IJIRSET
International Journal of Innovative Research in
SCIENCE | ENGINEERING | TECHNOLOGY

e-ISSN: 2319-8753 | p-ISSN: 2347-6710


# INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

## IN SCIENCE | ENGINEERING | TECHNOLOGY

Volume 13, Issue 4, April 2024

ISSN INTERNATIONAL STANDARD SERIAL NUMBER INDIA

Impact Factor: 8.423

9940 572 462 | 6381 907 438 | ijirset@gmail.com | www.ijirset.com

International Journal of Innovative Research in Science, Engineering and Technology IJIRSET)

|e-ISSN: 2319-8753, p-ISSN: 2347-6710| www.ijirset.com | Impact Factor: 8.423| A Monthly Peer Reviewed & Referred Journal |

|| Volume 13, Issue 4, April 2024 ||

| DOI:10.15680/IJIRSET.2024.1304170 |

# Deep Fake Detection Using Deep Learning

**Mohd Salim Shaikh[1], Lucky Nirankari[2], Vasant Pardeshi[3], Rupesh Sharma[4], Prof. Sunil Kale[5]**

Student, Department of Computer Engineering, Sandip Institute of Technology and Research Centre, Nashik, Maharashtra, India[1,2,3,4]

Professor, Department of Computer Engineering, Sandip Institute of Technology and Research Centre, Nashik, Maharashtra, India[5]

**ABSTRACT**: As technology progresses and society evolves, deep learning becomes increasingly accessible. Unfortunately, there are individuals exploiting deep learning technology to create fraudulent images and videos, posing serious threats to societal stability. Instances include fabricating statements from politicians, disseminating misinformation through face-swapping techniques, and generating counterfeit videos for financial gain. Recognizing this societal challenge, this paper proposes a novel approach utilizing the EfficientNet network to discern the authenticity of visual content, building upon existing fake face detection systems. Our method is applied to two prominent large-scale fake face datasets, showcasing the superior performance of EfficientNet compared to current detection networks through comprehensive training and testing evaluations. Ultimately, by enhancing the accuracy of the detection system in distinguishing genuine from counterfeit faces, we achieve notable improvements in visual fidelity during the identification of real pictures and videos.

**KEYWORDS:** Deep learning, Fraudulent images and videos, EfficientNet

## I. INTRODUCTION

With the rapid advancement of technology, particularly in the field of artificial intelligence, the generation of deepfake media has become increasingly prevalent. Deepfake media, which typically refers to manipulated videos or images created using deep learning algorithms, have raised concerns due to their potential to deceive and manipulate viewers. These media are often generated by training deep neural networks on vast amounts of data, including images and videos of the target individual, to learn and mimic their facial expressions, movements, and speech patterns.

Detecting and classifying deepfake media poses a significant challenge due to their increasingly sophisticated nature. Researchers have explored various characteristics of deepfake media that can be utilized for classification purposes. These characteristics may include inconsistencies in facial expressions, unnatural eye movements, or discrepancies in audio-visual synchronization. Additionally, approaches such as analyzing artifacts introduced during the generation process or detecting anomalies in facial landmarks have been employed to identify deepfake content. Despite these efforts, the arms race between creators of deepfake media and detection methods continues, underscoring the need for robust and efficient classification techniques.

In this research, we propose the utilization of the EfficientNet architecture for the classification of deepfake media. EfficientNet is a convolutional neural network architecture that aims to achieve superior performance while maintaining computational efficiency. It employs a compound scaling method that optimizes model depth, width, and resolution to achieve better accuracy and efficiency compared to traditional architectures. By leveraging the capabilities of EfficientNet, we aim to develop a classification system capable of accurately distinguishing between genuine and deepfake media, thereby contributing to the ongoing efforts to combat the spread of misinformation and manipulation in digital media.

