Fraudulent Face Image Detection

Rahul Awhad[1,*], *Saurabh* Jayswal[1,**], *Adesh* More[1,***], and *Jyoti* Kundale[1,****]

[1]Ramrao Adik Institute of Technology, Dept. of Information Technology, Mumbai, India

Abstract. Due to the growing advancements in technology, many software applications are being developed to modify and edit images. Such software can be used to alter images. Nowadays, an altered image is so realistic that it becomes too difficult for a person to identify whether the image is fake or real. Such software applications can be used to alter the image of a person's face also. So, it becomes very difficult to identify whether the image of the face is real or not. Our proposed system is used to identify whether the image of a face is fake or real. The proposed system makes use of machine learning. The system makes use of a convolution neural network and support vector classifier. Both these machine learning models are trained using real as well as fake images. Both these trained models will take an image as an input and will determine whether the image is fake or real.

1 Introduction

Many available software applications in the market can be used to modify and edit images. Some advanced software is capable of making an altered image which can seem to be realistic. Hence such alterations affect the authenticity and originality of the image and it becomes very difficult for a person to identify whether an image is altered or not. Such altered images can be of a person's face also. Altered images of the face can be used to create fake profile images on social media platforms and can be used to deceive other people. There are now emerging new technologies for solving such problems. Such technology is machine learning. Machine learning is a technique used to train a system and make a prediction for a problem-based on the trained system. Our proposed system is an attempt to check the originality of the image of a face. Our proposed system will help in identifying whether the face in the image is spliced or not.

2 Review of Literature

Fake Colorized Image Detection: Guo, Cao, Zhang, and Wang in the year 2018 have proposed that there are many techniques used for coloring gray-scale images. In this paper, they have used histogram-based fake colorized detection and encoding based fake colorized image detection techniques to identify whether the image is a colorized image or not. The histogram-based detection uses the hue and saturation channel of an image. The histogram for the hue channel of fake images is generally smooth while the histogram for the saturation channel for fake images shows different peak values. Their proposed system observes the statistical inconsistencies in the bright and dark channels of the image [1].

Detection of GAN-generated Fake Images Over Social Networks: Marra, Gragnaniello, Cozzolino, and Verdoliva in the year 2018 have proposed that generative adversarial network allows one to modify the image in a very realistic way. In this paper, they have studied the performance of many forgery detectors. The detection accuracy can be achieved by up to 95% by the use of conventional and deep learning detectors [2].

Fake Colorized Image Detection with Channel-wise Convolution based Deep-learning Framework: Zhuo, Tan, Zeng, and Lit in the year 2018 have proposed a system to identify a fake colorized image. They have introduced a Wider separate-then-reunion network, a deep learning-based data-driven color image steganalyzer to detect fake colorized images [3].

Image Splicing Detection Through Illumination Inconsistencies and Deep Learning: Pomari, Ruppert, Rezende, Rocha, and Carvalho in the year 2018 have proposed a system for detecting spliced images. They have made use of illuminant maps and deep neural network for training data to find hints of forgery. The system makes use of the deep neural network as well as the transfer learning process. The transfer learning process uses an already trained machine learning model for training the data for the required application. The proposed system eliminates the tedious engineering process and provides an accuracy of more than 96% [4].

Digital image tampering detection and localization using singular value decomposition technique: Mall, Roy, and Mitra in the year 2013 have proposed a way to identify the integrity of the digital images. The proposed

*e-mail: rahul.awhad28@gmail.com
**e-mail: srsjayswal@gmail.com
***e-mail: adeshmore98@gmail.com
****e-mail: jyoti.jadhav@rait.ac.in

system is based on the generation of a hash of the image using singular value decomposition. The generated hash value will be affected if any tampering is done on the image. Such a hash vector will help in localizing image tampering [5].

Fake Face Detection Based on Radiometric Distributions: Edmunds and Caplier in the year 2016 have proposed a method for fake face detection using radiometric distortions. In this paper, the modeling of radiometric distortions arising in the recapturing process is done to solve the crisis of fake image detection. In this technique, the detection process occurs after the face identification process. Based on twelve features, differentiation between real and fake faces is done [6].

Learning to detect fake face images in the wild: Hsu, Lee, and Zhuang in the year 2018 have proposed a system to detect computer-generated images. In this paper, a deep forgery discriminator is developed to detect computer-generated images. The system is trained using fake images with the help of a deep neural network. Fake images generated by (GAN) Generative Adversarial Network are successfully detected. This was the first work to solve problems of detecting fake images. The system successfully detected 94.7% of the fake images created by GAN techniques [7].

