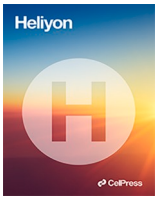
Heliyon 10 (2024) e37163

Contents lists available at ScienceDirect

Heliyon

journal homepage: www.cell.com/heliyon



Research article

Detection of real-time deep fakes and face forgery in video conferencing employing generative adversarial networks

Sunil Kumar Sharma [a,*], Abdullah AlEnizi [e,**], Manoj Kumar [b,d], Osama Alfarraj [c], Majed Alowaidi [e]

[a] Department of Information System, College of Computer and Information Sciences, Majmaah University, Majmaah, 11952, Saudi Arabia
[b] School of Computer Science, University of Wollongong in Dubai, Dubai Knowledge Park, Dubai, United Arab Emirates
[c] Computer Science Department, Community College, King Saud University, Riyadh, 11437, Saudi Arabia
[d] MEU Research Unit, Middle East University, Amman, 11831, Jordan
[e] Department of Information Technology, College of Computer and Information Sciences, Majmaah University, Majmaah, 11952, Saudi Arabia

ARTICLE INFO

Keywords:
Deep fake detection
Deep conditional generative adversarial networks
Ensemble discriminators

ABSTRACT

As facial modification technology advances rapidly, it poses a challenge to methods used to detect fake faces. The advent of deep learning and AI-based technologies has led to the creation of counterfeit photographs that are more difficult to discern apart from real ones. Existing Deep fake detection systems excel at spotting fake content with low visual quality and are easily recognized by visual artifacts. The study employed a unique active forensic strategy Compact Ensemble-based discriminators architecture using Deep Conditional Generative Adversarial Networks (CED-DCGAN), for identifying real-time deep fakes in video conferencing. DCGAN focuses on video-deep fake detection on features since technologies for creating convincing fakes are improving rapidly. As a first step towards recognizing DCGAN-generated images, split real-time video images into frames containing essential elements and then use that bandwidth to train an ensemble-based discriminator as a classifier. Spectra anomalies are produced by up-sampling processes, standard procedures in GAN systems for making large amounts of fake data films. The Compact Ensemble discriminator (CED) concentrates on the most distinguishing feature between the natural and synthetic images, giving the generators a robust training signal. As empirical results on publicly available datasets show, the suggested algorithms outperform state-of-the-art methods and the proposed CED-DCGAN technique successfully detects high-fidelity deep fakes in video conferencing and generalizes well when comparing with other techniques. Python tool is used for implementing this proposed study and the accuracy obtained for proposed work is 98.23 %.

1. Introduction

One of modern society's rising challenges is the prevalence of deep fakes. Face tampering is a significant problem for the safety of the global community and the reliability of human-biometric authentication and identification systems [1]. Deep fake replaces a

* Corresponding author.
** Corresponding author. College of Computer and Information Sciences, Majmaah University, Majmaah, 11952, Saudi Arabia.
E-mail addresses: s.sharma@mu.edu.sa (S.K. Sharma), aalenizi@mu.edu.sa (A. AlEnizi), wss.manojkumar@gmail.com (M. Kumar), oalfarraj@ksu.edu.sa (O. Alfarraj), m.alowaidi@mu.edu.sa (M. Alowaidi).

https://doi.org/10.1016/j.heliyon.2024.e37163
Received 26 March 2024; Received in revised form 16 August 2024; Accepted 28 August 2024
Available online 29 August 2024

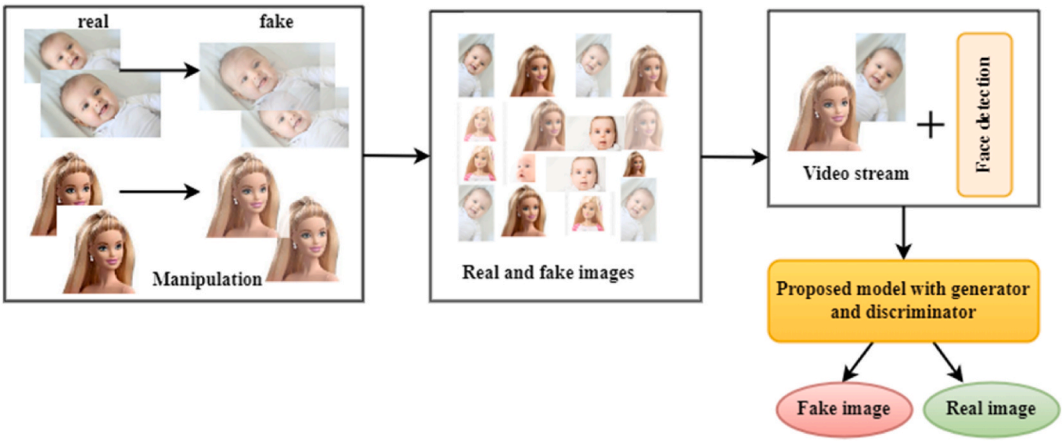


Fig. 1. Model for generating and recognizing deep fakes.

human face in a picture or video with a digitally created one. Deep fakes refer to realistic fake images that could be formed through computational modification [2]. There are various false media recognition techniques; however, the recent advance of unreal data technology has rendered the current analyzers useless [3]. As a result, the need for better technologies to identify fake accounts in films is pressing. In addition, auto-encoders and GANS have increased the prevalence of disingenuous video and picture content on the web [4]. The technology relying on GAN makes it easier for amateurs to create convincing fake face videos. Hence, hoax films quickly propagate over the web [5].

GANs and VAEs (Variational Autoencoders) are two examples of the sophisticated generative algorithms used to create convincing fake pictures and movies. Now that these tools are widely available and highly precise, it's next to impossible to tell a fake from the actual footage [6]. For a long time, video-deep fake identification has been accomplished with the help of Neural Networks (NNs), with the most significant results achieved with algorithms [7]. The classifier in the GAN architecture serves as an error function, giving the generators an accurate training signal to refine their settings. For the classifier in modern state-of-the-art GANs, deep Convolutional Neural Networks (CNNs) are commonly utilized, with the networks being trained by gradient descent optimization methods [8]. Typically, a CNN will have many convolutional, activation, and max pooling built on top of one another to help extract discriminative features for the recognition job. However, given that most algorithms have become increasingly precise in synthesizing highly realistic faces, this research focuses on deep fake detection of face images in video conferencing [9].

Fig. 1 displays an example of a phoney picture made with one of these applications. If given randomly, it's easier to tell which images are real or fake [10]. Anybody, regardless of age or level of education, this create phoney pictures. In light of these considerations, the study of how to identify deep fakes emerged as an excellent new topic of study [11]. It's not easy, and it still needs to be addressed. Deep fake enables the production of phony-convincing movies by replacing a person's facial in an existing clip with a different individual or by modifying facial features such as the eyes, mouth, and eyebrows to create a convincing illusion from another setting [12].

In the article, Deep Conditional Generative Adversarial Networks (DCGANs) have been tremendously successful in a wide range of information tasks, still experiencing significant difficulties in producing aesthetically accurate representations without the model collapsing or becoming unstable during GAN training. One widely recognized reason for the model collapse and instability is that the discriminator focuses on the most discriminative difference between actual and synthesized pictures while ignoring the less racist and discriminatory components, providing a low training signal to the generators [13,14]. Making DCGANs effective requires a significant challenge of getting the discriminator to have a more substantial representation so that a more accurate learning signal can be supplied to the generator to improve it. As a solution, a unique ensemble learning loss function and compact deep ensemble to construct a discriminator for GANs [15]. They are building a conditional GAN employing compact ensemble-based discriminators architecture and evaluating significant parameters, including non-convergence, mode collapse, and decreased gradients. Existing approaches for identifying deep fakes are insufficient because they cannot adapt to the constant development of deep fake generating techniques. Most of these technologies rely on static image analysis or traditional machine learning algorithms, which are inadequate when faced with excellent fake images and lack the ability to generalize to different types of modified information. Furthermore, most current solutions lack the ability to handle data that streams, which is a critical feature in real-time videotelephony. Filling these gaps is made possible by combining CED-DCGAN, which aims for real-time detection while adding feature analysis. CED-DCGAN outperforms current approaches for deep fake detection in video conferencing. It solves the problem of high-quality deepfakes and provides real-time analysis capabilities. It is possible that the proposed approach is effective against the deep fake creation methods employed in the study, but more study could be required to prove the approach's effectiveness against a larger range of deep fake generation approaches. CED-DCGAN can improve video conferencing security on platforms by accurately identifying deep fakes in real time. This can help to avoid imitations, safeguard user privacy, and assure participant authenticity. The study introduced a tight ensemble discriminator for assessing the generators from various angles. Many supplementary primarily categorized are integrated into a single deep model, taking cues from ensemble methods. An innovative ensemble loss function is created to guarantee proper coordination throughout

Table 1
List of abbreviation.

Abbreviation	Details
CNN	Convolutional Neural Network
3DCNN	3D convolutional neural network
DICNN	Dual-Input Convolutional Neural Network
GAN	Generative Adversarial Network
AGAN	Attention-Based Generative Adversarial Network
DCGANS	Deep Convolutional Generative Adversarial Network
CED-DCGAN	Compact Ensemble-based discriminators architecture using Deep Conditional Generative Adversarial Networks
CED	Compact Ensemble discriminator
CelebA-HQ	CelebFaces Attributes High Quality
MSE	Mean Squared Error
PIFR-GAN	Pose Invariant Face Recognition - Generative Adversarial Networks

training across the whole framework. While the results of a research may work well in specific situations or datasets, their application to detecting new or evolving kinds of deep fakes necessitates constant validation and adaption. Addressing these problems with strong evaluation frameworks, constant model improvement, and interdisciplinary collaboration might improve the generalizability as well as practical utility of deep fake detection systems in a variety of environments and applications.

The main components of the article are as follows.

1. This investigation used an innovative active forensic method by Compact Ensemble-based discriminators architecture trained on Deep Conditional Generative Adversarial Networks (CED-DCGAN) to detect deep fakes in real-time during video conferences
   2.Split real-time video pictures into frames with essential parts and utilize that bandwidth to train a compact ensemble-based discriminator classifier to recognize DCGAN-generated images.
3. The suggested method detects the high-fidelity, accuracy, and sensitivity of deep fakes in video conferencing and generalizes effectively to other GANs utilizing similar refined procedures.

The rest of the article is organized as follows: section 2 goes on the literature. The approach and information sources of deep fake detection and face forgery are then discussed in section 3. section 4 offers experimental data and analyses. Lastly, section 5 offers the conclusion and suggestions for further study.

