https://doi.org/10.33472/AFJBS.6.13.2024.1604-1617

AFJBS

African Journal of Biological Sciences

Journal homepage: http://www.afjbs.com

Research Paper Open Access

A REVIEW ON DETECTION OF REAL AND FAKE HUMAN FACES USING DEEP LEARNING TECHNIQUES

Goolla Mamatha[1], Dr. Rajchandar K[2]

[1]Research Scholar,
[2]Assistant Professor & Supervisor,

Email: [1]mamatha1993.phd@gmail.com, [2]rajchandark@gmail.com

Article Info

Volume 6, Issue 13, July 2024

Received: 02 June 2024

Accepted: 30 June 2024

Published: 24 July 2024

doi: 10.33472/AFJBS.6.13.2024.1604-1617

ABSTRACT:

The rapid growth of generative models and deep learning has raised concerns about "deepfakes," which are synthetic media that closely resembles real-world information. These media assets are subject to restrictions in a variety of fields, including entertainment, politics, and safety. As a result, there has been a lot of interest in designing systems to detect deepfakes. Numerous experts have developed a wide range of two-classification-based algorithms for identifying Deepfakes. This article gives a thorough examination of the most current advances in deepfake detection techniques. It gives a thorough examination of the underlying theories, the datasets utilized for training and testing, and the current issues that this rapidly developing profession faces. Significant scholarly work has been committed to investigating alternative techniques to tackling the issue that Deepfake raises. To assist research, the methodologies are divided into four categories: multimodality-based DFDT, video-based DFDT, image-based DFDT, and audio-based DFDT. Deepfake photos and movies were shared on social media sites using cutting-edge techniques that we developed. Scholars and researchers specializing in this discipline will be captivated by these methods. Several datasets, such as FaceForensics++, the DeepFake Detection Challenge (DFDC), and Celeb-DF, are used to test and improve deepfake detection systems. Discuss the model's features, adjustments, and constraints, highlighting the importance of using a wide range of real-world datasets to ensure its applicability across multiple situations.

Keywords: Deepfake, Artificial Intelligence, Machine Learning, Deep Learning, CNN, RNN, Multimodality, Forensics, Cyber Security.

1. INTRODUCTION

Artificial intelligence and deep learning enable intricate media manipulation. These alterations are known as "deepfakes." Deepfakes are images, films, and sounds created with powerful machine learning algorithms. It is not uncommon to manufacture fake news; DeepFake use machine learning to edit or create audio and video content that deceives people. Machine learning and deep learning are used to train autoencoders and GANs to create deepfakes. This is a long-standing issue. Historically, digital data had little impact on document authentication. The only permitted activities were paper review, verification, and corrections. It is impossible to ignore the daily exponential growth of digital data on the Internet, as well as its disruptive impact on health graphics, digital advertising, legal forensic imaging, and sensitive satellite image processing. Cybercrime is increasing as a result of the proliferation of digital data applications.

People have long been captivated by the alteration of music, pictures, and movies. Photoshop makes image editing simple. Changing the sounds and pictures is challenging. Frame by frame, video records are manipulated to create new cinematic effects. Personal computers have the ability to influence film images. Machine learning algorithms and artificial intelligence (AI) have enabled novel audio, video, and image editing techniques. The first evidence of this occurred when malicious Face Swap Machine Learning algorithms replaced renowned faces in inappropriate videos.

A satirical film cautions the US president about Deepfake techniques. Many people refer to visual and audio media that have undergone extensive deep learning-based processing as "deepfake." There are numerous methods for creating computer images that seem like photographs. Professionals can create these photographs and films, but tools make it simple for anyone to produce phony media. Non-human techniques can detect false material even in the absence of labeled data. Certain models can distinguish between false and true information better than CNNs and other trained models. Certain models using simple DNN and GAN offer higher accuracy, however they can only process one or two datasets. Deepfake's annual report highlights Deep Learning researchers' major generative modeling work. Computer vision experts propose Face2Face for facial repair. This technology immediately transforms a person's facial expressions into a computer-generated avatar.Cycle GAN, developed at UC Berkeley, changes images and videos.

Fig. 1. Understanding of Deep Fake.

Individuals use machine learning to identify media and spread false news, as shown in Figure 1. Deepfake is used to represent cybercrime. Recent news reports have revealed routine Deepfake exploits used to earn revenue, commit crimes, and achieve other goals. A Bloomberg writer named Maisy Kinsley, who is active on LinkedIn and Twitter, is most likely a Deepfake. Her image appears to have been generated with a computer. Given Maisy Kinsley's persistent efforts to communicate with Tesla stock naysayers, it appears that the goal of her social media account is to collect revenue. Smartphones, laptops, and cameras are becoming increasingly popular. As a result, unrestricted photo and video content has risen. In the recent decade, social media has enabled more people to have quick access to a diverse range of media. This facilitates the creation and distribution of multimedia. Grover, developed by the Allen Institute for Artificial Intelligence, is an online fake news and textual Deepfake detector. Researchers stated that the software detected deepfake words 92% of the time. Grover makes use of Common Crawl's free, open-source web crawler and archive test suites. Experts from Harvard and the MIT-IBM Watson Laboratory created the Giant language model test room website to assess the precision of AI-generated text.

