2022 10th International Conference on Reliability, Infocom Technologies and Optimization (Trends and Future Directions) (ICRITO)
Amity University, Noida, India. Oct 13-14, 2022

2022 10th International Conference on Reliability, Infocom Technologies and Optimization (Trends and Future Directions) (ICRITO) | 978-1-6654-7433-7/22/$31.00 ©2022 IEEE | DOI: 10.1109/ICRITO56286.2022.9964871

A Comprehensive Review on Fake Images/Videos Detection Techniques

Ruby Chauhan
Department of Computer Engineering, National Institute of Technology Kurukshetra
Haryana, India
rubyyy.rana@gmail.com

Renu Popli
Chitkara University Institute of Engineering and Technology
Chitkara University
Punjab, India
renu.popli@chitkara.edu.in

Isha Kansal
Chitkara University Institute of Engineering and Technology
Chitkara University
Punjab, India
isha.kansal@chitkara.edu.in

***Abstract*—Now that image creation and manipulation have advanced so quickly, there are serious questions about how this may affect society. At best, this leads to loss of trust in digital content. There are many existing algorithms such as Naive Bayes, CNN, RNN, Robust Hashing, GANs, SVM etc. which are being used for the detection of fake videos. Making and classifying deep fakes using Deep Neural Networks (DNN) nowadays have increased the interest of researchers in this field. Deep Fake is the regenerated media that is attained by edging in or replacing some information within the DNN model. In this work, survey withdrawn by various research groups focused the feasibility loopholes that need to be recovered for deep fakes. The use of above-mentioned techniques has been increased by a significant percentage in video game industries and cinema like enhancing visual stuff in pictures. In this paper, different types of datasets used by authors and various contemporary techniques used for fake image/video detection are described. Finally, various research gaps and the possible future directions are highlighted.**

Keywords—Fake image detection, Deep learning, Fake videos, CNNs, GANs.

I. Introduction

Massive advances have been made in the field of automatic video enhancement strategies over the previous few years. In particular, amazing success has been demonstrated in the direction of facial manipulation tactics [1]. Similar facial features can be reproduced now with the help of the advanced technologies which involves altering facial emotions from one film to another [2-3]. This makes it possible to switch between speaker identifications with little or no effort. Face manipulation systems and equipment's have improved to the point where even users with no prior knowledge in photo editing or digital arts may use them. Indeed, all the things are pre-written in some code form or other forms, which are freely available to the general public on a high rate [4-5]. On the one hand, technical advancements open up new creative prospects (*e.g.*, film making, visible effect, visible arts, and so forth).

Unfortunately, it also facilitates the technique of video forgery with the help of malevolent consumers. Spread of fake information and illegal activities or creating morphed images of some individuals for taking revenge are some of the potentially destructive areas of advanced face information and structure manipulation technology in the wrong hands. The capacity to determine whether or not a face in a videotape or photo series has become increasingly significant.[6], as the distribution of these types of manipulated films invariably has severe and Dangerous consequences (for example, reduced faith in the media, centralized opinion formation, cyberbullying, and so on). Detecting whether or not a video has been tampered with is not a new challenge. Researchers in multimedia forensics experts have been working on this content for many times, presenting a variety of solutions to specific problems [7-9]. For example, the authors of [10-11] analyze the coding records of motion movies. Exclusive strategies for detecting body duplication or deletion are proposed in [14-15]. The data set containing real and fake images after data processing is categorized in training phase and testing phase as shown in Figure 1 given below:

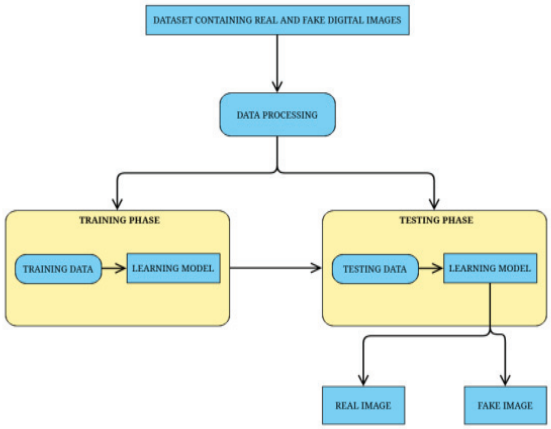

Fig. 1. Generalized Process Followed in Deep Fake Detection

All of the methods mentioned above work on the same base method: none of them can be reversed; action leaves an unseen, different type of imprint that may be used to discover the specific improvement. Forensic footprints, on the other hand, are typically dispersed and difficult to locate. Several unique face modification techniques exist (*i.e.*, entire face synthesis, attribute manipulation, identity swap, expression swap). At the end, modified motion pictures are commonly exchanged via social structures that practice resizing as well as coding procedures, impeding the overall performance of traditional forensic detectors [29]. Fake sample or fake set of images is

generated from every pristine face [12-13]. A sample of input dataset taken from FF++ and DFDC is shown in Figure 2 given below:

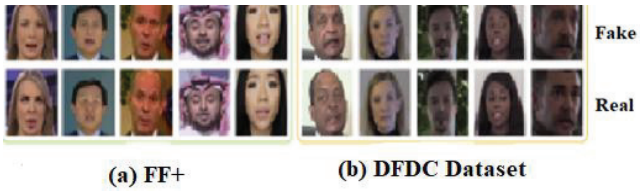

Fig. 2. Face sample inputs taken from FF++ and DFDC datasets [29]

There are few steps involved in making a face alteration video as shown in Figure 3 given below:

By using AI methods known as embedded, firstly employ thousands of two-person facial photos.

Embedded identifies and learns similarities between two faces, then restored to compressed images by Decoder.

By training one code to restore the first person's face then another code to restore the second person's face. Then insert the encoded photos into incorrect code to make a face alteration. Recorder then uses the contour and shape of Human face A to remodel human face B likewise a video.

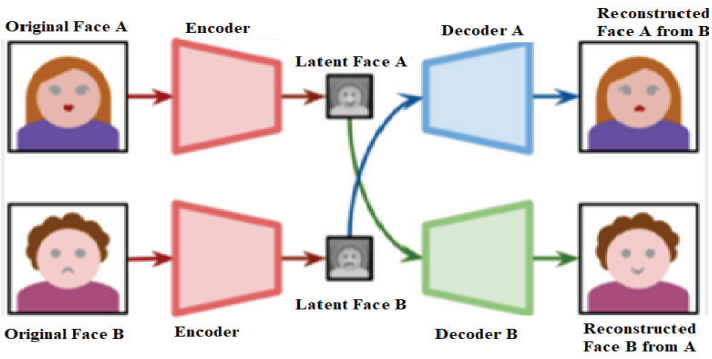

Fig. 3. Representation of the technology behind face manipulation [31]

Furthermore, Facial manipulation can be categorized in four main different categories from high to low level based on level of manipulations.

a) Synthesis of the Entire Face: With the help of advanced algorithms such as Generated Adversarial Network (GAN), completely non-existing face images can be created very easily [17]. With great quality, and realistic features, these images look very real, and the images and the results achieved are very surprising. Although, this feature has several benefits in certain industries, for example gaming industry and 3d model making industry, but there are several dark sides of these advancements too.
b) Identity Swap: As the name suggests, this is swapping the face of one individual with another individual, thus completely manipulating the identity of that individual. It basically uses two approaches:
   - Classical computer graphics-based techniques like Face Swap.
   - Various different types of deep learning methods such as deep Fakes e.g. ZAO mobile apps.

c) Attribute Manipulation: It is also known as retouching. It means to modify certain features of the image, (face particularly), so as to manipulate the image. With the help of several advanced techniques such as GAN [32]. For example, an app called Face App, used by consumers on a wide range of products similar as cosmetics, haircut in virtual terrain.
d) Expression Swap: As the name suggests, it means to manipulate the facial features or changing of facial expressions i.e. The GAN structure just replaces the expressions of the face of one person with another person. This kind of manipulation can create a lot of damage to the identity of an individual as well as to the society [10-11].

The principal motives of this method are given below:

- To bring in Deep Fake tools and technologies this can be used to change different aspects of pictures and videotapes.
- To bring in Deep Fake datasets and some traditional datasets for forensic evaluation
- To judge newly built Deep Fake detection techniques used in pictures and videotapes.

Whole process starts with providing some literature Review in Section II. Then, Description of Datasets is discussed in Section III, and Section IV proceeds to classification of different types of contemporary techniques used for fake image detection and fake video detection. Section V discusses the Evaluation parameters. Finally, Section VI gives concluding remarks and future directions.

