# FORGERY DETECTION OF LOW QUALITY DEEPFAKE VIDEOS

*M. Sohaib*[*], *S. Tehseen*[*]

**Abstract:** The rapid growth of online media over different social media platforms or over the internet along with many benefits have some negative effects as well. Deep learning has many positive applications like medical, animations and cybersecurity etc. But over the past few years, it is observed that it is been used for negative aspects as well such as defaming, black-mailing and creating privacy concerns for the general public. Deepfake is common terminology used for facial forgery of a person in media like images or videos.The advancement in the forgery creation area have challenged the researchers to create and develop advance forgery detection systems capable to detect facial forgeries. Proposed forgery detection system works on the CNN-LSTM model in which we first extracted faces from the frames using MTCNN then performed spatial feature extraction using pretrained Xception network and then used LSTM for temporal feature extraction. At the end classification is performed to predict the video as real or fake. The system is capable to detect low quality videos. The current system has shown good accuracy results for detecting real or fake videos on the Google deepfake AI dataset.

Key words: *convolutional neural networks, recurrent neural networks, long short term memory, multi-task cascaded neural networks*

*Received: April 21, 2021*
*Revised and accepted: April 30, 2023*

**DOI:** 10.14311/NNW.2023.33.006

## 1. Introduction

The media is the face of the state and media is the one which timely informs if any important event happens around the globe. Media plays a vital role in educating the people and also create awareness. Imagine if a government person speaks against the other country or order an attack against someone which might prove to be fake but widely shared online without checking the resource and its authenticity could result in devastating consequences which can lead to social media and 5-th generation warfare which is a major issue going on right now. The manipulations of images and videos have started way back when tools like Photoshop and video editing software were released. In the past, these tools have been used against mainly for black mailing people and defaming. The researchers over the past few

[*] Muhammad Sohaib; Samabia Tehseen – Corresponding author; Bahria University Computer science Department, Islamabad Pakistan, E-mail: 01-249182-013@student.bahria.edu.pk, stehseen@bahria.edu.pk, stehseen.buic@bahria.edu.pk

years have explored different ways for detecting these manipulations. But with the current advancement in technology, there is an urgent need that state of the art techniques are used to detect and stop the spread of forged images and videos.

Over the past decade, rapid growth in technology and social media is observed. Millions of images and videos are being uploaded daily and data is growing exponentially. The tremendous use of digital data has also given rise to the digital manipulation of these images and videos. This situation motivates for an urgent development of a forensic research authority which can tackle these issues. Since last few years, it is seen that multimedia forgery is expanding rapidly commonly named as 'deepfakes' on different platforms [22, 14]. It makes it almost impossible to identify the real and fake content through human eye inspection. The word 'deepfakes' refers to deep learning and fake which involves machine learning techniques to create and detect fake videos.

However, the tools and techniques required to create deepfakes are open sourced and are freely available online. FaceApp is a popular open source tool for face swapping [23]. It allows user to apply transformation to face images and let the user for change in hairs, gender and age attributes. FaceApp lets you create masks which can swap face on videos with artificial intelligence face2face technique [20]. face2face is a real time facial capture and reenactment method for videos. Faceswap [10] is another deep learning based technique which can swap the faces in the images and it is done via generative adversarial networks (GANs). Fundamental work flow of GANs for deepfakes is graphically presented in Fig. 1.

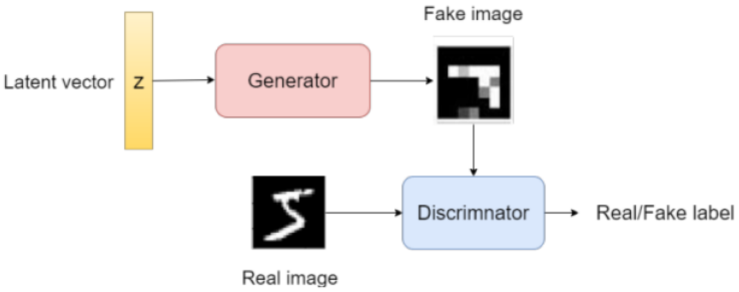


**Fig. 1** *Generative adverserial networks for deepfakes.*

Google CEO Sundar Pichai said, "Detecting deepfakes is one of the most important challenges ahead of us." Because of the importance of detecting deepfakes, tech giants are stepping in this area as well recently. Google has released a deepfake dataset [6] that can be used as benchmark which will help to detect fake videos and would help researchers to develop best possible deepfake detection model. Similarly Facebook teamed up with Microsoft corporation to launch an AI competition and Facebook is putting 10 million dollar into the kaggle "Deepfake Detection Challenge", which encourage researchers to work in the deepfake detection area.

