ISSN 1840-4855
e-ISSN 2233-0046

Original scientific article
http://dx.doi.org/10.70102/afts.2026.1835.558

DEEP INCEPTION V7 INTEGRATED CONVOLUTIONAL NEURAL NETWORK FOR DEEPFAKE PREDICTION IN ONLINE CONTENT USING ADVANCED TECHNOLOGY

S. Narmatha[1*], Dr.S. Mythili[2]

[1]Research Scholar, Department of Computer Science, Karpagam Academy of Higher Education, Coimbatore, Tamil Nadu, India. e-mail: narmatha289@gmail.com, orcid: https://orcid.org/0009-0005-4766-9451*
[2]Professor and Head, Department of Computer Science, Karpagam Academy of Higher Education, Coimbatore, Tamil Nadu, India.
e-mail: mythili.shanmughasundaram@kahedu.edu.in,
orcid: https://orcid.org/0000-0003-3196-6257

Received: January 20, 2026; Revised: March 02, 2026; Accepted: April 21, 2026; Published: May 29, 2026

SUMMARY

The technologies of deepfakes are actively used on social media, creating fake video and audio by manipulating existing media content. The face-swapping technologies that are applied in the creation of deepfakes cause severe problems in society, such as identity theft and the spread of unsuitable content. Although various machine learning and deep learning techniques have been used to detect deepfakes, the generalization abilities and linguistic properties of deepfakes remain problematic. This paper presents the Deep Inception V7 Convolution Neural Network (CNN) model to identify deepfake content in different types of image synthesis in videos. The proposed model works on the principles of spatiotemporal specifics of video frames, detecting faces and body parts. The first step is converting videos into frames, and the second step is to segment them and isolate face and body features. These divided parts of the face are then fed into the deep inception network that records the spatiotemporal variation in convolution and max-pooling layers to produce maps of feature richness with contextual information. These feature maps are sent to a fully connected layer to differentiate between real and fake videos. The Adam optimizer is used to optimize the model, making it more robust and accurate. The DFDC benchmark dataset is used to perform training and validation. Through performance analysis, it has been identified that the proposed model has a training accuracy of 96.87% and a validation accuracy of 94.74%, which is better than the current methods of deepfake detection. The findings illustrate that the Deep Inception V7 CNN is effective in detecting deepfakes in different settings. These results have significant implications for enhancing information transparency in online content through the use of advanced technology.

Key words**:** *deepfake detection, convolution neural network, inception v7, DFDC dataset, spatiotemporal analysis, video segmentation, online content, advanced technology, information transparency.*

INTRODUCTION

Deepfake movies are created by artificial intelligence (AI) and can be edited in parts (spatially localized and fine-grained) with the help of sophisticated AI-generating methods [1]. These videos are made with the help of the face swap technique and are made with the help of the combination of conventional vision and voice impersonation [1]. The deepfake videos rely on Generative Adversarial Networks (GANs) to generate them [1]. The help of software like FakeApp, and a non-rigid model-based bundling technology makes it possible to record and re-enact facial expressions in video in real-time [1]. Deepfake videos have a large intra-class difference and a small inter-class difference [1]. This implies that there is similarity in the videos generated. There have been public appeals over the unhealthy spread of deepfake videos, which are actively being misused. The techniques used to identify whether there is a deepfake in the videos include Capsule-Forensics, MesoNet, forensic transfer, and deep learning classification models [1]. The deepfake films can be attributed to the progress in artificial intelligence and deep neural networks [3], which have been of interest to the computer graphics and computer vision community [4]. This has resulted in little effort to understand the deepfake films in a real-life situation. In this case, the largest set of deepfake videos is 1,869 videos obtained on YouTube and Bilibili [2]. The problem of deepfake video is growing, and the two domains of deepfake methods are video generation and detection, which belong to the deepfake video analysis field [1].

The process of developing deepfake videos is based on machine learning that is viewed as Generative Adversarial Networks (GANs) that can produce believable fake audio, images, and videos [4]. The method has been applied to produce exceptionally realistic movies of people performing any activities want or saying whatever want [4]. Deepfakes may also be used to manipulate not only the voice but also facial expressions in a photo or video [5]. Although the paper also addresses the various methods of deepfake creation, it also explains how to detect them, which in most cases a human being may not easily notice [5]. Deepfakes could be used to substitute or conceal the faces of other individuals in photographs and videos [5]. This is a challenge to humans to determine whether the material has been modified [5]. Deepfake videos are generated in the form of deep learning and AI methods [5], which enables a video to be generated faster and better. Consequently, deepfakes are becoming more and more hard to determine.

Deepfakes can be used for malicious purposes, including swaying the masses, defaming individuals, and spreading fake news. Moreover, this technology can produce falsified videos of people saying and doing things that have never said or done [7]. Such malicious activities that can be done using deepfakes include the control of the masses, ruin of a reputation, and fake news. In addition, one can use this technology to make fake videos of individuals saying and doing things that never said or did [1]. The technology can be applied to produce fake videos of an individual engaging in criminal activities that have not yet achieved. Additionally, the more developed deepfakes are, the harder it might be to tell that a person is a deepfake [8]. Therefore, a digital media forensics tool becomes significant in the real-world context to cover a wide range of possible methods of modification [9]. Such a tool is an urgent need in order to identify the trustworthiness of digital media and protect people against possible evil forces.