1.1. Motivation

The fast development of video conferencing systems has transformed communication by making distant interactions more fluid and accessible. However, technological innovation has also resulted in considerable issues, notably in terms of security. One of the most important issues is the rise of deep fakes and face forging, in which sophisticated algorithms are utilized to modify video information, potentially leading to serious consequences. Identifying deep fakes as well as face forgeries during real time presents novel challenges. Traditional approaches sometimes fail due to the complex nature of modern forging techniques and the requirement for fast testing. The enormous amount of video data, along with the need for fast processing, complicates the detection process. As a result, there's an urgent need for novel technologies that can accurately and efficiently detect forgeries as they develop. Table 1 shows the list of abbreviation.

2. Literature survey

Liu et al. [16] introduced a unique Convolutional Neural Network (CNN) that uses periodic roughness enhancement and information compression to better characterize materials at different logical levels in pictures, increase resilience, and enhance global textured perceptions. Regrettably, the fundamental issue with existing fake face detection algorithms is that they must generalize to discriminate between various GANs. Nevertheless, results show that our model routinely exceeds the state-of-the-art in both single- and multi-domain image processing tasks, particularly when pictures are affected by Gaussian filters. Kohli et al. [17] proposed a lightweight 3D convolutional neural network (3DCNN) to identify common face-spoofing forms such as DeepFakes, Face2Face, and FaceSwap. For video authenticity verification, the suggested system studies similar interest characteristics. In addition, activation maps are analyzed extensively to learn more about 3DCNN's detection procedure. Furthermore, the widely-used FaceForensic++ dataset is utilized to learn from and test the suggested solution. Finally, the presented technique is tested. Its efficacy is compared to state-of-the-art methods regarding the number of learnable parameters and the precision with which it can recognize binary images. Bhandari et al. [18] presented a ten-fold cross-validation Dual-Input Convolutional Neural Network (DICNN) model to bridge the gap and broaden the network's perspective, we forced the model into 'SHapley Additive exPlanations (SHAP),' which can visually describe the research results and interconnectivity using comprehensible artificial intelligence. Perfectly shaped values, with an average recognition rate of 99.36 0.62 %, a test accuracy of 99.08 0.64 %, and a ground truth of 99.30 0.94 %. Having the suggested model acknowledged by forensics and security specialists is crucial since it has unique characteristics and is far more accurate than state-of-the-art approaches. Fig. 2 shows the architecture of GAN training cycle.

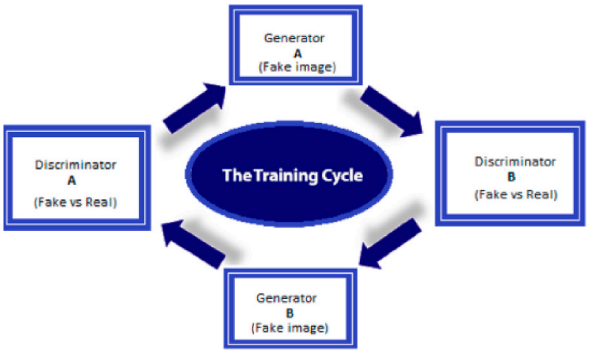


Fig. 2. GAN training cycle.

Chen et al. [19] introduced SS-GAN, or a Semi-Supervised Adversarial Generative Network, is used to avoid image counterfeiting. Then, the effect of anomalies is removed by using a representative frame selection module to filter them out of the training movie. Second, provide a single-frame feature anti-spoofing method that leverages a generative adversarial network (GAN) to better control the variable selection of the neural network via data preprocessing by a generator, thereby resolving the problem of inadequate training data. The results of our experiments show that our method is superior to most manually created characteristics, and it operates similarly to state-of-the-art deep learning methods. Ali et al. [20] proposed an in-network damage segmentation model, and an Attention-Based Generative Adversarial Network (AGAN) was created (IDSNet). Compared to other province networks, the generated IDSNet accuracy of an instrument is well. It has excellent diagnostic accuracy (0.952), a high F1-score (0.941), and fast response (0.942) over a test set, with a mean crossover over union probabilities of 0.900. With AGAN's contribution, IDSNet's mIoU grows by 12 %. As a result of its highly compact separative, IDSNet can analyze 640 x 480 x 3 thermal pictures at 74 frames per second instantaneously, with a total of 0.085 M trainable parameters. Hariharan et al. [21] proposed a generative adversarial network is the Deep Convolutional Generative Adversarial Network (DCGANS), which uses convolution and convolutional-transpose layers as the discriminator and generator. Despite the volume of real-time photos being processed, this DCGANS returns results quickly. Nevertheless, the quality of the photographs used must be reduced to generate these fakes, which could serve as a giveaway that they are fakes. The models in this project were trained with data from Kaggle's CelebA dataset.

Recent advancements in Machine Learning (ML) and Deep Learning (DL) have significantly enhanced various fields by providing innovative solutions and improving efficiency in handling complex tasks. In crime news retrieval, ontology-based systems address the limitations of traditional keyword-based methods by understanding the context and semantics of search queries, thus enhancing the accuracy and relevance of retrieved information [22]. In the medical field, ML and DL have transformed disease detection and diagnosis. Convolutional Neural Networks (CNNs) analyze skin images to identify cancerous lesions with high accuracy, facilitating early detection and effective treatment [23]. Similarly, expert systems integrating ML algorithms analyze patient data to diagnose heart conditions and suggest treatment options, thereby improving diagnostic accuracy and patient care [24]. In software engineering, ML models predict software defects by analyzing historical data, allowing developers to address potential issues early and enhance software reliability [25]. These advancements underscore the transformative impact of ML and DL, offering more accurate, efficient, and automated solutions across multiple domains.

Huang et al. [26] introduced Non-Intrusive Load Monitoring using a Deep Convolutional Generative Adversarial Network for Prediction (NILM-GAN). Propose a technique for modifying the feature space utilizing the EMBED dataset's visual representation and the deep convolutional GAN (DCGAN)—finally, the test presentation by training some basic classifiers and assessing their accuracy using the accuracy score. A series of tests assess the models' flexibility in applying to various devices. On comparing results across different training epoch counts, observe that the average performance is better than 70 % accurate. Wijaya et al. [27] introduced Generative Adversarial Networks (GANs) to identify the distracted driver and his distraction source using a machine learning approach. In this piece, GAN differentiates between distracted and safe driving. Developed the model by amassing a large amount of picture data and then training using various parameters to achieve optimal accuracy. Using the data from the experiment, found that our suggested model can reach D Loss = 0.0391 and G Loss = 5.7638. As a result, our model has the potential to be an excellent answer to the problem of distracted driver identification utilizing a more complex method. Bansal et al. [28] DFN (Deep Fake Network) is an unique design structure that integrates the major aspects of mobNet, such as a continuous stack of separable convolution, max-pooling layers using Swish as a perceptron, and XGBoost as a classifiers. DFN functionality was evaluated using the Deep Fake Detection Challenge (DFDC) dataset. Using this dataset, the suggested approach attained an accuracy of 93.28 % and a precision of 91.03 %. Moreover, the loss during training was 0.14, while the loss during validation was 0.17.

Kas et al. [29] introduce 2D PIFR-GAN (Pose Invariant Face Recognition - Generative Adversarial Networks) that uses GAN for translating images. The proposed Consolidated dataset consists of four data collections that provide identities and their corresponding upfront images. The GAN employed is a Pix2Pix paired architecture spanning multiple generator and discriminator models. The fractalization and categorization subsystems in our proposed architecture, the Combined-PIFR database, are separated with individual constraints in mind. The outcomes have been significantly improved by 33.57 % compared to the baseline, all down to the GAN-based fractalization.

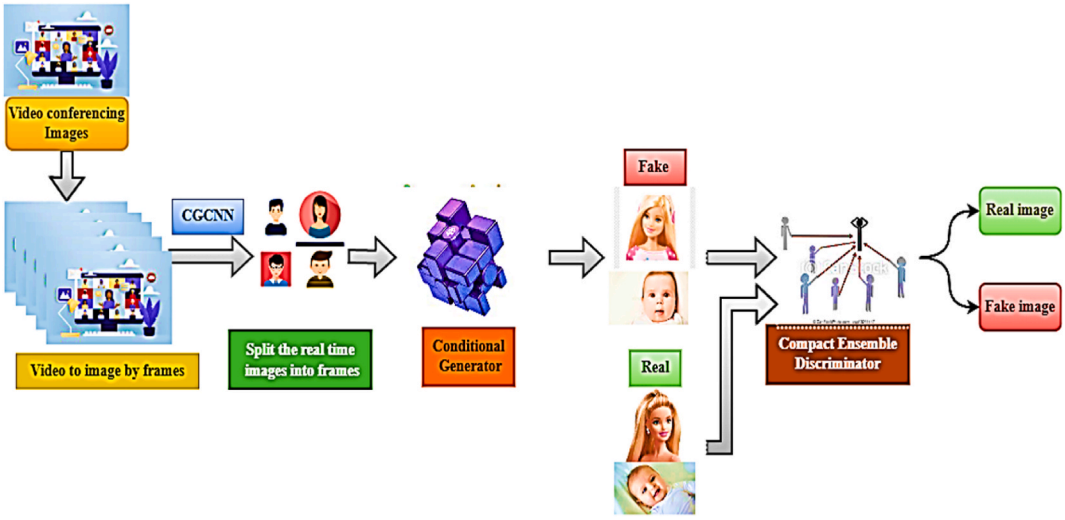


Fig. 3. Structure of the proposed CED-DCGAN model.

Heidari et al. [30] suggested an approach combines SegCaps and CNN approaches to improve visual feature extraction, and then includes capsule network (CN) training for better generalization. A new data normalizing technique is developed to address heterogeneity from several global data sources. Additionally, transfer learning (TL) as well as preprocessing methods are used to improve DL performance. Collaborative global training of models using blockchain and FL ensures data source privacy.

Ahamad et al. [31] suggested a hybrid face detection/object detection-based surveillance video system has been evaluated under extremely rigorous conditions within a cloud-IoT (Internet of Things) incorporated distributive computing environment, and the experimental study predicts that the proposed hybrid system would perform far better in the real world than existing systems. The authors discovered that the cloud-IoT hybrid technique over recording video of suspicious objects within distributed real-time surroundings has high accuracy, low overall error rate, and high average recall rate for both self-generated real-time data sets and benchmark datasets.

