Proceedings of the 7th International Conference on Computing Methodologies and Communication (ICCMC-2023)
IEEE Xplore Part Number: CFP23K25-ART; ISBN: 978-1-6654-6408-6

Deepfake Detection using Inception-ResNet-V2 Network

1st Dr. R. R. Rajalaxmi
Computer Science and Engineering
Kongu Engineering College
Perundurai, India
rrr.cse@kongu.edu

2nd Sudharsana P P
Computer Science and Engineering
Kongu Engineering College
Perundurai, India
sudharsanapp.19mcse@kongu.edu

3rd Rithani A M
Computer Science and Engineering
Kongu Engineering College
Perundurai, India
rithaniam@gmail.com

4th Preethika S
Computer Science and Engineering
Kongu Engineering College
Perundurai, India
spreethika72@gmail.com

5th Dhivakar P
Computer Science and Engineering
Kongu Engineering College
Perundurai, India
dhivakar782001@gmail.com

6th Gothai E
Computer Science and Engineering
Kongu engineering College
Perundurai, India
egothai.cse@kongu.edu

2023 7th International Conference on Computing Methodologies and Communication (ICCMC) | 978-1-6654-6408-6/23/$31.00 ©2023 IEEE | DOI: 10.1109/ICCMC56507.2023.10083584

***Abstract*—Deepfakes entails obscene videos in which a face can be changed with someone else's utilizing neural networks. Deepfakes are a public problem, thus developing methods to detect them is critical. With deepfake and human-level control, deep learning has already utilized to address complicated issues. Deepfake is a new branch of AI technology, which in that one's face is superimposed on another's face, which is popular on social media. Deep learning techniques were also employed in the development of software that affects national security, democracy and privacy. One of the most current applications that use deep learning is deepfakes. Algorithms for deepfake may generate fake photos and movies which are indistinguishable from real ones. As a result, the development of technology capable of automatically detecting and assessing the reliability of digital video is critical. This research describes algorithms for detecting deepfakes. By reviewing the influence of deepfakes and deepfake recognition systems, this work enables the creation of new and so many effective methodologies to cope with increasingly complex deepfakes. InceptionResNetV2 architecture in Convolutional Neural Networks (CNN) is utilized in this comparative study to distinguish real and deepfake images.**

Keywords—Deep Learning, Deepfake, CNN, Inception ResNet-V2

I. INTRODUCTION

1.1 ARTIFICIAL INTELLIGENCE

Artificial Intelligence (AI), a subdivision of computer science aims to establish intelligent machines (Russel,2009). These intelligent devices function and react in the same way as people do. Artificial intelligence research is almost typically very technical and specialized. The main objective of Artificial Intelligence is to create technology that enables intelligent machine and computer operation. Some of the features that have attracted the greatest attention in recent days are logical thinking, knowledge representation, general intelligence, social intelligence, natural language processing, planning, manipulation and motion and problem solving. AI techniques are utilized in robotics, scheduling, data mining, logistics, video games, healthcare, automotive, government, finance, and economics, among other fields.

1.2 DEEP LEARNING

Deep learning produces artificial neural networks that can learn and make intelligent judgments with the use of algorithms. Deep learning employs neural networks with several node layers between the input and output layers. Deep refers to the number of layers in the network; the more levels there are, the deeper the network. A series of layers between input and output identify features. The necessity for deep learning arose primarily as a result of the requirement to analyse vast amounts of data, perform difficult algorithms, achieve best performance with large amounts of data, and extract effective features.

1.2.1 Deep Learning Application

Healthcare, finance, deepfake detection, earthquake prediction, voice search and voice-activated help, automatic machine translation, photo recognition, and self-driving cars are just a few areas where deep learning is used.

1.2.2 Deep Learning in Deepfakes

Deep learning has had a lot of success in detecting deepfakes. Deepfakes manipulate photos and videos of people so that it is impossible for people to tell the difference between them and the actual thing using deep learning technology. Numerous studies have been done in recent years to learn how deepfakes operate, and many methods based on deep learning have been developed to identify deepfake movies or images.

