A Performance Enhancement of Deepfake Video Detection through the use of a Hybrid CNN Deep Learning Model

Original Scientific Paper

Sumaiya Thaseen Ikram
School of Information Technology and Engineering
Vellore Institute of Technology,
Vellore, Tamil Nadu, India.
isumaiyathaseen@vit.ac.in

Priya V
School of Information Technology and Engineering
Vellore Institute of Technology, Vellore,
Tamil Nadu, India.
vpriyacse@vit.ac.in

Shourya Chambial
School of Information Technology and Engineering
Vellore Institute of Technology, Vellore,
Tamil Nadu, India.
shourya39@gmail.com

Dhruv Sood
School of Information Technology and Engineering
Vellore Institute of Technology, Vellore,
Tamil Nadu, India.
sooddhruv2@gmail.com

Arulkumar V
School of Computer Science and Engineering
Vellore Institute of Technology, Vellore,
Tamil Nadu, India.
arulkumar.v@vit.ac.in

Abstract – *In the current era, many fake videos and images are created with the help of various software and new AI (Artificial Intelligence) technologies, which leave a few hints of manipulation. There are many unethical ways videos can be used to threaten, fight, or create panic among people. It is important to ensure that such methods are not used to create fake videos. An AI-based technique for the synthesis of human images is called Deep Fake. They are created by combining and superimposing existing videos onto the source videos. In this paper, a system is developed that uses a hybrid Convolutional Neural Network (CNN) consisting of InceptionResnet v2 and Xception to extract frame-level features. Experimental analysis is performed using the DFDC deep fake detection challenge on Kaggle. These deep learning-based methods are optimized to increase accuracy and decrease training time by using this dataset for training and testing. We achieved a precision of 0.985, a recall of 0.96, an f1-score of 0.98, and support of 0.968.*

Keywords: *Deepfake, Machine learning, Deep learning, Inception, Xception*

1. INTRODUCTION

Information sharing and broadcasting are now much easier and faster, thanks to the growth of social media platforms. With only one click, people may now access knowledge from around the globe. Regarding news consumption, social media platforms can be utilized for two different purposes: to alert the public of breaking news or, conversely, to disseminate false information [1].

DeepFakes is a popular concept with widespread application. Deepfakes ("fake") are synthetic media (AI-generated media) in which an existing image or clip of a person is superimposed with another person's image [2] [3].

To damage the character's reputation, deepfake technology is used to replace performers' faces in pornography, revenge porn, fake news, hoaxes, and financial fraud with the faces of celebrities. This has spurred business and government actions to identify and forbid their use. The three most risky ways to apply face-swapping algorithms identified are as follows: (i) Face-swap, in which one face is automatically superimposed on another; (ii) Lipsync, a technique in which only a portion of a person's face is altered, forcing them to utter things they have never said before; and (iii) puppet master, in which the face of the target individual is animated by a person sitting in front of the camera [4]. FakeApp, created by a Reddit user using the auto encoder-decoder pairing structure, was the first deepfake generation attempt. The face images are broken down into their parts in this manner by the autoencoder, which also extracts latent properties from the face images. Two encoder-decoder pairs, each trained on a different image set, are required to swap faces between the source and target images. The two network pairs share the encoder's parameters. Alternatively, two pairs share a common encoder network [5]. Many businesses, including Facebook Inc., Google, and the United States Defense Advanced Research Projects Agency (DARPA), have launched a research initiative to find and eliminate deep fakes [6] and [7]. Numerous deep learning methods, including long short-term memory (LSTM), recurrent neural networks (RNN), and even hybrid approaches, have been created to detect deep fakes in images and videos, and additional research has been conducted in this area [8] and [9].

Many studies have been done on deep-fake detection due to the quick development of face swaps and other video manipulation technologies. Various attempts have been made to find a solution to this problem. Visual artifacts, common among deep fakes [10], have been used frequently in solution strategies. The Deepfake Detection Challenge was developed in collaboration with META, Microsoft, and AWS on AI's Media Integrity Steering Committee and academics (DFDC). The challenge's purpose is to persuade scholars worldwide to create successful new techniques for detecting deep fakes and controlling the media. In another instance, Google researchers announced the AI Principles, stating that they are committed to developing AI models that reduce the risk of harm and misuse [11]. The researchers contributed a synthetic speech dataset in 2018 to aid in a big competition to build very effective fake audio detectors. In 2019, they contributed a sizable collection of visual deepfakes.

The primary goal of this paper is to examine the available approaches, highlight trends, and address the current issues in the investigations. Finally, the performance of the various techniques is analyzed. This paper proposes a new hybrid technique based on Inception Resnet V2 and Xception. Multiple input samples, such as positive, negative, and generated samples, are used to train the Xception and InceptionResnet v2 networks for classification. During the training process, a regularization loss is implemented to ensure the embedding space's inter-class proximity and intra- class regularity.

Spreading deep fakes over social media platforms has grown increasingly common, resulting in spamming and speculation based on inaccurate information. These deep lies will be terrifying and deceiving to the general population. Deep fake detection is essential for fixing this issue. As a result, we describe a unique deep learning-based technique for distinguishing between AI-generated false films (deep fake Movies) and true videos. It is vital to develop technologies capable of identifying forgeries to detect and prevent deep fakes from spreading over the internet.

Our work aims to develop a robust and efficient model to help reduce the threat posed by malicious users who try to exploit online and open-source images for unethical purposes and to malign a person's image. It also aims to reduce the false information spread by these fake videos.

Section 2 describes existing techniques for deepfake detection; Section 3 describes the background; Section 4 describes the methodology, which also includes dataset description, data preprocessing, the proposed technique, and the technology used; Section 5 includes results and analysis, and Section 6 presents conclusions and future work.