2.1. Problem statement

In this study, learning GANs for posture invariant face recognition needs large databases with a wide range of pose variations, which can be time-consuming in terms of data collection and computing resources. The objective of non-invasive load monitoring (NILM) with a deep convolutional GAN is to break down overall electrical consumption into specific appliances without the use of intrusive hardware. One of the main drawbacks is the need for a large amount of high-quality training data that is labeled. One significant concern with Attention-based GANs is the network architecture's increased complexity. Processing three-dimensional data require handling massive amounts of information, resulting in significant increases in computational burden and memory utilization when compared to 2D CNNs. This makes it difficult to deploy 3D CNNs in real-time applications or on devices with low computational power. OULU-NPU dataset is not be well suited to complex attacks such as deepfakes, and it predates popular mask use. Fake-Vs-Real-Faces dataset lacks details about deepfake techniques, which limits generalizability. Subjective difficulty labels can also be unreliable. To overcome such limitations, this proposed study introduced a two novel dataset and novel deep learning model for identifying fake face forgery.

3. System methodology

3.1. Compact ensemble-based discriminators architecture trained on Deep Conditional Generative Adversarial Networks (CED-DCGAN)

The GAN operates on three pillars: first, data to be created using some form of statistical representation, and second, the generator network can learn. It's relatively new to the field of deep learning and makes use of two systems, one of which creates pictures. Validating deep fakes' legitimacy has become increasingly challenging since GAN models are constantly refined. With the rapid expansion of such characteristics and the vulnerabilities identified by hackers, the techniques used to check their reliability could be more effective. In addition, the natural environment sometimes presents challenges, such as inadequate lighting and complicated backdrops. As a result, deep feature learning cannot begin until after the face pictures have been preprocessed to correct for alignment and normalization. A novel model architecture for deep fake video detection is shown in Fig. 3; this design is a realistic and optimal mix of two state-of-the-art models, Dense Net, and Res Net. In terms of both horizontal and vertical computing, sophistication, layer count, etc., as explained above, both approaches have advantages and disadvantages.

Regarding intra-class differences, the study presents a DCGAN-based technique for concurrently learning functional and

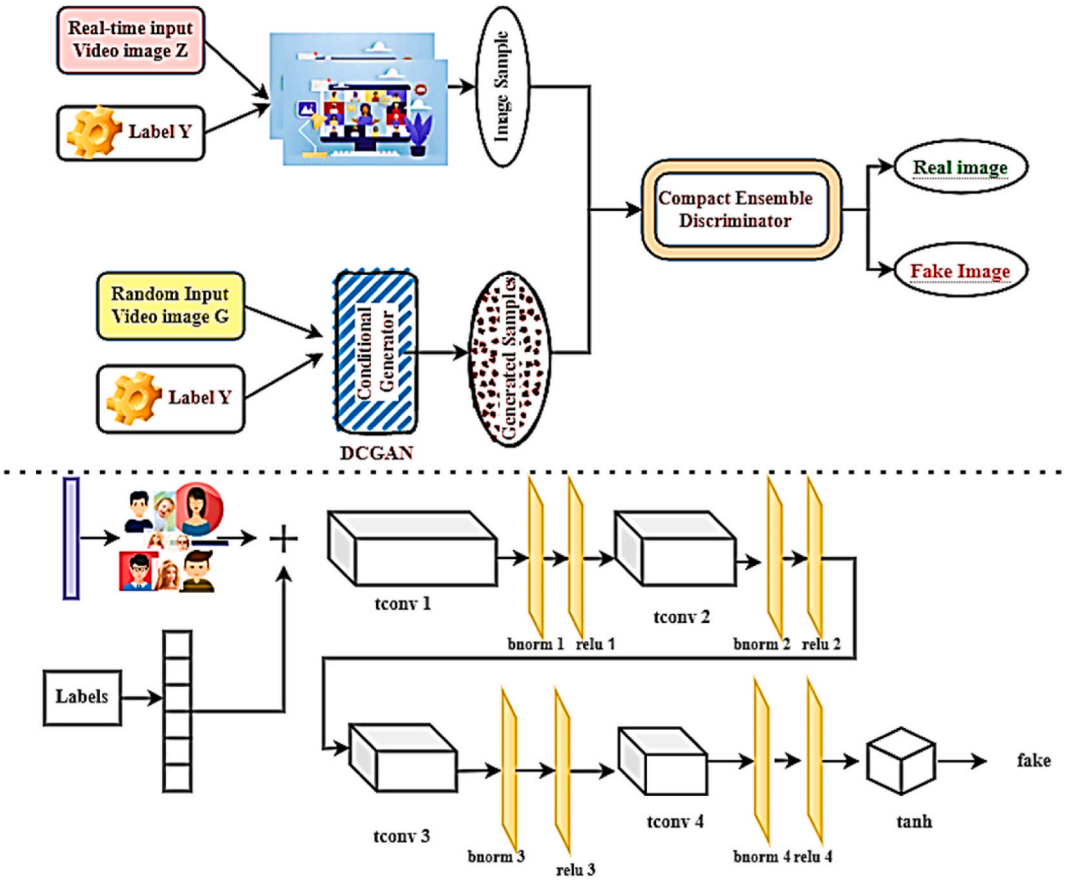


Fig. 4. a) Structure of DCGAN and 4b) Deep Conditional Generator working model.

exclusionary models, which could be utilized to identify fakes. The DCGAN is a variant of the GAN wherein the generator is fed extra information that could be used to exert more direct control over the generator's output. The suggested system comprises a Generator (G) and some discriminators. First, the conditional generator takes a face image as input. Then, it outputs a prototype face image based on a video chat image while maintaining the other features of the input face image.

Given a condition involving action units, the generator G is better able to focus on data that will help it identify Deep fakes and facial forgeries. DCGANs are a sophisticated deep learning model that specializes at producing accurate and manageable data. They extend the basis of GANs by including convolutional layers and conditional inputs. In addition to label synthesis, DCGAN has been used to successfully rebuild objects from edge maps and edit photographs. For DCGANs to be effective, the delivery produced must closely resemble the distribution of actual data. The DCGAN's generator G and ensemble discriminator D compete in multiple combinatorial events to reach this objective. In particular, the discriminator learns to distinguish between genuine and false samples, while the generators know to produce fake models to trick the discriminator.

The CED-DCGAN approach provides a strong defense mechanism for video conferencing against deep fakes. By identifying up sampling anomalies in the frames of a video, which is usually seen in generated content, deep fake can be detected even at its high quality in real-time. Moreover, since it can be applied to a range of contexts beyond video conferencing, it could be highly useful for identifying fake information in other spheres. The process repeats until the two components establish an Equilibrium state. Next, separate the video frames containing the facial area to discard any irrelevant data. First, Cross-Gradient Convolutional Neural Systems (CGCNN) are used to recognize deep fake faces and give the facial area frame since CNN has shown to be resilient and efficient for synchronization. Next, the resized images should be normalized, and the data should be supplemented using methods like a randomized switch. Normalizing the picture inputs has two purposes: first, it gets rid of the excessive frequencies, and second, it makes sure that each of the image's pixels follows the same distribution curve. In addition, the model's efficiency and generalizability benefit significantly from the increased sample size afforded by randomized switching. After carefully considering the characteristics of both models and efficiently combining them, a new model design has been devised that achieves excellent performance with just a slight increase in computing complexity. Besides designing a new architecture, the dataset boosts the model's real-time performance.

The anticipated error rate of a DCGAN based on a loss function is shown in equation 1

$$GD(G,Z) = Epdata(x)logG(x) + Ep(d)(1 - log\ \ log\, Z(G(x)))\ ) \tag{1}$$

G denotes the generator network; Z indicates the discriminator network; data(x) represents the dispersion of actual statistics; p(d) represents the dispersion of generator information; x represents a data point; Z represents a sample from p; Z(x) represents a discriminator network; and G(x) represents the generator system. The first term of the equation is the most excellent possible value of the cross-entropy (E), which is the proper data distribution (data (x)) multiplied by the discriminator (G(x)). CGCNNs are developing as a promising solution that makes advantage of information flow between genuine and modified samples to increase accuracy. This opens the door for DCGANs to play a role. DCGANs, which are known for their forgery creation skills, can be used to build more diversified and realistic deepfakes for training powerful real-time face forgery detection systems.

3.2. Deep conditional GAN

DCGAN structure is a more efficient deep learning methodology, is depicted in Fig. 4a. Several restrictions are placed on the DCGAN's ability to generate pictures. In the case, a new layer whose values come from a single hot-encoded picture. DCGANs are a strong tool for combating video-deep fake detection because of their ability to train and analyze certain aspects of a video. Compared with vanilla GANs, DCGANs can be conditioned on extra information, which is not the case for the former. In the context of deep fake detection, this conditioning information could be the original video of the person or a reference image. This makes it possible for the DCGAN to learn not only the general features of real videos but also to adjust to the subject or scene of the target video. Some of the features include.

- Subtle anomalies include blinking patterns, lip motions which do not match speech, and strange skin textures.
- Temporal artifacts are inconsistencies in light exposure or head posture that happen across video frames.
- Statistical oddities are patterns in levels of noise or color distribution that deviate from real footage.

In this case, a Conditional GAN's Discriminator doesn't get trained to tell the two types of data apart. From there, it learns to accept only really compatible pairings while rejecting any that have a fabricated history. During the training of the network, it noticed significant outcomes in correlated faces was unable to overfit the data. Whenever the generator advances while the discriminator fails repeatedly. Fig. 4b depicts the deep conditional GAN functions. At first, the video conferencing data are split into frames, and by using CGCNN the frames are separated images for the fine-tuning of the images. DCGANs are meant to contain specific data about pictures (y) associated with the conditional generators (CG) and the discriminator (D), which affect the class of the produced output DCGANs are a great option here because of their proven success with unlabeled data.

