# Integrating Deep Learning Architecture with Pufferfish Optimization Algorithm for Real-Time Deepfake Video Detection and Classification Model

**Sameer Nooh**[1,*]

[1] Information Systems Department, Faculty of Computing and Information Technology, King Abdulaziz University, Jeddah 21589, Saudi Arabia

Emails: snooh@kau.edu.sa

**Abstract**

Deepfake is a technology employed in making definite videos, which are operated utilizing an artificial intelligence (AI) model named deep learning (DL). Deepfake videos were normally videos that cover activities grabbed by definite people but with another individual's face. Substitute of people appearances in videos utilizing the DL model. The technology of Deepfake permits humans to operate videos and images utilizing DL. The outcomes from deepfakes are challenging to differentiate utilizing normal vision. It is a combination of the words DL and fake, and it mostly denotes material shaped by deep neural networks (DNNs), which is a subclass of machine learning (ML). Deepfake denotes numerous modifications of face models, and integrates innovative technologies, with computer vision and DL. The detection of a deepfake model can be assumed as a dual classification procedure that can be categorized as the original or deepfake class. It works by removing features from the videos or images that is employed to distinguish between original and deepfake content. Therefore, this study proposes Leveraging Pufferfish Optimization and Deep Belief Network for an Enhanced Deepfake Video Detection (LPODBN-EDVD) technique. The LPODBN-EDVD technique intends to detect fake videos utilizing the DL model. In the presented LPODBN-EDVD technique, the data preprocessing stages include splitting the video into frames, face detection, and face cropping. For the process of feature extraction, the EfficientNet model is exploited. Besides, the deep belief network (DBN) classifier can be executed for deepfake video detection. Finally, the pufferfish optimization algorithm (POA) is employed for the optimal hyperparameter selection of the DBN classifier. A wide range of simulations was involved in exhibiting the promising results of the LPODBN-EDVD method. The experimental analysis pointed out the enhanced performance of the LPODBN-EDVD technique compared to recent approaches

**Keywords:** Deepfake Video Detection; Deep Belief Network; Pufferfish Optimization Algorithm; EfficientNet; Deep Learning

## 1. Introduction

Technologies to alter videos, audio, and images are emerging fast. Techniques and technical knowledge to manipulate and create digital content are also readily available [1]. Now, it is promising to flawlessly produce hyperactive realistic digital images with a small number of resources and easy how-to-do instructions available on the internet. Deepfake is a method that intends to change the face of a targeted individual with the appearance of somebody else in a video [2]. It is made by merging the synthesized face area into the original image. The term additionally means to denote the last output of a hyper-realistic video generated [3]. Deepfakes are applied for the creation of hyper-realistic Virtual Reality (VR), Augmented Reality (AR), Computer Generated Imagery (CGI), Cinema, Animation, Arts, and Education. Nevertheless, since Deepfakes are pretended, they can also be employed for malicious uses [4]. Most of the people utilize deepfake technologies for negative intentions. In the past few years, social media has allowed people to rapidly connect verified multimedia content, inducing important developments in multimedia content output and availability [5]. Improved the speed with which erroneous and false information is manufactured and spread, identifying the truth and believing the information became more and more challenging, perchance resulting in disastrous repercussions [6].



DOI: https://doi.org/10.54216/FPA.180120
Received: July 18, 2024 Revised: October 22, 2024 Accepted: January 04, 2025

The growth of Deepfakes and the propagation of direct fake news discovered online links with nude content concentrated on celebrities [7]. The major aim beyond these videos is to spoil the status of the relevant individuals [8]. Cybercriminals carry on increasing their abilities, frequently using state-of-the-art algorithms to perform cyber-attacks like fraud, phishing, and hacking. New findings make it problematic to differentiate between fake and real. The amount of deepfakes carries on rising every day, and this occurs since the fast usage of open-source devices increases privacy fears and bullies' routines [9]. Then, there is a requirement to generate an efficient and reliable mechanism to prevent and detect possible loss to overcome the challenges produced due to this media. Most organizations, studies, and researchers have just concentrated on recognizing fake content [10]. Hence, a deepfake is material generated by DL that appears real in human eyes. The term deep fake is a combination of the words DL and fake, and it mostly relates to material generated by a deep neural network (DNN) that can be a subdivision of ML. This can be obtainable for some years owing to several user-friendly software products that permit pictures, video editing, and audio.

Therefore, this study proposes Leveraging Pufferfish Optimization and Deep Belief Network for an Enhanced Deepfake Video Detection (LPODBN-EDVD) technique. The LPODBN-EDVD technique intends to detect fake videos utilizing the DL model. In the presented LPODBN-EDVD technique, the data preprocessing stages include splitting the video into frames, face detection, and face cropping. For the process of feature extraction, the EfficientNet model is exploited. Besides, the DBN approach can be executed for deepfake video detection. Finally, the pufferfish optimization algorithm (POA) is utilized for the optimum parameter choice of the DBN classifier. A wide range of simulations was involved in exhibiting the promising results of the LPODBN-EDVD method.

## 2. Related Works

Javed et al. [11] introduce an advanced technique, which integrates the analysis of eye movement with a hybrid DL method to tackle the requirement for real time Deepfake recognition. The presented hybrid DL method combines 2 deep neural network (DNNs) structures, ResNet101 and MesoNet4, to optimize their relevant structure intensities for effectual Deepfake identification. MesoNet4 is a lightweight CNN method intended clearly to identify delicate operations in facial images. Simultaneously, ResNet101 manages robust feature extraction and intricate visual data. Uniting the limited feature learning of MesoNet4 with the deep, further widespread ResNet101 feature representations. Qadir et al. [12] present a hybrid method, which gathers input from videos of consecutive targeting frames after providing these frames to the ResNet-Swish-BiLSTM, an enhanced convolutional BiLSTM-based residual network for classification and training. To evaluate the strength of our presented method, the accessible Face Forensics deepfake collections (FF++) and deepfake detection challenge dataset (DFDC) are used. Talreja et al. [13] proposed a deepfake recognition method, containing picture analysis, ML methods, and scientific approaches. These methodologies' performance and efficiency have been evaluated based on their capability to precisely recognize and classify deepfakes. The study also inspects the problems and limitations of deepfake recognition, like the improvement of more composite and considerable deepfakes. Additionally, potential usages and upcoming probabilities for deepfake recognition study have been analyzed, with a focus on developing recognition skills and producing countermeasures that are more effectual.

Alhaji et al. [14] introduce an advanced technique to deepfake video recognition by incorporating features based on DL and ant colony optimizer-particle swarm optimizer (ACO-PSO) methods. The presented technique optimizes DL methods and ACO-PSO features to improve recognition robustness and accuracy. Then, these features have been utilized to train DL classifications to spontaneously differentiate between the true and deepfake videos. Bhat et al. [15] present an overall analysis of the approaches used for DeepFake recognition. It probes into the incorporation of various media forms (like speech, images, and videos) with ML to distinguish fake content. Furthermore, it deliberates the essential datasets used by scholars for assessing their DeepFake detection methods. This study investigated various approaches designed for detecting fake videos, images, and fake voices. They show that combining dissimilar methods, like incorporating videos and images or using various ML methods, could produce great efficient outcomes in DeepFake detection. Qiao et al. [16] presented to devise a completely unsupervised Deepfake detector. Especially, in the complete process of testing or training, the author does not know any information regarding the real sample labels. Initially, the newly designed pseudo-label creator was for identifying the training models, whereas the conventional handcrafted features were utilized to describe both kinds of instances. Next, the samples of training with the pseudo labels were served into the presented improved contrasting learner, in which the discriminating features were additionally removed and constantly developed by iterating on the direction of the contrasting loss.

Chen et al. [17] present a deepfake video recognition technique based on 3D spatio-temporal direction. In particular, the author uses a strong 3D method to create spatiotemporal motion features, combining the feature information from either 2D or 3D frames to alleviate the effect of insufficient lighting or enormous head rotation angles in frames. Additionally, facial expressions are separated from head movements and form a sequential study technique derived from stage space motion direction for exploring the feature variances among true and fake faces in deepfake videos. Omar et al. [18] present a DL bagging ensemble classifier to identify deployed faces in videos.