Detecting Both Machine and Human Created Fake Face Images In the Wild: Tariq, Lee, Kim, Shin, and Woo in the year 2018 has proposed a system to detect fake face images generated by both machine and human effort. Machine created images are GAN-created images. Human created fake images are images which are edited in software like Adobe Photoshop. For detecting GAN-created images, they have created a neural network model by training real and fake images of varying sizes from 64X64 to 1024X1024 pixel size images. For detecting human-created fake face images, they distinguish real ad fake images by using RGB channel information [8].

Image forgery detection based on physics and pixels: Kumar and Srivastava in the year 2017 have made a detailed discussion about various passive forgery detection techniques. The techniques discussed are copy-move based forgery detection, resampling-based forgery detection techniques, forgery detection based on splicing, and physics/lighting-based forgery detection techniques [9].

Classifying Genuine Face images from Disguised Face Images: Kim, Han, and Woo in the year 2019 have proposed a system for classifying fake face images. Their proposed system makes use of face detection classifiers like ShallowNet, VGG-16 and, Xception. The Xception model from the proposed system is giving an accuracy of 62% [10].

3 Background Study

3.1 Convolution Neural Network

Convolution Neural Network is a category of neural networks. It is mostly used for image recognition or image classification. The first layer in CNN is the convolution layer which is responsible for feature extraction. Then pooling is performed to reduce the dimensional size. The output in the vector form is then fed into the fully connected layer. The output of a fully connected layer classifies the image.

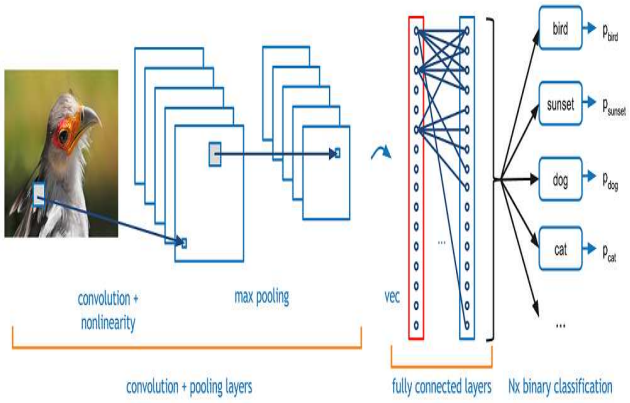

Figure 1. Working of Convolution Neural Network

3.2 Support Vector Classifier

Support Vector Classifier is a technique used in Support Vector Machine to classify the training data as per the labels. This technique is used in supervised learning. SVC algorithm classifies given training data with the help of a hyper-plane. Hyper-plane is a line that divides a plane into two parts thus separating data into different classes.

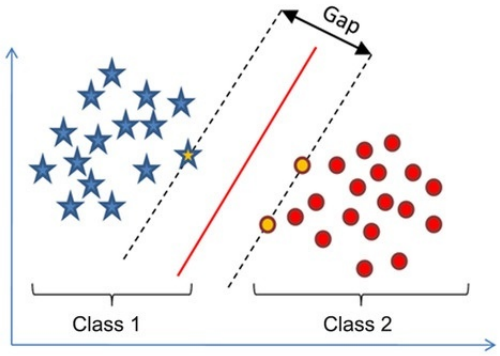

Figure 2. Classification Example

4 Proposed Method

The proposed system utilizes the concept of machine learning. There are various machine learning models

available for different applications. The proposed system makes use of a convolution neural network and support vector classifier. Both models give their results for the input image. The proposed system makes use of a convolution neural network and supports vector classifier models. The working of our proposed system is as follows:

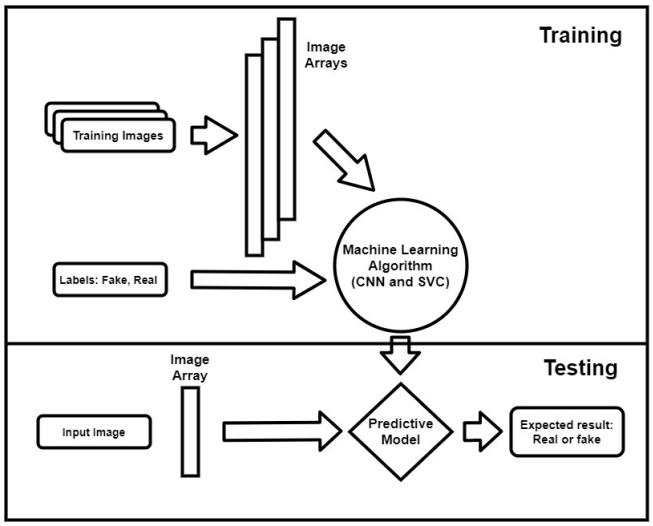

Figure 3. Block Diagram of the Proposed System

- The first step in the system involves collecting the training images and converting then into image arrays.
- Each fake or real image is labeled accordingly.
- The next step involves feeding the training data-set and the labels to the machine learning algorithm to form a model.
- After the model is created, an input image that is to be tested is taken and it is converted into an image array.
- The image array of the input image is then fed into the machine learning model to do the prediction.
- The model gives a label as a result of the prediction which is real or fake.

4.1 Data Collection

The data consists of images of different faces. Real as well as fake images are collected for training. The data can be obtained by manually altering images or by obtaining through online available sources.

4.2 Data Preparation

The images are resized to a fixed dimension. The images are resized to 50X50 pixel size to decrease the overall size and space required to train the data. Images are then stored in an array format for data training. After the model is created, the input image for testing is also resized to 50X50 pixel size.

4.3 Choosing a Model

The proposed system makes use of a convolution neural network and an SVC-based machine learning model. CNN model is generally used for image recognition and processing pixel data. The proposed neural network uses two activation functions. The two activation functions used are the ReLu and the Sigmoid activation function. The SVC classifier classifies the data into two classes. The two classes are "real" and "fake".

4.4 Sigmoid Activation Function

The curve of the Sigmoid function appears as an S-shape. The main reason we use sigmoid is that its value lies between 0 and 1.

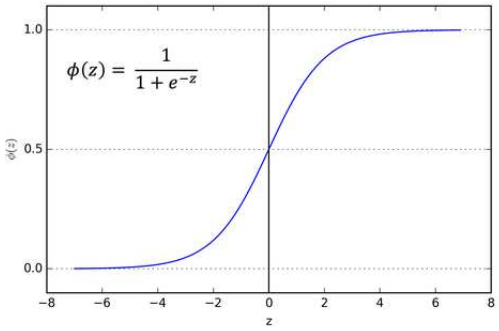

Figure 4. Sigmoid Graph

The curve is given by the following formula:

$$f(x) = \frac{L}{1 + e^{-k(x-x_0)}} \tag{1}$$

where:
L = Curve's maximum value
k = Steepness of the curve
x_0 = x value of sigmoid's midpoint

It is therefore often used in models where we have to estimate the likelihood of an output. Since the probability of the output or prediction lies only between the range of 0 and 1, the sigmoid activation function is preferred for classification.

4.5 ReLU Activation Function

ReLU stands for the Rectified Linear Unit. It is the most used activation function in many machine learning applications. Hence it is almost used in all convolution neural networks and deep learning. The ReLU function is halfway rectified from the bottom. The value of a function is zero when z is less than 0 and the value of a function is z if the value of z is above or equal to zero. The equation for the ReLu function can be given as:

$$f(z) = max(0, z) \tag{2}$$

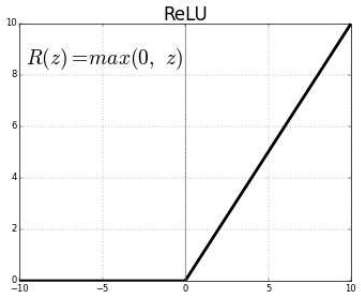

Figure 5. ReLU Graph

4.6 Training Model and Predicting

The convolution neural network is trained using the collected images using the above activation functions. The proposed system's code is written in python language. For training the data we used python's "sklearn" and "tensorflow" module. After the model is trained, it can be used to test the input images. The model will predict whether the input image is fake or real.

4.7 Training and Working of the CNN-Based Model

The training data consist of real and fake images. All the images are converted into gray-scale images. All the images are then resized into 50X50 pixel size images. This will reduce the size and space of data for training purposes. All the images are shuffled before giving input for the training. The machine learning model is created using the python module called "tensorflow".

