Intelligent Automation & Soft Computing
DOI: 10.32604/iasc.2023.030486
Article

Tech Science Press

Deep Fake Detection Using Computer Vision-Based Deep Neural Network with Pairwise Learning

R. Saravana Ram[1], M. Vinoth Kumar[2], Tareq M. Al-shami[3], Mehedi Masud[4], Hanan Aljuaid[5] and Mohamed Abouhawwash[6,7,*]

[1]Department of Electronics and Communication Engineering, Anna University, University College of Engineering, Dindigul, 624622, India
[2]Department of Computer Science and Engineering, Anna University, University College of Engineering, Dindigul, 624622, India
[3]Department of Mathematics, Faculty of Science, Sana'a University, Sana'a 13509, Yemen
[4]Department of Computer Science, College of Computers and Information Technology, Taif University, Taif, 21944, Saudi Arabia
[5]Computer Sciences Department, College of Computer and Information Sciences, Princess Nourah bint Abdulrahman University (PNU), Riyadh, 11671, Saudi Arabia
[6]Department of Mathematics, Faculty of Science, Mansoura University, Mansoura, 35516, Egypt
[7]Department of Computational Mathematics, Science, and Engineering (CMSE), Michigan State University, East Lansing, MI, 48824, USA
*Corresponding Author: Mohamed Abouhawwash. Email:abouhaww@msu.edu
Received: 27 March 2022; Accepted: 27 April 2022

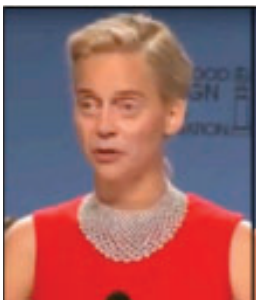
Abstract: Deep learning-based approaches are applied successfully in many fields such as deepFake identification, big data analysis, voice recognition, and image recognition. Deepfake is the combination of deep learning in fake creation, which states creating a fake image or video with the help of artificial intelligence for political abuse, spreading false information, and pornography. The artificial intelligence technique has a wide demand, increasing the problems related to privacy, security, and ethics. This paper has analyzed the features related to the computer vision of digital content to determine its integrity. This method has checked the computer vision features of the image frames using the fuzzy clustering feature extraction method. By the proposed deep belief network with loss handling, the manipulation of video/image is found by means of a pairwise learning approach. This proposed approach has improved the accuracy of the detection rate by 98% on various datasets.

Keywords: Deep fake; deep belief network; fuzzy clustering; feature extraction; pairwise learning

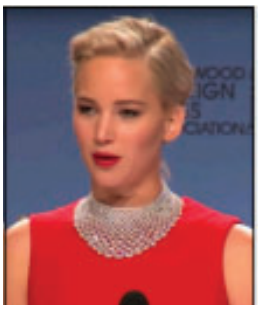
1 Introduction

DeepFake is a technological advancement that can synthesize the face of a person with a duplicate look as appearing on the original video using artificial intelligence concepts. It is created to manipulate the activity of the targeted person by saying the messages. Due to the development of smartphones and social media, deepfake technology poses a threat to digital content and spreads false information that is not identified

by human eyes [1]. Fig. 1 stipulates the real and fake images created from intelligent technology with an advanced network architecture that can also use a large amount of data to train the network.



Real Image



Fake Image

Figure 1: Real and deepfake image [1]

These kinds of forged videos and images created from deep learning algorithms can be used for facial authentication, which makes fake news on social media. This is easy to simplify the facial forgeries on the video of the targeted person [2,3] or ID proof [4,5]. Fake images can be existed in three different forms [6]; Lip Sync and Face Swap. Lip Sync is the source video that is modified by the recording of audio with a reliable movement in the mouth region. Puppet Master is a technique in the creation of animated expressions on faces, heads, and eye movements of the targeted person who is called a puppet. Moreover, this acting is performed in front of the camera to show the action of the puppet. Actually, Face Swapping replaces the source face with the destination face which is considered in source and destination videos. Face Swap is the standard deepfake method [7,8].

In order to create DeepFake, traditional approaches are used in visual effects and computer graphics methods. Various recent technological advancements with deep learning techniques including autoencoders and GAN (Generative Adversarial Networks) are used to create fake faces which are applied mainly in the computer vision area [9]. Furthermore, the DeepFake method needs a large volume of videos and image data to train the network model so as to create photo-realistic videos and images. Due to the advancement of social media, photos and videos of politicians and celebrities are available online. They are considered to be getting deep fakes. The first DeepFake video was created from the face swap of a celebrity to porn an actor in 2017. The incident has threatened for security reasons while these methods synthesize the speech of world leaders in making a fake speech for falsification needs [10].

There is also a positive side of deepfakes, including the applications in digital avatars, visual effects, creating videos of the last people for their relatives, Snapchat filters, and updating the episodes without reshooting for movies [11]. Comparatively, malicious uses have been increasing recently which cannot be distinguished from the real ones using deep neural networks with more sophisticated and developed computer algorithms. From the still image, one can create a new fake face [12]. Less work is required to produce such false images from tempered footage. Therefore, it is considered to be a threat affecting methods not only for celebrities and politicians but also for ordinary persons. This kind of falsification can create significant threats to violate the identity of persons and the privacy of human life. For example, the voice of the CEO is fake for $243000 [13]. DeepNude software is a recent release in the transformation of a person into non-consensual porn threats [14]. Similarly, the Chinese app Zao can synthesize the skilled user face into the bodies of movie stars and incorporate themselves with the

movies and television clips. For those reasons, truth-finding in the digital domain is critical and challenging.

