JCBI
Journal of Computing & Biomedical Informatics Volume 04 Issue 02
ISSN: 2710 - 1606 2023

Research Article
https://doi.org/10.56979/402/2023

Frequency-based Deep-Fake Video Detection using Deep Learning Methods

Mubasher H. Malik[1*], Hamid Ghous[1], Salman Qadri[2], Syed Ali Nawaz[3], and Anam Anwar[1]

[1] Department of Computer Science, Institute of Southern Punjab, Multan, Pakistan.
[2]Department of Computer Science, Muhammad Nawaz Sharif University of Agriculture, Multan, Pakistan.
[3]Department of Information Technology, The Islamia University of Bahawalpur, Bahawalpur, Pakistan.
[*]Corresponding Author: Mubasher H. Malik. Email: mubasher@isp.edu.pk

Received: December 29, 2022 **Accepted:** February 21, 2023 **Published:** March 29, 2023.

Abstract: Deep Learning (DL) is an advanced and effective technology widely used in diverse industries, including medical imaging (MI). Data Mining (DM), Image Processing (IP), and Machine Vision (DM). Deep-fake uses DL technology to alter videos to render them indistinguishable from the original humans. The effectiveness of deep-fake has recently obtained significant attention from researchers, and numerous DL-based techniques have been developed to identify deep-fake videos. In this paper, a novel deep-fake video detection method is proposed. The Deep Fake Detection Challenge (DFDC) and Face Forensic datasets were used in the research. In addition, frequency-based frame extraction was conducted on each video during the preprocessing stage. Convolutional Neural Networks (CNN) Long Short-Term Memory (LSTM) - CNN techniques were used to identify fake videos. The LSTM-CNN approach achieved an accuracy of 82%. To identify fake videos using DL techniques, this work will be helpful to researchers.

Keywords: Deep Fake; Deep Learning; CNN; LSTM; Fake Videos.

1. Introduction

Our facial features are our identity; someone can recognize us by our faces. As a result, when image and video forgery appear to explode, face manipulation becomes the focus. Face manipulation has grown considerably in the last two decades (Hashemifard, 2021), (Bindemann, 2020), (Werner, 2013). The words "deep learning" and "fake" are combined to form the term "deep fake." (Westerlund, 2019), (Nguyen, 2022). Deep-fake alters a person's identity by swapping their face for another person's in an already-existing video or image. With the powerful DL approach, anyone can easily develop high-potential fake content (Agarwal, 2020), (Suganthi, 2022) . Deep fakes have gained widespread attention because of their illegal use in making fake content like fake news and used for malicious purposes (Helmus, 2022). Face recognition is also employed in our daily lives for various applications. Face attendance, phone authentication systems, and ace payment are just a few biometric applications that use facial recognition technologies (Imaoka, 2021). As a result, it provides material for blackmail, which presents an inevitable threat to our society. Massive advancements in face identification methods also pave the way for a number of face manipulation applications that might be accidentally utilized maliciously (Gan, 2017), (Gangarapu, 2022).

Thus, creating efficient solutions against these sorts of forgeries attacks is essential to reduce the adverse influence on public and private security (Dang, 2020). Face matches can be avoided by using adversarial face attack, which creates high-quality, undetectable adversarial images. The variation autoencoder and GANs, which generate a whole or partial photorealistic facial image, can be used to launch digital manipulation attacks (Tolosana, 2022).

According to recent studies, deep fake videos and images are being widely shared on social media. So, it has become more crucial and significant to detect deepfake videos and images (Dasilva, 2021). Many

companies, including Facebook Inc., Google, and the United States Defense Advanced Research Projects Agency (DARPA), initiated a research program to help detect and stop deep fake content to entice researchers (Chesney, 2019). Deep fakes have attracted considerable attention due to their illegal usage in creating fake content, such as fake news, and for harmful objectives (Karnouskos, 2020). This work contributes to the design of a novel frequency-based deep-fake video detection using DL methods such as CNN and LSTM+ CNN. The Viola-Jones algorithm is used for the detection of faces from videos. OpenCV library extracts each 1/5 extracted facial frame from video datasets based on frequency. The proposed method is generalized on the most recent and challenging face Forensic++ (FF++) (Ma, 2015) and Deep Fake Detection Challenge (DFDC) (Dolhansky, 2019) datasets.

2. Literature Review

This section describes the contribution of researchers in the area of deep fake video detection. This contains complete detail of techniques and approaches adopted by researchers for the detection of fake videos.