DOI: https://doi.org/10.54216/FPA.180120
Received: July 18, 2024 Revised: October 22, 2024 Accepted: January 04, 2025

The coAtNet method can be vertically stacked into self-attention layers and depth-wise convolution layers in this manner, in which efficiency, generalization, and capacity were enhanced.

## 3. Materials and Methods

In this article, we have proposed a novel LPODBN-EDVD technique. The LPODBN-EDVD technique intends to detect fake videos using the DL model. To accomplish that, the LPODBN-EDVD technique has four distinct stages involving different levels of preprocessing, feature extraction, classification using DBN, and POA-based parameter tuning methods. Fig. 1 depicts the workflow of the LPODBN-EDVD model.

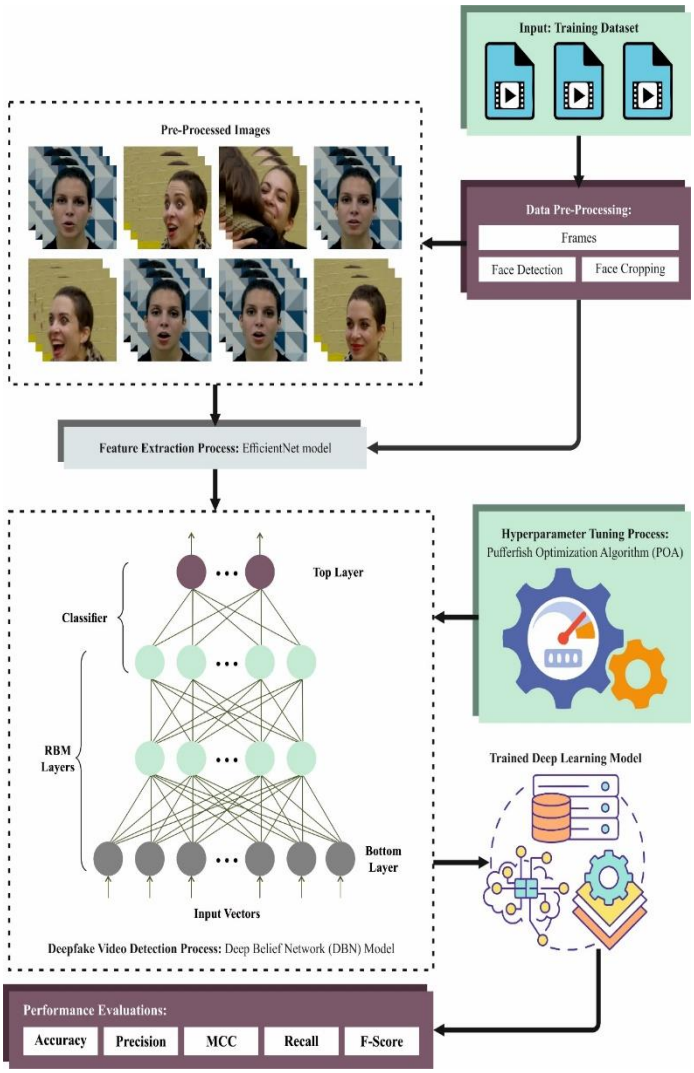


**Figure 1.** Workflow of LPODBN-EDVD technique

### A. Stage I: Preprocessing

Initially, the presented LPODBN-EDVD technique undergoes the data preprocessing stages including splitting the video into frames, face detection, and face cropping [19].

**Splitting Video into Frames:**

The initial stage in pre-processing for deepfake detection includes splitting the video into individual frames. This method changes the constant video stream into a sequence of static images, each representing a single instant in time. By examining frames individually, the model can inspect the visual content more carefully, permitting a detailed examination of all images. This segmentation is critical because deepfake manipulations frequently manifest in subtle frame-wise inconsistencies or distortions that might not be obvious. After viewing the video completely. By isolating each frame, the detection approach can better recognize these anomalies, resulting in more precise detection of manipulated content.



DOI: https://doi.org/10.54216/FPA.180120
Received: July 18, 2024 Revised: October 22, 2024 Accepted: January 04, 2025

**Face Detection Using Multitask Convolutional Neural Network:**

Once the video is decomposed into frames, the following pre-processing stage is face detection by using MTCNN. The MTCNN is a face detection structure depending on cascaded CNN. It employs multitask learning to identify face frames and at the same time find 5 main facial points [20]. PNet is a smaller, complete CNN. The backbone contains 3 convolutions and 1 pooling layer. It rapidly extracts the face candidate areas from the multiscale dense sliding box of the image pyramid and implements rough regression. RNet is a somewhat large network. The backbone involves 3 convolutions, 2 pools, and 1 full connection. It overwhelms various false detection examples made by PNet utilizing better classifications and tuning the face candidate boxes. ONet is one of the biggest networks. The backbone comprises 4 convolutions, 3 pools, and 1 full connection. It applies a stronger algorithm to perform the candidate boxes. It also removes a smaller amount of incorrect candidate models and detects faces. The frame and 5 main facial points have been precisely reverted. Amongst the 3 detections, non-maximum destruction can be utilized for merging the candidate frames with more significant intersections to avoid continual recognitions. The MTCNN designs smaller, larger, and medium CNNs in the medium, sparse, and dense recognition phases. This aids in balancing the computational overhead of every phase. It additionally utilizes a cascading framework that incorporates numerous samples to be rapidly removed and main samples to be slowly improved. This theory presents highly efficient balance and at present the most effective face detector.

In a real-time face detection and investigation method, it is essential to condense the face alignment, face recognition, face detection, facial critical point positioning, age, and gender recognition models into a short time. While MTCNN has specific efficacy levels, the detection time continued to occupy mostly. Consequently, it is required to optimize the MTCNN detection time. According to the time statistics consumed in every detection phase of MTCNN, the time consumed in the PNet phase explains above 60% of the detection time. Hence, enhancing PNet can be important to accelerating MTCNN. Though the prediction cost of PNet for a solitary sample is lower than that of ONet and RNet, after being confronted with the dense multiple-scale sliding box in the image pyramid, too many prediction samples produce the complete speed of PNet to be lower. Enhancing the PNet detection speed mostly depends on a lightweight backbone network and image pyramid sparseness.

**Face Cropping:**

After detecting the faces in each frame, the following pre-processing stage is face cropping. This method includes isolating and extracting the facial areas recognized by the MTCNN from the remaining image. By cropping out the faces, the method removes unrelated background information and decreases the computational complexities of more processing phases. This concentrated model improves the detection system's effectiveness by concentrating analysis and resources on the facial features, which are most likely to show manipulation signs. Face cropping is crucial to improve the performance of deepfake detection algorithms, as it guarantees that the following feature extraction and classification techniques are performed on the more relevant data.

### B. Stage II: Feature Extraction

For the process of feature extraction, the EfficientNet model is exploited. Convolutional Neural Networks (CNNs) operate recurrent learning models [21]. During the training stage, a CNN is assigned at random weights to each of the features finding the outcomes and matching them using the real outcomes to compute the error. According to the error, it adjusts or reassigns the weights of each feature. This can continuously gain better accuracy, owing to these features; people have understood how to improve the layer counts in the CNN to obtain greater accurateness. Dropout is applied at every layer for training various neurons on dissimilar aspects. When each of the neurons is trained on a similar feature, they will each offer an equal outcome for a hidden image. This decreases the model's accuracy. To avoid these conditions, the dropout is adjusted as needed. By increasing the layer counts, accuracy can be enhanced to a greater level. Numerous tests have been performed to discover the saturation stage of the layer counts of a CNN. By improving the layer counts, we enlarge the depth and, thus, the complexities of the model. Since the complexities improve, the time required for training the method too improves. This results in the necessity of higher computing resources such as memory, SSD, GPU, and so on.

To decrease the complexities and the requirement for computing resources, investigators discovered a solution regarding the horizontal scaling of the algorithm. They were generated by scaling up the CNN on all 3 features of resolution, depth, and width. This algorithm was called the EfficientNet model. They decided that though it is crucially important to balance each 3 features (resolution, width, and depth) in a specific algorithm when we scale up our method on all 3 features while preserving balance amongst all 3 features, we can gain higher accuracy with lower computational resources. These methods are 8.4×smaller and 6.1×faster impact on it than the greatest recent CNNs.

