*Article*

# Performance Analysis of Intelligent Neural-Based Deep Learning System on Rank Images Classification

**Muhammad Hameed Siddiqi[1,*], Asfandyar Khan[2], Muhammad Bilal Khan[2], Abdullah Khan[2], Madallah Alruwaili[1] and Saad Alanazi[1]**

[1]College of Computer and Information Sciences, Jouf University, Sakaka, Aljouf, 73211, Kingdom of Saudi Arabia
[2]Institute of Computer Science and Information Technology, ICS/IT FMCS, University of Agriculture, Peshawar, 25000, Pakistan
*Corresponding Author: Muhammad Hameed Siddiqi. Email: mhsiddiqi@ju.edu.sa

**Abstract:** The use of the internet is increasing all over the world on a daily basis in the last two decades. The increase in the internet causes many sexual crimes, such as sexual misuse, domestic violence, and child pornography. Various research has been done for pornographic image detection and classification. Most of the used models used machine learning techniques and deep learning models which show less accuracy, while the deep learning model ware used for classification and detection performed better as compared to machine learning. Therefore, this research evaluates the performance analysis of intelligent neural-based deep learning models which are based on Convolution neural network (CNN), Visual geometry group (VGG-16), VGG-14, and Residual Network (ResNet-50) with the expanded dataset, trained using transfer learning approaches applied in the fully connected layer for datasets to classify rank (Pornographic *vs.* Nonpornographic) classification in images. The simulation result shows that VGG-16 performed better than the used model in this study without augmented data. The VGG-16 model with augmented data reached a training and validation accuracy of 0.97, 0.94 with a loss of 0.070, 0.16. The precision, recall, and f-measure values for explicit and non-explicit images are (0.94, 0.94, 0.94) and (0.94, 0.94, 0.94). Similarly, The VGG-14 model with augmented data reached a training and validation accuracy of 0.98, 0.96 with a loss of 0.059, 0.11. The f-measure, recall, and precision values for explicit and non-explicit images are (0.98, 0.98, 0.98) and (0.98, 0.98, 0.98). The CNN model with augmented data reached a training and validation accuracy of 0.776 & 0.78 with losses of 0.48 & 0.46. The f-measure, recall, and precision values for explicit and non-explicit images are (0.80, 0.80, 0.80) and (0.78, 0.79, 0.78). The ResNet-50 model with expanded data reached with training accuracy of 0.89 with a loss of 0.389 and 0.86 of validation accuracy and a loss of 0.47. The f-measure, recall, and precision values for explicit and non-explicit images are (0.86, 0.97, 0.91) and (0.86, 0.93, 0.89). Where else without augmented data the VGG-16 model reached a training and validation accuracy of 0.997, 0.986 with a loss of 0.008, 0.056. The f-measure, recall, and precision values for explicit and non-explicit images are (0.94, 0.99, 0.97)

and (0.99, 0.93, 0.96) which outperforms the used models with the augmented dataset in this study.

## 1 Introduction

The use of the internet is increasing all over the world on a daily basis in the last two decades. In this modern society, the internet is used to access global information [1]. According to "We Are Social and Hootsuite" in 2019 [2], 45 of users now using the internet of the world's population. With increasing the number of internet users, the use of the internet increased. And the communication network resources are more actively used as increased the number of internet users. Due to active network communication, resource information is easily shared on the internet at any time. And people can easily promote various information tools and share information on the Internet [1]. Due to increasing the number of internet users the misuse of the internet is also increased. Using the internet various pornographic information has also been usually reported [3–6].

On the other side, the authors in [7] reported that in the past ten years with the use of the internet negative content such as pornography has significantly increased. The negative internet content leads to many moral issues and social problems and affects people's normal lives, especially teenagers, [8,9]. Moreover, various research has been done which described that with the use of the increase in internet negative content such as pornography causes sensual crimes, domestic violence, sexual manipulation, and child pornography at an upward rate [10]. Nowadays, majority of the people utilized different kinds of social media platforms to share various information and communicate with the world. Most young children use various kinds of social media platforms on their smartphones for playing games. While playing games, different pornographic material and ads are shared, which badly affect the mind of teenage kids. On the other hand, it is also widely reported that any information related to pornographic content is distributed on the internet.

Internet pornography content affects many people's life especially adolescents and creates many social problems and moral issues [1]. It is important to recognize internet pornography by using proper internet resources and principles and employing healthy development for using the internet. Therefore, it needs to plan a pornographic image recognition model, which effectively detected and recognized pornographic images [7,11,12]. Due to the developments in computer vision, content-based internet pornographic image recognition has been widely investigated [13]. In past decades many researchers have done various research to resolve this problem. However, there are still many challenges in the existing detection methods of pornographic images. Firstly, to classify the pornographic image and non-pornographic images, in which the pornographic image recognition task is considered as category A, and normal images are considered as category B, considered a binary classification task [1]. Currently, deep learning methods in the field of image recognition achieved great attention from researchers and attained great success for image-based pornography detection. Similarly, to automatically extract pornographic image features to detect pornography [14] presented multi-layer convolutional neural networks. Which shows better performance in the field of pornography detection.

Moreover, in [15], the authors developed an efficient model for child pornographic image detection which is based on a skin tone filter with a novel set of facial features systems to increase the image

recognition performance. Similarly, to further improve pornography detection, the authors of [16] presented a new model which used transfer learning with multiple feature fusion, which gives better performance for negative content image recognition, which showed the usefulness of transfer learning. Although [17] proposed a human pose model for sexual organ detection. Similarly, in [18], the authors used texture features and the maximum posterior approach for pornographic classification. The authors of [19] proposed a deformable part model and support vector machine (SVM) for sexual organ detection. Moreover, in [20], the authors used a high-level semantic tree model and SVM for Pornographic classification that achieved 87.6 accuracies. Furthermore, the authors of [21] implement Color features with convolutional neural networks (CNN) to classify Pornography and skirt images which show less accuracy. More specifically, in [22], the authors used artificial Neural networks (ANN) to increase the reliability of skin detectors which also did not perform better for skin detectors. Further, the authors of [23] proposed CNN training strategies for skin detection, which perform better than the ANN, Similarly, in [24], the authors implemented the combination of recurrent neural network (RNN) and CNN for human skin detection.

Similarly, for the Pornographic *vs.* Nonpornographic classification [14] proposed AlexNet & Google Net models achieved 94 accuracies. Furthermore, the authors of [25] used the Google Net, CNN, model, and Pornographic *vs.* color feature histogram for the Pornographic *vs.* Nonpornographic images classification on the NPDI dataset. Likewise, the authors of [26] used CNN with multiple instance learning (MIL) to detect exposed body parts on the NPDI dataset. The authors of [27] proposed a Google Net model for region-based recognition of sexual organs on the NPDI dataset. In [28], the authors proposed CNN and Google Net architecture for video pornography detection. Most of the used models as mentioned above used machine learning techniques and deep learning models which show less accuracy, while the deep learning model was used for classification and detection that show high performance. Therefore, it needs to develop a new deep learning model for the accurate classification of nudity images. To further enhance the deep learning model for accurate classifying this paper presents a novel deep learning model for the Pornographic *vs.* Nonpornographic classification in images. The deep learning models will be based on CNN, VGG-14, VGG-16, and ResNet-50 with expansion and VGG-16 without expansion, which are used to classify Pornographic *vs.* Nonpornographic. The main contributions of this study are:

- ■ This paper proposes ResNet-50, VGG-16, and its variant for the classification of rank images.
- ■ This paper describes the pre-trained model VGG-16 as a feature extractor for transfer learning on the NPDI images Dataset.
- ■ Further, the performance of the deep learning-based Convolutional Neural Network i.e., CNN, ResNet-50, VGGNet-14, and VGG-16 models are checked using accuracy, loss, precision, recall, and f-measure.
- ■ The proposed models use transfer learning approaches in the fully connected layer of the network.

