Computer Systems Science & Engineering
DOI: 10.32604/csse.2022.022723
Article

Tech Science Press

Transfer Learning on Deep Neural Networks to Detect Pornography

Saleh Albahli*

Department of Information Technology, College of Computer, Qassim University, Buraydah, Saudi Arabia
*Corresponding Author: Saleh Albahli. Email: salbahli@qu.edu.sa
Received: 17 August 2021; Accepted: 25 November 2021

Abstract: While the internet has a lot of positive impact on society, there are negative components. Accessible to everyone through online platforms, pornography is, inducing psychological and health related issues among people of all ages. While a difficult task, detecting pornography can be the important step in determining the porn and adult content in a video. In this paper, an architecture is proposed which yielded high scores for both training and testing. This dataset was produced from 190 videos, yielding more than 19 h of videos. The main sources for the content were from YouTube, movies, torrent, and websites that hosts both pornographic and non-pornographic contents. The videos were from different ethnicities and skin color which ensures the models can detect any kind of video. A VGG16, Inception V3 and Resnet 50 models were initially trained to detect these pornographic images but failed to achieve a high testing accuracy with accuracies of 0.49, 0.49 and 0.78 respectively. Finally, utilizing transfer learning, a convolutional neural network was designed and yielded an accuracy of 0.98.

Keywords: Pornographic video detection classification; convolutional neural network; InceptionV3; Resnet50; VGG16

1 Introduction

The internet has become available to everyone regardless of their age. As the internet has expanded in the 21st century, so too has the amount of online publicly available pornographic content. As a result, the detection of pornographic content has become very important with the increasing consumption of online media. Due to the increasing consumption of electronic media [1], especially among children, there is a growing interest in such an application. This type of application could also be useful for law enforcement agencies to detect illegal videos. Previous attempts at detecting pornographic content in video consisted of first detecting possible nudity or human skin exposure [2,3], then proceeding to create an appropriate filter depending on the type of nudity or percentage of skin exposure [4]. Although such an approach is intuitive, it is naive to assume that porn content can be detected by nudity and skin exposure alone.

There are different classifiers in the analysis of numerous detections of computer vision and recognition. Each classifier is unique and must be trained on a limited dataset. The performance of each classifying model is entirely dependent upon the training dataset. Regardless of similarity, each classifying model requires different amounts of time to effectively train, there is a variation of certainty in the training cycles of the models. It can be seen that the operations in an overall context, the operations of these classifying models tend to be based on the level and direction of operations determined by the dataset and algorithm that it operates on. Moreover, classifiers are normally used in cases of deep transfer learning. There are certain limitations to accuracy when the normal classifiers were concerned, removed when the gaze shifted on to deep learning. This historically was not present in phases where deep learning was not involved. Due to the potential of these classifiers remaining at their peak, the accuracy of the models could not be improved.

Previous research found that, along with a number of vector models for image identification and classification, deep transfer learning increased the performance of detection in computer vision in a certain problem based on the efforts of a series of nodes applied in deep learning [5]. Further studies illustrate the application of neural networks in the framework [6]. Additionally, when more than one non-linear algorithm is applied in an overall concept the deviation of the route tends to be illustrated and normal optimizers tend to take [7]. After the potential lead is combined into layers of classifiers, they are then divided into nodes. All of this is termed as deep transfer learning. The overall performance of deep transfer learning is evaluated on the basis of error and return. Whereas error needs to be minimized, the return is in fact maximized [8]. The first element from the node is termed as the optimal classifier [9], whereas the next layer increases its predictability by transformation of the data at hand [10]. Though this does not increase the predictability of the model, it does have effect on the time taken to compute it.

Given the early stages of academic research on pornographic content detection, additional research is needed on the concepts of accuracy, computational overhead, and training data. Therefore, this paper proposed four innovative deep learning models which could effectively detect videos that contain nudity. More specifically, the models proposed in this paper are as followed: VGG16, InceptionV3, ResNet50, and a Convolutional Neural Network (CNN). These models were trained and tested to get the most effective model for detecting publicly obtainable pornographic content via online media platforms.

This study proposes a method based on deep CNNs specially designed to fit the problem of detecting pornographic content. Current convolutional neural networks are designed and trained on image datasets that rarely, if ever, have any form of nudity, making it be very hard to use train these models to detect pornographic videos. This research, therefore, is aimed at finding a way to utilize deep learning techniques with deep learning to create and train a model that will automatically detect nudity in videos and take appropriate actions. In this regard, this study's major contributions are as follows:

1. We tested three different pre-trained CNNs to see if they could perform well in detecting pornographic content.
2. We designed a CNN fine-tuned to detect pornographic content that achieved the best accuracy ever recorded.
3. We designed a model that could differentiate between a pornographic and a non-pornographic video.

The rest of the paper is organized into five main sections.: Section 2 highlights previous work on pornographic content detecting algorithms and models. Next, Section 3 describes the methodology used to analyze the models examined in this study. Results and discussion are presented in Section 4. Finally, we conclude our work in Section 5.

2 Related Work

Garcia et al. [11] attempted to design a model that identified pornographic images and videos by utilizing a pixel-based recognition algorithm. Their model was specifically trained to filter images and videos for skin color. The model was trained using 1,239 multimedia images consisting of 986 images and 253 videos collected from the web. The model got a precision of 0.9033 and an accuracy of 0.8023 after final testing.

The model designed by Jansohn et al. [12] trained their model to predict if a video was going to display content considered pornographic in nature. In their design, Janson and colleagues trained the model to characterize skin, visual words, and motion information. Regarding motion analysis, the model inspect the videos by utilizing Periodicity detection (PER), Sliding window periodicity (PERWIN), and Motion holograms (MHIST). After various combinations, the error of the model was reduced from 0.099 to 0.06.

Izzah et al. [13] used support vector machine and naïve bayes algorithms to classify pornographic content on twitter. The model selection was based unigram and bigram features, classification algorithms and k-cross fold validation with F1 score being the mode of evaluation. The highest F1 score was obtained by a model which combined support vector machine, most common words, and a combination of unigram and bigram, which returned an F1 score of 91.14

Chung et al. [14] used high- and low-quality images to design an obscene image detection algorithm. They used Canny Edge to analyze the fine grains of the image to determine whether the image is of high or low quality, then used this to determine whether an image passes their final obscene test. The proposes model had an accuracy of 0.868 which is 0.01 more than the accuracy of the existing method used.