Motivation

Images communicate more than words, and throughout the early digital age, everyone used photographs to promote themselves. It has already been deployed in several sites. Various GAN-based machine learning techniques are being used to create false photos. Because legitimate and counterfeit images cannot be distinguished in databases, this is precisely the goal. Several research institutes, government agencies, and well-known IT corporations have achieved substantial advances in multimedia forensics over the previous fifteen years. In 2016, the US Department of Defense's DARPA conducted an extensive digital forensic study to improve media integrity research. The approach and standard dataset yielded remarkable results. Originating in the Medi Digital media validation ensures the accuracy of digital, physical, and semantic data for taxonomy. Deep learning algorithms are clearly superior. They are displacing the bulk of technologies as large IT firms and institutions rapidly adopt them.

Adversarial generative networks and autoencoders may produce deepfake photos and movies using computer vision and deep learning. Individuals with ill intent or a poor knowledge of machine learning can distort data. This may produce an image or video that is indistinguishable to both humans and automated systems. It's tough to understand how easily this modern technology may be used in 2018 to spread falsehoods, mimic politicians, and insult innocent people. Certain "deepfakes" are now significantly enhanced.

Evaluation of Deep Fake techniques

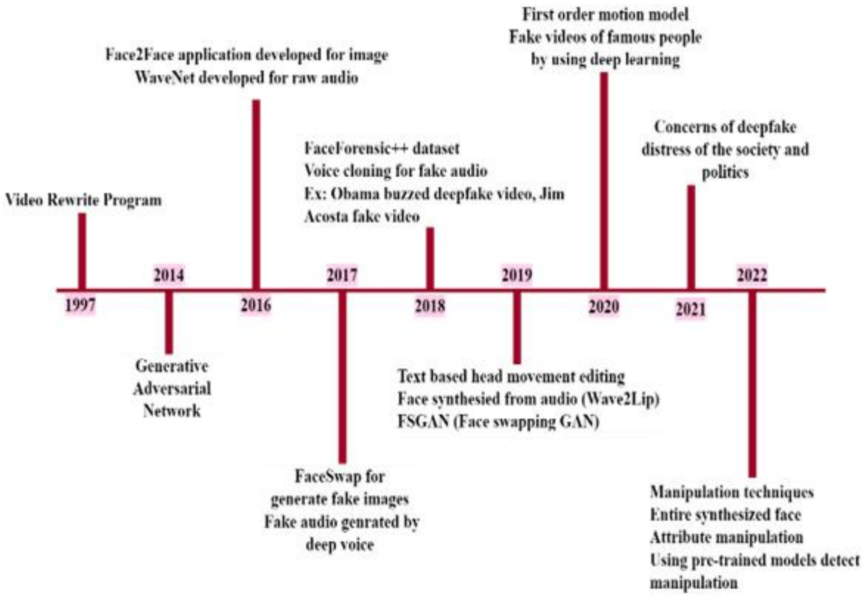

Fig. 2. Evaluation of Deep Fake detection techniques.

Figure 2 shows the timeline of the DFD approach. The DFDT explanation includes the progression and current deepfake detection technologies. For generations, photography has featured foraged or digitally changed images.

2. LITERATURE REVIEW

This section will compare and explain tactics so that you may comprehend their actual use. Recent surveys, datasets, and models were also discussed throughout the discussion. This section explains deepfake and categorization with artificial intelligence and different neural network designs.

Process of DeepFake detection

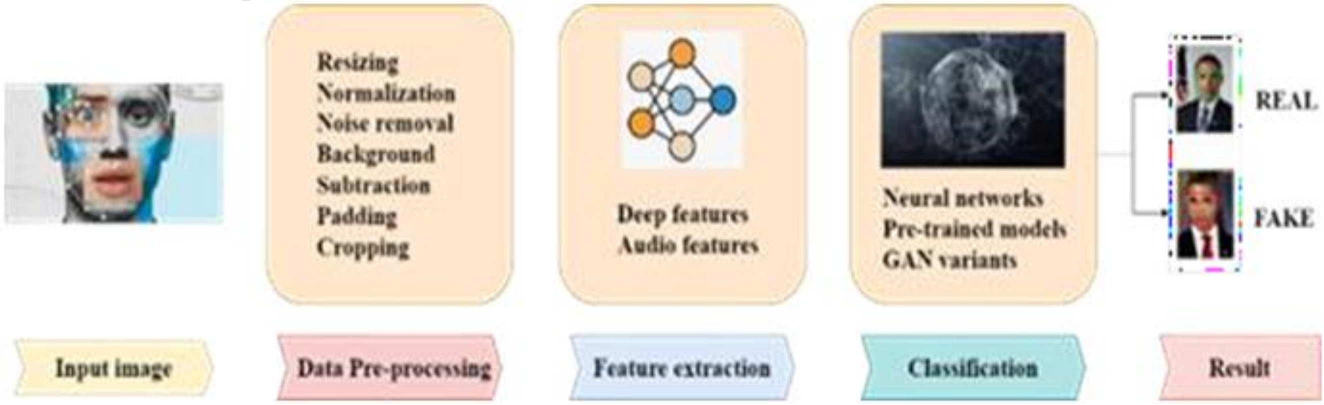

Fig.3. Generalized diagram of DeepFake detectiontechnique

Figure 3 illustrates a full Deepfake detection pipeline. The majority of deepfake detection algorithms rely on feature extraction based on deep learning or manually created features. There are several ways that mix meticulously built components with visual and audible data to aid manipulation.