The rest of the paper is organized as Section 2 will discuss the related work. Similarly, next Section 3  will describe the methodology of the paper. Further, Section 4 explained the result and discussion. Finally, Section 5 gives the conclusion of this research.

## 2  Related Work

A lot of existing research has previously shown that machine learning algorithms performed better and produced findings that were more accurate than those obtained. This section includes

some of the best and most efficient learning strategies that shed light on earlier developments that numerous researchers had suggested on how to enhance the learning efficacy of their networks to obtain some favorable and promising results for this category. The early two decades of the 21st century have witnessed that more information is available than ever before on the internet. The use of the internet is increasing explosively and may offer people universal access to searching web images, as a significant information transmission medium. However, pornography images are widely available on the internet which affects our young generation badly, due to the lack of control over information sources. Various studies have been conducted to detect Pornographic *vs.* Nonpornographic images. The authors of [29] suggested a new approach using Multi-Layer Perceptron combine with fuzzy integral-based information fusion for identifying pornographic images. Similarly, in [30], the authors proposed principal component analysis (PCA) to gain high accuracy and detect the desired area properly with a high rate of feature extraction. The result showed that the proposed method increases the accuracy up to 4.0 and decreases the false positive ratio (FPR) by 20.6 respectively, and the authors of [27] worked in nude image classification. For this purpose, a local-context-aware network (LocoaNet) was proposed to classify the nude image. By developing a multi-task learning scheme, LocoaNet and employing an obscene detection network can extract negative features for obscene image classification. The result analysis shows that the desired model achieved better results having high processing speed on both datasets. The authors of [19] developed a new model for the detection of obscene image detection. The author used traditional detection whose design is based on shape features. The author represents sexual organs using a histogram of gradient-based shapes. The sexual organ detector is trained by the color-saliency preserved mixture deformable part model (CPMDPM). The evaluation shows that the performance of CPMDPM is superior to shape feature-based detectors. In [18], the authors developed a new model named Adult Content Recognition with Deep Neural Networks (ACRDNN) for adult video classification. ACRDNN is a combination of convolutional networks. (ConNet) and long short-term memory (LSTM). The performance evaluation shows that the proposed model performs better than the state-of-the-art for the detection of obscene images. Furthermore, in [27], the authors designed multiple instance learning (MIL) approaches to detect pornographic images. The author proposed weighted MIL under the CNN framework. The dataset consisted of 138 K obscene images and 205 K normal images. The overall result shows that the desired model achieved 97.52 accuracies. The authors of [1] designed a porn images and text detection system for gambling websites. The proposed study uses a decision mechanism. The performance evaluation shows that the model achieved highly satisfactory results, and the framework is feasible and effortless for the detection of such items. Likewise, the authors of [31] proposed the combination of COCO-trained you only look once (YOLOv3) for nudity detection. In addition, they use CNN models such as AlexNet, Google Net, VGG16, Inception v3, and ResNet for feature extraction. The proposed model classifies images into two classes nude and normal. The performance evaluation shows that the proposed model of YOLO-CNN achieved high efficiency as compared to standard CNN Models. Table 1 shows the comparative study of previous research in this paper.

**Table 1:** Cooperative study

| Existing works | Methodology | Dataset | Result |
| --- | --- | --- | --- |
| [15] | Skin tone filter with a novel set of facial features systems | For child pornographic image detection | 80% accuracy rate TP 83% |
| [16] | Transfer learning with multiple feature fusion, recognition, | And negative content images | With 0.91% accuracy |
| [17] | Proposed a human pose model for sexual organ detection. | The sexual organ detection dataset | 89.50% detection rate & with an 18.64% error rate |
| [18] | Maximum posterior approach for pornographic classification. | Pornographic classification | With a 0.8 low error rate |
| [19] | Proposed deformable part model and SVM for sexual organ detection. | Sexual organ detection | 50% higher result in the used models |
| [20] | The semantic tree model and SVM | Pornographic classification | Achieved 87.6% accuracy. |
| [21] | Furthermore, the authors of implementing color features with CNN | Pornography, and upskirt images dataset | Show 90.23% accuracy |
| [22] | ANN to increase the reliability of detectors skin | detectors skin | Achieved 95.6176% TP & 0.8795% FP rate |
| [23] | CNN | skin detectors | Achieved 91% accuracy which is better than ANN |
| [24] | RNN and CNN | For human skin detection. | Show better performance |
| [25] | Google Net, CNN, model and pornographic *vs.* color feature histogram | Pornographic *vs.* nonpornographic classification on the NPDI dataset. | The proposed model reaches 99.31%, accuracy which is 2.67% higher than the CNN |
| [26] | CNN with multiple instance learning (MIL) | To detect exposed body parts on the NPDI dataset. | Show remarkable accuracy with 97.01% TP at 1% FP, |

(Continued)

**Table 1 (continued)**

| Existing works | Methodology | Dataset | Result |
|---|---|---|---|
| [27] | Proposed Google Net model for region-based recognition. | Sexual organs on the NPDI dataset | The claimed accuracy is 97% with a rate of true positive and 1% with a rate of false positive |
| [28] | proposed CNN and Google Net architecture for | Video pornography detection. | 97.8% is the claimed accuracy along with 64.3% error reduction |
| [29] | Multi-layer perceptron combine with fuzzy integral- | pornographic images | During training, the claimed accuracy is 93.1% in true positive and 8.2% in false positive; while in testing, 87.1% and 5.4% respectively |
| [30] | (PCA) | Pornographic images | Increase the accuracy by up to 12% and decrease the FP by about 16%, |
| [31] | YOLOv3 for nudity detection. | Nudity image detection. | The accuracy is improved by 4% from 85.5% to 89.5%. Moreover, the AUC is also increased from 93% to 97%. |

## 3 Learning Models

### 3.1 CNN Overview

The generalized structure of the CNN network was proposed by [32]. Due to computation hardware limits the researcher did not widely use this network before for training. In the 1990s, the authors of [33] proposed CNNs with gradient-based learning algorithms for handwritten digit classification problems. After attaining successful results in different recognition tasks many works have been done by different researchers to further improve CNNs for achieving more state-of-the-art results. Fig. 1 shows the overall CNNs architecture. The CNN architecture is made up of a variety of layers, including completely linked, max-pooling, and convolution layers. Max-pooling layers are in the network's middle level while convolutional layers are located at a low level. By performing convolution operations upon the input images, each node of the convolution layer retrieves the features of the input images [34]. Two Conv2D layers having Relu activation functions, two Max-pooling layers, a Flatten layer, & two dense layers with Relu and sigmoid activation functions make up the tiny CNN model's detailed architecture. Conv2D has a kernel size of (3332) and a bias of 32. Batch input is in the form of 224, 224, 3, and 32-filter layers with Relu activation and Maxpooling layers. Similar to