They demonstrated their higher accuracy against different other recent methods. Over a wide study, they resultant an optimum equation with the succeeding coefficients to retain a balance amongst all 3 features. This represents that to scale up the CNN algorithm, the layer's depth should rise by 20%, resolution by 15%, and width by 10% to



DOI: https://doi.org/10.54216/FPA.180120
Received: July 18, 2024 Revised: October 22, 2024 Accepted: January 04, 2025

attain the maximum efficacy likely while increasing the application and increasing the precision of the CNN method. The new output layer includes 1000 outputs. They add a linear fully connected (FC) layer using the activation function of ReLU to acquire the dual output.

The efficientnet model utilizes the technique of complex scaling that employs a composite co-efficient $\varphi$ to scale up each of the parameters such as network depth, resolution, and width equally in an upright manner:

$$Depth: d = \alpha^{\varphi} \tag{1}$$

$$Width: w = \beta^{\varphi} \tag{2}$$

$$Resolution: r = \gamma^{\varphi} \tag{3}$$

$$s.t. \alpha * \beta^2 * \gamma^2 \approx 2$$

$$\alpha \geq 1, \beta \geq 1, \gamma \geq 1$$

Whereas, $\beta$, and $\gamma$ are constants that are decided by a smaller search grid. Instinctively $\varphi$ represents a user-specified co-efficient. They applied these in several types of efficient methods from B0-7. This has grown the model's complexities. Therefore, B7 is the maximum complexities method and B0 is the lowest complexities version of the algorithm. Fig. 2 represents the architecture of EfficientNet.

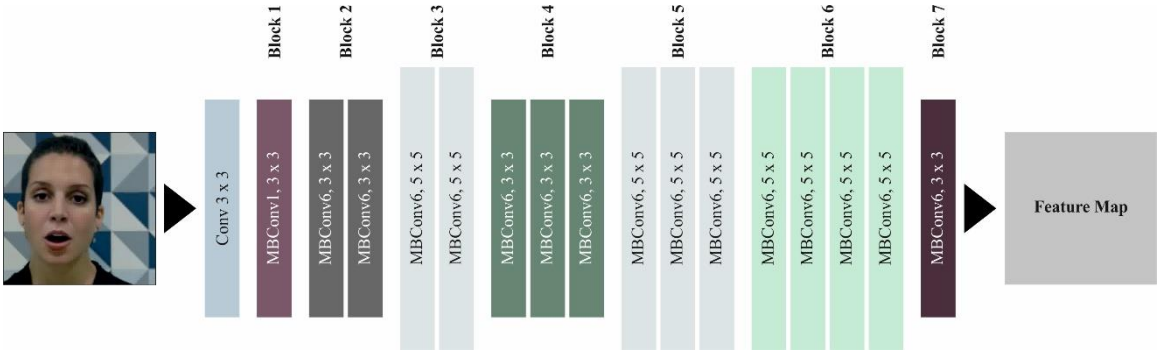


**Figure 2.** Framework of EfficientNet

### C. Stage III: Classification using DBN

Besides, the DBN model is applied for deepfake video detection. DBN is composed of numerous stacks of RBM and a classifier on the higher stage [22]. Every RBM can be made up of a hidden layer (HL) and a visible layer (VL). These VL and HL are associated with weights, and the neurons within the similar layer are autonomous to one another. DBN constantly removes the related features of the earlier layer of output over RBM, which understands the step-to-step training from the bottom-up approach. During the process of fine-tuning the complete structure of DBN, as well as training, the optimum maps between every RBM are gained. In the process of training, the neuron's state in the RMB using VL is stated as $v = \{v_1, v_2, \cdots, v_j\}$, and the neuron state in the HL is specified as $h = \{h_1, h_2, \cdots, h_j\}$. The RBM's ability function is shown as:

$$E(v,h) = -\sum_{i=1}^{n} a_j \, v_j - \sum_{j=1}^{m} b_j \, h_j - \sum_{i=1}^{n}\sum_{j=1}^{m} w_{ij} \, v_j h_j \tag{4}$$

Whereas $a$ defines the balancing vector of the VL; $b$ denotes a balancing vector of the HL; $w$ denotes the weight matrix connecting the HL and VL.

If each parameter is established, Eq. (4) is stated by the joined probability distribution function amongst the HL and VL, as represented in Eq. (5):

$$P(v,h) = \frac{1}{\sum_{i=1}^{m}\sum_{j=1}^{n} e^{-E(v,h)}} e^{E(v,h)} \tag{5}$$

The normally utilized activation functions in DL contain ReLU, Tanh, Sigmaid, and so on. If tanh and sigmoid functions model negative and positive infinities, the gradient vanishes, and the weight upgrade speed reduces. It's a plane and contains a quicker convergence speed rate. Hard swish function illustrations are:

$$Hard\ swish(x) = x\frac{ReLU6(x+3)}{6} \tag{6}$$



DOI: https://doi.org/10.54216/FPA.180120
Received: July 18, 2024 Revised: October 22, 2024 Accepted: January 04, 2025

$$ReLU6 = min\,(ReLU, 6) \tag{7}$$

The conditional distribution probability for HL and the VL is:

$$P(v_i = 1|v) = Hard\ swish\left(b_i + \sum_{j=1}^{n} v_j\, w_{ij}\right) \tag{8}$$

$$P(v_i = 1|h) = Hard\ swish\left(a_i + \sum_{j=1}^{n} h_j\, w_{ij}\right) \tag{9}$$

The new sigmoid function is substituted by a hard swish function in the work. This might resolve the difficulty in which the gradient vanishes and the weight upgrades the speed declines. Hence, the precision of the DBN has been increased.

### D. Stage IV: Parameter Tuning Process

Eventually, the POA was deployed for the optimum hyperparameter choice of the DBN classifier. The developed POA technique is a population-based model that may attain effectual solutions for optimizer issues by utilizing its population exploration influence in the problem-solving space in an iteration procedure [23]. Every member of the POA defines the significance of the decision variable of an issue as per its location in the search space. Thus, every member of POA is a candidate solution, which is demonstrated from an arithmetical viewpoint utilizing a vector, whereas every element resembles a decision variable. From an arithmetic standpoint, the group of these vectors is displayed utilizing a matrix as per Eq. (10). The main location at the start of the technique is set utilizing Eq. (11).

$$X = \begin{bmatrix} X_1 \\ \vdots \\ X_i \\ \vdots \\ X_N \end{bmatrix}_{N\times m} = \begin{bmatrix} x_{1,1} & \cdots & x_{1,d} & \cdots & x_{1,m} \\ \vdots & \ddots & \vdots & \iddots & \vdots \\ x_{i,1} & \cdots & x_{i,d} & \cdots & x_{i,m} \\ \vdots & \iddots & \vdots & \ddots & \vdots \\ x_{N,1} & \cdots & x_{N,d} & \cdots & x_{N,m} \end{bmatrix}_{N\times m}, \tag{10}$$

$$x_{i,d} = lb_d + r \cdot (ub_d - lb_d), \tag{11}$$

Where $X$ denotes a matrix form of POA population, $X_i$ signifies the $ith$ POA member, $x_{i,d}$ refers to its $dth$ dimension in the searching space, $N$ and $m$ indicate the number of population members and decision variables, respectively; $r$ denotes a value generated at random in the range $0\ and\ 1$, and $lb_d$ and $ub_d$ refer to lower and upper bounds of the decision variables, correspondingly. The set of assessed values of the issue can be signified utilizing a vector as per Eq. (12).

$$F = \begin{bmatrix} F_1 \\ \vdots \\ F_i \\ \vdots \\ F_N \end{bmatrix}_{N\times 1} = \begin{bmatrix} F(X_1) \\ \vdots \\ F(X_i) \\ \vdots \\ F(X_N) \end{bmatrix}_{N\times 1} \tag{12}$$

In this equation, $F$ denotes a vector of assessed objective function (OF) and $F_i$ signifies the estimated OF depends upon the $ith$ POA member. The assessed values are appropriate norms to compute the excellence of the candidate solution projected by every member of the POA. The finest assessed value for the OF relates to the finest members and the worst assessed value relates to the worst member.