1.3 CONVOLUTIONAL NEURAL NETWORK

CNNs, a unique deep learning architecture, are designed for certain tasks like picture categorization. CNNs were designed using the network of neurons seen in the frontal cortex of animal brains. A CNN has an input layer as one of its parts. The input for basic image processing, however, typically comprises of a two-dimensional collection of nerves that represent the pixel values of the image. It has an output layer as well, which is normally made up of a group of one-dimensional output units. To process the incoming images, CNN combines pooling layer with sparse connections. To further decrease the number of neurons needed in the network's subsequent layers, they additionally incorporate down sampling levels known as pooling layers.

1.3.1 InceptionResNetV2 Architecture

A Convolutional Neural Network named Inception-ResNet-v2 was trained using more than a million photos from the ImageNet collection. The 164-layer network can

categorise photos into 1000 different item categories. The network has therefore acquired rich feature representations for a variety of pictures. The network receives a 299 by 299 pixel picture as input, and it outputs a list of predicted class probabilities.

It is constructed using both the Residual connection and the Inception structure. Multiple convolutional filters of various sizes are mixed with residual connections in the Inception-Resnet block. In addition to avoiding the degradation issue brought on by deep structures, the inclusion of residual connections shortens training time. The basic Inception-Resnet-v2 network architecture is depicted in the Fig. 1

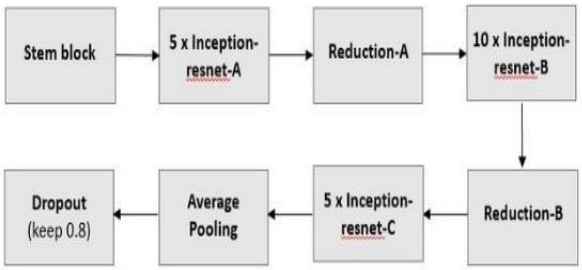

Fig. 1 Basic InceptionResNetV2 Architecture

II. LITERATURE REVIEW

Ricard Durall, et al., proposed a method [1] based on a standard frequency domain analysis, then a simple classifier. This technique exhibited extremely strong outcomes only using just few labeled training examples and even obtained acceptable accuracies in wholly unsupervised circumstances, in contrast to previous systems that require to be fed with enormous volumes of labelled data. A new benchmark called Faces-HQ was created by combining numerous public data sets of genuine and synthetic faces for the evaluation of high-resolution face photos. When trained on as few as 20 labelled examples using such high-resolution photos, this method achieves a flawless classification accuracy of 100%.

Johansson, Emil introduced [2] a feature for identifying false or changed video content, particularly those where faces have previously been altered. The four different types of altering movies that were considered were FaceSwap, DeepFakes, Face2Face, and Neural Textures. An LSTM network and a CNN are used in the proposed model to extract the features. The dataset and accompanying benchmark from FaceForensics++ were used. The model was able to recognize videos more effectively than or on par with other models in the benchmark, but it was unable to match the most sophisticated detectors.

Tackhyun, Jung, Sangwon Kim, et al., [3] introduced a novel method to examine a substantial change in the blinking pattern of eyes, which is an activity performed voluntarily and without conscious effort, to detect Deepfakes produced by the GANs model. The normal eye movement rate is known to vary substantially depending on a person's overall physical and cognitive activity. It carries out integrity checking by keeping track of notable variations in a subject's eye blinking pattern during a video. However, since cyber-security attack and defences grow regularly, there are a lot of ways to improve this.

Pavel Korshunov, Sebastien Marcel produced Deepfake films, [4] in which a person's face is robotically replaced with that of another. Many huge datasets of deepfake movies and numerous methods to detect them have been presented in response to the threat that these modifications potentially pose to the confidence in video evidence. They offer a subjective study carried out in a setting like crowdsourcing, which systematically assesses how challenging it is for people to determine if a video is authentic or not.