the first conv2D layer, the second conv2D layer was followed by Maxpooling layers. Additionally, the Relu activation function was utilized in the first dense layer, which has a kernel size of $(93312 * 128)$, a bias value of 128, and a kernel size. The last dense layer, on the other hand, employed a kernel size of (1281), a bias of 1, and a sigmoid activation function. The next model used in the paper is the ResNet-50 model which is discussed in the next section.
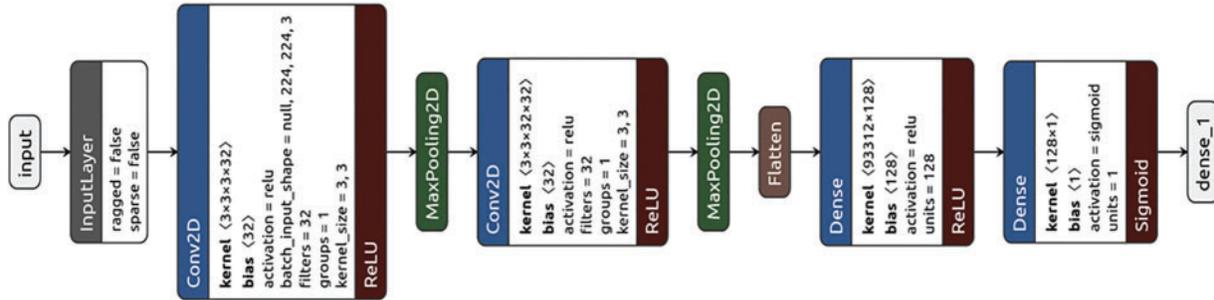
**Figure 1:** The structure of CNN [34]

### 3.2 Residual Network (ResNet-50)

The ResNet network model was developed by Kaiming [35,36]. This model is consisting of designing ultra-deep networks which solve the vanishing gradient problem that exists in the previous model. Various researchers have developed ResNet with many different numbers of layers; such as 101, 34, 50, 152, and even1202. One of the most efficient models among all of these models is the ResNet-50 network model. The ResNet-50 model consists of a total of 50 layers. Among all the 50 layers 49 layers are convolution layers while at the end of the network one fully connected layer is used. There were 3.9 million MACs and 25.5 million weights in the entire network. Fig. 2 depicts the fundamental block diagram of the residual block inside the ResNet architecture. A residual connection makes up each of the basic residual blocks in the ResNet model. Based on the outputs, it is possible to define the output of a residual layer that is created from the output of the preceding layer after performing various operations, such as convolution with various filter sizes and batch normalization (BN) following through an activation function, like a Relu. However, based on the various architectures of residual networks, the operations inside the residual block can vary [35].
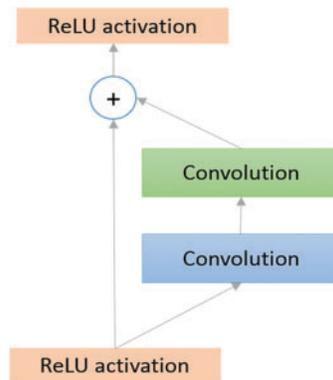
**Figure 2:** Residual block diagram

## 4 Methodology

### 4.1 Dataset

This study used Explicit and Non-Explicit Pornography datasets containing nearly 3,800 obscene and 3,200 non-obscene images, which are collected from the Explicit Content Detection System [37]. The dataset consists of certain classes of obscenity and depicts actors of multiple ethnicities, including multi-ethnic ones. The non-obscene images were chosen at random and were selected from textual search queries like "beach," "wrestling," and "swimming," which contain body skin but not nudity or porn. In total, there are 7,000 frames (images). In this research, 70%, of the dataset is used for training purposes while the rest of the dataset is used for testing the model. Figs. 3 and 4 highlight some statistics and samples of this dataset, respectively. All the images are taken from the link given https://drive.google.com/drive/folders/1T2lwjWcW3L2DQw27ruwm8hyLT9teIUb9?usp=sharing.
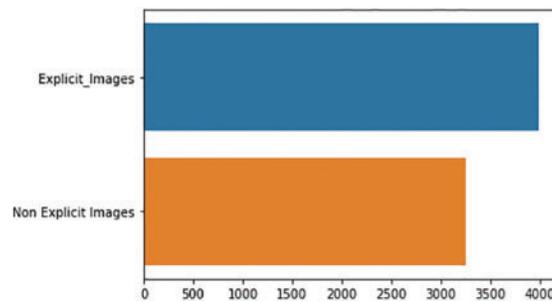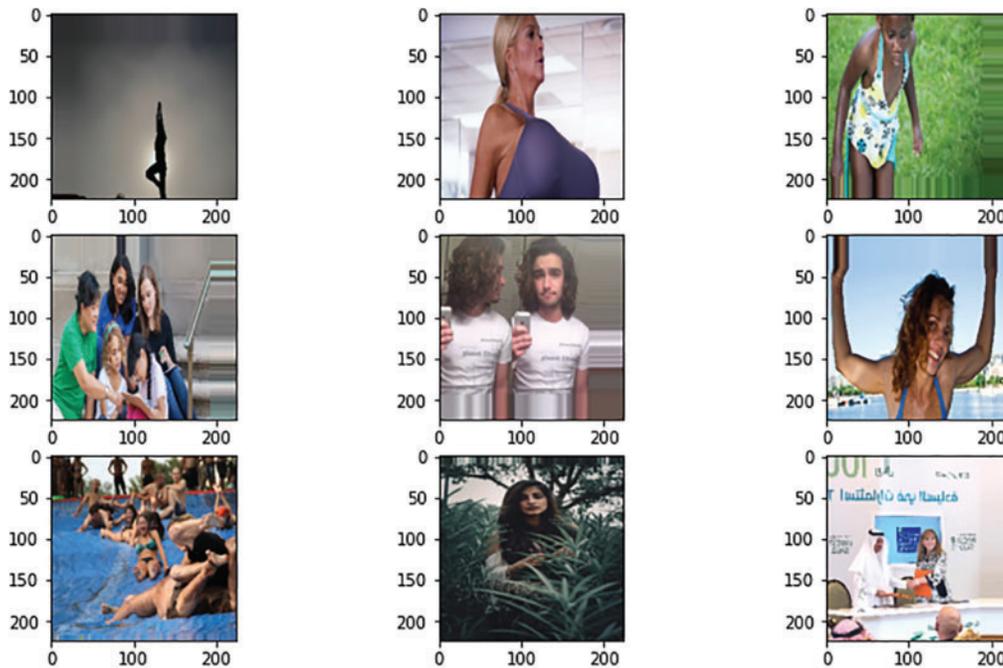


**Figure 3:** Explicit and non-explicit images



**Figure 4:** Body skin but not nudity images

### 4.2 Processing Image Data

To easily set up Python generators that automatically transform picture files in preprocessed tensors fed straight into models during training, utilize the preprocessing Keras ImageDataGenerator() method. On the training dataset, picture alteration operations including translation, rotation, and zooming are employed to produce new iterations of existing images. This study added these extra images to the model during training. It effectively completes the duties listed below for us: Recognize the RGB pixel grids from the JPEG data. Then transform these into floating point tensors. The image is randomly zoomed by 0.1 using the zoom range parameter. The image is rotated by a random 10 degrees using the rotation range parameter.

The height_shift_range and the width_shift_range parameters are used to translate the image randomly horizontally or vertically of the image's width or height by a 0.1 factor. The sheer range parameter is used to apply shear-based transformations randomly. Rescale (between 0 and 255) the pixel values to the [0, 1] interval. It takes the image pixel values and normalizes them to have a 0 mean and 1 standard deviation. The data expansion operation was applied to 30% testing dataset. After applying these five data generator functions mentioned in Table 2, on testing data, the total number of augmented data reach up to 9265 images. The main purpose of applying the data generator function is to check the performance of the used models with augmented and without augmented datasets. Table 2 gives the ImageDataGenerator function detail.