During this project of the projected POA technique, the population member's location is upgraded depending upon the model of usual conduct among pufferfish and its predators. So, in every iteration, the location of POA was upgraded in dual stages such as exploration depending upon the model of predator's attack near and exploitation depending on the method of the pufferfish defending mechanism.

Phase of Exploration

In the initial phase, the population member location was upgraded depending on the model of the predator-attacking tactic near the pufferfish. In the design of POA, every population member is a hunter, the location has a superior value for the OF is measured as the the candidate pufferfish location for violence. Every population member is recognized by utilizing Eq. (13).

$$CP_i = \{X_k : F_k < F_i\ and\ k \neq i\}, whereas\ i = 1,2,\ldots,N\ and\ k \in \{1,2,\ldots,N\}, \tag{13}$$



DOI: https://doi.org/10.54216/FPA.180120
Received: July 18, 2024 Revised: October 22, 2024 Accepted: January 04, 2025

While $CP_i$ denotes a collection of candidate pufferfish positions for the $ith$ predator, $X_k$ refers to a superior value of the OF, and $F_k$ Denotes an OF value. Next, if the OF value was improved in the novel location and it substitutes the preceding location of the corresponding member as per Eq. (15).

$$x_{i,j}^{P1} = x_{i,j} + r_{i,j} \cdot \left(SP_{i,j} - I_{i,j} \cdot x_{i,j}\right), \tag{14}$$

$$X_i = \begin{cases} X_i^{P1}, F_i^{P1} \leq F_{i;}. \\ X_i, else, \end{cases} \tag{15}$$

Where, $SP_i$ denotes a nominated pufferfish for the $ith$ predator, which can be selected at random from the collection of $CP_i$ ($i.e., SP_j$ refers to an element of $CP_i$ set), $SP_{i,j}$ is its $jth$ dimension, $X_i^{P1}$ indicates a novel location computed for the $ith$ predator depending upon 1st phase, $x_{i,j}^{P1}$ refers to a $jth$ dimension, $F_i^{P1}$ denotes it's OF value, $r_{i,j}$ are randomly generated values from the range of $0\ and\ 1$, and $I_{i,j}$ are numbers which are chosen at random as 1 or 2.

Phase of exploitation

In the second stage, the location is upgraded depending upon the method of a pufferfish's defense device besides hunter assaults. Depending upon the demonstration of the predator's location alteration after getting away from the hunter, a novel location was computed for every POA member utilizing Eq. (16). Next, this novel location, if it enhances the OF values, substitutes the equivalent member as per to Eq. (17).

The Eq. (17) was made to increase the technique. When a new position is computed for the POA member, then it is tested from a contrast of the OF value. If it is positive, then the novel place is adequate for the equivalent member of POA, or else the novel location is unsuitable (due to it mains to a poorer solution) and the equivalent member rests in the preceding location. So, Eq. (17) displays that the upgrade procedure for every POA member is uncertain about enhancing the OF value.

$$x_{i,j}^{P2} = x_{i,j} + \left(1 - 2r_{i,j}\right) \cdot \frac{ub_j - lb_j}{t}, \tag{16}$$

$$X_i = \begin{cases} X_i^{P2}, F_i^{P2} \leq F_{i;}. \\ X_i, else, \end{cases} \tag{17}$$

Where $X_i^{P2}$ denotes a novel location intended for $ith$ predator depending upon the 2nd stage of the projected POA, $x_{i,j}^{P2}$ refers to its $jth$ dimension, $F_i^{P2}$ indicates a value of an OF, $r_{i,j}$ are randomly generated and valued in the range of $[0,1]$, and $t$ refers to several iterations. The fitness selection is a substantial feature inducing the performances of the POA. The parameter selection process has the solution encoding technique for estimating the effectiveness of the candidate solutions. In this study, the POA studies precision as the main criterion to model the fitness function (FF).

$$Fitness\ = \max\left(P\right) \tag{18}$$

$$P = \frac{TP}{TP + FP} \tag{19}$$

Here, $FP$ and $TP$ demonstrate the false and true positive rates.

## 4. Result Analysis and Discussion

In this section, the experimental validation of the LPODBN-EDVD method is tested under the FaceForensics++ dataset [24]. The dataset contains 300 videos under dual classes as displayed in Table 1.

**Table 1:** Details on Dataset

| FaceForensics++ Dataset | |
|---|---|
| Class | No. of Videos |
| Real | 150 |
| Deep Fake | 150 |
| Total Videos | 300 |



DOI: https://doi.org/10.54216/FPA.180120
Received: July 18, 2024 Revised: October 22, 2024 Accepted: January 04, 2025

Fig. 4 represents the classifier results of the LPODBN-EDVD model in terms of the FaceForensics++ dataset. Figs. 4a-4b displays the confusion matrices with precise classification and identification of all 16 classes on a 70%TRAP and 30%TESP. Fig. 4c illustrates the investigation of PR, demonstrating maximum performance across each class. Ultimately, Fig. 4d demonstrates the ROC study, showing efficient outcomes with great ROC outcomes for distinct classes.

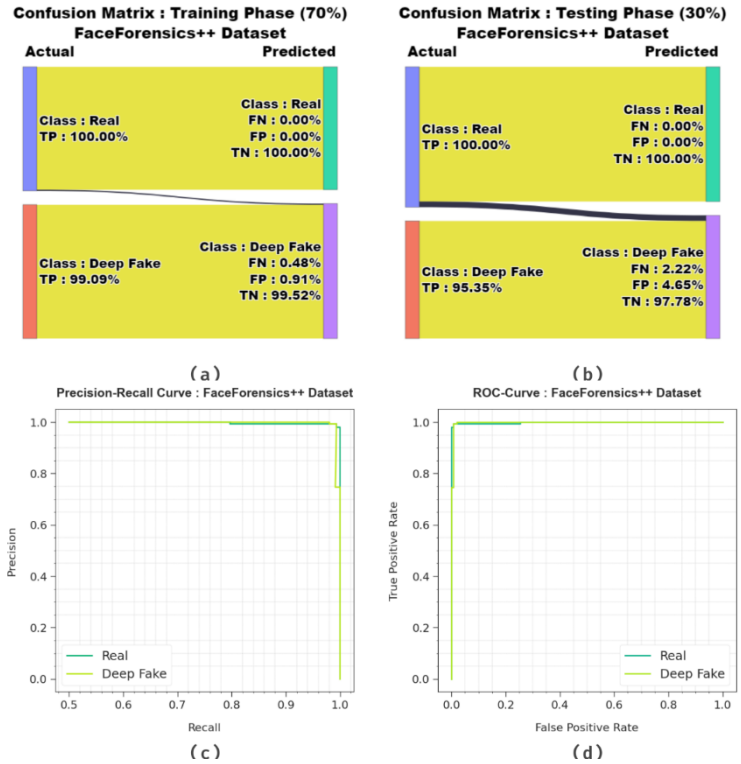


**Figure 4.** FaceForensics++ dataset (a-b) 70%TRAP: 30%TESP of confusion matrices and (c-d) Curves of PR and ROC

The deepfake video detection result of the LPODBN-EDVD algorithm on the FaceForensics++ dataset is exhibited in Table 2 and Fig. 5. The results denote that the LPODBN-EDVD model accurately identified the instances. On 70%TRAP, the LPODBN-EDVD technique achieves an average $accu_y$ of 99.50%, $prec_n$ of 99.55%, $reca_l$ of 99.50%, $F_{score}$ of 99.52%, and MCC of 99.05%. Furthermore, on 30%TESP, the LPODBN-EDVD system reaches an average $accu_y$ of 97.96%, $prec_n$ of 97.67%, $reca_l$ of 97.96%, $F_{score}$ of 97.77%, and MCC of 95.63%.