2. RELATED WORK

2.1 DETECTION BASED ON ML

Xin et al. [12] developed a system against exposing AI-generated fraudulent face pictures or videos and compared head locations computed using all visual indicators to those judged using only the center area. Li et al. [13] identified blinking of eyes in films, a behavioral indication poorly represented in the bogus film. Falko et al. [14] presented a collection of simple characteristics for recognizing produced faces, deep fakes, and Face2Face pictures in the eyes, teeth, and facial contours. Guarnera et al. [15] examined bogus videos of human faces to develop a novel discernment approach capable of detecting a forensics trail buried in photos.

2.2 DETECTION BASED ON CNN

A. Facial Tampering

Guera and Delp [16] devised a solution consisting of key components of a convolutional neural network and long short-term memory. After combining the attributes of many consecutive frames, CNN creates a collection of features for each frame in a particular picture sequence and provides them to the LSTM for analysis. The suggested model underwent training on 600 videos and attained an accuracy of 97.1 percent.

Li and Lyu [17] established a technique for identifying distorted images in manipulated films with an accuracy of up to 99 percent when trained with four distinct deep-learning models on legitimate and modified photos. Zhou et al. [18] proposed a multi-stream network for facial recognition modification in the deepfake. A deep learning face classification model is being trained in the first stream to collect evidence of tampering with artifacts. In the second stream, a steganographic model-based multi-layer network is trained to regulate functions that collect leftover noise evidence nearby. Afshar et al. [19] built MesoNet, a CNN, to distinguish between the actual and Deepfake- modified faces. Meso-4 and MesoInception-4 are two models based on inception used in the network, along with layers linked with the max-pooling function. Khalid and Simon [20] developed a one-class approach for identifying deepfakes and achieved 97.5% accuracy on the face forensics++ dataset without having any fake images in the training samples. The authors in [22] developed a strategy for creating a Deepfake detector dubbed FakeCatcher (FC), which emphasizes using features derived from face regions to recognize synthetic portrait films. Missing reflections and minute features in the facial areas are exploited, and characteristics from the face are retrieved from the essential facial features and supplied into machine learning classifying models for identifying them as fake or real films.

B. Digital Media Forensics

Oza and Patel [23] developed a One-class convolutional Neural Network as an instance of a one-class-based technique (OC-CNN). The primary notion behind OC-CNN is to employ a negative class of zero-centered Gaussian noise in the hidden space and train the network using cross-entropy loss. Cozzolino et al. [24] proposed ForensicTransfer (FT), an architecture based on autoencoders that distinguish legitimate from tampered photos. The ForensicTransfer contacts multiple tests and results with an accuracy rate of up to 80% to 85%. Nguyen et al. [25] suggested an aggregate deep-learning method for simultaneously detecting and dividing altered pictures and clips. The suggested system includes an encoder that encodes binary classification characteristics and a Y-shaped decoder that adopts the results from one of its sub-branch to partition the modified areas. The authors in [26] reported a deep learning model that detects Deepfake using a capsule network (CN). Furthermore, it detects replay assaults and computer-generated images.

3. BACKGROUND

A. Generic Overview

This section attempts different ways to determine whether videos are fake. The annotations are saved in a JSON file in the train sample videos folder, and a video dataset is used for analysis.

The stages involved are as follows:-

- Reading and collecting images from the videos.
- The image is placed in the correct folder after reading the label from the JSON file.
- After converting the image to an array, the data is divided into train and test groups.
- Using InceptionResNetV2 and Xception to train data and customize them.
- Testing for accuracy and outcomes.

The dataset is pre-processed first in the low-level design, and then the model is trained, tested, and results in predictions. The DFDC dataset is used for experimental analysis. The dataset is pre-processed. The model's initial state consists of frames of real and fake images being generated under the real and fake folders, respectively, and these images will be the input for the model. Finally, the model is tested using test videos and produces the desired output.

High level: This is the overview of the system. In the proposed approach, both real and fake images are considered. Fake images are generated using a generator and a discriminator to discriminate between fake and real images. The low-level and high-level diagrams of the proposed approach are shown in Fig. 1 and Fig. 2.

The primary contribution of Inception-V3 and Xception is that they mix numerous convolution filters, such as Conv (1–1), Conv (3–3), and Conv (5–5), in a multi-extractor. Typically, the Inception-V3 design has 22 convolutional layers and 5 pooling layers. Because of the variety of Inception and its high memory requirements in V3, a more optimized version of the creation family known as Xception has been proposed to reduce computational complexity. Separable convolutions have been suggested in this variant [27].