**Table 2:** Image data generator function

| ImageDataGenerator name | Selected value |
|---|---|
| Rescale | 1/255 = 0.0039 |
| Width_shift_range | 0.2 |
| Heigh_shift_range | 0.2 |
| Shear_range | 0.2 |
| Zoom_range | 0.2 |
| Rotation_range | 10 |

### 4.3 Processing VGG-16 Model

VGG-16 and VGG-14-based CNN architectures are proposed in this study. The VGG-14 model has 14 layers. While the VGG-16 has 16 layers. The ImageNet database was used to train the model. On various image classification & image recognition datasets, this model showed good classification performance. VGG-16 and VGG-14 can be implemented using a high-level library that supports Tensor Flow & Theano Backend called the Keras framework. Thirteen convolutional layers and a dense or linked layer were used to build the VGG-14 model. While the VGG-16 was built using 3 fully connected layers and 13 convolutional layers. Up till the dense prediction layer, each of the 14 levels carries a weight that is incrementally transferred. Whereas the convolutional layers need a $3 \times 3$ frame with such a Relu activation function for each hidden layer, the pooling layer uses a $2 \times 2$ window. The SoftMax activation function is used in the dense prediction layer or the final layer. The fully connected layer receives the bottleneck characteristic from the final activation layer. To cut down on training time, the bottleneck feature can be used as a pre-training weight. The completely connected layer is where the transfer learning takes place. In our experiment, Fig. 5 displays the hyper-parameters for the suggested max-pooling model's diagram. We made the following adjustments to the VGG-14 model:

■ Remove the dense fully connected layer, which employs 1000 classes of SoftMax.
■ Replace the dense fully connected layer with two sigmoid classes.
■ All CNN layers are trainable before the fully connected layers because we want to train all CNN layers, and our modified VGG-14 is also trained on fully connected layers.
■ To improve the training's efficiency, use Adam's [38] learning algorithm.
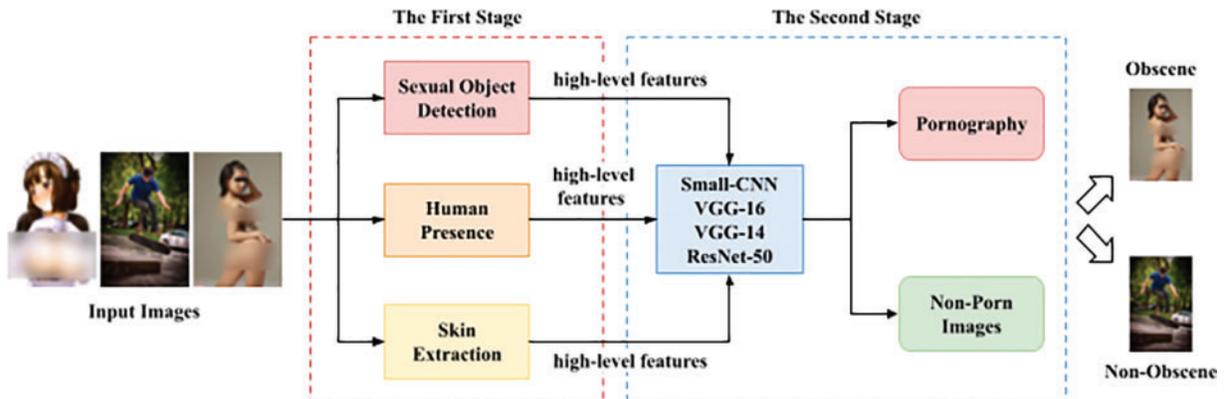■ Finally, the Optimizer Adam learning rate is 5e-5, and binary cross-entropy is used to evaluate losses.



**Figure 5:** The architecture of the proposed models

## 5 Evaluation Parameters

To check the performance of the proposed model is compared to the state-of-the-art models in terms of various parameters which is used by the various researcher is discuss. The confusion matrix ($n \times n$) which might be known as the likelihood table or error matrix is utilized to describe the performance of any identification analysis of having various drawbacks. The size of such a matrix relies on the produced number of classes. Therefore, the summation of the entire recovered positive values (such as true positive (TP) and true negative (TN)) is reflected as accurate identification. In true positive, the subject is presented as associated with a class; however, it does not belong to that class. While, in false positive (FP), the element is presented as not associated with a class, but does not belong to that class. On the other side, the rest of the cases are considered as rejected which are the combination of false positive and false negative (FP + FN).

### 5.1 Recall

The recall is known as the rate of true positive rate, which is the whole number of TP values divided by the respective summation of the whole values of the TP and FN. Recall is a quantitative measure that presents the extensiveness of the results. The rate of the TP is an accurately identified number, which means high recall is accurately retrieved along with high efficiency.

$$\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}} \tag{1}$$

### 5.2 Precision

The precision describes the whole number of the value that is positively retrieved, which means that the whole number of retrieved values is correlated and belongs to a specific class. Precision might

be measured by dividing the number of TP values by the respective summation of the whole values of the TP and FP.

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}} \tag{2}$$

### 5.3 F-Measure

F-measure uses average harmonic instead of using average arithmetic. F-Measure is calculated from recall and precision is given as.

$$\text{F} - \text{Measure} = \frac{2 * \text{Recall} * \text{Precision}}{\text{Recall} + \text{Precision}} \tag{3}$$

### 5.4 Accuracy

Accuracy presents the efficiency of the proposed model, which shows that the model is safe and compact with the detection of positive and negative values. Also, the proposed approach can retrieve accurate values, which are presented in Eq. (4).

$$\text{Accuracy} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{TN} + \text{FP} + \text{FN}} \tag{4}$$

## 6 Simulation Results

This section explains the simulation result of the various models used in this study. First, the random hyperparameter search experiment results, with the Small-CNN model, are presented. Then the result is followed by VGG16, VGG-14, and ResNet-50, models trained with transfer learning using Image Net pre-trained weights.

### 6.1 Small CNN

CNN architecture is the first proposed model of this research which is self-defined architecture, that's why we called it small CNN. The small CNN model consists of two convolution layers with max pooling, one flatten layer, and two dense layers. The kernel size is $3 \times 3$, the input shape of the image is $224 \times 224$, the feature maps are 32, and Relu has used an activation function. While the output layer consists of a sigmoid activation function. Table 3 shows the hyperparameter of the small CNN model and Table 4 describes the small CNN architecture.