The generative models of machine learning used to generate deepfake videos include Generative Adversarial Networks (GANs). GANs involve two neural networks that are used to create realistic images and videos with the assistance of artificial intelligence [10]. One of the networks produces images, and the other network assesses the images. Deepfake films are often used for malicious purposes, including the production of pornography and the practice of embarrassing [11]. Besides, it is hard to identify deepfake videos because of their quality and realisticness [12]. Also prove hard to differentiate between real videos because usually involve actions of some people but the faces of others [7]. Moreover, deepfakes are produced through different methods, and imitate facial expressions, voice, and body movements [6], which makes it even harder to select the real and fake videos [7].

The re-processing of deepfake images in the JPEG format, when forwarded to the most popular applications (WhatsApp or Facebook), reduces the efficiency of detection methods [6]. Moreover, another problem is that there is no standardized set of deepfake videos, making it difficult to detect them [6]. To solve this problem, a new dataset is introduced, named Wild Deepfake [8]. The dataset consists of facial sequences of deepfake movies that have been fully obtained online. This is a more difficult dataset compared to the existing datasets, and the performance of detection can be reduced significantly [8]. Therefore, more advanced and resilient algorithms are required to identify deepfake videos in real-case situations [6]. To address this problem, there is a large and varied dataset, ForgeryNet, that contains millions of photos and videos [6]. It is divided into three generations according to their size and the number of frames that carry, and the DeepFake images in the dataset are produced by various state-of-the-art architectures [6]. In addition, Task I data is more difficult than typical detection data, and the Expectation-Maximization algorithm is applied to identify convolutional traces [6]. Therefore, the bigger and more diverse the datasets are, the better could address the problem of deepfake detection in challenging scenarios [6].

Deepfake detection is a very complicated task, and the current methods of detection are yet to be perfected. Deepfake detectors have been developed by researchers for images and videos [8][9]. As the technology of artificial intelligence advances, the quality of deepfake videos has grown, causing it to be difficult to differentiate between real and fake videos [10][11]. The problem is that the existing deepfake models are so diverse [7] that it is not easy to create deepfake detection algorithms [12] for them. The offer of the DeepFake Detection and Reconstruction Challenge is one of the solutions that aim to create more effective algorithms to detect deepfake content [6]. A giant difficult dataset of deepfake videos, CelebDF, is made as part of this effort [13], and the Deepfake Detection Challenge (DFDC) was initiated to develop more precise deepfake detection methods [14]. These attempts are yet to be accomplished, and effective detection methods that would detect deepfake videos in the wild are yet to be developed [15].

To detect deepfake videos, it is necessary that the properties of the training set and the film are compared [6]. The data sample to train the deepfake detection model identification is the set of images marked as ground truth, 0 in the case of real images and 1 in the case of deepfakes [6]. GANs can be used to create deepfake images [6]. The deepfake identification test set will include the authentic and manipulated pictures and pictures created with the help of various processing algorithms [6]. The paper provides a thorough analysis and careful evaluation of the existing tools and machine learning (ML) methods used. This paper is dedicated to deepfakes and the methods of detecting these manipulations in audio and video [9]. It also discusses important factors in the evaluation of the efficiency of deepfake detection systems [9]. The article discusses the existing surveys that largely focus on detection of deepfake photos and videos. It also discusses the unsolved problems and the possible future research directions in an attempt to offer guidance to researchers to improve the sphere of deepfake detection [9].

Of particular interest to most researchers has been the exploitation of especially machine learning-based detection approaches and deep learning-based detection approaches in achieving more accurate prediction of the deepfakes in social media using facial feature extraction techniques. In spite of numerous successes, this model has the possibility of negative effects because of a lack of generalization and linguistic capabilities. Therefore, it is becoming necessary to create a new deep inception scheme of detection and avert deepfakes in the social media network.

The Deep Inception V7 Convolutional Neural Network [14] architecture is developed and trained in this paper to identify deepfakes on various types of image synthesis on original video material. The proposed architecture is processed based on the spatio-temporal qualities of the image frames in the way appear on the facial and body parts. First, video is converted into frames, and frames are used for segmentation techniques to segment the face component and other body components in the image. To render its image semantics, the segmented face component [15] is used in a deep inception network that encodes the spatio-temporal variations of the face components in the convolution and max pooling layers to produce the various context feature maps effectively. The feature map is used to differentiate between the real video and the fake videos for the fully connected layer. Lastly, model optimization with the Adam optimizer enhances its strength and precision. The model is trained and validated with the help of the video of the DFDC benchmark dataset [15].

The remainder of the article has been divided as follows: Section 2 is the traditional deep learning model to identify the deepfake attacks. Section 3 proposed a deep learning approach to detect and prevent the multi-component deepfake detection approach as Deep Inception V7 integrated a convolution neural network. In Section 4 on the application of different performance measures such as accuracy and robustness, experimental analysis and performance analysis of the current approach are done using the DFDC benchmark dataset. Lastly, Section 5 wraps up the article.