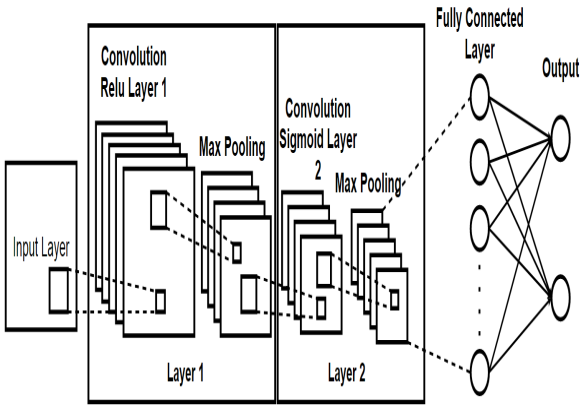

Figure 6. Training of CNN model

For the proposed system, the neural network model used is called the Sequential model. This model has two main layers. There are two types of activation functions used which are "ReLU" and "Sigmoid". Each layer consists of a finite number of neurons. These neurons hold different values after the training process. When the test image is input to the model, the model identifies the pattern for a real or a fake image using the values stored in the neural network model. The last layer will give the value which is used to determine the result.

4.8 Training and Working of the SVC-Based Model

SVC machine learning model is created using python's "sklearn" module. This model is also trained using 50X50 pixel-sized and gray-scale images. This model creates a graph of points and one or more hyper-plane that classifies these points in the graph. This model can classify the points linearly or non-linearly which depends on the application. For the proposed system, classification is done non-linearly. We can specify the way in which we want to classify the data in the kernel parameter. We have used the "rbf" type of non-linear classification which is specified in the kernel parameter. After creating the model, the input image is classified as per the graph formed from the trained data.

5 Results

The proposed system is trained using 2041 images out of which 960 are fake images and 1081 are real images. Figure 7 and figure 8 display the result of the prediction done by the SVC and neural network model for the real and fake image. We have tested the proposed system for 409 images. The SVC-Based model is giving an accuracy of 65% and the CNN-Based model is giving an accuracy of 94%. The 409 tested images include both real and fake images. Figure 6 represents opening an image from the graphical

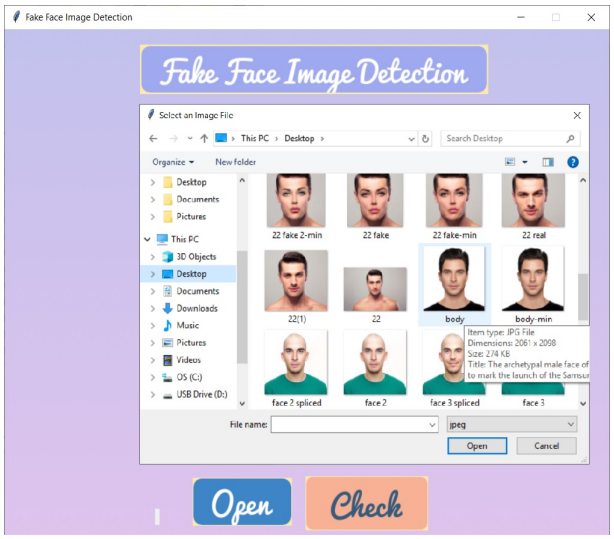

Figure 7. Graphical user interface of the proposed system

user interface of the system. The 'Open' button will open a window from where you can select an image from your computer to be checked. Figure 7 displays the result of the proposed system for a real image. The 'Check' button will run the script for checking the image. Figure 8 displays the process of splicing an image. Here, the term splicing