The inputs in Fig. 4 are scrambled, and a new picture is produced by applying random information (noise) to each. All the data pictures broken up are put through this procedure. Many different kinds of GANs are used to create deep fakes by creating new samples that are convincing simulations of an existing data set. The Deep conditional GAN model combines the generative character of learning with the racist and discriminatory nature of classification to turn low-dimensional random noise into photorealistic pictures. Fig. 4a depicts the DCGAN's generating and discriminator network designs, respectively. Fig. 4b illustrates the 18 layers that make up the generator (G): an input node (noise), a construction and reorganization layer, an encoding layer, an embedding layer, five transcribed convolutional (tconv1-tconv4), four batch normalization layers (bnorm1-bnorm4), four rectified linear unit (ReLU) layers (relu1-rel (tanh). Specifically, the network uses to construct and restructure to transform a 100-sample random stream into a $4 \times 4 \times 256$ array. Then it scales up the matrices using inverted conditional layers, batch normalization layers, and ReLU layers until they reach 128x128x3.In particular, five filters with properly selected numbers and sizes were used to produce 128 x 128 pictures through transpose conditional neural networks. It's important to note that the number of translated conditional neural networks will change depending on the target size. This means that five, there will be only four transposed conditional networks if the target size is 64x 64. The number of epochs is decreased by using regularization in each layer for normalization, reducing computational expenses. 'Tanh' is utilized as the activating feature at the logits layer since it is compatible with the output layer of the network, is represented in equation (2).

$$tanhx = \frac{2}{1 + e^{-2x}} - 1 \tag{2}$$

Tanh activation functions are preferred over sigmoid ones because of the sharper and more constant slopes they produce over time.

A 100-sample random vector is projected onto a restructure layer, where it is then reorganized into a 44x512 matrix. Then, it reformats the category labels into a 4 by 4 array of embedding vectors. The network then combines the pictures from the two inputs along the channel dimension to produce a 4×4x 512 matrix. Finally, it uses a stack of transposed conditional layers, batch normalization, and ReLU layers to scale the output arrays up to 128 by 128 by 3. As summarized, the generator uses a collection of training pictures to teach its network to make convincing false content, while the discriminator determines if an incoming image is genuine or fake. The cost functions of the Deep conditional generator are given in equation 3

$$C_g = min\left[\frac{1}{n}\sum_{i=1}^{n} 1 - log\ \ log\, Z(G(x))\right] \tag{3}$$

Where n is the noise factor in the deep conditional generator(x) represents the generator random input. This DCGAN uses conditional data on the input to drive its feature learning. The collected characteristics are then sent into a CED, which utilizes the strengths of numerous smaller discriminators to perform the final classification.

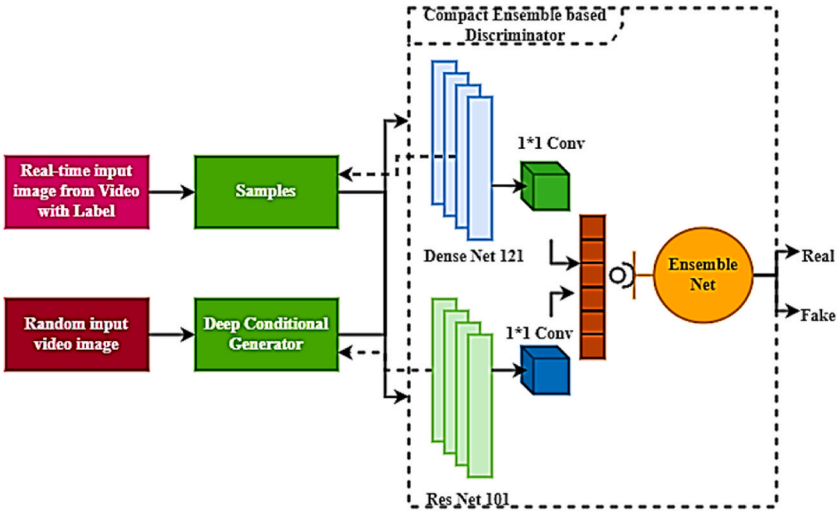


Fig. 5. Compact Ensemble learning-based Discriminator.

3.2.1. *Compact ensemble discriminator*

Fig. 5 depicts the suggested ensemble discriminator structure. The Compact Ensemble Discriminator (CED) is a new architecture proposed within Deep Conditional Generative Adversarial Networks (DCGANs) that was specifically built to handle the challenges of deep fake detection. Compared to traditional discriminators in DCGANs, this often function on a single, monolithic network, the CED employs a collection of smaller, specialized discriminators. Each discriminator in the ensemble identifies a particular group of distinguishing qualities between genuine and synthetic images. This distributed technique enables the CED to detect deepfakes with greater accuracy through utilizing the skills from multiple feature detectors. There are two classifiers and two sources in this architecture. Both the DenseNet and the ResNet are used as discriminators therefore to their excellent potential in a wide variety of tasks. Fully linked layers are applied to these discriminators, combining their resulting image features. One of the generators relies on a static, already-trained network. On the other hand, the deep conditional generator is trained from the beginning using the discriminator's ensemble feedback. The quality and generalizability of an ensemble model's false pictures to be greatly improved by increasing the variety amongst networks in an ensemble discriminator. This was accomplished using two distinct network topologies, Dense Net and Res Net. At a further step, a fully connected layer is linked to the combined feature maps of the two networks' outputs.

The proposed generative adversarial ensemble learning system is comprised of two discriminators (DenseNet and ResNet). The completely connected layers on top of both descriptors combine their increased properties as well as the output confidentially for genuine and fraudulent images. The FaceForensics++ and CelebA-HQ datasets serve as the genuine samples for compact ensembled discriminator training, while face photos created by two different generators serve as the false samples. Here, throughout adversarial learning, one generator uses CelebA-pre-trained HQ's system settings to produce conventional but otherwise interesting fake images for classification methods to evaluate. The other one is taught from scratch using the judgments of the ensembled discriminators. Tricking the discriminators throughout training produce more convincing phoney pictures. The excellent precision of this system makes it ideal for use in the identification of face froggery and false images. ResNet's residual function, when expressed using a shortcut link that skips across several layers, looks as f(x) + x, whereby f(x) and x stand for excess and identification translation, respectively. Developing a very deep conditional network cause disappearing grades, however, using speed connectors can compensate. Integrating DenseNet and ResNet feature maps is necessary for enhanced real-vs-fake picture discrimination. Extraction of 1024-dimensional output features using globally averaged mixing and training sets with filters requires feeding the same picture into both networks and then performing the extraction. The output characteristics of both systems are then concatenated to produce a feature vector of 2048 dimensions. The Discriminator employs two layers and the sigmoid function as an input layer in the output layer to determine if the simulated picture is genuine or not. Construct a channel picture from the output of a final conditional layer using a kernel of size and a hyperbolic tangent function (Tanh). The generator should be spectral-normalization-pretrained on the CelebA-HQ datasets, while the other should be trained from scratch using the adversarial-learning-based discriminators. The cost functions of the deep conditional generator combined with the compact ensemble discriminator are given in equation 4

$$C_d = max\left[\frac{1}{n}\sum_{i=1}^{n} logG(x) + 1 - log\ \ log\, Z(G(x))\right] \tag{4}$$

Cross-entropy loss stimulation of the nonlinear activation type is referred to as sigmoid cross-entropy. Losses in the generator and discriminator are calculated by utilizing the nonlinear activation cross-entropy of the observed logits given in equations (5) and (6)

$$CE_{Loss(G,Z)} = -t_1 logf(s_1) - (1 - t_1)(1 - logf(s_1)) \tag{5}$$

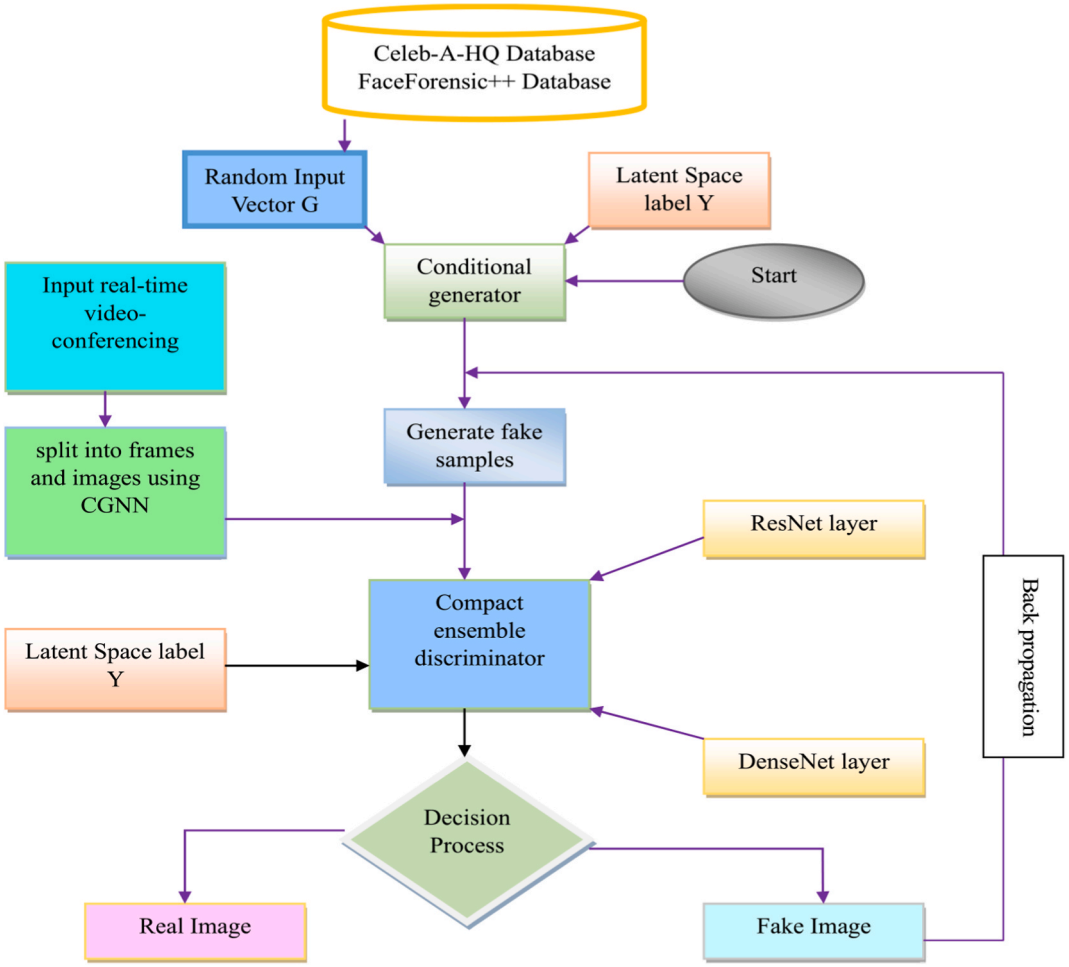


Fig. 6. Flow chart representation of proposed model.

$$f(s)_1 = \frac{1}{1 + e^{-s_1}} \tag{6}$$

Here, $f(s_1)$ represents the scalar factor of the discriminator output. The sensitivity for generator loss is less than or equal to that for discriminator error.