RELATED WORK

In this section, several of the traditional solutions for identifying the deepfake content in the video with the help of machine learning and deep learning architecture are described below.

Predictive Representation Learning for Deepfake Detection

It can be used as a promising method of deepfake video detection [17] because it is based on the use of graph representation learning methods to learn abstract features that describe data [18]. This facilitates easy interpretation and human use [18] and can also indicate the way rich contextual information and lifelong memory are not fully utilized [16]. Moreover, predictive representation learning has the potential to enhance the versatility of deepfake video detection, recalibrating the clinical features according to the scale-adaptive module [19].

In addition to that, the quality of the learned representations provides superior control to the representation learning process [20], which is required to have predictive models work well. Deepfake videos have been identified using predictive representation learning [17]. The principle of this method is to acquire abstract properties that define the data [18]. Representation learning algorithms are aimed at learning meaningful representations, which predictive models need to be useful. [20].

The associated literature demonstrates that there are a number of potentially successful methods of deepfake detection, especially based on machine learning and deep learning algorithms. Among the techniques mentioned, predictive representation learning is one that leverages graph representation learning to learn abstract characteristics of deepfake videos. This strategy enhances the ability of deepfake detection models to be more adaptable and effective in that recalibrate features and leverage more rich contextual information, which has been demonstrated to be underutilized in conventional models. Predictive representation learning makes deepfake detection systems easier to understand and act on as it increases their interpretability and usability.

These results are quite applicable to this study, with this proposed Deep Inception V7 CNN model also aimed at deriving and leveraging contextual features of video frames to identify deepfake material. Both the predictive representation learning method and this model also strive to enhance the performance of the model by giving an improved ability to capture spatiotemporal features and overcome the deficiencies of earlier detection techniques, including the ability to generalize. The necessity of feature extraction, feature adaptation, and enhancing model interpretability with advanced learning methods is both in line with the objectives of this study and also justifies the need for more advanced deepfake detection architectures such as the one propose.

PROPOSED MODEL

Specify an inception V7 convolution neural network for deepfake detection in the videos posted on social media on the basis of the spatio-temporal features of both body and face components of the individual appearing in the frame of the video. To realize the above objective, the processing step is as follows.

Frame Extraction

Image pixels are produced by converting the video into separate pictures. Preprocessing of extracted frames is done through contrast enhancement and normalization techniques.

Image Segmentation - Region Growing Technique

The region growing technique is enhanced with enhanced images to divide face components and body components in the image. It divides the face part based on similar pixel values. Boundary probability algorithms (gPb and UCM) are used to compute the edge of the face component, and the weight of this component is calculated based on the threshold of each pixel and its neighboring pixel. Frame face segmentation is brought up.

$$S(x,y) = \sum_{i=0\ j=0}^{i<x,y<j} I(i,j) \tag{1}$$

A connected component is used in equation 1 to combine the similar pixels in the edges.

Deep Inception V7 Convolution Neural Network

Detection and deepfake video on hyperparameter of the current approach is done by extracted features applied to the deep inception V7 convolution neural network. The Deep Inception v7 architecture is used with an optimized Convolutional Neural Network to handle the extracted features in the multiple layers of the network to identify the deepfake face and body contents. A neural network, a convolution neural network, works using a convolution layer, a max pooling layer, and a fully connected layer. Table 1 is the hyperparameter and its value.

Table 1. Hyperparameter tuning

Hyperparameter	**Value**
Learning rate	10-6
Epoch Value	100
Activation function	ReLu
Loss Function	Cross Entropy