Aarti Karandikar, Vedita Deshpande et al., proposed a methodology for Deepfake which is an artificial intelligence-based method for synthesizing human images. [5] Utilizing Deepfake, existing pictures are merged and superimposed onto the source images. Those are convincingly fabricated videos that are impossible to tell apart with the naked eye. Using a combination of dense and convolutional neural network layers, deepfakes were binary classified.

Brian Dolhansky, Joanna Bitton and et al., [6] described that with the use of the current off-the-shelf modification method known as deepfakes, anyone may switch two characteristics in a single video. A number of GAN-based face swapping techniques have also been released with corresponding code, in addition to Deepfakes. In order to combat this growing threat, a very large face swap video dataset was created to facilitate the training of detection models, and a related DeepFake Detection Challenge (DFDC) Kaggle competition was held. The DFDC dataset, which has over 100,000 total clips obtained from 3,426 paid actors and was created using a variety of Deepfake, GAN-based, and non-learned approaches, is by far the largest face swap video dataset that is currently and openly accessible.

Worku Muluye Wubet aimed to look at deep-fake difficulties and find deep-fake films utilizing eye-blink detection. [7] Deepfake detections are techniques for determining if social media videos and photos are genuine or fraudulent. To train the detection models for deepfake detection, authentic and faked image or video datasets are required. This paper initially addressed deepfake technique and its difficulties before identifying the video datasets that were readily available. To measure the height and breadth of open and closed eyes as well as to identify blinking intervals, the eye aspect ratio was utilized. Using eye blinking to distinguish between real and fraudulent videos, the model trained on the UADFV dataset finds 18.4 blinks per minute in authentic videos and 4.28 blinks per minute in phony films. Overall, 93.23% and 98.30%, respectively, of actual and false videos could be detected.

Ruben Tolosana, Ruben Vera-Rodriguez and et al., conducted a survey [8] which describes in detail how to manipulate facial photos, including DeepFake techniques, as well as how to spot similar modifications. Four specific forms of facial manipulation are examined: i) identity swap (DeepFakes), ii) entire face synthesis, iii) expression swap, and iv) attribute manipulation Each manipulation group provides information on the manipulation methods used, the public databases that are currently in use, and important benchmarks for the technical evaluation of false detection

systems, along with a summary of the findings from those studies. Of all the topics included in the survey, the most recent DeepFakes generation receives specific emphasis, emphasizing both its advancements and difficulties in false identification.

Young-Ju Choi, Young-Jin, Heo, et al., created a CNN feature-based positioning model [9] using patches that acquires to communicate with so many locations in order to discover the artefact region and solve the false negative problem. Without the ensemble approach, the model achieves 0.978 AUC and 91.9 f1 score, whereas the recent SOTA model achieves 0.972 AUC and 90.6 f1 score under the identical conditions.

Nicolo Bonettini, Edoardo Daniele Cannas and et al., proposed an approach for addressing the issue of face alteration recognition in video frames focusing on contemporary facial manipulation techniques. [10] Using two distinct principles, several models are produced from a basic network (i.e., EfficientNetB4): (i) Siamese training; (ii) Attention layers. Demonstrating that merging these networks produces positive results for face modification detection on two datasets that are openly accessible and contain more than 119000 videos.

Sreeraj Ramachandran, Aakash Varma Nadimpalli and et al., evaluated by employing various loss functions and deepfake generating strategies, researchers assessed the effectiveness of deep facial recognition in detecting deepfakes. Deep face recognition is more effective at recognizing deepfakes than two-class CNNs and the ocular modality, according to experimental studies on the difficult FaceForensics++ and Celeb-DF datasets. [11] Using face recognition on the Celeb-DF dataset, reported findings indicate an Equal Error Rate (EER) of 7.1% and a peak Area Under Curve (AUC) of 0.98 for recognizing deepfakes. When compared to the EER computed for the 2 different CNN and the visual modality on the Celeb-DF dataset, this EER is reduced by 16.6%. On the FaceForensics++ dataset, further analysis yielded an AUC of 0.99 and an EER of 2.04%.