## II. RELATED WORK

"Faceforensics: A Large-Scale Video Dataset for Forgery Detection in Human Faces," by Andreas Rossler, Davide Cozzolino, Luisa Verdoliva, Christian Riess, Justus Thies, and Matthias Nießner, from the Technical University of Munich, University Federico II of Naples, and University of Erlangen Nuremberg in March 2018, presents a significant contribution to the field of forgery detection in video content. [1]

|e-ISSN: 2319-8753, p-ISSN: 2347-6710| www.ijirset.com | Impact Factor: 8.423| A Monthly Peer Reviewed & Referred Journal |

|| Volume 13, Issue 4, April 2024 ||

| DOI:10.15680/IJIRSET.2024.1304170 |

"Deepfake Detection Challenge (DFDC) Preview Dataset" paper, authored by Brian Dolhansky, Russ Howes, Ben Pflaum, Nicole Baram, and Cristian Canton Ferrer from the AI Red Team at Facebook AI, was published in October 2019. The paper introduces a dataset created for the Deepfake Detection Challenge, aiming to spur advancements in deepfake detection technologies. The dataset includes manipulated videos with facial deepfakes, providing a preview for researchers and participants in the challenge. [2]

"Celeb-DF: A Large-Scale Challenging Dataset for Deepfake Forensics" was authored by Yuezun Li, Xin Yang, Pu Sun, Honggang Qi, and Siwei Lyu in March 2020. The research was conducted at the University at Albany, State University of New York, USA, and the University of Chinese Academy of Sciences, China. [3]

"An Improved Dense CNN Architecture for Deepfake Image Detection," authored by Yogesh Patel, Sudeep Tanwar (Senior Member, IEEE), Innocent Ewean Davidson (Senior Member, IEEE), and Thokozile F. Mazibuko, and published in March 2023, introduces an advanced convolutional neural network (CNN) architecture tailored for detecting deepfake images. The authors leverage their expertise to propose enhancements, addressing contemporary challenges in deepfake detection. [4]

"Exposing Deepfake Videos by Detecting Face Warping Artifacts," authored by Yuezun Li and Siwei Lyu from the University at Albany, State University of New York (May 2019), focuses on detecting face warping artifacts as a means of unveiling deepfake videos. The authors contribute to the field of deepfake detection by proposing methods to identify distortions in facial features. This research addresses the challenges of early deepfake detection, offering insights into the limitations of existing methods. [5]

"Deep Fake Video Detection Using Recurrent Neural Networks" by Abdul Jamsheed V. and Janet B., published in April 2021, likely focuses on employing Recurrent Neural Networks (RNNs) for the purpose of detecting deep fake videos. The primary objective appears to be the development and application of a model that utilizes the temporal information in videos, a characteristic addressed by RNNs, to identify manipulated content.[6]

“Deep Detection for Face Manipulation” by Disheng Feng, Xuequan Lu, Deakin University, Australia published in September 2020, the model first learns features with the aid of a contrastive loss (i.e., triplet loss) which can metrically push the forgery faces away from the real faces. Then design a simple linear classification network to obtain binary results, shows that the method is effective, and outperforms state-of-the-art techniques in most cases. [7]

“Fighting Deepfake by Exposing the Convolutional Traces on Images” by Luca Guarnera, Oliver Giudice And Sebastiano Battiato published in 2020 stated that an algorithm based on Expectation-Maximization was employed to extract the Convolutional Traces (CT): a sort of unique fingerprint useable to identify not only if an image is a Deepfake but also the GAN architecture that generated it. Obtained results demonstrate also to overcome the state-of-the-art with a technique simple and fast to be computed. [8]

