicit and nonexplicit images. Moreover, a fuzzy integral based information fusion is applied to combine MLP (Multi-Layer Perceptron) and NF (Neuro-Fuzzy) outputs. The proposed method attained precision was 93% in TP and 8% in FP on training dataset, and 87% and 5.5% on test dataset.

In [8], deeper neural network is implemented to detect explicit content. Deeper neural networks are usually difficult to train. The proposed work makes use of residual based learning to ease the training process and is substantially deeper than the conventional ones. The identification and detection accuracy is improved by increasing the depth. In addition, Yahoo had also developed a model referred to as Open_NSFW. The Yahoo model uses CaffeOnSpark (framework for distributed learning) for training models for experiment. We are using this model into our system for identifying whether the input image is explicit or not.

A pornographic image recognition using skin probability and principle component analysis (PCA) on YCbCr color space is proposed in [9, 10]. The paper aims to optimize the accuracy and false rejection rate of the skin probability and

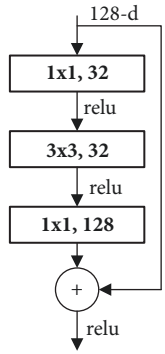

Figure 1: Basic building block of Resnet-50 1 by 2 architecture.

fusion descriptor based recognition system using PCA. The proposed method experimentally proves that it can increase the accuracy by about 4.0% and decreases the FPR to 20.6% of those of pornographic recognition using fusion descriptors when tested on large size dataset. The proposed method also works fast for recognition, which requires 0.12 seconds per image.

The above discussed explicit content detection models are exposed to few issues such as improper detection of skin tone, lightning issue, and inaccurate prediction, and most of the models are limited to image content only. Moreover, most of the research works are just limited to modelling and simulation level, no proper implementation at application level is available.

In particular, our proposed model is based on the residual network as adapted by Yahoo to aid the detection of explicitness in image and video. The key difference is that the proposed explicit content detection (ECD) system is not only limited to modelling and simulation level, but also implemented at application level. The proposed system can detect any kind of NSFW media contents on the storage device. Furthermore, we retrained the model of Yahoo NSFW, which is based on CaffeOnSpark framework that uses the thin Resnet-50 1 by 2 (based on Resnet-50) [11] architecture as shown in Figure 1 as the pretrained network, to fine-tune the weights of the above model based on our own created dataset of images and videos.

3. Overview of the Proposed Explicit Content Detection (ECD) System

From the earlier discussions, it is essential to develop explicit content detection (ECD) system that not only classifies the contents as explicit, nonexplicit, and suspicious but should also need to be scalable, fast, and accurate.

In our proposed ECD system as shown in Figure 2, we take a file or directory (i.e., multiple files) as an input and check for the content types' (i.e., image/video/non-image file). Each identified image/video file will be passed to ECD-CNN model for the classification of content as explicit, nonexplicit based on predefined threshold probability.

(i) In case of image file the system simply forwards the file to ECD-CNN model for content classification.

(ii) In case of video file the system simply extracts the frames from the video and treats each extracted frame as an image. Analyzing each frame from a video can be tedious and troublesome. Mostly there are at least 15 fps (i.e., frames per second) in each video (i.e., a 5-minute video would have 4500 frames). To encounter such issue the video frames are randomly shuffled and optimal stopping criteria are applied on them that will help in boosting the software system. Later, it will be forwarded to ECD-CNN model for content classification.

(iii) If the content is an image or a video but the extension is tampered, then the system simply labels it as suspicious. In case of non-image file the system simply ignores it and continues with further processing.

An object of Tika class invokes detect (File file) method which returns a string containing the file type (i.e., assigned file extension) and file content type (i.e., extracted from the header of the file) which represents the original file type separated by "/". If both parts of the string represent same type then our application marked it as normal file (i.e., non-image file), whereas if the first part is a non-image file type and second part represents an image file, then our application marks the file as suspicious.

In the following subsections, we will discuss different phases of the proposed ECD system in case of image and video files.

3.1. Phase 1: Searching and Content Type Detection. In this phase, the selected directory is forwarded for scanning of media files. During scanning, the content type (image/video/non-image) of the file is checked. In case of non-image file, if the header of the file corresponds to any legitimate image format but the file extension is tempered, then the proposed system simply labels it as suspicious (i.e., we have incorporated Apache Tika API to validate the content of the file with the assigned extension); else it ignores the file (i.e., it is not a valid image/video file; therefore no further action is required) as shown in Figure 2.

3.2. Phase 2: Classification into Explicit and Nonexplicit Content. Once the content type is identified as either image or video, the contents are passed through the next phase i.e., explicit content detection module as shown in Figure 2. This phase works differently for each content type (image, video) and can be categorized into image explicit content detection module for image and video explicit content detection module for video, explained as follows.

3.2.1. Explicit Content Detection. The contents of the scanned images are first encoded into BASE64 string that are then one by one transmitted to the web server. On the server side, the received BASE64 encoded image string is decoded into image file and passed to our ECD-CNN model (i.e., our ECD-CNN model is developed using CaffeNet deep learning framework [8]) for classification into explicit/nonexplicit as shown in Figure 3. The proposed ECD-CNN is a Resnet-50 1 by 2 convolutional deep learning neural network which