The problem of deepfake identification was addressed by Hina Fatima Shahzad, Furqan Rustam, et al., [12]. The detection of face alteration has been the subject of extensive research. This paper also covers mitigation strategies for deepfake technology's negative effects. For deepfake prevention, content authentication, deepfake identification and technical development is desired. This research tries to emphasize current advances in the detection of deepfakes in images and videos, including deepfake creation, several recognition algorithms on benchmark datasets and custom datasets already in existence.

The Face Deepfake Detection and Reconstruction Challenge [13] was put up by Oliver Giudic, Luca Guarnera, et al.. The participants offered two activities: (I) establishing a Deepfake detector that might work in reality; and (II) developing a technique for recovering authentic Deepfake photos. Natural photographs from CelebA and FFHQ as well as Deepfake images obtained by StyleGAN, StyleGAN2, StarGAN, StarGAN-v2, AttGAN, and GDWCT were used.

Maryam Taeb and Hongmei Chi proposed Deep-learning algorithms [14] that run over massive datasets until they have learned how to solve the given problem might produce phoney media that looks realistic. The growing deterioration of public discourse is being caused by the vast production of such content and modification technology. The capacity to improve performance while utilizing fewer computer resources by augmenting data to better understand how the CNN algorithms used in this study should be understood.

S.Sathvikaa, M.S.Supriya, et al., have seen a significant increase in media manipulation due to the advancement of technology and simplicity of producing fake information. [15] Deepfake media, or AI-altered films, have become a severe threat to media credibility. Deepfake media is frequently made and spread on social media platforms, and it is regarded to be extremely difficult to identify. InceptioResNetV2 in a CNN method with LSTM, is used to detect Deepfake movies. Deep-Learning (DL) model developed obtained 91% accuracy on the Celeb-Df dataset

Vijay Chahar, Jatin Sharma, et al., recommended CNN based algorithm [16] to recognize fake facial photos. Applying GANs and data augmentation, the face dataset is created to discriminate between authentic and fraudulent faces. The proposed model makes advantage of transfer learning methods using already trained convolutional networks like VGG16 and ResNet50. Three benchmark datasets are used to assess the performance of the proposed method: Real and Fake Face Detection, Fake Faces and 140k Real and Fake Faces. Furthermore, the results show that the suggested model achieves other models.

Wanying Ge, Jose Patino and et al., demonstrated the tool's ability to expose unexpected classifier performance, the artefacts that have the most impact on classifier outputs, and variations in how various spoofing detection models behave. The tool is effective and applicable, easily adaptable to a variety of various architecture models as well as connected, various applications. Using open-source software, it is possible to replicate every finding mentioned in the study. [17]

Shichao Dong, Jin Wang and et al., [18] aimed to understand how deepfake detection models, which are only guided by binary labels, learn the artefact properties of photos. In order to do this, the following three assumptions are put out from the viewpoint of image matching. 1. When recognizing such visual ideas as artifact-relevant, deepfake detection algorithms distinguish between real and fake pictures on the basis of ideas that are not source-relevant nor target-relevant. 2. Deepfake detection models implicitly acquire artifact-relevant visualizations through the FST-Matching (i.e., the matching fake, source, and target pictures) in the training set, in addition to the supervision of binary labels. 3. The FST-Matching in the original training data set exposes implicitly learnt artefact visual notions that are susceptible to video compression.

Even though several work has been proposed by the researchers, deepfake detection using prebuilt CNNs has not been addressed. So, in this work we utilized Inception-Resnet-V2 for deepfake detection.

III. PROPOSED WORK

3.1 FLOW DIAGRAM

In the Fig. 2, the flow diagram depicts the proposed work in which the videos in the dataset are pre- processed. Then frames are extracted from videos and images are recognized using OpenCV. Videos are labelled as real and fake and stored in different folders while extracting frames. Next, extracted frames are given as input for InceptionResNetV2 model which is build and get trained. After training, a test video is given as input and the result is predicted.