The video pornography detection model proposed by Perez et al. [15] yielded a high accuracy (0.979). Their CNN was created to detect pornography through a blend of information ranging across pictures (static) and motion (dynamic) by incorporating MPEG motion vectors and optical flow.

Mallmann et al. [16] proposed the first architecture to detect pornographic features of private parts in real time when viewing a video. Their CNN model, called Private Parts Censor (PPCensor), did not require end-user device processioning beyond 35 concurrent connections on desktop computers. With 50,000 manually labelled objects forming the dataset used to train the PPCensor model, PPCensor addresses the detection of pornographic content as an object detection problem.

Agastya et al. [17] designed a convolutional neural network for pornographic image detection by modifying the last layer of a VGG-16 model so as to fit the task of detecting pornographic images. The model is trained using NPDI dataset which is a public repository that contains almost 8 h of 400 pornographic and 400 non-pornographic images. The proposed model had an accuracy of 0.938.

Wijaya et al. [18] used skin probability and eigenporn of skin ROIs images to design a pornographic image recognition model, the research used K-nearest neighbor to optimize the accuracy and false rejection rate of the skin probability and fusion descriptor-based recognition system. The best accuracy gotten was 0.8699.

Thaweekote et al. [19] designed an automatic nipple detection system using random forest based on face detection and ideal proportion female. The result of this algorithm shows the high accuracy and reducing the computational time when compares with the existing method. Li et al. [20] used pre-trained convolutional neural networks to detect and depornize images using transfer learning. The research also introduces a depornize algorithm to cover these sensitive and pornographic content. Experimental results demonstrate the efficiency of both algorithms proposed in this paper. Ashan et al. [21] used transfer learning for pornographic content detection using VGG-16, MobilNet, Inception V3, Xception and ResNet 50. All of these models are fine tune to detect pornographic images from NPDI database. The models' results were very high and efficient.

Influenced by the Late Fusion model and the computational complexity theory, Li et al.'s [20] deep learning model was novel by utilizing stream mode to detect pornographic content. In order to design their model, Li and colleagues combined frame sampling, which resulted in a higher accuracy level when compared to the naïve form of deep learning model.

Chen et al. [22] proposed one of the most successful pornographic detection systems to date. Designed to detect pornographic related content on both porn and gambling websites from common wireless routers often used in schools and homes, their decision mechanism yielded measures over 0.99 for accuracy, precision, and F-measure. Their detection system composed of several steps. After using Doc2Vet to discover and analyze textual features, spatial relationships, and visual words in website hypertext markup language source code, the model was trained to classify pornographic content via text and image features. Finally, the decision mechanism's final prediction of pornographic context across the categorical features of text and image was based on a data fused, logistic regression statistical algorithm.

As seen from Tab. 1, there has been a lot of work in field of artificial intelligence with respect to pornographic content and how to automatically detect them. To address these problems, we designed a deep neural network that could automatically detect whether a video is pornographic or non-pornographic.

3 Description and Analysis of Models

3.1 Background

Previous approaches to classification of videos mainly involved multiple stages [23]. First, local visual features from image frames of the videos would be extracted at a pre-determined frequency [24]. Next, video level descriptors would be created for these features. These descriptors would then result in a "*bag of visual words*," which would be classified using a classifier [25].

Biologically inspired CNNs can perform the work of all three stages with a single neural network. Convolutional Neural Networks are typically multi-layered neural networks that can extract topological properties from an image or 2D matrix [26]. CNNs are especially useful as they reduce the number of parameters required with parameter sharing and can identify spatial patterns even if they are shifted or rotated. The equation below defines the operations of a typical CNN.

$$f_{ij}^{xy} = f_n\left(b_{ij} + \sum_{m}\sum_{p=0}^{Pi-1}\sum_{q=0}^{Qi-1} w_{ijm}^{pq} f_{(i-1)m}^{(x+p)(y+q)}\right) \tag{1}$$

In the equation above, the position (x, y) is on the j^{th} filter from the i^{th} layer, filters are indexed by m, and the bias is b. The parameter's value at position (p, q) is w^{pq}. The height and width of the filter are represented by P_i and Q_i, respectively [27]. The f_n is a non-linear function which can introduce some form of non-linearity. Generally, f_n is either a *Rectified Linear Unit* or a *Sigmoid* function. The ability of CNN to capture features independent of its location and orientation is mainly due to pooling operations [28]. Pooling operations captures values in from the neighboring cells and outputs the dominant values within the region [29,30].

Table 1: Comparison table of previous methodologies

Reference	Methodology	Classifier	Findings	Gaps identified
Garcia et al. [11]	Nudity filtering flow chart.	Pornographic image and video filtering application	Precision of 0.9033 Accuracy of 0.8023	While it did well in identifying nudity, sometimes nudity doesn't show a lot of skin. The model doesn't detect audio nudity.

(Continued)

Table 1 (continued)

Reference	Methodology	Classifier	Findings	Gaps identified
Jansohn et al. [12]	Skin detection and a bag of visual words unigram and bigram features, classification algorithms and k-cross fold validation	Motion analysis	Error of 0.06 F1 score of 0.9114	The model works well.
Izzah et al. [13]	Canny Edge	Support vector machine and Naïve bayes	Accuracy of 0.868	Obscene videos cannot be detected
Chung et al. [14]	Using static (picture) and dynamic (motion) information	Canny Edge test	Accuracy of 0.979	Better than previous model as it considers that not all skin exposure is nudity.
Perez et al. [15]	PPCensor	Deep convolutional neural network	Very high accuracy	Misclassification of object that look like nudes.
Mallmann et al. [16]	Convolutional neural network	Convolutional neural network	Accuracy of 0.938	The model performs well.
Agastya et al. [17]	Eigenporn extraction	VGG-16	Accuracy of 0.8699	The model performs well.
Wijaya et al. [18]	An automatic nipple detection system	K-nearest neighbor	High accuracy	The model cannot detect nonsymmetrical nipples.
Thaweekote et al. [19] Villán [31]	Detect and depornize images	Random forest	High accuracy	The model performs well.
Chen et al. [22]	Transfer learning	Pre-trained convolutional neural networks VGG-16, MobilNet, Inception V3, Xception and ResNet 50	Efficient accuracies	The model performs well.

