Indonesian Journal of Electrical Engineering and Computer Science
Vol. 32, No. 2, November 2023, pp. 1168∼1176
ISSN: 2502-4752, DOI: 10.11591/ijeecs.v32.i2.pp1168-1176

A survey on deepfake video detection datasets

Md Fahimuzzman Sohan, Md Solaiman, Md Aumit Hasan
Department of Software Engineering, Daffodil International University, Dhaka, Bangladesh

Article Info

Article history:

Received June 15, 2023
Revised July 8, 2023
Accepted August 15, 2023

Keywords:

Datasets
Deep learning
Deepfake detection
Deepfakes video
Survey

ABSTRACT

Deepfake video has usefulness in entertainment and multimedia technology, however, the danger of deepfake is significant to the social, economical, and political sectors so far. Specifically, to diverge any public opinion by generating fake news and spreading misleading information, national security may be under risk due to misrepresenting statements given by political leaders. The creation of such manipulated videos are getting easier day by day and at the same time it is necessary to detect and prevent them. In order to do that, researchers are creating challenging fake video databases for artificial intelligence (AI) based detection models to contribute to the research. This paper reviewed the existing deepfake video detection datasets available online and used in the previous research articles. We analyzed the literature from two different perspectives, datasets and detection models. The goal of this study is to introduce all publicly available datasets in this field including the discussion of techniques used to generate the data. In addition to our contribution, we showed a result comparison among different deepfake datasets and discussed the findings.

Corresponding Author:

Md Fahimuzzman Sohan
Department of Software Engineering, Daffodil International University
Dhaka, Bangladesh
Email: fahimsohan2@gmail.com

1. INTRODUCTION

In Figure 1, the first picture (left) is taken from a Twitter video published a few days later of starting the Russia-Ukraine war in 2022, where apparently it seems the Ukrainian president Volodymyr Zelenskyy is asking his soldiers to surrender and give up their arms, however, it was a deepfake video and the president never gave that speech. A deepfake video is a video generated by artificial intelligence (AI) and machine learning algorithms to create realistic videos of people saying and doing things that they never did. So, the question is how such videos are created? Actually face manipulation is not a new concept, years ago makeup artists worked to manipulate the looks of actors in the film and drama industry. Similarly, many tools and techniques are used to manipulate pictures and videos in the era of current digital image [1]. However, there is a fundamental difference between deepfake and traditional face manipulation. Usual approach requires skilled people who operate the editing tools and deal with the graphical operation [2]. On the other hand, deepfake is produced in an automatic way, as bread is manufactured in an automatic machine from raw materials. For example, to produce a deepfake video it just needs to input two videos, source and destination, rest action will be performed by the automatic AI imaging art which replaces source face with targeted face (face swapping) using very advanced graphic design. Therefore, the mystery of the deepfake shows in Figure 1 is replacing someone's face with actual Volodymyr Zelenskyy who is in the right picture. Although deepfake has many hopeful prospects that can be used for entertaining and digital media, they can also be dangerous and catastrophic for not only politicians but also celebrities and regular civilians. We discussed the hazard of deepfake video earlier in this section but this is not the only area to concern about, there are other types of deepfake, it can be done by audio where voice can be changed to its sounds like someone else, similarly writing can also be distorted by AI oriented tools. Hence, no doubt such activities are a great potential threat for future society because it can start to dispute any reality by manipulating people's perception where common people can not define what truth is. However, these growing threats are not yet extensively in use; besides several giant technology companies are aware of this harmful potential, they are spending money in research and coming up with powerful tools and technologies to detect this kind of malicious activities. Mainly deep learning techniques are employed to detect deepfake and interestingly the same techniques were used to create them. As usual, deep learning based detection models require dataset which has an important role for building robust machine learning (ML) models since ML algorithms build mathematical models based on sample data known as training data to make detection. Therefore, several deepfake video datasets have been published by different platforms such as YouTube, Facebook to contribute to detect and prevent the videos online [3].

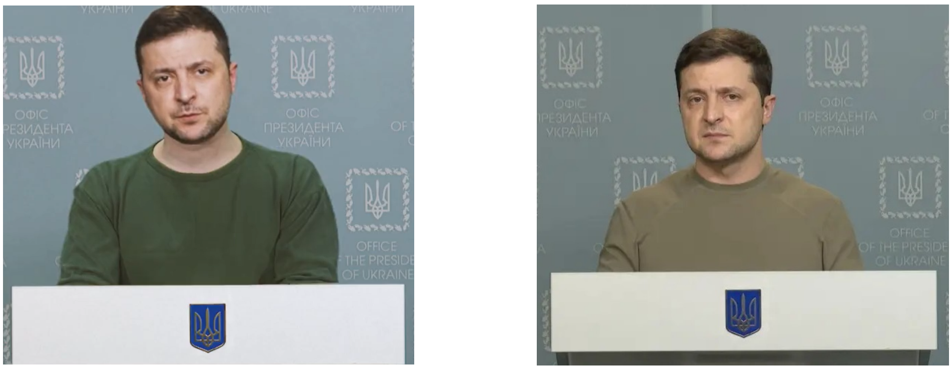


Figure 1. An example of deepfake, collected from two different videos: it looks like both of the pictures represent Ukrainian president Volodymyr Zelenskyy, however, the picture in left is fake and taken from a deepfake video where someone's face is replaced by Volodymyr Zelenskyy

In this study, we discovered and reviewed the published deepfake video detection datasets used in the research field from various perspectives. Mainly we focused on two integral parts of deepfake detection, different datasets and ML models used in the research area. Practically, The researchers who wish to contribute to the datasets are trying to build more realistic fake videos that will be difficult to identify by any ML model. In contrast, deepfake detection research is taking this challenge by trying to construct more precise detection models. The following contribution we included in this survey, i) listing and discussion of all publicly available deepfake video datasets, ii) brief description on how these datasets are created (techniques), and iii) collecting a group of deepfake detection research papers and compared the results according to the used datasets.