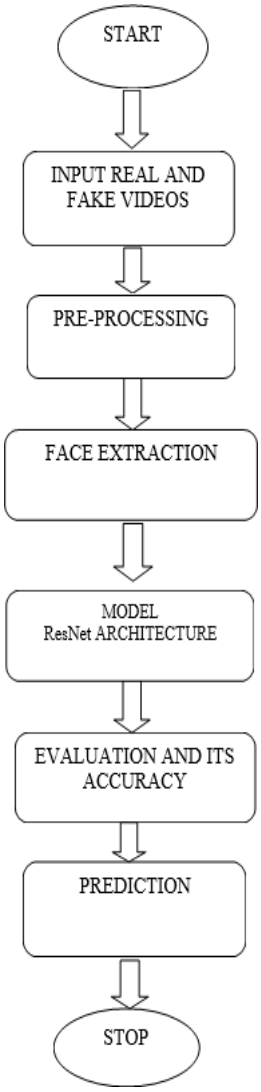

Fig. 2 Proposed Work

3.2 DATASET DESCRIPTION

Deepfake Detection Challenge Dataset (DFDC) is used here. It consists of both real and fake video files. It contains 400 training videos with a metadata.json file which contains 323 fake videos and 77 real videos, 400 testing videos and a CSV file. The sample dataset is shown in Fig. 3.

Fig. 3 Sample Dataset

3.3 PRE-PROCESSING

Data preparation is a standard first step in the machine learning pipeline to convert raw data into the form that the networks can understand. The length of an image input layer, for instance, might be affected by resizing the image input. You can also pre-process data to reduce artefacts that could skew the network or enhance desired features. For instance, you may normalise or clean up the incoming data's noise.

Pre-processing picture input with operations like scaling is possible with the use of datastores and functions.

3.3.1 Videos and video processing

The visual component of something transmitted is called video. No video in this collection had more than one modification. The dataset contains videos of length 10 s.

The videos are separated in two folders as real and fake with the help of metadata.json file. 10 frames per video is created, labelled as real and fake along with video file name and frame number (fake_aagfhgtpmv_1.png) and stored in the real and fake folders respectively.

3.3.2 Frames and frames extraction

The frame is a single image in a series of photographs. The word comes from the historical evolution of film stock, in which when single sequentially recorded images are examined, they resemble a framed picture. In general, one second of video is made up of 24 or 30 frames per second, abbreviated as FPS. The frame is made up of the image and the time it was exposed to the view.

The goal of extracting video frame aims to display as much video material as feasible with the fewest number of video frames possible, to remove unnecessary frames from a video, and to decrease processing time.

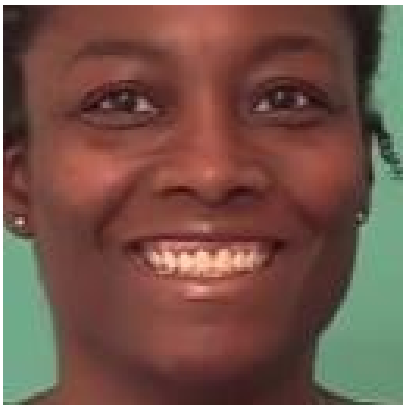

Fig. 4 Extracted Frame

OpenCV is used to extract frames from a video. The number of frames extracted from each video is 10. From those frames, the image i.e., a human face is captured as shown in Fig. 4 and the file is saved in a .png format.

3.3.3 Resize the image

Resolution is defined as an image's height and width in pixels. The resolution of each frame in a video is the same, for example, (500x480) or (620x540), etc. An image's dimensions are changed when it is cropped, changing them to other resolutions like (47x50) or (35x48), etc. A predetermined number of resolutions are sent to the CNN model for each frame. The image size will be altered to so that it can be used as an input with the Classification algorithm (128 x128).

3.3.4 Label and split the data

The deep learning algorithm will essentially operate in two stages, the first of which is model training and the second of which is model testing. In the training stage, typical picture feature qualities are isolated, and each classification category is given its own distinct description. These feature-space partitions will be utilized to categorize image features in the future testing stage. In supervised learning, CNN models will be utilized, which depend on the label data to quickly and effectively learn the data. The data will be divided into two categories: fake photos and real images. In the training phase, the data will be divided into real video frames and fake video frames. However, the data cannot be labelled throughout the testing stage in order to test the model.