Typically, CNNs are followed by a dense layer and an output that predicts the probability of an particular image belonging to a class. The difference in predicted output and true labels are calculated by a loss function. For the purpose of this study, categorical cross-entropy (log loss) was used as the loss function [32]. The loss function can be described as below:

$$loss = -y \log \hat{y} - (1-y)\log(1-\widehat{y}) \tag{2}$$

In this particular function, y are the true labels (0,1) and y^ are the predicted probability of the outcome belonging to label 1. The parameters in such a network can be optimized with backpropagation algorithm [33] to minimize the above loss function. Currently a lot of modified algorithms are available [34], which are an improvement over the simple backpropagation algorithm. For the purpose of the study, RMSProp [35], Adam optimizer [36], and Stochastic Gradient Descent algorithms were used.

A vast majority of the models discussed above find it hard to get excellent accuracies of pornographic detection. Given the delicate nature of the task of pornographic detection, it requires a carefully selected dataset and a well-tailored start of the art deep learning model. Most pre-trained models are not suitable for this task because they are trained on a dataset that doesn't have any pornographic content, hence they have absolutely no experience when it comes to dealing with such content.

3.2 Models

VGG16: VGG16 is a CNN model proposed by University of Oxford scholars Simonyan et al. [37]. This model was used for recognizing large scale images and outperformed other models at the ImageNet challenge. The architecture of the original VGG16 model is described in Fig. 1.

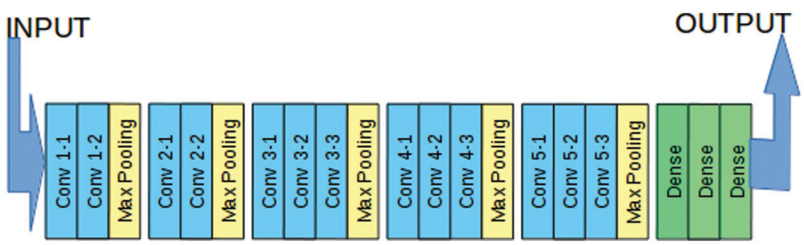

Figure 1: Original VGG16 architecture

The number of CNN layers proposed in original paper [37] has been kept the same, but the number of parameters is different. This is because the original model had an input size 224 × 244 × 3, but the data-set used here has images of size 75 × 75 × 3. Instead of three dense layers at the ending, only one is kept. The number of parameters in each layer is shown in Tab. 2.

Table 2: Parameters in each layer

Layers	Parameters
conv1-1	1792
conv2-1	36928
maxpooling	0
conv2-1	73856
conv2-2	147584
maxpooling	0
conv3-1	295168
conv3-2	590080
conv3-3	590080
maxpooling	0

(Continued)

Table 2 (continued)

Layers	Parameters
conv4-1	1180160
conv4-2	2359808
conv4-3	2359808
maxpooling	0
conv5-1	2359808
conv5-2	2359808
conv5-3	2359808
maxpooling	0

Since transfer learning is implemented, the parameters associated with convolutional layers are pre-loaded with weights from ImageNet training. Only parameters associated with the dense layer is updated via backpropagation. The log loss function is used to calculate loss and Adam optimizer algorithm is used to update parameter values.

InceptionV3: This CNN model was introduced by GoogleNet [38]. It was further modified with the introduction of batch normalization and then factorization in its last version. InceptionV3 outperformed all the models at ILSVRC (ImageNet Large Scale Visual Recognition Competition) 2015. Transfer learning is used with InceptionV3 model. Only the last dense layer of the model is trained. The rest of the model has pre-trained weights. The name inception comes from the fact that there are mini-models within the larger models. The advantage of having multiple models with different filter sizes is that we do not need to vary the filter size hyper-parameter too much. It also allows extraction of local features and larger abstract features via smaller and larger convolutions simultaneously [39]. This model also included batch normalization steps. The model contains a total of 23,115,554 parameters. But only 1,312,770 from the last dense layer were updated while training. The log loss function is used to calculate loss and Adam optimizer algorithm is used to update parameter values.

Resnet50: Regarding the Resnet50 CNN model, resnet stands for residual network. In residual networks, some weights from previous layers are added to a later layer bypassing layer in between. These bypassing layers contain weights influenced by immediately preceding layers as well as from layers from deeper layers [40]. This solved the vanishing gradient problem and training the deep neural network was feasible. It stops the degradation of network during training [41]. This model also involves batch normalization. The total number of parameters in this model was 10,711,938. Transfer learning was used, and hence only the topmost layer was trained. Thus 8,611,712 parameters were updated using Stochastic Gradient Descent algorithm.

Proposed Network: This CNN contains three layers as shown in Fig. 2. Each layer has convolution network followed by a relu non-linear activation function. Then a maxpooling function followed by batch normalixation. After the three layers there are two dense layers before output is generated. Fig. 2 shows the suggested architecture of the proposed model. The log loss function was used to calculate the loss. The model was train using Adam optimization algorithm.

The proposed model used transfer learning involving the reuse of an already trained model with gathered knowledge. This proposed model initially took in inputs with their respective labels and passed them into the first convolutional layer to the next until it got to the last layer where key features were extracted for the classification task. The classification task was passed through two dense layers and evaluated with the log loss matric to test effectiveness. Parameters of the model were transferred into training another model, thereby enhancing its effectiveness by applying transfer learning. This process of transferring parameters was repeated until an optimal model was yielded. This final model was also evaluated using the log loss metric.