II. Literature Review

This section comprises the study related to techniques of Fake image/ video detection. Philip S.Yu*et al.,* (2018) [27] aimed to use TI-CNN. TI-CNN is trained with both text and picture input at the same time by projecting explicit and latent characteristics into a unified feature space. Aswini Thota *et al.,* (2018) [28] described a method for fake news detection to obtain an accuracy of 94.21 percent on test data, a precisely calibrated Tiff-IDF – Dense neural network (DNN) model used in it. Research fraternity has been using Deep learning Methods as well as Artificial Intelligence Techniques for identifying forgery contents. Connor Shorten *et al*., (2019) [18] When the models are assessed on supplemented test data, they achieve 50.99 percent accuracy on the CIFAR-10 dataset against 70.06 percent accuracy on the CIFAR-10 dataset. Krithi Dinesh *et al*., (2019) [26] proposed an approach of fake news detection with the use of SVM, Naive Bays and Logistic Regression dataset. Andreas Rossler*et al*., (2019) [32] focused on the impact of compression on the detestability of state-of-the-art manipulation algorithms in this paper, and a standardized baseline for future research is proposed.

Moreover, Bowen Dong *et al*., (2019) [31] explained about Fake detector is a framework that consists of two primary components: portrayal of characteristics learning and credibility label inference, which combined form a deep diffusive network model. Hyeong-Jun Kim *et al.*, 2019 created a model for identifying false news, several processes based on "Fast text" and "Shallow-and-wide CNN" was applied and transformed [29]. Njood Mohammed *et al*., (2019) [14] aims to develop a model for classifying Instagram content in order to detect threats and forged images The model was built using deep algorithms learning whichis CNN, Alexnet network and transfer learning using Alexnet. Md Rafiqul *et al*., (2020) [17] told that GRU (Gated Recurrent Unit) has the best accuracy in both datasets, at 0.88 and 0.91, respectively. GRU, LSTM, and Tech: tan-RNN (GRU). Furthermore, a strategy given by (Chi-Chung Hsu *et al.*, (2020) [4] beats previous state-of-the-art systems in terms of accuracy, recall rate, according to experimental results. The research work done by (Worku Muluye *et al.*, (2020) [20] gave the fact that when attempting to detect deep fake movies, the image quality metrics and the lip-syncing approach with Support Vector Machine (SVM) reveal an error [31]. Neetu Pillai *et al.*, (2020) [21] False colorized picture detection using convolution neural networks (CNN) outperforms fake colorized image detection using histograms and feature extraction. Aarti Karandikar*et al.*, (2020) [22] used a convolution neural network to detect deep fake and resulted that the accuracy of the model reported in the paper is roughly 70%. Suhail Yousaf *et al*., (2020) [30] used a CNN based approach for deep fake detection with attention target specific regions and manual distillations extraction , Using ensemble approaches and a variety of linguistic feature sets, able to categories as true or false.

Miki Tanaka *et al.,* (2021) [23] Dataset for Image Manipulation hashing approach was chosen because of its high robustness against image compression and resizing

T.T. Nguyen *et al*., (2021) [16] suggested method which has a promising performance in detecting false videos, which can be further enhanced by taking into account dynamic patterns of blinking, such as excessively recurrent blinking can be a sign of tampering. Bhutanese Singh *et al*., (2021) [30] beyond B0, using higher scaled forms of Efficient Net results in over learning and decreased accuracy 85.3 percent and 81.2 percent, respectively.

Jamal Abdul *et al*., (2021) [25] proposed an innovative hybrid deep learning model for false news classification that blends convolutional and recurrent neural networks The model was verified effectively on two fake news datasets (ISO and FA-KES), yielding detection results that outperformed existing non-hybrid baseline approaches. I.M.V. Krishna *et al*., (2021) [19] claimed in their research that a model on the basis of count victimizer or an if matrix *(i.e*. word censuses relative to how frequently they're used in other papers in your dataset) can help in finding applicable papers. Another important research presents a new way of predicting fake image in social media well in advance. In this, Md Shohel *et al.,* (2022) provided an updated overview of work done in deep fake by conducting systematic literature review (SLR) and analyses them by grouping into categories like deep learning, classical ML, statistical, blockchain based techniques.