**Table 3:** Hyperparameters value rates

| Hyperparameter | Selected value | Value range |
|---|---|---|
| Convolutional layers | 2 | {1, 2} |
| Fully connected layers | 2 | {1, 2} |
| Kernel size | 3 | {3, 5, 7, 9, 11} |
| Feature maps | 32 | {16, 32, 48, 64} |
| Learning rate | 0.00001 | $10 - 5 - 10 - 4$ |
| Input size | 224 | {64, 96, 128, 224} |

**Table 4:** The architecture of the small CNN

| Layer | K | AF | FM | Input size | Output size | Parameters |
|---|---|---|---|---|---|---|
| Input size | | | | [224 × 224 × 3] | [224 × 224 × 3] | 0 |
| Conv2d max_pooling2d | 3 × 3 | Relu | 32 | [224 × 224 × 3] | [222 × 222 × 32] | 896 |
| | | | | [222 × 222 × 32] | [111 × 111 × 32] | 0 |
| Conv2d_1 max_poolg2d1 | 3 × 3 | Relu | 32 | [111 × 111 × 32] | [109 × 109 × 32] | 9,248 |
| | | | | [109 × 109 × 32] | [54 × 54 × 32] | 0 |
| Flatten | | | | [54 × 54 × 32] | 93312 | 0 |
| Dense | | Relu | | 93312 | 128 | 11,944,064 |
| dense_1 | | sigmoid | | 128 | 1 | 129 |
| | | | | | | **11,954,337** |

Fig. 6 displays the CNN model's total accuracy and loss. It demonstrates the training performance of CNN. The learning algorithm is archived once CNN training has been completed. When training starts on the first epoch its accuracy is 0.64 and gradually the accuracy increased with 8 epochs the final accuracy reached 0.776. The early halting policy caused the training to end after 8 epochs. After the last epoch, the model achieved an accuracy for the training of 0.776 and an accuracy for validation of 0.79 with losses of 0.48 & 0.46. Similarly, Table 5 presents the performance assessment of the CNN on the training dataset. Table 5 contains the explicit Images and Non_Clear_Pictures According to testing data in terms of precision, recall, and $f_1$-measure. Here the average clear image precision is 0.81 and the average Non_Clear_Pictures is 0.79 on testing data. The explicit images recall average is 0.81 and the average is Non_Clear_Pictures 0.79 on testing data. F-measure average explicit Images is 0.81 while the average micro is 0.79. While Figs. 6 and 7 describe the training performance and confusion matrix performance of the CNN model.



**Figure 6:** Training performance of small CNN

**Table 5:** Precision recall and $f$-measure of the performance of the CNN model

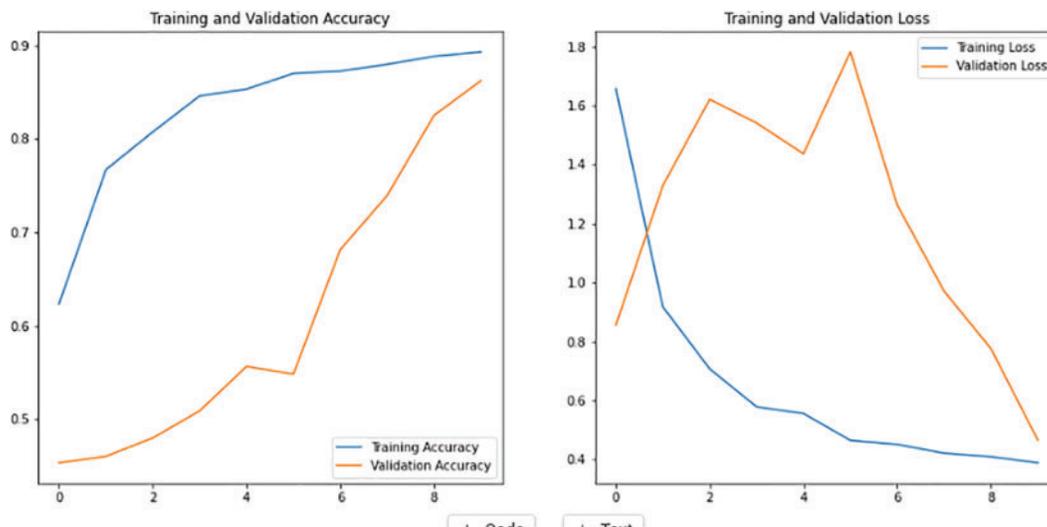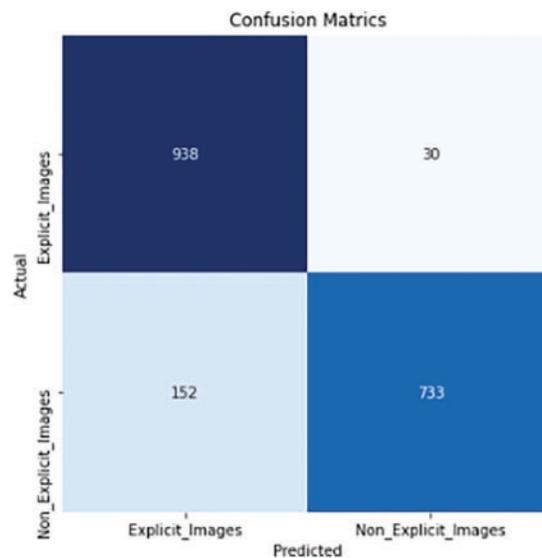| Dataset | Support | Precision | Recall | F-measure |
|---|---|---|---|---|
| Clear pictures | 967 | 0.81 | 0.81 | 0.81 |
| Non_clear pictures | 886 | 0.79 | 0.80 | 0.79 |
| Macro_avg | 1853 | 0.79 | 0.79 | 0.79 |
| Weighted_avg | 1853 | 0.79 | 0.79 | 0.79 |



**Figure 7:** Confusion matrix of small CNN

## 6.2 ResNet-50

One of the most efficient models among all of the ResNet models is the ResNet-50 network model. The ResNet-50 model has a total of 50 layers. Among all the 50 layers, 49 layers are convolutional layers of the total of layers, and finally 1 dense layer. For the entire network, there are 25.5 M weights and 3.9 M MACs, respectively. Fig. 2 depicts the fundamental block diagram of the residual block in the ResNet architecture. A conventional feed-forward network with a residual connection is called ResNet-50. Table 6 shows the ResNet-50 performance during training. On the final epoch, the model reached a training accuracy of 0.89 with a loss of 0.389 and 0.86 of validation accuracy and a loss of 0.47. Similarly, Table 6 shows the performance evaluation of the ResNet-50 model, which contains the explicit Images and Non_Clear_Pictures precision, recall, and $f$-measure on testing data. Here the average explicit images precision is 0.86 and the average Non_Clear_Pictures is 0.96 using validation data. Recall on clear images is 0.97 and on Non_Clear_Pictures is 0.83. Similarly, the $f$-measure on average explicit Images is 0.91, and on Non_Clear_Pictures is 0.89. The overall accuracy and loss of the ResNet-50 model are shown in Fig. 8. While Fig. 9 shows the confusion matrix performance of the ResNet-50 model.

**Table 6:** Precision recall and *f* measure the performance of the ResNet-50 model

| Dataset | Support | Precision | Recall | F-measure |
|---|---|---|---|---|
| Clear pictures | 967 | 0.87 | 0.98 | 0.92 |
| Non-clear pictures | 886 | 0.97 | 0.84 | 0.90 |
| Macro_avg | 1853 | 0.91 | 0.90 | 0.90 |
| Weighted_avg | 1853 | 0.91 | 0.90 | 0.90 |