“Multi-attention-based approach for deepfake face and expression swap detection and localization” produced by Saima Waseem, Syed Abdul Rahman Syed Abu Bakar, Zaid Omar, Bilal Ashfaq Ahmed, Saba Baloch and Adel Hafeezallah stated that to enhance the performance and provide localization of face forgery, introduce a strategy involving the combination of the encoder and preceding layer decoder with dual attention block scAB. This approach localizes the manipulated facial regions at the pixel level. Through extensive experiments on deepfake benchmarks, Localization results, demonstrated that the model tends to focus on the forgery regions instead of unwanted biases and artifacts, leading to more accurate predictions. [9]

“Xception: Deep Learning with Depthwise Separable Convolutions” by François Chollet, Google, Inc. in 2017 presented a novel architecture based on this idea, named Xception, which has a similar parameter count as Inception V3. Compared to Inception V3, Xception shows small gains in classification performance. [10]

“Combining EfficientNet and Vision Transformers for Video Deepfake Detection” by Davide Coccomini, Nicola Messina, Claudio Gennaro, and Fabrizio Falchi in 2023 demonstrated the effectiveness of mixed convolutional transformer networks in the Deepfake detection task using pretrained convolutional networks, such as the widely used EfficientNet B0, to extract visual features, and relied on Vision Transformers to obtain an informative global description for the downstream task. [11]

|e-ISSN: 2319-8753, p-ISSN: 2347-6710| www.ijirset.com | Impact Factor: 8.423| A Monthly Peer Reviewed & Referred Journal |

|| Volume 13, Issue 4, April 2024 ||

| DOI:10.15680/IJIRSET.2024.1304170 |

## III. METHODOLOGY

Our methodology for detecting manipulated faces in videos revolves around employing a CNN based approach, leveraging the strengths inherent in various CNN-based classifiers. This strategic choice aligns with the well-established principle that amalgamating multiple models often leads to heightened predictive performance. Our primary objective is to train a diverse CNN-based classifiers capable of capturing unique high-level semantic information, thereby amplifying the ensemble's efficacy for this specific task.

We leverage the groundwork laid by the EfficientNet family of models, which has been lauded as a pioneering method for automatically scaling CNNs in previous research [12]. These architectures stand out for their exceptional accuracy and efficiency compared to other cutting-edge CNNs, rendering them exceptionally well-suited for tackling the hardware and time constraints inherent in identifying manipulated faces.

Our endeavor to enhance EfficientNet models for ensemble learning encompasses two primary strategies. Firstly, we integrate an attention mechanism, a feature that not only enhances classification accuracy but also furnishes analysts with valuable insights by spotlighting the most informative regions within video frames. Secondly, we delve into the integration of siamese training strategies into the learning process, an additional technique designed to extract supplementary information from the data, thereby enriching the models' comprehension of facial manipulation cues.

### A. EfficientNet and attention mechanism

The EfficientNetB4 model is chosen as the baseline due to its favourable trade-off between parameters, runtime, and classification performance. Compared to XceptionNet, another model considered for the task, EfficientNetB4 achieves higher accuracy with fewer parameters and lower computational cost.[12],[13]

EfficientNet is a convolutional neural network that is based on the notion of "compound scaling." This idea tackles the age-old trade-off between model size, accuracy, and computing efficiency. Compound scaling is intended to scale three critical aspects of a neural network: breadth, depth, and resolution.

1)Width: The number of channels in each layer of the neural network is referred to as width scaling. By enlarging the width, the model may capture more complicated patterns and characteristics, resulting in higher accuracy. Reducing the width, on the other hand, results in a more lightweight AI model that is ideal for low-resource applications.
2)Depth: The total number of layers in the network is referred to as depth scaling. Deeper models can capture more complex data representations, but they also need more computing resources. Shallower models, on the other hand, are more computationally efficient but may forfeit accuracy.
3)Resolution: Resolution scaling entails changing the size of the input image. Higher-resolution images give more detailed information, which may result in improved performance. They do, however, need larger memory and processing power. Lower-resolution images, on the other hand, use fewer resources but may lose fine-grained information.