The remaining parts of this article describe: section 2 presents the information of all included datasets and the data creation methods. The performance achieved by different deepfake datasets are discussed in section 3. Lastly, section 4 concludes this study with necessary discussions.

2. DATASETS

AI is increasingly becoming more powerful than human editing, it gets better at generating deepfakes and we must also get better at identifying them. Deepfake detection using deep learning actually works like "using a thorn to remove a thorn", that is to say such manipulated videos will be identified by similar videos data in ML approach. So, dataset is the primary focus of deepfake video detection, a challenging dataset can compete with critical situations more accurately. Already a number of deepfake video datasets have been published in different repositories for research and development. In this section we discussed the publicly available datasets used in this research field from 2018 to recent. More specifically, first we have included details of all available datasets, then provided an overview of every data generation technique, third discussed the features of the datasets, and finally performance of the ML models using the datasets are compared. We intensively investigate dozens of research articles to extract all available deepfake video detection datasets online. Mainly three types of data exist in the literature, most of them are constructed by video, some are image, and audio data. Besides, a group of datasets are public, freely accessible to anyone and some researchers used personally built datasets for their models [1]. In this article we briefly described the commonly used public datasets found in different repositories. Table 1 listed 13 datasets which we accumulated from our findings based on the use and availability.

Table 1. Fundamental information of existing deepfake video detection datasets

No.	Dataset	Release	Source	Real		Fake	
				Video	Frame	Video	Frame
[4]	Faceforensics	Mar 2018	YouTube	1,004	519.127K	1,004	521.406K
[5]	Faceforensics++	Jan 2019	YouTube	1,000	509.914K	4,000	18M
[6]	DeepFakeDetection	Sep 2019	Paid participants	363	315.4K	3,068	2242.7K
[7]	UADFV	Nov 2018	YouTube	49	-	49	-
[8]	DeepfakeTIMIT	Dec 2018	YouTube	-	-	620	68K
[9]	Celeb-DF	Sep 2019	YouTube	408	-	795	-
[10]	Celeb-DFv2	Nov 2019	YouTube	590	230.100K	5,639	2199.210K
[11]	DFDC preview	Oct 2019	Paid participants	1,131	-	4,119	-
[12]	DFDC	Jun 2020	Paid participants	23.654K	-	104.500K	-
[13]	DeeperForensics-1.0	Jan 2020	Paid participants and YouTube	50K	12.6M	10K	5M
[14]	WildDeepfake	Oct 2020	Collected from online	3,805	440.5k	3,509	739.6k
[15]	KoDF	Mar 2021	Volunteer participants	62.166K	-	175.776K	-
[16]	ForgeryNet	Jul 2021	-	99.630K	1,438.201K	121.617K	1,457.861K

Since deepfake is comparatively a new fenomena, the resources are still inadequate and researchers are trying to contribute regularly. Therefore, databases get updated with different versions as the table shows. Faceforensics is the first deepfake video dataset found according to the publication date of associated research articles in the field. The dataset was created by collecting videos from youtube without preserving the participants rights. 1,004 fake and equal number of real videos of 480p resolutions are used in this widely applied dataset. Following this one Faceforensics++ is an updated database where researchers mainly increased the number of manipulated videos than the previous version. The initial set included only low quality (854×480 pixels resolution) videos, however, this time high definition (HD), full HD videos are also considered. In addition, Faceforensics++ database is constructed by four different face manipulation techniques (i.e., DeepFakes, Face2Face, FaceSwap, NeuralTextures) while only Face2Face were used previously. Similarly, Deepfakedetection is another extended part of FaceFoensics dataset invented by collaboration with Google and Jigsaw [17].

