IRJAEH

International Research Journal on Advanced Engineering Hub (IRJAEH)
e ISSN: 2584-2137
Vol. 04 Issue: 01 January 2026
Page No: 163-175
https://irjaeh.com
https://doi.org/10.47392/IRJAEH.2026.0023

Machine Learning Based Datasets Collection and Preprocessing of Deepfake Video

Prof. Vikram Singh[1], *Naresh Kumar*[2]

[1,2]*Department of Computer Science & Engineering, Chaudhary Devi Lal University, Sirsa (Haryana), India.*

Email ID: *vikramsingh@cdlu.ac.in*[1], *nranga386@gmail.com*[2]

Abstract

In a matter of developing deepfake technology is a major challenge in methods to detect manipulated videos. This study is dedicated to the deepfake dataset review and discusses the possible strategies in deepfake Collecting and formatting data for the establishment of a reliable deepfake detection model. We survey different datasets available for deepfake research and refer to the preprocessing techniques that aid in the performance of deepfake detection models and provide an exhaustive account of the existing deepfake video datasets. The study suggests how can selecting the right data sets and methodologies for preprocessing in order to increase the accuracy and efficacy of deepfake. present the issues and limitations of current datasets and preprocessing methods and envisage future work such as the creation of novel datasets and sophisticated preprocessing methods.

Keywords: *Detection deepfakes, Machine Learning, Transfer Learning, Deep Learning, Forgeries, Deeper Forensics.*

1. Introduction

The training of deepfake detection systems requires robust and high-performance collections of datasets and preprocessing techniques, and is also valuable for the advancement of research in multimedia forensics. Any deepfake detection method is dependent on the availability of high-quality diverse and representative datasets. Collection methods are usually based on the pooling of the real and manipulated videos from a broader base of sources to ensure the diversity of data [1,2]. While developing the dataset, it is necessary to have standardized protocols for labeling, annotating, and organizing. Credible metadata like the identities of the subjects, the ways of manipulating, and quality scores, guarantee that later model training and validation can be reproduced and are scientifically rigorous. Dataset curation is commonly challenged with the problem of finding the right balance between the amount of real and fake content. It also concerns the inclusion of different synthesis techniques such as face swapping, morphing, and audio-visual manipulation in order to make models not overfit to a particular generation. It is a process that helps deep learning networks by standardizing input which will lead to better algorithmic performance and reproducibility of the experiments. Besides that, by adding some synthetic distortions (for example, compression, blurring, and noise) to the data one can imitate real-world situations and thus make the model more robust. The cutting-edge preprocessing techniques can even get the spatiotemporal features by utilizing convolutional neural networks (CNNs) for the spatial part and recurrent networks (RNNs) for the temporal one. The twofold method helps to detect small changes from one to the next video frame as well as the trace left by the manipulation which are not visible for the static images and thus, the detection rates are considerably increased. What is more, great attention to the ethical and privacy issues during data gathering is a prerequisite, for instance, employing the methods of release by consent and the methods of anonymization so that they are in accordance with the legal and societal standards. The successful creation of a deepfake video dataset is contingent upon the availability of content that is diverse and of high quality, which has been collected using systematic protocols and processed using advanced spatial-temporal techniques to retain the relevant forgery cues. The thorough annotation, spatiotemporal preprocessing, and ethical measures taken together provide a solid foundation for the development of datasets that are necessary for the creation of powerful and generalizable deepfake detection technologies [2,3,4,5,6,7].

2. Background

Deepfake technology uses deep learning algorithms to create very visually manipulated videos, which complicates it to a great extent to choose between true and false information. Markers of the history of deepfake video datasets are the shift of simple to use single-technique sets to large scale, complex sets designed to assist in accurate detection, and cross environmental classification. The academic circle desires to keep pace with the synthesis of the media with this development and keep the realism of the digital content as well. The majority of the recent work including FaceForensics++ and FFIW10K is aimed at capturing more difficult multi-face scenarios and give accurate annotations to both spatial and temporal features. The ongoing enhancement of these corpora owes to the similar advance in their creation that must be managed by detection research using data that are equally well-rounded and extensive. [4].

3. Objective

The first purpose of making and sharing of deepfake video datasets is to enable accurate algorithm development for training and validation that can make a distinction between real and fake videos. In order to verify some robustness and efficiency of models, scientists must subject their detection models to various kinds of manipulation approaches, particular face modifications and changes in light environments. This is made possible by extensive datasets. In addition to making the data more accessible to the algorithm, transformations of the deepfake video media such as conversion to frame-based images, facial detection, alignment and highlighting also highlight likely areas of forgery in order to speed up feature extraction for the analysis that follows. Establishing standard input data for machine learning algorithms is another significant objective that provides consistency and equity in evaluation within investigation. By carefully selecting and analyzing these datasets, the study group seeks to improve digital media security, increase ethical utilization of artificial materials and increase public confidence in visual information□□To enable the enhanced detection of deepfake videos through the collection of deepfake video datasets, different preprocessing measures have been applied by the researchers. They accomplished this by devising systematic strategies covering data acquisition, cleaning, and advanced transformation techniques. As a result, collection best practices usually obtain videos from diverse established datasets such as FaceForensics++, Deeper Forensics, DFDC, and BioDeepAV to broaden the variety of content manipulation methods, environments, and actor demographics. For instance, the ExDDV dataset that combines the usage of thousands of real and fake videos produced by different face-swapping and generative methods and also gives the splits for training, validation and testing carefully to ensure generalizability and robust benchmarking. Recent work on the topic has led to the development of different preprocessing methods that essentially point out the difference between camera-specific artifacts and GAN-generated fingerprints or use physiological signals such as eye blinks as sophisticated deepfake indicators. In neural network pipelines, some studies prove the advantage of using different preprocessing methods which spatially and temporally resolve the features extracted to further increase the accuracy of the model in a complex deepfake scenario. The integration of multi-source data, rigorous preprocessing, and context-aware augmentation collectively constitute the core of efficient modern deepfake video dataset strategies. [8,9].