Although the foundation of generative adversarial ensemble learning is in DCGAN, improving detectability is the primary focus of this kind of memorization. In contrast to the goal of traditional GAN approaches, which is to enhance generating capacity, this approach focuses on improving the quality of the generated content.

Pre-processing

Fig. 6 shows the flowchart representation of the CED-DCGAN. The initial step in video preprocessing is to break the video into frames. The face is identified in each frame after the movie is divided into frames, and the frame is truncated into images. The frame size can refer to the resolution or file size of each video frame. This is determined by the input video's resolution. If the provided video is 1920x1080 (Full HD), each frame will be 1920x1080 pixels. Each frame's file size is determined by the format that involves it was saved. Reduced frame rate allows the system to process fewer frames every second, lowering computing load. Develop techniques to detect corrupt frames. Checksums and hash comparisons are two techniques that can help identify corrupted frames. Afterward, the chopped image is turned into a new video by integrating all of the video's frames. The method is repeated for every video, resulting in the generation of a treated dataset comprising only videos with faces. During pretreatment, the image that does not include the face is disregarded. Like with traditional GANs, the first one can create a wide variety of facial pictures thanks to its prior training on the CelebA-HQ dataset.

Training Process

Table 2
Hyperparameter value.

Hyperparameter	Value
Batch Size	64
Learning Rate	0.0002 (for both generator and discriminator)
Optimizer	Adam
β1	0.5
β2	0.999
Epochs	200
Latent Dimension	100
Label Smoothing	0.9
Weight Initialization	Xavier initialization

The adversarially trained discriminator-based generator, on the other hand, is not designed to generate facial features, rather to learn features to exclude artifacts. The developments in image-generating technologies have made it difficult to distinguish between genuine and synthetic facial pictures in terms of form, color, texture, and individual facial features. In fact, artificial traces surrounding synthetic areas are a primary indicator utilized to distinguish between genuine and false face images in the context of facial recognition. For this reason, a discriminator must be trained to spot these traces during DCGAN, while a generator must be taught to produce artificial pictures devoid of artifacts. Hence, generators is be used in a variety of ways, and the diversification of false pictures can be used to further enhance discriminators. The CED-DCGAN is used to address the real-time deep fake detection in video conferencing. The architecture consists of two primary components: the Generator Network and the CED. The generator is responsible for generating fake video frames, while the discriminator evaluates the frames and determines whether the content is real or fake.

Generator Network

The generator of CED-DCGAN was designed for producing high-quality false video frames. It employs the following configuration:
Input: Randomized input vector (G), latent space label (Y), and layers.
Dense Layer: FC layer that projects and reshapes the input vector.
Transposed Convolutional Layers: A set of layers that up sample the input, resulting in higher quality images.
Batch normalization: It is applied after every transposed convolutional layer for stabilizing training.
ReLU Activation: Applied in all layers but the output layer.
Tanh *Activation*: Utilize in the output layer for generating images within the range $[-1, 1]$.

Discriminator Network

The CED uses ResNet along with DenseNet layers to improve detection capability. The architecture contains:
Input: Actual video frames divided into separate images.
ResNet Blocks: It has Convolutional Layers, which extract features via residual connections.
Batch Normalization: It stabilizes and accelerates training.
ReLU Activation: Assures non-linearity.
DenseNet Block: Dense Layers are efficient feature propagation; layers should have dense connections.
Batch normalization and ReLU activation are similar to the ResNet block. The Ensemble Layer combines features from the ResNet and DenseNet blocks.
Fully Connected Layer: The final layer that produces the categorization conclusion (genuine or fake).

Training procedure

During the preprocessing stage, real-time video is split into segments and normalized to the range $[-1, 1]$. Generator training involves utilizing a random vector (G) and a label (Y) for generating fake images. The generator is designed to use binary cross-entropy loss to reduce the discriminator's capacity to differentiate between actual and fake images. For discriminator training, both actual and produced fake images are used, and a discriminator is trained on maximizing correct classification while also optimizing with binary cross-entropy loss. Backpropagation is the process of iteratively enhancing both networks by updating the generator's weights with the discriminator's gradients. Table 2 shows the hyperparameter value.

Hyperparameters

The detectability is enhanced by averaging the feature maps produced by two separate discriminators, to get a confidence score for distinguishing between actual and false images. Several artificial face pictures are sent to the integrated discriminators as D by means of deep conditional generators. The CelebA-HQ dataset has a variety of human faces that were used to pre-train a generator. Hence, the conditional generator is taught to generate generic fake images that are eerily similar to real-life facial photographs. During the adversarial learning phase, a conditional generator is not further tweaked since doing so would diminish generator generality and lead

Table 3
CED-DCGAN system parameters setting.

Number of video images	1000
Batch size	64
Generator learning rate	0.003
Discriminator learning rate	0.001

to less accurate detection. Contrarily, G is taught in an adversarial manner from the ground up. Compact ensembles discriminator is used to assessing the conditional generator's picture quality. The ensemble discriminator image DCGAN's quality was checked using the index of Structural Resemblance Measure (ISRM) given in equation (7).

$$ISRM = \frac{\left(2\mu_z\mu_y + b_1\right)\left(2\sigma_{zy} + b_2\right)}{\mu_z{}^2 + \mu_y{}^2 + \sigma_z{}^2 + \sigma_y{}^2} \tag{7}$$

Where, σ_{zy} denotes the co-variance of Z and Y, σ and μ represent the average and standard deviation of the original image Z, and label Y and b_1 & b_2 are the constants.

The capability of the structure is decreased during training to prevent overfitting, and the insertion of the ensemble net layers in the discriminator networks aids in this process. The original dataset was used to train the CED-DCGAN using the settings detailed in Table 3. The number of epochs was determined experimentally to provide acceptable results in terms of picture quality. To make the CED-DCGAN more robust, the learning rates of the generator and discriminator were adjusted to be different. As a means of improving the model's predictive power in real-time. Next collected information from a variety of publicly accessible datasets, such as Face-Forensic++ and CelebA-HQ. In addition, combined data from other sources to develop a dataset, using to recognize video genres accurately and in real-time.

In the proposed model images taken into account 50 % real and 50 % false videos images to prevent the model from being biased during training. Nevertheless, the use of audio deepfake in the context of the CelebA-HQ dataset is outside the scope of this article. We have preprocessed the CelebA-HQ dataset, extracted 300 real and 400 false films, and combined them with 100 real and 200 fake videos from the FaceForensic++(FF) dataset for a total of 1000 videos.

4. Results and discussion

4.1. Dataset description

The proposed system utilized two datasets for training as CelebA-HQ dataset and FaceForensics++. https://paperswithcode.com/dataset/celeba-hq. The CelebA-HQ (CelebFaces Attributes High Quality) collection is a large scale, high-quality face image dataset derived from the CelebA dataset. CelebA-HQ has 30,000 celebrity images, each with 40 binary labels or features such as gender, age, and other visual characteristics such as smiling or wearing spectacles. The images are larger (1024 × 1024 pixels) than the initial images in the CelebA dataset, making them ideal for jobs demanding a high level of detail, such as image synthesis, super-resolution, as well as face attribute management. This dataset is widely used in machine learning and computer vision to train and test a variety of models, including GANs and other deep neural networks.

https://www.kaggle.com/datasets/sorokin/faceforensics FaceForensics++ offers a large-scale dataset designed to address the issues of detecting fraudulent and forged material in facial photographs. It contains over 1000 video sequences that have been manipulated using four different techniques: face swap, deep fakes, face2face, as well as neural textures. They range from fundamental computer graphics techniques to advanced deep learning-based machine learning. It is also worth noting that the dataset includes both modified and original movies, as well as frame-level annotation, making it ideal for training and testing models that detect face modifications. FaceForensics++ is popular amongst researchers looking to construct forensic instruments and algorithms to identify deepfakes and evaluating the content of digital video.

4.2. Performance metrics

The deep conditional generator and the compact ensemble discriminator machine learning framework were used to create the suggested system. Recall, accuracy, Sensitivity and the F1-score were employed as measures to evaluate the effectiveness of the suggested model. CED-picture DCGAN's quality was checked using the index of the Mean Squared Error (MSE). Below are the two equation (8) that were used to determine the two metrics:

$$MSE = \frac{1}{MN}\sum_{n=1}^{m}\sum_{m=1}^{n}\widehat{GD}(m,n) - GD(m,n) \quad m = 1,2\ldots n\,; n = 1,2\ldots n \tag{8}$$

Where $\widehat{GD}(m,n) - GD(m,n)$ denotes the images in the generator and discriminator.

Accuracy, precision (specificity), recall (sensitivity), and F1-score [47] are the most popular measures used to assess deep CGAN-CED learning models. The sigma coefficient is be used to check how well the labeled map corresponds to the raw data. It's utilized to

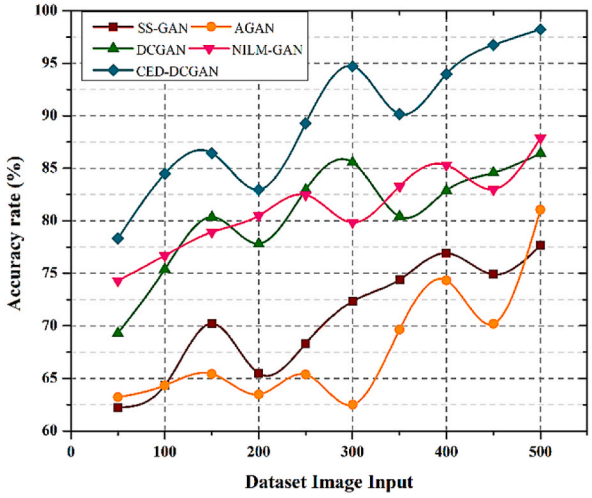


Fig. 7. Accuracy rate.