The modification to the standard EfficientNetB4 architecture is inspired by advancements in natural language processing and computer vision, particularly the use of attention mechanisms in works like the transformer and residual attention networks [14],[15].

An attention mechanism is explicitly implemented into the EfficientNetB4 model to enable the network to focus on the most relevant portions of feature maps. This involves selecting feature maps from a specific layer, processing them with a convolutional layer and sigmoid activation function to obtain an attention map, and then multiplying this map with the selected feature maps.

This simple mechanism enables the network to focus only on the most relevant portions of the feature maps. It also provides us with a deeper insight on which parts of the input the network assumes as the most informative. Indeed, the obtained attention map can be easily mapped to the input sample, highlighting which elements of it have been given more importance by the network. The result of the attention block (fig. 1) is finally processed by the remaining layers of EfficientNetB4. The whole training procedure can be executed end-to-end, and we call the resulting network EfficientNetB4Att.

|e-ISSN: 2319-8753, p-ISSN: 2347-6710| www.ijirset.com | Impact Factor: 8.423| A Monthly Peer Reviewed & Referred Journal |

|| Volume 13, Issue 4, April 2024 ||

| DOI:10.15680/IJIRSET.2024.1304170 |

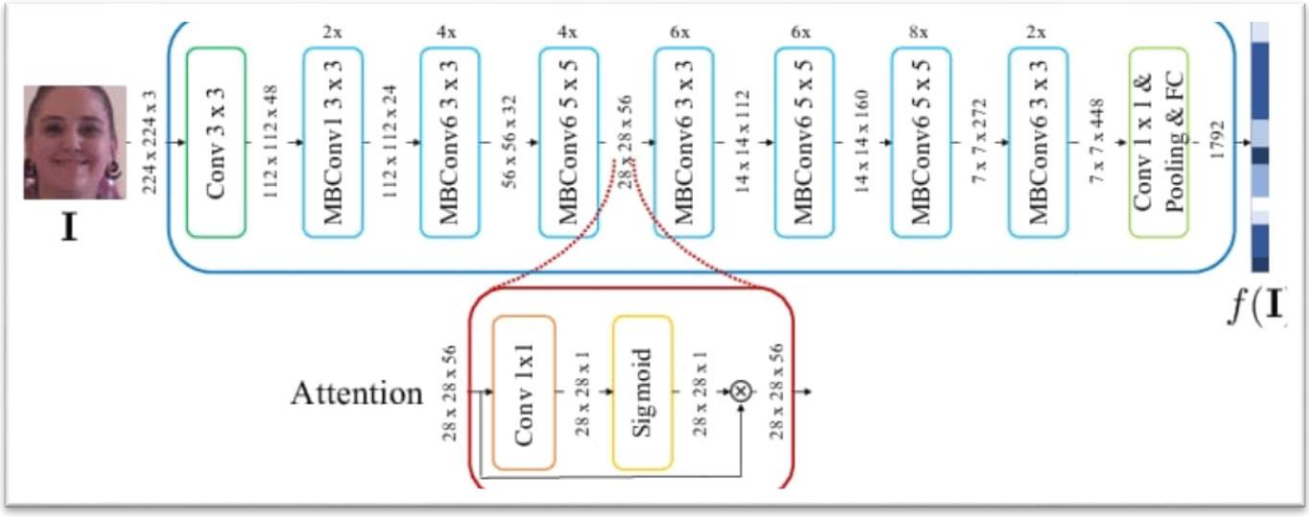


Fig. 1. EfficientNetB4 with Attention Mechanism

### B. Network training

We employ two different approaches for training the model: end-to-end training and Siamese training. The former represents a more classical training strategy, also used as evaluation metrics in the contest of DFDC. The latter aims at exploiting the generalization capabilities offered by the networks in order to obtain a feature descriptor that privileges the similarity between samples belonging to the same class. The ultimate goal is to learn a representation in the encoding space of the network's layers that well separates samples (i.e., faces) of the real and fake class.