III. DESCRIPTION OF DATASETS

Datasets play a major role in detecting and predicting deep fakes *i.e.* a large number of pictures and videotapes to feed the data to a ML algorithm as illustrated in Table 1

TABLE I. DATASET USED IN FAKE VIDEO DETECTION.

Sr. No.	Dataset	Links
1	Celeb-DF fake processed videos	Celeb-DF1
2	Celeb-DF Real processed videos	Celeb-DF2
3	Face Forensics++ Real and fake processed videos	FF++
4	DFDC Fake processed videos	DFDC1
5	DFDC Real processed videos	DFDC2

Traditional and deep fake datasets are the two main categories into which forensics datasets can be divided. Traditional forensics datasets are prepared manually with a lot of labour under strictly controlled settings including rotation detection, in painting, splicing, and camera imperfections. A number of datasets with picture alterations were suggested. Traditional forensics datasets have been produced manually, and under very strict controls. Table 2 illustrates a dataset description used in fake image detection.

TABLE II. DATASET DESCRIPTION USED IN FAKE IMAGE DETECTION.

S.No	**Dataset Description**	**Dataset Link**
1	Dataset consists of video sequences which are manipulated using automated face manipulation methods.	FaceForecncis+
2	Over 70,000 images of human faces having very good quality of 1024 * 1024.	Flickr Faces
3	Unique new dataset created by experts to benchmark deep fake detection models.	DFDC
4	Non-commercial dataset of over 367,000 faces annotated using 3,100 subjects' facial points.	UMDFaces
5	Deep fake TIMIT swaps videos using a GAN-based approach developed from the encoder-based Deep fake algorithm.	DeepFakeTIMIT
6	Dataset containing people of all ages, ethnicity and gender.	UTKFace
7	Collective of 3 datasets A, B, C CAISA Gait includes 20 people with 4 sequences for each dimension.	CASIA Gait

8	Google AI dataset of over 156,000 facial images meticulously annotated by six human annotators.	GFEC
9	Commercial research purpose dataset having over 200,000 celebrity images.	CelebA
10	VidTIMIT consists of videos of people reciting short phrases. Used for research on face recognition.	VidTIMIT
11	Dataset of over 10,000 images of people from across 15 countries aged 4 and 70 years.	TuftsFace
12	Over 10,000 images from seven different cameras were converted to greyscale and scaled to 512*512.	BossBase

IV. CLASSIFYING TECHNIQUES OF FAKE IMAGE / VIDEO DETECTION

Computer vision nowadays is using a ton of machine learning and deep learning [24, 28] for detection and prediction of deep fakes. Alike stereotypes have been used in the prediction of fake image detection. In this section, Table 3 illustrates an overview of many modern ways, used by different authors for fake video detection and Table 4 illustrates classification of contemporary techniques for Fake Image Detection.

TABLE III. CLASSIFICATION OF MODERN TECHNIQUES USED FOR FAKE VIDEO DETECTION.

Ref.	**CFFN**	**CNN**	**GAN**	**LSTM**	**RNN**	**Robust Hashing**	**Alex Net**	**SVM**	**NaiveBayes**	**DNN**
T.T. Nguyen *et al*.[16]	✓	✓	✓							
Md. Rafiqul et al.,[17]				✓	✓					
C. Shorten *et al.[18]*			✓							
B. Singh*et al, .[33]*					✓					
C.C. Hsu*et al.[4]*	✓									
W.M. Wubet*et al.,[20]*	✓			✓						
N. Pillai *et al*.,[21]		✓								
A.Karandikar*et al.,[22]*		✓		✓	✓					
M. Tanaka*et al.,[23]*						✓	✓			
N.M. AlShariah*et al.,[14]*		✓			✓					
J.A. Nasir*et al.[25]*								✓	✓	
K.D. Kottary*et al.[26]*		✓								
P.S. Yu *et al.,[27]*										✓
A.Thota*et al, [28]*				✓	✓				✓	
K.A. Kumar *et al*.[29]				✓					✓	
I.M.V. Krishna *et al,[19].*									✓	
S. Yousaf *et al*., [30]		✓								
B.D. Jiawei *et al*.,[31]										✓
H.J. Kim *et al.[24]*		✓			✓					
A. Rossler*et al.[32]*		✓						✓		

TABLE IV. CLASSIFICATION OF CONTEMPORARY TECHNIQUES FOR FAKE IMAGE DETECTION.