Data Pre-processing

Age, height, and weight are some of the feature variables used in exploratory data analysis. It is important to normalize and compress visual information. Building a model needs several data preprocessing steps. It is possible to recognize any sort of data, including text, numerical, structured, unstructured, audio, PNG/JPG images, and time series. Photo processing is used to capture, process, segment, collect, and categorize photos of medical equipment, cars, fruits, and digital text, among other things. Fundamental data preparation includes feature normalization, imputation, encoding (including one-hot encoding), engineering, and selection. This may necessitate using dimensionality reduction techniques.

Enhancing a photograph may result in better image quality. This strategy is commonly used. Illumination degrades contrast, brightness, and noise in satellite and digital photos. Image improvement improves fuzzy and noisy photos by reducing noise and sharpening features. The current study looks at histogram equalization, text photo and ID number sharpening, and noise reduction for salt and pepper distortions. Data that has been properly structured is easier to analyze. Deepfake's preprocessing approaches include changing characteristics and inverting words and identities. Figure 4 depicts the procedures for data pre-processing.

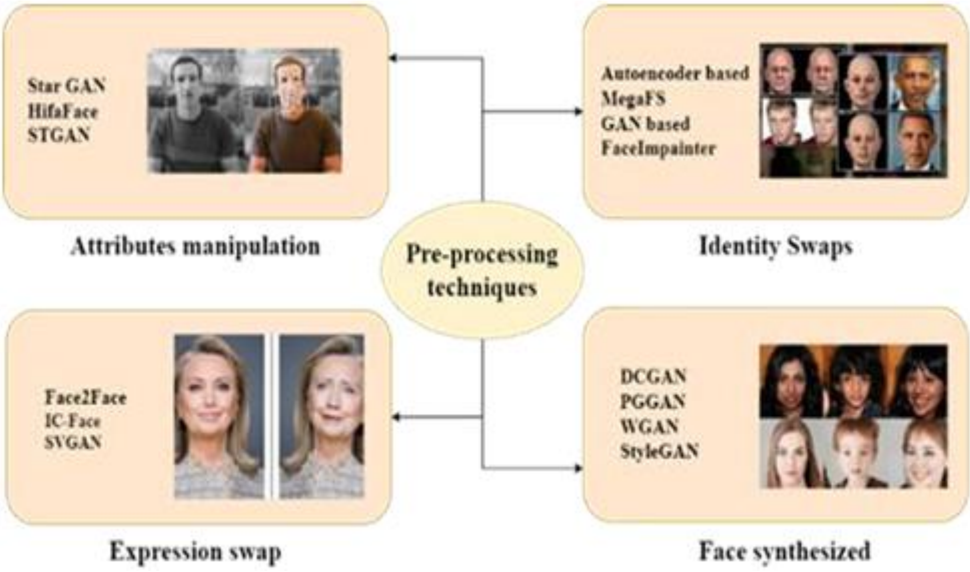

Fig. 4. Overview of pre-processing techniques

Feature extraction

The focus of feature extraction is dimensionality reduction, which is the partition of a set of raw data into smaller groupings. As a result, the operation will be hastened. Feature extraction is the process of reducing the dimension of an unprocessed dataset by breaking it into more manageable subsets. A number of unique elements contribute significantly to the properties of these massive datasets. A large number of computer instruments are required to process variables. The most valuable components of extraordinarily big data sets are found using feature extraction, which involves choosing and combining factors. Its goal is to decrease the amount of data while preserving the most important information. These features are easy to use and correctly reflect the actual dataset. Such a diagram.