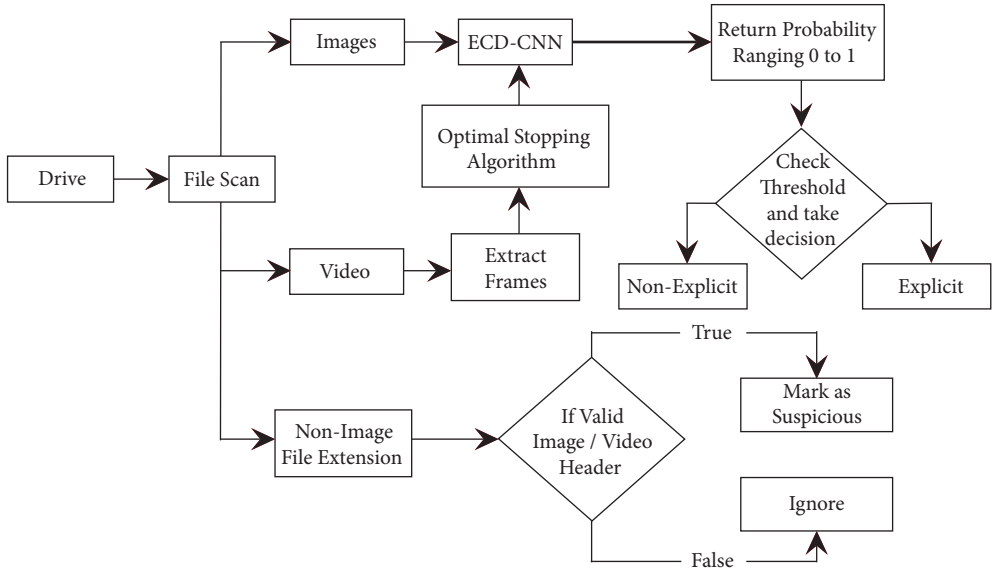

Figure 2: Block diagram of the proposed ECD system.

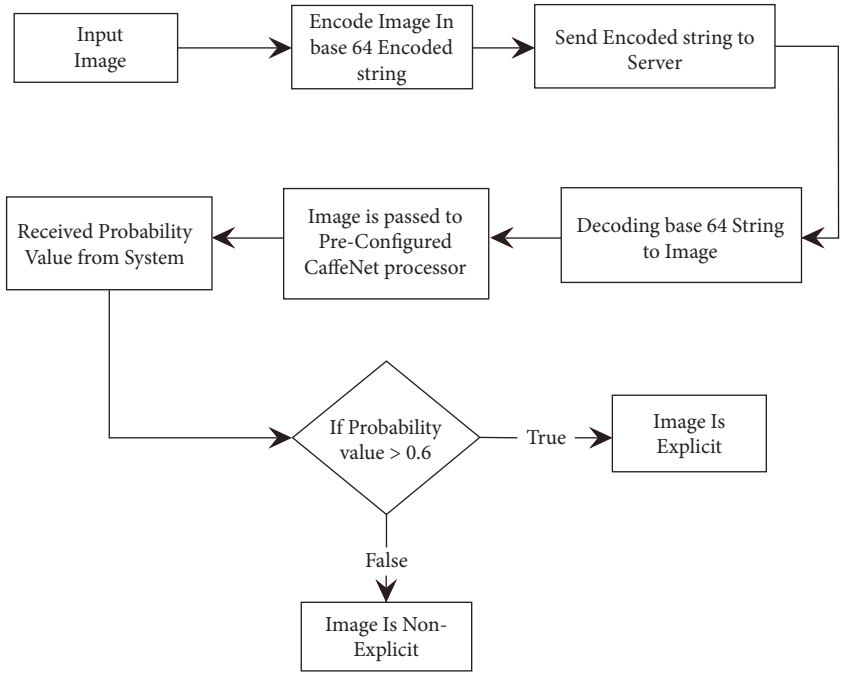

Figure 3: Sending images to system.

is the extended form of actual Resnet-50 architecture [8]. CaffeNet is a powerful framework that allows using existing models proposed by other researchers, extending any existing models, or building a new model from scratch. We can extend or create a new model by defining the desired number and types of layers in CaffeNet configuration file (i.e., CaffeNet configuration file must have .prototxt file extension).

Figure 4 defines the details about proposed Resnet-50 1 by 2 architecture built using CaffeNet. The proposed ECD-CNN (i.e., Resnet-50 1 by 2) consists of 50 layers divided into 5 convolutions in total, with each layer having stride size of 2. Relu function is applied to each layer to gain output for the said layer, which becomes the input for the next layer and to the layer after that. Each color differentiates one convolutional layer from others. After 1st convolutional layer maximum pooling is done and output is forwarded as input to the next convolution layer. For the remaining convolutional layers (i.e., from 2nd to 5th convolutional layer), output of the current convolutional layer will become input for the next convolutional layer. The dotted line after each convolution is called shortcut which is added to prevent the vanishing gradient problem. The proposed Resnet-50 1