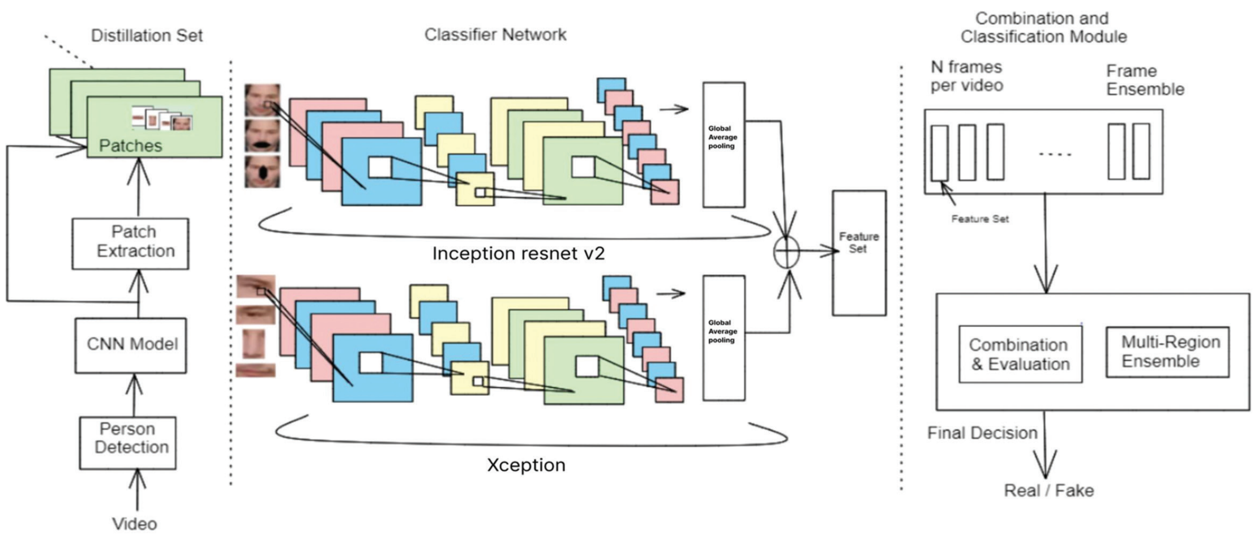

Fig. 1. Low–level diagram of modules

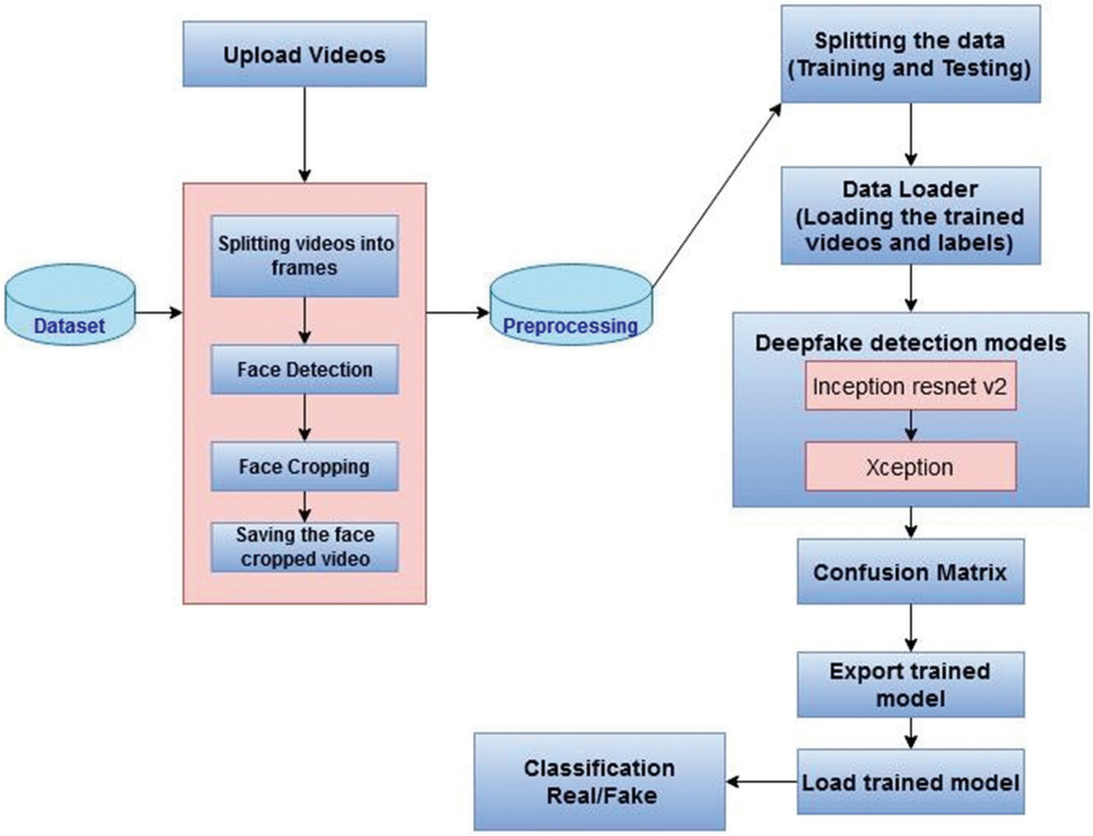

Fig. 2. High–level diagram

B. About Inception ResNet v2

Inception-ResNet-v2 is a convolutional neural network trained on millions of images from the image net collection. It has a network of over one hundred and fifty layers and has been used to classify pictures into thousand distinct items such as flowers, food, aeroplanes, etc. Therefore, as a consequence, the framework has learned a wide range of rich feature interpretations for a wide range of pictures. Consider an input image with dimensions of around 300 by 300 pixels to generate a list of anticipated class probabilities.

C. Working of Inception ResNet V2

The basis of the model is based on the structure of Inception, ensembled with the residual connection. Several combinations are made between residual connections and convolutional filters of various sizes. The residual connections have been utilized to handle the degradation problem and have even helped reduce the training time by fifty percent. The output from the inception model is added to the current input connections of the residual network.

The input and output dimensions must be in sync to perform the residual addition. One by one, convolution has been utilized to match the depth size. The Inception network has three modules, A, B, and C, to form the entire network. The pooling layer will be replaced by a one-by-one convolution layer along with the residual connection network.

D. Pseudocode for Inception Network followed by ResNet

This pseudocode is used to identify and classify images.

Input: clips and frames of images.

Output: The face is detected using a boundary-based box.