In 2023, a deep fake video detection system was proposed. Video and audio frames were extracted. For extracting features XceptionNet model and modified InceptionResNetVw model were adopted. Features extracted using both methods were fused to produce bio-modal information-based feature representation. FakeAVCeleb dataset was adopted to detect forgery from video, audio, or both (Elpeltagy, 2023). In 2022, a Deep Fake Predictor approach was proposed. The model was based on VGG16 and CNN. The benchmark dataset of fake and real images was taken for experiments. The proposed model achieved 94% accuracy. (Raza, 2022). In 2022, a DL-based hybrid model for detecting deep fake videos was proposed. Multilayer perception is used to learn the differences between real and fake videos. CNN was used to extract features. DFDC and dessa datasets were taken for experiments. The proposed model achieved 87% accuracy (Kolagati, 2022).

In 2021, a DL-based model for detecting fake videos was proposed. XGBoost was adopted for detection. CelebDF and FF++ datasets were used for experiments. YOLO face detector, CNN, and Inception Resnet was used for face area extraction. The proposed model produced 80% accuracy (Ismail, 2021). In 2021, a deep fake video detection model was proposed. CNN was deployed for the detection of fake videos. FF++ and DFDC datasets were adopted for experiments. The model achieved 97.6% accuracy using the FF++ dataset (Zhao, 2021) . In 2020, a DL-based deep fake video classification model was proposed. Mobile Net and Xception techniques were deployed. An experiment was performed on the FF++ dataset. The proposed model achieved 91% accuracy (Pan, 2020). In 2020, a multimodal DL technique for deep fake video detection was proposed. The FDCD dataset was taken for experiments. LSTM was adopted in the multimodal network, which achieved 61% accuracy (Lewis, 2020).

In 2020, a DL-based deep fake image detection model was proposed. Experiments were performed using the CelebA dataset. GAN was adopted for pairing fake and real images. Dense Net and Fake Feature Network (FFN) detected fake images. The proposed model achieved 90% accuracy (Hsu, 2020). In 2018, a deep fake video detection model was proposed. Frame-level features were extracted using CNN. Recurrent Neural Network (RNN) was trained using features extracted by CNN. The model produced promising results on videos collected from websites (Guera, 2018).

The next section describes the proposed model for the detection of fake videos. It contains preprocessing steps along with the detection phase.

3. Materials and Methods

This section contains the details of the frequency-based deep fake detection model. Firstly, FF++ and DFDC datasets were taken for input. Secondly, videos were manipulated for extraction of frequency-based frames. These image frames were resized accordingly and converted into grayscale. Finally, CNN and LSTM+CNN were adopted for the detection of fake videos. The experimental results showed that LSTM-CNN produced promising results as compared to CNN. The proposed model is depicted in (Figure 1).