Ref.	NLP	Frequency Analysis	CNN	ML	Resnet50	SVD	SVM	Deep Learning	Feature Learning	VGG
Y. Li et al. [1]			✓		✓					✓
P.Korshunov*et al.* [2]							✓			✓
D. Archer *et al.* [3]			✓							
C. Hsu *et al.* [4]									✓	
K.Kuruvilla *et al.* [5]				✓				✓		
E. Cuvee et al. [6]	✔									
X. Zhang et al. [7]				✓						
F. Mara et al. [8]			✔				✔			
L. Nat raj et al. [9]				✓						
H. Chen et al. [10]									✔	
N. Abad et al. [11]						✔				
J. Frank et al. [12]		✔								
R. Shankar et al. [13]			✔	✓						
N.AlShariaet al.[14]			✔						✔	
L. Zhou et al. [15]			✔							

V. CONCLUSION AND FUTURE SCOPE

Various researchers have been working in the field of deep fake image/video detection by using state of the art approaches such as Machine Learning Based techniques, analytical techniques, deep learning-based techniques. In this paper, various contemporary techniques used for fake image/video detection are classified which will help the research fraternity to choose the best approach based on their requirements. After careful analysis and review of various research techniques, it is observed that deep learning techniques have gained maximum accuracy but work can be done to improve this accuracy. Deepfake detection still face many challenges, so, this paper provides a good resource for the researchers to develop effective detection methods and alternative solutions for it.

REFERENCES

[1] Y. Li and S. Lyu, "Exposing deepfake videos by detecting face warping artifacts,"*Openaccess.Thecvf.com*, pp. 46-51," In Proceedings of IEEE, 2018.

[2] P. Korshunov and S. Marcel, "DeepFakes: a new threat to face recognition assessment and Detection,"*Arxiv.org*, vol. 31, no. 2, pp. 1-5, December 2018.

[3] D. Afchar, V. Nozick, J. Yamagishi and I. Echizen, "MesoNet: a compact facial video forgery detection network," In Proceedings of *Ieeexplore.ieee.org*, pp 1-6, IEEE, 2018.

[4] C. Hsu, Y. Zhuang and C. Lee, "Deep fake image detection based on pairwise learning," *MDPI*, pp. 1-14, January 2020.

[5] K. Kuruvilla and Y. Gaikwad, "Fake Image detection using machine learning," *IRACST*, vol. 7, No. 2, pp. 19-22, April 2017.

[6] E. Cueva, G. Ee, A. Iyer, A. Pereira, A. Roseman and D. Martinez, "Detecting fake news on twitter using machine learning models, " *Soe.rutgers.edu*, pp 1-12, July 2020.

[7] X. Zhang, S. Karaman and S. Chang, "Detecting and simulating artifacts in GAN fake images," *Arxiv.org*, pp. 1-10, October 2019.

[8] F. Marra, D.Gragnaniello, D. Cozzolino and L. Verdoliva, "Detection of GAN-generated fake images over social networks," *Researchgate*, pp 1-6, April 2018.

[9] Nataraj, L., Mohammed, T.M., Manjunath, B.S., Chandrasekaran, S., Flenner, A., Bappy, J.H. and Roy-Chowdhury, A.K., 2019. Detecting GAN generated fake images using co-occurrence matrices. *Electronic Imaging*, *2019*(5), pp.532-1.

[10] H. Chen, K. Zhang, S. Hu, S. You and C. Kuo, "Geo-DefakeHop: high-performance Geographic fake image detection, " *arXiv.org*, pp. 1-12, 2021.

[11] N. Abbadi, A. Hassan and M. AL-Nwany, "Blind fake image detection," *IJCSI*, vol. 10, No.1, pp.180-186, July 2013.

[12] J. Frank, T. Eisenhofer, L. Schonherr, A. Fischer, D. Kolossa and T. Holz, "Leveraging frequency analysis for deep fake image recognition,"*arXiv.org*, vol. 119, pp 1-12, 2020.

[13] R. Shankar, A. Srivastava, G. Gupta, R. Jadhav and U. Thorate, "Fake image detection using Machine learning," *Ijcrt.org*, ISSN: 2320-2882, vol. 8, pp. 295-302, May 2020.

[14] N. AlShariah and A. Saudagar, "Detecting fake images on social media using Machine learning,"*Thesai.org*, vol. 10, No. 12, pp. 170-176, 2019.