4. Comprehensive Review of Existing Deepfake Video Datasets

One major area where deepfake video datasets have changed is after the rise of synthetic media technologies in their size, diversity and application have increased significantly. The first deepfake datasets generations such as UADFV and DFTIMIT contained a few numbers of manipulated videos that were only used to research the identification of facial liveness through blinking and lip movement. For instance, UADFV contains 98 samples of real and fake content that were obtained from YouTube, while DF-TIMIT has 640 videos based on GAN face swapping that are divided into low and high-resolution groups. Although they have been instrumental, these early datasets were constrained by their absence of diversity, small participant numbers and poor representation of the real-world conditions. [10]. Later on, deepfake datasets like FaceForensics and its advanced version FaceForensics++, were created as a result of subsequent changes. These datasets contain a significantly larger amount of data than before more than 1,000 original YouTube videos have been manipulated in multiple ways using cutting-edge techniques such as Face2Face, FaceSwap, DeepFakes, and NeuralTextures. FaceForensics++ offers both low and high-resolution video samples along with the ground-truth segmentation masks thus making it possible to a much greater extent to validate the models for both classification and forgery localization. Moreover, the Celeb-DF dataset resolved the quality problems of the earlier datasets and presented more than 5,600 high-quality manipulated videos of 59 different celebrity subjects, which were recorded at 30 FPS at the resolution of 256x256 pixels with careful synthesis to ensure that there were no visible artifacts. The dataset introduced more diversity in terms of ethnicity, age, gender, lighting, and backgrounds, thus becoming like the present-day social media communities. [11,12]. In the author hand specialized datasets such as DF-Mobio mimic real situations like video calls and they provide both fake and authentic samples that represent the goal of anti-spoofing systems testing. In an effort to reveal bias and robustness issues, researchers get more and more datasets constructed with thorough, detailed annotations of demographic, visual, and algorithmic features, which is evident from recent large-scale annotation projects for FaceForensics++, Celeb-DF, and DFDC datasets [13]. According to the studies, each dataset performs differently under various sets of criteria: DFDC and DeeperForensics 1.0 are good for large-scale and diverse challenging scenarios, FaceForensics++ is a multimodal benchmark for forgery localization, WildDeepfake provides a wider range of naturalistic samples, and Celeb DF is a dataset of high-quality and social media realism. The merge of datasets and standard benchmarking practices that continues is a way of generalization problems that have previously existed between synthetic and real-world manipulations are being solved, thus, academic and applied deepfake detection research are getting strengthened in combination.

5. Impact of Dataset Variety on Detection Model Performance

The range of datasets can greatly influence the detection of deepfake models. It determines their stability and the extent to which they can be applied to new manipulations or untested attacks. If the MIT reality models are trained solely on one dataset or on a few methods of synthetic generation, then they demonstrate high effectiveness in the evaluation of the dataset but their accuracy decreases significantly when they are tested on videos from unknown datasets or with new methods of manipulation. The term generalization gap indicates that it is very important to have diverse content, various manipulation techniques, and extensive demographic representation in the training mode. The rise in dataset diversity not only alleviates the problem of overfitting to specific artifacts or production chains but also allows models to get further through consistent and more profound cues to forgery which, for example, can be physiological inconsistencies, lighting mismatches, or subtle GAN fingerprints. As an example, experiments reveal that the use of cross-domain data together with the application of augmentation strategies can lead to a high degree of improvement in cross-dataset generalization as well as in endurance against the direction of attacks. The combination of supervised-reinforcement learning and information decomposition framework is another idea that can be utilized to realize the detection enhancement of deepfake-related features across variable domains and manipulation styles by the training networks. Whereas insufficient dataset variety can generate bias issues which lead to uneven execution in different demographic groups and low fairness in the results of detection. According to experimental evidence, the detection backbone models which form the basis of unbalanced data and are thus trained, may misattribute certain facial features or demographic characteristics, resulting in biased or unreliable outcomes in the populations of different ethnicities. Therefore, to enhance the model's trustworthiness it's not only the increase in data that is necessary but also the careful dataset curation, source domain balancing, and deepfake techniques incorporation through continual updates. The existence of a well-balanced, well-balanced, and constantly updated deepfake video dataset is the key to the development of detection models that are not only able to keep high performance levels but also be fair and resistant to manipulations that have been both previously and unexpectedly discovered [14,15,16].