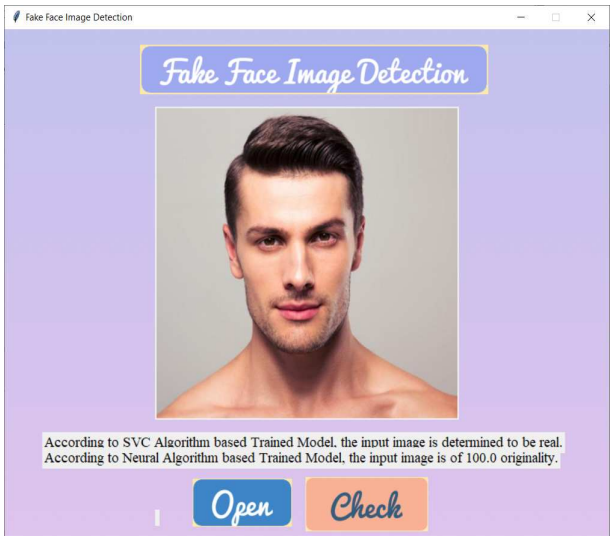

Figure 8. Result of a real image

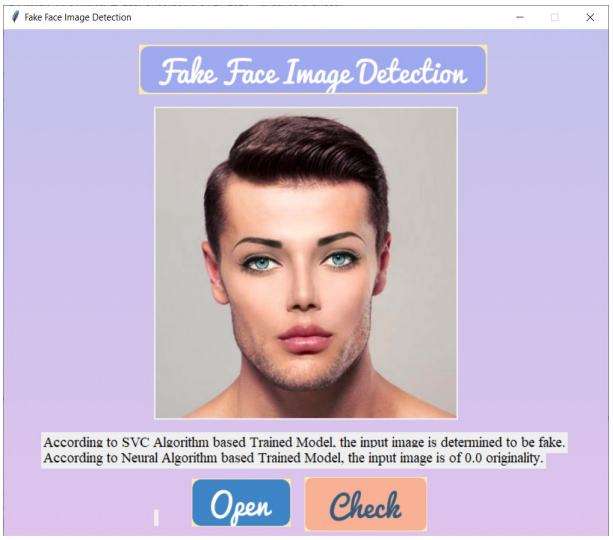

Figure 10. Result of a fake image

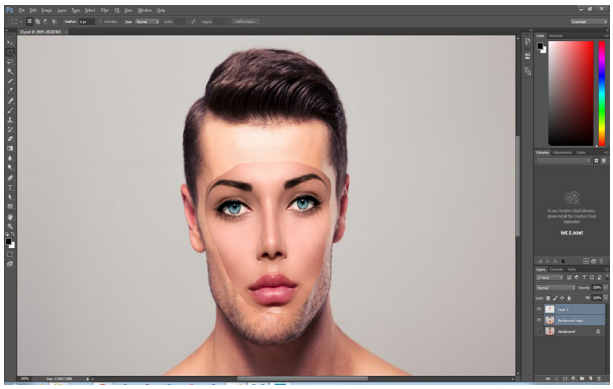

Figure 9. Process of splicing an image

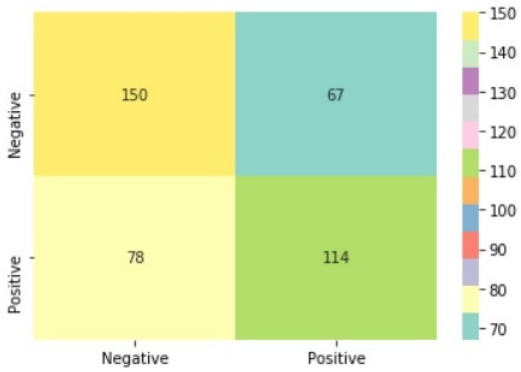

Figure 11. Confusion matrix for the SVC-based model

	precision	recall	f1-score	support
0	0.66	0.69	0.67	217
1	0.63	0.59	0.61	192
accuracy			0.65	409
macro avg	0.64	0.64	0.64	409
weighted avg	0.64	0.65	0.64	409

Figure 12. Classification report for the SVC-based model

means taking a part of the image from any other resource and replacing it on to the target image. Figure 9 displays the result of a fake image that is given as input to the system. The confusion matrix for the SVC-based model is shown in Figure 11. The classification report for the SVC-based algorithm is shown in Figure 12. The confusion matrix for the Neural Network based model is shown in Figure 13. The classification report for the Neural Network algorithm is shown in Figure 14.

6 Comparison with a Existing System

Our proposed system is similar to Pomari et.al's [4] method of determining inconsistencies in an image. Pomari et.al's [4] system makes use of a deep neural network and a transfer learning process whereas our proposed system makes use of a convolutional neural network. Pomari et.al's [4] system utilizes the ReLu activation function while our proposed system makes use of ReLu as well as Sigmoid activation function. Pomari et.al's [4] system trains the data using the transfer learning process which involves an already trained model for a different application. Our proposed system trains the data from scratch without the help of a pre-built machine learning model. Pomari et.al's [4] system finds inconsistencies in all types of images whereas our proposed system focuses on the images consisting of a single face of a person. Pomari et.al's [4] system gives a classification accuracy of 96%. Our proposed system has an accuracy of 65% for the SVC-based model and accuracy of 94% for the CNN-based model.