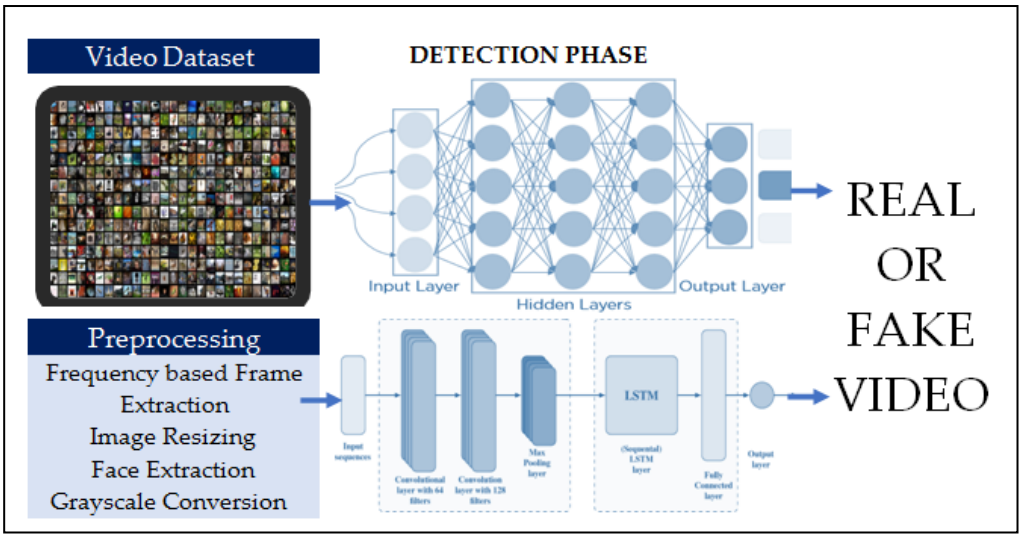

Figure 1. Frequency-based Deepfake Video Detection Model

3.1. Dataset

Face Forensic++ (FF++), and DFDC (Wodajo, 2021) datasets were taken for experiments. FF++ is a facial forgery dataset containingorged facial videos (Rossler, 2019). This dataset has four automated face manipulation methods known as "Deep-fakes," "Face2Face", "Neural texture," and "Face swap" (Ramachandran, 2021). From these four methods, there are two Computer graphics-based approaches (Face2Face and Face swap) and two learning-based approaches (Deep-fakes and neural texture) (Xu Y. a., 2022), (Xu B. a., 2021). FF++ consists of 1000 original video sequences manipulated from these four face-manipulated methods and 1000 veritable videos (R{\"o}ssler, 2018). The resolutions of these videos are 480p, 720p, and 1080p.

3.2. Preprocessing

In preprocessing, Firstly, all videos were used to extract frames based on frequency. Open Source Computer Vision (OpenCV) is an open-source image and video analysis library. (Culjak, 2012). OpenCV was used to extract frequency-based frames from the videos of FF++ and DFDC datasets. Each 1/5th key frame fps is selected for further preprocessing to avoid data redundancy. After saving frames number of operations can be performed on these frames. These frames are used for the face extraction process. Secondly, all extracted frames were resized for a standard image size of 256x256. Thirdly, faces were extracted from extracted frames. This face extraction is performed by the Viola-Jones algorithm (Wang, 2014), (Vikram, 2017). Finally, all images were converted into a grayscale images.

3.3. Detection Phase

In the detection phase, CNN and LSTM+CNN are deployed to detect real or fake videos. CNN is mainly used for object classification and image recognition. It is an effective tool for analyzing pattern recognition problems, which learn spatial hierarchies of features from low to high-level patterns.CNN comprises multiple layers of artificial neurons, each with different activation functions passed to the next layers. Three convolutional layers, activational layers, and two pooling layers (used as optional) are followed by a fully connected layer, and one output layer is used. The deep learning CNN model used for training and testing purposes takes each input image and passes it through convolutional layers with kernels. A common size of these kernels is 3*3 or 5*5. In CNN Conv2 layer 32, the kernel size is (3*3), followed by the MaxPool2D layer. After this, Conv2 layer 64 with kernel size (3*3) followed by MaxPoll2D layer and ReLU activation function. In the CNN model, after this, Conv2 layer 128 with kernel size (3*3) followed by MaxPoll2D layer and ReLU activation function and followed by flattened and Dense layers with a sigmoid activation function as shown in (Figure 2). Then its pooling layer is used for basic feature extraction like horizontal and diagonal edges. And after that fully connected layer maps these features into the final output for more complex extraction like corners or combinational edges. As we work deeper into this network, it can also detect more complex features like objects and faces. In the end, the Softmax function is applied to classify this object, probably using a classification layer with 0 or 1 (Afridi, 2021), (Sustika, 2018). The architecture of CNN is depicted in (Figure 2).

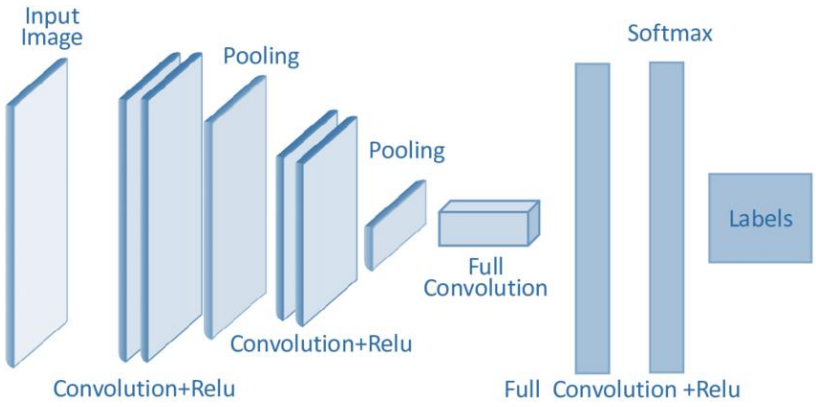

Figure 2. Convolutional Neural Networks (CNN) Architecture (Zarandy, 2015)