- A one-by-one convolution non-activation layer is added to the network to match the depth.
- Summation layers are not a part of the batch normalization process; apart from them, normalizations are used everywhere.
- Residual values were then scaled down before being added to the prior layer activation, which helped to stabilize the training. Scaling values of 0.1 to 0.3 were chosen to scale the residuals.
- A combination of residual connections with several-sized filters happens in the network block. This brings down the training time by fifty percent.
- Factorize five-dimensional convolution into two three-dimensional convolution processes to increase computing speed. A five-dimensional convolution costs around three times as much as a three-dimensional convolution, which may look contradictory. Stacking two three-dimensional convolutions enhances performance as a consequence.

- The dimensions would be considerably decreased if the module was made deeper instead, resulting in information loss. The filter banks of the module were thus expanded to remove the above factor.

The architecture of Inception ResNet V2 is shown in Figure 3. Inception Architecture ResNet V2, The architecture of Inception ResNet V2, is shown in Figure 3.

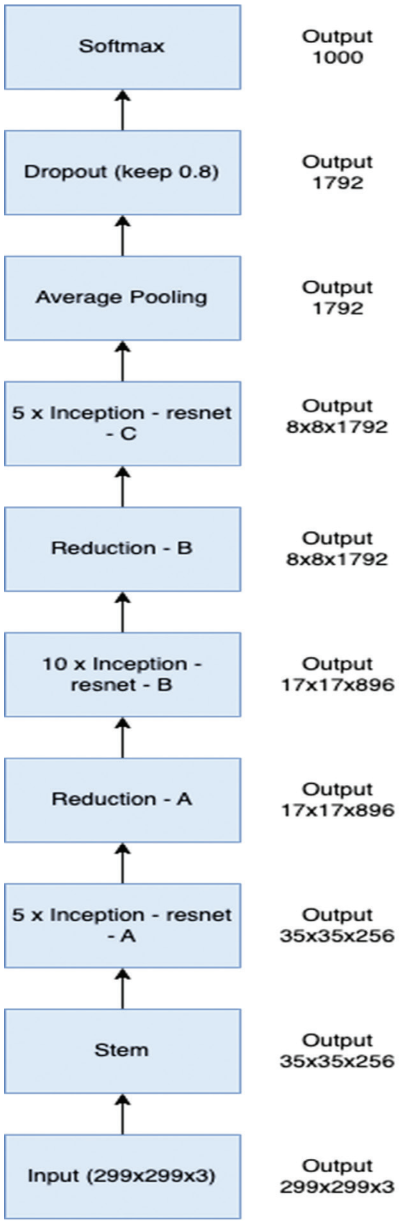

Fig. 3. Architecture of InceptionResNet V2

E. About Xception

Firstly, the data is passed through the input flow, through the middle flow eight times, and finally, through the exit flow. Every layer of separable convolution and convolution is subject to batch normalization, and the architecture is shown in Fig 4.

F. Working of Xception

Xception is a very efficient deep-learning model that depends on the following:

- Depth-wise Separable Convolution
- As in ResNet, there are shortcuts between Convolution blocks.

The architecture of Xception is made up of Depth wise separable convolution blocks and max-pooling, all of which are coupled via shortcuts in the same way as ResNet implementations. A Pointwise convolution does not follow the Depth wise convolution in Xception; instead, the sequence is inverted.

G. Pseudocode for Xception

- All the necessary layers must be imported Necessary functions must be written for
- Conv-BatchNorm block
- SeparableConv- BatchNorm block
- For each of the three flows (Entry, Middle, and Exit), write a separate function.
- Utilize these features to create the full model.

4. PROPOSED METHODOLOGY

A. Dataset

The DFDC dataset is used for the experiments. Many deepfake or face swap datasets include films shot in non-natural environments like news or briefing rooms. Worse, the people in these films may not have consented to have their faces modified. With over 100,000 total clips collected from 3,426 paid actors and produced using a variety of Deepfake, GAN-based, and non-learned algorithms, the DFDC dataset is by far the largest currently and publicly available face swap video dataset.

Each of the 100,000 forged videos in the DFDC Dataset is a one-of-a-kind target/source switch. DF-1.0 consists of 1,000 distinct bogus videos, despite the disruptions. The DFDC dataset includes movies of people in indoor and outdoor situations, with a wide range of lighting situations.

The various datasets available for Deepfake detection have been tabulated in Table 1.

B. Our Approach

Deepfakes news is influencing the globe because individuals worldwide use it for various purposes, including face swapping, reproducing pornographic movies with someone's face or body, and manufacturing and disseminating fake news.

Deep Fakes are increasingly harming democracy, privacy, security, religion, and people's cultures. Deep Fakes are becoming more common, yet there is no standard for evaluating deep fake detection systems. Since 2018, the number of deep fake movies and photos discovered online has nearly doubled. For more than ten years, the Massachusetts Institute of Technology (MIT) evaluated 126,000 news stories shared by 3,000,000 individuals. Finally, they determined that bogus news travels 1,500 times faster than accurate news. Deepfakes create fake news, photos, videos, and terrorist events. Deepfake undermines public faith in the media and contributes to social and financial fraud. Religions, organizations, politicians, artists, and voters are all affected by deepfake. People will disregard the truth as deepfake videos and pictures proliferate on social media.