[15] L. Zhou, S. Tan, J. Zeng and B. Lit, "Fake colorized image detection with channel-wise convolution based Deep-learning framework,"*Ieeexplore.ieee.org*, pp. 733-736, November 2018.

[16] Nguyen, Thanh Thi, Cuong M. Nguyen, Dung Tien Nguyen, Duc Thanh Nguyen, and SaeidNahavandi. "Deep learning for deepfakes creation and detection," *arXiv preprint arXiv:1909.11573* 1 (2019)

[17] Nataraj, Lakshmanan, Tajuddin Manhar Mohammed, B. S. Manjunath, Shivkumar Chandrasekaran, Arjuna Flenner, Jawadul H. Bappy, and Amit K. Roy-Chowdhury. "Detecting GAN generated fake images using co-occurrence matrices,"*Electronic Imaging*2019,no.5(2019):532-1.

[18] Connor Shorten, Taghi M. Khoshgoftaar. "A survey on image data augmentation for deep learning," Journal of big data 6, No. 1 (2019): 1-48.

[19] I.M.V. Krishna, S.Sai Kumar. "Fake News Detection Using Naive Bayes Classifier," In Proceedings of International Journal of Creative Research Thoughts, pp. e757-e761. 2021.

[20] WorkuMuluyeWubet. "The deepfake challenges and deepfake video detection,"In proceeding of International Journal of Innovative Technology and Exploring Engineering (IJITEE) 9, No. 6 (2020): 7-796.

[21] Neetu Pillai, "Fake colorized and morphed image detectionusingconvolutionalneuralnetwork,"ACCENTSTransactionson ImageProcessingandComputerVision6,No.18(2020):8.

[22] Rashmi Welekar, Aarti Karandikar and ShubhangiTirpude. "Emotion Categorization Using Twitter," In Proceedings of International Journal 9 No. 3 (2020).

[23] Miki Tanaka, Sayaka Shiota and Hitoshi Kiya. "A detection method of operated fake-images using robust hashing," Journal of Imaging 7, No. 8 (2021): 134.

[24] Dong-Ho Lee, Yu-Ri Kim, Hyeong-Jun Kim, Seung-Myun Park, and Yu-Jun Yang. "Fake news detection using deep learning," Journal of Information Processing Systems 15, No. 5 (2019): 1119-1130.

[25] Jamal Abdu Nasir, Osama Subhani Khan, and IraklisVarlamis. "Fake news detection: A hybrid CNN-RNN based deep learning approach," International Journal of Information Management Data Insights 1, No. 1 (2021): 100007.

[26] Anusha Achaya, AshwithaJathan,Deepa Anchan, Krithi Dinesh Kottary, and Mr Sunil BN. "Fake News Detection Using Machine Learning," (2019).

[27] Zonghan Wu, Shirui Pan, Fengwen Chen, Guodong Long, Chengqi Zhang, and S. Yu Philip. "A comprehensive survey on graph neural networks," IEEE transactions on neural networks and learning systems 32, No. 1 (2020): 4-24.

[28] Aswini Thota, Priyanka Tilak, Simrat Ahluwalia and NibratLohia. "Fake news detection: a deep learning approach," SMU Data Science Review 1, No. 3 (2018): 10.

[29] Arun Kumar, G. Preethi and K. Vasanth. "A Study of Fake News Detection usingMachieLearning Algorithms," Int.J.Technol.Eng. Syst 11, No. 1 (2020): 1-7.

[30] Iftikhar Ahmad, Muhammad Yousaf, Suhail Yousaf, and Muhammad Ovais Ahmad. "Fake news detection using machine learning ensemble methods, "Complexity (2020).

[31] Zhang, Bowen Dong Jiawei, and S. Yu Philip. "Fakedetector: Effective fake news detection with deep diffusive neural network," In proceeding of 36th international conference on data engineering (ICDE), pp. 1826-1829. In Proceedings of IEEE 2020.

[32] Andreas Rossler, Davide Cozzolino, Luisa Verdoliva, Christian Riess, Justus Thies, and Matthias Niebner. "Faceforensics++: Learning to detect manipulated facial images," In Proceedings of IEEE/CVF international conference on computer vision, pp. 1-11. 2019.

[33] B. Singh and D.S. Sharma, "Predicting image credibility in fake news over social media using multi-modal approach," part of Springer Nature 2021.