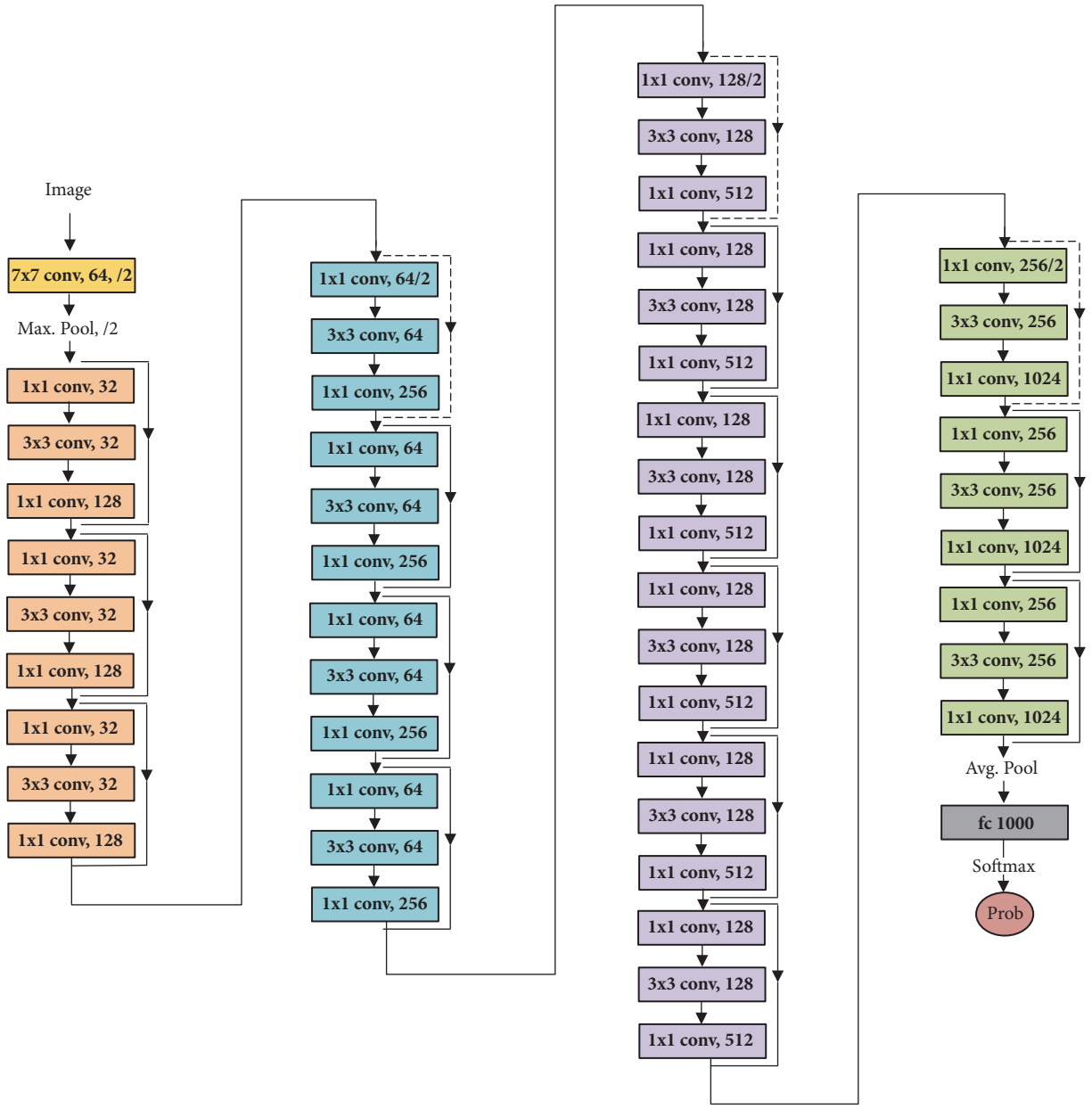

Figure 4: Block diagram of Resnet-50 1 by 2 architecture.

by 2 architecture uses half (i.e., 1/2) number of inputs as compared with the existing Resnet-50 architecture. In the proposed ECD-CNN architecture (i.e., Resnet-50 1 by 2) the number of inputs is reduced to achieve better computational complexity without compromising the classification accuracy of the model (i.e., we have proved the accuracy of the model using our prediction accuracy in the result section).

2nd convolution has 3 Residual Networks each having three layers. 3rd layer has 4 Residual Networks each having three layers. 4th layer has 6 Residual Networks each having three layers. 5th layer has 3 Residual Networks each having three layers. After 5th convolutional layer average pooling is done and a softmax function is applied in the output layer to gain a probability ranging from 0 to 1.

In fourth step, we train and optimize model and get the trained model in a file with extension .caffemodel. After training now we can use the trained model for the prediction. The database used for the testing of the proposed model can be found at [12].

Our residual network model after processing the image returns a probability value ranging between 0 and 1 which is the NSFW score of the image. The score is then compared with the threshold value that is set as 0.6 (nonexplicit $< 0.6 \geq$ explicit) to define the image content as explicit or nonexplicit. The threshold value 0.6 is selected by extensively running hit and trial method. A dataset of 2000 explicit/nonexplicit images is used to train the system to find the appropriate threshold. The system is tested at different threshold value

```
FUNCTION ImageModule(ImageFilePath)
IF extension is NOT a legitimate image extension
        //label the file as SUSPICIOUS
ELSE
        Image.Encode(); /*Encode it to base64 String
        Base64 String passed to the web service.
        The Web Service would return the probability of the
        Image i.e either file is NSFW or SFW. */
        IF returnedProbabiltyValue>= 0.6 AND < 0.7 THEN
                model.status = LOW; /* Image contains low Level Explicit Content */
        ELSE IF returnedProbabiltyValue>= 0.7 AND < 0.8 THEN
                model.status = MEDIUM; /* Image Contains Medium Level Explicit Content */
        ELSE IF returnedProbabiltyValue>= 0.8 AND < 1 THEN
                model.status = HIGH; /* Image Contains High Level Explicit Content */
        ELSE
        model.status = Non_Explicit; /* Image Doesn't contain any Explicit Content */
ENDFUNCTION
```

Pseudocode 1: Image content based system.

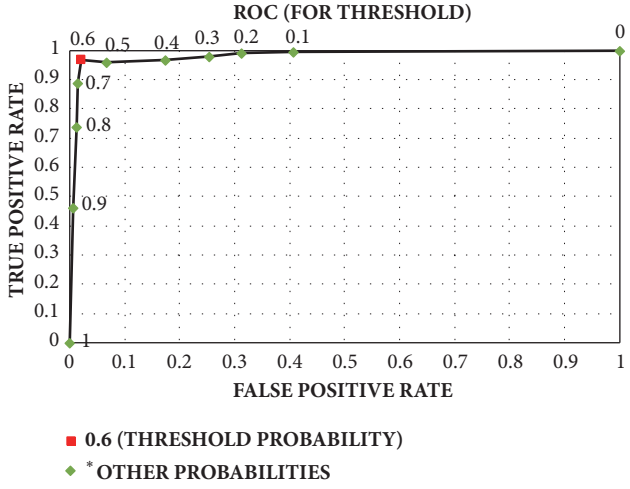

Figure 5: ROC for (explicit/nonexplicit) images predictions.

Table 1: Threshold level criteria.

Level	Threshold
Low	Probability value >= 0.6 < 0.7
Medium	Probability value >= 0.7 < 0.8
High	Probability value >= 0.8 < 1

from 0.1 to 1 with an interval of 0.1 and a Receiver Operating Characteristic (ROC) Curve as shown in Figure 5 is obtained that shows the preeminent threshold value to differentiate between explicit and nonexplicit images.