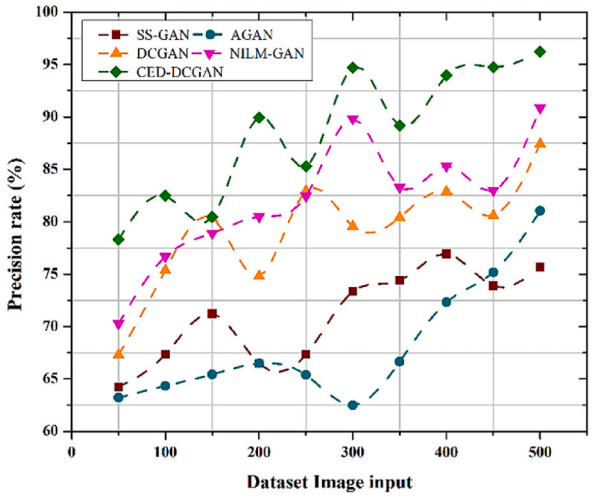


Fig. 8. Precision rate.

keep under wraps any cases that could have gotten their classification right by luck. The measures were selected for the investigation. equations (9)–(12) for the aforementioned metrics with

$$Accuracy = \frac{TP + TN}{TP + FN + TN + FP} \tag{9}$$

$$Precision = \frac{TP}{TP + FP} \tag{10}$$

$$Sensitivity = \frac{TP}{TP + FN} \tag{11}$$

$$F1\ score = \frac{2 * (Precision \times Recall)}{Precision + Recall} \tag{12}$$

The other statistics are based on the counts of True Positives (TP), True Negatives (TN), False Positives (FP), and False Negatives (FN). Finally, the confusion matrix is used to illustrate the inter-category relationships.

4.2.1. Accuracy rate

Fig. 7 shows a graph of the suggested mode's accuracy rate. Many alternative GAN models for detecting fraudulent images are

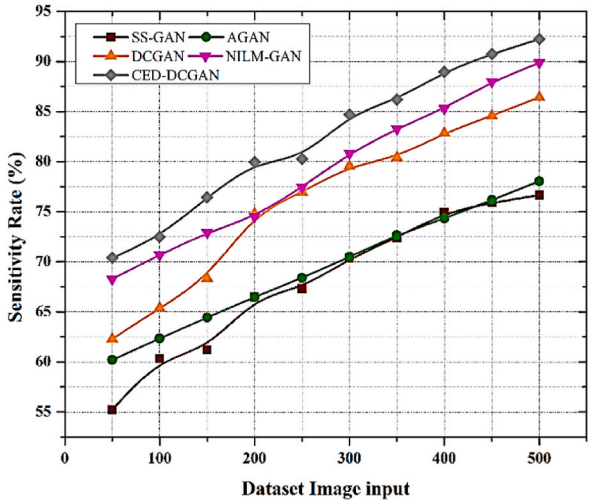


Fig. 9. Sensitivity rate.

compared with the proposed model (CED-DCGAN), including SS-GAN, AGAN, DCGAN, and NILM-GAN. One way in which categorization models of real and fake images are be assessed by accuracy. Informally, the success rate of our model is defined as the proportion of correct predictions. Adopt the accurate positive rate and true negative rate as performance indicators for exploratory modelling methods to ensure their efficacy, where TP, T N, FP, and FN represent the numbers of immediately described fake face images, detected real face face images, mis-categorized actual face specimens, and incorrectly classified fake face images, respectively. All investigative models, once trained, are able to provide very good TPR and TNR results on both datasets. Overall, the suggested system performed better, and it was especially effective at telling the difference between real and fabricated pictures. The last test was run again, with the generator module and discriminator output classes removed. The generator and the fake emotion classes are used to improve reliability. The study utilized just 500 dataset images from the actual world and was able to get an accuracy of 98.23 %. A 2 % improvement in accuracy and variance in the examined photos was found to be a result of using false images in the categorization process.

4.2.2. Precision rate

The original dataset reported was used to train the CED-DCGAN model, which can distinguish between genuine and false photos. CED-DCGAN precision rate graph (Fig. 8). A model with a precision of 1.0 yields no false positives. Examining both precision and recall is necessary for a comprehensive evaluation of a model's performance. Yet, there is often a conflict between accuracy and memory. Simply put, when accuracy increases, recall decreases, and vice versa. As a result of the training procedure in terms of the loss scores of the generator and the discriminator. Almost 16 h were spent in training. Image quality was used to determine the optimal number of epochs (1000) began with 100 epochs and increased until satisfactory results were achieved. The NILM-GAN, SS-GAN, AGAN, and DCGAN are all models compared to the suggested model. According to the chart, the recommended model is the most accurate option.

4.2.3. Sensitivity rate

The dataset used for training images is seen in Fig. 9 below. Sensitivity refers to a measure of how well information can be recovered from a database, database, archive, or specimens in the context of pattern recognition, information retrieval, object identification, and categorization. The responsiveness of a test is the chance of obtaining a positive result if the patient really is affirmative. The likelihood of a negative test result, given that the subject is, in fact, negative, is known as the sensitivity (actual negative rate). The percentage of relevant occurrences that were recovered is known as recall (or sensitivity). The generator and discriminator engage in an adversarial competition, with the generator attempting to lower the loss function to trick the discriminator and the discriminator attempting to raise it to better discriminate between genuine and fraudulent pictures. There is a comparison between the proposed model and other, more sensitive GAN models.

4.2.4. F1 score

The percentage of f1 scores in the training set is shown in Graph 10 and is calculated using equation (12). A model's goodness of fit is measured using the F1-score. Data classification techniques that could be used to separate "outstanding" from "terrible" results are analyzed. Measures of model efficacy are often summarized by their F1-score, the geometric average of their accuracy and recall. The F1-score is often used for many different sorts of image detections and for assessing information retrieval systems like search engines. The F-score is to modified such that it prioritizes accuracy over recall, or recall over precision. The F0.5-score and F2-score, in addition to the normal F1-score, are often used adjustments to the F-score. The images were checked using the Train on Simulated technique, in which the CelebA-HQ dataset and FaceForensics++ model were trained on image sequences and then evaluated on real ones. The CED-DCGAN was able to create pictures that were comparable to the genuine photos, as seen by the greater levels on the CelebA-HQ dataset

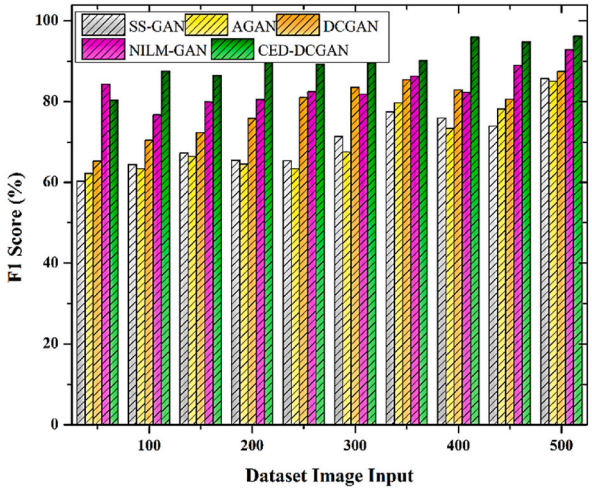


Fig. 10. F1 score.

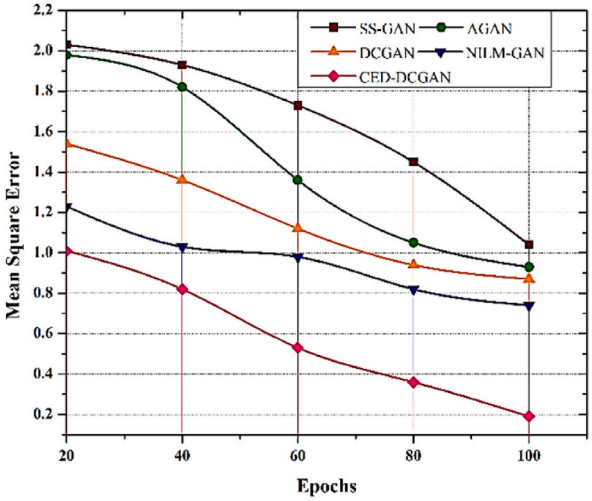


Fig. 11. Mean square error.

and FaceForensics++ performance metrics. In the subsequent step, fake image identification was accomplished by combining the produced photos with the actual images from the training and testing datasets. The suggested GAN model improved upon the prior proposals in terms of F1score(as per Fig. 10).

4.2.5. Mean Square Error (MSE)

An experiment was performed whereby several error metrics (mean squared error, MSE) were employed to guarantee that the produced pictures were unique. With that goal in mind, chose 500 photos at random to compute the aforementioned metrics. MSE, computed using formula 8, reveals how comparable two sets of data are when it is equal to zero. There will be less resemblance if the value is more than one, and it will keep growing as the average intensity difference between pixels widens. The MSE score for negative (at −1) or positive (at +1), with +1 indicating perfect similarity. Typically, the MSE is 0.23677 on average. The CED-DCGAN proved successful in producing unique pictures since the MSE value is larger than one and the SSIM value was close to zero. Fig. 11 provides examples of the MSE values comparing one picture to the rest of the photos so that you can get a feel for the overall findings.

4.2.6. Training accuracy and training loss

Training accuracy and training loss are two of the most basic metrics which are used to give clear and understandable information about the learning process of the machine learning model during the training process on the given dataset. Training accuracy is simply a measure of how well the model is able to learn the patterns and the relationships between the features in the training dataset. A high training accuracy is usually considered desirable as it would mean the model is learning well from the examples which have been fed to it. In contrast, training loss measures the size of the error committed by the model during training. This implies that the model is

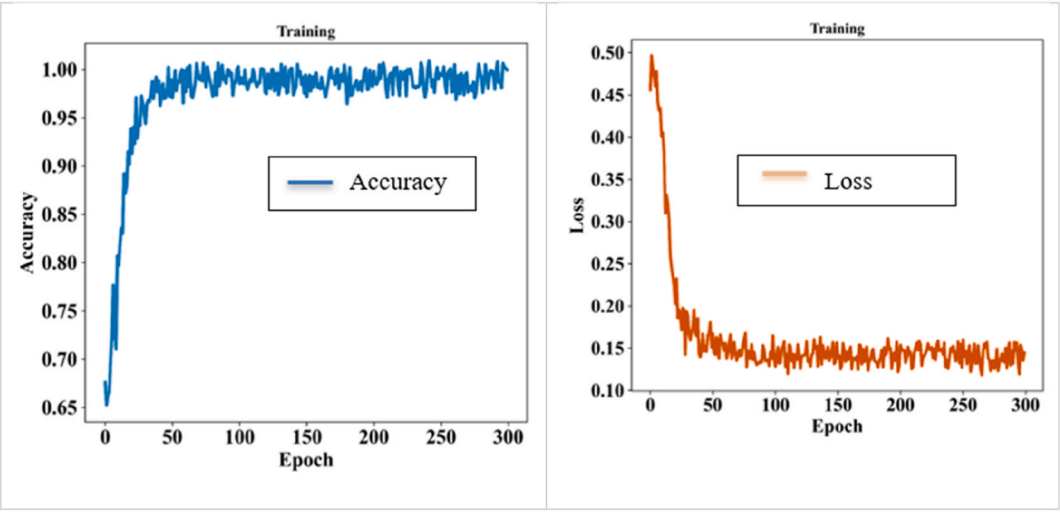


Fig. 12. Training Accuracy and loss curve.