- Anyone can create fake images using readily available deep learning tools and machine learning approaches. So far, various methods have been proposed to detect deepfakes [15] based on deep learning methods. To address this problem, the United States Defense Advanced Research Project Agency (DARPA) has started research on media forensics to hasten the fake digital media detection approaches. The Deepfake detection challenge was recently created in order to detect and prevent counterfeit images by Facebook inc. join with Microsoft or partner with AI coalition. Detailed surveys about artificial face creation and available intelligent detection methods are discussed in the paper using deep learning approaches. Even though most of the ways are developed in the detection of deepfakes, an improvement in accuracy and a reduction in net loss are challenging tasks. This work concentrates on viewing the fake images with a deep learning model. The contributions of the paper are as follows:
- Input image is preprocessed with Gabor filter-based Gaussian rule to remove the image noises and enhance its quality for better detection.
- Computer vision features are extracted using a fuzzy clustering approach, and the extracted relevant features are fed as input in the classification process.
- To detect the fake images, a deep belief network-based classifier is used, and this network is enhanced with a loss handling mechanism with pairwise learning. This approach will strengthen the energy function and reduces the net loss of the detection.
- The proposed system is implemented and compared with various traditional fake detection systems with four deepfake datasets. The experimental outcome has proven that the proposed artificial detection system has improved the detection accuracy and reduced the error rate.

The rest of the article is structured as shown; Section 2 reviews the work related to deepfake detection. Section 3 has the proposed methods and materials. Section 4 illustrates the experimental results and evaluations. Section 5 explains the conclusion of the proposed system with future suggestions.

2 Related Work

A paper for various DeepFake detection methods has been proposed recently. This uses a Deep Neural Network (DNN)using a Convolutional Neural Network (CNN) with tiny noises in the image to detect the fake images using a Convolutional Neural Network (CNN) by extracting the eye blinks and Long Short Term Memory (LSTM) has been used for the detection. [16] used LSTM and CNN for extracting the frame image of the video features. Li et al. [17] had detected the distorted face with a Residual Network (ResNet) 50 and Visual Geometry Group (VGG) 16 based CNN. Yang et al. [18] had extracted 68 landmarks from images of the face and detected them using SVM.

The compact features using a bag of words with various classifiers include random forest, SVM, and multi-layer perceptron to detect the fake and real images. Preprocessing steps are used to remove the noise from GAN images using Gaussian blur and Gaussian noise. This increases the statistical measure of the real and fake images at the pixel level and helps the classifiers to learn about the intrinsic features with a generalization facility than the traditional approaches or image step analysis networks. The proposed two methods use deep learning for deepfake detection. Phase one extracts the features using a standard fake network and architecture for Siamese network detection.

The proposed deep learning for detecting the image editing process using convolution layers learned with features. The blockchain approach sees the fake videos assuming that the videos are accurate, which

are associated to the parent videos, and every parent is hierarchically linked to their children. This blockchain helps the users to trace easily the natural association between the parent and the child video and to know that the video is copied multiple times. This model has been tested with a dataset consisting of greyscale images. It has obtained accuracy for multi-class classification as 99.10% and binary classification as 99.31%.

They have discussed the operational structure of the failed techniques with face swapping of images in a high-value precision. The Generative Adversarial Networks (GANs) with two neural network structures with generator and discriminator are used. The Generator section creates false images from the given data set. Conversely, discriminators and neural networks are used for image evaluation, synthesized using a generator, and verified authenticity. The main issues in the Deepfake are hard because of insulting the individual and assassination, spreading fake news in the society. The fake detection using k nearest neighbor (k-NN), Latent Dirichlet Allocation (LDA) and Support Vector Machine (SVM) methods are extracted by the pipeline features. They used attention GAN, Group-wise Deep Whitening and Coloring Method (GDWCT), StarGAN, StyleGAN, and StyleGAN2 datasets for evaluation. Neves had used CNN for deepfake detection using 100K faces datasets. Dang et al. [19] had used CNN and Autoencoder for fake detection with Dirichlet Free Form Deformation (DFFD) (ProGAN, StyleGAN) datasets. CNN and LSTM for deep fake detection with UADFV, Celeb-DF, FaceForensics++, Celeb-DF, DeepFake Detection Challenge datasets.

3 Proposed DeepFake Detection Methodology

Fig. 2 shows the proposed general architecture deep fake detection system. The input image is preprocessed using Gabor filter-based Gaussian rule for noise removal and appearance enhancement. The preprocessed image has been fed as input to the feature extraction process. The computer vision features are extracted and provided information to the classification process using fuzzy clustering method. Deep belief network is computed to identify the input image as fake or real. The net loss is handled using feature learning approach. DBN with feature learning will detect the input image as fake or real with an improved accuracy.

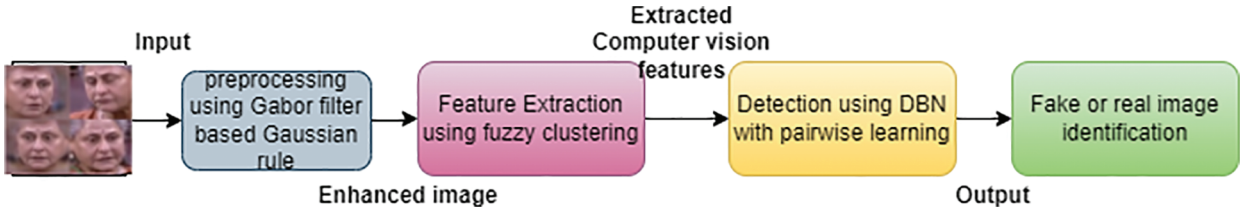


Figure 2: Proposed deepfake image detection architecture

3.1 Preprocessing

The input image from the dataset is preprocessed with Gabor Filter based Gaussian filter for noise removal and enhanced the image to improve the feature extraction and classification process. Gabor filter is a multi-resolution filter [20–27] for the removal of noise. It can be applied in various orientations and frequencies [28–37]. The steps to remove the noises using Gabor filter based gaussian rule are stated below.

Algorithm 1: Gabor Filter based Gaussian Rule

Step 1: Input image is converted into image as $A_{i,j}$, where i = 1,2,.. m and j = 1, 2, … n

Step 2: Image is divided as 16 ×16 sub-blocks.