3.4 IMPLEMENTATION

From the videos, frames are captured and stored as images. Using OpenCV images are identified and processed. The obtained frames are given into the model to be pre-processed. After pre-processing, the InceptionResNetV2 model is trained with real and fake images. InceprionResNetV2 replaces the loss layer with an output layer that detects deepfake loss, termed as deepfake detection loss output layer. The network's precise tuning limits variances from either the data detected in the dataset or the predicted performance. This model has been compiled for 20 epochs, 0.00001 learning rate and 100 batch size to master the training dataset. The model employs the Sigmoid activation function, which is useful for neural networks. This function converts required graph data into a value between 0 and 1. The confusion matrix is used to guide further evaluation. After the model is trained, a test data is given to the model for classification.

The parameters used in Inception-ResNet-V2 are shown in TABLE I.

TABLE I. PARAMETERS FOR INCEPTION-RESNET-V2

S.No.	*Parameters*	*Values*
1	Number of layers	467
2	Learning Rate	0.00001
3	Number of Epochs	20
4	Batch Size	100
5	Optimizer	Adam

The Inception-ResNet-V2 model is implemented using python in Google colab cloud platform. We have imported the necessary libraries like Sklearn, NumPy, Pandas, Tensorflow, Keras and Matplotlib. We ran the model on GPU with 128GB RAM, GPU RAM 32GB, 4TB hard disk capacity.

IV. RESULTS AND DISCUSSION

4.1 PARAMETERS FOR EVALUATION

The following metrics are used for the evaluation.

- Confusion Matrix
- Accuracy

4.1.1 Confusion Matrix

A table called a confusion matrix is used to describe how well a classification system performs. The output of a classification algorithm is summarised in a confusion matrix which is shown in Fig. 5. The confusion matrix is made up of four main properties (numbers) that are utilised to define the classifier's measuring metrics. These four numbers are:

- True positive (TP): The data is expected to be positive and is in fact positive.
- False positive (FP): The data is expected to be positive but turned out to be negative.
- True negative (TN): The data is expected to be negative and is in fact negative.
- False negative (FN): The data is expected to be negative but turned out to be positive.

TABLE II. CONFUSION MATRIX RESULT

S.No.	*Measures*	*Values*
1	TP	2890
2	FP	96
3	FN	91
4	TN	668

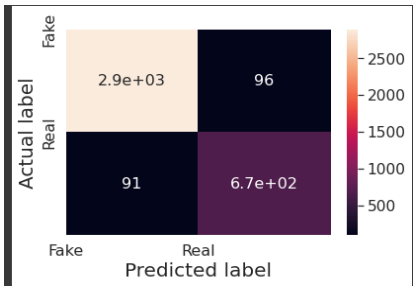

Fig. 5 Confusion Matrix

4.1.2 Accuracy

Accuracy is the parameter used for evaluating the dataset. It is one of the metrics for evaluating classification models. Formally, accuracy is expressed as the ratio of accurate predictions divided by the total number of predictions as shown in equation (4.1). Informally, the accuracy of the model is the percentage of correct predictions.

$$ACCURACY = (TP+TN) / (TP+FN+FP+TN) \quad (1)$$

The designed model is tested for 20 epochs with batch size 100 because of limitation in runtime and achieved an training accuracy of 98.37% and validation accuracy of 87.06%. The resultant graphs which is shown in Fig. 6 were obtained after implementation.