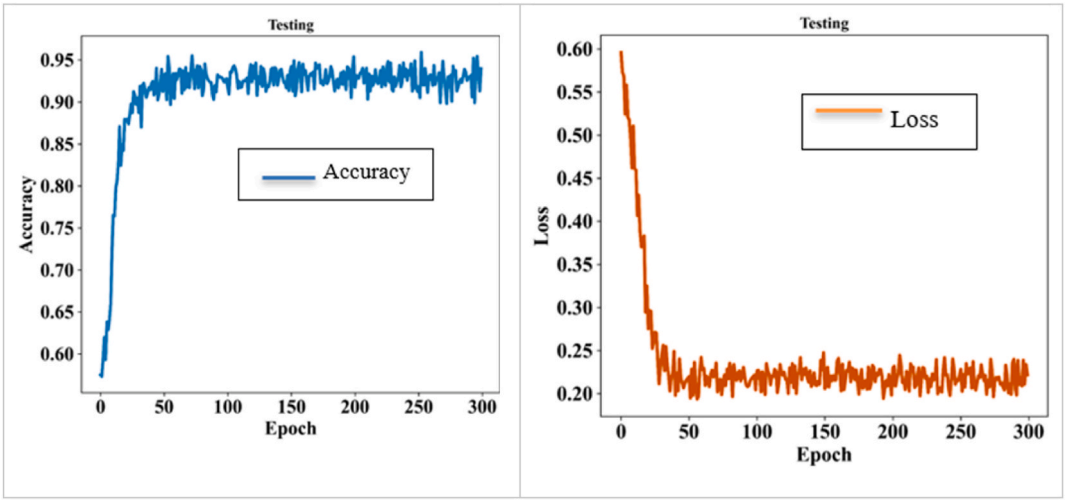


Fig. 13. Training and testing loss.

gradually learning the actual values it is expected to predict since lower training loss indicates that the model's predictions are closer to the actual values. There are many types of loss functions and, while they could differ in formulation, their aim is to penalize the model for wrong predictions. Fig. 12 shows the training accuracy and loss of proposed work.

Testing accuracy and loss are important parameters for assessing the effectiveness of a ML network. Testing accuracy is the fraction of accurately predicted cases over all instances within the test set. It gives a direct indication of how effectively the model applies to new inputs. High testing accuracy indicates that the model can make accurate predictions on fresh, previously unknown data, which is the ultimate objective of model training. In contrast, testing loss measures the error among the model's predictions with the actual targets in the test set. It measures how well the model's predictions fit the actual results. When the epoch value is 0, the testing loss start rising, when the epoch value is 50, the testing accuracy increase and remains constant for other all epoch values. Fig. 13 shows the training and testing loss of proposed work. Fig. 14 (a) for Fig. 14 (b) shows the comparison graph of both proposed and existing approaches in terms of accuracy, precision, F-score and sensitivity, respectively (see Fig. 14).

The outcome comparison table demonstrates the usefulness of the suggested approach for detecting actual time deep fakes. The proposed model beats previous GAN-based techniques including SS-GAN, AGAN, DCGAN, and NILM-GAN in a number of specifications. It achieves the maximum accuracy (98.546 %), exhibiting superior overall performance. The proposed approach also performs well in terms of F-score (98.039 %), precision (97.578 %) and sensitivity (97.986 %), demonstrating its capacity to correctly detect both real and fake images while minimizing false positives and negatives. This thorough enhancement shows the suggested model's performance in real-time deep fake identification for videoconferencing. Table 4 shows the overall comparison analysis.

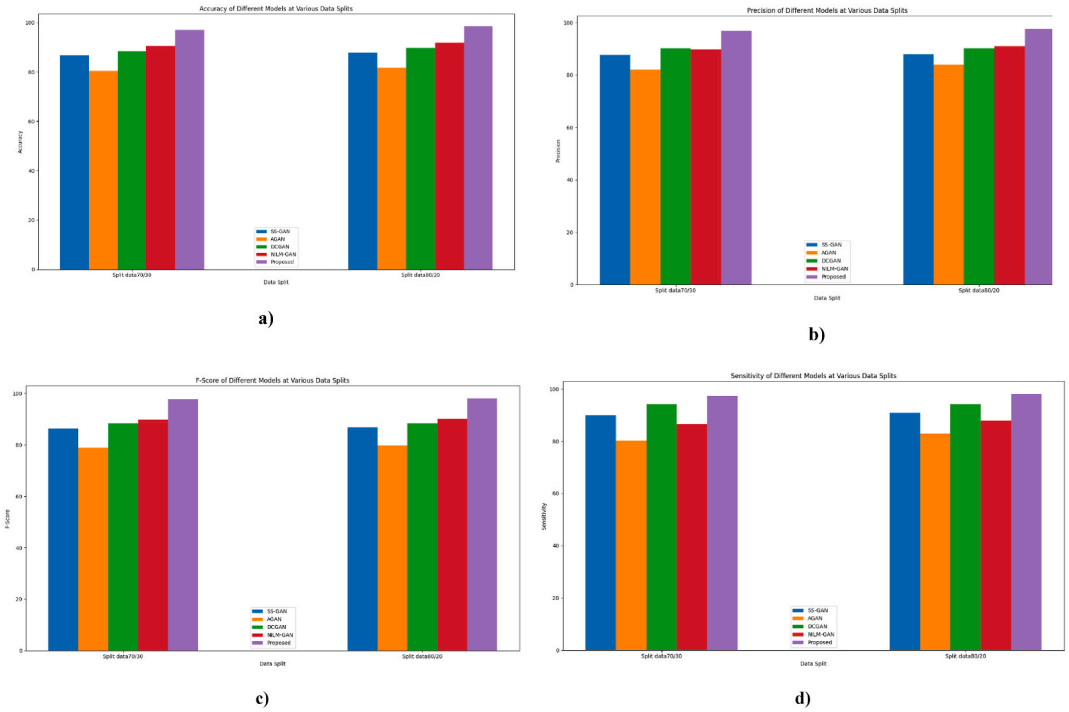


Fig. 14. Overall comparison analysis of proposed and existing approaches.

Table 4
Comparison analysis of Proposed and existing approach.

Model	Accuracy	Precision	F-Score	Specificity	Sensitivity
SS-GAN	87.766	87.898	86.979	85.959	90.853
AGAN	81.786	83.908	79.803	82.897	82.894
DCGAN	89.659	90.206	88.351	90.312	94.205
NILM-GAN	91.887	90.969	90.074	88.905	87.896
Proposed	98.546	97.578	98.039	96.982	97.986

4.3. Discussion

The video conferencing technology has developed as one of the most important forms of communication in the current day. However, the use of deepfake software to create fake films and audios compromises the trust and safety of these interactions. Identifying deep fakes and face fabrication in real time when video conferencing is critical for avoiding fraud and protecting society. This research describes a reasonable approach for quickly detecting and addressing these hazards using GANs. Mode collapse is a prevalent issue with GAN-based generators, particularly DCGANs, StyleGANs, as well as vanilla GANs. This happens when the generator gets caught in a loop, generating a limited set of outputs rather than the complete range of possibilities. This results in a lack of variation in the collected data, limiting its use in real-world applications. Training GANs was notoriously challenging. The adversarial process among the generator and discriminator could become unstable, resulting in sluggish convergence or even inability to learn properly. Furthermore, fine-tuning hyperparameters and determining the optimal balance between the two networks necessitates significant skill. While some architectures, like StyleGANs, have produced impressive results, producing very realistic and comprehensive data remains difficult.

Many deep generative models, like 3DCNNs and DCGANs, require a large amount of highquality training data to work well. This can be an important obstacle for applications that require limited data. Furthermore, overfitting to training data is possible, resulting in models that are unable to generalize to new data. Researchers are continually looking at ways to avoid these restrictions. These consists of regularization methods, improved network architectures that prevent mode collapse, as well as semi-supervised learning approaches that make greater use of fewer datasets. In this study, a unique forensic technique for detecting real-time deep fakes in video conferencing that use a Compact Ensemble-based Discriminator (CED) framework within Deep Conditional Generative Adversarial Networks (DCGAN). Given the rapid evolution of fake production technology, DCGAN focuses on important characteristics when identifying video deep fakes. Real-time video frames are divided and analyzed for training the CED, which detects spectral anomalies in GAN up sampling processes. The CED effectively distinguishes between artificial and natural images, generating a robust training

Table 5
Existing method Comparison.

Author	Techniques	Parameters	Dataset
Proposed	CED-DCGAN	Accuracy- 98.23 %	CelebA-HQ dataset and FaceForensics++
[16]	GAN-based generators	–	–
[17]	3D Convolutional Neural Network (3DCNN)	Accuracy – 95 %	FaceForensics++
[18]	Dual Input Convolution Neural Network (DICNN)	Accuracy- 98 %	Fake-Vs-Real-Faces
[19]	Semi-supervised Generative Adversarial Network (GAN)	Accuracy – 93 %	OULU-NPU dataset
[20]	Attention-based Generative Adversarial Network and IDSNet	Accuracy- 95.2 %, F1 score- 94.1 %	–
[21]	Hybrid DCGANs and StyleGANs	Accuracy – 90 %	CelebA dataset
[26]	Deep Convolutional Generative Adversarial Network (DCGAN)	Accuracy- 80 %	EMBED dataset
[27]	Generative Adversarial Networks (GAN)	D_Loss = 0.0391 and G_Loss = 5.7638	–
[28]	Deep Fake Video Detection Techniques (unspecified)	Accuracy-93.28 %, precision – 91.03 %	DFDC dataset
[29]	Generative Adversarial Networks (GAN) with 2D-based CNN	Accuracy – 97 %	Combined-PIFR dataset
[30]	BFLDL	Accuracy- 98 %	CelebDF
[31]	HASSO	Accuracy- 95 %	Kaggle

signal. Research findings on public datasets reveal that this strategy beats existing innovative methods for detecting high-fidelity deep fakes. Table 5 shows the existing method with proposed comparison.