Although, it is comparatively a smaller database than the preliminary one, created with 363 fake and 3,068 real videos. The promising feature of this dataset is they preserved the image rights of each participant. A tiny dataset UADFV, consists of 49 fake and 49 real videos with 294×500 pixel elements. They used a mobile application named FakeApp to create the synthetic videos [18]. Deepfake-TIMIT is prepared from previously published YouTube VidTIMIT dataset. They collected 16 pairs of individual faces from the VidTIMIT based on the similarities of different visual features between each pair. Since each individual face has 10 videos, 16 pairs compiled 320 videos by generative adversarial network (GAN)-based face-swapping algorithm. The dataset is compiled twice, once with 320 low quality (64×64) and again with 320 high quality (128×128) fake videos. The first version of Celeb-DF data presented in September 2019. The authors used the original DeepFake Synthesis algorithm to create 795 forged videos from 408 genuine YouTube videos of 59 celebrities. Celeb-DF-v2 is the larger database than the previous version, a total 5,639 fake videos were created by the improved synthesis algorithm from 590 normal videos of different ages, ethnic groups and genders. Similarly deepfake detection challenge (DFDC) Preview and DFDC are two corresponding datasets, mainly they differ in the number of samples. The preview dataset consists of 1,131 real and 4,119 fake videos while DFDC is a large-scale database where 23,654 clean and 104,500 manipulated videos exist. A faceswap technique was used initially to create deepfake videos and eight different methods are considered for DFDC dataset. Both of the datasets include image rights, all the participants (66 in DFDC preview and 960 in DFDC) are allowed to use their face in this particular task. DeeperForensics is another dataset containing thousands of videos with millions of frames, a collection of 10,000 real and 50,000 fake videos. The source data are collected from 100 paid participants from various countries, age, and facial expressions. WildDeepfake dataset is different from other competitive dataset. The researchers did not create any fake videos by their own, instead they directly collected deepfake videos that are available on various video sharing websites. KoDF is the largest deepfake video dataset we found, consisting of 175,776 manipulated videos. The contributors mainly emphasized the Korean ethnic group. The face manipulation was performed by six different synthesis models, such as FaceSwap, DeepFaceLab, face swapping GAN (FSGAN), FOMM, ATFHP, and Wav2Lip. A recent enormous dataset ForgeryNet released with 99,630 real and 121,617 fake videos. The authors claim they used 15 video forgery approaches. In addition, they maintained diversity in terms of several visual aspects while collecting the source samples from four different databases. Table 2 shows the tools and techniques used to generate the datasets.

Table 2. List of deepfake creation techniques used to prepare the datasets

No.	Dataset	Method
1	Faceforensics	Face2Face
2	Faceforensics++	DeepFakes, Face2Face, FaceSwap, NeuralTextures
3	DeepFakeDetection	DeepFakes, Face2Face, FaceSwap, NeuralTextures
4	UADFV	FakeApp mobile application
5	DeepfakeTIMIT	FSGAN
6	Celeb-DF	Original DeepFake synthesis algorithm
7	Celeb-DFv2	Improved DeepFake synthesis algorithm
8	DFDC preview	Faceswap
9	DFDC	DF-128, DF-256, MM/NN, NTH, FSGAN, StyleGAN, Refinement, and Audio Swaps
10	DeeperForensics-1.0	DF-VAE
11	WildDeepfake	Various sources on the internet
12	KoDF	FaceSwap, DeepFaceLab, FSGAN, FOMM, ATFHP, and Wav2Lip
13	ForgeryNet	15 techniques*

*Detailed description of the forgery approaches is provided in [16]

2.1. Deepfake video generation

Several previous review articles in the field classified the deepfake generation in different categories [19]. Following the literature and our investigation we divided the method of creating deepfake video dataset in the four different classes. They are: face reenactment, identity swap, entire face synthesis, and attribute manipulation.

- The face reenactment mainly transfers the facial expression of a person to another, thus it is also known as expression swap method. Traditional computer graphics and deep learning both approaches are applied to manipulate expression. For example Face2Face is a conventional facial expression replacement technique proposed by Thies *et al.* [20]. More specifically suppose expressions of person ‘T’ (target video) is possible to replace by another person ‘S’ (source video). It is able to create animation of legacy video footage without changing the identity of the target person. The source faces are collected by a red, green, and blue (RGB) sensor and go through the transfer functions for real time expression transformation. The approach also emphasizes on the background features of target video during image synthesis. Moreover, their mouth synthetic approach follows the target mouth shape to ensure fair adjustment with the source data. NeuralTextures is another facial reenactment technique. According to Thies *et al.* [21] the method combines traditional graphics based image synthesis with machine learning. One of their features was transferring facial expression from source to target video. They generated an altered urban vector (UV) map of the target face using the expression of the source face. The UV map is further associated with the neural texture map. Lastly, the background image of the target face to the neural renderer used for final output imagery.
- Identity swap or face swapping mainly transfers the face of a person to another. Following the previous method researchers used computer graphics and deep learning approach for their datasets. For instance faceswap is a computer graphics-based Python application that uses face alignment, Gauss Newton optimization and image blending techniques to swap the face of a target face with a source face. In order to do that, first the system detects the face region and locates the facial landmarks from the input image. After that the landmarks are fitted into a 3D template model. In the third step, pygame is used to render the 3D models. In addition color correction is used during the blending operation of the source face with the targeted one. Furthermore, faceswap github is a Python program to produce manipulated videos. The technique mainly uses deep learning to replace the source face to a target face. However, in this process the original and target face conduct an encoder-decoder operation. Practically two original faces participate in an operation where each face has their own encoder. After that both encoders produce respective latent faces. Lastly,

the decoders use exchanged latent faces to make the reconstructed faces. Generative adversarial networks (GNA) based modes are parallelly popular in making deepfake videos. Traditional supervised ML settings consist of one AI model to make a prediction or detection. While GNA is unsupervised and uses two neural network modes, generator and discriminator. The generator creates fake samples and locates to the second neural network model which decides whether the image is real or fake by analyzing the real sample from the database. Faceswap-GAN is an example of such fake video generator developed from previously described encoder-decoder based faceswap GitHub algorithm additionally includes the discriminator model.