End-to-end training: We feed the network with a sample face, and the network returns a face-related score yˆ. Notice that this score is not passed through a Sigmoid activation function yet. The weights update is led by the commonly used LogLoss function

$$L_L = -\frac{1}{N}\sum_{1=1}^{N}[y_i log\,(S(\hat{y}_i)) + (1 - y_i) log(1 - S(\hat{y}_i))]$$

where $\hat{y}_i$ represents the i-th face score, $y_i \in \{0, 1\}$ the related face label. Specifically, label 0 is associated with faces coming from real pristine videos and label 1 with fake videos. N is the total number of faces used for training and S (·) is the Sigmoid function.

Siamese networks typically consist of two identical subnetworks (or twins) that share the same architecture and parameters. These subnetworks take two input samples (often referred to as "anchor" and "positive" samples) and process them to obtain feature representations. This process optimizes the network parameters to better discriminate between features, enhancing its effectiveness in tasks such as similarity measurement or classification. This fine-tuning process enhances the network's ability to extract meaningful features and make accurate predictions tailored to the target domain.

During training, Siamese networks are trained using pairs of samples along with their corresponding labels indicating whether the samples are similar or dissimilar. The network is optimized to minimize a loss function that penalizes differences between similar pairs and encourages differences between dissimilar pairs.We use a loss function called triplet loss given by $L_T = \max(0, \mu + \delta_+ - \delta_-)$. This loss function operates on triplets of samples: an anchor sample, a positive sample (similar to the anchor), and a negative sample (dissimilar to the anchor). The network is trained to minimize the distance between the anchor and positive samples while maximizing the distance between the anchor and negative samples.

|e-ISSN: 2319-8753, p-ISSN: 2347-6710| www.ijirset.com | Impact Factor: 8.423| A Monthly Peer Reviewed & Referred Journal |

|| Volume 13, Issue 4, April 2024 ||

| DOI:10.15680/IJIRSET.2024.1304170 |

## IV. EXPERIMENTAL RESULTS

### C. Dataset

We conducted experiments using our proposed methodology on two distinct datasets: FF++ and DFDC. FF++ is a comprehensive facial manipulation dataset crafted using cutting-edge automated video editing techniques. We employed the Frame Separation method on a meticulously curated selection of 1000 high-quality pristine videos. These videos feature predominantly front-facing subjects with minimal occlusions, ensuring optimal data quality. Each sequence consists of no fewer than 280 frames, culminating in a database comprising over 1.8 million images extracted from 4000 manipulated videos. To replicate real-world conditions, we compressed the videos using the H.264 codec, generating both high-quality and low-quality variants with constant rate quantization parameters set to 23 and 40, respectively.

On the other hand, DFDC serves as the training dataset provided by Meta (formerly Facebook) for the Kaggle challenge. It encompasses more than 119,000 video sequences specifically curated for this competition, encompassing a mix of authentic and manipulated content. The authentic sequences feature diverse actors across various demographics (gender, skin tone, age, etc.) filmed against varied backgrounds to introduce visual diversity. The manipulated videos, or DeepFakes, are derived from the authentic footage and involve the application of various face-swapping algorithms and other DeepFake techniques.

### D. Networks

In our experimentation, we incorporate the following neural networks:
1)EfficientNetB4, chosen for its superior blend of accuracy and efficiency when compared to alternative methods.
2)EfficientNetB4Att, specifically designed to find out relevant features within facial samples while filtering out irrelevant ones.

Each neural network undergoes individual training and evaluation across the designated datasets. Specifically, within the FF++ dataset, we exclusively focus on videos generated using a constant rate quantization parameter set to 23. The two EfficientNet models undergo training utilizing both traditional end-to-end and siamese approaches as delineated in Section III-B. Consequently, we produce four distinct trained models: EfficientNetB4 and EfficientNetB4Att trained through the conventional end-to-end process, alongside EfficientNetB4ST and EfficientNetB4AttST, which undergo training employing the siamese methodology. Thus, the model consists of 4 Neural Networks that can be used to classify a video as genuine or deepfake.