5. Conclusion

The article introduces a novel active forensic technique, Compact Ensemble-based discriminators architecture employing Deep Conditional Generative Adversarial Networks (CED-DCGAN), for detecting real-time deep fakes in video conferencing. By dividing real-time video into crucial frames, it teaches ensemble-based discrimination to identify deep fakes by focusing on features and spectral abnormalities. Experimental findings on public datasets show that CED-DCGAN outperforms modern approaches for identifying high-fidelity deep fakes as well as generalizing well comparing to other techniques. However, this proposed approach contains some limitations like data complexity, black box nature, need more data etc. To overcome such limitations, attempting to establish a link between people's emotional outbursts over time and the development of their health status is one potential avenue for future study. The research requires participants to be observed at irregular intervals in their natural environment through a mobile device, computer, or smart TV camera to determine their feelings in various settings. In the future, an innovative architecture should be chosen that takes into account each of these variables. Additional audio-visual qualities to help identify deepfakes. This model serves as a standard for detecting deepfakes and can be used by future researchers. Because of noise and additional factors, audio characteristics have not yet been incorporated into the design.

Data availability statement

The proposed system utilized two datasets for training as CelebA-HQ dataset and FaceForensics++. https://paperswithcode.com/dataset/celeba-hq. The CelebA-HQ (CelebFaces Attributes High Quality) collection is a large scale, high-quality face image dataset derived from the CelebA dataset. CelebA-HQ has 30,000 celebrity images, each with 40 binary labels or features such as gender, age, and other visual characteristics such as smiling or wearing spectacles.

CRediT authorship contribution statement

Sunil Kumar Sharma: Conceptualization. **Abdullah AlEnizi:** Writing – original draft, Supervision, Funding acquisition, Formal analysis. **Manoj Kumar:** Software, Project administration, Methodology, Conceptualization. **Osama Alfarraj:** Writing – review & editing, Visualization, Validation, Software, Methodology. **Majed Alowaidi:** Validation, Software, Project administration, Investigation, Funding acquisition, Data curation, Conceptualization.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

The authors extend their appreciation to the deputyship for research & innovation, ministry education in Saudi Arabia for funding this research work through the project number (IFP-2022-37)

References

[1] Chi Ho Chan, Muhammad Atif Tahir, Josef Kittler, Matti Pietikainen, Multiscale local phase quantization for robust component-based face recognition using kernel fusion of multiple descriptors, IEEE Trans. Pattern Anal. Mach. Intell. 35 (5) (2012) 1164–1177.
[2] Jia Deng, Gaoyang Pang, Zhiyu Zhang, Zhibo Pang, Huayong Yang, Geng Yang, cGAN based facial expression recognition for human-robot interaction, IEEE Access 7 (2019) 9848–9859.
[3] Candice R. Gerstner, Hany Farid, Detecting real-time deep-fake videos using active illumination, in: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2022, pp. 53–60.
[4] Hai Thanh Nguyen, Tinh Cong Dao, Thao Minh Nguyen Phan, Tai Tan Phan, Fake face detection in video using shallow, deep learning architectures, Int. J. Intell. Syst. Technol. Appl. 20 (6) (2022) 469–486.
[5] Ziyu Xue, Xiuhua Jiang, Qingtong Liu, Zhaoshan Wei, Global–Local facial fusion based GAN generated fake face detection, Sensors 23 (2) (2023) 616.
[6] Maryam Taeb, Hongmei Chi, Comparison of deepfake detection techniques through deep learning, Journal of Cybersecurity and Privacy 2 (1) (2022) 89–106.
[7] Zhimao Lai, Yufei Wang, Renhai Feng, Xianglei Hu, Haifeng Xu, Multi-feature fusion based deepfake face forgery video detection, Systems 10 (2) (2022) 31.
[8] Ammar Elhassan, Mohammad Al-Fawa'reh, Mousa Tayseer Jafar, Mohammad Ababneh, Shifaa Tayseer Jafar, DFT-MF: enhanced deepfake detection using mouth movement and transfer learning, SoftwareX 19 (2022) 101115.
[9] Zhiqing Guo, Gaobo Yang, Dewang Wang, Dengyong Zhang, A data augmentation framework by mining structured features for fake face image detection, Comput. Vis. Image Understand. 226 (2023) 103587.
[10] Shiming Ge, Fanzhao Lin, Chenyu Li, Daichi Zhang, Weiping Wang, Dan Zeng, Deepfake video detection via predictive representation learning, ACM Trans. Multimed Comput. Commun. Appl 18 (2s) (2022) 1–21.
[11] Shengqing Pei, Yu Wang, Bende Xiao, Shengchang Pei, Yaqi Xu, Yifan Gao, Jianbin Zheng, A bidirectional-LSTM method based on temporal features for deep fake face detection in videos, in: 2nd International Conference on Information Technology and Intelligent Control (CITIC 2022), vol. 12346, SPIE, 2022, pp. 311–318.
[12] Chongyang Zhang, Yu Qi, Hiroyuki Kameda, Multi-scale perturbation fusion adversarial attack on MTCNN face detection system, in: 2022 4th International Conference on Communications, Information System and Computer Engineering (CISCE), IEEE, 2022, pp. 142–146.
[13] Zhiqing Guo, Gaobo Yang, Dewang Wang, Dengyong Zhang, A data augmentation framework by mining structured features for fake face image detection, Comput. Vis. Image Understand. 226 (2023) 103587.
[14] Xinyi Ding, Zohreh Raziei, Eric C. Larson, Eli V. Olinick, Paul Krueger, Michael Hahsler, Swapped face detection using deep learning and subjective assessment, EURASIP J. Inf. Secur. (1) (2020) 1–12, 2020.
[15] K.H. Teoh, R.C. Ismail, S.Z.M. Naziri, R. Hussin, M.N.M. Isa, M.S.S.M. Basir, Face recognition and identification using deep learning approach, in: Journal of Physics: Conference Series, vol. 1755, IOP Publishing, 2021 012006, 1.
[16] Wei Liu, Jingdong Sun, General forgery face detection: against generative adversarial networks using knowledge distillation, in: 2022 5th International Conference on Data Science and Information Technology (DSIT), IEEE, 2022, pp. 1–5.
[17] Aditi Kohli, Abhinav Gupta, Light-weight 3DCNN for DeepFakes, FaceSwap and Face2Face facial forgery detection, Multimed. Tool. Appl. 81 (22) (2022) 31391–31403.
[18] Mohan Bhandari, Arjun Neupane, Saurav Mallik, Loveleen Gaur, Hong Qin, Auguring fake face images using dual input convolution neural network, Journal of Imaging 9 (1) (2022) 3.
[19] Junting Chen, Jiwen Dong, Qingtao Hou, Shenyuan Li, Xizhan Gao, Sijie Niu, Semi-supervised generative adversarial network for face anti-spoofing, in: The International Conference on Image, Vision and Intelligent Systems (ICIVIS 2021), Springer Nature Singapore, Singapore, 2022, pp. 121–130.
[20] Rahmat Ali, Young-Jin Cha, Attention-based generative adversarial network with internal damage segmentation using thermography, Autom. ConStruct. 141 (2022) 104412.
[21] B. Hariharan, S. Karthic, E. Nalina, PN Senthil Prakash, Hybrid deep convolutional generative adversarial networks (DCGANS) and style generative adversarial network (STYLEGANS) algorithms to improve image quality, in: 2022 3rd International Conference on Electronics and Sustainable Communication Systems (ICESC), IEEE, 2022, pp. 1182–1186.
[22] Fiaz Majeed, Afzaal Ahmad, Muhammad Awais Hassan, Muhammad Shafiq, Jin-Ghoo Choi, Hamam Habib, Ontology-based crime news semantic retrieval system, CMC-COMPUTERS MATERIALS & CONTINUA 77 (1) (2023) 601–614.
[23] Tehseen Mazhar, Inayatul Haq, Allah Ditta, Syed Agha Hassnain Mohsan, Faisal Rehman, Imran Zafar, Jualang Azlan Gansau, Lucky Poh Wah Goh, The role of machine learning and deep learning approaches for the detection of skin cancer, in: Healthcare, vol. 11, MDPI, 2023, p. 415, 3.
[24] Tehseen Mazhar, Qandeel Nasir, Inayatul Haq, Mian Muhammad Kamal, Inam Ullah, Taejoon Kim, Heba G. Mohamed, Alwadai Norah, A novel expert system for the diagnosis and treatment of heart disease, Electronics 11 (23) (2022) 3989.
[25] M. Ali, T. Mazhar, T. Shahzad, Y.Y. Ghadi, S.M. Mohsin, S.M.A. Akber, M. Ali, Analysis of feature selection methods in software defect prediction models, IEEE Access 11 (2023) 145954–145974, https://doi.org/10.1109/ACCESS.2023.3343249.
[26] Qifeng Huang, Hanmiao Cheng, Kaijie Fang, Wenbin Yu, Cheng Fan, Yangsong Li, Non-intrusive load monitoring based on deep convolutional generative adversarial network prediction, in: 2022 IEEE 5th International Conference on Electronics Technology (ICET), IEEE, 2022, pp. 1050–1054.
[27] Nurhadi Wijaya, Sri Hasta Mulyani, Albertus Christian Noviadi Prabowo, DeepDrive: effective distracted driver detection using generative adversarial networks (GAN) algorithm, Iran Journal of Computer Science 5 (3) (2022) 221–227.
[28] Nency Bansal, Turki Aljrees, Dhirendra Prasad Yadav, Kamred Udham Singh, Ankit Kumar, Gyanendra Kumar Verma, Teekam Singh, Real-time advanced computational intelligence for deep fake video detection, Appl. Sci. 13 (5) (2023) 3095.
[29] M. Kas, Y. El-merabet, Y. Ruichek, R. Messoussi, Generative adversarial networks for 2D-based CNN pose-invariant face recognition, International Journal of Multimedia Information Retrieval (2022) 1–13.
[30] Arash Heidari, Nima Jafari Navimipour, Hasan Dag, Samira Talebi, Mehmet Unal, A novel blockchain-based deepfake detection method using federated and deep learning models, Cognitive Computation (2024) 1–19.
[31] Rayees Ahamad, Kamta Nath Mishra, Hybrid approach for suspicious object surveillance using video clips and UAV images in cloud-IoT-based computing environment, Cluster Comput. 27 (1) (2024) 761–785.