6. Preprocessing Techniques and their Impact on Detection Accuracy

Preprocessing techniques are the main contributors to the improvement of deepfake detection models. Such techniques enable a model to extract appropriate features, lessen noise, and obtain uniformity in the dataset. Here, we detail the different preprocessing techniques and their influence on the detection accuracy. Frame extraction is essential in the preprocessing of deepfake videos. It is about choosing the exact frames from a video to analyze the time aspect and obtain the traces of the falsification. The way frame extraction is done can have a great effect on the detection performance description of a dataset that is commonly used for videos manipulated by different methods. The temporal method means that frames are taken at fixed time intervals (e.g., every 10 frames). In this way, the frames are evenly spaced throughout the video, thus the temporal changes are captured properly. On the other hand, if the intervals are too large manipulations may be missed. The impact on detection accuracy of the frame extraction method can be very different. Other methods, in particular have been found to increase detection capabilities by providing a balanced representation of temporal information and frame differences [17]. Face detection and extraction are the processes of identifying the face area in every video or image frame and then cropping that portion of the frame to concentrate on the regions of interest. It is a necessary procedure because, as a rule, deepfake editing only changes the facial features of the person. There are numerous ways and toolkits for detecting faces, and each of them has certain advantages and disadvantages. Full frame refers to the use of the whole video frame for the purpose of analysis. With this approach, the whole scene is captured, but it may contain a lot of irrelevant information, thus, the model's focus on the face region is lowered. Mask this is the facial region detected by the face detection model. With this method you concentrate only on the face but it is possible that some contextual information will be lost. Detected face with a little bit of an added margin. This method not only focuses on the face but also includes the surrounding area, which may be helpful for the identification of the manipulation of the image. The way faces are extracted has a major influence on whether detection accuracy will be high or not. The most effective method has proven to be that of face which uses the detected face with an added margin, as it allows for a good compromise between concentrating on the face and, at the same time bringing in the relevant contextual information [18]. Normalization methods are necessary to keep the dataset uniform and to upgrade the quality of the images. Some common normalization methods are resizing images, histogram equalization, and pixel value normalization. Resizing images to a uniform size is a way of making the whole dataset consistent and also it is a method of lowering the computational complexity. Contrast of the images is enhanced by the histogram equalization technique making it easier to detect even the most subtle manipulation artifacts. Converting pixel values to a certain range (for instance, 0 to 1) is a way of stabilizing the training process and model performance is also getting better. Normalization methods allow the dataset to be of a higher standard detection models will find it easier to learn the relevant features. Correct normalization may have a major effect in detection precision by noise reduction and making the dataset consistent [19]. Temporal analysis includes looking at the sequence of frames to find temporal inconsistencies that might signify manipulation. This method is especially effective in discovering deepfakes because, as a rule, many manipulation techniques bring temporal artifacts. The optical flow method studies the movement of pixels in a frame that is compared with the previous one. It is capable of finding fake or manipulated videos where the motion is not natural. The frame difference method is used to find differences between the frames. The method can highlight areas where manipulation has occurred. Influence on detection accuracy temporal analysis techniques can play a major role in detection accuracy when used in conjunction with other techniques because they can locate the manipulation artifacts that are invisible to the naked eye from single frames. Such methods are especially efficient in revealing deepfakes generated by techniques that result in temporal inconsistencies. [20]. Data cleaning is the process of getting of corrupted or irrelevant data from the dataset. It is a crucial step that makes the dataset consistent and error-free, thus preventing these errors from negatively affecting the detection model. It is a task of removing corrupted frames and recognizing those which are corrupted and have errors. Making sure that all frames in the dataset are consistent in terms of resolution, format and other attributes is called data cleaning. It raises the overall quality of the dataset which in turn makes it easier for detection models to learn relevant features. A clean and consistent dataset has the potential to increase detection accuracy to a great extent by cutting down on noise and allowing the model to concentrate on the most relevant information [20,21].

7. Overview of Deepfake Datasets

Deepfake technology has substantially improved that it has numerous datasets created to support the development and evaluation of deepfake detection models. These datasets differ in their volume, quality, and the types of manipulations they are appropriate for different research requirements. Here, we present a review of a few deepfake datasets that are commonly used and their contributions. Face-Forensics++ (FF++): Face Forensics++ is the most widely referenced dataset in deepfake studies. It comprises a variety of videos that have been altered by different methods, such as Deepfakes, Face2Face, Face Swap, and Neural Textures. The dataset features genuine and fabricated videos with each falsification technique being performed on a separate subset of the data. The videos are taken from YouTube, so a broad range of natural variations in appearance and background are guaranteed. This dataset is partitioned into training, validation and test sets can be used for any model evaluation. [22].

Celeb-DF: Celeb-DF is a large-scale dataset with high-quality videos of celebrities. It consists of both real and fake videos with the fake ones created by different deepfake methods. The dataset is aimed at being difficult for the detection models due to the high-resolution videos and the wide range of facial expressions. Celeb-DF has two versions: Celeb-DF-v1 and Celeb-DF-v2 where the second one has more videos and more intricate manipulations. The dataset serves as a great instrument for testing the strength of the detection models against high-quality deepfakes [22].

Deep Fake Detection Challenge (DFDC): The DFDC dataset was the principal element of the deep fake detection challenge an open competition designed to identify the best deepfake detection techniques. It comprises a large number of original videos and manipulated ones where the fakes are created by various AI methods. The dataset tries to be very diverse as videos of different people are used. The DFDC dataset divides its data into training, validation, and test sets, thus providing a full benchmark for deepfake detection studies. [22].

Deeper Forensics 1.0: Deeper Forensics 1.0 is a large-scale dataset meant to overcome the constraints of current datasets by including more lifelike and varied manipulations. The dataset covers an extensive array of perturbation methods used on the fake videos thus it is also a very challenging benchmark for deepfake detection. The dataset consists of training, validation and test sets separately and in addition detailed annotations are provided for each video. DeeperForensics-1.0 can, therefore, be considered as a source of great value in testing the resilience of detection models to highly intricate and lifelike deepfakes [22].