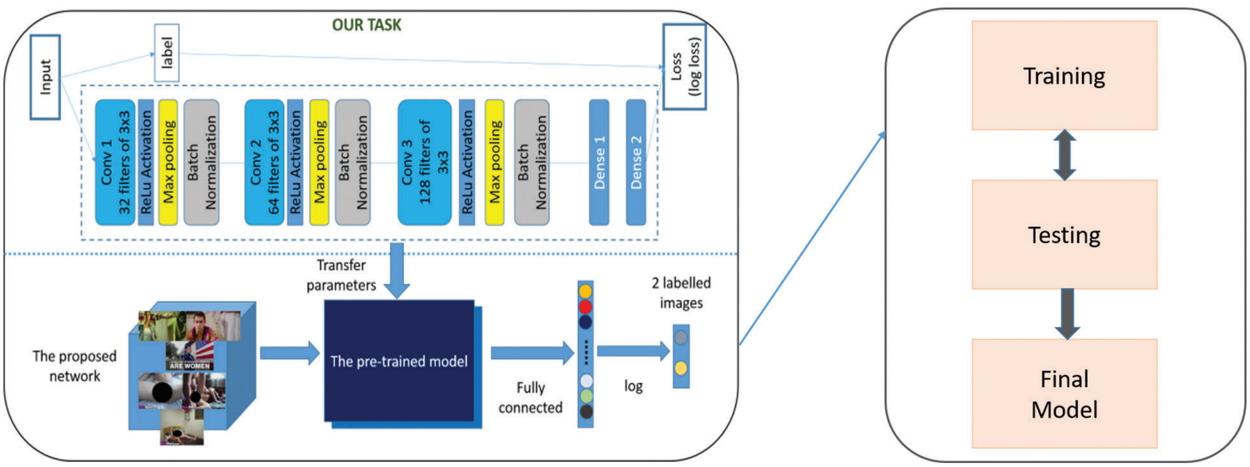

Figure 2: Architecture of the proposed network

4 Results and Discussion

4.1 Dataset

The dataset was produced by accumulating videos from various websites and loaded on Kaggle's dataset repository [42]. The content for the non-pornographic videos comprised of generic human activity and movements in public places in a confined location such as a room and buildings Fig. 3. Pornographic videos were gathered from torrents and RTA labelled websites. This dataset contained almost 190 videos. All videos were trimmed to remove the non-active and out-of-frame parts. The videos in total accumulated to over 5 GB of data.

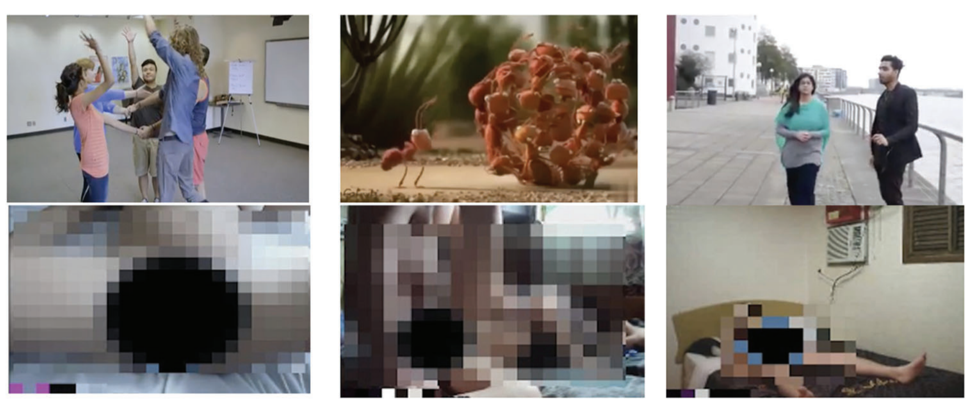

Figure 3: Example frames from the pornography and non-pornographic dataset, illustrating the diversity of the non-pornographic (top row) and pornographic content (bottom row)

4.2 Videos

Processed at a frame-rate of 5, videos generated more than 62,000 images. OpenCV [31] was used for the purpose of processing the videos. The image frames extracted from the video were then uploaded in a Google drive folder, separately as porn and non-porn as seen in Fig. 3-bottom row-, available to the research community and the public at large. Every video was trimmed and only frames which displayed movements involving some activities were retained. These frames were valuable as they were used for training the model. Any frame in a pornographic video that did not contribute to the classification of video as pornographic could cause the model to learn incorrect patterns and may initiate overfitting. All the processing work was carried out on Kaggle platform, with a 13 GB RAM, 20 GB of Memory and TeslaK80 GPUs. As a skewed data set can hamper the ability of the model to learn parameters appropriately, a balanced data set was compiled. After the pre-processing process, a total of 11,310 images for non-porn and 11,310 images for porn were collected. These images were split into training and testing sets in the ratio 80:20, meaning 80% of the total dataset was reserved for training the deep learning models while the remaining 20% was used to test and validate the models.

4.3 Models

4.3.1 VGG16

The VGG16 model was trained for 15 epochs. The loss and accuracy over the epochs are shown in Fig. 4 and 5. After the first epoch, there was no improvement in accuracy and the loss did not decrease any further. When predicting the labels of test data, the accuracy of the model was fairly low (0.499). The confusion matrix of the prediction of VGG16 model is shown in Fig. 6. The precision and recall score of ‘pornographic’ class was 0.5 and 1.0 respectively, while the precision and recall score of ‘non-pornographic’ class was 0.0 for both metrics.

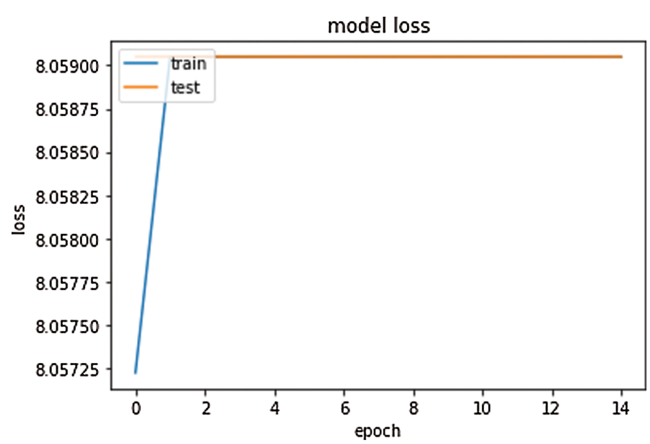

Figure 4: VGG16 model loss for train and test data

4.3.2 Inception-V3

The InceptionV3 model was trained for 15 epochs. The loss and accuracy over the epochs are shown in Figs. 7 and 8. Although there wasn’t significant increase in accuracy of prediction of the test data, the accuracy of train data continuously improved. There is a large difference between the train and test data accuracy. The accuracy of the model was again fairly low (0.499) when it predicted the labels of test data after 15 epochs. The confusion matrix of the prediction of InceptionV3 model is shown in Fig. 9. The precision and recall score of ‘pornographic’ class was 0.5 and 0.93 respectively, while the precision and recall score of ‘non-pornographic’ class was 0.49 and 0.06.