The rapid growth of users in social media platforms are facing a challenge of fake images and videos. It is also posing as a threats to the people like defaming and blackmailing. The videos and images available online can be used for manipulation using different vision and deep learning techniques. These manipulations are difficult to identify by a normal person and it will be a challenging task to detect such videos. Artificial intelligence techniques are being misused by the people in generating fake images and videos because these methods are capable of creating fake facial masks. If these videos are of low quality it is difficulty level also increases because detection of facial landmarks become a great challenge for deep learning model because important features might be lost. To identify low quality fake compressed videos generally of news and people on social media platforms is difficult and rapid detection methods are required for these fake videos for early detection.

Most of the research work for fake detection is targeted on images only. While work for videos is being going on with slow pace. The main challenges when dealing with videos are huge amount of data to process, temporal correlations, low resolution and non-availability of data. Recently the data problem is solved as Google has released a large video dataset of high quality and low quality deepfakes. Low quality videos in which it is difficult to detect face for fake videos therefore an efficient algorithm is required to tackle this problem.

The development of a deep vision algorithm for detecting fake videos will help the people from getting defamed or blackmailed by any other person or video being challenged in the court of law and evidence cannot be used against him or her. This algorithm will also help to detect videos and news on social media platforms and prevent its rapid sharing by people which will overcome the concerns of Microsoft, Facebook and Google which thinks rapid sharing of deepfakes is a threat to users. To handle these issues the proposed work will identify Fake videos with the help of deep learning and vision techniques. Proposed method will detect forged videos based on classified features extracted by CNN and after that LSTM feature extraction and classification will identify if video is fake or real.

To allow researchers to further work in this area and to explore the deepfake dataset this paper has following contributions:

- Development of a fake video detection model which detect low quality videos using spatial and temporal feature extraction,
- The images used is of lower size 160 × 160 to check how well our model performs on lower resolution and small size of image as input,

Proposed architecture based on neural networks will extract facial features from video frames based on temporal features proposed architecture will identify facial landmarks and detection of manipulated areas in video to determine the probability if the video is original or fake.

## 2. Related work

Current deep learning and vision techniques are being used for manipulating a face in a video and spreading fake news on different online platforms. Face plays a vital role for identification and it can get manipulated with current algorithms it is easy to manipulate an image or video and changing a face in video without affecting audio and face movement [17]. Dataset plays an important role in training the model to identify fake videos and testing your model for accuracy. The quality of video plays a vital role in detection of deepfakes low resolution of synthesized faces, color inconsistency by which major facial details were lost are easy to identify missing facial skin details [17, 18]. These effects can be seen in Fig. 2. Visible facial parts help the AI model to figure out fake targets easily and to train a model efficiently [21].

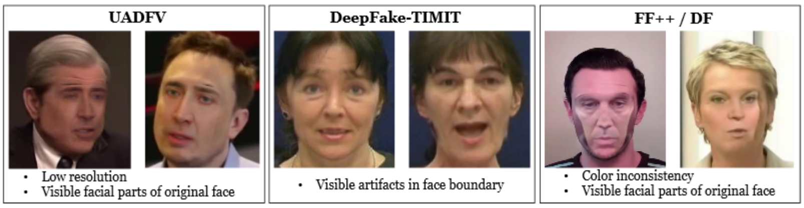

UADFV
- Low resolution
- Visible facial parts of original face
DeepFake-TIMIT
- Visible artifacts in face boundary
FF++ / DF
- Color inconsistency
- Visible facial parts of original face

**Fig. 2** *Low quality effects in video.*

As we are entering into artificial intelligence era we are seeing advancements in the technology which are for good and bad like the use of generative adversarial networks which were used for creating animations and arts. These networks are being misused today for manipulating videos and are used for facial forgeries in videos [8].

## 2.1 Deepfake creation

The fake creation is also done through paired encoder-decoder [17] as shown in Fig. 3 which creates a great concern about how a good technology is being used for a bad purpose. Today we are seeing AI generative models are passing beyond researcher's scope and are posing challenge for detection of fake facial replicas. An approach presented by the authors to detect synthesized videos is based on biological signals of face after which they will transform and extract features to classify if the video portraits are real or fake which will add the contribution in the area of fake videos detection [17].