UADFV: The UADFV dataset comprises a limited number of diverse real and fake videos. The collection is intended to be difficult for the detection models and hence the videos have been taken from different online sources. The dataset splits into training, validation and test sets, thus offering complete benchmark deepfake detection research. UADFV mainly serves as a resource for testing the stability of the models in the face of various manipulative techniques [22].

Ding et al. Swapped Face Dataset: In this dataset, there are 420,053 images of celebrities, which cover 156,930 real images and 263,123 fake face-swapped images. Two different methods and auto encoder GAN were used to create the fake images. The dataset is aimed at offering a large and varied set of images to the deepfake detection models to be used for their training and evaluation.Faces-HQ: Faces-HQ consists of 40,000 high-resolution images, where half of them are real, and the other half are deepfake. The images were obtained from four sources Celeb A-HQ, Flickr Faces-H or 100K-Faces, and thispersondoesnotexist.com. The dataset is intended to be a diverse set of high-resolution images for the training and evaluation of deepfake detection models. [23]. Diverse Fake Face Dataset: The data collection comprises 299,039 images in total, out of which 58,703 are real images while 240,336 are fake images. The fake images in the dataset depict the four different facial manipulation types the changes in the identity, expression, attribute, and complete synthesis. The dataset is intended to expose deepfake detection models to a varied image set for their training and evaluation [23]. IFakeFaceDB: iFakeFaceDB comprises 87,000 224×224 face images that were created with the GAN-fingerprint Removal approach (GANprintR). The collection of data is intended to offer a wide-ranging set of artificially created pictures for deepfake detection models training and their performance evaluation. ifake face DB serves as an excellent resource, especially, when a detection model's resistance to synthetic images needs to be tested. [23,24] Shown in Table 1.

Table 1 Dataset Comparison Table of Different Dataset

Data Set	Real Videos	Fake Videos	Total Videos	Cleared	Total Subject	Deepfake Method
FaceForensics++	1000	4000	5000	NO	N/A	2
Deep Fake Detection	1000	1000	2000	NO	N/A	1
DFDC	361	4119	2003	YES	26	3
Ding et al. Swapped Face Dataset	156,930	263,123	420,053	N/A	N/A	2
iFakeFaceDB	N/A	87,000	87,000	N/A	N/A	1
Faces-HQ	20,000	20,000	40,000	N/A	N/A	1
Celeb A Spoof	N/A	625,537	625,537	N/A	N/A	1
Diverse Fake Face Dataset	58,703	240,336	299,039	N/A	N/A	4

8. Strategies for Effective Deepfake Videos Data Collection

Effective deepfake video data collection greatly influences the progress of detection model. This is because the quality and variety of datasets determine how well the models developed for detection can be generalized and how strong they are to different types of deepfakes videos. A thorough programming starts with getting the most credible and the least manipulated videos from as much different real-life places and media as possible. Often the performance evaluation of deepfake detection techniques facing real data issues is affected by representational biases because numerous benchmarks have depended on outdated methods of generation or have excessively targeted single-person portrait manipulations. Since novel generative technologies like diffusion models and transformers appear quickly data should also advance to encompass computer-generated content that is indistinguishable from the dominating sources of the social discourse such as fake collection scenes deepfake videos of natural disasters and political-leaning deepfake videos. A platform for employees to contribute in product development such as open fake field is a creative way to overcome the feature of same datasets that have not been altered. Such open unrealistic field datasets can be continuously benchmarked, and therefore, wherever the generative processes are headed, it will always be current because a user is able to generate their adversarial synthetic media and submit it to the most recent classifiers. Contemporary collection methods do not focus on the quantity of data but also the diversity as they would have sampled various races, genders, backgrounds, complexities and age groups. This is the method that combats the dataset-based biases and thus makes detection models to the real world. Corresponding hand in hand data curation.in hand with normalization is among the most significant elements in the mending of over-fitting to pre-processing artifacts. The point is made that assessing sample stability under various normalization can be used to assist with the detection of strong forensic indicators of manipulation. It achieves better generalization to unseen deepfake methods through acquiring and improving data using normalization invariant samples. The appropriate collection also addresses the ethical aspect of issues such as the right permission to use it and ensures that other individuals, mainly the ones involved, have their privacy with public figures cooperation with domain experts in psychology, facial biometrics and signal processing to locate understated manipulation signals and broaden dataset coverage for complex situations such as multi-face scenes or partial forgeries can be very helpful. Ongoing benchmarking, adaptive collection platforms, extensive annotation efforts, normalization-consistent curation, and interdisciplinary collaboration are the components of a comprehensive strategy that envisages deepfake video dataset collection and hence, it is the backbone for the continuous development of dependable and robust detection technologies [25,26,27,28]. Effective annotation of deepfake datasets at a deep level is a hybrid process involving manual labeling, whereby experts pinpoint the areas of manipulation and describe the artifacts and automated tools that generate labels for features for human validation. Agreement labeling and difficulty scoring are used to confirm the trustworthiness of results, while detailed attribute tagging like the kind and the strength of the manipulation allows for a more refined analysis. Quality annotation results from well-established instructions, the work of multiple annotators and pixel-level or attribute-specific labels. All these measures combined lead to the creation of robust datasets which are the basis for sophisticated explainable deepfake detection. The increases in fake content creation methods has caused a lot of anxiety in legislative bodies as well as among regulatory authorities that are apprehensive about the use of fake multimedia for illegal and manipulation of the opinion of the mass's purposes. Detecting and categorizing the latest deep fakes the most advanced tools for identification of deep fakes urgently need to be addressed by those who are seeking effective ways of prediction to be able to avert political and social crises of a harmful nature. This research delves into the deep learning-based and transfer learning techniques for image and video manipulation that have been deeply investigated.