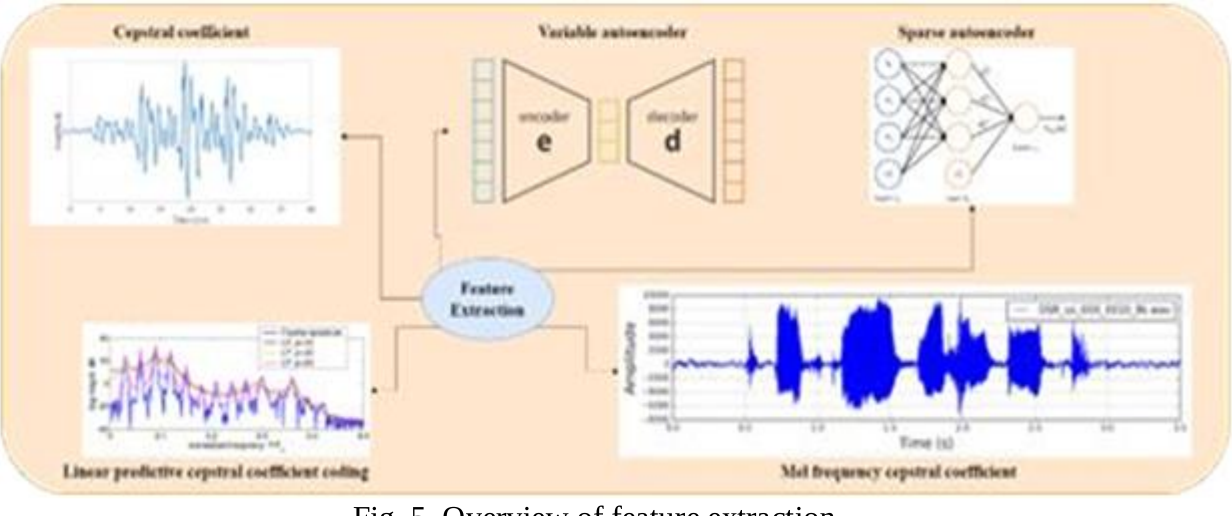

Fig. 5. Overview of feature extraction

Convolutional and synthetic media variable autoencoders are the most popular feature reduction techniques. Variational autoencoders (VAEs) are generative models that debuted in 2013. The model described above produces reliable data and important representations from input. Variational Autoencoder (VAE) employs Bayesian inference. VAE uses a latent variable to calculate the probability distribution p(x) of input data. As stated in this paper, the VAE is a probabilistic neural network with PCA and non-linearity.

To address visual quality difficulties raised in previous research, we developed three key principles for developing face-swapping systems that provide high-fidelity results. To produce a big volume of high-quality movies, the system must be extensible and adaptable to many circumstances. Facial expression incongruity caused by differences must be addressed. Video temporal coherence must be investigated.

Variational autoencoders contain two networks. The encoder or inference network converts significant data attributes into a latent representation of the input. A decoder or generative network transforms the encoded latent representation into recovered data.

A partially processed image is utilized to extract computer vision features via a sparse autoencoder-based LSTM model. The auto-encoder technique does not allow for direct transmission of data from memory to a layer for feature identification. Unsupervised deep learning encodes data for feature extraction using a single hidden layer. This function calculates the error after extracting the utterance from the buried layer and applying the most significant attributes. The method employs a sparse autoencoder, or autoencoder variant. Unsupervised deep learning encodes input with a single hidden layer for feature extraction. This function computes the error and extracts the concealed layer phrases with the most important attributes.

Encoders and decoders create a compression space in an SAE. The encoder translates input into parameter space using buried layers, and the decoder transfers data from parameter space to output layer. Autoencoders cannot comprehend negative values, which limits the network's training capacity, hence sigmoid activation functions were chosen over ReLU activation functions. SAE decreases the gap between input and regenerated data. Due to the sparsity, fewer hidden layers are used. Overfitting in this situation can be avoided by repeatedly scaling down larger datasets.Detective frameworks

Deep learning has proven effective at detecting bogus photos. Unfortunately, video compression diminishes pixel data, making it harder to employ deep learning systems to detect fake films in photographs. The following section divides Deepfake video recognition research into two categories: temporal and spatial feature analysis and biological signal analysis.

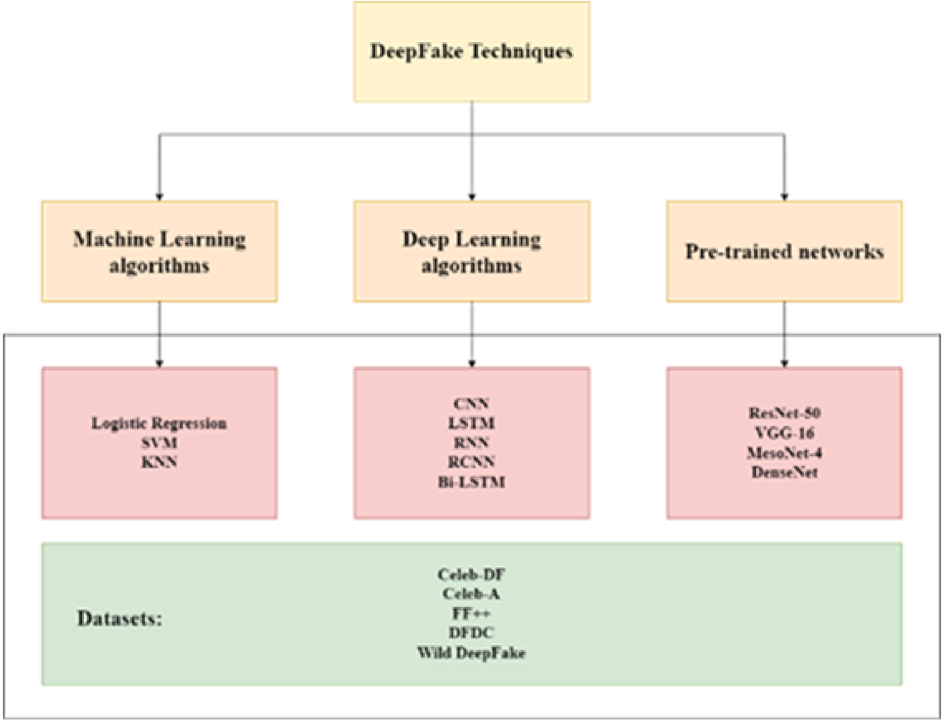

Fig. 6. Some existing detection methods of DeepFake

Figure 6 illustrates the current locations of deepfakes. It has networks learned in datasets, machine learning, and deep learning. Few studies use machine learning.