In addition, if the image is marked as explicit the image content-based system also checks the level of explicitness in the image content, i.e., low, medium, or high based on the threshold criteria defined in Table 1.

The overall process of the image content-based system is defined in Pseudocode 1.

3.2.2. Video Content-Based System. In case of video contents, the explicit video content module as shown in Figure 6 is selected. It initially extracts all the frames (considered as images) using "*javacv.FfmpegFrameGrabber*" from the video that are randomly shuffled and an optimal stopping criteria is applied on them that will improve the efficiency of the software system as extracting and processing of video requires a lot of computing.

The optimal stopping criteria algorithm first selects $\sqrt{\text{Frames\_Count}}$ frames from the video that are encoded (BASE 64) and then decodes into image file at the server side for NSFW probabilities determination. At the server side, the decoded image file is passed through CaffeNet model that finds the maximum probability among all first $\sqrt{\text{Frames\_Count}}$ frames and uses this probability to continue further analysis. The ECD-CNN model takes image (i.e., a video frame) as input and returns the probability of the associated explicitness in the given image.

If the maximum probability of first $\sqrt{\textbf{Frames\_Count}}$ is below 0.6 then it stops further processing and declares the current video as nonexplicit (i.e., this becomes the best case in terms of computational time because we have scanned only $\sqrt{\textbf{Frames\_Count}}$ frames). However, if the probability is 0.6 or more then this probability will be used as a threshold probability for the remaining frames in the video. It further checks probability of each frame one by one with threshold probability. If any frame exceeds threshold probability (i.e., set by the $\sqrt{\textbf{Frames\_Count}}$ frames) then it immediately stops further execution and declares current video as explicit (i.e., this becomes average case in terms of computational time because on average we have scanned only **Frames_Count/2** frames). Then, if none of the remaining frames exceeds threshold probability then current video is declared as nonexplicit (i.e., this becomes worst case in terms of computational

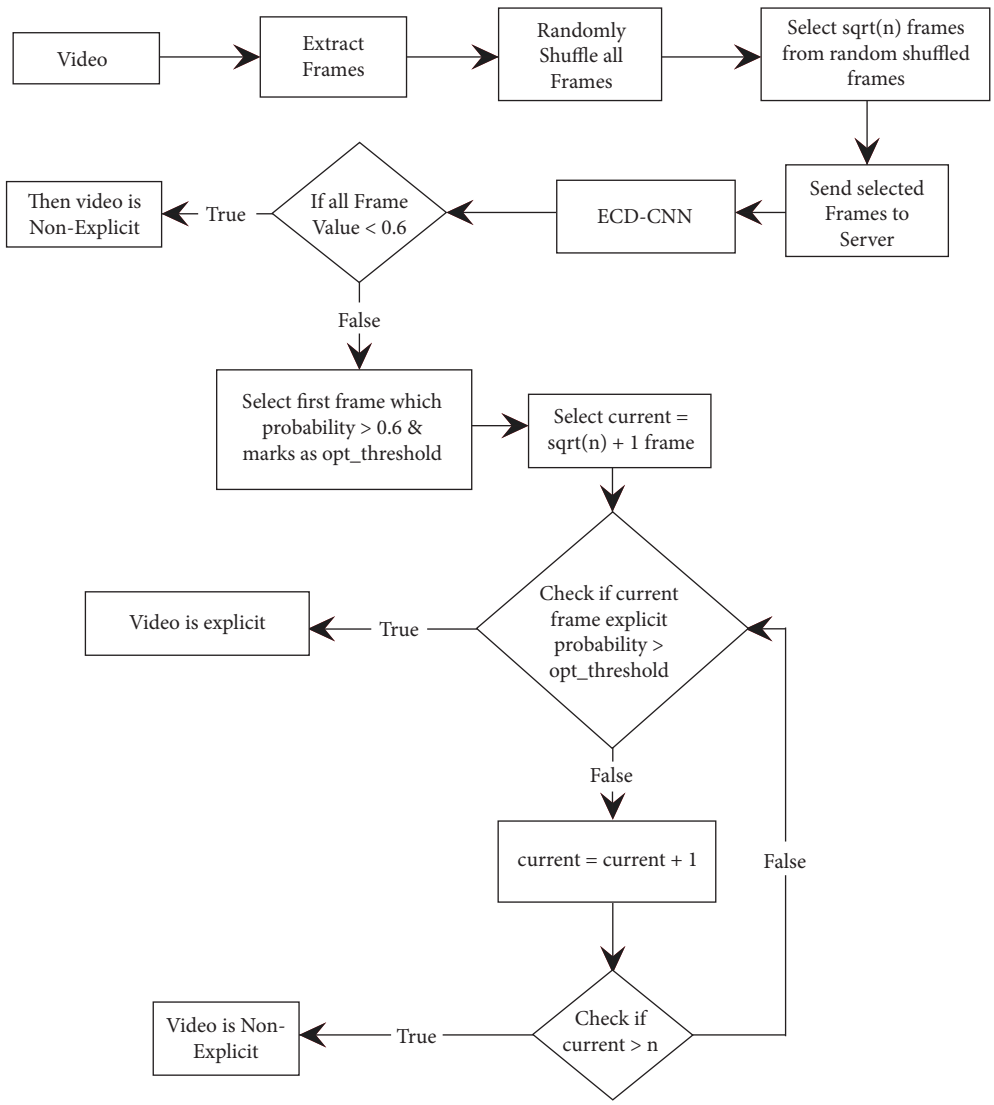

Figure 6: Checking video content.

time because we scan all individual frames of the video that need to be scanned).

The overall process of the video content-based module is defined in Pseudocode 2.

The overall process of the ECD system is defined in Pseudocode 3.