– Entire face synthesis works with the entire image during the reconstruction process. Indeed, background style, lighting, movement may influence the synthesis process and this has been effectively implemented in several research. StyleGAN is one of them, an improved face synthetic method that considers the entire face. This powerful GAN based system is able to process high-level attributes such as pose, identity of human face and finally generate high, quality realistic manipulated facial images.
– Attribute manipulation targets some specific regions of the face to modify the particular attributes. This editing may include various facial appearances such as changing the eyeglass, skin tone, viewpoint, modifications of hairstyle, age, gender and so on. Since GAN has popularity and implementation materials are publicly available for everyone, most of the attribute manipulation methods were constructed with GAN approaches. Choi *et al.* [22] have proposed StarGAN, which uses a single model to transfer one image to another. During the training session image and a selected target domain are inputted into the generator for each iteration. To illustrate, facial attributes such as hair color is inputted for some iterations and facial expression such as happiness is inputted for other iterations by repeating this the generator can be trained under multiple domains.

3. PERFORMANCE OF THE DATASETS

The advances in the field of deepfakes are equally convenient and alarming. Deepfake detection is extensively important and we must also get better at identifying them to prevent the evil impact on the social and financial sector. Sometimes it is difficult for the innate human eye to detect the results of deepfake technology while a ML or deep learning based method can efficiently differentiate fake videos. Deepfake video detection is used to be a binary classification where the classifiers give the video the response of either fake or real. A prominent ML algorithm SVM was used in the detection models earlier but it has high error rate issues. However, deep learning based detection techniques are gaining momentum as they achieve better results. Under such that technique, first the video which is to be tagged is gathered by the input video. Then selected and divided into multiple frames and the frames are subjected to face alignment in the pre-processing step, finally a ML model is applied to the frames which checks if there are any inconsistencies between the extracted frames and actual reference to the video frames. In Table 3, we show a collection of articles (Reference) and their average scores according to the datasets.

Table 3. References of articles with average accuracy and AUC scores achieved by the different datasets. Some spaces are empty due to the unavailability of data

No.	Dataset	Accuracy		AUC	
		Reference	Average score	Reference	Average score
1	Faceforensics	[23]	98	-	-
2	Faceforensics++	[24], [25], [26], [27], [28], [29], [30], [31], [32], [33], [34], [35], [36], [37], [38]	94.2	[39], [40], [41], [42], [43]	93.55
3	DeepFakeDetection	[44]	90.80	-	-
4	UADFV	[45], [46], [47]	93.4	[48], [49]	98.7
5	DeepfakeTIMIT	[38]	99.45	[39], [44], [50], [51]	92.96
6	Celeb-DF	[25], [29], [33], [35], [36], [41], [45], [47], [52]	85.79	[27], [43], [49], [50], [53], [54]	82.23
7	Celeb-DFv2	[34]	99.31	[40], [32], [31]	88.01
8	DFDC preview	[24], [45], [55], [56], [57]	91.61	[51]	84.4
9	DFDC	[29], [35], [42], [58]	83.27	[29], [31], [41], [42], [59]	89.3
10	DeeperForensics-1.0	[30]	62.46	-	-
11	WildDeepfake	[34], [35]	85.21	[41]	85.11
12	KoDF	-	-	[51]	89
13	ForgeryNet	-	-	-	-

According to the previous study the manipulated video detection models are divided into two different forms, within-domain and cross-domain. Within-domain also can be addressed by within-dataset approach since the data is used from the same database during train, validation, and test of the detection models. On the other hand, cross-domain is mainly a cross-dataset setting where different datasets' data are used to train, test or validate the models. The second approach is comparatively more complex because of the diversity in characteristics of different datasets, thus very few research materials are available on cross domain research (for example [39], [40], and [60]). Although whatever domain is used the deepfake detection is mainly carried out by deep learning algorithms, particularly CNN is the numerously considered in fake video detection [61]. The detection performance of such deep learning models are diverse across the datasets. Practically the performance widely depends on which dataset is used, besides the ML model also has a significant impact on the results. However, for this study we collected a bunch of research articles on deepfake video detection research and considered different datasets found from our study. We included scores of two commonly used performance measures, accuracy and area under the curve (AUC). It is clearly observable from the table that Faceforensics++ is the most frequently used dataset in this community followed by Celb-DF and DFDC. The average accuracy and AUC scores of these three datasets are 92.2%, 93.55%; 85.79%, 82.23%; and 83.27%, 89.3% respectively.