LSTM-CNN is a CNN-based method specially designed for sequence prediction, like videos and images. In the LSTM-CNN model, CNN layers are used for feature extraction, and then these features are combined with the LSTM network with input data for prediction. In the LSTM-CNN model, add LSTM 30 layer with ReLU activation function, followed by Dens 100 layer and ReLU activation function. Then the model adds Dens 10 layer with Softmax activation function, followed by CNN model layers. Conv2D 128 layers with kernel size (3*3) followed by MaxPool2D layer *(Livieris, 2020)* as shown in (Figure 3).

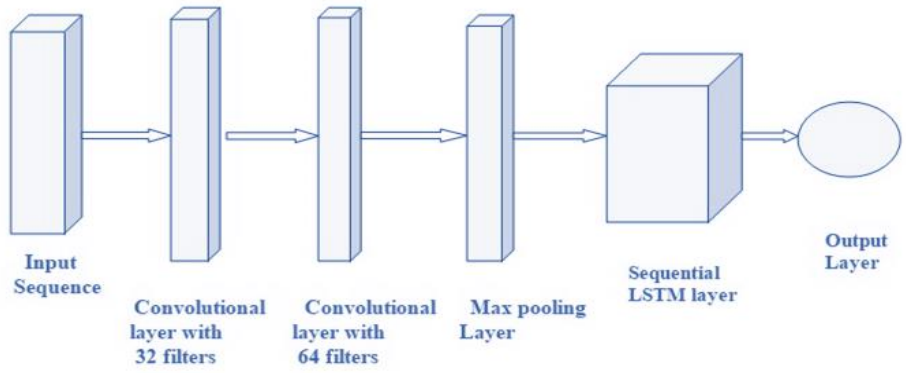

Figure 3. LSTM-CNN Architecture (Islam, 2020)

4. Results

The FF++ dataset consists of two thousand videos, from which one thousand are real, and one thousand are fake. These videos were converted into frames based on frequency. One hundred Two thousand two hundred and sixteen images were used for experiments. Firstly, the whole dataset splits into a training-testing ratio of 80%-20%. Eighty-one Thousand Seven Hundred Seventy-two real images were used for training, and twenty-four hundred forty-four were used for testing. The model's performance in terms of accuracy is exhibited in (Table 1). LSTM-CNN produced 82% accuracy, while CNN produced 75% accuracy. Secondly, the whole dataset splits into a training-testing ratio of 70%-30%. Seventy-one thousand five hundred and fifty-one images were taken for training, and Thirty thousand six hundred sixty-five were taken for testing. The model's performance in terms of accuracy is shown in (Table 1). CNN produced 82.0% accuracy, while LSTM-CNN produced 66.0% accuracy. LSTM-CNN produced higher accuracy on an 80-20% training-testing split, while CNN produced higher accuracy on a 70-30%.

Table 1. Accuracy results based on the FF++ dataset using 80-20% and 70-30% split.

Method	Training – Testing Ratio 80%-20% Accuracy	Training – Testing Ratio 70%-30% Accuracy
CNN	75.0%	82.0%

LSTM-CNN	82.0%	66.0%

The graphical representation of the performance evaluation of the proposed model using the FF++ dataset is shown in (Figure 4). The graph showed the frequency of both CNN and LSTM-CNN methods.

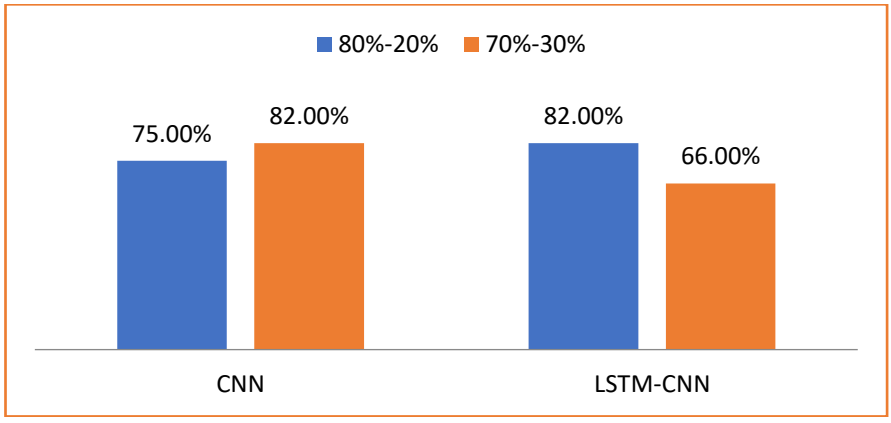

Figure 4. Graphical Representation of FF++ dataset accuracy using CNN and LSTM-CNN