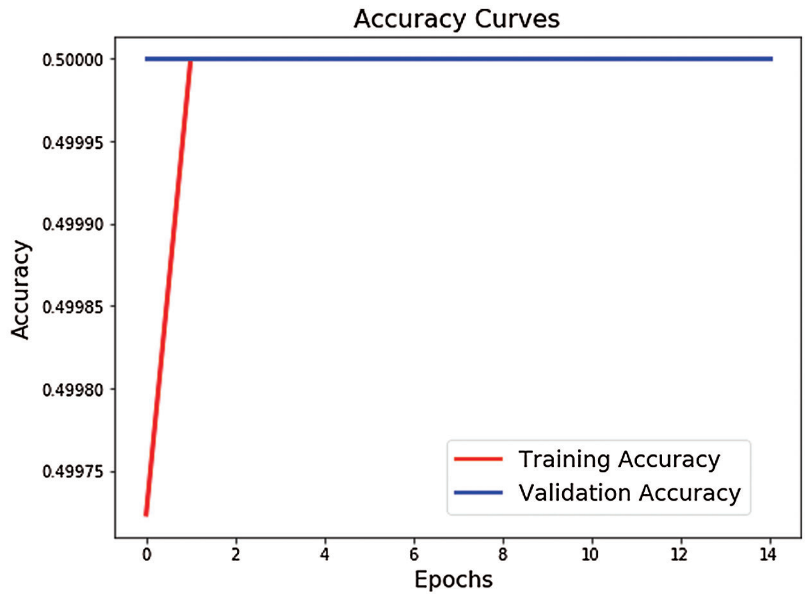

Figure 5: Accuracy of VGG16 model prediction for train and test data

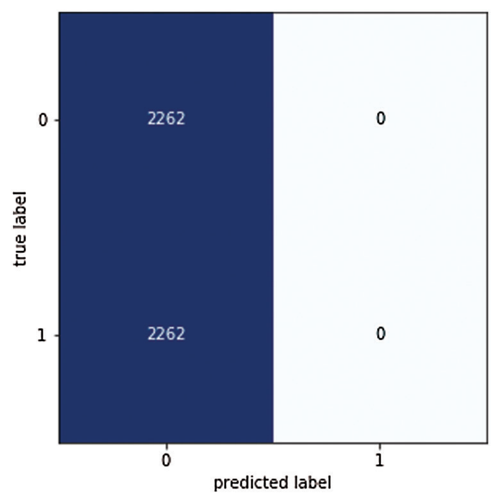

Figure 6: Confusion matrix for prediction by VGG16 model

4.3.3 Resnet50

The ResNet50 model was also trained for 15 epochs. The loss and accuracy over the epochs are shown in Figs. 10 and 11. Although accuracy score and loss improved as the epochs increased, there was a large difference between the train and test data loss. There was also a wide contrast regarding accuracy score as well. The accuracy of the model when it predicted the labels of test data after 15 epochs was 0.78, which is an improvement over the previous models. The confusion matrix of the prediction of ResNet50 model is shown in Fig. 12. The precision and recall score of 'pornographic' class was 0.98 and 0.58 respectively, while that of 'non-pornographic' class was 0.70 (precision) and 0.99 (recall). Except for recall score of 'non-pornographic' class, the other three metrics improved over the last two models.

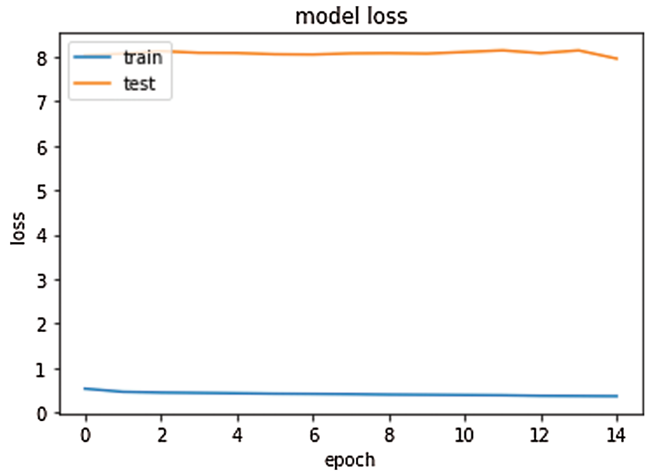

Figure 7: InceptionV3 model loss for train and test data

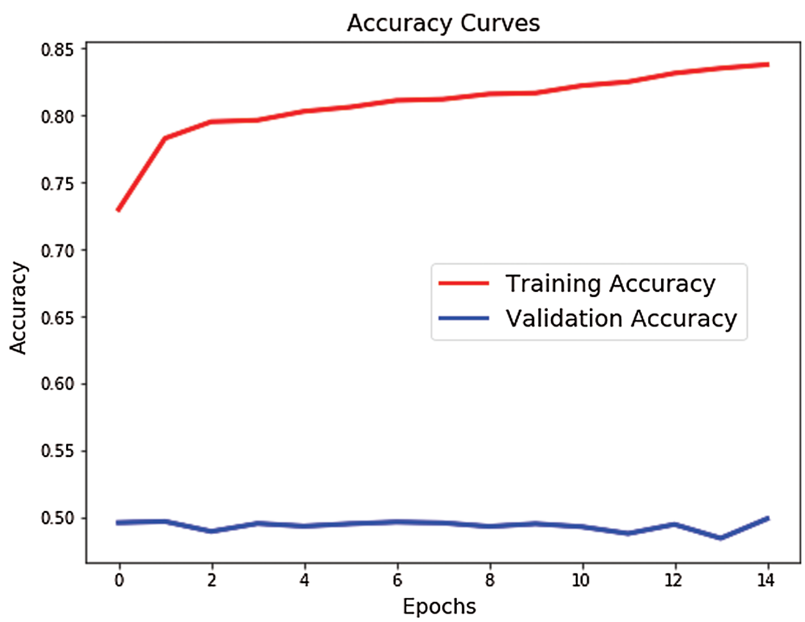

Figure 8: Accuracy of inceptionV3 model prediction for train and test data

4.3.4 Proposed Model

The loss and accuracy over the epochs for the proposed model are shown in Figs. 13 and 14. As described by Tab. 3, the loss decreased and the accuracy increased for both test and train data. The difference between accuracy of test and train data was also lesser than previous InceptionV3 and ResNet50 models. The accuracy of the model when it predicted the labels of test data after training was 0.98, which was a significant improvement over the previous models. The confusion matrix of the prediction of proposed model is shown in Fig. 15.