4. CONCLUSION

Face manipulation is not a new advancement, it has been practiced in the graphics community for decades. Currently deep learning based face enhancement methods (deepfake) add a lot of attention to the interested people in different sectors because of its cost and time affordable and easy to implement characteristics. However, not all deepfake videos are useful for human beings on digital platforms. Therefore, malicious deepfakes need to be identified immediately and prevent future attacks. This is why deepfake video detection is necessary to protect online space. To swift the deepfake video detection research we essentially need to build large-scale, multi-dimensional, and challenging datasets. Till now several deepfake video detection datasets have been published and widely used in the research community. In this survey we assembled all of the datasets used in deepfake video detection research. Our investigation includes a brief explanation of the datasets, the methods followed to create them, and the performance achieved by the datasets under different ML models. We intensively investigated all of the platforms to collect potential publicly available datasets and found 13 datasets that are used in deepfake video detection research. In conclusion, the following discussion will highlighted our investigation:

- The datasets: Our investigation ended with 13 deepfake video datasets which are publicly available and widely considered in the research. We observed most of the datasets have multiple versions. Besides, we found an ongoing competition among the contributors, particularly they focused on dataset size which offers larger datasets in recent periods. Some of them considered multi ethnicity which provides a heterogeneous outlook for the datasets. However, the quantity is increasing continuously but there is a question about quality. We investigated hundreds of deepfake videos from the datasets and found most of them can be identified easily by open eyes. Hence, the major drawbacks include eye blinking, lips movement, and replaced faces were not matched properly. In addition, most of the data were collected from interview or news presentation based videos which may not be effective in an universal detection model since fake video creation technology has improved a lot.
- Deepfake creation methods: Fake video creation mainly follows two methods in technical aspect, graphical and deep learning. Computer graphic based approach is comparatively cost and time consuming where deep learning approach is automated and easy to handle. One of the commonly considered methods is FaceSwap and available in both forms (i.e., graphic based face swap and deep learning based face swap). Other commonly used algorithms are Face2Face, DeepFakes, and different GAN oriented generation approachs. However, we observed limitations from our practical experience in deep learning based tools, though it is automated but takes a long time to process the data, even sometimes several hours for a single video. Moreover, it requires high configured computer set-up during data preprocessing and model generation.
- Performance analysis: The impact of dataset on the deepfake video detection is undeniable. The results show detection scores of the datasets are high for some models and also have lower performance scores. It is difficult to conclude with a concrete remark on the overall performance, because new datasets are added to the community regularly, thus the datasets were not used uniformly among the articles. However, datasets that have a long history in the research also have higher accuracy and AUC, for instance, Faceforensics to

DeepfakeTIMIT, all have achieved more than 90% scores. On the other hand, it is noticeable that recently released datasets are challenging compared to the previous fleet since prepared fake video quality is increasing in the datasets. It is very expected that further competitive deepfake detection models will comes up in future articles with most accurate results.

This paper presented a survey where we focused on deepfake video detection research. We included our review outcomes with detailed discussion of existing fake video datasets, the methods used to prepare the videos, and performance achieved by the datasets under different detection models. We expect this study has successfully covered all datasets available in the research domain. Moreover, our discussion will assist to explore relevant information of the data for future contributors.

REFERENCES

[1] P. Yu, Z. Xia, J. Fei, and Y. Lu, "A survey on deepfake video detection," *IET Biometrics*, vol. 10, no. 6, pp. 607–624, 2021, doi: 10.1049/bme2.12031.

[2] S. H. Silva, M. Bethany, A. M. Votto, I. H. Scarff, N. Beebe, and P. Najafirad, "Deepfake forensics analysis: An explainable hierarchical ensemble of weakly supervised models," *Forensic Science International: Synergy*, vol. 4, p. 100217, 2022, doi: 10.1016/j.fsisyn.2022.100217.

[3] Z. Akhtar, "Deepfakes generation and detection: A short survey," *Journal of Imaging*, vol. 9, no. 1, p. 18, 2023, doi: 10.3390/jimaging9010018.

[4] A. Rössler *et al.*, "FaceForensics: A large-scale video dataset for forgery detection in human faces," arXiv.org, doi.org/10.48550/arXiv.1803.09179 (accessed Jun. 15, 2023).

[5] A. Rossler *et al.*, "FaceForensics++: Learning to detect manipulated facial images," in *Proceedings of the IEEE/CVF international conference on computer vision*, 2019, pp. 1-11, doi: 10.1109/iccv.2019.00009.

[6] N. Dufour, A. Gully, Contributing data to deepfake detection research, Google AI Blog.

[7] M. Li, B. Liu, Y. Hu, and Y. Wang, "Exposing deepfake videos by tracking eye movements," in *2020 25th International Conference on Pattern Recognition (ICPR)*, 2021, pp. 5184-5189, doi: 10.1109/icpr48806.2021.9413139.

[8] P. Korshunov and S. Marcel, "Deepfakes: A new threat to face recognition? assessment and detection," arXiv.org, doi.org/10.48550/arXiv.1812.08685 (accessed Jun. 15, 2023).

[9] Y. Li, X. Yang, P. Sun, H. Qi, and S. Lyu, "Celeb-DF: A new dataset for Deepfake Forensics," 27-Sep-2019. arXiv.org, doi.org/10.48550/arXiv.1812.08685 (Accessed Apr. 30, 2023).

[10] Y. Li, X. Yang, P. Sun, H. Qi, and S. Lyu, "Celeb-DF: A large-scale challenging dataset for deepfake forensics," in textitProceedings of the IEEE/CVF conference on computer vision and pattern recognition, 2020, pp. 3207-3216, doi: 10.1109/CVPR42600.2020.00327.