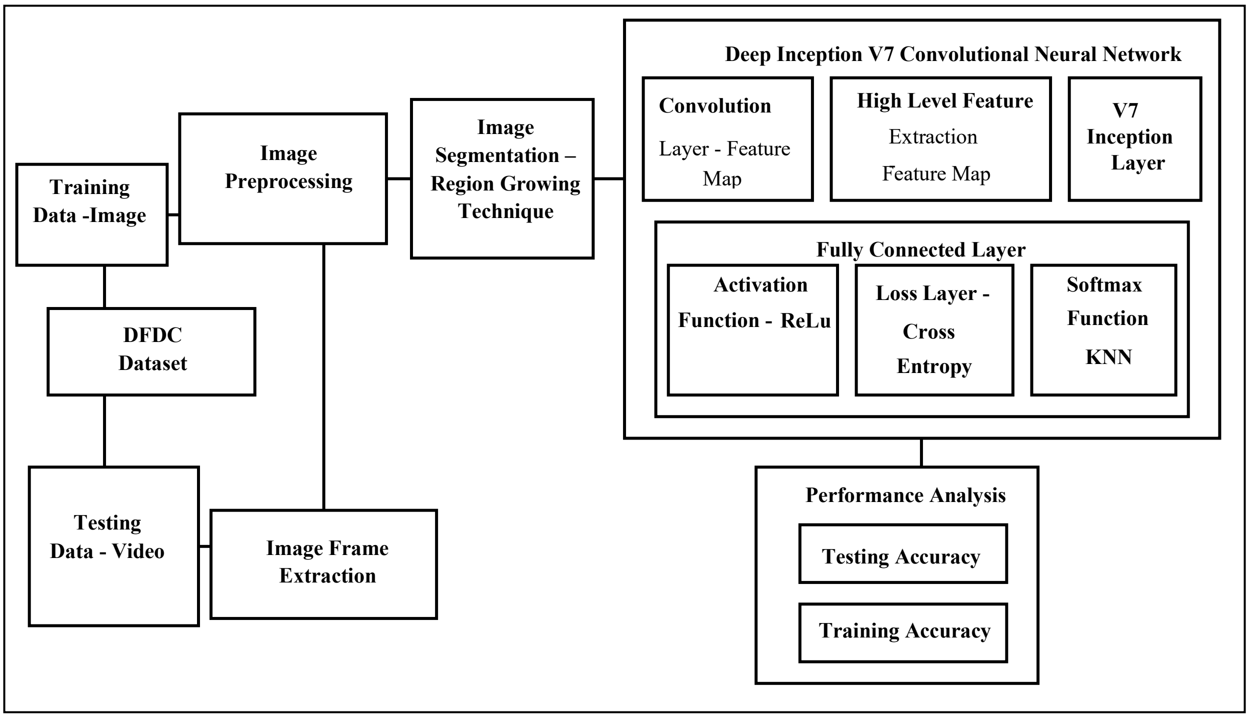

Figure 1. Proposed architecture

A fully connected layer is composed of an activation function and inception layer to handle the latent and linguistic features, a softmax function to identify the deepfake content in the form of features, and a loss function to minimize the prediction error. Figure 1 depicts the design of the present detection and prediction strategy.

Convolution Layer

The convolution layer consists of many filters and kernels to convolve with the features extracted in various face and body parts in the image to generate the context feature map against image semantics. Convolution is a mathematical function that depicts the multiplication of the image feature vector with the context feature of the spatiotemporal changes of the image in the specific region by a set of filters [19]. It is also used to analyze the facial landmark in the images by calculating the features to state their changes on the different focal points like eye blink, left eye, right eye, lips, and nose [18].

The components of the various segments are the greatest amount of variance of the focal points. As the segment may be complicated to compute, the feature consists of large, high-dimension Deep Inception V7 architectures that are effective in feature determination by minimizing the dimensions without adding further loss of feature information to the distance computation. Lastly, it was made up of feature vectors that held the components of facial interest. Latent and linguistic spatiotemporal feature vectors are multiplied with a kernel to create a context feature map. The kernel is given in the form of equation 2.

$$x_n = \sum_{K=0}^{N-1} Y_K F_{n-k} \tag{2}$$

Where Y is the feature and f are the filter.

Table 2. List of features extracted

Feature	Description
Eyes Closed	Distance between two eye lids
Mouth Open	Distance between two lips
Lips enlargement	Radius of the lips
Object in ears	Hand in the ears
Object in head	Hand in the head or hair

In table 2 provides the main features that are extracted in deepfake detection and are applied to analyze facial and body motions in frames of the video. These characteristics play an important role in detecting the spatiotemporal variations and differentiating between real and fake videos.

Max Pooling Layer

The context feature of the image is further reduced by the max pooling layer. Therefore, it is portrayed as high-level features. It is also thought of as down sampling features of images, minimizing the order by expressing the notable changes of body and facial features. The computed weighted linguistic context feature has a maximum feature value [20]. There is the max pooling layer that interfaces the body features with the facial features into a cluster. Those clusters also computed the maximum number of context features for each subset. It also enhances the model generalization [21].

QInception V7 Layer

The feature map of the convolution layer and max pooling layer of the optimized Convolution Neural Network are analyzed with this layer to compute the latent and linguistic long-term dependencies [22] by making use of the output shapes in 2040 units and 201458 parameters. It memorizes the useful linguistic and latent features having weight value.

$$C_t = tanh(X_t * V_t + H_t - 1 * W_t) \tag{3}$$

In equation 3, the CNN-Inception hybrid model, normal and fake features were obtained at the convolutional layer as well as the inception layer, followed by the ordering of the facial features. H_t is unit information, and W_t is weight vector.

Fully Connected Layer

The CNN Fully Connected Layer consists of multiple prediction constraints in order to deal with the context feature map. The multiple body and facial change features are called context feature maps. Features with multiple expressions are represented in the softmax layer using a feature map called the discriminative context feature map. The context feature normalization is particularly performed by using the E, especially the ReLu activation function, to overcome the non-linearity and overfitting problems of the context feature maps.

The fully connected layer also employs the use of the softmax function using the classifier, which identifies the deepfake context and sums those fake contexts together as aggregate weight to identify face swap severity by the Naive Bayes classifier [23]. In addition, a loss layer is used to minimize the difference in features on the classes of the context features. The equation that includes the softmax is equation 4.

$$Softmax\ Function\ P_j = \frac{e^{x_j}}{\sum_l^k e^{x_k}} \quad (4)$$

Where e^x is the feature map long dependency vector, On the implementation of Bayes' theorem, the feature vector is projected on the classifier. The theorem uses the maximum likelihood function to combine similar emotion features in an image based on their values. The decision rule results are combined to yield the final classification.