Step 3: Gaussian rule of standard deviation is applied onto the image Ai,j with various angles such as 0°, 20°,40°, 60°, 120° and various orientations such as 60,80,120 and 140 using the Eq. (1)

$$A_{i,j}(i,j,\sigma) = e^{-\frac{(i-i_c)^2}{2\sigma_i^2}-\frac{(j-j_c)^2}{2\sigma_j^2}} e^{j\left(\omega_{ic}i+\omega_{jc}j\right)} \tag{1}$$

where,

ω_{ic}, ω_{jc} -Image centre frequency of i and j the direction.

σ_i, σ_j-Gaussian function standard deviation

I,j-Position of the image in pixels

Step 4: Gabor filter-based Gaussian rule is stated in Eq. (2)

$$\varphi(i,j,\omega,\sigma,\ \theta) = e^{-\frac{\left(i\cos\theta_p - j\sin\theta_p\right)^2}{2\sigma_i^2}+\frac{\left(-i\sin\theta_p - j\cos\theta_p\right)^2}{2\sigma_j^2}} e^{xj\left(\omega_{ic}i\cos\theta_p+\omega_{jc}j\sin\theta_p\right)} \tag{2}$$

$$i\theta_p = i\cos\left(\theta_p\right) + j\sin\left(\theta_p\right) \tag{3}$$

$$j\theta_p = i\sin\left(\theta_p\right) + j\sin\left(\theta_p\right) \tag{4}$$

Step 5: Until all the pixels are evaluated, repeat the steps 3 and 4.

3.2 Feature Extraction Fuzzy Clustering (Fuzzy C Means)

A feature vector has been generated from the preprocessed image using fuzzy clustering approach [38–42]. It is an unsupervised learning method that identifies the similarity of pixel values without knowing the attribute label used in the cluster overlapping of the image pixel [43–48]. Step by step procedure for fuzzy clustering-based feature extraction is stated as follows:

Algorithm 2: Fuzzy clustering-based feature extraction

Step 1: Using the Gabor filter based Gaussian rule, noise is removed from the input mages. It enhances the image for further processing.

Step 2: Read the input preprocessed image superscript and k = 1,2,…, n.

Step 3: Initialize the random weight w^t for each pixel value of the image between [0,1]

Step 4: Evaluate new centroid as $c(t)$, k = 1,2,..,n

$$c(t) = \sum_{k=1}^{n} w^t A^{(k)} \tag{5}$$

Step 5: Update the weight with centroid value and minimize the total weight mean square error using

$$new_w = (w^t,\ c(t)) = \sum_{p=1}^{m}\sum_{k=1}^{n} (w)^p \left|\left|A^k - c(t)\right|\right|^2 \tag{6}$$

Step 6: Each pixel of the image is clustered based on the maximum weight.

Step 7: End.

The extracted computer vision features are listed in Tab. 1. MSE (Mean Square Error) is the measure to represent the image similarity using the intensity of pixel differences between the two images. PSNR (Peak Signal to the Noise Ratio) is for loss evaluation of image quality by focusing on the differences in numbers based on MSE. When the value of MSE is 0, PSNR is also set as 0. SSIM (Structural Similarity Index Measure) evaluates the difference in humans such as contrast, structural data and luminance. Red, Green, Blue (RGB), and Hue, Saturation, and Value (HSV) represent the color in the image space. The histogram represents the hue distribution of the image. Luminance is the brightness of the image. Variance evaluates the image brightness variance. Edge density is the edge component ratio of all pixels. DCT (Discrete Cosine Transform) is the image sharpness. Since deep fake images synthesize the target image, computer vision features are caused by unnatural changes, and deepfake image creation is limited with size transformation and resolution to fit the source image. The feature extraction process is depicted in Fig. 3.

Table 1: Computer vision extracted features using fuzzy clustering

Features	Description
MSE	Squared value of estimated average value and actual value
PSNR	Ratio of the maximum power signal to corrupting noise
SSIM	Perceived cinematic picture and digital television with quality
RGB	Percentages of red, green and blue colors in the image
HSV	Percentage of hue, saturation, and value
Histogram	Presence of Brightness pixels in the image
Luminance	Image of total brightness
Variance	Image variance
Edge density	Ratio of edge to the total image pixel
DCT	DCT image bias

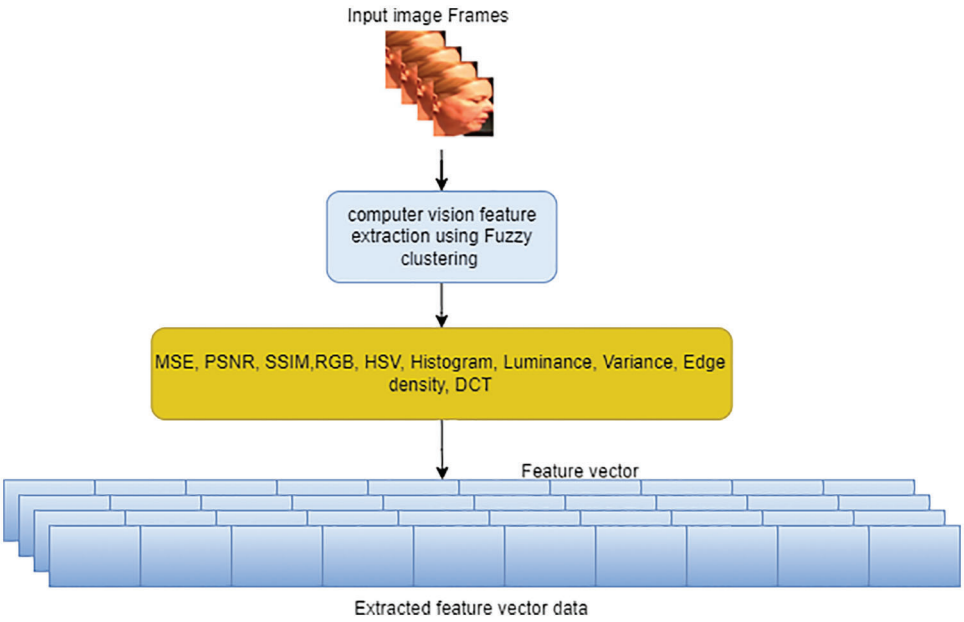


Figure 3: Computer vision features [20]