TABLE III. TABLE FOR ACCURACY

S.No.	*Types*	*Accuracy (%)*
1	Training accuracy	98.37
2	Validation accuracy	87.06

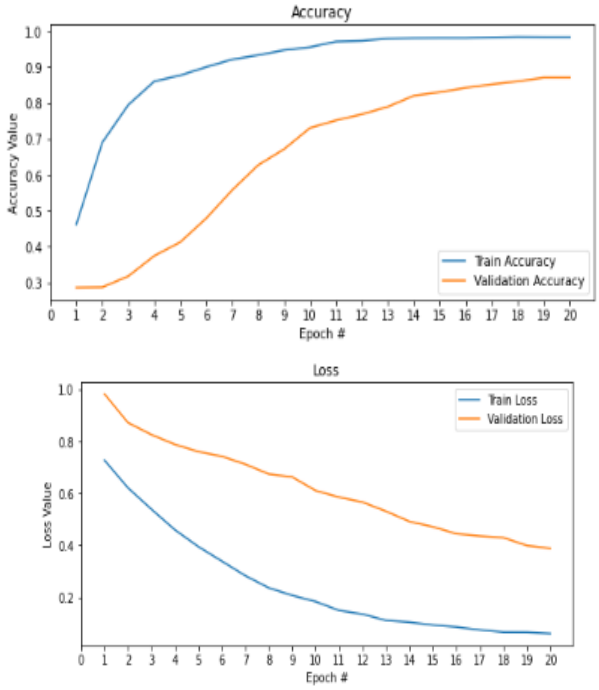

Fig. 6 Accuracy and Loss Graphs

4.2 COMPARATIVE ANALYSIS

In the existing model [19], DFDC dataset with 249 videos, out of which 199 fake videos and 53 real videos and YouTube dataset from Dessa containing 66 real videos, in total 318 videos, out of which 199 fake and 119 real videos were used. Around 3114 frames including 2189 fake and 925 real video frames were extracted and given as input to CNN and MLP and the results of those were given to concatenation layer. This model gave accuracy of 84%.

In our model, DFDC dataset with 323 fake videos and 77 real videos, in total 400 videos are used. Around 3745 frames that includes 2986 fake and 759 real video frames are extracted. Extracted images are given as input to InceptionResNetV2 architecture. The model is trained with a greater number of frames when compared to existing model which achieved an training accuracy of 98.37% and validation accuracy of 87.06%.

CONCLUSION AND FUTURE WORK

Due to Deepfakes, the general public's faith has begun to erode as the streaming video no longer seems genuine and real. In this project, a deep learning-based method for automatically identifying deep fakes is done. In deepfakes, the target face only appears for a brief period of time in a video. As a result, the model splits the user video into frames, and then further pre-processes these frames using InceptionResNetV2. The approach offered a high degree of accuracy and dependability. The suggested method can analyse any video using convolutional InceptionResNetV2 and aids in spotting deepfake faces that have been altered, preventing people from defaming others. To get more precision, experiments can also been run with a greater number of epochs and learning rates.

In future, one can expand this work by exploring more architectures like LeNet Architecture, VGGNet Architecture, AlexNet Architecture, etc. and also this project can be explored with different datasets like UADFV, FaceForensics++, and Celeb-DF. This work can also be extended further by taking face landmarks like eyes, nose, lips, mouth from the extracted face image and train the model accordingly.

ACKNOWLEDGMENT

The authors acknowledge and thank the Department of Science and Technology (Government of India) for sanctioning the research grant (Ref. No.SR/FST/COLLEGE-096/2017 dated 16.01.2018) under Fund for Improvement of S&T Infrastructure (FIST) program for completing this work.

REFERENCES

[1] R. Durall, M. Keuper, F.-J. Pfreundt, and J. Keuper, "Unmasking DeepFakes with simple Features," Nov. 2019, [Online]. Available: http://arxiv.org/abs/1911.00686

[2] E. Johansson, "Detecting Deepfakes and Forged Videos Using Deep Learning."

[3] T. Jung, S. Kim, and K. Kim, "DeepVision: Deepfakes Detection Using Human Eye Blinking Pattern," *IEEE Access*, vol. 8, pp. 83144–83154, 2020, doi: 10.1109/ACCESS.2020.2988660.