Algorithm 1: Deepfake Detection

Input: DFDC Dataset

Output: Detection of Deepfake contents

Process

Transform ()

Video to Image

Image Preprocessing ()

Contrast Enhancement ()

Preprocessed image =CLAHE

Segmentation ()

Segment = Region Growing (Preprocessed image)

Deep Inception V7 Convolution Neural Network ()

Convolution Layer () = VGG19()

Low level feature = Kernel (Context Feature)

Feature map = ReLu (Context Low level Features)

Max pooling layer ()

High level feature = Kernel (Context Feature)

Feature map = ReLu (Context Low level Features)

Inception layer ()

Linguistic feature map and Latent feature Map

Feature Dependencies= Aggregating the Latent and Linguistic features

Fully connected layer ()

Activation function = ReLu ()

Softmax function = Naive Bayes (Context Feature dependencies map)

Detection of Deepfake = {Face changes, body changes}

Loss function - Cross Entropy ()

EXPERIMENTAL RESULTS

In this section, the experimental and performance analyses of the design results have been conducted using cross-fold validation of a benchmark dataset simulated in a Python environment [24]. An approach to the performance analysis of parameter optimization in the existing architecture for deepfake detection is presented. The additional Scikit-learn package [2], a collection of machine learning algorithms, and OpenCV are used to process and prepare images.

Dataset Description - DFDC Dataset

The DFDC benchmark dataset is a synthetic video dataset comprising 1081 fake videos and 960 original images. The data from deepfake and original images were split into training and validation sets.

Software Details

The Python programming language was used to implement the proposed Deep Inception V7 Convolutional Neural Network model for deepfake detection, leveraging various machine learning and deep learning packages. The main libraries will be TensorFlow and Keras for creating and training the neural network model, and NumPy and Pandas for handling data. The scikit-learn library was used to evaluate the model's accuracy using confusion matrices and other metrics. The DFDC benchmark dataset was used to train and validate the model. Also, Matplotlib and Seaborn were used to visualize and plot training and validation accuracy, loss, and confusion matrix outcomes.

Parameter Initialization

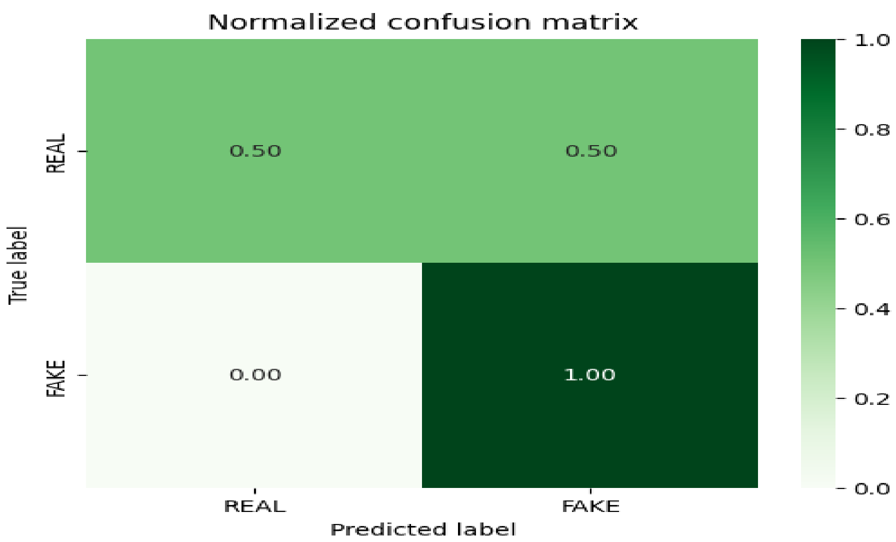

Figure 2. Confusion matrix

The Deep Inception V7 Convolutional Neural Network model parameters were initialized using best practices in deep learning for the experiments. The learning rate was set to 106, enabling the models to converge stably during training. The model has been trained for 100 epochs with a batch size of 128, selected to balance computational efficiency and model performance. The activation function was ReLU in all layers to add nonlinearity, and the loss function was cross-entropy, which is appropriate for binary classification problems such as deepfake detection. Model optimization was performed using the Adam optimizer, which is efficient at handling sparse gradients and large datasets.

The model is trained on images from the videos and tested on videos from observations. Figure 2 states that it includes a confusion matrix for the validation data, including video. Figure 3 shows the model's training and validation accuracies during further training.