Standard machine learning (ML) algorithms have the ability to explain any option that is understandable to humans. Deepfake's data and process understanding makes these tactics effective. Model designs and hyperparameters are more easily modified. Extremely randomized trees, decision trees, and random forests are examples of tree-based machine learning algorithms that use trees to represent the decision-making process. Despite the model's reliance on a small number of variables, standard machine learning techniques yielded the most optimal results.

LSTM (Long short-term memory)

An element of recurrent neural networks is LSTM. LSTM artificial recurrent neural networks demonstrate exceptional performance in the domain of long-term relationship management. The input channels of LSMT deliver data encompassing the entire stream. Numerous fields utilize time series data to classify and make predictions, and LSTM is employed in this regard. LSTM consists of input, forget, and output gates. Prior numbers are stored in the LSTM cell state for future use. For determining the range of the sigmoid function, the neglect gate was utilized.

CNN (Convolutional neural network)

Deep learning models typically use convolutional neural networks. Convolutional neural networks, like conventional neural networks, have buried layers in both the input and output layers. The upper layer of this network sends information to the secret layers prior to convolution of the data. In this sense, "convolution" means matrix multiplication. CNN employs a Rectified Linear Unit (RELU). It is a matrix multiplication-based nonlinear activation function. More complicated layers follow, including the flooded layer. Pooling layers diminish the multidimensionality of unprocessed data and limit the amount of data that can be sent and received at various levels. Each link is assigned a weight, which reflects how two items are related.

In recent years, deep learning algorithms have substantially enhanced the quality of synthetic voices. It can now convert into other accents, translate them with semantic comprehension, learn from a small set of samples, and build self-aware networks. These enhancements are quite important. They are still unable to develop words that sound more natural and resemble human speech due to a dearth of training examples in a range of scenarios. Guarnera et al. used deep learning to develop a system for detecting changes in photos. To train the naive classifier to distinguish between legitimate and counterfeit photos, information about the images was gathered using the Expectation Maximization (EM) approach. While the accuracy of deepfake identification has improved, its use is currently confined to static images. Convolutional neural networks are an example of deep neural network architecture. The top layer of a Convolutional Neural Network (CNN) receives unprocessed data, which is then processed convolutionally in the hidden layers. The process of doing dot product calculations or merging matrices is referred to as "convolution" in this context. CNN uses the rectified linear unit (RELU), a nonlinear activation function made up of pooling and convolution layers after matrix multiplication.

RNN (Recurrent neural network)

Specifically, this form of neural network can process sequential input and extract data features. Recurrent Neural Networks, or RNNs, are made up of a large number of non-observable layers. Similar to how weights and biases are assigned to layers in neural networks. Recurrent neural networks, or RNNs, are distinguished by the order in which node connections are created inside a directed cyclic graph. RRN has the advantage of being able to recognize activity with temporal dynamics. Unlike feed-forward networks, this particular neural network uses private storage to maintain the sequence of data from previous inputs. This feature makes it useful in a wide range of applications, including natural language processing and voice recognition. RNNs may process temporal sequences because they provide a recurrent concealed state that contains the links between multiple time scales. They are thus able to process temporal sequences. The input layer, as the neural network's initial layer, is responsible for receiving and relaying data to the layer immediately above it. The input data is made up of four components: A1, A2, A3, and A4. The next stratum contains concealed layers, which are interconnected clusters of synthetic neurons arranged in a stacked pattern.

Techniques based on deep features.

Researchers discovered the internal GAN process and devised a range of algorithms for detecting changes in facial traits. Based on previous research, the author proposed that intracellular brain activity may be used to discern between genuine and false faces. This is performed by layering multiple combinations of cell activations to provide a complete set of visual properties. To calculate the deep features, the FakeSpotter approach used a variety of deep learning-based face recognition frameworks such as OpenFace, VGG-facial, and FaceNet. The data acquired was used to train the SVM algorithm to distinguish between authentic and counterfeit faces. While the study successfully identified changes in facial features, its performance was substandard when applied to groups under dramatically different lighting circumstances. The YOLO face identification method was used to video frames to extract facial areas. Transfer learning can be used to build predictive models that have already been trained. Transfer learning allows for the use of previously acquired skills to generate exceptionally exact predictions. Transfer learning-based network fine-tuning may entail retraining a segment of the network on a new dataset.

GANs (Generative Adversarial Networks)

GAN is a more advanced way to training generative models that uses data distribution to produce convincing instances. GANs usually include both discriminator and generator networks. By comparing these neural networks using a dynamic mini-max game. This is because D seeks to distinguish between actual and sample data, whereas the generator tries to fabricate data. Given enough time to contend, both models will advance. As a result, the generator mimics the distribution of the data, while the discriminator calculates the likelihood that the data came from the training examples.