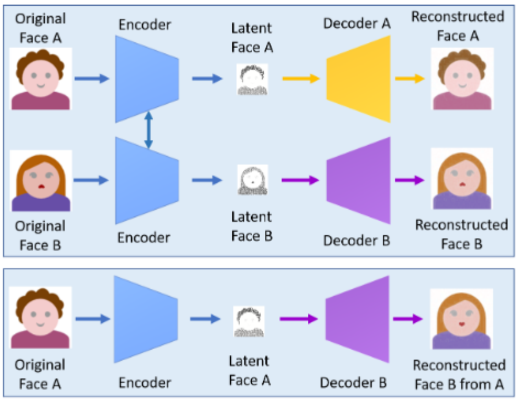


**Fig. 3** *Face swapping using auto encoders.*

The approaches for manipulating images and videos are improving a lot. Therefore exposing these fake videos and images becomes an urgent need of day. Different clues are used for detection of deepfakes like fake parts of the face, head orientation and position [24]. Segmenting approach can also be used for detecting manipulated images which commonly detects removal, copy-move and splicing attacks. Similarly semantic segmentation can be used to detect spoofing regions which returns boxes representing manipulated regions [24]. Forged videos can also be detected through eye blinking which captures eye blinking state from the video frames and can determine the originality of video [11]. Local noise analysis can also help in capturing hidden tampered areas and can be vital for increasing the efficiency to detect fake images [9].

## 2.2 Deepfake detection

Detection of fake news is getting difficult now a days with rapid sharing on different platform and its detection is becoming a challenge for researchers. Current deep learning and vision models are detecting high quality forged videos with great accuracy. But when it comes to detect face landmarks in low quality compressed videos make it difficult for the model to detect with great accuracy. With current rise in forged videos on online platforms, an urgent need of AI model is required to detect every type of forged video and stop it's rapid sharing on online platforms. There are various type of facial manipulations which could help to detect fake videos, few types are full facial synthesis, identity swapping, attribute manipulations, expression swapping [21]. These manipulations can be seen in Fig. 4. Different approaches are used for manipulations using GAN architectures [8] are famous for the manip-

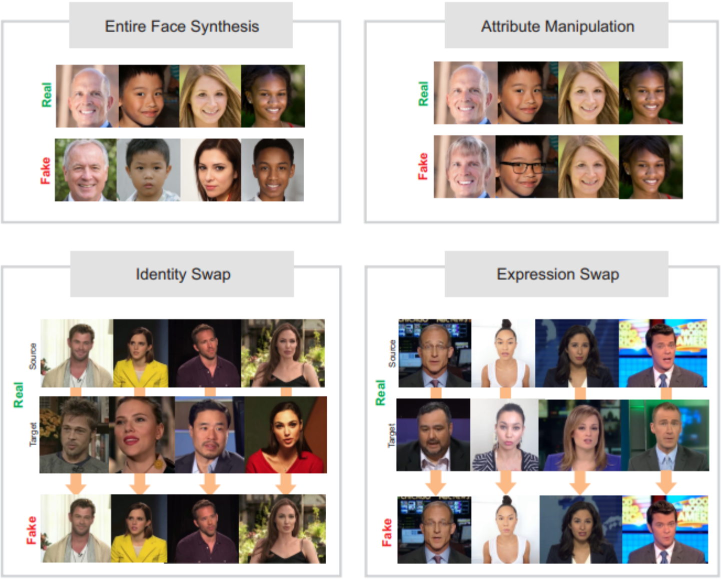


**Fig. 4** *Facial manipulations.*

ulation of full face producing high quality face images. Celeb-DF [13] and FF++ [18] are databases which consists of facial manipulations. Identity swap on the other hand replaces a face of a single person in a video with another person. Face Swap [10] a computer based manipulation app is used for these type of manipulation. Attribute manipulation consists of manipulating human characteristics such as hair, color, age etc. FaceApp is an example of this manipulation. Expression Swap refers to manipulate a facial expression of a person face2face is an example of this type of manipulation.

In the study [21] few important detection techniques mentioned which highlight the inconsistencies left behind in creating the fake videos and can be used as a detection feature are:

- Deepfake videos are mostly facial manipulated most attention should be given to facial manipulations.
- Facial affected areas are often cheeks, skin attribute manipulations and incongruousness on different dimensions.
- Eye blinking is an important factor and fake videos lack this attribute.
- Shadow and lightening can play an important role in detection mostly of eyebrows.
- Facial hair in forged videos lacked the naturalness and can be detected.

Different techniques have also been used to detect deepfakes and the technology is evolving from time to time like visual artifacts [12] eye blinking [11] head pose movement [24] which identifies movement distance between synthesized and original head poses. Two-stream CNN network [9] which uses a classification model of LeNet based architecture for model training. MesoNet [1] based on shallow architectures having inception module tends to learn different features from the frames. Matern [15] proposed deepfake detection system detect synthesized face based on missing reflections and details of the eye. Deepfakes created by splicing synthesized face regions may introduce errors and can be detected by 3D head pose estimations. Agarwal and Farid [2] proposed method works both with facial expression like facial muscles such as nose, cheek area and head movements distance which was classified using SVM for final classification.