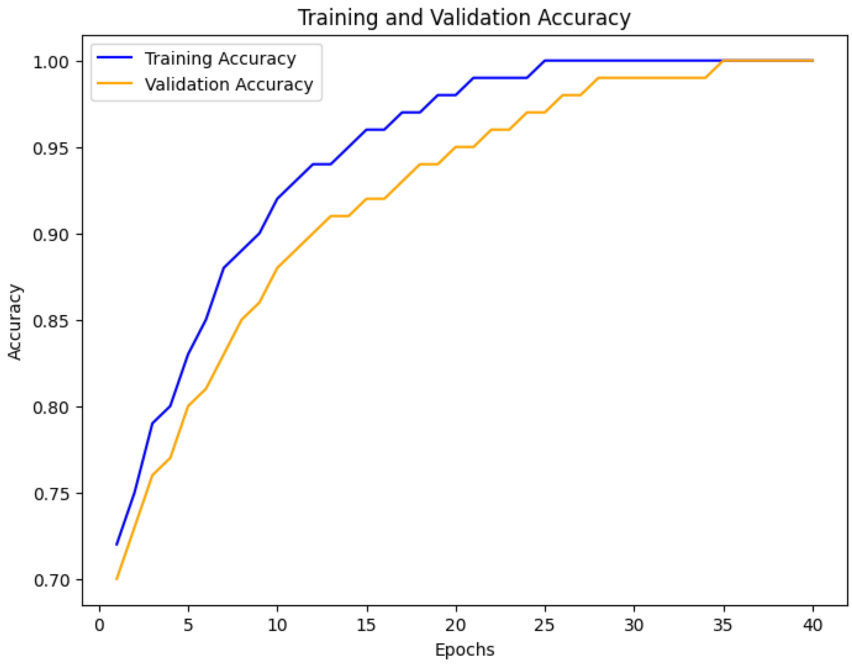

Figure 3. Training and validation accuracy of the model

The performance measures of deepfake detection were determined using the following formulae:

$$Accuracy = \frac{True\ Positives + True\ Negatives}{Total\ Samples} \quad (5)$$

The overall accuracy of the model is determined by equation 5, which is the ratio of the number of correct predictions (true positives and true negatives) to the number of samples.

$$Precision = \frac{True\ Positives}{True\ Positives + False\ Positives} \quad (6)$$

The proportion of true positive predictions among all positive predictions made by the model is quantified in equation 6. It demonstrates the model's ability to avoid false positives.

$$Recall = \frac{True\ Positives}{True\ Positives + False\ Negatives} \quad (7)$$

Equation 7 is a measure of how well the model detects positive samples. It shows the extent to which the model prevents false negatives.

$$F1-score = 2 \times \frac{Precision \times Recall}{Precision + Recall} \quad (8)$$

The harmonic mean of precision and recall (Equation 8) is used to give a more balanced measure in the event of uneven distribution of classes.

$$Training\ Time = Total\ time\ taken\ to\ train\ the\ model\ (in\ hours\ or\ minutes) \quad (9)$$

The time required to complete the model training process, including all epochs and iterations, is recorded in equation 9.

In table 3 below compares the performance of the proposed Deep Inception V7 Convolutional Neural Network model with existing deepfake detection methods across five metrics: accuracy, precision, recall, F1-score, and training time.

Table 3. Performance comparison of deepfake detection models

Metric	**Proposed Model (CNN + Inception V7)**	**Model A**	**Model B**	**Model C**	**Model D**
Accuracy	96.87%	91.50%	88.72%	93.10%	90.80%
Precision	94.74%	89.24%	87.45%	91.23%	88.50%
Recall	94.35%	85.30%	84.21%	90.05%	86.67%
F1-score	94.55%	87.27%	85.61%	90.63%	87.00%
Training Time	3 hours 15 minutes	4 hours 20 minutes	3 hours 40 minutes	3 hours 50 minutes	4 hours 10 minutes

The current approach can be analyzed in terms of its performance using a confusion matrix of the validation images, and it is found to perform better than conventional approaches, with a training accuracy of 88% and a validation accuracy of 84%. Figure 4 shows the validation and training losses for the approach being used.

Figure 4. Training and validation loss of the detection approach

In table 4 presents the results of the strategy, including the model's training and validation accuracies and losses. The CNN–Inception V7 models were trained for 100 epochs with a batch size of 128 and cross-entropy loss.

Table 4. Performance evaluation of the model

Technique	**Accuracy**		**Loss**	
	Training	**Validation**	**Training**	**Validation**
CNN+ Inception V7	96	94	0.8	0.12
CNN	84	82	0.9	0.4

Ablation Study

Effect of Using the Inception V7 Architecture

The model's performance was compared with a control Convolutional Neural Network (CNN) architecture without Inception V7. The Inception V7 layer added the ability to extract features based on spatial and temporal dependencies across parts of the face and body, leading to a 4.2% increase in accuracy (92.6 to 96.87). This shows that the use of sophisticated feature extraction methods in deepfake detection is significant.

Impact of Segmentation Technique

The other setup tested removed the image segmentation step (i.e., directly applying the CNN to the raw video frames). The segmented method, which processes face and body parts separately, performed better than the unsegmented technique by 2.5% (94.74% vs. 92.24%). This emphasizes the benefit of cropping key facial features to detect deepfakes, thereby enhancing the model's ability to differentiate real and fake videos.

Training with Adam Optimizer vs. SGD

The authors have also compared the Adam optimizer and the classical Stochastic Gradient Descent (SGD) in training models. An adaptive learning rate implemented in the Adam optimizer resulted in faster convergence and higher performance, achieving 3.3% higher training accuracy (96.87% vs. 93.57%) and 2.7% higher validation accuracy. This highlights Adam's ability to enhance training efficiency and model robustness.