3.3 Classification: Proposed DBN with Loss Handling using Pairwise Learning

A deepfake technique has used Generative Adversarial Networks (GANs) to generate the fake images. The deep learning-based classifiers are used to detect the fake images. In this proposed face detection, Deep Belief Network (DBN) is used for classifying the fake and natural images from the input dataset. During the training process, the net loss is handled with a feature learning approach to improve the accuracy of the classification process. DBN has been proved to be one of the best machine learning approaches for classification with the capability of handling larger network structures and fast implications. It consists of multi-layer hidden units with one visible layer, and it is assumed to be the generative model. The visual layer transfers the input features to the hidden section for the process based on Restricted Boltzmann Machine (RBM). Each RBM layer is communicated to its previous and subsequent layers. Each RBM also contains restricted visible and hidden layers as sub-layers. The process of activating the visual to the hidden layer is done using sigmoid activation function as in Eq. (7) based on the learning rule of RBM. The RBM architecture is shown in Fig. 4, which has four stacked RBMs. RBM1 consists of a visible section and hidden layers, RBM 2 has a hidden one layer and hidden two layers, RBM3 are the hidden layers 1,2 and 3 and RBM 4 consists of hidden layer three and output layer.

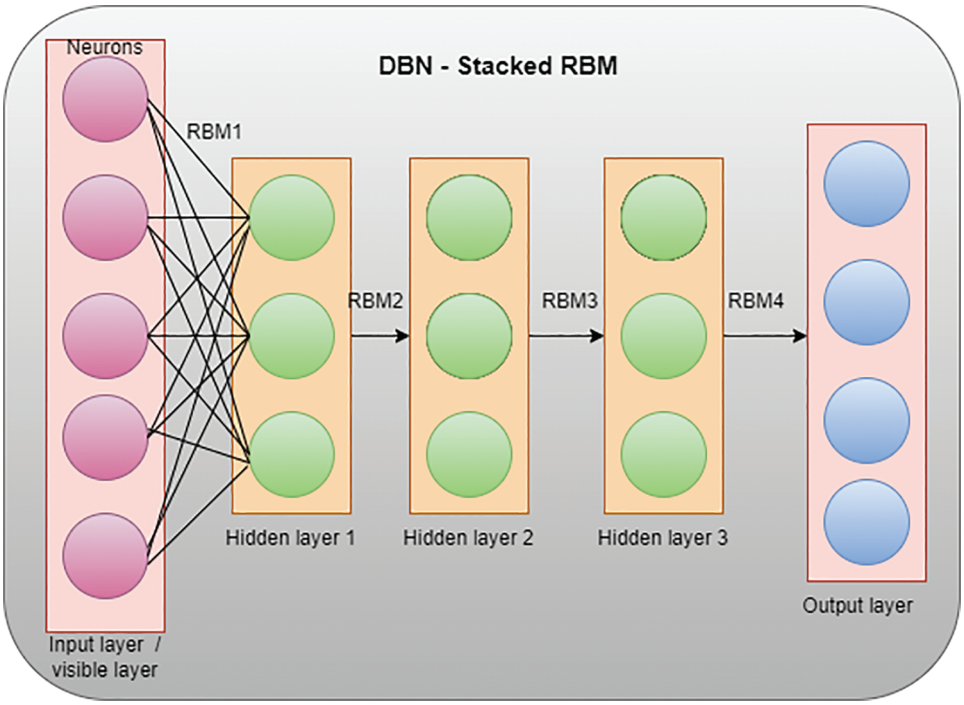


Figure 4: Architecture of DBN- RBM

Training of DBN includes training the RBM with the learning rule with the parameters such as synaptic weigh of the layers, states, and bias of the neurons. The neuron state is formed based on the bias transformation and state of the previous layer neuron weight to the next layer neuron.

$$Pb(st_i = 1) = \frac{1}{1 + \exp(-b_i - \sum_j st_j w_{ij})} \tag{7}$$

Input training data is consisting of two steps as positive and negative. The positive step is responsible for the conversion of visible layer data into hidden layer data. The negative step is responsible for the conversion

of hidden layer data into respective visible layer data. Positive and negative steps of individual activation function are stated as in Eqs. (8) and (9) sequentially.

$$Pb(v_i = 1|hd) = sigm(-b_i - \sum_j hd_j w_{ij}) \tag{8}$$

$$Pb(hd_i = 1|v) = sigm(-c_i - \sum_j hd_j w_{ij}) \tag{9}$$

The parameters of weight are optimized until the arrival of maximum training epochs using the Eq. (10).

$$W' = update\left(w_{ij} + \frac{\eta}{2} \times \left(positive\left(Ed_{ij}\right) - negative\left(Ed_{ij}\right)\right)\right) \tag{10}$$

where,

$positive\left(Ed_{ij}\right)$-Positive statistics of edge $Ed_{ij} = (hd_j = 1|v)$

$negative\left(Ed_{ij}\right)$-Positive statistics of edge $Ed_{ij} = P(vd_j = 1|hd)$

η-learning rate

The mentioned training process is for one RBM. The same training process is executed till all the RBM training processes get over. The parameters of weight are optimized with the loss function by adding contrastive loss to enhance the proposed classification system performance by introducing pairwise learning approach to train the RBM. The pairwise inputs are added to the network using Siamese network structure as shown in Fig. 5.