**Figure 8:** Training performance of ResNet-50 with expansion data



**Figure 9:** Training performance of ResNet-50 with expansion data

### 6.3 VGG-16

VGG-16-based CNN architectures are proposed in this study. This study further validated the VGG-14 model which has 14 layers. The model was trained using the Image Net database. The VGG-14 model is trained with expansion data. While the VGG-16 model is trained with both augmented and without augmented data. This model demonstrated strong classification performance on a variety of image classification and image recognition datasets. The Keras framework, tensor flow supported with Theano high-level library, is used to implement VGG-14 and VGG-16 models. The VGG-14 model was constructed having 13 convolutional layers with one dense or connected layer. Each of the 14 layers has a weight that is gradually increased to a dense prediction layer. The pooling layer employs a $2 \times 2$ window, while the convolutional layers use a $3 \times 3$ frame with a Relu activation function in each hidden layer. The SoftMax gradient descent algorithm is used in the dense or the prediction layer. The bottleneck feature is sent down to the fully connected layer from the final activation layer. To reduce training time, pre-trained weights are assigned to the bottleneck feature. Fully connected layers produce transfer learning Tables 7 and 8 show the performance of VGG-14, and VGG-16 models in terms of loss, accuracy, precision-recall, and f-measure. The VGG-16 model with augmented data reached a training and validation accuracy of 0.97, 0.94 with a loss of 0.079, 0.16. The recall, f-measure and precision values for explicit, and non-explicit images are (0.94, 0.94, 0.94) and (0.94, 0.94, 0.94). Where else without augmented data the VGG-16 model reached a training and validation accuracy of 0.997, 0.986 with a loss of 0.008, 0.056. Further, the precision, recall, and f-measure values for explicit and non-explicit images without augmented data are (0.94, 0.99, 0.97) and (0.99, 0.93, 0.97). Similarly, the VGG-14 model contains the explicit Images and Non_Clear_Pictures of precision, recall, and $f_1$-measure on testing data. The VGG-14 model with augmented data reached a training and validation accuracy of 0.98, 0.96 with a loss of 0.059, 0.11. The f-measure, recall, and precision values for explicit and non-explicit images are (0.98, 0.98, 0.98) and (0.98, 0.98, 0.98). Similarly, Fig. 10 shows the accuracy and loss of the VGG-16 model without an augmented dataset. The overall accuracy and loss of the VGG-16 are shown in Fig. 11 with the augmented dataset. While Fig. 12 shows the performance evaluations of the VGG-14 model with the augmented dataset in terms of accuracy and loss. Further, Figs. 13 and 14 consist of the confusion matrix of the VGG-16 and VGG-14 model respectively. Where the CNN give the evaluation performed on the average Explicit Images in term of precision is 0.80 and the average Non_Clear_Pictures is 0.78 on testing data. The explicit images recall average is 0.80 and the average is Non_Clear_Pictures 0.79 on testing data. F-measure average explicit images is 0.80 while the average micro is 0.78. In the ResNet-50 performance during training on the final epoch, the model reached a training accuracy of 0.89 with a loss of 0.389 and 0.86 of validation accuracy and a loss of 0.47. Here the average explicit images precision is 0.86 and the average Non_Clear_Pictures is 0.96 on testing data. Recall on explicit images is 0.97 and on Non_Clear_Pictures is 0.83. Similarly, the $f$-measure on average clear Images is 0.91, and on Non_Clear_Pictures is 0.89.

**Table 7:** Accuracy and loss of VGG-16, and VGG-14 models

| Models | Tr-Acc | Val-Acc | Tr-Loss | Val-Loss |
|---|---|---|---|---|
| VGG-16 with expansion | 0.98 | 0.95 | 0.080 | 0.17 |
| VGG-14 with expansion | 0.99 | 0.97 | 0.060 | 0.12 |
| VGG-16 Without expansion | 0.99 | 0.98 | 0.008 | 0.05 |

(Continued)

**Table 7 (continued)**

| Models | Tr-Acc | Val-Acc | Tr-Loss | Val-Loss |
|---|---|---|---|---|
| CNN with expansion | 0.78 | 0.78 | 0.48 | 0.46 |
| ResNet-50 with expansion | 0.89 | 0.86 | 0.39 | 0.47 |

**Table 8:** Precision-recall and *f*-measure performance of VGG-16 and VGG-14 models

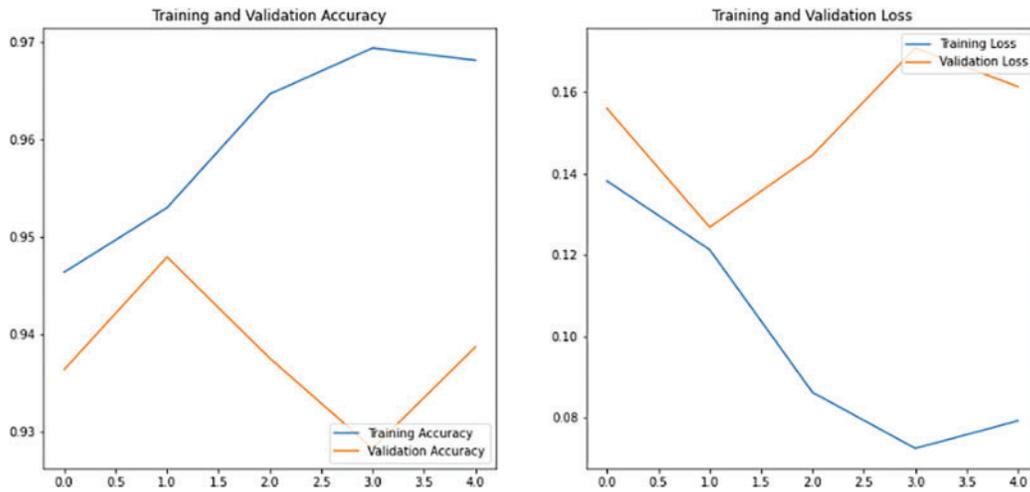| Dataset | Support | Precision | Recall | F-measure |
|---|---|---|---|---|
| VGG-16 with augmentation data | | | | |
| Clear pictures | 967 | 0.95 | 0.95 | 0.95 |
| Non-clear pictures | 886 | 0.95 | 0.95 | 0.95 |
| Macro_avg | 1853 | 0.94 | 0.94 | 0.94 |
| Weighted_avg | 1853 | 0.94 | 0.94 | 0.94 |
| VGG-14 with expansion | | | | |
| Clear pictures | 969 | 0.99 | 0.99 | 0.99 |
| Non-clear pictures | 886 | 0.99 | 0.99 | 0.99 |
| Macro_avg | 1853 | 0.98 | 0.98 | 0.98 |
| Weighted_avg | 1853 | 0.98 | 0.98 | 0.98 |
| VGG-16 without expansion | | | | |
| Clear pictures | 969 | 0.95 | 0.98 | 0.98 |
| Non-clear pictures | 886 | 0.98 | 0.94 | 0.96 |
| Macro_avg | 1853 | 0.97 | 0.96 | 0.96 |
| Weighted_avg | 1853 | 0.96 | 0.96 | 0.96 |
| CNN with expansion | | | | |
| Clear pictures | 969 | 0.81 | 0.81 | 0.81 |
| Non-clear pictures | 886 | 0.79 | 0.80 | 0.79 |
| Macro_avg | 1853 | 0.96 | 0.97 | 0.97 |
| Weighted_avg | 1853 | 0.96 | 0.95 | 0.96 |
| ResNet-50 with expansion | | | | |
| Clear pictures | 969 | 0.87 | 0.98 | 0.92 |
| Non-clear pictures | 886 | 0.97 | 0.84 | 0.90 |
| Macro_avg | 1853 | 0.95 | 0.96 | 0.97 |
| Weighted_avg | 1853 | 0.96 | 0.97 | 0.96 |