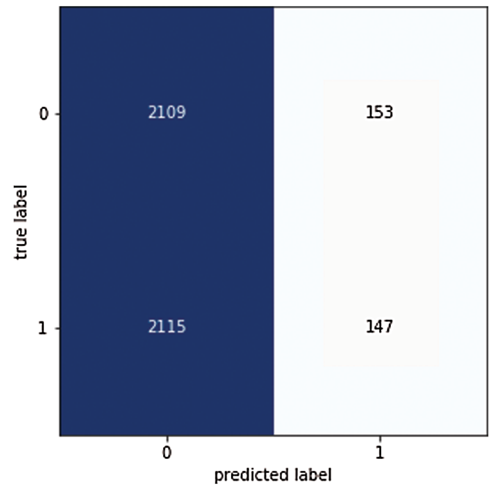

Figure 9: Confusion matrix for prediction by InceptionV3 model

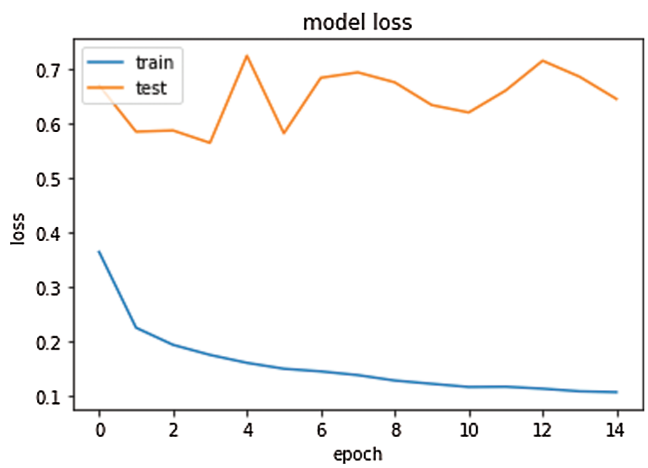

Figure 10: ResNet50 model loss for train and test data

The precision and recall score of 'pornographic' class was 0.99 and 0.96 respectively, while that of 'non-pornographic' class was 0.96 (precision) and 0.99 (recall). As shown in Tab. 3, the proposed got a better score for every measured metric .. 3. All the steps in the above analysis are presented and supported with graphs to show how well the various models used for this research performed. This performance is summarized in Tab. 3, which shows the training time, test loss, test accuracy, precision, recall, F1 score, and support of each of the trained models. The proposed model had an accuracy of 0.98, which is 0.20 more than the best accuracy yielded by the other trained models. This shows that the proposed CNN can effectively detect whether a video has pornographic content or not.

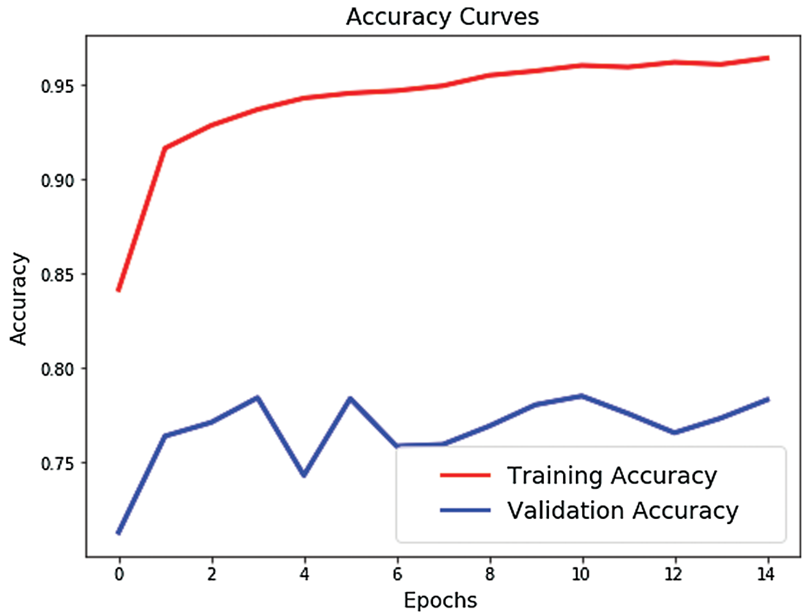

Figure 11: Accuracy of ResNet50 model prediction for train and test data

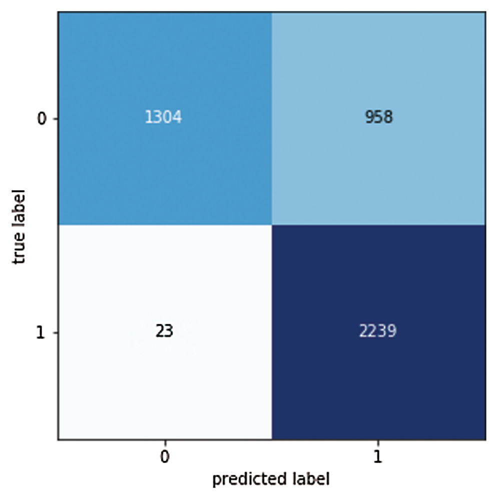

Figure 12: Confusion matrix for prediction by ResNet50 model

Tab. 3 shows the proposed CNN model results compared to the other three pre-trained models. As noted earlier, pre-trained models are generally trained on a wide variety of dataset which rarely involves pornographic content, hence, the models have absolutely no experience as regards pornographic content. On the other hand, the proposed CNN model in this study was specifically, designed, tailored, and trained to solve the problem of pornographic content detection.