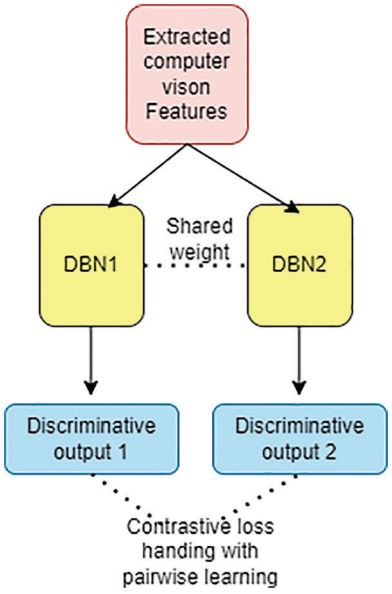


Figure 5: Proposed DBN pairwise learning based on Siamese network structure and contrastive loss

Contrastive loss is added to the weight updating function. For the given image feature pair (f1,f2) and the pairwise label called p where p = 0 denotes the imposter pair and p = 1 denotes the genuine pair. The energy function between the networks is declared as in Eq. (11).

$$E_{W'}(f1,f2) = \|f_{DBN1}(f1) - f_{DBN2}(f2)\|_2^2 \tag{11}$$

The energy function is minimized with constant mapping by calculating the l2 norm distance between the networks. This constant mapping increases the net loss and reduces the loss, the contrastive loss is introduced as shown in Eq. (12).

$$L(W', (Pb,f1,f2)) = 0.5 \times \left(p_{ij}E_{W'}^2\right) + \left(1 - p_{ij}\right) \times \max(0, (m - E_{W'})_2^2 \tag{12}$$

where m is the threshold value. If the input is genuine, the cost function reduces the energy and loss of the network. If the input is impostor, the loss minimizes the function $max(0, (m - E_{W'})$. The energy is maximized if the network feature distance is small for impostor pair and m. Thus, the fake images are trained. The fake images are classified while $f_{DBN}(f1)$ is similar to $f_{DBN}(f2)$ at $p_{ij} = 1$. Therefore, the network is iteratively trained with this contrastive loss to enhance the detection process with an improved accuracy.

4 Performance Evaluation

This section discusses the experimental evaluation of the proposed system with the hyperparameter by changing the optimizer. The dataset used for this evaluation is discussed as follows:

4.1 Dataset Used

Three datasets were used for evaluation, such as Face2Face and FaceSwap provided by FAceForensics++. 100 K faces consisted of 100,000 human face images generated using StyleGAN and DeepFake detection challenge dataset from Kaggle. The size of the dataset was 470 GB. The dataset characters were collected from various genders, races, and shooting circumstances. Face2Face dataset consisted of deep fake videos, and in this experimental study, 205 videos were used. From FaceSwap, 210 videos were used. From the videos, face images were extracted using MTCNN. To extract the computer vision features from the extracted faces, python library called OpenCV was used. The proposed network was implemented using Python 3 to train the machine learning models Keras used.

4.2 Evaluation

The proposed FC-DBNPL detection system is evaluated with the comparison between Mesonet using CNN method and SVM in terms of accuracy stated in Eq. (13). Tab. 2 shows the evaluated results.

$$accuracy = \frac{TP + TN}{TP + TN + FP + FN} \times 100 \tag{13}$$

Table 2: Comparison of deepfake detection systems

Detection models	Face2Face	Faceswap	100KFaces	DFDC
Mesonet with CNN	91.2	93.23	91.02	93.15
SVM	56.2	53.4	67.3	59.2
Proposed model	97.54	96.75	98.1	96.26

The accuracy percentage of Mesonet with CNN for all the datasets are around 90% of detection. SVM has obtained the accuracy of less than 70%. The proposed system has obtained the improved accuracy of approximately 98% for all the datasets. Tab. 3 shows the accuracy of the proposed model by changing

the optimizer and the number of hidden layers to decide the respective hyperparameter. The highest accuracy of 97.54% has been obtained while implementing the proposed FC-DBNPL with Adam optimizer. While increasing the number of hidden layers, accuracy rises gradually.

Table 3: Proposed system performance to find hyperparameters

Optimizer	No of hidden layers	Network loss	Accuracy
SGD	3	0.5643	68.43
	5	0.4323	75.34
	7	0.3123	83.42
AdaGrad	3	0.6423	64.23
	5	0.6612	68.01
	7	0.6341	71.91
Adam	3	0.1432	93.64
	5	0.0651	97.54
	7	0.1028	95.08

The evaluation in terms of precision and recall for all the detectors with various dataset is shown in Tab. 4. This comparative analysis proves that the proposed FC-DBNPL is significantly better for all the datasets than the other approaches. With the implementation of pairwise learning, the loss is handled, which improves the performance of the detection system. The ROC comparison of the evaluated models is shown in Fig. 6.

Table 4: Performance comparison of proposed and other deep fake detector approaches

Detection models	Face2Face		Faceswap		100KFaces		DFDC	
	Precision	Recall	Precision	Recall	Precision	Recall	Precision	Recall
Mesonet with CNN	0.769	0.789	0.781	0.81	0.7823	0.7912	0.7682	0.7234
SVM	0.6725	0.6682	0.5921	0.6729	0.7102	0.6992	0.7231	0.6872
Proposed model	0.9231	0.9143	0.9462	0.9374	0.9457	0.9413	0.9248	0.9102

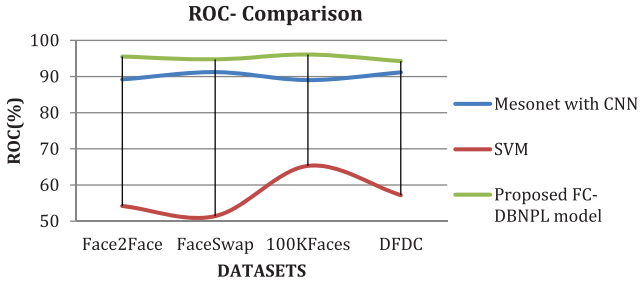


Figure 6: ROC comparison of deep fake detection systems

The performance comparison. Fig. 5 shows that the proposed system has obtained improved ROC values for the datasets such as Face2Face, FaceSwap, 100KFaces and DFDC as 95.54%, 94.75%, 96.1% and 94.26%, respectively. The other approaches such as Mesonet and SVM have secured 89.2%, 91.23%, 89.02%, 91.15% and 54.2%, 51.4%, 65.3% and 57.2% respectively. The error detection and computation time comparisons are shown in Figs. 7 and 8.