Missing reflection and illumination details in the eyes or teeth features plays an important role in detection of synthesized videos. On the other side some researched work [16] also shown the out performance of deep architectures over the shallow network. Different deepfake detection approaches used can be effective for high quality videos but poses a challenge for low quality videos. Proposed method of Guera and Delp [7] highlighted that deepfake videos contain intra-frame inconsistencies and temporal inconsistencies between frames. Which can be detected effectively in High quality videos but will cause problem in detection of low quality videos. Yisroel mentions mouth, gaze, pose, expression and body as key factors for deepfake detection [16].

Deepfake created often leave some artifacts which might be difficult for human to recognize but can easily be recognized by machine. A hypothesis was also proposed in 2014 by researchers that computer generated faces can be detected by heart rate signals, blood volume patterns in skin [5]. Inconsistencies like lightening and color remained after deepfakes creation is an important factor for detection and quality measure and frequency analysis can be used to detect these inconsistencies [5]. Tampered facial areas can be exposed by predicting masks learned from ground truth.

### 2.3 Deepfake datasets

Celebrities are greatly affected by the fake videos a new dataset celeb-DF was launched to help researchers to develop a model to detect fake celebrity images and videos it has improved many factors as compare to other datasets UADFV, Vid TIMIT like low resolution of synthesized faces, color inconsistency by which major facial details were lost because of low quality data.

The Google deepfake AI dataset [6] used consists of 360 real and 360 fake video dataset and is divided into two classes fake and real having total size of approx. 4 GB consisting of 28 different actors on different scenes of the videos. These videos are from 28 actors being available in raw (original form), high quality (compressed 23 times) low quality (compressed 40 times).

FaceForensics++ [18] first generation dataset contains 1000 real and 4000 forged videos the manipulated videos have been generated using state of the art techniques based on deepfakes, face2face, face swap and neural textures as prominent representatives for facial manipulations at random compression level.

Celeb-DF [13] contains both real and synthesized videos total 795 fake and 500 original videos collected from YouTube on different subjects created through public having standard 30 fps frame rate.

Deepfake TIMIT [19] is a dataset consist of faces in video which are swapped using GAN manipulation. This dataset consists of 16 similar pair with 32 subjects trained on low quality image model and high quality image model consisting of 10 videos per person and total of 320 videos total with no manipulation in voice.

The advancements in creation of deepfakes is making the detection of deepfakes difficult. In our research we have found that most of the research is going on in high quality deepfake videos because of high accuracy of results but the accuracy is low in case of low quality compressed videos and it also makes it difficult for the detection model to classify and needs enhancement in the fake detection pipeline because the detection model must be capable of detecting low or high quality deepfakes in real world scenario to stop it's rapid share. The other downside of detection model is that it might not show good results if it has been given a manipulated video created of an unseen technique

## 3. Methodology

In this section we will discuss about the working of our proposed methodology. The proposed architecture can be seen in Fig. 5. Proposed methodology can be divided into four major steps including data pre-processing, spatial feature extraction, temporal feature extraction and classification. Details of each phase is described in the following sections.