7 Conclusion

The proposed system determines whether the image of a face is real or fake. The fake image can be a spliced image in which some part of a face is replaced with a part from another face. The proposed system makes use of a convo-

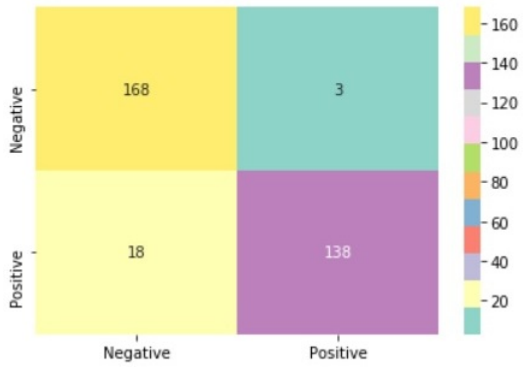

Figure 13. Confusion matrix for the CNN-based model

	precision	recall	f1-score	support
0	0.90	0.98	0.94	171
1	0.98	0.88	0.93	156
accuracy			0.94	327
macro avg	0.94	0.93	0.94	327
weighted avg	0.94	0.94	0.94	327

Figure 14. Classification report for the CNN-based model

lution neural network and support vector classifier. Both these algorithms extract the features from training images and forms two models. One model is based on CNN and the other is based on the support vector classifier. Both of these models predict their own results. The proposed system displays result from both the models. The currently proposed system has an accuracy of 65% for the SVC-based model and 94% for the CNN-based model used in the proposed system. Hence, the proposed system can play an important role in identifying any altered face image which can help people from getting misguided on social media platforms.

8 Future Scope

The proposed system can be included with a face recognition algorithm to identify the presence of any face before checking the image. The accuracy of the system can be increased by providing more data-set of real and fake images. The accuracy can also be increased by altering the parameters used for training the data. The parameters include the number of layers of neurons, activation functions, etc. The proposed system can be applied to social media platforms to check the authenticity of a facial image. Hence, the proposed system can help people to identify if any facial image is altered or not on social media platforms.

References

[1] Y. Guo, X. Cao, W. Zhang and R. Wang, "Fake Colorized Image Detection," in IEEE Transactions on Information Forensics and Security, vol. 13, no. 8, pp. 1932-1944, August 2018.

[2] F. Marra, D. Gragnaniello, D. Cozzolino and L. Verdoliva, "Detection of GAN-Generated Fake Images over Social Networks," 2018 IEEE Conference on Multimedia Information Processing and Retrieval (MIPR), Miami, FL, pp. 384-389, 2018.

[3] L. Zhuo, S. Tan, J. Zeng and B. Lit, "Fake Colorized Image Detection with Channel-wise Convolution based Deep-learning Framework," 2018 Asia-Pacific Signal and Information Processing Association Annual Summit and Conference (APSIPA ASC), Honolulu, HI, USA, pp. 733-736, 2018.

[4] T. Pomari, G. Ruppert, E. Rezende, A. Rocha and T. Carvalho, "Image Splicing Detection Through Illumination Inconsistencies and Deep Learning," 2018 25th IEEE International Conference on Image Processing (ICIP), Athens, pp. 3788-3792, 2018.

[5] V. Mall, A. K. Roy and S. K. Mitra, "Digital image tampering detection and localization using singular value decomposition technique," 2013 Fourth National Conference on Computer Vision, Pattern Recognition, Image Processing and Graphics (NCVPRIPG), Jodhpur, pp. 1-4, 2013.

[6] T. Edmunds and A. Caplier, "Fake face detection based on radiometric distortions," 2016 Sixth International Conference on Image Processing Theory, Tools and Applications (IPTA), Oulu, pp. 1-6, 2016.

[7] C. Hsu, C. Lee and Y. Zhuang, "Learning to Detect Fake Face Images in the Wild," 2018 International Symposium on Computer, Consumer and Control (IS3C), Taichung, Taiwan, pp. 388-391, 2018.

[8] Tariq, Shahroz Lee, Sangyup Kim, Hoyoung Shin, Youjin Woo, Simon. (2018). Detecting Both Machine and Human Created Fake Face Images In the Wild. 81-87. 10.1145/3267357.3267367.

[9] Kumar, Manoj Srivastava, Sangeet. (2017). Image forgery detection based on physics and pixels: a study. Australian Journal of Forensic Sciences. 51. 1-16. 10.1080/00450618.2017.1356868.

[10] J. Kim, S. Han and S. S.Woo, "Classifying Genuine Face images from Disguised Face Images," 2019 IEEE International Conference on Big Data (Big Data), Los Angeles, CA, USA, 2019, pp. 6248-6250.