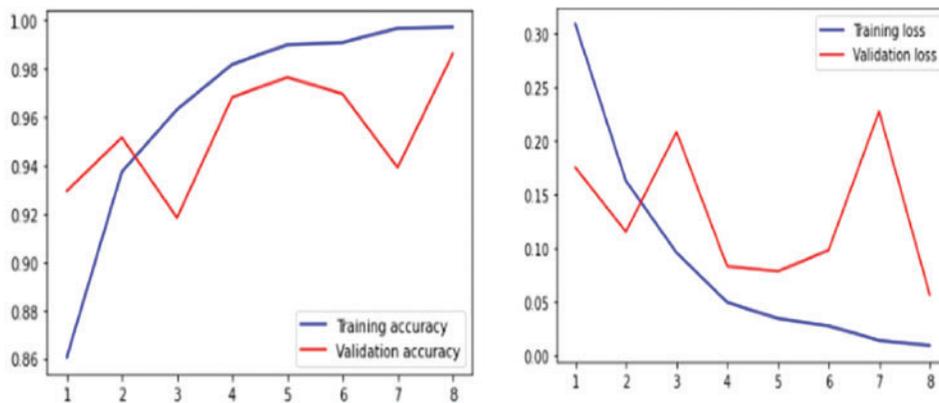**Figure 10:** Training performance of VGG-16 without augmented data



**Figure 11:** Training performance of VGG-16 with augmented data
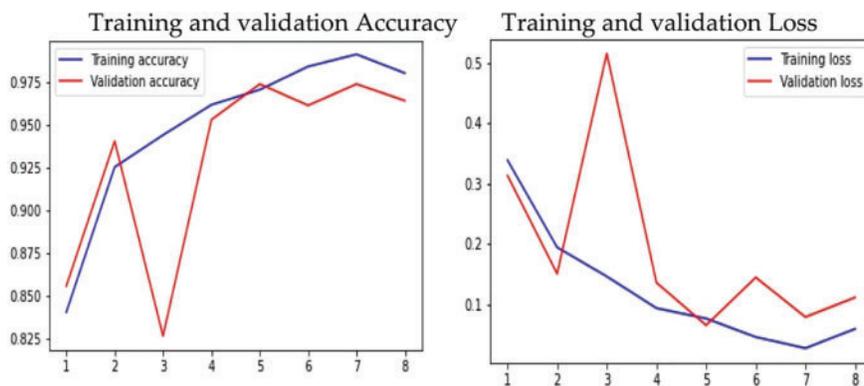


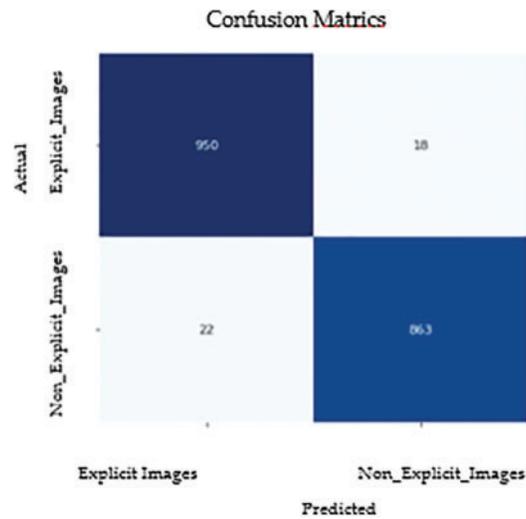**Figure 12:** Training performance of VGG-14 with expanded data
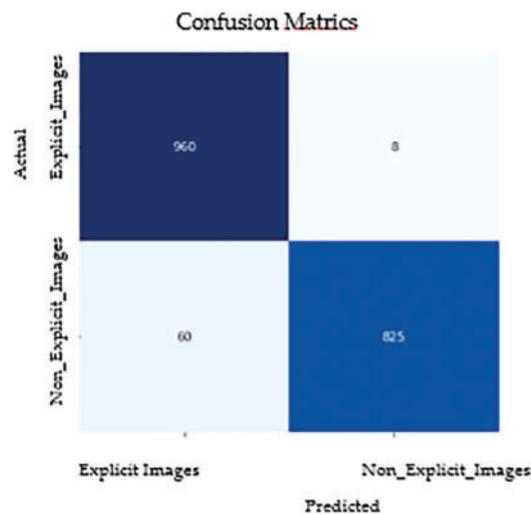
**Figure 13:** Confusion matrix of VGG-16



**Figure 14:** Confusion matrix of VGG-14

## 7 Conclusions

Nowadays, majority of the people utilized different kinds of social media platforms to share various information and communicate with the world. Most young children use various kinds of social media platforms on their smartphones for playing games. While playing games, different pornographic material and ads are shared, which badly affect the mind of teenage kids. On the other hand, it is also widely reported that any information related to pornographic content is distributed on the internet. While playing games, different pornographic material and ads are shared, which badly affect the mind of teenage kids. Various research has been done on pornographic images. Therefore, this research proposed deep learning models will be based on CNN, VGG-14, and ResNet-50 with the expanded dataset, and the VGG-16 model was trained with transfer learning occurs in the fully connected layer with both augmented and without augmented dataset are used to classify Pornographic *vs.*

Nonpornographic. The simulation result shows that VGG-16 performed better than the used model in this study with augmented data. From the overall simulation result, it shows that the VGG-16 model with augmented data reached a training and validation accuracy of 0.97, 0.94 with a loss of 0.070, 0.16. The precision, recall, and $f_1$-measure values for explicit and non-explicit images are (0.94, 0.94, 0.94) and (0.94, 0.94, 0.94). Where else without augmented data the VGG-16 model reached a training and validation accuracy of 0.997, 0.986 with a loss of 0.008, 0.056. Further, the precision, recall, and f-measure values for explicit and non-explicit images without augmented data are (0.94, 0.99, 0.97) and (0.99, 0.93, 0.97). Similarly, the VGG-14 model comprises the clear pictures and Non_Clear_Pictures of precision, recall, and $f_1$-measure on assessment data. The VGG-14 model with augmented data reached a training and validation accuracy of 0.98, 0.96 with a loss of 0.050, 0.11. Form overall analysis shows that the VGG-16 performed better without augmented data as compared with the used model in this research. In the future, this study will be extended to detect and blur the Pornographic both in video and image pornography detection using deep learning to enhance the YOLO-4 model.

**Conflicts of Interest:** The authors declare that they have no conflicts of interest to report regarding the present study.

**References**

[1] J. Chen, G. Liang, W. He, C. Xu, J. Yang *et al.,* "A pornographic images recognition model based on deep one-class classification with visual attention mechanism," *IEEE Access*, vol. 8, pp. 122709–122721, 2020.

[2] Global digital report, 2019. [Online]. Accessed: Jan. 30, Available. https://wearesocial.com/global-digital-report-2019

[3] D. K. Braun-Courville and M. Rojas, "Exposure to sexually explicit web sites and adolescent sexual attitudes and behaviors," *Journal of Adolescent Health*, vol. 45, no. 2, pp. 156–162, 2009.

[4] J. D. Brown and K. L. L'Engle, "X-rated: Sexual attitudes and behaviors associated with US early adolescents' exposure to sexually explicit media," *Communication Research*, vol. 36, no. 1, pp. 129–151, 2009.

[5] H. Chen, Y. Wu and D. J. Atkin, "Third person effect and internet pornography in China," *Telematics and Informatics*, vol. 32, no. 4, pp. 823–833, 2015.

[6] G. Tyson, Y. Elkhatib, N. Sastry and S. Uhlig, "Measurements and analysis of a major adult video portal," *ACM Transactions on Multimedia Computing, Communications, and Applications (TOMM)*, 2016, vol. 12, no. 2, pp. 1–25, 2016.