The generator and discriminator are represented by the symbols G and D, respectively. BigGAN, Style-GAN, Wasserstein GAN, progressive expanding GAN, and deep convolutional GAN (DCGAN) have all been designed to improve training, losses, and designs.

ANNs were used to adjust properties of the SAGAN model. Using spatial attention, this model limited alteration to a single area. A Generative Adversarial Network (GAN) also made changes to characteristics.

PA-GAN utilized GAN's progressive attention method to integrate encoder and attribute features in a specific attribute area. We created an attention mask that smoothly transitions from high to low levels of feature. Increased feature levels and resolution allow for more exact attribute manipulation and attention masking. This method effectively modifies a large number of variables while maintaining insignificant features within a model.

Current Surveys Using Various Modalities

This section covers DeepFake approaches for image, video, and hybrid datasets. Document-based DFDT. There are several approaches for deep neural networks to recognize Generative Adversarial Network (GAN) images. Tariq et al. proposed using deep neural networks to detect fraudulent images. Performing a statistical analysis prior to processing increases the detection of manually made false face photos. Nhu et al. provide an alternate method for detecting GAN-generated fake images. This system is powered by a strong convolutional neural network. This technique uses a deep learning network to extract facial features from face recognition networks. Then, facial traits are altered to differentiate between genuine and counterfeit images. These strategies give good results for contest validation.

To address the disparity, the superficial layer's textural properties are improved to better depict local components. These properties now provide enhanced semantic data. The author can improve the network's ability to learn knowledge from different attention locations by deliberately allocating high reaction attention to veiled regions throughout the training phase. A number of datasets are used to evaluate this strategy. The following mathematical formula is used in this study to supplement attention-guided data: Table 1 presents a comparison of picture data.

Table 1: Existing work based on Image data.

Year	Methodology	Advantages	Disadvantages
2021 [1]	Author creates many attention heads using deep semantics to predict spatial attention maps. Classify edited photos precisely. This study advised attention-guided data augmentation.	Identify fake photos for classification. Simple to grasp.	Since datasets have different formats, precision will vary.
2021 [2]	Unsupervised genuine and fake data classification. Two picture augmentations enable unsupervised contrastive learning. Learn backbone network preprocessing.	solves unsupervised contrastive learning with unlabeled datasets.	Labeling is laborious.
2022 [3]	This study examines text-to-image synthesis, attribute manipulation, picture augmentation, and shifting. StyleGAN, YOLO, and RCN CNN models detect fake data.	The picture collection methods help create a precise model.	This work requires a long time to edit photos.
2022 [4]	Facial swapping technologies like Deepfake change the face while retaining the environment. Differences suggest manipulative exploitation.	Increase dataset size. Alter image context to produce phony visualizations.	Data types complicate generated-original data combinations.
2023 [5]	The Dual Shot Face Detector collects facial features from movies and photographs. XceptionNet, MesoNet, FWA, and capsule identify deepfakes.	Photographs with and without resolution.	Higher-level computation is needed for training.

3. DISCUSSION RELATED IMAGE BASED DFDT

Different GANs simplified the development of Deepfake images using AI. TensorFlow also includes a large number of pre-trained detection models, making neural network training and picture classification simple. Various models were used in image-based detection. Table 6 shows the dataset performance of several models.

Audio based DFDT

In computer graphics, cross-modal audio-driven talking head animation has a long history. Prior strategies for generating real videos might be divided into two categories. The rigging characteristics of facial models are linked to 3D vertex coordinates utilizing these nonphotorealistic methods. Determine whether source waveforms correspond to facial motions. Rigging settings usually demand creative professionals or high-quality 4D face capture data. Pre-trained models, audio analysis tools, and training audio datasets that do not contain video are used. Using audio data, Table 2 compares various DFDT algorithms.

Table 2: Existing work based on Audio data.

Year	Methodology	Advantages	Disadvantages
2021 [6]	A Deepfake video dataset is combined with genuine audio. A mel-spectrogram was made from sounds. Used to make fake sound with vocoders. Author streams video using R(2+1)D 18. The author uses a simple 1D convolutional network to assess audio stream 1D raw wave-form data.	In this study, the authors examined audio, video, and integrated audio-video datasets.	The merged dataset is less accurate than separate datasets.
2022 [7]	A Mel-frequency cepstral coefficient parameter is introduced. In this initial stage, we use NTU and bispectral analysis to examine datasets like a machine learning cycle. Using spectral roll-off, Mel-frequency Cepstral Coefficients (MFCCs), RMS, zero crossing, and chroma frequency, I trained many RNN models.	minimum difficult math.	Audio samples are two seconds long.
2022 [8]	SVM, K-Nearest Neighbours (KNN), and other machine learning approaches can differentiate authentic and fake audio with 97% accuracy. CNN is another deep learning method with 99% accuracy and 2% misclassification. The author employed ResNet34, LSTM, and RNN CNN models correctly.	These methods can detect audio duplication.	Languages other than English have few audio Deepfake detection methods.
2020 [9]	Randomly identify and target speech segment to match lip synchronization. Multiple discriminator frames are used.	The Wav2Lip model outperforms current methods. Learn vocal synchronization.	Additional evaluation measures are needed to choose the model. Strengthening is essential.
2023 [10]	Systematically identify false audio and video with AVFakeNet. Unique Dense Swin Transformer Net was built for feature extraction.	Novel method to determine multimodality.	More processing power is needed.