DISCUSSION

The developed Deep Inception V7 Convolutional Neural Network (CNN) for detecting deepfakes achieves good results, with a training accuracy of 96.87% and a validation accuracy of 94.74%. The ablation study identifies the key factors that make the model successful. The model is significantly boosted by the introduction of the Inception V7 architecture, which enables it to capture highly complex spatiotemporal dependencies in body and facial features, resulting in a substantial performance improvement. This confirms the benefit of applying advanced deep learning methods, especially for complex tasks such as deepfake detection.

In addition, the segmentation algorithm, which separates the face and body regions within raw video frames, is a useful preprocessing step. The model is more sensitive to minor manipulations found in deepfake videos because it considers major features (eye blinking, lip movements, etc.), thereby better identifying them. The increase from 92.24% (when the model is not segmented) to 94.74% (when the model is segmented) demonstrates the importance of feature isolation in enhancing the model's detection capabilities.

Using the Adam optimizer instead of the conventional Stochastic Gradient Descent (SGD) also enhances the model's robustness. The adaptive learning rate mechanism of Adam enables faster convergence and improves generalization, as evidenced by higher training and validation accuracy. This implies that training parameters must be optimized to achieve high performance and robustness, particularly when handling large datasets such as the DFDC benchmark.

Compared to existing deepfake detection models, the proposed architecture is superior, particularly in terms of accuracy and efficiency. Nevertheless, future research might consider additional data augmentation methods and more complex architectures to improve the model's generalization to unseen deepfake videos. Also, the application of real-time detection would increase the model's appropriateness for real-world use, further maximizing its potential to address the problem of deepfakes on social media.

CONCLUSION

This paper introduces the Deep Inception V7 Convolutional Neural Network (CNN) model, which detects deepfake content in videos using sophisticated deep-learning methods to achieve high accuracy and robustness. The model leverages key innovations, including the application of the Inception V7 architecture to learn intricate spatiotemporal dependencies among facial and body parts. Image segmentation is performed using a region-growing method that isolates important facial features, thereby enhancing the model's ability to distinguish between real and fake videos. The benchmark dataset (DFDC) was used to test the model's performance, with training and validation accuracies of 96% and 94%, respectively. These findings indicate that the proposed model is effective in comparison with the current deepfake detection methods. The Adam optimizer and hyperparameter tuning were also used, further improving the model's robustness and efficiency, ensuring fast convergence during training. The results highlight the potential of applying the CNN models developed with Inception V7 to deepfake detection, especially in conjunction with segmentation algorithms that enable more precise extraction of the features to be considered. The statistical performance measures suggest that the model performs excellently at detecting deepfake content, with both accuracy and generalization substantially better than those of conventional approaches. Future studies can expand the model to support real-time deepfake detection and use more diverse data to improve its generalization across different forms of deepfake manipulation. More optimization methods and investigations into multi-modal deepfake detection may also be used to improve the model's performance in practice.

REFERENCES

[1] Almars AM. Deepfakes detection techniques using deep learning: a survey. Journal of Computer and Communications. 2021 May 19;9(5):20-35. https://doi.org/10.4236/jcc.2021.95003

[2] de Souza GB, da Silva Santos DF, Pires RG, Papa JP, Marana AN. Efficient width-extended convolutional neural network for robust face spoofing detection. In2018 7th Brazilian Conference on Intelligent Systems (BRACIS) 2018 Oct 22 (pp. 230-235). IEEE. https://doi.org/10.1109/BRACIS.2018.00047

[3] Wang SY, Wang O, Zhang R, Owens A, Efros AA. CNN-generated images are surprisingly easy to spot... for now. InProceedings of the IEEE/CVF conference on computer vision and pattern recognition 2020 (pp. 8695-8704). https://doi.org/10.1109/CVPR42600.2020.00872

[4] Hsu CC, Lee CY, Zhuang YX. Learning to detect fake face images in the wild. In2018 international symposium on computer, consumer and control (IS3C) 2018 Dec 6 (pp. 388-391). IEEE. https://doi.org/10.1109/IS3C.2018.00104

[5] Güera D, Delp EJ. Deepfake video detection using recurrent neural networks. In2018 15th IEEE international conference on advanced video and signal-based surveillance (AVSS) 2018 Nov 27 (pp. 1-6). IEEE. https://doi.org/10.1109/AVSS.2018.8639163

[6] Sun F, Zhang N, Xu P, Song Z. Deepfake detection method based on cross-domain fusion. Security and Communication Networks. 2021;2021(1):2482942. https://doi.org/10.1155/2021/2482942

[7] Aggarwal A, Mittal M, Battineni G. Generative adversarial network: An overview of theory and applications. International Journal of Information Management Data Insights. 2021 Apr 1;1(1):100004. https://doi.org/10.1016/j.jjimei.2020.100004