DFDC dataset consists of eight hundred videos, from which four hundred are real, and four hundred are fake. These videos were converted into frames based on frequency. Twenty-four thousand images were used for experiments. Firstly, the whole dataset splits into a training-testing ratio of 80%-20%. Nineteen thousand two hundred images were taken for training, while forty-eight hundred were taken for testing. The model's performance is exhibited in (Table 2) in terms of accuracy. LSTM-CNN produced higher accuracy on an 80%-20% training-testing split. Secondly, the whole dataset splits into a training-testing ratio of 70%-30%. Sixteen thousand eight hundred real and fake videos and seventy-two hundred real and fake images were taken simultaneously. The model's performance is exhibited in (Table 2) in terms of accuracy

Table 2. Accuracy results based on the DFDC dataset using 80-20% and 70-30% split.

Method	**Training – Testing Ratio 80%-20% Accuracy**	**Training – Testing Ratio 70%-30% Accuracy**
CNN	57.0%	75.0%
LSTM-CNN	72.0%	72.0%

The graphical representation of the performance evaluation of the proposed model using the FF++ dataset is shown in (Figure 5). The graph showed the frequency of both CNN and LSTM-CNN methods.

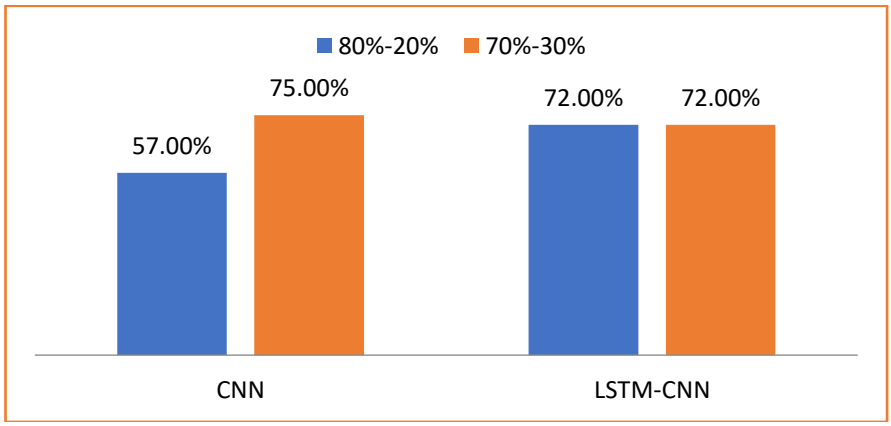

Figure 5. Graphical Representation of DFDC dataset accuracy using CNN and LSTM-CNN

Hence, experiments showed that CNN and LSTM-CNN methods deployed on FF++ and DFDC datasets. Both datasets contain videos that were converted into frames based on frequency. Experiments were performed using an 80%-20% and 70%-30% training-testing split. The results showed that using the FF++ dataset, CNN produced higher accuracy on a 70%-30% split, while on an 80%-20% split, LSTM-CNN performed outstandingly. On the other hand, using the DFDC dataset, CNN produced higher on both 70%-30% and 80%-20% split. Furthermore, the proposed model's comparison with existing techniques is shown in (Table 3).

Table 3. Comparison of the proposed model with existing methods.

Author	Method & Dataset	Accuracy
Aditi Kohli et al. (Mitra, 2021)	CNN & FF++	51.0%
Hadi. M et.al (Mansourifar, 2020)	GAN & DFDC	67.0%
ILKE.D et.al (Demir, 2021)	DNN & FF++	79.0%
Our Proposed Model	LSTM-CNN & FF++	82.0%
Our Proposed Model	CNN & DFDC	72.0%

5. Conclusions

Deep-fake becomes popular due to the quality of tampered videos and the easy-to-use ability of these applications with the number of users with minimal computer skills from professional to novice. Face images contain rich personal identity information and weak privacy, making them easily manipulated. And deep fake is used widely to change a person's identity, which can cause distress and negative effects on those targeted persons. This is becoming critical nowadays as the techniques for creating deep fakes are increasing daily. And social media platforms are spreading those fake content quickly. Although many face-manipulated detection techniques have been proposed, the issue remains unsolved. In our proposed work, DFDC and FF++ video dataset is used. This targeted video is then split into frames using Opencv on a frequency base. This group of frames is then used for the face extraction procedure using the viola Jones algorithm. Then these extracted face frames are resized and converted into grayscale images. CNN and LSTM-CNN were deployed to identify the forged face from the targeted video for detection purposes. The accuracy of the proposed model CNN with 80-20 splits is achieved at 77.0%, and that of 70-30 splits is 82.0%.