Discussion related Audio based DFDT

Mel-spectrum is a common technique for detecting audio changes. The procedure converts audio data into an unprocessed state, allowing for classification using CNN, RNN, or LSTM. A variety of sound-based detection methods were implemented. Table 6 shows how the models' performance compares to datasets. The lip is moved by picking frames from a video or transcript that correspond to the emotions defined in prior lip-syncing research.

Video based DFDT

Deep learning has lately shown efficacy in the detection of false data. Deep learning approaches used in image analysis are currently incapable of detecting counterfeit videos due to significant information loss caused by video compression. Current research in deepfake video recognition is focused into two categories: 1) biological signal analysis; 2) time and spatial analysis. The video dataset's target pictures are constructed using the space-time conditioning volume formulation . Table 3 presents a detailed evaluation of DFDT algorithms for video data.

Table 3: Existing work based on Video data.

Year	Methodology	Advantages	Disadvantages
2018 [11]	Encoders and decoders identify deepfake videos. Integration of LSTM with CNNs achieves this. Note the updated LSTM model. LSTM employs 20, 40, or 80 frames.	used to detect films with several LSTM variants.	Two autoencoders cannot operate independently during training.
2018 [12]	This adaptive method detects bogus videos by measuring eye blinking. Blinking and eye movement are detected using convolutional and recursive neural networks.	The research showed that the proposed technology can recognize fake photos.	Converting to frames was tedious, but necessary.
2018 [13]	First, transfer an actor's orientation, head position, eye gaze, and facial expression. A breakthrough rendering-to-video network creates realistic and temporally consistent movies from computer images.	These approaches initially showed extremely accurate reenactment outcomes in several situations.	Video-based classification makes it laborious.
2018 [14]	Resource and production time constraints limit an algorithm's face photo size. Affine warping aligns the output picture's face features with the source.	Overcome dataset generation duration constraints.	require more time than trained networks.
2023 [15]	To detect fake videos, hierarchical file structure and feature selection methods like gray wolf optimization and Vortex search algorithm are utilized.	improved HFS technique precision across numerous datasets.	A difficult process.

Discussion related Audio based DFDT.

Deepfake film production is tough because to time restrictions. These require CNN and RNN neural networks for classification. After converting video material to frames, neural networks were trained. Optimal precision was achieved through extensive training. Initially, eye-blinking datasets were created and made available to the public in an attempt to detect it. These binary classification approaches can discriminate between real and false instances in complex datasets. Table 6 compares the performance of various video-based detection methods on their individual datasets.

Multimodality in DFDT.

This section discusses multimodality in deepfake detection and the relevant approaches. We evaluated and investigated datasets that used a single data mode and the accompanying techniques. Multimodal data analysis: How can the truth of media be verified? One approach mixes pictures and sound, while the other blends video and audio. A variety of strategies exist for reviewing and creating different datasets. Pre-trained networks are used for integrated datasets, while audio and images are processed using MFCC, CNN, and RNN. In Table 4, various modalities are compared.

Table 4: Existing work based on Multimodality data.

Year	Methodologygy	Advantages	Disadvantages
2021[16]	DFDC mixes real and fake music and video without revealing it. Multimodal models included Xception, Efficient Net, and VGG16.	Labels are unnecessary.	Current detection algorithms only analyze audio or video files.
2022[17]	Predict audio and video with LSTMs and GANs. Audio detection apps use SincNet to process unprocessed waveforms without preprocessing or data loss.	use a step-by-step paradigm for maximum precision.	Individual modality categorization increases training and feature extraction time.
2022[18]	StyleGan fakes data utilizing face synthesis, attribute modification, expression shifting, and others.	Creating deepfake datasets with viewing angle in mind.	Compliance with labeling is needed.
2022[19]	The investigation used dual-stream networking. Deepfake detection with M2TR.	Multi-scale transforms reveal image frauds.	Faces need trimming..
2023[20]	Extract audio-visual components from monomodal datasets using time-aware neural networks.	Model training does not require current datasets.	Training with disconnected monomodal datasets is needed.s.

4. CONCLUSION AND FUTURE SCOPE

Deep learning has become popular across industries. Many deep learning-based algorithms have been developed to address this issue and detect phony photographs and videos. Current deep learning approaches must be improved to detect phony photographs and videos. The proposed background employs a variety of attention mapping approaches to inspect different regions and improves deep layer texture features to detect more minor defects. Attention mapping is then utilized to create semantic and textural features. An attention-guided data augmentation technique with an independence loss function teaches embarrassing multiple awareness. Both approaches are in development. Our method produces positive results across a variety of KPIs. The following can be used to summarize the investigation: The Face Forensic ++ dataset is the most widely utilized in DFDT testing for deep learning algorithms. CNN models make up the majority of deep learning models. The most common performance parameter is detection. The alternative result demonstrates that deep learning can detect deepfakes. In terms of performance, deep learning trumps non-deep learning models.