[11] B. Dolhansky, R. Howes, B. Pflaum, N. Baram, and C. C. Ferrer, "The deepfake detection challenge (DFDC) preview dataset," arXiv.org, doi.org/10.48550/arXiv.1910.08854 (Accessed Apr. 30, 2023).

[12] B. Dolhansky, J. Bitton, B. Pflaum, J. Lu, R. Howes, M. Wang, and C. C. Ferrer, "The deepfake detection challenge (DFDC) dataset," arXiv.org, doi.org/10.48550/arXiv.2006.07397 (Accessed Apr. 30, 2023).

[13] L. Jiang, R. Li, W. Wu, C. Qian, and C. C. Loy, "DeeperForensics-1.0: A large-scale dataset for real-world face forgery detection," in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 2020, pp. 2889-2898, doi: 10.1109/cvpr42600.2020.00296.

[14] B. Zi, M. Chang, J. Chen, X. Ma, and Y.-G. Jiang, "WildDeepfake: A challenging real-world dataset for deepfake detection," Proceedings of the 28th ACM International Conference on Multimedia, 2020. doi:10.1145/3394171.3413769.

[15] P. Kwon, J. You, G. Nam, S. Park, and G. Chae, "KODF: A large-scale Korean deepfake detection dataset," in *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 2021, pp. 10744-10753, doi: 10.1109/iccv48922.2021.01057.

[16] Y. He *et al.*, "ForgeryNet: A versatile benchmark for comprehensive forgery analysis," *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 2021, pp. 4360-4369, doi: 10.1109/cvpr46437.2021.00434.

[17] M. Taeb and H. Chi, "Comparison of deepfake detection techniques through deep learning," *Journal of Cybersecurity and Privacy*, vol. 2, no. 1, pp. 89–106, 2022. doi: 10.3390/jcp2010007.

[18] L. A. Passos, D. Jodas, K. A. P. da Costa, L. A. S. Júnior, D. Colombo, and J. P. Papa, "A review of deep learning-based approaches for deepfake content detection," arXiv.org, doi.org/10.48550/arXiv.2202.06095 (Accessed Apr. 30, 2023).

[19] Y. Mirsky and W. Lee, "The creation and detection of deepfakes: A survey,“ *ACM Computing Surveys (CSUR)*, vol 54, no 1, pp 1–41, 2021, doi: 10.1145/3425780.

[20] J. Thies, M. Zollhofer, M. Stamminger, C. Theobalt, and M. Niessner, "Face2Face: Real-time face capture and reenactment of RGB videos," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016, pp. 2387-2395, doi: 10.1109/cvpr.2016.262.

[21] J. Thies, M. Zollhöfer, and M. Nießner, "Deferred neural rendering: Image synthesis using neural textures," *ACM Transactions on Graphics*, vol. 38, no. 4, pp. 1–12, 2019, doi: 10.1145/3306346.3323035.

[22] Y. Choi, M. Choi, M. Kim, J.-W. Ha, S. Kim, and J. Choo, "Stargan: Unified generative adversarial networks for multi-domain image-to-image translation," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2018, pp. 8789-8797.

[23] D. Afchar, V. Nozick, J. Yamagishi and I. Echizen, "MesoNet: A compact facial video forgery detection network," *2018 IEEE International Workshop on Information Forensics and Security (WIFS)*, Hong Kong, China, 2018, pp. 1-7, doi: 10.1109/WIFS.2018.8630761.

[24] A. Mitra, S. P. Mohanty, P. Corcoran, and E. Kougianos, "A machine learning based approach for deepfake detection in social media through key video frame extraction," *SN Computer Science*, vol. 2, no. 2, pp. 1-18, 2021. doi: 10.1007/s42979-021-00495-x.

[25] A. Chintha *et al*., "Recurrent convolutional structures for audio spoof and Video deepfake detection," in *IEEE Journal of Selected Topics in Signal Processing*, vol. 14, no. 5, pp. 1024–1037, 2020, doi: 10.1109/JSTSP.2020.2999185.

[26] I. Amerini, L. Galteri, R. Caldelli and A. Del Bimbo, "Deepfake video detection through optical flow based CNN," in *Proceedings of the IEEE/CVF international conference on computer vision workshops*, 2019, pp. 1205-1207, doi: 10.1109/ICCVW.2019.00152.

[27] X. Sun, B. Wu, and W. Chen, "Identifying invariant texture violation for robust deepfake detection," arXiv.org, doi.org/10.48550/arXiv.2012.10580 (Accessed Apr. 30, 2023).

[28] M. S. Rana and A. H. Sung, "DeepfakeStack: A deep ensemble-based learning technique for deepfake detection," *2020 7th IEEE International Conference on Cyber Security and Cloud Computing (CSCloud)/2020 6th IEEE International Conference on Edge Computing and Scalable Cloud (EdgeCom)*, New York, NY, USA, 2020, pp. 70-75, doi: 10.1109/CSCloud-EdgeCom49738.2020.00021.

[29] P. Saikia, D. Dholaria, P. Yadav, V. Patel, and M. Roy, "A hybrid CNN-LSTM model for video deepfake detection by Leveraging Optical Flow features," in *2022 International Joint Conference on Neural Networks (IJCNN)*, 2022, pp. 1-7, doi: 10.1109/ijcnn55064.2022.9892905.