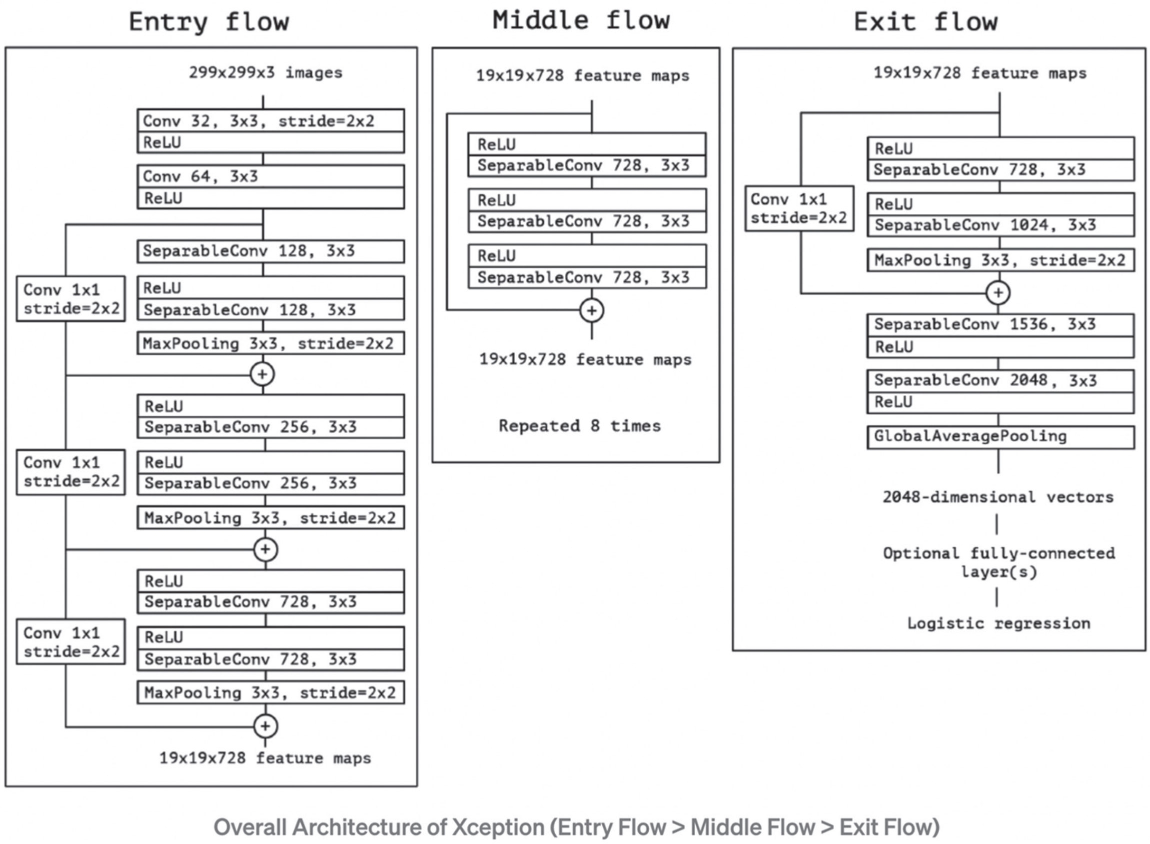

Fig. 4. Architecture of Xception [28]

	Train Data		Test Data	
	Real	**Fake**	**Real**	**Fake**
UADFV	35 videos (13976 frames)	35 videos (13638 frames)	14 videos (3353 frames)	14 videos (3353 frames)
Celeb-DF	370 videos (158992 frames)	733 videos (290043 frames)	38 videos (16409 frames)	62 videos (22834 frames)
Deepfake Detection	254 videos (202723 frames)	2148 videos (1678558 frames)	109 videos (94437 frames)	920 videos (681550 frames)

For learning temporal aspects of facial data from training films, a hybrid deep learning model employing CNN (Convolutional Neural Network) models consisting of Inception Resnet v2 followed by Xception is proposed. We have suggested a CNN-based model that learns different patterns between Deep-Fake and actual videos. Pixel distortion, discrepancies with facial superimposition, skin color variances, blurring, and other visual aberrations are among these distinguishing characteristics. Using a frame-based technique based on the aforementioned different properties, the suggested approach has successfully trained a CNN (convolutional neural network) to discern DeepFake films. The proposed work, which involves an ensemble of Inception and xception, shows the viability of our model's ability to identify deep fake faces in a specific video source accurately. This will help security applications used by social media platforms combat the growing threat of "deepfakes" by accurately determining the authenticity of videos, allowing them to be flagged or removed before they cause harm that cannot be repaired.

The dataset is imported and converted based on metadata training and labeling in a JSON file. All face frames were cropped, aligned, and reduced to 256x256 pixels after internal face tracking and alignment were utilized to preprocess the source videos. 5,000 face frames were used to train models. The Inception ResNet v2 model feeds temporal features to the Xception model. The Xception model's feedback architecture may learn from consecutive inputs. We trained our model with 10 epochs and 25 batches. An "epoch" is a machine learning term that describes how many rounds the algorithm did across the full training dataset. Once educated, the ".h5" file can be downloaded. Hybridization successfully leverages many model layers to boost learning performance.

Table 2. Accuracy Achieved

Model	Accuracy
Inception_ResNet_v2	89.41%
Xception	93.85%
Hybrid Model	95.75%

C. Our Methodology

In the proposed approach, both real and fake images are considered. Fake images are generated using a generator, and then a discriminator is used to differentiate between fake and real images. In the low-level design, the dataset is pre-processed first, and then the model is trained, tested, and the results are determined. The DFDC dataset is used for experimental analysis. The dataset is pre-processed. The model's initial state consists of frames of real and fake images generated under the real and fake folders, respectively, and these images will be the input for the model. Finally, the model is tested using test videos and produces the desired output.

The low-level diagram depicting the flow of events is shown in Fig. 5.