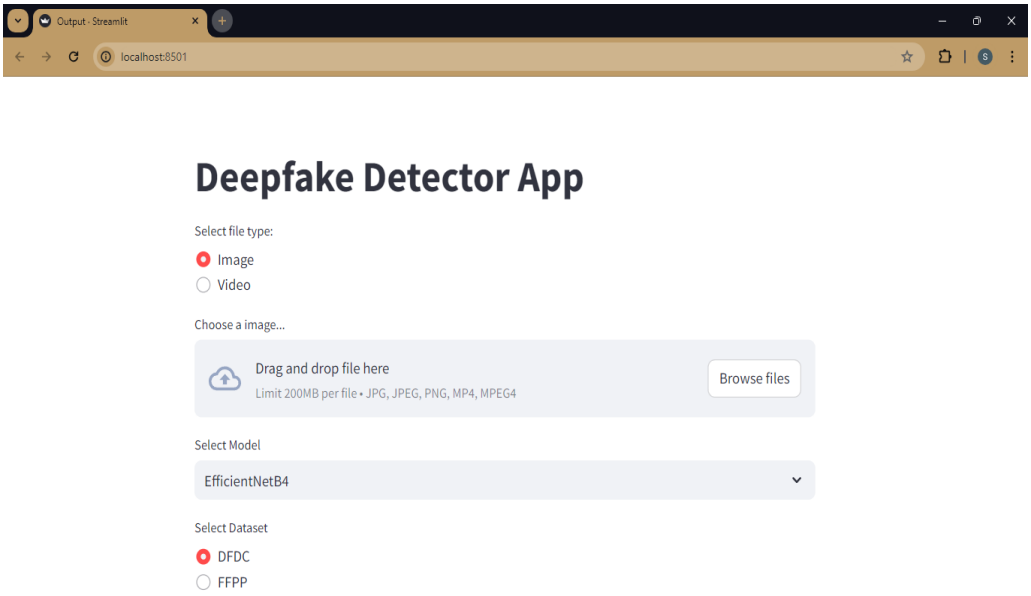


Fig. 2. User Interface

|e-ISSN: 2319-8753, p-ISSN: 2347-6710| www.ijirset.com | Impact Factor: 8.423| A Monthly Peer Reviewed & Referred Journal |