On the other hand, using DFDC higher accuracy of the proposed model on 80-20 splits is 57.0%, and 70-30 splits are 75.0%. The accuracy of the proposed model LSTM-CNN with 80-20 splits achieved on neural texture is 82.0%, and that of 70-30 splits is 66.0%. While on DFDC higher accuracy of the proposed model on 80-20 splits is 72.0%, and 70-30 splits are 75.0%. In the future, there is a need to develop a technique to reduce the features of an image and introduce a robust model for bigger videos dataset.

Funding: Please add: "This research received no external funding."
Conflicts of Interest: "The authors declare no conflict of interest."

References

1. Afridi, T. H.-K. (2021). A multimodal memes classification: A survey and open research issues. In Innovations in Smart Cities Applications Volume 4: The Proceedings of the 5th International Conference on Smart City Applications (pp. 1451--1466). Springer.
2. Agarwal, S. a.-G.-N. (2020). Detecting deep-fake videos from appearance and behavior. 2020 IEEE international workshop on information forensics and security (WIFS), 1--6.
3. Bindemann, M. a. (2020). Understanding face identification through within-person variability in appearance: Introduction to a virtual special issue. Quarterly Journal of Experimental Psychology, 73(12), NP1--NP8.
4. Chesney, B. a. (2019). Deep fakes: A looming challenge for privacy, democracy, and national security. Calif. L. Rev., 107, 1753.
5. Culjak, I. a. (2012). A brief introduction to OpenCV. 2012 proceedings of the 35th international convention MIPRO, 1725--1730.
6. Dang, H. a. (2020). On the detection of digital face manipulation. In Dang, Hao and Liu, Feng and Stehouwer, Joel and Liu, Xiaoming and Jain, Anil K (pp. 5781--5790).
7. Dasilva, J. P. (2021). Deepfakes on Twitter: which actors control their spread? In 9 (Ed.), Media and Communication (1, Trans., p. 301). Cogitatio Press.
8. Demir, I. a. (2021). Where do deep fakes look? synthetic face detection via gaze tracking. ACM Symposium on Eye Tracking Research and Applications, 1--11.
9. Dolhansky, B. a. (2019). The deepfake detection challenge (dfdc) preview dataset. arXiv preprint arXiv:1910.08854.
10. Elpeltagy, M. a. (2023). A Novel Smart Deepfake Video Detection System. (IJACSA) International Journal of Advanced Computer Science and Applications, 407-419.
11. Gan, J. a. (2017). 3d convolutional neural network based on face anti-spoofing. 2017 2nd international conference on multimedia and image processing (ICMIP), 1--5.
12. Gangarapu, K. (2022). Ethics of Facial Recognition: Key Issues and Solutions. Learn. g2. com.
13. Guera, D. a. (2018). Deepfake video detection using recurrent neural networks. In 2018 15th IEEE international conference on advanced video and signal based surveillance (AVSS) (pp. 1--6). IEEE.
14. Hashemifard, S. a. (2021). A compact deep learning model for face spoofing detection. arXiv preprint arXiv:2101.04756.
15. Helmus, T. C. (2022). Artificial Intelligence, Deepfakes, and Disinformation: A Primer.
16. Hsu, C.-C. a.-X.-Y. (2020). Deep fake image detection based on pairwise learning. Applied Sciences, 10(1), 370.
17. Imaoka, H. a. (2021). The future of biometrics technology: from face recognition to related applications. APSIPA Transactions on Signal and Information Processing, 10, e9.
18. Islam, M. Z. (2020). A combined deep CNN-LSTM network for the detection of novel coronavirus (COVID-19) using X-ray images. Informatics in medicine unlocked, 20, 100412.
19. Ismail, A. a. (2021). A new deep learning-based methodology for video deepfake detection using XGBoost. Sensors, 21(16), 5413.
20. Karnouskos, S. (2020). Artificial intelligence in digital media: The era of deepfakes. IEEE Transactions on Technology and Society, 1(3), 138--147.
21. Kolagati, S. a. (2022). Exposing deepfakes using a deep multilayer perceptron--convolutional neural network model. International Journal of Information Management Data Insights, 2(1), 100054.
22. Lewis, J. K.-A. (2020). Deepfake video detection based on spatial, spectral, and temporal inconsistencies using multimodal deep learning. In 2020 IEEE Applied Imagery Pattern Recognition Workshop (AIPR) (pp. 1--9). IEEE.
23. Livieris, I. E. (2020). A CNN--LSTM model for gold price time-series forecasting. Neural computing and applications, 17351--17360.
24. Ma, C. a.-B.-H. (2015). Hierarchical convolutional features for visual tracking. In Proceedings of the IEEE international conference on computer vision (pp. 3074--3082).
25. Mansourifar, H. a. (2020). One-shot gan generated fake face detection. arXiv preprint arXiv:2003.12244.
26. Mitra, A. a. (2021). A machine learning based approach for deepfake detection in social media through key video frame extraction. SN Computer Science, 2(Springer), 1--18.
27. Nguyen, T. T.-T.-V. (2022). Deep learning for deepfakes creation and detection: A survey. Computer Vision and Image Understanding, 233, 103525.
28. Pan, D. a. (2020). Deepfake detection through deep learning. In 2020 IEEE/ACM International Conference on Big Data Computing, Applications and Technologies (BDCAT) (pp. 134--143). IEEE.
29. R{\"o}ssler, A. a. (2018). Faceforensics: A large-scale video dataset for forgery detection in human faces. arXiv preprint arXiv:1803.09179.
30. Ramachandran, S. a. (2021). An experimental evaluation on deepfake detection using deep face recognition. In 2021 International Carnahan Conference on Security Technology (ICCST)} (pp. 1--6). IEEE.
31. Raza, A. a. (2022). A Novel Deep Learning Approach for Deepfake Image Detection. Applied Sciences, 12(19), 9820.
32. Rossler, A. a. (2019). Faceforensics++: Learning to detect manipulated facial images. In Proceedings of the IEEE/CVF international conference on computer vision (pp. 1--11).
33. Suganthi, S. A. (2022). Deep learning model for deep fake face recognition and detection. PeerJ Computer Science, 8, e881.