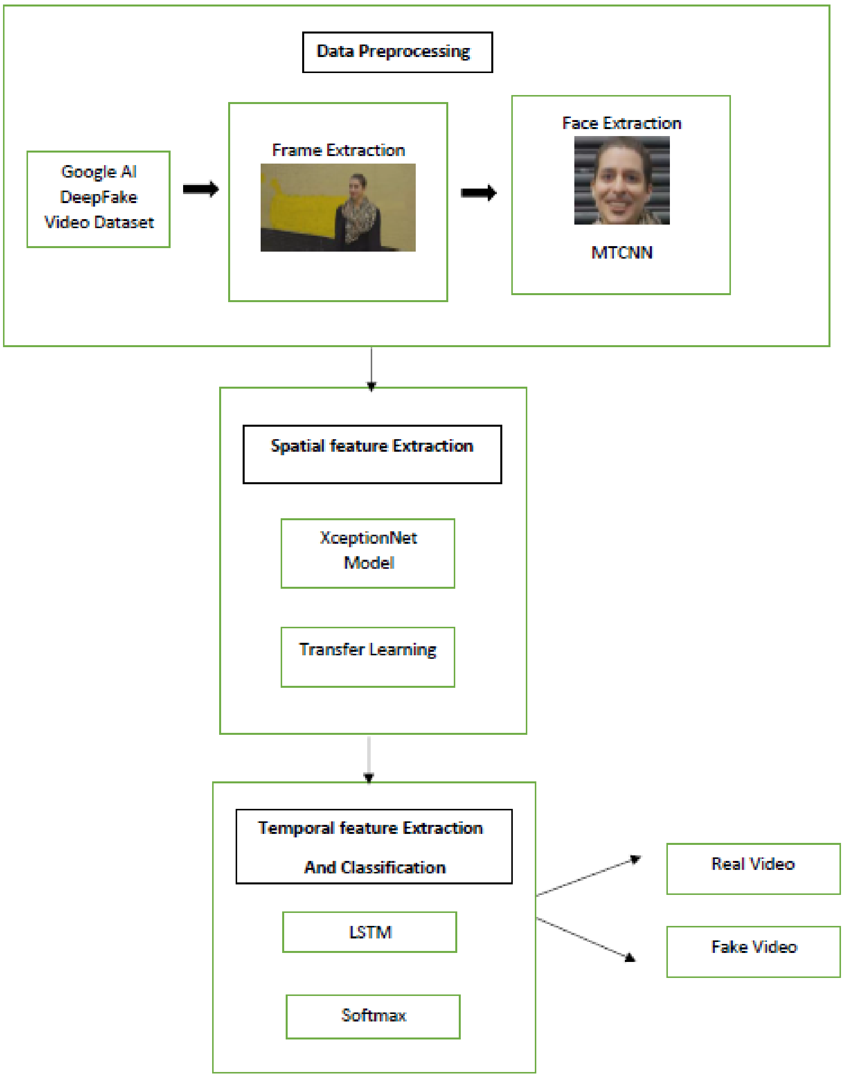


**Fig. 5** *Proposed architecture.*

## 3.1 Data preprocessing and preparation

We have used Google deepfake AI dataset consisting of 360 real and 360 fake video dataset is used which is divided into two parts labeled as 0 and 1 where 0 refers to fake class and 1 refers to real class. Training/testing data is split into 80:20 ratio. Moreover, data is preprocessed by extracting frames from the videos at the frame rate of 5 to capture frequent change. In the next step faces have been extracted from the frames using MTCNN (multi-task cascaded convolutional networks) [25], as shown in Fig. 6. MTCNN captures faces effectively during extraction.

In the proposed methodology the input data of frames were passed as input to the MTCNN architecture which consist of three networks P-Net, R-Net, O-Net. P-Net works by taking input as a 12 × 12 kernel that goes over the image in order to find the face. The output of P-Net is passed to R-Net as probability of face in bounding box along with the coordinates of box. R-Net further improves the coordinates of the bounding box. In the last network O-Net which takes the input of R-Net and output the following data

- Box: It returns the $x, y$, width and height of box
- Confidence: It returns the face matching probability
- Key points: It returns the key points for nose, left mouth, right mouth, left eye and right eye.

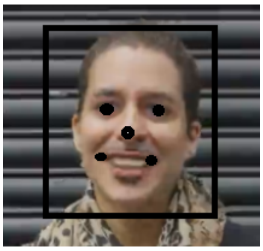


**Fig. 6** *MTCNN face extraction.*

## 3.2 Feature extraction and classification

In this study, We used CNN to extract features from the frames of videos while LSTM is utilized for temporal features and classification. LSTM can exploit the temporal inconsistencies due to the introduction of manipulation in videos.

### 3.2.1 Spatial feature extraction

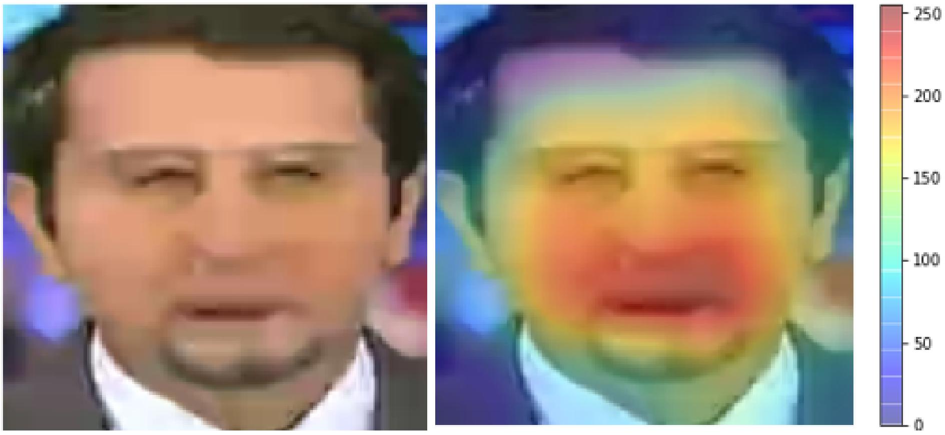
We have utilized the pre-trained Xception [4] model for spatial feature extraction because of it's great performance in recent studies [18]. Xception instead of splitting the data into blocks it maps spatial correlations for each output separately and further performs 1 × 1 depthwise convolution to map cross-channel correlations [4].