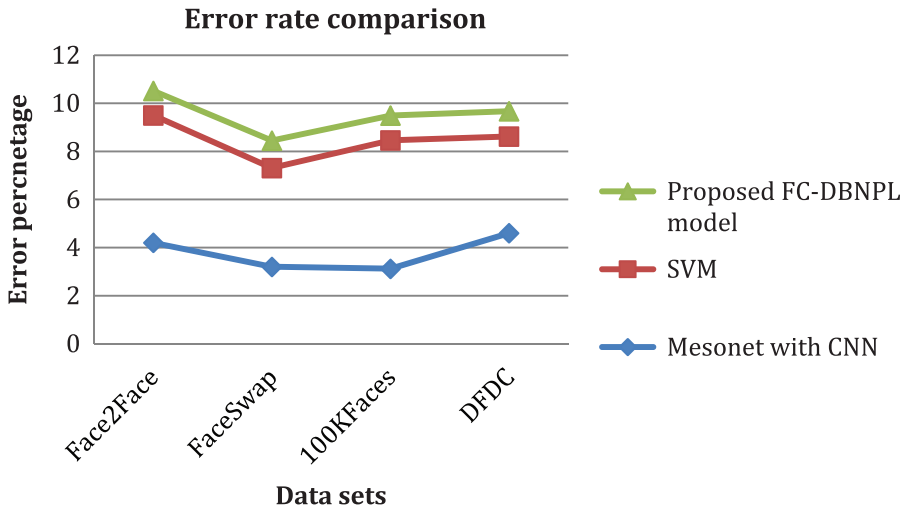


Figure 7: Error rate comparison of proposed *vs.* existing deep fake detection systems

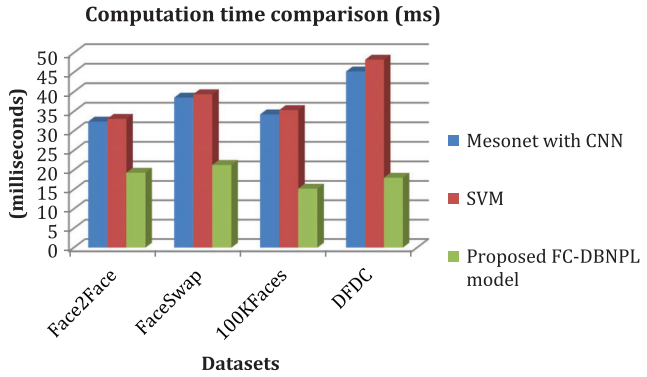


Figure 8: Computation time comparison of proposed *vs.* existing deep fake detection systems

The evaluation results of the error rate Fig. 8 show that the proposed system has secured the minimum error rate for all the datasets with the implementation of loss handling approach and has secured the rate for the datasets Face2Face, FaceSwap, 100KFaces and DFDC as 1.02%, 1.14%, 1.031% and 1.051% respectively. Other approaches such as Mesonet and SVM have secured 4.2%, 3.2%, 3.12%, 4.6% and 5.3%, 4.11%, 5.34% and 4.02% respectively. The computation time evaluation, Fig. 7 proves that the proposed system has secured less time than the other approaches. For various datasets such as Face2Face, FaceSwap, 100KFaces and DFDC, the obtained computation timings are 19.32, 21.3, 15.24 and 18.03 ms, respectively. Other approaches such as Mesonet and SVM have secured 32.4, 38.53, 34.23, 45.23 ms and 33.1, 39.4, 35.34, 48.23 ms, respectively. Thus, the proposed FC-DBNPL approach detects the fake images with high accuracy, precision, and recall with minimum error and takes less computation time through all the kinds of evaluation.

5 Conclusion

This paper has proposed a fuzzy clustering-based feature extraction from the input deepfake image. Initially, the input image was preprocessed with Gabor filter-based Gaussian rule to remove the noise and enhanced the image with better quality for classification. Further, the classification method called deep belief network has enhanced with loss handling approach called pairwise learning. Due to the advantage of this methodology, the error rate for the classification is reduced. The proposed FC-DBNPL has obtained the accuracy in detecting the deepfake image with 97.54%, 96.75%, 98.1%, and 96.26% on various datasets such as Face2Face, FaceSwap, 100KFaces and DFDC, respectively. Improved ROC values for the datasets such as Face2Face, FaceSwap, 100KFaces and DFDC using the proposed system are 95.54%, 94.75%, 96.1% and 94.26%, respectively. The proposed system has obtained the minimum error rate for Face2Face, FaceSwap, 100KFaces and DFDC as 1.02%, 1.14%, 1.031% and 1.051%, respectively. For various datasets such as Face2Face, FaceSwap, 100KFaces and DFDC, the obtained computation timings are 19.32, 21.3, 15.24 and 18.03 ms, respectively. Hence, compared to another DeepFake detection system, the proposed fuzzy clustering-based deep learning approach called FC-DBNPL has improved the accuracy in detecting the deep fake images from the real images. In the future, the proposed system can be extended to detect deepfake using edited audio and video by changing the speech of the person with lip sing. The proposed method can also be applied to larger datasets.

Acknowledgement: Princess Nourah bint Abdulrahman University Researchers Supporting Project number (PNURSP2022R54), Princess Nourah bint Abdulrahman University, Riyadh, Saudi Arabia.

Funding Statement: Princess Nourah bint Abdulrahman University Researchers Supporting Project number (PNURSP2022R54), Princess Nourah bint Abdulrahman University, Riyadh, Saudi Arabia.

Conflicts of Interest: The authors declare that they have no conflicts of interest to report regarding the present study.

References

[1] G. Lee and M. Kim, "Deepfake detection using the rate of change between frames based on computer vision," *Sensors*, vol. 21, no. 3, pp. 1–14, 2021.

[2] J. Thies, M. Zollhofer, M. Stamminger, C. Theobalt and M. Nießner, "Face2face: Real-time face capture and reenactment of RGB videos," in *Proc. IEEE Conf. on Computer Vision and Pattern Recognition*, San Juan, USA, pp. 2387–2395, 2016.