34. Sustika, R. a. (2018). Evaluation of deep convolutional neural network architectures for strawberry quality inspection. Int. J. Eng. Technol, 7(4), 75--80.
35. Tolosana, R. a.-R.-G. (2022). An introduction to digital face manipulation. In Handbook of Digital Face Manipulation and Detection: From DeepFakes to Morphing Attacks (pp. 3--26). Springer International Publishing Cham.
36. Vikram, K. a. (2017). Facial parts detection using Viola Jones algorithm. 2017 4th international conference on advanced computing and communication systems (ICACCS), 1--4.
37. Wang, Y.-Q. (2014). An analysis of the Viola-Jones face detection algorithm. Image Processing On Line, 4, 128--148.
38. Werner, N.-S. K. (2013). The neuroscience of face processing and identification in eyewitnesses and offenders. Frontiers in behavioral neuroscience, 7, 189.
39. Westerlund, M. (2019). The emergence of deepfake technology: A review. Technology innovation management review, 9(11).
40. Wodajo, D. a. (2021). Deepfake video detection using convolutional vision transformer. arXiv preprint arXiv:2102.11126.
41. Xu, B. a. (2021). DeepFake videos detection based on texture features. CMC-COMPUTERS MATERIALS \& CONTINUA, 68(1), 1375--1388.
42. Xu, Y. a. (2022). When Handcrafted Features and Deep Features Meet Mismatched Training and Test Sets for Deepfake Detection. arXiv preprint arXiv:2209.13289.
43. Zarandy, A. a. (2015). Overview of CNN research: 25 years history and the current trends. In 2015 IEEE International Symposium on Circuits and Systems (ISCAS) (pp. 401--404). IEEE.
44. Zhao, H. a. (2021). Zhao, H.; Zhou, W.; Chen, D.; Wei, T.; Zhang, W.; Yu, N. Multi-attentional deepfake detection. In Proceedings of the IEEE/CVF. In Proceedings of the IEEE/CVF conference on computer vision and pattern recognition (pp. 2185--2194).