4. Explicit Content Detector (ECD) System Software

In this section, the proposed ECD system implementation as an application software is discussed. The software front end contains a progress bar, select drive, start search button, browse location, cancel button and information area as shown in Figure 7.

4.1. Progress Bar. The progress bar is used to shows the scanned status of the folder or drive that is in search for explicit content. The progress is shown in Figure 8.

4.2. Select Drive. It allows the user to select the drive which has to be scanned for explicit content.

4.3. Start Search. Once the drive or folder is selected the start search button is used to start the scanning of selected drive or folder as shown in Figure 9.

4.4. Browse Location. Browse location selects the certain folder from the drive to scan. After the browse location button is clicked a dialogue box will appear showing list of images to be selected from the folder.

4.5. Information. Figure 10 shows the scanning progress of each media file with location, file name, and type (i.e., either explicit or nonexplicit) of selected folder/drive being scanned by the ECD system.

4.6. Cancel. During the process of scanning if the cancel button is clicked, the scanning will be stopped and the output result till the time of cancellation will be displayed.

```
FUNCTION VideoModule()
IF model is detected as a VIDEO
        IF extension is NOT a legitimate video extension THEN
                Label the file as SUSPICIOUS
        ELSE
                numberOfFrames=Get NumberOfFrames(Video_Path)
              /*Do Sampling taking square root of total number of frames*/
                n = doRandomShuffle(numberOfFrames)
                /*Select rFrames = (√n) frames to the server and gets their Probability value ranging from 0-1*/
                ArrayList<double>rFramesVals new ArrayList<>();
/* To store tempVar no of Probability Values */
FOREACH f in rFrames
rFramesVals.add(getExplicitProbability(f))
                /*Now, Take probability of each frame in rFrames and store it in rFramesVals List */
                IF all the values in rFramesVals< Threshold THEN
                        model.isExplicit = False;
                ELSE
                        /*Take the value which is just higher than the threshold from the array rFramesValsand
store it in variable opt_threshold */
          /*Now pick the remaining frames singularly and compare the value with opt_thresholduntil all frames
checked or any frame greater than opt_thresholdis found */
                IF Any Frame Greater than opt_thresholdis Found THEN
                        model.isExplicit = true;
                ELSE
                        model.isExplicit = false;
ENDFUNCTION
```

Pseudocode 2: Checking video content function.

```
Input: a file or a directory
Output: report identifying the SFW, NSFW and Suspicious Content.
FUNCTION ECD
WHILE Directory NOT Fully Traversed
      ArrayList<Model> data = getFiles(Paths.get(currentFilePath)); /*Getting All  files in the selected directory*/
End WHILE
FOREACH model in data
     get the data type of each model by using ApacheTika.detect(CurrentFilePath)
End FOREACH
IF CurrentFileisMediaTHEN
        IF CurrentFileisImage THEN
        ImageModule();
        ELSE IF CurrentFileisVideo THEN
              VideoModule();
ELSE
         //IGNORE and Go to next
GenrateReport();
ENDFUNCTION
```

Pseudocode 3: ECD system.

4.7. Output Result. This box shows the output result of scanned files and drives.

5. Experimental Results

In the following, the evaluation of the proposed ECD system for image and video contents is presented. A set of test media data (i.e., images and videos) is applied to evaluate its performance. In order to obtain a significant amount of testing data for the proposed approach, we have created datasets of 2000 explicit/nonexplicit images and 1000 explicit/nonexplicit videos from the Internet with the content categorized as low, medium, and high. The explicit content contains naked as well as seminaked people with various skin tones whereas nonexplicit content by its nature does not contain explicit content and includes dressed people, trees,

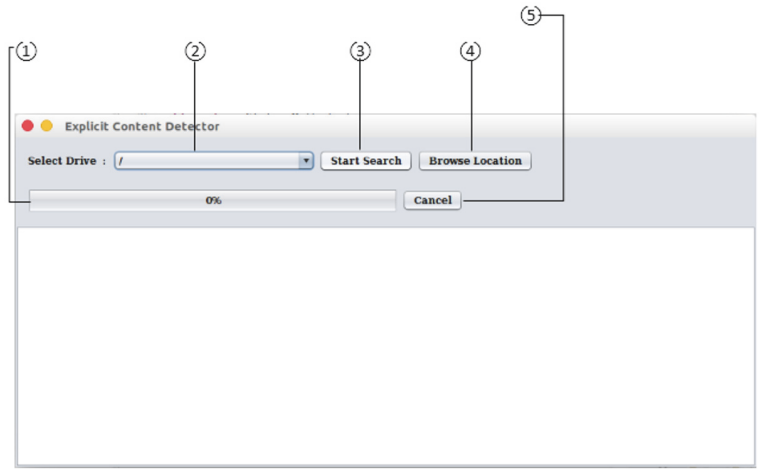

Figure 7: Front end of ECD software.

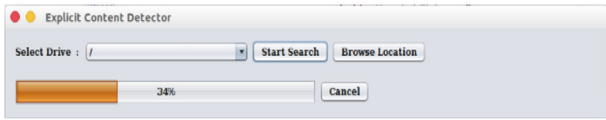

Figure 8: Searching progress of explicit content from the drive.

wildlife, flowers, automotive, cartoons, and landscapes. The images are about 640 × 480 pixels (cropped to the field of interest, still more challenging than standard ones), while the videos are of 1 MB each containing 100 frames. Each image and video was selected with different exposures, skin tones, and lighting conditions and are labeled as "highly explicit", "medium explicit", and "low explicit" for high content, medium content, and low content, respectively, to efficiently train the proposed ECD system for accuracy. In addition, an optimal stopping criterion is applied for video after running ECD on dataset to get balanced output.

The experimental setup involves the implementation of the proposed ECD system for image and video contents. The ECD system requires both efficiency and accuracy; yet, both are vice versa. In order to obtain significant amount of efficiency and accuracy ECD has to be checked on various videos and images through which ECD can be trained from which efficiency and accuracy will be obtained. In addition, classification error by using confusion matrix is also calculated to evaluate the performance of the proposed system. The classification error permits more detailed analysis of the data than accuracy, as the accuracy might lead to certain misleading results if the dataset is unbalanced (large variation in the observations of different classes).