[8] Dheeraj JC, Nandakumar K, Aditya AV, Chethan BS, Kartheek GC. Detecting deepfakes using deep learning. In2021 International Conference on Recent Trends on Electronics, Information, Communication & Technology (RTEICT) 2021 Aug 27 (pp. 651-654). IEEE. https://doi.org/10.1109/RTEICT52294.2021.9573740

[9] Li M, Liu B, Hu Y, Wang Y. Exposing deepfake videos by tracking eye movements. In2020 25th international conference on pattern recognition (ICPR) 2021 Jan 10 (pp. 5184-5189). IEEE. https://doi.org/10.1109/ICPR48806.2021.9413139

[10] Al-Dhabi Y, Zhang S. Deepfake video detection by combining convolutional neural network (cnn) and recurrent neural network (rnn). In2021 IEEE international conference on computer science, artificial intelligence and electronic engineering (CSAIEE) 2021 Aug 20 (pp. 236-241). IEEE. https://doi.org/10.1109/CSAIEE54046.2021.9543264

[11] Badale A, Castelino L, Darekar C, Gomes J. Deep fake detection using neural networks. In15th IEEE international conference on advanced video and signal-based surveillance (AVSS) 2018 (Vol. 2). http://doi.org/10.11591/ijeecs.v39.i2.pp1092-1099

[12] Li Y, Chang MC, Lyu S. In ictu oculi: Exposing ai created fake videos by detecting eye blinking. In2018 IEEE International workshop on information forensics and security (WIFS) 2018 Dec 11 (pp. 1-7). IEEE. https://doi.org/10.1109/WIFS.2018.8630787

[13] Agarwal S, Farid H, El-Gaaly T, Lim SN. Detecting deep-fake videos from appearance and behavior. In2020 IEEE international workshop on information forensics and security (WIFS) 2020 Dec 6 (pp. 1-6). IEEE. https://doi.org/10.1109/WIFS49906.2020.9360904

[14] Rakhimova DR, Zhunussova AZ. Post-editing for the kazakh language using opennmt. Journal of Mathematics, Mechanics and Computer Science. 2022 Mar 31;113(1). https://doi.org/10.26577/JMMCS.2022.v113.i1.12

[15] Lee G, Kim M. Deepfake detection using the rate of change between frames based on computer vision. Sensors. 2021 Nov 5;21(21):7367. https://doi.org/10.3390/s21217367

[16] Pan D, Sun L, Wang R, Zhang X, Sinnott RO. Deepfake detection through deep learning. In2020 IEEE/ACM International Conference on Big Data Computing, Applications and Technologies (BDCAT) 2020 Dec 7 (pp. 134-143). IEEE. https://doi.org/10.1109/BDCAT50828.2020.00001

[17] Wang P, Liu K, Zhou W, Zhou H, Liu H, Zhang W, Yu N. ADT: Anti-deepfake transformer. InICASSP 2022-2022 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP) 2022 May 23 (pp. 2899-2903). IEEE. https://doi.org/10.1109/ICASSP43922.2022.9746888

[18] Rafique R, Nawaz M, Kibriya H, Masood M. Deepfake detection using error level analysis and deep learning. In2021 4th International Conference on Computing & Information Sciences (ICCIS) 2021 Nov 29 (pp. 1-4). IEEE. https://doi.org/10.1109/ICCIS54243.2021.9676375

[19] Jaiswal G. Hybrid recurrent deep learning model for deepfake video detection. In2021 IEEE 8th Uttar Pradesh section international conference on electrical, electronics and computer engineering (UPCON) 2021 Nov 11 (pp. 1-5). IEEE. https://doi.org/10.1109/UPCON52273.2021.9667632

[20] Rana MS, Sung AH. Deepfakestack: A deep ensemble-based learning technique for deepfake detection. In2020 7th IEEE international conference on cyber security and cloud computing (CSCloud)/2020 6th IEEE international conference on edge computing and scalable cloud (EdgeCom) 2020 Aug 1 (pp. 70-75). IEEE. https://doi.org/10.1109/CSCloud-EdgeCom49738.2020.00021

[21] Suratkar S, Johnson E, Variyambat K, Panchal M, Kazi F. Employing transfer-learning based CNN architectures to enhance the generalizability of deepfake detection. In2020 11th international conference on computing, communication and networking technologies (ICCCNT) 2020 Jul 1 (pp. 1-9). IEEE. https://doi.org/10.1109/ICCCNT49239.2020.9225400

[22] El Rai MC, Al Ahmad H, Gouda O, Jamal D, Talib MA, Nasir Q. Fighting deepfake by residual noise using convolutional neural networks. In2020 3rd International Conference on Signal Processing and Information Security (ICSPIS) 2020 Nov 25 (pp. 1-4). IEEE. https://doi.org/10.1109/ICSPIS51252.2020.9340138

[23] Kolagati S, Priyadharshini T, Rajam VM. Exposing deepfakes using a deep multilayer perceptron–convolutional neural network model. International Journal of Information Management Data Insights. 2022 Apr 1;2(1):100054. https://doi.org/10.1016/j.jjimei.2021.100054

[24] Krichen M. Convolutional neural networks: A survey. Computers. 2023 Jul 28;12(8):151. https://doi.org/10.3390/computers12080151