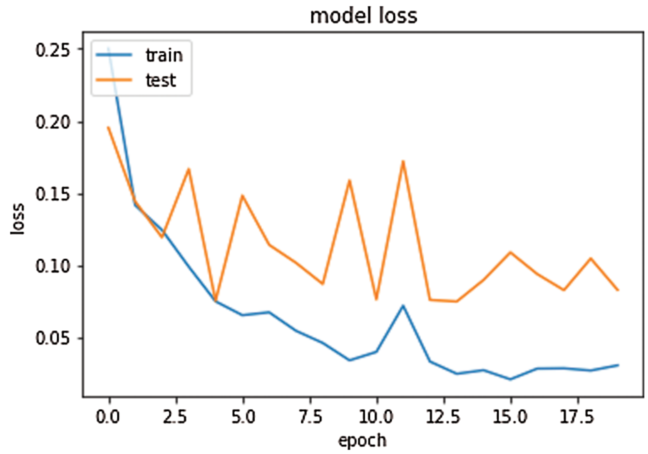

Figure 13: Proposed model loss for train and test data

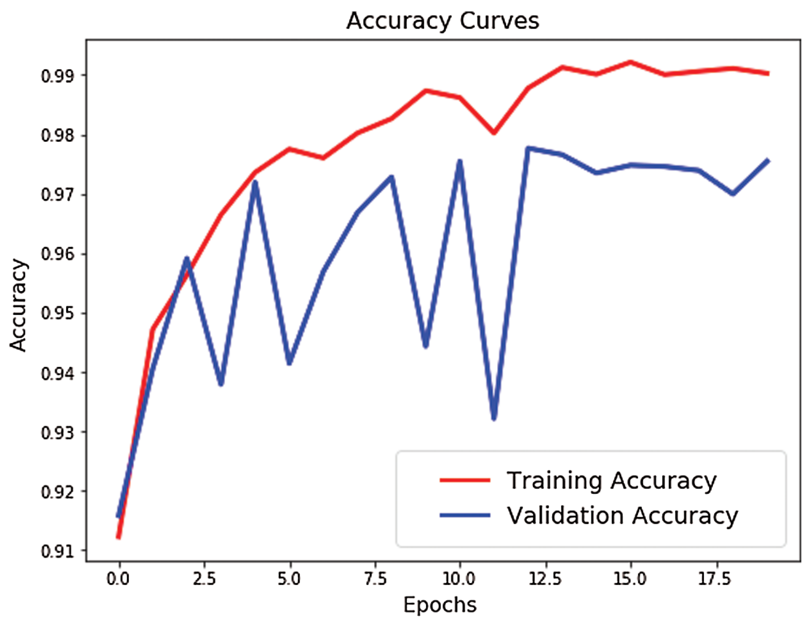

Figure 14: Accuracy of proposed model prediction for train and test data

Table 3: Comparison of models training and testing accuracy with metric evaluation

Model	Training time	Test loss	Test accuracy	Precision	Recall	F1 score	Support
VGG16	414	8.05	0.499	0.50	1	0.62	2262
InceptionV3	247	7.96	0.498	0.50	0.93	0.65	2262
Resnet50	315	0.643	0.789	0.98	0.58	0.73	2262
Proposed CNN	101	0.082	0.975	0.99	0.96	0.98	2306

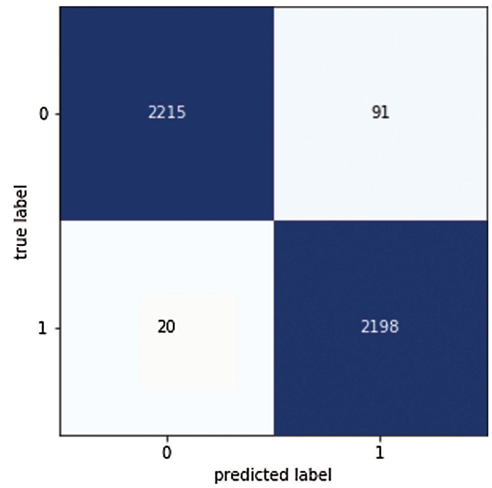

Figure 15: Confusion matrix for prediction by proposed model

5 Conclusion

Pornographic content in videos has always been a delicate issue, especially regarding the content being viewed and accessed by minors. Several research endeavors have attempted to automatically detect pornographic content and taking appropriate actions. The CNN model analyzed in this study outperforms all previous models in the aspects of training time, test loss, test accuracy, precision, recall, F1 score, and support. The testing was performed on 18,096 frames and tested on 4,524 frames for three state-of-the-art models initially trained and evaluated with the highest accuracy being 0.78. Then proposed CNN model was designed, trained, and fine-tuned to detect pornographic videos specifically. Indicated by a high accuracy evaluation (0.98), this model performed well. As a result, this proposed model can effectively detect whether a video has pornographic content in it. With an accuracy of 98%, if properly implemented and integrated on web browsers and media players, this algorithm would eradicate the problem of illegal content reaching the wrong audience. For example, parents could have peace of mind as to what their offspring interact with without supervision.

It is important to note that the concept of pornography can vary across different cultures, nationalities, and countries. As a result, is important for future researchers to design models that can detect particular kinds of pornographic content in either a picture or a video.

Acknowledgement: The researchers would like to thank the Deanship of Scientific Research, Qassim University for funding the publication of this project.

Funding Statement: The authors received no specific funding for this study.

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

References

[1] J. da Silva, "Children and electronic media: How much is too much?," in *The American Psychology Association*, 2015. http://www.apa.org/pi/about/newsletter/2015/06/electronic-media.aspx.

[2] H. Zheng, M. Daoudi and B. Jedynak, "Blocking adult images based on statistical skin detection," *Electronic Letters on Computer Vision and Image Analysis (ELCVIA)*, vol. 4.2, pp. 1–14, 2004.

[3] H. Zuo, W. Hu and O. Wu, "Patch-based skin color detection and its application to pornography image filtering," in *Proc. of the 19th Int. Conf. on World Wide Web, WWW '10, ACM*, New York, NY, USA, pp. 1227–1228, 2010.

[4] H. Bouirouga, S. El Fkihi, A. Jilbab and D. Aboutajdine, "Skin detection in pornographic videos using threshold technique," *Journal of Theoretical and Applied Information Technology*, vol. 35, no. 1, pp. 7–19, 2012.

[5] Z. Liu, Q. Pan, J. Dezert, J. W. Han and Y. He, "Classifier fusion with contextual reliability evaluation," *IEEE Transactions on Cybernetics*, vol. 48, no. 5, pp. 1605–1613, 2017.

[6] S. Albahli, N. Ayub and M. Shiraz, "Coronavirus disease (COVID-19) detection using X-ray images and enhanced DenseNet," *Applied Soft Computing*, vol. 110, no. 6, pp. 107645, 2021.

[7] Di. Miao, M. Zhang, M. Sun and Z. Tan, "Bin-based classifier fusion of iris and face biometrics," *Neurocomputing*, vol. 224, no. 4, pp. 105–118, 2017.

[8] K. Ludmila, "A theoretical study on six classifier fusion strategies," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 24, no. 2, pp. 281–286, 2002.

[9] R. Dymitr and G. Bogdan, "An overview of classifier fusion methods," *Computing and Information Systems*, vol. 7, no. 1, pp. 1–10, 2020.

[10] K. LudmilaI, W. Christopher, S. Catherine and D. Robert, "Limits on the majority vote accuracy in classifier fusion," *Pattern Analysis and Applications*, vol. 6, no. 1, pp. 22–31, 2017.

[11] M. B. Garcia, T. F. Revano, B. G. M. Habal and J. O. Contreras, "A pornographic image and video filtering application using optimized nudity recognition and detection algorithm," in *The 10th Int. Conf. on Humanoid, Nanotechnology, Information Technology, Communication and Control, Environment and Management (HNICEM)*, Baguio City, Philippines, pp. 1–5, 2018.

[12] C. Jansohn, A. Ulges and T. M. Breuel, "Detecting pornographic video content by combining image features with motion information," in *Proc. of the 17th ACM Int. Conf. on Multimedia (MM '09)*, New York, NY, USA, Association for Computing Machinery, pp. 601–604, 2009.