9. Data Preprocessing Techniques

Preprocessing techniques are playing important role in improving the performance and the ability of a deepfake video detection model to generalize. Usually, the first step is frame extraction through which videos are broken down into individual frames at optimal intervals using scene detection algorithms or fixed sampling rates so that relevant temporal information is kept without too much redundancy. Most of the time face detection and cropping are used to remove facial regions, which are the main sources of manipulation and hence, concentrate further analysis on these regions. Facial alignment guarantees the same pose and scale thus the model training is less affected by the variability of different frames. Additional advanced preprocessing may involve multimodal feature extraction. By way of example, the spatial features for objects and textures are derived from the pretrained models like ResNet or Inception whereas the temporal features that capture the movement and the dynamics come from 3D CNNs, LSTMs, or transformer-based models like Times former. The audio streams are isolated for deepfake videos with manipulated speech and usually, this is done by means of Mel-Frequency Cepstral Coefficients (MFCCs) or deep audio embeddings extraction. At the same time, dimensionality reduction techniques like principal component analysis (pca) are implemented to shorten the feature spaces that are large thus the computational load is decreased though the discriminative information necessary for robust detection is still preserved. Quality control in preprocessing means that there is no corrupt or incomplete data, the frames with occlusions are filtered, and the low-resolution content or artifacts are removed. The standard augmentations such as horizontal flipping, random cropping, rotation, and color upset are used to enlarge the training set and make it possible for the deep learning models to generalize the manipulations not encountered previously. Preprocessing pipelines could also eliminate the interaction of feature extraction and indexing which implies that the original frames and the processed embeddings are saved separately, the model may thus be updated in a sequence without retransformation of the data [29,30,31,32,33].

10. Improvement of Different Preprocessing Methods Deepfake Videos Detection

Different preprocessing techniques significantly influence deepfake detection accuracy, affecting both baseline performance and generalization to new manipulations. Face detection cropping, and alignment ensure models focus on manipulated regions, with studies showing substantial accuracy gains when preprocessing isolates faces before classification. Normalization steps, such as resolution standardization and pixel value scaling, reduce domain shift and improve cross-dataset results, supporting models in adapting to varied video sources. Data augmentation pipelines incorporating color jitter, rotation, compression artifacts and synthetic transformations consistently boost generalization with the addition of diverse augmentations raising detection AUC by up to 9% across benchmarks. Not all augmentations are equally beneficial and careful selection is crucial; for instance, aggressive affine transformations may decrease performance while Gaussian blur and autoencoder-based augmentation can enhance robustness under noisy conditions. Pairing real and fake samples from the same source, as well as including diverse content in training mitigate shortcut learning and foster better generalization to new fabrication techniques [34,35,36,37,38]. Deepfake detection accuracy as well as the ability to handle new types of attacks can be greatly enhanced by the use of very specific preprocessing steps that involve face region isolation, normalization, tailored augmentation and temporal selection.

11. Preprocessing Techniques Vary Between Image and Video Deepfake Detection

Preprocessing techniques for image and video deepfake detection differ primarily in handling the temporal dimension and data structure. For image-based detection, preprocessing focuses on single images face detection alignment, cropping, normalization (resizing, pixel scaling), and possibly color space adjustment are standard steps. Noise reduction and image sharpening techniques such as unsharp masking may also be applied to enhance artifacts indicative of manipulation. On the other hand, video deepfake detection needed additional temporal preprocessing steps. First, videos are broken down into frames and then each frame undergoes face detection and alignment. To address computational resource limitations, key frame selection or frame sampling might be used to eliminate frames with redundant information and keep only representative frames for further processing. Besides that, temporal normalization, for example ensuring that each video clip has the same number of frames, is very important for the compatibility of input with sequential models such as RNNs and transformers. Video-specific preprocessing might also involve locating facial landmarks in different frames to detect movement consistency, using scene detection to change the sampling rate, and combining frame-level predictions to get a video-level decision. In the end, even though both image and video pipelines have some common spatial preprocessing steps, the video pipeline specifically deals with temporal coherence, sequence organization, and frame aggregation, thus allowing the use of motion cues and spatiotemporal inconsistencies arising from deepfakes [39,40].