**Table 2:** Deepfake videos detection of LPODBN-EDVD technique under FaceForensics++ dataset

| Class | $Accu_y$ | $Prec_n$ | $Reca_l$ | $F_{score}$ | $MCC$ |
|---|---|---|---|---|---|
| TRAP (70%) | | | | | |
| Real | 99.01 | 100.00 | 99.01 | 99.50 | 99.05 |
| Deep Fake | 100.00 | 99.09 | 100.00 | 99.54 | 99.05 |
| Average | 99.50 | 99.55 | 99.50 | 99.52 | 99.05 |
| TESP (30%) | | | | | |
| Real | 95.92 | 100.00 | 95.92 | 97.92 | 95.63 |
| Deep Fake | 100.00 | 95.35 | 100.00 | 97.62 | 95.63 |
| Average | 97.96 | 97.67 | 97.96 | 97.77 | 95.63 |



DOI: https://doi.org/10.54216/FPA.180120
Received: July 18, 2024 Revised: October 22, 2024 Accepted: January 04, 2025

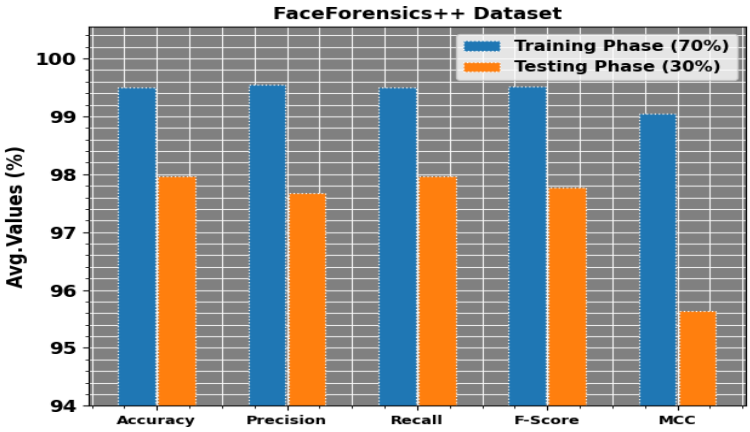


**Figure 5.** Average of LPODBN-EDVD technique under FaceForensics++ dataset

In Fig. 6, the training (TRA) and validation (VLA) $accu_y$ results of the LPODBN-EDVD system on the FaceForensics++ dataset are exhibited. The $accu_y$ values are computed for 0-300 epoch counts. The result underlined that the TRA and VLA $accu_y$ values show a rising trend that indicates the capability of the LPODBN-EDVD method with superior performance across different iterations. Besides, the TRA and VLA accuracies stay adjacent across the epoch counts, which illustrations the lowest minimal overfitting and presents the greater performance of the LPODBN-EDVD methodology, guaranteeing consistent prediction on unseen samples.

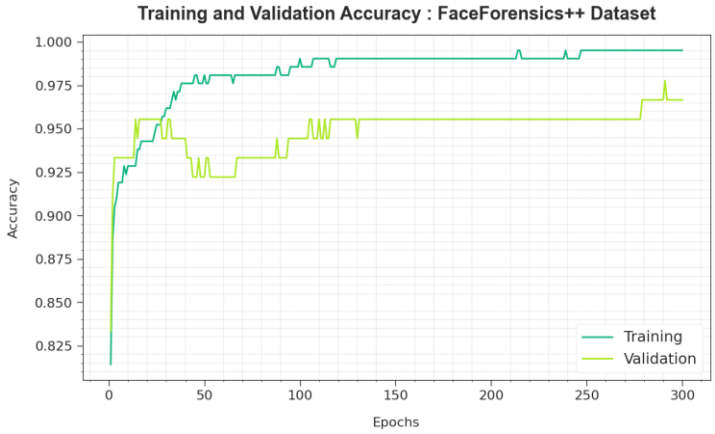


**Figure 6.** $Accu_y$ Curve of LPODBN-EDVD technique under FaceForensics++ dataset


**Figure 7.** Loss curve of LPODBN-EDVD methodology at FaceForensics++ dataset



DOI: https://doi.org/10.54216/FPA.180120
Received: July 18, 2024 Revised: October 22, 2024 Accepted: January 04, 2025

In Fig. 7, the TRA and VLA loss graph of the LPODBN-EDVD system on the FaceForensics++ dataset are presented. The loss values are computed for 0-300 epoch counts. It can be denoted that the TRA and VLA $accu_y$ rates illustrate a lower trend, which notifies the ability of the LPODBN-EDVD model to balance a trade-off between generalization and data fitting. The continual decline in loss rates also pledges a better solution of the LPODBN-EDVD methodology and fine-tuning the prediction results on time.

Table 3 and Fig. 8 study the comparative results of the LPODBN-EDVD model using the current approaches under the FaceForensics++ dataset [11]. The results highlighted that the ResNet101, MesoNet4, Data-Fusion-Cascade, EDL-Det and Facial Image Inpainting methods have stated worse performance. Meanwhile, MesoNet4 + ResNet-101, and One-Class learning methods have closer results. Additionally, the LPODBN-EDVD approach reported enhanced performance with maximum $prec_n$, $reca_l$, $accu_y$, and $F_{score}$ of 99.55%, 99.50%, 99.50%, and 99.52%, respectively.

**Table 3:** Comparative outcome of LPODBN-EDVD method with existing models under FaceForensics++ database

| FaceForensics++ | | | | |
|---|---|---|---|---|
| Algorithm | $Accu_y$ | $Prec_n$ | $Reca_l$ | $F_{score}$ |
| LPODBN-EDVD | 99.50 | 99.55 | 99.50 | 99.52 |
| ResNet101Model | 93.54 | 93.40 | 95.30 | 94.34 |
| MesoNet4 Model | 93.04 | 92.75 | 93.89 | 93.32 |
| MesoNet4 + ResNet-101 | 98.73 | 98.71 | 99.44 | 99.07 |
| One-Class Learning | 95.36 | 94.29 | 96.41 | 95.20 |
| Data-Fusion-Cascade | 90.42 | 91.07 | 88.24 | 90.29 |
| EDL-Det Classifier | 92.56 | 93.25 | 91.25 | 92.09 |
| Facial Image Inpainting | 91.29 | 92.42 | 90.52 | 91.14 |

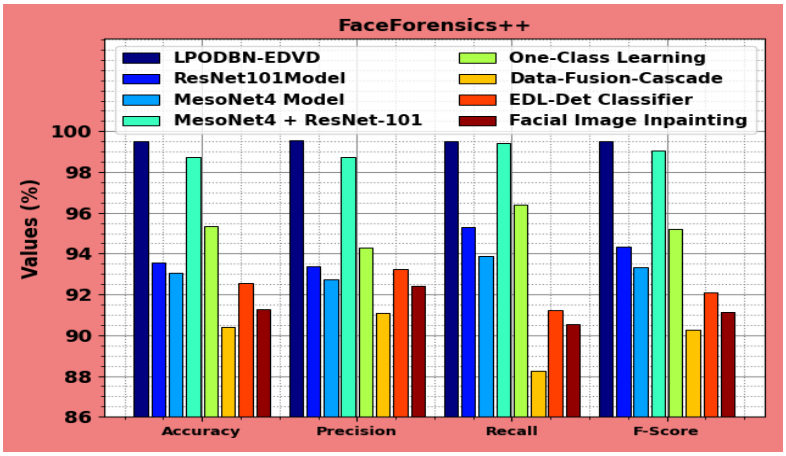


**Figure 8.** Comparative analysis of LPODBN-EDVD technique under FaceForensics++ database

In Table 4 and Fig. 9, the comparative results of the LPODBN-EDVD approach under the FaceForensics++ database are stated with respect to execution time (ET). The outcomes suggest that the LPODBN-EDVD methodology gets larger performance. With respect to ET, the LPODBN-EDVD approach offers lesser ET of 15s where the ResNet-101, MesoNet-4, Hybrid (MesoNet-4+ResNet-101), 1-Class Learning, Data Fusion Cascade, EDLDet, and Facial Image In painting methods obtain higher ET values of 60s, 43s, 56s, 50s, 35s, 83s, and 79s, respectively.