5.1. Application of ECD System on Explicit/Nonexplicit Image Dataset. In this section, the performance of the proposed ECD system is evaluated for explicit/nonexplicit image dataset to validate its efficiency and accuracy.

The results presented in Figure 11 shows that the proposed system when applied to a dataset of 1000 explicit images not only detects the explicit images correctly but also categorizes them based on the content type as low, medium, and high (shown as blue dots). Moreover, it is also observed that few of the images were erroneously detected (shown as red dots).

Figure 12 shows the results of proposed ECD system when applied to a dataset of 1000 nonexplicit images. The proposed scheme detects the nonexplicit images accurately (shown as blue dots) with only few instances of inaccurate images' detection (shown as red dots).

In Figure 13, the complete dataset of 2000 explicit/nonexplicit images is used to evaluate the performance of the proposed system. The proposed scheme performs well and detects the explicit (blue dots) and nonexplicit images accurately (red dots) as well as highlighting the explicit content images as low, medium, and high (shown as black dotted lines). In addition, it should also be noted that only few instances of inaccurate images' detection (green dots with black circle) are also present. The inaccuracy is because the dataset contains various images with different skin tone, lightening, and exposure that might lead to some prediction errors.

Table 2 represents the classification accuracy of explicit, nonexplicit, and combined. From the table it can be observed that the proposed system provides 91.3% and 98.5% accuracy for datasets of explicit and nonexplicit images, respectively. In addition, the combined classification accuracy of both explicit and nonexplicit is 95%.

Table 3 shows the confusion matrix along with TPR, FPR, Precision, accuracy, and f1 score of the proposed ECD system on test dataset of explicit and nonexplicit images. The test dataset contains equal number of explicit and nonexplicit images (i.e., 1000 explicit and 1000 nonexplicit images). From the confusion matrix, it can be observed that the proposed approach correctly classified 913 explicit images and 981 nonexplicit images. While only 87 explicit and 19 nonexplicit images were misclassified by the proposed system.

Table 4 shows the confusion matrix along with TPR, FPR, Precision, accuracy, and f1 score of the proposed YCBCR based classification algorithm on same test dataset of explicit

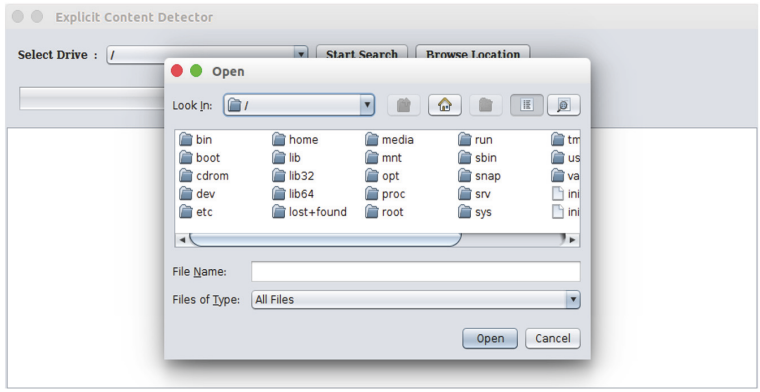

Figure 9: Select the drive/folder to scan for the explicit contents.

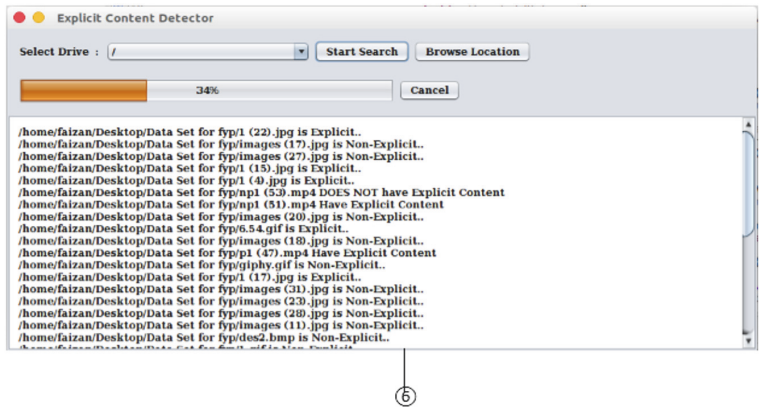

Figure 10: Scanning progress of explicit/nonexplicit content from the selected drive/folder.

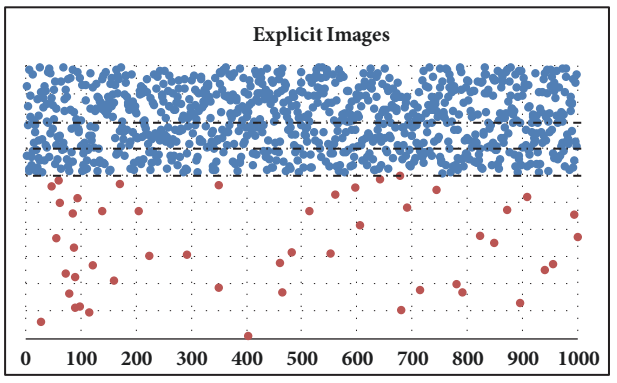

Figure 11: Result graph of application of the proposed ECD system on 1000 explicit dataset images.