12. The Accuracy of Difference Between Deepfake Detection Algorithms

Deepfake detection algorithms mostly employ deep neural networks that differ widely in their accuracy. The accuracy of these methods depends on the extent to which the models can detect subtle spatial, temporal, and physiological cues in the videos or images. In particular, variations of EfficientNet, most notably EfficientNe have dramatically improved performance with accuracy figures of more than 97% being reported on common benchmarks. The reason for this is their precisely tuned scaling and compound structure, which provides an optimal balance of model depth, width, and resolution. The main idea of FaceForensics++ is the use of ensembles and fusion features from multiple backbone networks; thus, it can be found second in the comparison with a reported accuracy of 96.8%. The method is capable of generalizing well across challenging deepfake manipulations such as face swapping and reenactment. Xception, a model that extensively uses depth-wise separable convolutions, is always around 96.5%, and it is the backbone of the detection in the popular FaceForensics++ benchmark used to show the model's stability against compression artifacts and various attack methods. CapsuleNet, which involves capsule-based routing mechanisms, reaches 96.0% by understanding the natural part–whole relationships in the facial structures thus, it is difficult for adversarial perturbations to fool it. YOLOv8's real-time architecture for object detection and now used for face tampering, provides quick inference with an accuracy of up to 95.5%, which is very advantageous for edge device usage. MesoNet and VGGNet, the first CNN-based models for deepfake detection that are still on par with the state of the art, have an accuracy of about 94-95% and are thus, quite popular due to their low complexity and easy deployment, although they are slightly less powerful in confronting subtle attacks. On the temporal side, recurrent models like LSTM (Long Short-Term Memory) and GRU (Gated Recurrent Unit) prove that using frame-sequential data can enhance video-level detection; nevertheless, their accuracy is typically lower than that of cutting-edge CNN with LSTM being approximately 93.2% and GRU - 92.8%. Furthermore, physiological-based methods, e.g., those employing remote photoplethysmography to derive signals, open up an additional angle of the problem by estimating the heartbeat-driven facial blood flow with the accuracy of up to 93.9% but these are still extremely limited when the face is hidden and in environments with lots of noise [41,42] Shown in Figure 1.

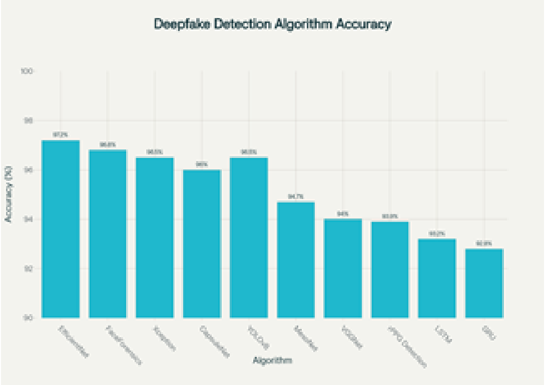

Figure 1 The Level of Accuracy Changes in Deepfake Detection Algorithm and Data Set

13. Experimental Setup

Very large data sets such as FaceForensics++ are or reflect the experimental configuration of deepfake detection. There are CelebDF that have been broken down into training, validation,and test sets. The preparation for the experiment includes frame extraction, face alignment, cropping, and normalization. The backbone models (EfficientNet, ResNet, or hybrid CNN-RNNs) are trained with the help of these prepared data. Data loads and feed input through the network and loss functions such as cross-entropy are employed along with optimizers models. The models' effectiveness is indicated by the metrics for instance, accuracy, F1-score, and AUROC; besides this, ablation studies show the architectural impacts. Cross-dataset tests serve the purpose of generalization of models and their robustness in the real world. A deepfake detection system is a sequential process that merges several sophisticated AI and computer vision components. It starts with data acquisition when extensive datasets comprising of both authentic and manipulated media are gathered for training and testing. Further data preprocessing is done for extraction and alignment of faces; image resolution is normalized and frames from videos are taken for temporal analysis. After that machine learning and deep learning algorithms like CNN, RNN, or transformers are used to extract spatial and temporal features. These features are then used to find the irregularities in facial movements, lighting or expression, and the synchrony of the frames. The system goes on to check the audio-visual consistency by comparing the lip movements with the given video to find the mismatches, which indicate the forgery. Some deepfake detection technologies in addition to the above-mentioned method of checking lip movement also check the frequency of eyes blinking to fool detection. A person who is watching can hardly see the fake yet this technique provides an additional layer of detection against very high-quality fakes. Among the common deepfake detection tools, one can mention FaceForensics++ which is a benchmark suite most often used for the training and evaluation stages of algorithms dealing with manipulated image and video datasets. Deep ware Scanner and Sensity AI provide commercial APIs facilitating quick detection and forensic analysis of the suspicious media in a real-time environment. Open-source resources like DeepFaceLab support the research and adversarial training community by providing the means for the controlled creation of deepfake samples, while the platforms of MesoNet and Xception serve as the base models for the feature extraction and anomaly detection process [43,44,44,45,46].

14. Results and Discussion

Significant improvements in classification accuracy have been made through recent deepfake detection innovations. These advancements have also, however, pointed out certain issues such as dataset quality, lighting conditions and computational scalability that are still considerable. Usually, state-of-the-art models like transfer learning frameworks and ensemble deep neural networks obtain more than 90% accuracy on standard benchmarks such as Celeb-DF and FaceForensics++ even when there are variations in lighting conditions. The detection methods find it easier to operate on bright images because they possess a greater signal to noise ratio as well as heavily textured images that suggest the presence of the artifacts in a manipulation attempt. One of the ways to make the method more sensitive to very tiny forgery traces and, at the same time, reduce the computational load is by changing the color channels and reducing the number of early pooling layers in the detection method, as identified by researchers. Even the video level accuracy can be quite high and the time required in the inference can be significantly reduced using confidence aggregation schemes, e.g. dynamic frame sampling with efficient encoding. However, deepfake generation technologies are quickly improving, and poor uncertainty quantification is highly important to the systems to be deployed in the real world. The findings render technical diversity and adaptation of testing environments to be among the most important factors in the facilitation of reliable and scalable deepfake detection.