[4] P. Korshunov and S. Marcel, "Deepfake detection: humans vs. machines," Sep. 2020, [Online]. Available: http://arxiv.org/abs/2009.03155

[5] A. Karandikar, "Deepfake Video Detection Using Convolutional Neural Network," *International Journal of Advanced Trends in Computer Science and Engineering*, vol. 9, no. 2, pp. 1311–1315, Apr. 2020, doi: 10.30534/ijatcse/2020/62922020.

[6] B. Dolhansky *et al.*, "The DeepFake Detection Challenge (DFDC) Dataset," Jun. 2020, [Online]. Available: http://arxiv.org/abs/2006.07397

[7] W. M. Wubet*, "The Deepfake Challenges and Deepfake Video Detection," *International Journal of Innovative Technology and Exploring Engineering*, vol. 9, no. 6, pp. 789–796, Apr. 2020, doi: 10.35940/ijitee.E2779.049620.

[8] R. Tolosana, R. Vera-Rodriguez, J. Fierrez, A. Morales, and J. Ortega-Garcia, "DeepFakes and Beyond: A Survey of Face Manipulation and Fake Detection," Jan. 2020, [Online]. Available: http://arxiv.org/abs/2001.00179

[9] Y.-J. Heo, Y.-J. Choi, Y.-W. Lee, and B.-G. Kim, "Deepfake Detection Scheme Based on Vision Transformer and Distillation," Apr. 2021, [Online]. Available: http://arxiv.org/abs/2104.01353

[10] N. Bonettini, E. D. Cannas, S. Mandelli, L. Bondi, P. Bestagini, and S. Tubaro, "Video Face Manipulation Detection Through Ensemble of CNNs," Apr. 2020, [Online]. Available: http://arxiv.org/abs/2004.07676

[11] S. Ramachandran, A. V. Nadimpalli, and A. Rattani, "An Experimental Evaluation on Deepfake Detection using Deep Face Recognition," Oct. 2021, [Online]. Available: http://arxiv.org/abs/2110.01640

[12] H. F. Shahzad, F. Rustam, E. S. Flores, J. Luís Vidal Mazón, I. de la Torre Diez, and I. Ashraf, "A Review of Image Processing

Techniques for Deepfakes," *Sensors*, vol. 22, no. 12. MDPI, Jun. 01, 2022. doi: 10.3390/s22124556.

[13] L. Guarnera, O. Giudice, and S. Battiato, "DeepFake Detection by Analyzing Convolutional Traces." [Online]. Available: https://deepfakedetectionchallenge.ai/

[14] M. Taeb and H. Chi, "Comparison of Deepfake Detection Techniques through Deep Learning," *Journal of Cybersecurity and Privacy*, vol. 2, no. 1, pp. 89–106, Mar. 2022, doi: 10.3390/jcp2010007.

[15] V. V. V. N. S. Vamsi *et al.*, "Deepfake detection in digital media forensics," *Global Transitions Proceedings*, vol. 3, no. 1, pp. 74–79, Jun. 2022, doi: 10.1016/j.gltp.2022.04.017.

[16] J. Sharma, S. Sharma, V. Kumar, H. S. Hussein, and H. Alshazly, "Deepfakes Classification of Faces Using Convolutional Neural Networks," *Traitement du Signal*, vol. 39, no. 3, pp. 1027–1037, Jun. 2022, doi: 10.18280/ts.390330.

[17] W. Ge, J. Patino, M. Todisco, and N. Evans, "Explaining deep learning models for spoofing and deepfake detection with SHapley Additive exPlanations," Oct. 2021, [Online]. Available: http://arxiv.org/abs/2110.03309

[18] S. Dong, J. Wang, J. Liang, H. Fan, and R. Ji, "Explaining Deepfake Detection by Analysing Image Matching," Jul. 2022, [Online]. Available: http://arxiv.org/abs/2207.09679

[19] S. Kolagati, T. Priyadharshini, and V. Mary Anita Rajam, "Exposing deepfakes using a deep multilayer perceptron – convolutional neural network model," *International Journal of Information Management Data Insights*, vol. 2, no. 1, Apr. 2022, doi: 10.1016/j.jjimei.2021.100054.