|| Volume 13, Issue 4, April 2024 ||

| DOI:10.15680/IJIRSET.2024.1304170 |

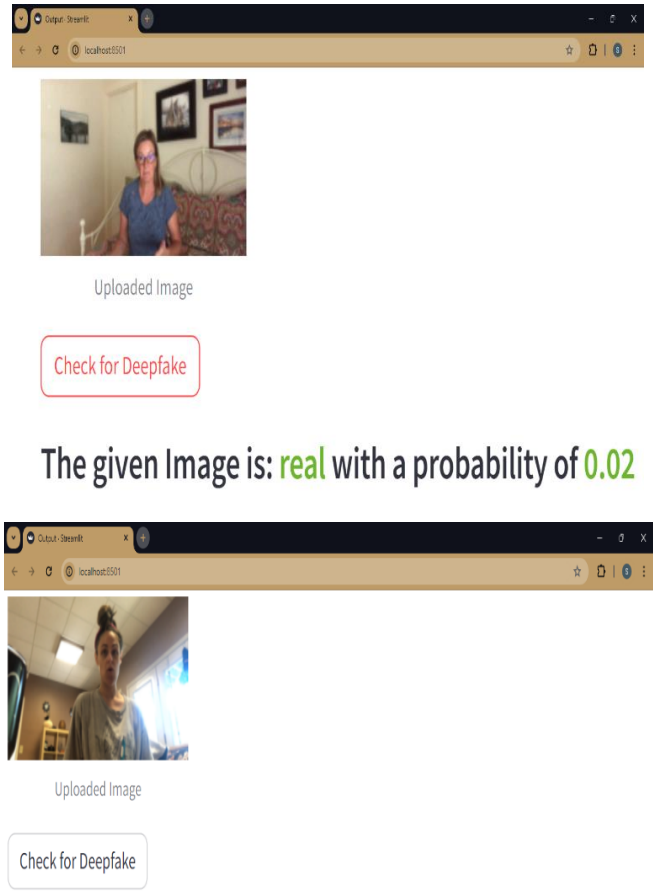


Fig.3. Real Image
Fig. 4. Fake Image

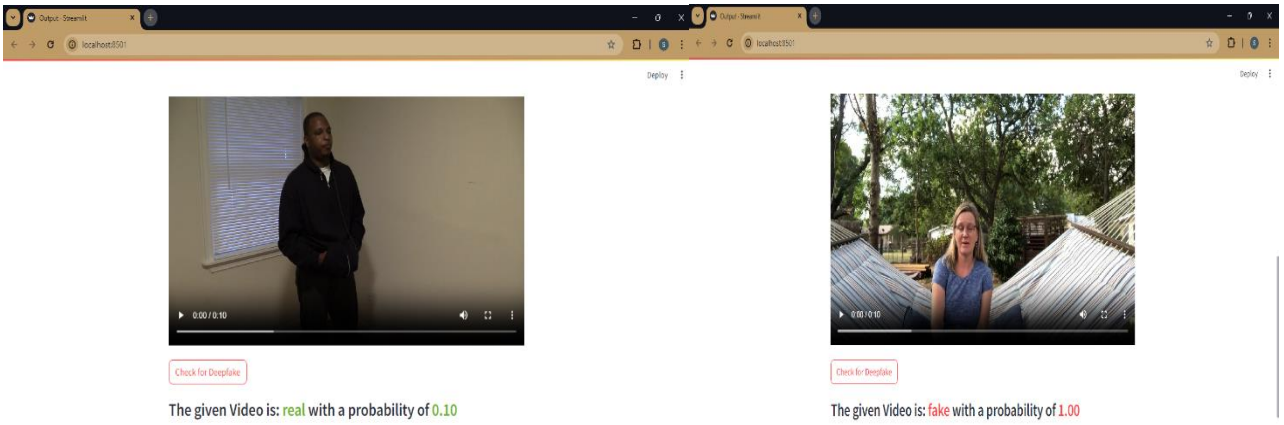


Fig. 5. Real Video

Fig. 6. Fake Video

|e-ISSN: 2319-8753, p-ISSN: 2347-6710| www.ijirset.com | Impact Factor: 8.423| A Monthly Peer Reviewed & Referred Journal |

|| Volume 13, Issue 4, April 2024 ||

| DOI:10.15680/IJIRSET.2024.1304170 |

## V. CONCLUSION

Detecting whether a video has been altered is crucial nowadays, considering how much impact videos have on our daily lives and communication. In this study, we focus on identifying facial manipulation in video clips, specifically targeting both traditional computer-generated imagery and deep learning-generated fake videos. Our approach draws inspiration from EfficientNet models and builds upon a recent method. We explore models trained using two key techniques: an attention mechanism that provides human-readable insights into the model's decisions, enhancing its learning capacity, and a triplet siamese training strategy that extracts deep features from data to improve classification accuracy. Evaluation on two publicly available datasets comprising nearly 120,000 videos demonstrates that our approach effectively detects facial manipulation in videos.