Conclusion

With the emergence of generative diffusion models and the classical, the deepfake space has significantly changed evidence like pixel noise or lighting discontinuities comes almost to a standstill. Consequently, the modern-day detection systems focus on locate dynamic approaches, including spatiotemporal pattern learning, adaptive training pipelines and explainable AI systems that provide interpretability of classification results in a verifiable way. Serving as an ethical concern, the AI community expressed the importance of exercising ethical AI practices other than detection accuracy, particularly in sectors with a high-risk risk like journalism, finance, and governance. The detectors should be always a step ahead of the forgers since, deepfake creation tools are getting quicker and more accessible to the public. So, detection systems need to be continuously updated through dynamic learning, cross-domain datasets, and adversarial retraining. The sole dependence on isolated models is gradually being overhauled as detection success is proven to hinge on globally collaborative frameworks, real-time benchmarking, and embedded safeguards within content platforms and devices. Later on, research on deepfake detection will be centered around the creation of solid multi-modal models that can incorporate the cues from images, videos, and audios to improve detection accuracy and real-time responsiveness. Expanding datasets and cross-domain benchmarking will be instrumental in enhancing generalization in the real world, whereas adversarial robustness, and automatic thresholding will be the key factors in combating sophisticated attacks.

References

[1]. E. Altuncu, et al., "Deepfake: definitions, performance metrics and standards,"2024[OnlineAvailable: https://pmc.ncbi.nlm.nih.gov/articles/PMC11408348/

[2]. S. Son, et al., "Enhancing Deepfake Detection: Spatial-Temporal Preprocessing," ACM, 2023

[3]. A. Malik, et al., "DeepFake Detection for Human Face Images and Videos," IEEE Access, vol. 10, pp. 18757–18775, 2022

[4]. ACM, "A Large-Scale Dataset for Evaluating Video Deepfake Detection," ACM, 2025

[5]. A. Qadir, et al., "An efficient deepfake video detection using robust deep learning approaches," 2024

[6]. B. Spatiotemporal Feature Analysis, "Exploring Deepfakes - Creation Techniques, Detection Strategies and Emerging Challenges," IJRASET, 2023

[7]. MFA Sheikh, et al., "Analyzing the Impact of Deepfake Videos on Social Media," IEEE, 2024

[8]. C. Chen et al., “A New Dataset for Explainable Deepfake Detection in Video,” arXiv:2503.14421, 2025

[9]. L. Berjawi et al., “Optimization of DeepFake Video Detection Using Image Preprocessing,” ACTEA, 2024

[10]. Deepfake Datasets: Role and Practical Application — Antispoofing Wiki,”2023. [Online]. Available: https://antispoofing.org/deepfake-datasets-role-and-practical-application/

[11]. J. Upadhyay et al., “Deepfake Detection: A Comprehensive Review of Techniques and Challenges,” International Journal for Multidisciplinary Research, vol. 7, no. 2, 2025

[12]. P. Terhoer, “Massively Annotated DeepFake Databases,” IEEE Trans. on Technology and Society, 2024.

[13]. https://github.com/pterhoer/DeepFakeAnnotations

[14]. K. Korshunov and S. Marcel, "DeepFake video detection: Insights into model generalisation," Sci. Direct, 2025

[15]. S. Yang et al., "CrossDF: Improving Cross-Domain Deepfake Detection with Deep Information Decomposition," Frontiers in Big Data, 2022.

[16]. A. Nadimpalli and A. Rattani, "On Improving Cross-dataset Generalization of Deepfake Detectors," arXiv:2204.04285, 2022

[17]. Y. Zhou and X. Li, "Deepfakes: A New Challenge for Face Recognition and Beyond," in IEEE Transactions on Information Forensics and Security, vol. 14, no. 11, pp. 2867-2878, Nov. 2019.

[18]. A. Lyu and J. Farid, "Exposing Deepfakes Using Inconsistent Head Poses," in IEEE Transactions on Information Forensics and Security, vol. 15, no. 1, pp. 1-12, Jan. 2020.

[19]. J. Chen et al., "Deepfake Detection: A Survey," in IEEE Transactions on Circuits and Systems for Video Technology, vol. 31, no. 4, pp. 1234-1248, April 2021.

[20]. R. Roessler et al., "FaceForensics++: Learning Robust Models for Fake Video Detection," in Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV), Seoul, South Korea, Oct. 2019, pp. 1038-1047.

[21]. D. Cozzolino, L. Verdoliva, and G. Poggi, "In the Wild Deepfakes Detection Dataset," in Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), Long Beach, CA, USA, June 2019, pp. 0-0.

[22]. A Comprehensive Benchmark of Deepfake Detection. (2023). NeurIPS Proceedings. Retrieved from https://proceedings.neurips.cc/paper_files/paper/2023/file/0e735e4b4f07de483cbe250130992726 PaperDatasets_and_Benchmarks.pdf

[23]. Y. Yu et al., “MSVT: Multiple spatiotemporal views transformer for deepfake video detection,” IEEE Trans. Circuits Syst. Video Technol., 2023. [Online]. Available: https://doi.org/10.1109/TCSVT.2023.3281448

[24]. A. Heidari, "Deepfake detection using deep learning methods: A Review," Wiley Online Library, 2024.