DOI: https://doi.org/10.54216/FPA.180120
Received: July 18, 2024 Revised: October 22, 2024 Accepted: January 04, 2025

**Table 4:** ET outcome of LPODBN-EDVD technique with recent models under FaceForensics++ dataset

| FaceForensics++ | |
|---|---|
| Algorithm | Execution Time (sec) |
| LPODBN-EDVD | 15 |
| ResNet101Model | 60 |
| MesoNet4 Model | 43 |
| Hybrid (MesoNet4 + ResNet-101) | 56 |
| One-Class Learning | 50 |
| Data-Fusion-Cascade | 35 |
| EDL-Det Classifier | 83 |
| Facial Image Inpainting | 79 |

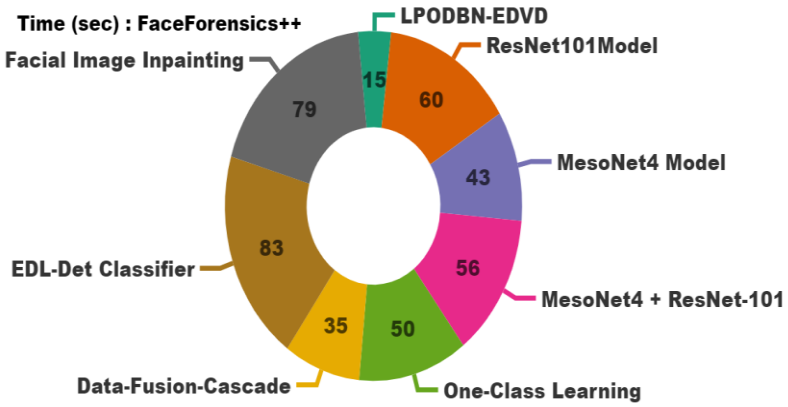


**Figure 9.** ET outcome of LPODBN-EDVD technique with recent models under FaceForensics++ dataset

The performance assessment of the POADEL-ID system is verified under the Celeb-DF v1 dataset [25]. The dataset contains 600 videos under dual-class labels as represented in Table 5.

**Table 5:** Details of Celeb-DF v1 dataset

| Celeb-DF v1 Dataset | |
|---|---|
| Class | No. of Videos |
| Real | 300 |
| Deep Fake | 300 |
| Total Videos | 600 |

Fig. 10 represents the classification results of the LPODBN-EDVD approach under the Celeb-DF v1 dataset. Figs. 10a-10b displays the confusion matrix with precise recognition and classification of all 16 classes on a 70%TRAP and 30%TESP. Fig. 10c displays the PR examination, representing maximum performance across the each class. Lastly, Fig. 10d exhibits the ROC study, indicating proficient results with high ROC outcomes for several class labels.



DOI: https://doi.org/10.54216/FPA.180120
Received: July 18, 2024 Revised: October 22, 2024 Accepted: January 04, 2025

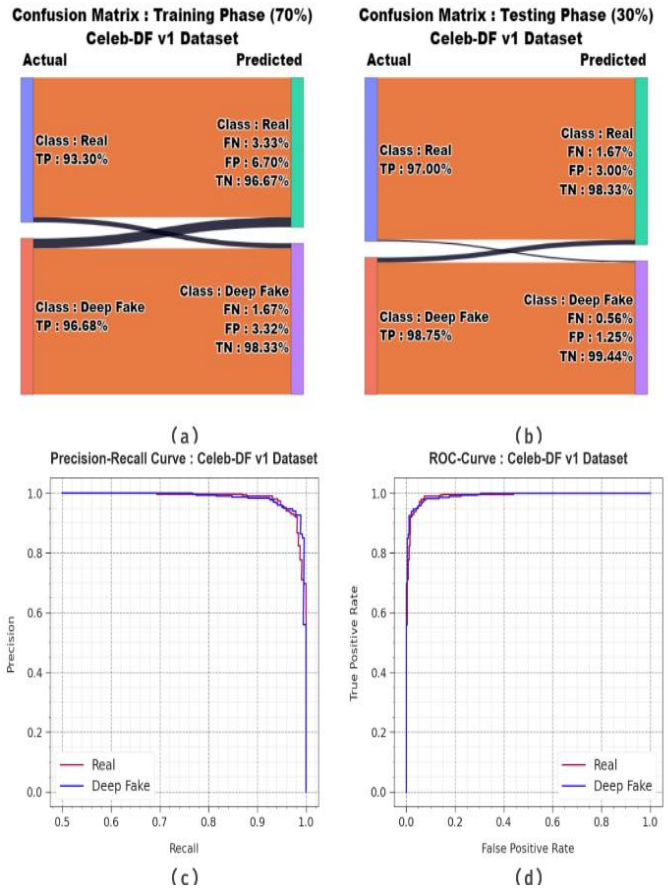


**Figure 10.** Celeb-DF v1 dataset (a-b) 70%TRAP: 30%TESP of confusion matrices and (c-d) Curves of PR and ROC

The deepfake video detection outcome of the LPODBN-EDVD system at the Celeb-DF v1 dataset is exposed in Table 6 and Fig. 11. The table values denote that the LPODBN-EDVD method correctly identified the samples. On 70%TRAP, the LPODBN-EDVD algorithm reaches an average $accu_y$ of 95.06%, $prec_n$ of 94.99%, $reca_l$ of 95.06%, $F_{score}$ of 95.00%, and MCC of 90.05%. Furthermore, on 30%TESP, the LPODBN-EDVD system attains an average $accu_y$ of 97.66%, $prec_n$ of 97.88%, $reca_l$ of 97.66%, $F_{score}$ of 97.76%, and MCC of 95.54%.

**Table 6:** Deepfake videos detection of LPODBN-EDVD technique under Celeb-DF v1 dataset

| Class | $Accu_y$ | $Prec_n$ | $Reca_l$ | $F_{score}$ | $MCC$ |
|---|---|---|---|---|---|
| TRAP (70%) | | | | | |
| Real | 96.53 | 93.30 | 96.53 | 94.89 | 90.05 |
| Deep Fake | 93.58 | 96.68 | 93.58 | 95.10 | 90.05 |
| Average | 95.06 | 94.99 | 95.06 | 95.00 | 90.05 |
| TESP (30%) | | | | | |
| Real | 98.98 | 97.00 | 98.98 | 97.98 | 95.54 |
| Deep Fake | 96.34 | 98.75 | 96.34 | 97.53 | 95.54 |
| Average | 97.66 | 97.88 | 97.66 | 97.76 | 95.54 |



DOI: https://doi.org/10.54216/FPA.180120
Received: July 18, 2024 Revised: October 22, 2024 Accepted: January 04, 2025

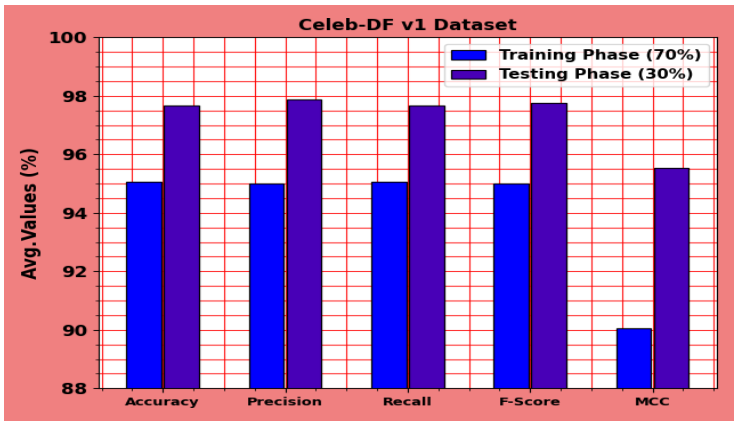


**Figure 11.** Average of LPODBN-EDVD technique under Celeb-DF v1 dataset

In Fig. 12, the TRA and VLA $accu_y$ results of the LPODBN-EDVD technique on the Celeb-DF v1 database are displayed. The $accu_y$ values are computed throughout 0-300 epoch counts. The result emphasized that the TRA and VLA $accu_y$ values display a rising trend, which notified the ability of the LPODBN-EDVD approach with enhanced performance under several iterations. Additionally, the TRA and VLA $accu_y$ remains closer over the epochs, which specifies the least minimum overfitting and shows improved performance of the LPODBN-EDVD model, promising consistent prediction on unidentified samples.