We have used the transfer learning approach for feature extraction by freezing the convolutional base because it saved a lot of training time and resources. After passing input to the the network we get the features from the output of convolutional base which played a vital role in spatial feature extraction . Xception uses weights pretrained on the ImageNet dataset. Since the ImageNet dataset is not similar to facial images we have used the weights fine-tuned on deepfake detection dataset's facial data [6]. This shifting increases the efficiency of network and extract features effectively . Xception model is inspired by the Inception based model. Xception network is 71 layers deep it has 36 convolution layers and 14 modules and trained on million of images of ImageNet database. Xception network consists of residual connections and the network is in the form of a linear stack of depthwise separable convolution layers. Xception network gives better results for new images and has shown better accuracy as compare to VGG-16, ResNet-152 and Inception V3 models [4].

In the first phase we have fine-tuned the CNN pretrained model Xception to learn the feature extraction and then used the model to extract feature from facial images and further fed this input to LSTM. The training of two components have been done separately due to low amount of facial data extracted from frames as input and the reason for training two components separately was to prevent the exclusion of useful hand designed components which might have lost in the end-to-end training.

The input data passed to the Xception network is of face image which is used to perform spatial feature extraction. The input images were of $160 \times 160 \times 3$ since we are not classifying and only extracting features we will extract spatial features in the end. Categorical cross entropy is used as a loss function the total parameters are 22 Million approximately. The features are stored in the disk in the sequence for every individual video.

Since global approach is used instead of local attribute recognition approach the results of full face can be checked using grad-CAM to identify the facial features using heat-maps as shown in Fig. 7.



**(a)** *Input image.* **(b)** *Heat-map of input image.*

**Fig. 7** *Facial feature result.*

### 3.2.2 Temporal feature extraction

In the following part, we utilized RNN variant LSTM and bi-directional LSTM network which is used to get the temporal information from input sequence in the proposed architecture. Since our dataset is large and network will require more and more memory to perform computations in this case LSTM can solve the problem due to its longer features storing advantage. LSTM works on the previously recorded features and pass them to next state. Extracted features are passed from CNN to this model. LSTM is trained on the sequential data and the weights of the model are saved which is used for temporal feature extraction to get better accuracy results. In our model the input shape for the LSTM is (25,512) where 25 represents the time frame and 512 are the features.

| Layer (type) | Output |
|---|---|
| Input layer | (25, 512) |
| LSTM | 32 |
| Dense | 2 |

**Tab. I** *LSTM layers in proposed architecture.*

| Layer (type) | Output |
|---|---|
| Input layer | (25, 512) |
| Bi-directional LSTM | 256 |
| LSTM | 32 |
| Dense | 2 |

**Tab. II** *Bi-directional LSTM layers in proposed architecture.*

### 3.2.3 Classification

After performing spatial and temporal feature extraction in the last phase we have performed the classification part using the activation function softmax using 2 neurons because the input is of 2D of target labels.

## 4. Result and analysis

We will discuss about the dataset used, experimental results along with the parameters used in the proposed approach. The main objective of the experiment is to identify forged videos based on hidden information in face.

The results were obtained using the Google AI deepfake dataset. The experiment is performed on the Nvidia Tesla K80 GPU. The model is trained on Adam optimizer with batch size of 32 and learning rate of 0.002.

In proposed architecture we have performed transfer learning using Xception network model which was used for spatial features extraction. To perform the experiment the input facial image of $160 \times 160 \times 3$ was passed to the model to extract features and pass these features as a sequence to LSTM to get temporal patterns from the video. We have performed the transfer learning and used categorical cross entropy loss. Early stopping was applied to avoid over-fitting. Evaluation metrics used are listed in equation 1–4. Aaccuracy was used to examine proportion of true results. $F1$ gives the success rate for *Precision* and *Recall.* In which *Recall* refers to the correct prediction ratio and *Precision* refers to the actual matching ratio. Below are the counters used in the metric calculations.