[3] I. Korshunova, W. Dambre and L. Theis, "Fast face-swap using convolutional neural networks," in *Proc. IEEE Int. Conf. on Computer Vision*, Cambridge, USA, pp. 3677–3685, 2017.

[4] A. Tewari, M. Zollhoefer, F. Bernard, P. Garrido, H. Kim *et al.,* "High-fidelity monocular face reconstruction based on an unsupervised model-based face autoencoder," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 42, no. 2, pp. 357–370, 2020.

[5] J. Lin, "FPGAN: Face de-identification method with generative adversarial networks for social robots," *Neural Networks*, vol. 133, no. 3, pp. 132–147, 2021.

[6] R. Chesney and D. Citron, "Deepfakes and the new disinformation war: The coming age of post-truth geopolitics," *Foreign Affairs*, vol. 13, no. 3, pp. 1–14, 2019.

[7] S. Lyu, "Deepfake detection: Current challenges and next steps," in *Proc. IEEE Int. Conf. on Multimedia & Expo Workshops (ICMEW)*, London, United Kingdom, pp. 1–6, 2020.

[8] M. T. Jafar, M. Ababneh, M. A. Zoube and A. Elhassan, "Forensics and analysis of deepfake videos," in *Proc. 11th Int. Conf. on Information and Communication Systems (ICICS)*, Jorden, pp. 53–58, 2020.

[9] M. A. Younus and T. M. Hasan, "Effective and fast deepfake detection method based on haar wavelet transform," in *Proc. Int. Conf. on Computer Science and Software Engineering (CSASE)*, Kurdistan Region, Iraq, pp. 186–190, 2020.

[10] D. Afchar, V. Nozick, J. Yamagishi, I. Echizen and A. MesoNet, "Compact facial video forgery detection network," in *Proc. IEEE Int. Workshop on Information Forensics and Security (WIFS)*, Hong Kong, China, pp. 1–7, 2018.

[11] Y. Li, M. Chang and S. Lyu, "Exposing AI created fake videos by detecting eye blinking," in *Proc. IEEE Int. Workshop on Information Forensics and Security (WIFS)*, Hong Kong, China, pp. 1–7, 2018.

[12] D. Güera and E. J. Delp, "Deepfake video detection using recurrent neural networks," in *Proc. 15th IEEE Int. Conf. on Advanced Video and Signal Based Surveillance (AVSS)*, Auckland, New Zealand, pp. 1–6, 2018.

[13] S. Agarwal, H. Farid, O. Fried and M. Agrawala, "Detecting deep-fake videos from phoneme-viseme mismatches," in *Proc. IEEE/CVF Conf. on Computer Vision and Pattern Recognition Workshops (CVPRW)*, Seattle, WA, USA, pp. 2814–2822, 2020.

[14] L. Zheng, S. Duffner, K. Idrissi, C. Garcia and A. Baskurt, "Siamese multi-layer perceptrons for dimensionality reduction and face identification," *Multimedia Tools and Applications*, vol. 75, no. 9, pp. 5055–5073, 2016.

[15] H. R. Hasan and K. Salah, "Combating deepfake videos using blockchain and smart contracts," *IEEE Access*, vol. 7, no. 12, pp. 41596–41606, 2019.

[16] B. Bayar and M. C. Stamm, "A deep learning approach to universal image manipulation detection using a new convolutional layer," in *Proc.4th ACM Workshop Inf. Hiding and Multimedia Security*, New York, United States, pp. 5–10, 2016.

[17] J. Esther and M. M. Sathik, "An analytical study on query integration in image retrieval system," *International Journal of Advanced Research in Computer Science*, vol. 3, no. 2, pp. 1–15, 2012.

[18] G. E. Hinton, S. Osindero and Y. W. Teh, "A fast learning algorithm for deep belief nets," *Neural Computation*, vol. 18, no. 3, pp. 1527–1554, 2006.

[19] S. Chopra, R. Hadsell and Y. LeCun, "Learning a similarity metric discriminatively, with application to face verification," in *Proc. IEEE Conf. on Computer Vision and Pattern Recognition*, San Diego, CA, USA, vol. 1, pp. 539–546, 2005.

[20] L. Nataraj, T. Mohammed, B. Manjunath, S. Chandrasekaran, A. Flenner *et al.,* "Detecting GAN generated fake images using co-occurrence matrices," *Electronic Imaging*, vol. 3, no. 5, pp. 1–7, 2019.

[21] S. Mahajan, A. Raina, M. Abouhawwash, X. Gao and A. K. Pandit, "COVID-19 detection from chest X-Ray images using advanced deep learning techniques," *Computers, Materials and Continua*, vol. 70, no. 1, pp. 1541–1556, 2022.

[22] V. Kandasamy, P. Trojovský, F. Machot, K. Kyamakya, N. Bacanin *et al.,* "Sentimental analysis of COVID-19 related messages in social networks by involving an N-gram stacked autoencoder integrated in an ensemble learning scheme," *Sensors*, vol. 21, no. 22, pp. 7582, 2021.

[23] M. Abouhawwash and A. Alessio, "Develop a multi-objective evolutionary algorithm for pet image reconstruction: Concept," *IEEE Transactions on Medical Imaging*, vol. 40, no. 8, pp. 2142–2151, 2021.

[24] M. Abouhawwash, "Hybrid evolutionary multi-objective optimization algorithm for helping multi-criterion decision makers," *International Journal of Management Science and Engineering Management*, vol. 16, no. 2, pp. 94–106, 2021.

[25] M. Abdel-Basset, R. Mohamed, M. Abouhawwash, R. K. Chakrabortty and M. J. Ryan, "EA-MSCA: An effective energy-aware multi-objective modified sine-cosine algorithm for real-time task scheduling in multiprocessor systems: Methods and analysis," *Expert Systems with Applications*, vol. 173, no. 4, pp. 114699, 2021.