[30] T. Wang, H. Cheng, K. P. Chow, and L. Nie, "Deep convolutional pooling transformer for deepfake detection," *ACM Transactions on Multimedia Computing, Communications and Applications*, vol. 19, no. 6, pp. 1–20, 2023, doi: 10.1145/3588574.

[31] S. Ge *et al*., "Deepfake video detection via predictive representation learning," *ACM Transactions on Multimedia Computing, Communications, and Applications (TOMM)*, vol. 18, no. 2s, pp. 1–21, 2022, doi: 10.1145/3536426.

[32] N. Wu *et al*., "Multisemantic path neural network for deepfake detection," *Security and Communication Networks*, 2022, doi: 10.1155/2022/4976848.

[33] L. Kuang, Y. Wang, T. Hang, B. Chen, and G. Zhao, "A dual-branch neural network for deepfake video detection by detecting spatial and temporal inconsistencies," *Multimedia Tools and Applications*, vol. 81, no. 29, pp. 42591–42606, 2022, doi: 10.1007/s11042-021-11539-y.

[34] A. Khormali and J.-S. Yuan, "DFDT: An end-to-end deepfake detection framework using vision transformer," *Applied Sciences*, vol. 12, no. 6, p. 2953, 2022, doi: 10.3390/app12062953.

[35] Z. Liu, H. Wang, and S. Wang, "Cross-domain local characteristic enhanced deepfake video detection," in *Proceedings of the Asian Conference on Computer Vision*, 2023, pp. 3412-3429, doi: 10.1007/978-3-031-26348-4_12.

[36] Y. Xu, K. Raja, L. Verdoliva, and M. Pedersen, "Learning pairwise interaction for generalizable deepfake detection," in *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision*, 2023, pp. 672-682, doi: 10.1109/wacvw58289.2023.00074.

[37] C.-M. Yu, K.-C. Chen, C.-T. Chang, and Y.-W. Ti, "SegNet: A network for detecting deepfake facial videos," *Multimedia Systems*, vol. 28, no. 3, pp. 793–814, 2022, doi: 10.1007/s00530-021-00876-5.

[38] X. H. Nguyen, T. S. Tran, V. T. Le, K. D. Nguyen, and D.-T. Truong, "Learning spatio-temporal features to detect manipulated facial videos created by the Deepfake Techniques," *Forensic Science International: Digital Investigation*, vol. 36, p. 301108, 2021, doi: 10.1016/j.fsidi.2021.301108.

[39] X. Li *et al*., "Fighting against deepfake: Patch&pair convolutional neural networks (PPCNN)," in *Companion Proceedings of the Web Conference 2020*, 2020, pp/ 88-89, doi: 10.1145/3366424.3382711.

[40] L. Bondi, E. Daniele Cannas, P. Bestagini, and S. Tubaro, "Training strategies and data augmentations in CNN-based deepfake video detection," *2020 IEEE international workshop on information forensics and security (WIFS)*, 2020, pp. 1-6, doi: 10.1109/wifs49906.2020.9360901.

[41] Z. Gu *et al*., "Delving into the local: Dynamic inconsistency learning for DeepFake video detection," in *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 36, no. 1, pp. 744–752, 2022, doi: 10.1609/aaai.v36i1.19955.

[42] S. A. Khan and D.-T. Dang-Nguyen, "Hybrid transformer network for deepfake detection," in *Proceedings of the 19th international conference on content-based multimedia indexing*, 2022, pp. 8-14, doi: 10.1145/3549555.3549588.

[43] Z. Xia *et al*., "Deepfake video detection based on Mesonet with preprocessing module," *Symmetry*, vol. 14, no. 5, p. 939, 2022, doi: 10.3390/sym14050939.

[44] Y. Li and S. Lyu, "Exposing deepfake videos by detecting face warping artifacts," arXiv.org, 22-May-2019. [Online]. Available: doi.org/10.48550/arXiv.1811.00656 (Accessed Apr. 30, 2023).

[45] R. Tolosana, S. Romero-Tapiador, J. Fierrez, and R. Vera-Rodriguez, "Deepfakes evolution: Analysis of facial regions and fake detection performance," in *international conference on pattern recognition*, Cham: Springer International Publishing, pp. 442–456, 2021, doi: 10.1007/978-3-030-68821-9_38.

[46] M. A. Younus and T. M. Hasan, "Effective and fast deepfake detection method based on Haar wavelet transform," in *2020 International Conference on Computer Science and Software Engineering (CSASE)*, 2020, 186-190, doi: 10.1109/csase48920.2020.9142077.

[47] A. Elhassan, M. Al-Fawa'reh, M. T. Jafar, M. Ababneh, and S. T. Jafar, "DFT-MF: Enhanced deepfake detection using mouth movement and transfer learning," *SoftwareX*, vol. 19, p. 101115, 2022, doi: 10.1016/j.softx.2022.101115.

[48] L. Trinh, M. Tsang, S. Rambhatla, and Y. Liu, "Interpretable and trustworthy deepfake detection via dynamic prototypes," in *Proceedings of the IEEE/CVF winter conference on applications of computer vision*, 2021, pp. 1973-1983, doi: 10.1109/wacv48630.2021.00202.