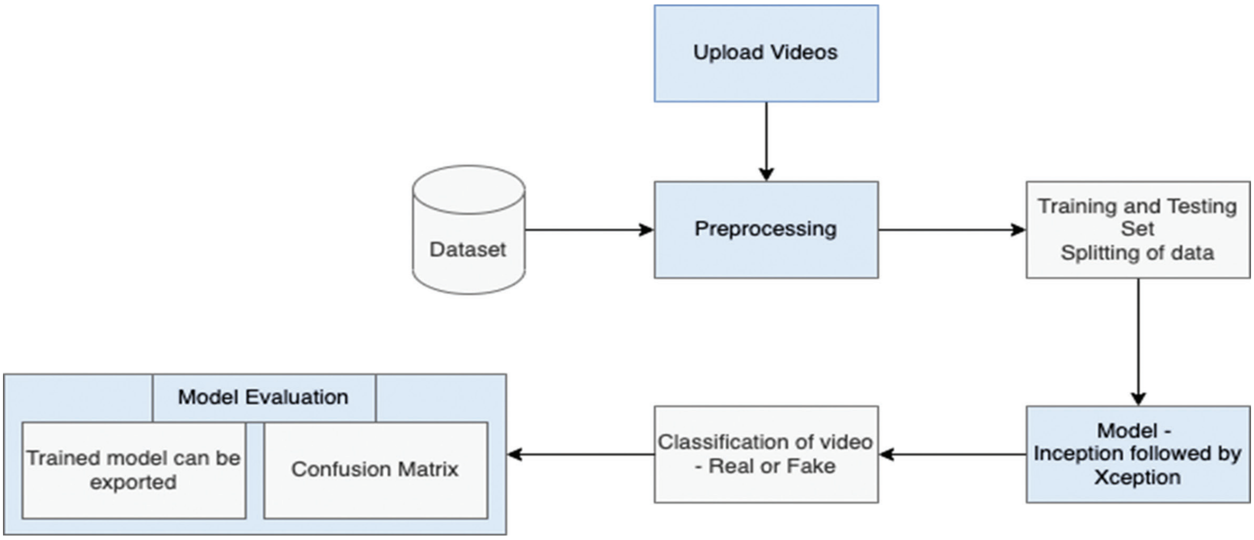

Fig. 5. Low-level diagram depicting the flow of events

The flow diagram is shown in Fig. 6.

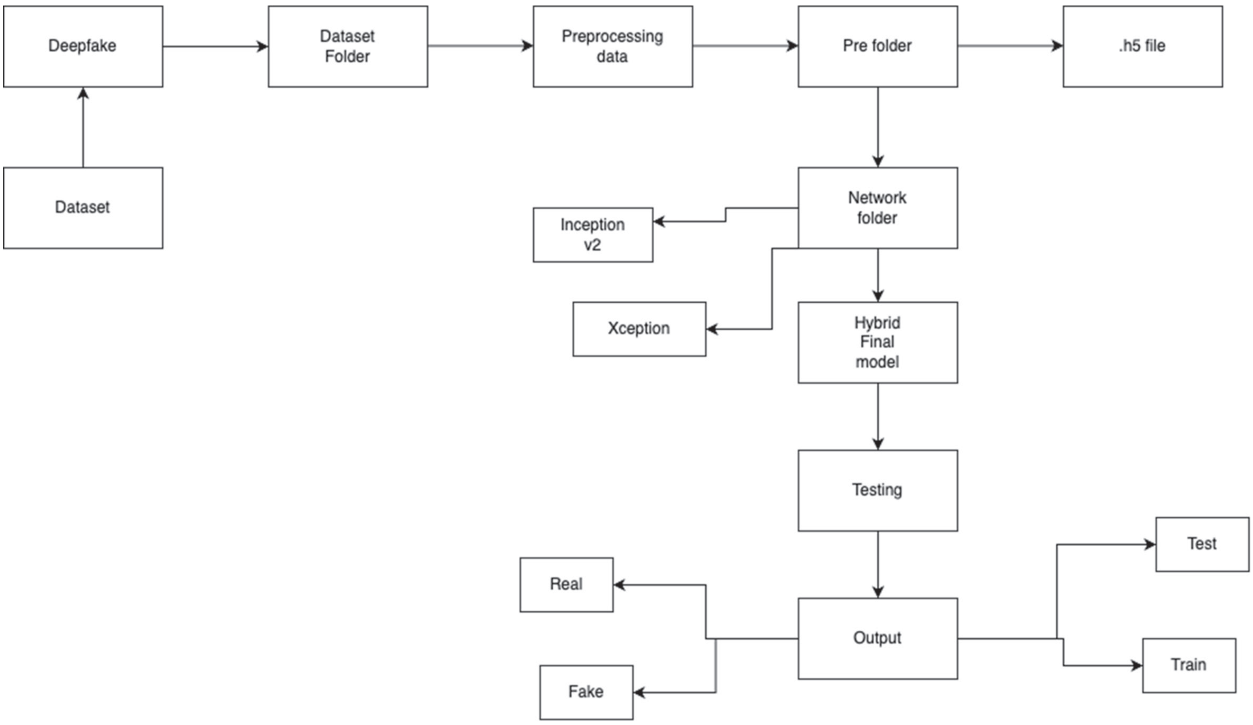

Fig. 6. Flow diagram

The pipeline design of our architecture is shown in Fig. 7.