[13] N. Izzah, I. Budi and S. Louvan, "Classification of pornographic content on Twitter using support vector machine and Naive Bayes," in *The 4th Int. Conf. on Computer and Technology Applications (ICCTA)*, Istanbul, pp. 156–160, 2018.

[14] M. Chung, J. Ko and D. Jang, "Obscene image detection algorithm using high-and low-quality images," in *The 4th Int. Conf. on New Trends in Information Science and Service Science*, Gyeongju, pp. 522–527, 2019.

[15] 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, no. 1, pp. 279–293, 2017.

[16] K. Mallmann, A. O. Santin, E. K. Viegas, R. Robson, J. Geremias *et al.,* "Architecture for real- time pornography detection in video streaming," *Future Generation Computer Systems*, vol. 112, no. 5, pp. 945–955, 2020.

[17] I. M. Agastya, A. Setyanto and D. O. D. Handayani, "Convolutional neural network for pornographic images classification," in *The Fourth Int. Conf. on Advances in Computing, Communication Automation (ICACCA)*, Subang Jaya, Malaysia, pp. 1–5, 2018.

[18] I. G. Wijaya, I. B. Widiartha and S. E. Arjarwani, "Pornographic image recognition based on skin probability and eigenporn of skin ROIs images," *Telkomnika*, vol. 13, no. 3, pp. 985–969, 2017.

[19] V. Thaweekote, P. Songram and C. Jareanpon, "Automatic nipple detection based on face detection and ideal proportion female using random forest," in *The IEEE Int. Conf. on Computational Intelligence and Cybernetics (CYBERNETICSCOM)*, Yogyakarta, pp. 11–15, 2013.

[20] H. Li, X. Guo, H. Wang, B. Yu and P. Zhao, "Application of computational complexity theory to pornographic detection in live based on frame sampling," in *IEEE 6th International Conference on Computer and Communications (ICCC)*, Chengdu, China, pp. 1334–1338, 2020.

[21] B. Ashan, H. Cho and Q. Liu, "Performance evaluation of transfer learning for pornographic detection," *Advances in Natural Computation*, vol. 1075, no. 3, pp. 143–149, 2020. https://doi.org/10.1109/ICCC51575.2020.9345284.

[22] Y. Chen, R. Zheng, A. Zhou, S. Liao and L. Liu, "Automatic detection of pornographic and gambling websites based on visual and textual content using a decision mechanism," *Sensors*, vol. 20, no. 1, pp. 95–109, 2020.

[23] H. Wang, M. M. Ullah, A. Klaser, I. Laptev and C. Schmid, "Evaluation of local spatio-temporal features for action recognition," in *Bmvc 2009-British Machine Vision Conf.*, London, UK, BMVA Press, pp. 124–133, 2009.

[24] H. Wang and C. Schmid, "Action recognition with improved trajectories," in *Proc. of The IEEE Int. Conf. on Computer Vision*, Sydney, Australia, pp. 3551–3558, 2013.

[25] S. Avila, N. Thome, M. Cord, E. Valle and A. D. A. AraúJo, "Pooling in image representation: The visual codeword point of view," *Computer Vision and Image Understanding*, vol. 117, no. 5, pp. 453–465, 2013.

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

[27] J. Wehrmann, G. Simes, R. Barros and V. Cavalcante, "Adult content detection in videos with convolutional and recurrent neural networks," *Elsevier Science Publishers B. V.*, vol. 272, no. C, pp. 432–438, 2018.

[28] D. Scherer, A. Müller and S. Behnke, "Evaluation of pooling operations in convolutional architectures for object recognition," in *Artificial Neural Networks–ICANN*, Springer, Thessaloniki, Greece, pp. 92–101, 2010.

[29] C. Lee, P. W. Gallagher and Z. Tu, "Generalizing pooling functions in convolutional neural networks: Mixed, gated, and tree," in *Artificial Intelligence and Statistics (PMLR)*, pp. 464–472, 2016.

[30] S. Albahli, F. Alhassan, W. Albattah and R. U. Khan, "Handwritten digit recognition: Hyperparameters-based analysis in Applied Sciences," *Applied Sciences*, vol. 10, no. 17, pp. 5988, 2020.

[31] A. F. Villán, *Mastering OpenCV 4 with Python: A Practical Guide Covering Topics from Image Processing, Augmented Reality to Deep Learning with OpenCV 4 and Python 3.7*. 1st ed., vol. 1. The Packt Publishing Ltd., pp. 1–532, 2019.

[32] K. P. Murphy, *Machine Learning: A Probabilistic Perspective*. 1st ed., vol. 1. Boston, USA: The MIT press, 2012.

[33] I. Goodfellow, Y. Bengio and A. Courville, "Back-propagation and other differentiation algorithms," *Deep Learning*, vol. 23, no. 4, pp. 200–220, 2016.

[34] S. Ruder, "An overview of gradient descent optimization algorithms," arXiv preprint arXiv: 1609.04747, 2017.

[35] T. Tieleman and G. Hinton, "Lecture 6.5-rmsprop: Divide the gradient by a running average of its recent magnitude," *COURSERA: Neural Networks for Machine Learning*, vol. 4, no. 2, pp. 26–31, 2012.

[36] D. Kingma and J. Ba, "Adam: A method for stochastic optimization," in *Int. Conf. on Learning Representations (ICLR)*, Banff, Canada, pp. 123–132, 2014.

[37] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," arXiv preprint arXiv: 1409.1556, 2014.

[38] C. Szegedy, V. Vanhoucke, S. Ioffe, J. Shlen and Z. Wojna, "Rethinking the inception architecture for computer vision," in *Proc. of The IEEE Conf. on Computer Vision and Pattern Recognition (CVPR)*, Las Vegas, NV, USA, pp. 2818–2826, 2016.

[39] J. Bankar and N. Gavai, "Convolutional neural network based inception v3 model for animal classification," *International Journal of Advanced Research in Computer and Communication Engineering, India*, vol. 7, no. 5, 2018.

[40] N. Alrobah and S. Albahli, "A hybrid deep model for recognizing arabic handwritten characters," *IEEE Access*, vol. 9, pp. 87058–87069, 2021.

[41] S. Albahli, M. Shiraz and N. Ayub, "Electricity price forecasting for cloud computing using an enhanced machine learning model," *IEEE Access*, vol. 2, no. 8, pp. 71–81, 2020.

[42] S. N. Rahul, "Profane video dataset," 2020. [Online]. Available: https://www.kaggle.com/dataset/e580519d04c54280abe560be09b71e3ac2133663d00578d893f5144ec6da9db8.