## REFERENCES

[1] FaceForensics: A Large-scale Video Dataset for Forgery Detection in Human Faces Andreas Rossler, Davide Cozzolino, Luisa Verdoliva, Christian Riess, Justus Thies, Matthias Nießner – Mar 2018
[2] The Deepfake Detection Challenge (DFDC) Preview Dataset Brian Dolhansky, Russ Howes, Ben Pflaum, Nicole Baram, Cristian Canton Ferrer AI Red Team, Facebook AI – Oct 2019
[3] Celeb-DF: A Large-scale Challenging Dataset for DeepFake Forensics Yuezun Li, Xin Yang, Pu Sun, Honggang Qi and Siwei Lyu 2019
[4] An Improved Dense CNN Architecture for Deepfake Image Detection, Yogesh Patel, Sudeep Tanwar (Senior Member, IEEE), Innocent Ewean Davidson (Senior Member, IEEE), and Thokozile F. Mazibuko, and published in March 2023
[5] Exposing Deepfake Videos by Detecting Face Warping Artifacts, Yuezun Li and Siwei Lyu from the University at Albany, State University of New York May 2019
[6] Deep Fake Video Detection Using Recurrent Neural Networks, Abdul Jamsheed V. and Janet B., April 2021
[7] Deep Detection for Face Manipulation, Disheng Feng, Xuequan Lu, Deakin University, Australia, September 2020
[8] Fighting Deepfake by Exposing the Convolutional Traces on Images, Luca Guarnera, Oliver Giudice And Sebastiano Battiato, 2020
[9] Attention-based approach for deepfake face and expression swap detection and localization, Saima Waseem, Syed Abdul Rahman Syed Abu Bakar, Zaid Omar, Bilal Ashfaq Ahmed, Saba Baloch and Adel Hafeezallah 2023
[10] Xception: Deep Learning with Depthwise Separable Convolutions, François Chollet, Google, Inc. in 2017
[11] Combining EfficientNet and Vision Transformers for Video Deepfake Detection, Davide Coccomini, Nicola Messina, Claudio Gennaro, and Fabrizio Falchi 2023
[12] EfficientNet: Rethinking Model Scaling for Convolutional Neural Networks Mingxing Tan 1 Quoc V. Le 1
[13] FaceForensics++: Learning to Detect Manipulated Facial Images, Andreas Rossler, Davide Cozzolino, Luisa Verdoliva, Christian Riess, Justus Thies, Matthias Nießner
[14] Attention Is All You Need: Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N. Gomez, Łukasz Kaiser Aug 2023
[15] Residual Attention Network for Image Classification, Fei Wang, Mengqing Jiang, Chen Qian, Shuo Yang, Cheng Li, Honggang Zhang, Xiaogang Wang, Xiaoou Tang Apr 2017
[16] On the Detection of Digital Face Manipulation Hao Dang, Feng Liu, Joel Stehouwer, Xiaoming Liu Anil Jain Oct 2020
[17] Recurrent Convolutional Strategies for Face Manipulation Detection in Videos Ekraam Sabir, Jiaxin Cheng, Ayush Jaiswal, Wael AbdAlmageed, Iacopo Masi, Prem Natarajan May 2019
[18] Deepfake Detection: A Systematic Literature Review Md Shohel Rana, Mohammad Nur Nobi, Beddhu Murali , And Andrew H. Sung, Jan 2022
[19] MesoNet: a Compact Facial Video Forgery Detection Network , Darius Afchar, Vincent Nozick, Junichi Yamagishi, Isao Echizen Sept 2018
[20] Deepfake Detection Based On The Xception Model Jamal E. Abdelkhalki, Mohamed Ben Ahmed, Aanouar Abdelhakim Boudhir Jan 2022

INTERNATIONAL STANDARD SERIAL NUMBER INDIA

INNO SPACE SJIF Scientific Journal Impact Factor

doi cross ref

निस्केयर NISCAIR

# INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

## IN SCIENCE | ENGINEERING | TECHNOLOGY

9940 572 462 6381 907 438 ijirset@gmail.com

www.ijirset.com

Scan to save the contact details