5. REFERENCES

1. A. Malik, M. Kuribayashi, S. M. Abdullahi and A. N. Khan, "Deepfake Detection for Human Face Images and Videos: A Survey," in IEEE Access, vol. 10, pp. 18757-18775, 2022, doi:10.1109/ACCESS.2022.3151186.
2. Hasam Khalid, Minha Kim, Shahroz Tariq, Simon S. Woo, “Evaluation of an Audio-

Video Multimodal Deepfake Dataset using Unimodal and MultimodalDetectors", 2021

3. Y. Zhou and S. Lim, "Joint Audio-Visual Deepfake Detection," in 2021 IEEE/CVF International Conference on Computer Vision (ICCV), Montreal, QC, Canada, 2021 pp. 14780-14789.
4. D. Güera and E. J. Delp, "Deepfake Video Detection Using Recurrent Neural Networks," 2018 15th IEEE International Conference on Advanced Video and Signal Based Surveillance (AVSS), 2018, pp. 1-6, doi:10.1109/AVSS.2018.8639163.
5. H. Zhao, T. Wei, W. Zhou, W. Zhang, D. Chen and N.Yu, "Multi-attentional Deepfake Detection," 2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2021, pp. 2185-2194, doi: 10.1109/CVPR46437.2021.00222.
6. Ahmed S Abdulreda; Ahmed J Obaid. "A landscape view of Deepfake techniques and detection methods". International Journal of Nonlinear Analysisand Applications, 13, 1, 2022, 745-755. doi: 10.22075/ijnaa.2022.5580
7. Khan, Madeeha B.; Goel, Sanjay; Katar Anandan, Jaswant; Zhao, Jersey; and Naik, Ramavath Rakesh, "Deepfake Audio Detection" (2022). AMCIS 2022Proceedings. 23.
8. Almutairi, Z.; Elgibreen, H. A Review of Modern Audio Deepfake Detection Methods: Challenges and Future Directions. Algorithms 2022, 15, 155. https:// doi.org/10.3390/a15050155
9. Sheldon Fung, Xuequan Lu, Chao Zhang, Chang-Tsun Li. DeepfakeUCL: Deepfake Detection via Unsupervised Contrastive Learning. https://doi.org/10.48550/arXiv.2104.11507.
10. Nirkin Y, Wolf L, Keller Y, Hassner T. Deepfake Detection Based on Discrepancies Between Faces and Their Context. IEEE Trans Pattern Anal Mach Intell. 202 Oct;44(10):6111-6121. doi: 10.1109/TPAMI.2021.3093446. Epub 2022 Sep 14.PMID: 34185639.
11. Tariq, S., Lee, S., Kim, H., Shin, Y. and Woo, S.S. (2018) Detecting Both Machine and Human Created Fake Face Images in the Wild. Proceedings of the 2nd International Workshop on Multimedia Privacy and Security, Toronto, 15 October 2018, 81-87.
12. Lomnitz, Michael & Hampel-Arias, Zigfried & Sandesara, Vishal & Hu, Simon. (2020). Multimodal Approach for Deepfake Detection. 1-9. 10.1109/AIPR50011.2020.9425192.
13. Prajwal K, Mukhopadhyay R, Namboodiri VP, Jawahar C (2020) A lip sync expert is all you need for speech to lip generation in the wild. In: Proceedings of the 28th ACM international conference on multimedia, pp 484–492.
14. Kim H, Garrido P, Tewari A, Xu W, Thies J, NiessnerM, Pérez P, Richardt C, Zollhöfer M, Theobalt C(2018) Deep video portraits. ACM Trans Graph 37:163–177.
15. Li, Yuezun, and Siwei Lyu. "Exposing Deepfake videos by detecting face warping artifacts." arXiv preprint arXiv:1811.00656 (2018).
16. Wang, Junke, et al. "M2tr: Multi-modal multi-scale transformers for Deepfake detection." Proceedings of the 2022 International Conference on Multimedia Retrieval. 2022.
17. Hatmaker Taylor, "DARPA is funding new tech that can identify manipulated videos and Deepfake", web blog post. May 01, 2018.
18. Ilyas, Hafsa, Ali Javed, and Khalid Mahmood Malik. "AVFakeNet: A unified end-to-end Dense Swin Transformer deep learning model for audio–visual deepfakes detection." Applied Soft Computing 136 (2023): 110124.
19. Mohiuddin, S., Sheikh, K.H., Malakar, S. et al. A hierarchical feature selection strategy for deepfake video detection. Neural Comput & Applic 35, 9363– 9380 (2023). https://doi.org/10.1007/s00521-023- 08201-z.
20. Salvi, D.; Liu, H.; Mandelli, S.; Bestagini, P.; Zhou, W.; Zhang, W.; Tubaro, S. A

Robust Approach to Multimodal Deepfake Detection. J. Imaging 2023, 9,122. https://doi.org/10.3390/jimaging9060122.