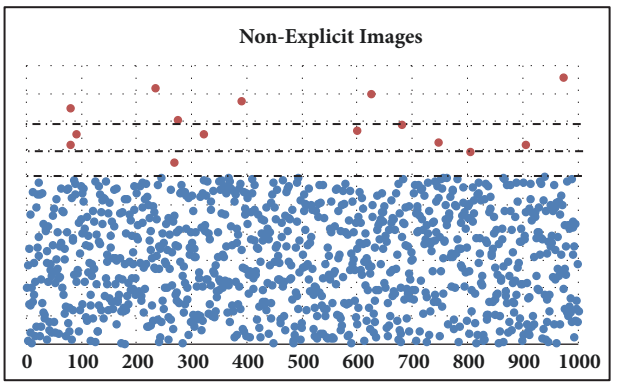

Figure 12: Result graph of application of the proposed ECD system on 1000 nonexplicit dataset images.

and nonexplicit images as used by our ECD system. From the confusion matrix, it can be observed that the YCBCR approach correctly classified 578 explicit images and 780 nonexplicit images. While, 422 explicit and 220 nonexplicit images were misclassified.

It is important to mention here that to the best of our knowledge (i.e., literature search on renowned research literature repositories) none of the other schemes have published their code and datasets due to the nature of the data. Therefore, we have implemented YCBCR algorithm and tested it on the same dataset to compare with our proposed ECD approach. The results for YCBCR as shown in Table 4 were obtained after implementing the approach mentioned in [6] and compared with results shown in Table 3 of the proposed ECD approach. The comparative analysis shows that the proposed approach provides significant improvement in

Table 2: Statistical analysis of explicit, nonexplicit, and combined image dataset.

Type	Total Images	Correct output	Wrong Output	Accuracy
Explicit	1000	913	87	91.30%
Non-Explicit	1000	981	19	98.10%
Combined	2000	1894	106	94.70%

Table 3: Confusion matrix for images dataset using proposed ECD system.

N=2000	Actual EXPLICIT	Actual NON-EXPLICIT
Predicted: EXPLICIT	913	19
Predicted: NON-EXPLICIT	87	981
True Positive Rate (Sensitivity) = 0.913, **False Positive Rate**= 0.019		
Precision= 0.9796, **Accuracy**= 0.9470, **F1 Score**= 0.9451		

Table 4: Confusion matrix for image dataset using YCBCR.

N=2000	Actual: EXPLICIT	Actual: NON-EXPLICIT
Predicted: EXPLICIT	578	220
Predicted: NON-EXPLICIT	422	780
True Positive Rate (Sensitivity) = 0.578, **False Positive Rate**= 0.2200		
Precision= 0.7243, **Accuracy**= 0.6790, **F1 Score**= 0.6429		

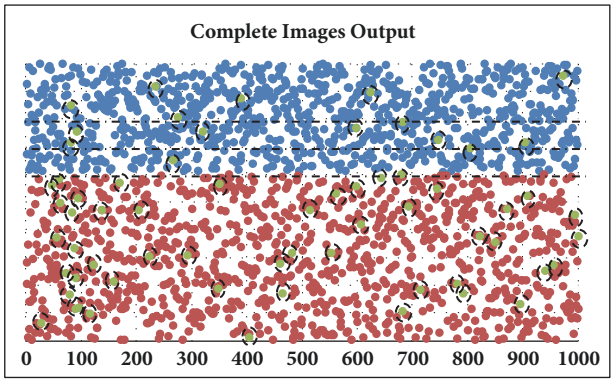

Figure 13: Result graph of application of proposed ECD system on complete (explicit/nonexplicit) dataset images.

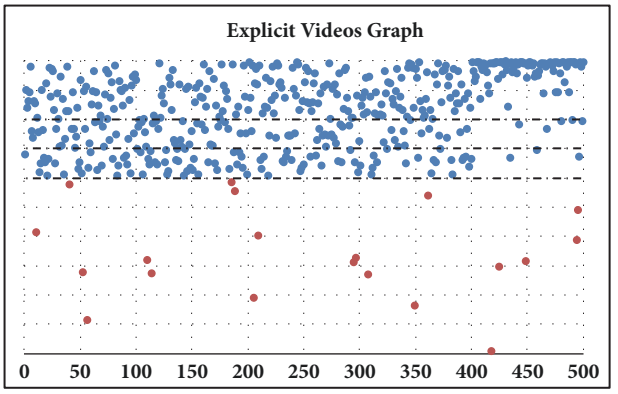

Figure 14: Result graph of application of the proposed ECD system on 500-explicit video dataset.

terms of TPR which is 0.913 versus 0.578, FPR which is 0.019 versus 0.22, Precision which is 0.9796 versus 0.7243, Accuracy which is 0.9470 versus 0.6790, and F1 score which is 0.9451 versus 0.6429 over YCBCR approach.

5.2. Application of ECD System on Explicit/Nonexplicit Video Dataset. In this section, the performance of the proposed ECD system is evaluated for explicit/nonexplicit video dataset to validate its efficiency and accuracy.

The results presented in Figure 14 show that the proposed system when applied to dataset of 500 explicit videos detects the explicit videos based on the content type as low, medium, and high (shown as blue dots) correctly. Moreover, it can also be seen that few of the videos were incorrectly detected (shown as red dots). The dataset accuracy was found to be 93% with only 7% incorrect detection.

Figure 15 shows the results of proposed ECD system when applied to a dataset of 500 nonexplicit videos. The proposed scheme detects the nonexplicit images accurately, i.e., 97% (blue dots), with only little inaccurate images' detection, i.e., 3% (red dots).

In Figure 16, the combined dataset of 1000 explicit and nonexplicit videos is used to evaluate the performance of the proposed system. The proposed scheme efficiently detects the explicit (blue dots) and nonexplicit (red dots) videos, with the explicit videos tagged as low, medium, and high (shown as black dotted lines) based on the contents. Further, it should also be noted that only few instances of inaccurate images' detection (green dots with black circle) are also present. The percentage of accurate detection is on the higher side, i.e., 95% with only 5% of inaccurate detection. The inaccuracy is due to the various factors (skin tone, lightening, and

Table 5: Confusion matrix for video dataset.