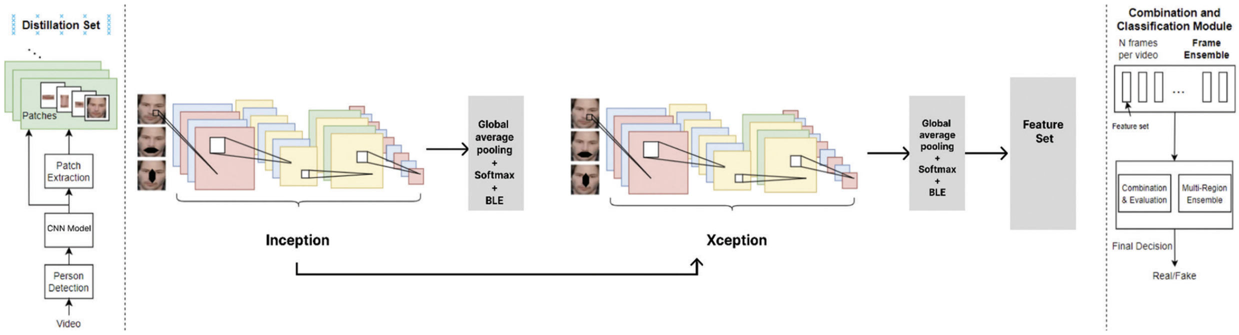

Fig. 7. Architecture of the proposed method

5. RESULTS

The method described in this paper is a way to identify deepfake pictures by utilizing the sign of the source to highlight irregularities inside the manufactured pictures. It is based on the theory that images distinguished by source features can be protected and removed after going through best-in-class deep-learning processes. The work presented here presents a smart portrayal of the learning approach, known as pairwise self-consistency learning (PCL), used for preparing convolutional networks to separate the origin highlights and distinguish bogus pictures. It is combined with an irregularity picture generator (I2G) method to produce clear information for PCL. The ROC curve for the proposed work is shown in Fig. 8. It is a plot of the true positive rate as a function of the false positive rate for various cut-off points of a parameter.

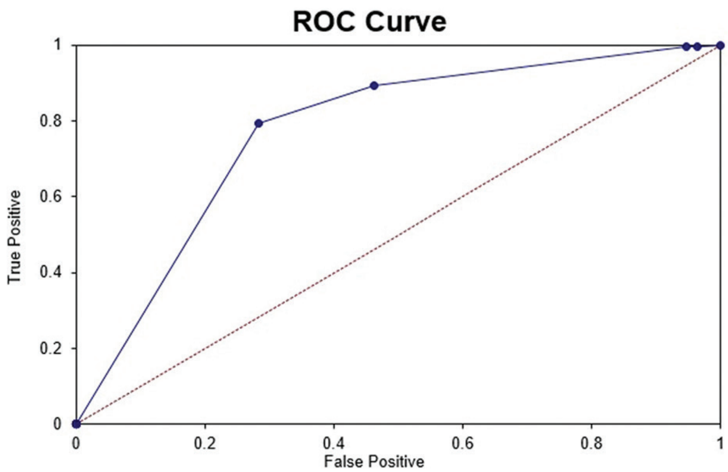

Fig. 8. ROC Curve for Training and Validation

Exploratory outcomes on Inception_resnet_v2, Xception, and hybrid are tabulated below. The evaluation metrics of the Inception, Xception, and Hybrid models are shown in Tables 3(a), 3(b), and 3(c), respectively. The hybrid model received a training accuracy of 0.98 and a validation accuracy of 0.93, respectively.

Table 3. Performance of Base and Hybrid Model

(a) Evaluation metrics for the Inception

	Inception			
	precision	recall	f1-score	support
0(FAKE)	0.98	0.97	0.97	712
1(REAL)	0.88	0.91	0.89	185
Accuracy			0.96	897
Macro avg	0.93	0.94	0.93	897
Weighted avg	0.96	0.96	0.96	897

(b) Performance of Inception Model

	Xception			
	precision	recall	f1-score	support
0(FAKE)	0.98	0.99	0.98	712
1(REAL)	0.97	0.91	0.94	185
Accuracy			0.98	897
Macro avg	0.97	0.95	0.96	897
Weighted avg	0.98	0.98	0.98	897

(c) Performance of Hybrid Model

	Hybrid			
	precision	recall	f1-score	support
0(FAKE)	0.98	1.00	0.99	1486
1(REAL)	0.99	0.92	0.96	451
Accuracy			0.98	1937
Macro avg	0.99	0.96	0.97	1937
Weighted avg	0.98	0.98	0.98	1937

6. CONCLUSION

This approach uses a CNN-based model to uncover the bogus clips. The model performed well on the DFDC dataset, including low- and high-quality movies. The outputs of tampered videos highlight that by adopting a hybrid network of Inception Resnet v2 and Xception, it can identify whether a clip has ever been deceived. This work is an effective first line of defense in detecting bogus media made with online technologies. In addition, the model can attain competitive output by adopting a pipeline design, which is also demonstrated. In the future, we can use subtle tactics during training to see how we can strengthen the system against false accusations. The experimental analysis demonstrates that the enhancements have greatly improved deepfake detection results, with maximum precision, recall, and f1-score of 0.98, respectively. Simultaneously, because video forgery technology and the caliber of video are still developing, it will be possible to facilitate the proposed model.

7. REFERENCES

[1] S. Senhadji, R. A. San Ahmed, "Fake news detection using naïve Bayes and long short term memory algorithms", IAES International Journal of Artificial Intelligence, Vol. 11, No. 2, 2022, pp. 748-754.

[2] K. N. Ramadhani, R. Munir, "A Comparative Study of Deepfake Video Detection Method", Proceedings of the 3rd International Conference on Information and Communications Technology, November 2020, pp. 394-399.

[3] D. Pan, L. Sun, R. Wang, X. Zhang, R. O. Sinnott, "Deepfake Detection through Deep Learning", Proceedings of the IEEE/ACM International Conference on Big Data Computing, Applications and Technologies, December 2020, pp. 134-143.