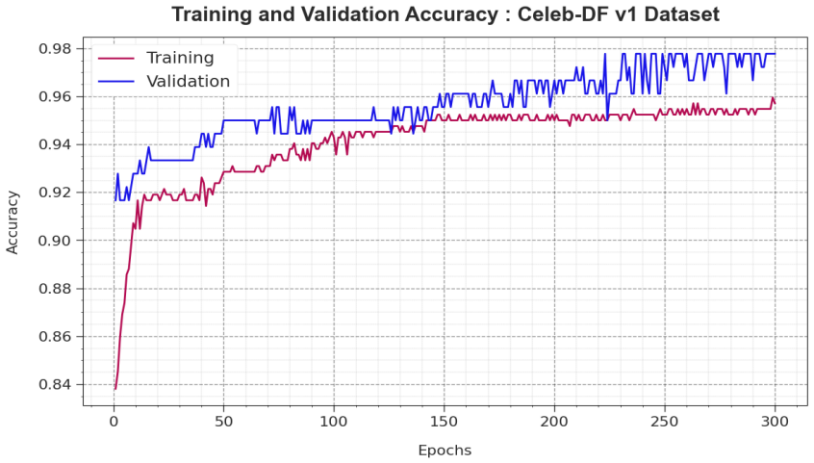


**Figure 12.** $Accu_y$ Curve of LPODBN-EDVD technique under Celeb-DF v1 dataset

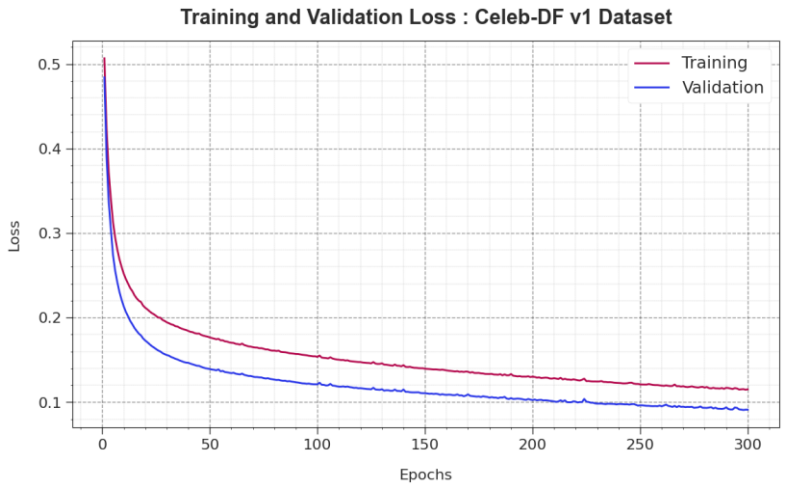


**Figure 13.** Loss curve of LPODBN-EDVD method at Celeb-DF v1 dataset



DOI: https://doi.org/10.54216/FPA.180120
Received: July 18, 2024 Revised: October 22, 2024 Accepted: January 04, 2025

In Fig. 13, the TRA and VLA loss graph of the LPODBN-EDVD method on the Celeb-DF v1 dataset is demonstrated. The loss values are computed throughout 0-300 epoch counts. It is denoted that the TRA and VLA $accu_y$ values demonstrate a lower trend, which notifies the abilities of the LPODBN-EDVD system to balance a trade-off between generalization and data fitting. The constant decrease in loss rates additionally pledges the improved outcome of the LPODBN-EDVD methodology and fine-tuning of the prediction outcomes on time.

Table 7 and Fig. 14 study the comparative investigation of the LPODBN-EDVD approach with the current methods under Celeb-DF v1 dataset. The results highlighted that the LPODBN-EDVD approach stated enhanced performance with maximum $prec_n$, $reca_l$, $accu_y$, and $F_{score}$ of 97.88%, 97.66%, 97.66%, and 97.76%, respectively. Whereas, the ResNet101 method has attained closer results of $prec_n$, $reca_l$, $accu_y$, and $F_{score}$ of 94.2%, 93.21%, 91.27%, and 93.69%, correspondingly. Whereas other methods MesoNet4, MesoNet4+ResNet-10, One-Class Learning, Data-Fusion-Cascade, EDL-Det, and Facial Image Inpainting approaches have reported worse performance.

**Table 7:** Comparative analysis of LPODBN-EDVD technique with existing models under Celeb-DF v1dataset

| Celeb-DF v1 Dataset | | | | |
|---|---|---|---|---|
| Algorithm | $Accu_y$ | $Prec_n$ | $Reca_l$ | $F_{score}$ |
| LPODBN-EDVD | 97.66 | 97.88 | 97.66 | 97.76 |
| ResNet101Model | 91.27 | 94.2 | 93.21 | 93.69 |
| MesoNet4 Model | 89.94 | 87.13 | 96.68 | 91.83 |
| MesoNet4+ResNet-101 | 96.89 | 97.54 | 97.29 | 97.42 |
| One-Class Learning | 92.012 | 91.016 | 93.012 | 92.016 |
| Data-Fusion-Cascade | 87.02 | 88.02 | 85.016 | 87.02 |
| EDL-Det Classifier | 89.013 | 90.013 | 88.02 | 89.016 |
| Facial Image Inpainting | 88.02 | 89.018 | 87.019 | 88.019 |

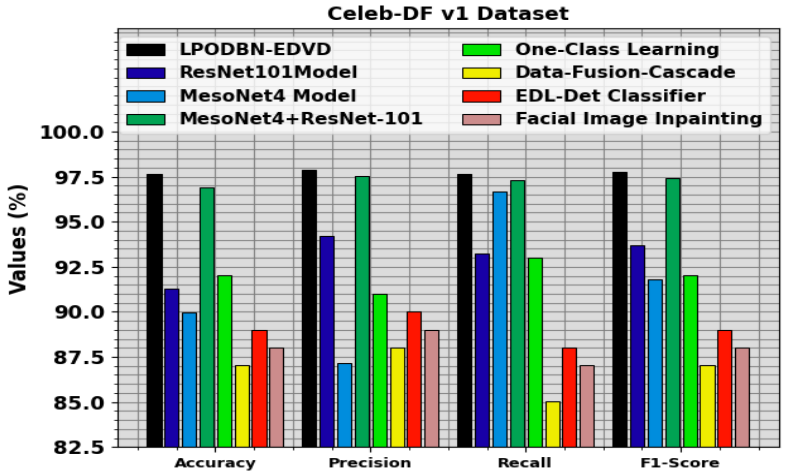


**Figure 14.** Comparative analysis of LPODBN-EDVD technique under Celeb-DF v1dataset

In Table 8 and Fig. 15, the comparative outcomes of the LPODBN-EDVD algorithm under Celeb-DF v1dataset are stated interms of computation time (CT). The table values suggest that the LPODBN-EDVD approach gets superior performance. Based on CT, the ResNet101, MesoNet4, MesoNet4+ResNet-101, One-Class Learning, Data-Fusion-Cascade, EDL-Det, and Facial Image Inpainting models have attained the worst performance. Where the presented LPODBN-EDVD algorithm got the least CT of the 30s when compared to other recent methods.