[26] K. Deb, M. Abouhawwash and J. Dutta, "Evolutionary multi-criterion optimization: 8th International conference," in *EMO 2015, Proc., Part II, Springer International Publishing*, Cham, Guimarães, Portugal, pp. 18–33, 2015.

[27] A. Nayyar, S. Tanwar and M. Abouhawwash, *Emergence of cyber physical system and IoT in smart automation and robotics: Computer engineering in automation*. Springer, USA, 2021.

[28] A. Garg, A. Parashar, D. Barman, S. Jain, D. Singhal *et al.,* "Autism spectrum disorder prediction by an explainable deep learning approach," *Computers, Materials & Continua*, vol. 71, no. 1, pp. 1459–1471, 2022.

[29] A. H. El-Bassiouny, M. Abouhawwash and H. S. Shahen, "New generalized extreme value distribution and its bivariate extension," *International Journal of Computer Applications*, vol. 173, no. 3, pp. 1–10, 2017.

[30] A. H. El-Bassiouny, M. Abouhawwash and H. S. Shahen, "Inverted exponentiated gamma and its bivariate extension," *International Journal of Computer Application*, vol. 3, no. 8, pp. 13–39, 2018.

[31] A. H. El-Bassiouny, H. S. Shahen and M. Abouhawwash, "A new bivariate modified weibull distribution and its extended distribution," *Journal of Statistics Applications & Probability*, vol. 7, no. 2, pp. 217–231, 2018.

[32] M. Abouhawwash and M. A. Jameel, "KKT proximity measure versus augmented achievement scalarization function," *International Journal of Computer Applications*, vol. 182, no. 24, pp. 1–7, 2018.

[33] H. S. Shahen, A. H. El-Bassiouny and M. Abouhawwash, "Bivariate exponentiated modified Weibull distribution," *Journal of Statistics Applications & Probability*, vol. 8, no. 1, pp. 27–39, 2019.

[34] M. Abouhawwash and M. A. Jameel, "Evolutionary multi-objective optimization using benson'skarush-kuhn-tucker proximity measure," in *Int. Conf. on Evolutionary Multi-Criterion Optimization*, East Lansing, Michigan, USA, Springer, pp. 27–38, 2019.

[35] M. Abouhawwash, M. A. Jameel and K. Deb, "A smooth proximity measure for optimality in multi-objective optimization using benson's method," *Computers & Operations Research*, vol. 117, no. 2, pp. 104900, 2020.

[36] M. Masud, G. Gaba, K. Choudhary, M. H.ossain, M. Alhamid *et al.,* "Lightweight and anonymity-preserving user authentication scheme for IoT-based healthcare," *IEEE Internet of Things Journal*, vol. 24, no. 2, pp. 1–12, 2021.

[37] A. Ali, H. A. Rahim, J. Ali, M. F. Pasha, M. Masud *et al.,* "A novel secure blockchain framework for accessing electronic health records using multiple certificate authority," *Applied Sciences*, vol. 11, no. 21, pp. 1–14, 2021.

[38] M. Abouhawwash, K. Deb and A. Alessio, "Exploration of multi-objective optimization with genetic algorithms for PET image reconstruction," *Journal of Nuclear Medicine*, vol. 61, no. 3, pp. 572, 2020.

[39] M. Masud, G. S. Gaba, K. Choudhary, R. Alroobaea and M. S. Hossain, "A robust and lightweight secure access scheme for cloud-based e-healthcare services," *Peer-to-Peer Networking and Applications*, vol. 14, no. 3, pp. 3043–3057, 2021.

[40] M. Masud, G. Gaba, S. Alqahtani, G. Muhammad, B. Gupta *et al.,* "A lightweight and robust secure key establishment protocol for internet of medical things in COVID-19 patients care," *IEEE Internet of Things Journal*, vol. 8, no. 21, pp. 15694–15703, 2021.

[41] S. Ibrahim, H. Alhumyani, M. Masud, S. S. Alshamrani, O. Cheikhrouhou *et al.,* "Framework for efficient medical image encryption using dynamic S-boxes and chaotic maps," *IEEE Access*, vol. 8, no. 13, pp. 160433–160449, 2020.

[42] M. Rawashdeh, M. Zamil, S. M. Samarah, M. Obaidat and M. Masud, "IoT-based service migration for connected communities," *Computers & Electrical Engineering*, vol. 96, no. 2, pp. 1–10, 2021.

[43] M. Abouhawwash and K. Deb, "Karush-kuhn-tucker proximity measure for multi-objective optimization based on numerical gradients," in *Proc. of the 2016 on Genetic and Evolutionary Computation Conf. Companion*, Denver, USA, pp. 525–532, 2016.

[44] Y. Wang, J. Ma, A. Sharma, P. K. Singh, G. Singh *et al.,* "An exhaustive research on the application of intrusion detection technology in computer network security in sensor networks," *Journal of Sensors*, vol. 2021, no. 3, pp. 1–11, 2021.

[45] M. Masud, M. Alazab, K. Choudhary and G. S. Gaba, "3P-SAKE: Privacy preserving and physically secured authenticated key establishment protocol for wireless industrial networks," *Computer Communications*, vol. 175, no. 4, pp. 82–90, 2021.

[46] M. AbdelBasset, R. Mohamed, M. Abouhawwash, R. K. Chakrabortty and M. J. Ryan, "A simple and effective approach for tackling the permutation flow shop scheduling problem," *Mathematics*, vol. 9, no. 3, pp. 270–282, 2021.

[47] P. Singh, M. Masud, M. S. Hossain and A. Kaur, "Cross-domain secure data sharing using blockchain for industrial IoT," *Journal of Parallel and Distributed Computing*, vol. 156, no. 3, pp. 176–184, 2021.

[48] S. M. M. Rahman, M. Masud, M. A. Hossain, A. Alelaiwi, M. M. Hassan *et al.,* "Privacy preserving secure data exchange in mobile P2P cloud healthcare environment," *Peer-to-Peer Networking and Applications*, vol. 9, no. 5, pp. 894–909, 2016.