[25]. V. Hondru et al., "A New Dataset for Explainable Deepfake Detection in Video," arXiv preprint arXiv:2503.14421, Mar. 2025. [Online]. Available: https://arxiv.org/html/2503.14421v1

[26]. OpenFake Team, "OpenFake: An Open Dataset and Platform Toward Large-Scale Deepfake Detection," arXiv preprint arXiv:2509.09495, 2019. [Online]. Available: https://arxiv.org/html/2509.09495v1

[27]. J. He et al., "Normalization-consistent data curation for generalizable deepfake detection," Neurocomputing, vol. 567, Oct. 2025. [Online]. Available: https://www.sciencedirect.com/science/article/abs/pii/S092523122502510X

[28]. Deepstrike.io, "Deepfake Statistics 2025: AI Fraud Data & Trends," Sep. 2025. [Online]. Available: https://deepstrike.io/blog/deepfake-statistics-2025

[29]. Milvus Team, “What are the best practices for video data preprocessing in search pipelines?”, Milvus.io, Oct. 2025. [Online]. Available: https://milvus.io/ai-quick-reference/what-are-the-best-practices-for-video-data-preprocessing-in-search-pipelines

[30]. S. Sharma et al., "Design and development of an efficient RLNet prediction model for video deepfake detection," Frontiers in Big Data, 2025. [Online]. Available: https://www.frontiersin.org/journals/bigdata/articles/10.3389/fdata.2025.1569147/full

[31]. M. Alrashoud, "Deepfake video detection methods, approaches, and challenges," J. King Saud Univ.-Comput. Inf. Sci., 2025. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S111001682500465X

[32]. W. Choi et al., "Enhancing practicality and efficiency of deepfake detection using multimodal temporal modeling," PMC, Dec. 2024. [Online]. Available: https://pmc.ncbi.nlm.nih.gov/articles/PMC11680869/

[33]. K. Maharana, "Data pre-processing and data augmentation techniques," Data Science Journal, 2022. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S2666285X22000565 [34] T. Nguyen, J. Yamagishi, and I. Echizen, “Deep learning-based approach for detecting manipulated facial videos,” J. Artif. Intell. Res., 2021.

[34]. P. Charitidis et al., "Investigating the Impact of Pre-processing and Prediction Aggregation on the DeepFake Detection Task," arXiv:2006.07084, 2020. [Online]. Available: https://arxiv.org/abs/2006.07084

[35]. J. Bondi et al., “Training Strategies and Data Augmentations in CNN-based DeepFake Video Detection,” Sci-Hub Red, 2021. [Online]. Available: https://sci-hub.red/downloads/2021-05-26/0c/bondi2020.pdf

[36]. L. Mihai et al., “Autoencoder-based Data Augmentation for Deepfake Detection,” AIMultimediaLab, 2023. [Online]. Available: https://bionescu.aimultimedialab.ro/index_files/pub/AE_Augmentation_MAD_23.pdf

[37]. L. Yan et al., "Deepfake Detection that Generalizes Across Benchmarks," arXiv:2508.06248, 2023. [Online]. Available: https://arxiv.org/html/2508.06248v1

[38]. A. S. Charitidis, E. Kordopatis-Zilos, I. G. Tsironi, et al., "Combating Digitally Altered Images: Deepfake Detection," arXiv preprint arXiv:2508.16975, May 2020. [Online]. Available: https://arxiv.org/html/2508.16975v1

[39]. X. Li, T. Nguyen, Y. Wang, et al., "Enhancing Deepfake Detection: Spatial-Temporal Techniques for Videos," ACM Trans. Multimedia Comput. Commun. Appl., 2024. [Online]. Available: https://dl.acm.org/doi/fullHtml/10.1145/3639592.3639597

[40]. S. Hossen, S. Rahman, “A survey on deep-fake detection algorithms,” World Journal of Advanced Research and Reviews, vol. 25, pp. 168–176, 2025. [Online]. Available: https://journalwjarr.com/sites/default/files/fulltext_pdf/WJARR-2025-2251.pdf

[41]. S. Bonk et al., "Comparative Analysis of Deepfake Detection Models on FaceForensics++ Dataset," hrcak.srce.hr, 2025. [Online]. Available: https://hrcak.srce.hr/file/471370

[42]. S. Sharma et al., "Design and development of an efficient RLNet prediction model for video deepfake detection," Frontiers in Big Data, vol. 8, 2025. [Online]. Available: https://www.frontiersin.org/journals/bigdata/articles/10.3389/fdata.2025.1569147/full

[43]. Innovatrics, “Deepfake Detection,” Oct. 2025. [Online]. Available: https://www.innovatricsInnovatrics, “Deepfake Detection,” Oct. 2025. [Online]. Available: https://www.innovatrics.com/glossary/deepfake-detection-technology/.com/glossary/deepfake-detection-technology/

[44]. DSCI, “Advanced Tools and Techniques for Deepfake Detection,” May 2022. [Online]. Available: https://ccoe.dsci.in/blog/Deepfake-detection

[45]. S. Nair et al., “A Comprehensive Review on Deepfake Generation, Detection, Challenges, and Future Directions,” IJRASET, May 2025. [Online]. Available: https://www.ijraset.com/best-journal/a-comprehensive-review-on-deepfake-generation-detection-challenges-and-future-directions

[46]. S. Bose et al., “Deepfake Detection Systems: A Comprehensive Survey of Algorithms and Techniques,” Journal of AI Research, vol. 12, no. 3, Apr. 2025. [Online]. Available: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=5240416