- True positive (TP): Deepfakes detected as deepfakes
- True negative (TN): Non-deepfakes detected as non-deepfakes.
- False positive (FP): Non-deepfakes detected as deepfakes.
- False negative (FN): Deepfakes detected as non-deepfakes.

$$Accuracy = \frac{\mathrm{TP} + \mathrm{TN}}{\mathrm{TP} + \mathrm{FN} + \mathrm{TN} + \mathrm{FP}} \tag{1}$$

$$Precision = \frac{\mathrm{TP}}{\mathrm{TP} + \mathrm{FP}} \tag{2}$$

$$Recall = \frac{\mathrm{TP}}{\mathrm{TP} + \mathrm{FN}} \tag{3}$$

$$F_1 = 2 \cdot \frac{Precision \cdot Recall}{Precision + Recall} \tag{4}$$

Proposed model has shown good evaluation results for bi-directional LSTM as shown below despite of low resolution video and the small size of facial images given as input to train the model. Tab. III shows the results on both the variants. Bi-directional LSTM is giving superior results. Results still needs improvement as 10 % are misclassified in the reported results. This is due the challenging research problem and artefacts due to low resolution videos.

| Model | Accuracy | Precision | Recall | F1-Score |
|---|---|---|---|---|
| LSTM | 88.0 | **1.0** | 0.78 | 0.87 |
| Bi-directional LSTM | **90.0** | 0.93 | **0.87** | **0.90** |

**Tab. III** *CNN-LSTM prediction results.*

Hyper-parameters plays a major role in training process and can be adjusted as per the requirements which helps to increase the efficiency. Different design parameters have been used to check the efficiency of model. Learning rate is one of the important factor higher learning rate can speed up the accuracy but lower Learning rate can be good but might lead to a chance of getting local minima. In this research, batch size of 32 and learning rate of 0.02 is used to get the accuracy of 90 %. We have performed experiment by changing frame per second sequence length 25 and 40 fps which is the input shape to LSTM and have achieved great results on the 25 fps.

Proposed model has shown great results as compare to deep distribution transfer technique [3] which is a new transfer learning approach. In our approach we have achieved higher accuracy in detecting low resolution videos because of efficient facial extraction technique along with spatial and temporal feature extraction technique.

| Model | Learning rate | Batch size | Accuracy | Precision | Recall | F1-Score |
|---|---|---|---|---|---|---|
| LSTM | 0.02 | 32 | 85.0 | 0.89 | 0.81 | 0.85 |
| LSTM | 0.02 | 64 | 83.8 | **1.0** | 0.68 | 0.81 |
| LSTM | 0.01 | 32 | 87.0 | **1.0** | 0.75 | 0.86 |
| Bi-directional | **0.02** | **32** | **90.0** | 0.93 | **0.87** | **0.90** |

**Tab. IV** *Parameter modification results.*

| Model | Accuracy |
|---|---|
| Bi-directional LSTM (Ours) | **90.0** |
| Deep distribution transfer [3] | 81.21 |

**Tab. V** *Comparison results.*

## 5. Conclusion and future directions

Currently the synthesized media creation is offering a great challenge for the researchers. There is an urgent need to detect these forgeries because the fake media creation tools are openly available and are advancing day by day. These fake contents must be detected through state of the art AI techniques. The main focus of our research is to detect the low resolution deepfake videos using effective facial extraction. The quality of deepfake video creation is continuously increasing and there is a need for effective detection method for low resolution videos along with high quality videos as well. Currently the researchers are focusing on the weakness of deepfake creation pipelines like color inconsistency, shadow and facial features. The proposed methodology has used spatial as well as temporal features for fake video detection. The proposed methodology has used CNN based Xception network with transfer learning for spatial feature extraction. Bi-directional LSTM is employed for temporal feature extraction. The future direction of this research can be the real-time forgery detection. Sequential networks like transformer networks can also be utilized for the forgery detection.

## References

[1] AFCHAR D., NOZICK V., YAMAGISHI J., ECHIZEN I. Mesonet: a compact facial video forgery detection network. In: *2018 IEEE International Workshop on Information Forensics and Security (WIFS)*, 2018, pp. 1–7.

[2] AGARWAL S., FARID H., GU Y., HE M., NAGANO K., LI H. Protecting World Leaders Against Deep Fakes. In: *CVPR Workshops*, 2019, pp. 38–45.

[3] ANEJA S., NIESSNER M. *Generalized Zero and Few-Shot Transfer for Facial Forgery Detection.* 2020. Available also from: arXiv: 2006.11863 (cs.CV).