[4] A. A. Maksutov, V. O. Morozov, A. A. Lavrenov, A. S. Smirnov, "Methods of deepfake detection based on machine learning", Proceedings of the IEEE Conference of Russian Young Researchers in Electrical and Electronic Engineering, January 2020, pp. 408-411.

[5] T. T. Nguyen, Q. V. Nguyen, D. T. Nguyen, "Deep learning for deep fakes creation and detection: A survey", Computer Vision and Image Understanding, Vol. 223, 2022, pp. 1-19.

[6] A. O. Kwok, S. G. Koh, "Deepfake: A Social Construction of Technology Perspective", Current Issues in Tourism, Vol. 24, No. 13, 2020, pp. 1798-1802.

[7] M. Westerlund, "The Emergence of Deepfake Technology: A Review", Technology Innovation Management Review, Vol. 9, No. 11, 2019, pp. 40-53.

[8] Y. Li, S. Lyu, "Exposing Deepfake Videos by Detecting Face Warping Artifacts", arXiv:1811.00656, 2018.

[9] A. M. Almars, "Deepfakes detection techniques using deep learning: a survey", Journal of Computer and Communications, Vol. 9, No. 5, 2021, pp. 20-35.

[10] N. S. Ivanov, Arzhakova, V. G. Ivanenko, "Combining deep learning and super-resolution algorithms for deep fake detection", Proceedings of the IEEE Conference of Russian Young Researchers in Electrical and Electronic Engineering, January 2020, pp. 326-328.

[11] K. Zhu, B. Wu, B. Wang, "Deepfake detection with clustering-based embedding regularization", Proceedings of the IEEE fifth International Conference on Data Science in Cyberspace, July 2020, pp. 257-264.

[12] F. Matern, C. Riess, M. Stamminger, "Exploiting visual artifacts to expose deep fakes and face manipulations", Proceedings of the IEEE Winter Applications of Computer Vision Workshops, January 2019, pp. 83-92.

[13] E. Sabir, J. Cheng, A. Jaiswal, W. AbdElmageed, I. Masi, P. Natarajan, "Recurrent convolutional strategies for face manipulation detection in videos", Interfaces (GUI), Vol. 3, No. 1, 2019, pp. 80-87.

[14] D. Güera, E. J. Delp, "Deepfake video detection using recurrent neural networks", Proceedings of the 15th IEEE International Conference on Advanced Video and Signal Based Surveillance, November 2018, pp. 1-6.

[15] Y. Li, S. Lyu, "Exposing deepfake videos by detecting face warping artifacts", arXiv:1811.00656, 2018.

[16] Y. Li, M. C. Chang, S. Lyu, "In ictu oculi: Exposing ai created fake videos by detecting eye blinking", Proceedings of the IEEE International Workshop on Information Forensics and Security, December 2018, pp. 1-7.

[17] D. Afchar, V. Nozick, J. Yamagishi, I. Echizen. Mesonet, “A compact facial video forgery detection network”, Proceedings of the IEEE International Workshop on Information Forensics and Security, December 2018, pp. 1-7.

[18] P. Zhou, X. Han, Morariu, L. S. Davis, “Two-stream neural networks for tampered face detection”, Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops, July 2017, pp. 1831-1839.

[19] B. Schölkopf, J. C. Platt, J. Shawe-Taylor, A. J. Smola, R. C. Williamson, “Estimating the support of a high-dimensional distribution”, Neural Computation, Vol. 13, No. 7, 2001, pp. 1443-71.

[20] H. Khalid, S. S. Woo, “OC-FakeDect: Classifying deepfakes using one-class variational autoencoder”, Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops, 2020, pp. 656-657.

[21] M. F. Ahmed, M. S. Miah, A. Bhowmik, J. B. Sulaiman, “Awareness to Deepfake: A resistance mechanism to Deepfake”, International Congress of Advanced Technology and Engineering, July 2021, pp. 1-5.

[22] U. A. Ciftci, I. Demir, L. Yin Fakecatcher, “Detection of synthetic portrait videos using biological signals”, IEEE Transactions on Pattern Analysis and Machine Intelligence, 2020, pp. 1-17.

[23] P. Oza, V. M. Patel, “One-class convolutional neural network”, IEEE Signal Processing Letters, Vol. 26, No. 2, 2018, pp. 277-81.

[24] D. Cozzolino, J. Thies, A. Rössler, C. Riess, M. Nießner, L. Verdoliva, "Forensictransfer: Weakly-supervised domain adaptation for forgery detection", arXiv:1812.02510, 2018.

[25] H. H. Nguyen, F. Fang, J. Yamagishi, I. Echizen, “Multi-task learning for detecting and segmenting manipulated facial images and videos”, Proceedings of the 10th International Conference on Biometrics Theory, Applications and Systems, September 2019, pp. 1-8.

[26] H. H. Nguyen, J. Yamagishi, I. Echizen, “Capsule-forensics: Using capsule networks to detect forged images and videos”, Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing, May 2019, pp. 2307-2311.

[27] A. Kherraki, R. El Ouazzani, “Deep convolutional neural networks architecture for an efficient emergency vehicle classification in real-time traffic monitoring”, IAES International Journal of Artificial Intelligence, Vol. 11, No. 1, 2022, pp. 110-120.

[28] S. H. Tsang, “Review: Xception-with depthwise separable convolution, better than inception-v3 (image classification)”, https://towardsdatascience.com/review-xception-with-depthwise-separable-convolution-better-than-inception-v3-image-dc967dd42568 (accessed: 2022)