DOI: https://doi.org/10.54216/FPA.180120
Received: July 18, 2024 Revised: October 22, 2024 Accepted: January 04, 2025

**Table 8:** CT of LPODBN-EDVD technique under Celeb-DF v1dataset

| Celeb-DF v1 Dataset | |
|---|---|
| Algorithm | Computational Time (sec) |
| LPODBN-EDVD | 30 |
| ResNet101Model | 80 |
| MesoNet4 Model | 66 |
| MesoNet4+ResNet-101 | 80 |
| One-Class Learning | 53 |
| Data-Fusion-Cascade | 90 |
| EDL-Det Classifier | 57 |
| Facial Image Inpainting | 81 |

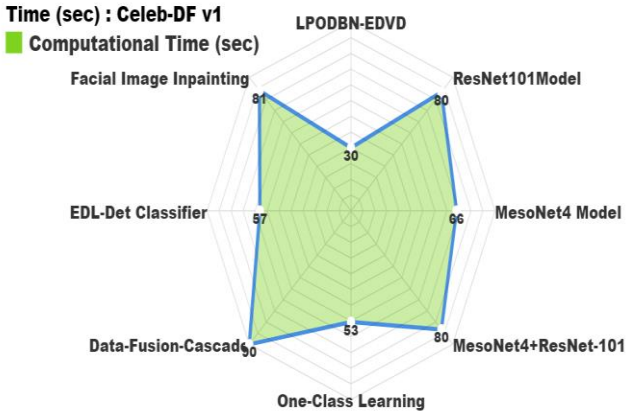


**Figure 15.** CT of LPODBN-EDVD model under Celeb-DF v1dataset

## 5. Conclusion

In this paper, we have proposed a new LPODBN-EDVD technique. The LPODBN-EDVD system intends to detect fake videos using the DL model. To accomplish that, the LPODBN-EDVD technique has four distinct stages involving different levels of preprocessing, feature extraction, classification using DBN, and POA-based parameter tuning processes. Initially, the presented LPODBN-EDVD technique undergoes data preprocessing stages including splitting the video into frames, face detection, and face cropping. For the process of feature extraction, the EfficientNet model is exploited. Besides, the DBN model is applied for deepfake video detection. Finally, the POA is used for the optimum parameter choice of the DBN model. A wide range of simulations was involved in exhibiting the promising results of the LPODBN-EDVD method. The experimental analysis pointed out the enhanced performance of the LPODBN-EDVD technique compared to existing systems.

**Funding:** "This research received no external funding"

**Acknowledgement:** The author gratefully acknowledges technical support provided by the Faculty of Computing and Information Technology, King Abdulaziz University, Jeddah, Saudi Arabia.

**Conflicts of Interest:** "The authors declare no conflict of interest."



DOI: https://doi.org/10.54216/FPA.180120
Received: July 18, 2024 Revised: October 22, 2024 Accepted: January 04, 2025

**References**


**[1]** D. Myvizhi and J. M. J. Pamila, "Extensive analysis of deep learning-based deepfake video detection," *Journal of Ubiquitous Computing and Communication Technologies*, vol. 4, no. 1, pp. 1–8, 2022.

**[2]** N. M. Alnaim, Z. M. Almutairi, M. S. Alsuwat, H. H. Alalawi, A. Alshobaili, and F. S. Alenezi, "DFFMD: A Deepfake Face Mask Dataset for Infectious Disease Era With Deepfake Detection Algorithms," *IEEE Access*, vol. 11, pp. 16711–16722, 2023.

**[3]** A. Mitra, S. P. Mohanty, P. Corcoran, and E. Kougianos, "Detection of deep-morphed deepfake images to make robust automatic facial recognition systems," in *2021 19th OITS International Conference on Information Technology (OCIT)*, Dec. 2021, pp. 149–154.

**[4]** S. Guefrachi et al., "Deep learning based DeepFake video detection," in *2023 International Conference on Smart Computing and Application (ICSCA)*, Feb. 2023, pp. 1–8.

**[5]** A. Agarwal, A. Agarwal, S. Sinha, M. Vatsa, and R. Singh, "MD-CSDNetwork: Multi-domain cross stitched network for deepfake detection," in *2021 16th IEEE International Conference on Automatic Face and Gesture Recognition (FG 2021)*, Dec. 2021, pp. 1–8.

**[6]** P. Liang, G. Liu, Z. Xiong, H. Fan, H. Zhu, and X. Zhang, "A facial geometry based detection model for face manipulation using CNN-LSTM architecture," *Information Sciences*, vol. 633, pp. 370–383, 2023.

**[7]** Z. Akhtar, M. R. Mouree, and D. Dasgupta, "Utility of deep learning features for facial attributes manipulation detection," in *2020 IEEE International Conference on Humanized Computing and Communication with Artificial Intelligence (HCCAI)*, Sep. 2020, pp. 55–60.

**[8]** R. Rafique et al., "Deep fake detection and classification using error-level analysis and deep learning," *Scientific Reports*, vol. 13, no. 1, p. 7422, 2023.

**[9]** A. Ismail, M. Elpeltagy, M. S. Zaki, and K. Eldahshan, "A new deep learning-based methodology for video deepfake detection using XGBoost," *Sensors*, vol. 21, no. 16, p. 5413, 2021.

**[10]** A. H. Khalifa, N. A. Zaher, A. S. Abdallah, and M. W. Fakhr, "Convolutional Neural Network Based on Diverse Gabor Filters for Deepfake Recognition," *IEEE Access*, vol. 10, pp. 22678–22686, 2022.

**[11]** M. Javed, Z. Zhang, F. H. Dahri, and A. A. Laghari, "Real-Time Deepfake Video Detection Using Eye Movement Analysis with a Hybrid Deep Learning Approach," *Electronics*, vol. 13, no. 15, p. 2947, 2024.

**[12]** A. Qadir, R. Mahum, M. A. El-Meligy, A. E. Ragab, A. AlSalman, and M. Awais, "An efficient deepfake video detection using robust deep learning," *Heliyon*, vol. 10, no. 5, 2024.

**[13]** S. Talreja, A. Bindle, V. Kumar, I. Budhiraja, and P. Bhattacharya, "Security Strengthen and Detection of Deepfake Videos and Images Using Deep Learning Techniques," in *2024 IEEE International Conference on Communications Workshops (ICC Workshops)*, Jun. 2024, pp. 1834–1839.

**[14]** H. S. Alhaji, Y. Celik, and S. Goel, "An Approach to Deepfake Video Detection Based on ACO-PSO Features and Deep Learning," *Electronics*, vol. 13, no. 12, p. 2398, 2024.

**[15]** M. Bhat, P. Agrawal, and C. Gupta, "DFDA: An Analysis of Deep Learning Models to Detect Deepfake Videos," 2024.

**[16]** T. Qiao, S. Xie, Y. Chen, F. Retraint, and X. Luo, "Fully unsupervised deepfake video detection via enhanced contrastive learning," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2024.

**[17]** Z. Chen, X. Liao, X. Wu, and Y. Chen, "Compressed Deepfake Video Detection Based on 3D Spatiotemporal Trajectories," *arXiv preprint arXiv:2404.18149*, 2024.

**[18]** K. Omar, R. H. Sakr, and M. F. Alrahmawy, "An ensemble of CNNs with self-attention mechanism for DeepFake video detection," *Neural Computing and Applications*, vol. 36, no. 6, pp. 2749–2765, 2024.

**[19]** E. Sabir et al., "Recurrent convolutional strategies for face manipulation detection in videos," *Interfaces (GUI)*, vol. 3, no. 1, pp. 80–87, 2019.

**[20]** S. Jia and Y. Tian, "Face Detection Based on Improved Multi-task Cascaded Convolutional Neural Networks," *IAENG International Journal of Computer Science*, vol. 51, no. 2, 2024.

**[21]** N. Bansal et al., "Real-time advanced computational intelligence for deep fake video detection," *Applied Sciences*, vol. 13, no. 5, p. 3095, 2023.

**[22]** G. Jia, Y. Meng, and Z. Qin, "Bearing Fault Diagnosis Based on Optimized Feature Mode Decomposition and Improved Deep Belief Network," *Structural Durability & Health Monitoring (SDHM)*, vol. 18, no. 4, 2024.

**[23]** O. Al-Baik et al., "Pufferfish Optimization Algorithm: A New Bio-Inspired Metaheuristic Algorithm for Solving Optimization Problems," *Biomimetics*, vol. 9, no. 2, p. 65, 2024.

**[24]** H. Le, "FaceForensics Dataset," *Kaggle*, 2023. [Online]. Available: https://www.kaggle.com/datasets/hungle3401/faceforensics.

**[25]** Y. Li, "Celeb-DeepfakeForensics Dataset," *GitHub*, 2023. [Online]. Available: https://github.com/yuezunli/celeb-deepfakeforensics.




DOI: https://doi.org/10.54216/FPA.180120
Received: July 18, 2024 Revised: October 22, 2024 Accepted: January 04, 2025