[7] F. Nian, T. Li, Y. Wang, M. Xu and J. Wu, "Pornographic image detection utilizing deep convolutional neural networks," *Neurocomputing*, vol. 210, pp. 283–293, 2016.

[8] R. Cohen-Almagor, "Online child sex offenders: Challenges and counter-measures," *The Howard Journal of Criminal Justice*, vol. 52, no. 2, pp. 190–215, 2013.

[9] M. L. Ybarra, V. C. Strasburger and K. J. Mitchell, "Sexual media exposure, sexual behavior, and sexual violence victimization in adolescence," *Clinical Pediatrics*, vol. 53, no. 13, pp. 1239–1247, 2014.

[10] M. L. Ybarra, K. L. Mitchell, M. Hamburger, M. Diener-West and P. J. Leaf, "X-rated material and perpetration of sexually aggressive behavior among children and adolescents," *Aggressive Behavior*, vol. 37, no. 1, pp. 1–18, 2011.

[11] I. M. A. Agastya, A. Setyanto and D. O. D. Handayani, "Convolutional neural network for pornographic images classification," in *Proc. of 4th Int. Conf. on Advances in Computing, Communication & Automation*, Taylor's University Lakeside Campus, Malaysia, pp. 1–5, 2018.

[12] J. Wehrmann, G. S. Simões, R. C. Barros and V. F. Cavalcante, "Adult content detection in videos with convolutional and recurrent neural networks," *Neurocomputing*, vol. 272, pp. 432–438, 2018.

[13] M. S. Nadeem, V. N. Franqueira, X. Zhai and F. Kurugollu, "A survey of deep learning solutions for multimedia visual content analysis," *IEEE Access*, vol. 7, pp. 84003–84019, 2019.

[14] M. Moustafa, "Applying deep learning to classify pornographic images and videos," ArXiv preprint arXiv: 1511.08899, 2015.

[15] N. Sae-Bae, X. Sun, H. T. Sencar and N. D. Memon, "Towards automatic detection of child pornography," in *Proc. IEEE Int. Conf. on Image Processing (ICIP)*, Paris, France, pp. 5332–5336, 2014.

[16] X. Lin, F. Qin, Y. Peng and Y. Shao, "Fine-grained pornographic image recognition with multiple feature fusion transfer learning," *International Journal of Machine Learning and Cybernetics*, vol. 12, no. 1, pp. 73–86, 2021.

[17] X. Shen, W. Wei and Q. Qian, "A pornographic image filtering model based on erotic part," in *Proc. 3rd Int. Congress on Image and Signal Processing*, Yantai, China, pp. 2473–2477, 2020.

[18] B. Choi, S. Han, B. Chung and J. Ryou, "Human body parts candidate segmentation using laws texture energy measures with skin color," in *Proc. 13th Int. Conf. on Advanced Communication Technology*, Gangwon-Do, South Korea, pp. 556–560, 2011.

[19] C. Tian, X. Zhang, W. Wei and X. Gao, "Color pornographic image detection based on color-saliency preserved mixture deformable part model," *Multimedia Tools and Applications*, vol. 77, no. 6, pp. 6629–6645, 2018.

[20] L. Lv, C. Zhao, H. Lv, J. Shang, Y. Yang *et al.,* "Pornographic images detection using high-level semantic features," in *Proc. 7th Int. Conf. on Natural Computation*, Shanghai, China, vol. 2, pp. 1015–1018, 2011.

[21] Y. Huang and A. W. K. Kong, "Using a CNN ensemble for detecting pornographic and upskirt images," in *Proc 8th Int. Conf. on Biometrics Theory, Applications and Systems (BTAS)*, NY, USA, pp. 1–7, 2016.

[22] Y. T. Alaa and A. J. Hamid, "Increasing the reliability of skin detectors," *Scientific Research and Essays*, vol. 5, no. 17, pp. 2480–2490, 2010.

[23] Y. Kim, I. Hwang and N. I. Cho, "Convolutional neural networks and training strategies for skin detection," in *Proc. IEEE Int. Conf. on Image Processing (ICIP)*, Beijing, China, pp. 3919–3923, 2017.

[24] H. Zuo, H. Fan, E. Blasch and H. Ling, "Combining convolutional and recurrent neural networks for human skin detection," *IEEE Signal Processing Letters*, vol. 24, no. 3, pp. 289–293, 2017.

[25] L. Huang and X. Ren, "Erotic image recognition method of bagging integrated convolutional neural network," in *Proc. of the 2nd Int. Conf. on Computer Science and Application Engineering*, Hohhot, China, pp. 1–7, 2018.

[26] Y. Wang, X. Jin and X. Tan, "Pornographic image recognition by strongly-supervised deep multiple instance learning," in *Proc. IEEE Int. Conf. on Image Processing (ICIP)*, Arizona, USA, pp. 4418–4422, 2016.

[27] X. Jin, Y. Wang and X. Tan, "Pornographic image recognition via weighted multiple instance learning," *IEEE Transactions on Cybernetics*, vol. 49, no. 12, pp. 4412–4420, 2018.

[28] M. Perez, S. Avila, D. Moreira, D. Moraes, V. Testoni *et al.,* "Video pornography detection through deep learning techniques and motion information," *Neurocomputing*, vol. 230, pp. 279–293, 2017.

[29] S. M. Kia, H. Rahmani, R. Mortezaei, M. E. Moghaddam and A. Namazi, "A novel scheme for intelligent recognition of pornographic images," arXiv preprint arXiv: 1402.5792., 2014.

[30] 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.

[31] N. A. Dahoul, H. A. Karim, M. H. L. Abdullah, M. F. A. Fauzi, A. S. B. Wazir *et al.,* "Transfer detection of YOLO to focus CNN's attention on nude regions for adult content detection," *Symmetry*, vol. 13, no. 1, pp. 26, 2020.

[32]  K. F. GI, "A hierarchical neural network capable of visual pattern recognition," *Neural Network*, vol. 1, pp. 90014, 1989.

[33]  Y. LeCun, I. Bottou, Y. Bengio and P. Haffner, "Gradient-based learning applied to document recognition," *Proceedings of the IEEE*, vol. 86, no. 11, pp. 2278–2324, 1998.

[34]  M. Z. Alom, T. M. Taha, C. Yakopcic, S. Westberg and P. Sidike, "A State-of-the-art survey on deep learning theory and architectures," *Electronics*, vol. 8, no. 3, pp. 292, 2019.

[35]  K. He, X. Zhang, S. Ren and J. Sun, "Deep residual learning for image recognition," in *Proc. of the IEEE Conf. on Computer Vision and Pattern Recognition*, NV, USA, pp. 770–778, 2016.

[36]  S. Wu, S. Zhong and Y. Liu, "Deep residual learning for image steganalysis," *Multimedia Tools and Applications*, vol. 77, no. 9, pp. 10437–10453, 2018.

[37]  A. Q. Bhatti, M. Umer, S. H. Adil, M. Ebrahim, D. Nawaz *et al.,* "Explicit content detection system: An approach towards a safe and ethical environment," *Applied Computational Intelligence and Soft Computing*, vol. 2018, pp. 1–13, 2018.

[38]  I. M. A. Agastya, A. Setyanto and D. O. D. Handayani, "Convolutional neural network for pornographic images classification," in *Proc. of the 4th Int. Conf. on Advances in Computing, Communication & Automation (ICACCA)*, Taylor's University Lakeside Campus, Malaysia, pp. 1–5, 2018.