N=1000	Actual: EXPLICIT	Actual: NON-EXPLICIT
Predicted: EXPLICIT	465	15
Predicted: NON-EXPLICIT	35	485
True Positive Rate (Sensitivity) = 0.9300, **False Positive Rate**= 0.0300		
Precision= 0.9688, **Accuracy**= 0.9500, **F1 Score**= 0.9490		

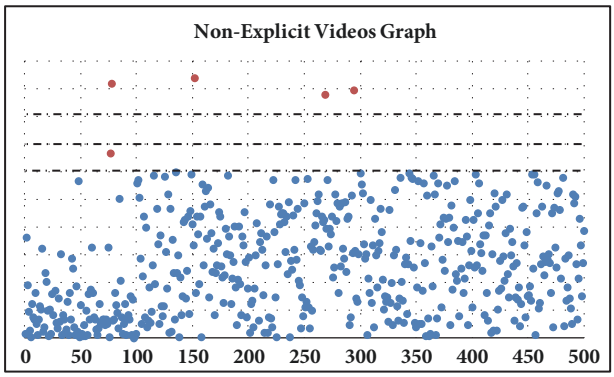

Figure 15: Result graph of application of the proposed ECD system on 500-nonexplicit video dataset.

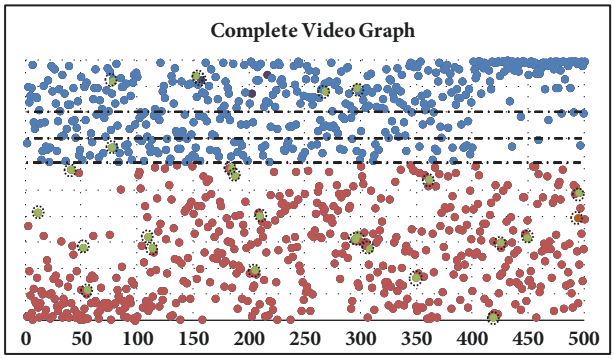

Figure 16: Result graph of application of proposed ECD system on complete (explicit/nonexplicit) video dataset.

exposure) that the video dataset contains that might lead to some prediction errors.

In Table 5, the classification error of the proposed ECD system is analyzed by using confusion table for video dataset. The results presented in Table 3 are for datasets of 500 explicit and 500 nonexplicit videos. From the results, it can be observed that the proposed work correctly classified 465 explicit videos and only 35 were incorrectly classified as nonexplicit out of 500. Similarly, for the 500 nonexplicit videos, 485 were classified correctly and 15 were incorrectly classified as explicit.

6. Conclusion

The main purpose of this proposed system is to provide parental monitoring over those contents or materials that are not ethically good for the societies. The key points of the proposed ECD system are (i) being an open source system, (ii) being computationally efficient, (iii) being highly robust, and (iv) ease of deployment in multiple modes which includes standalone on individual desktop, with proxy server, and dedicated server as content filtering system to detect or restrict the explicit content.

The results obtained by applying proposed system on real data (i.e., images/videos) significantly proved its accuracy (i.e., ~95%) in classifying NSFW contents from non-NSFW content. Further, the accuracy of the proposed technique also significantly outperforms the accuracy of YCBCR based approach on the same test dataset [12]. Therefore, we have strong recommendation for the deployment of proposed system in real environment.

Data Availability

The data used to support the findings of this study are available from the corresponding author upon request.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

References

[1] Y. C. Lin, H. W. Tseng, and C. S. Fuh, "Pornography detection using support vector machine," in *Proceedings of the 16th IPPR Conference on Computer Vision, Graphics and Image Processing (CVGIP '03)*, vol. 19, pp. 123–130, 2003.

[2] H. Zhu, S. Zhou, J. Wang, and Z. Yin, "An algorithm of pornographic image detection," in *Proceedings of the 4th International Conference on Image and Graphics, ICIG '07*, pp. 801–804, August 2007.

[3] P. Kakumanu, S. Makrogiannis, and N. Bourbakis, "A survey of skin-color modeling and detection methods," *Pattern Recognition*, vol. 40, no. 3, pp. 1106–1122, 2007.

[4] J. A. M. Basilio, G. A. Torres, G. S. Perez, L. K. T. Medina, H. M. P. Meana, and E. E. Hernadez, "Explicit content image detection," *Signal and Image Processing: International Journal*, vol. 1, no. 2, pp. 47–58, 2010.

[5] J. A. Marcial-Basilio, G. Aguilar-Torres, G. Sánchez-Pérez et al., "Detection of pornographic digital images," *International Journal of Computers*, vol. 5, no. 2, pp. 298–305, 2011.

[6] J. A. M. Basilio, G. A. Torres, G. S. Pérez, L. K. T. Medina, and H. M. P. Meana, "Explicit image detection using YCbCr space color model as skin detection," *Applications of Mathematics and Computer Engineering*, pp. 123–128, 2011.

[7] S. M. Kia, H. Rahmani, R. Mortezaei, M. E. Moghaddam, and A. Namazi, "A Novel Scheme for Intelligent Recognition of Pornographic Images," https://arxiv.org/abs/1402.5792.

[8] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2016*, pp. 770–778, July 2016.

[9] I. G. P. S. Wijaya, I. B. K. Widiartha, and S. E. Arjarwani, "Pornographic image recognition based on skin probability and eigenporn of skin ROIs images," *TELKOMNIKA Telecommunication Computing Electronics and Control*, vol. 13, no. 3, pp. 985–995, 2015.

[10] I. G. P. S. Wijaya, I. B. K. Widiartha, K. Uchimura, and G. Koutaki, "Pornographic image rejection using eigenporn of simplified LDA of skin ROIs images," in *Proceedings of the 14th International Conference on QiR (Quality in Research), QiR '15*, pp. 77–80, idn, August 2015.

[11] A. B. Burgess and C. A. Mattmann, "Automatically classifying and interpreting polar datasets with Apache Tika," in *Proceedings of the 15th IEEE International Conference on Information Reuse and Integration, IEEE IRI '14*, pp. 863–867, 2014.

[12] Explicit and Non-Explicit Dataset, http://drive.google.com/open?id=1ivxuwwNQuFjxLy4fY2OUo3ehNwj2oNQG.