[4] CHOLLET F. Xception: Deep Learning with Depthwise Separable Convolutions. In: *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2017, pp. 1800–1807. doi: 10.1109/CVPR.2017.195.

[5] CIFTCI U.A., DEMIR I., YIN L. FakeCatcher: Detection of Synthetic Portrait Videos using Biological Signals. *IEEE Transactions on Pattern Analysis and Machine Intelligence*. 2020, pp. 1–1, doi: 10.1109/TPAMI.2020.3009287.

[6] *Deep fake Dataset by google*. 2019. Available from: https://ai.googleblog.com/2019/09/contributing-data-\\to-deepfake-detection.html.

[7] GÜERA D., DELP E.J. Deepfake Video Detection Using Recurrent Neural Networks. In: *2018 15th IEEE International Conference on Advanced Video and Signal Based Surveillance (AVSS)*, 2018, pp. 1–6. doi: 10.1109/AVSS.2018.8639163.

[8] GOODFELLOW I., POUGET-ABADIE J., MIRZA M., XU B., WARDE-FARLEY D., OZAIR S., COURVILLE A., BENGIO Y. Generative adversarial nets. In: *Advances in neural information processing systems*, 2014, pp. 2672–2680.

[9] HAN X., MORARIU V., LARRY DAVIS P.I. Two-stream neural networks for tampered face detection. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*, 2017, pp. 19–27.

[10] KORSHUNOVA I., SHI W., DAMBRE J., THEIS L. Fast face-swap using convolutional neural networks. In: *Proceedings of the IEEE International Conference on Computer Vision*, 2017, pp. 3677–3685.

[11] LI Y., CHANG M.-C., LYU S. In ictu oculi: Exposing ai generated fake face videos by detecting eye blinking. *arXiv preprint arXiv:1806.02877*. 2018.

[12] LI Y., LYU S. Exposing deepfake videos by detecting face warping artifacts. *arXiv preprint arXiv:1811.00656*. 2018.

[13] LI Y., YANG X., SUN P., QI H., LYU S. Celeb-df: A new dataset for deepfake forensics. *arXiv preprint arXiv:1909.12962*. 2019.

[14] MARAS M.-H., ALEXANDROU A. Determining authenticity of video evidence in the age of artificial intelligence and in the wake of Deepfake videos. *The International Journal of Evidence & Proof*. 2019, 23(3), pp. 255–262.

[15] MATERN F., RIESS C., STAMMINGER M. Exploiting visual artifacts to expose deepfakes and face manipulations. In: *2019 IEEE Winter Applications of Computer Vision Workshops (WACVW)*, 2019, pp. 83–92.

[16] MIRSKY Y., LEE W. The Creation and Detection of Deepfakes: A Survey. *arXiv preprint arXiv:2004.11138*. 2020.

[17] NGUYEN T.T., NGUYEN C.M., NGUYEN D.T., NGUYEN D.T., NAHAVANDI S. Deep learning for deepfakes creation and detection. *arXiv preprint arXiv:1909.11573*. 2019, 1.

[18] ROSSLER A., COZZOLINO D., VERDOLIVA L., RIESS C., THIES J., NIESSNER M. Faceforensics++: Learning to detect manipulated facial images. In: *Proceedings of the IEEE International Conference on Computer Vision*, 2019, pp. 1–11.

[19] SANDERSON C., LOVELL B. *Deepfaek TIMIT Dataset*. (2009). Available from: https://www.idiap.ch/dataset/deepfaketimit.

[20] THIES J., ZOLLHOFER M., STAMMINGER M., THEOBALT C., NIESSNER M. Face2face: Real-time face capture and reenactment of rgb videos. In: *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016, pp. 2387–2395.

[21] TOLOSANA R., VERA-RODRIGUEZ R., FIERREZ J., MORALES A., ORTEGA-GARCIA J. Deepfakes and beyond: A survey of face manipulation and fake detection. *arXiv preprint arXiv:2001.00179*. 2020.

[22] WESTERLUND M. The emergence of deepfake technology: A review. *Technology innovation management review.* 2019, 9(11).

[23] WIRELESS LAB. *FaceApp Open-source software Tool available.* 2016. Available from: https://www.faceapp.com.

[24] YANG X., LI Y., LYU S. Exposing deep fakes using inconsistent head poses. In: *ICASSP 2019-2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2019, pp. 8261–8265.

[25] ZHANG K., ZHANG Z., LI Z., QIAO Y. Joint Face Detection and Alignment Using Multitask Cascaded Convolutional Networks. *IEEE Signal Processing Letters.* 2016, 23(10), pp. 1499–1503, doi: 10.1109/LSP.2016.2603342.