[49] K. Zhu, B. Wu, and B. Wang, "Deepfake detection with clustering-based embedding regularization," in *2020 IEEE fifth international conference on data science in cyberspace (DSC)*, 2020, pp. 257-264, doi: 10.1109/dsc50466.2020.00046.

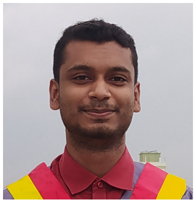
[50] M. T. Jafar, M. Ababneh, M. Al-Zoube and A. Elhassan, "Forensics and analysis of deepfake videos," *2020 11th International Conference on Information and Communication Systems (ICICS)*, Irbid, Jordan, 2020, pp. 053-058, doi: 10.1109/ICICS49469.2020.239493.

[51] T. Mittal, U. Bhattacharya, R. Chandra, A. Bera, and D. Manocha, Emotions don't lie: An audio-visual deepfake detection method using affective cues," in *Proceedings of the 28th ACM international conference on multimedia*, 2020, pp. 2823-2832, doi: 10.1145/3394171.3413570.

[52] P. Charitidis, G. Kordopatis-Zilos, S. Papadopoulos, and I. Kompatsiaris, "Investigating the impact of pre-processing and prediction aggregation on the Deepfake Detection Task," arXiv.org, doi.org/10.48550/arXiv.2006.07084 (Accessed Apr. 30, 2023).

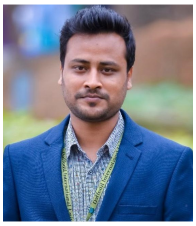
[53] A. A. Maksutov, V. O. Morozov, A. A. Lavrenov and A. S. Smirnov, "Methods of deepfake detection based on machine learning," *2020 IEEE Conference of Russian Young Researchers in Electrical and Electronic Engineering (EIConRus)*, St. Petersburg and Moscow, Russia, 2020, pp. 408-411, doi: 10.1109/EIConRus49466.2020.9039057.
[54] D. H. Choi, H. J. Lee, S. Lee, J. U. Kim and Y. M. Ro, "Fake video detection with certainty-based attention network," *2020 IEEE International Conference on Image Processing (ICIP)*, Abu Dhabi, United Arab Emirates, 2020, pp. 823-827, doi: 10.1109/ICIP40778.2020.9190655.
[55] D. M. Montserrat *et al.*, "Deepfakes detection with automatic face weighting," in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition workshops*, 2020, pp. 668-669, doi: 10.1109/cvprw50498.2020.00342.
[56] X. Li *et al.*, "Sharp Multiple Instance Learning for deepfake video detection," in *Proceedings of the 28th ACM international conference on multimedia*, 2020, pp. 1864-1872, doi: 10.1145/3394171.3414034.
[57] A. Singh, A. S. Saimbhi, N. Singh, and M. Mittal, "Deepfake video detection: A time-distributed approach," in *SN Computer Science*, vol. 1, no. 4, p. 212, 2020, doi: 10.1007/s42979-020-00225-9.
[58] X. Liu, Y. Yu, X. Li, Y. Zhao, and G. Guo, "TCSD: Triple complementary streams detector for comprehensive deepfake detection," *ACM Transactions on Multimedia Computing, Communications, and Applications*, vol. 19, no. 6, pp. 1-22, 2022, doi: 10.1145/3558004.
[59] Y.-J. Heo, W.-H. Yeo, and B.-G. Kim, "Deepfake detection algorithm based on improved Vision Transformer," *Applied Intelligence*, vol. 53, no. 7, pp. 7512–7527, 2022, doi: 10.1007/s10489-022-03867-9.
[60] T. Zhao *et al.*, "Learning self-consistency for deepfake detection," in *Proceedings of the IEEE/CVF international conference on computer vision*, 2021, pp. 15023-15033, doi: 10.1109/iccv48922.2021.01475
[61] M. S. Rana, M. N. Nobi, B. Murali and A. H. Sung, "Deepfake detection: A systematic literature review," in *IEEE Access*, vol. 10, pp. 25494-25513, 2022, doi: 10.1109/ACCESS.2022.3154404.

BIOGRAPHIES OF AUTHORS



Md Fahimuzzman Sohan received the B.Sc. degree in Software Engineering from the Daffodil International University, Bangladesh, in 2019. He has published several articles in reputed journals and conferences. His research interests include machine learning, computer vision, and image processing. He can be contacted at email: fahimsohan2@gmail.com.

Md Solaiman received his B.Sc. in Software Engineering from Daffodil International University, Dhaka. He is very keen research aspirant in the area of computer vision, image and video processing, deep learning, and data science. He can be contacted at email: solaiman35-1107@diu.edu.bd.



Md Aumit Hasan received B.Sc. in Software Engineering degree from Daffodil International University, Bangladesh, in 2018. He worked as a Research Assistant at Dept. of Software Engineering at Daffodil International University in 2018 and 2019. Currently, he is working as a Full-Stack Software Engineer at SELISE Digital Platforms expert in Python, JavaScript, Angular, React, Java, C++, and Spring Boot. His research interests include computer vision, machine learning, deep learning, and artificial intelligence. He can be contacted at email: aumitse@gmail.com.