SCIENTIFIC CULTURE
www.sci-cult.com

SCIENTIFIC CULTURE, Vol. 12, No. 4, (2026), pp. 8178-8192
Open Access. Online & Print

DOI: 10.5281/zenodo.12426813

A PERFORMANCE ENHANCEMENT OF DEEPFAKE VIDEO DETECTION THROUGH THE USE OF A CNN DEEP LEARNING MODEL

Mallikarjun Gachchannavar[1], Dr.Naveen KumarJ.R.[2]

[1]Selection Grade 1 Lecturer in Electronics and Communication Engineering Government Polytechnic, Belagavi, Research scholar, Srinivas Institute of Technology, Mukka, Srinivas University, Mangaluru. Email: mallikarjung346@gmail.com
[2]Professor and HOD in Electronics and Communication Engineering, Srinivas University Institute of Engineering and Technology, Mukka, Mangaluru. Email: naveenjr36@gmail.com

Received: 03/06/2025
Accepted: 20/01/2026

Corresponding Author: Mallikarjun Gachchannavar (mallikarjung346@gmail.com)

ABSTRACT

In the current era, many fake videos and images are created with the help of various software and new AI (Artificial Intelligence) technologies, which leave a few hints of manipulation. There are many unethical ways videos can be used to threaten, fight, or create panic among people. It is important to ensure that such methods are not used to create fake videos. An AI-based technique for the synthesis of human images is called Deep Fake. They are created by combining and superimposing existing videos onto the source videos. In this paper, a system is created that uses a Convolutional Neural Network (CNN) with EfficientNet and LSTM designs to pull out details from each frame. The Multi-Task Cascaded Convolutional Neural Network (MTCNN) algorithm is used for face detection and alignment in images. It utilizes a series of cascaded convolutional neural networks to efficiently locate faces and facial landmarks. Experimental analysis is performed using the DFDC deepfake detection challenge on Kaggle. These deep learning-based methods are optimized to increase accuracy and decrease training time by using this dataset for training and testing. We achieved a precision of 0.985, a recall of 0.96, an F1 score of 0.98, and support of 0.968.

KEYWORDS: Deepfake detection, Convolutional Neural Network, EfficientNet Architecture, LSTM Architecture, MTCNN.

1. INTRODUCTION

Information sharing and broadcasting are now much easier and faster, thanks to the growth of social media platforms. With only one click, people may now access knowledge from around the globe. Regarding news consumption, social media platforms can be utilized for two different purposes: to alert the public of breaking news or, conversely, to disseminate false information [1]. DeepFakes is a popular concept with widespread application. Deepfakes ("fake") are synthetic media (AI-generated media) in which an existing image or clip of a person is superimposed with another person's image [2] [3].

To damage the character's reputation, deepfake technology is used to replace performers' faces in pornography, revenge porn, fake news, hoaxes, and financial fraud with the faces of celebrities. This has spurred business and government actions to identify and forbid their use. The three most risky ways to apply face-swapping algorithms identified are as follows: (i) Face-swap, in which one face is automatically superimposed on another; (ii) lipsync, a technique in which only a portion of a person's face is altered, forcing them to utter things they have never said before; and (iii) puppet master, in which the face of the target individual is animated by a person sitting in front of the camera [4]. FakeApp, created by a Reddit user using the autoencoder-decoder pairing structure, was the first deepfake generation attempt. The face images are broken down into their parts in this manner by the autoencoder, which also extracts latent properties from the face images. Two encoder-decoder pairs, each trained on a different

image set, are required to swap faces between the source and target images. The two network pairs share the encoder's parameters. Alternatively, two pairs share a common encoder network [5]. Many businesses, including Facebook Inc., Google, and the United States Defense Advanced Research Projects Agency (DARPA), have launched a research initiative to find and eliminate deep fakes [6] and [7].

Numerous deep learning methods, including EfficientNet Architecture, long short-term memory (LSTM), and Convolutional Neural Network (CNN), have been created to detect deep fakes in images and videos, and additional research has been conducted in this area [8] and [9].The process of extracting the face from the frame using the MTCNN algorithm demonstrates the overall workflow of the proposed detection system for determining whether videos are genuine or fabricated. This approach detects and extracts facial features from several frames using MTCNN. After identifying the face regions, a binary classifier is trained with EfficientNet to extract features that will distinguish between real and fake faces.

The drive for flawless, realistic images and quality has resulted in the creation of more advanced and precise computer graphics (CG) techniques for generating digital environments. With the increased processing capabilities of modern computers, these methods can produce outcomes so lifelike that viewers may find it hard to differentiate them from reality. However, as highlighted in [10], once the goal of perfect CG image generation is achieved, it introduces a challenge in the field of technology distinguishing between a photo-generated (PG) image and a CG method. Deepfake is an image/video manipulation technique that employs various algorithms and image processing methods. It involves replacing one person’s face in a video with another person’s face.

One of the deep learning-based applications that has recently emerged is Deepfake. Deepfake facilitates the automatic creation and generation of fake video content. This technology is contentious, raising various societal concerns, including its potential to influence elections and facilitate cyber bullying. The preprocessing involves two stages: a data augmentation layer and an image filtering layer. In the first stage, the dataset was processed with transformations like horizontal and vertical flips, random cropping, rotation, compression, Gaussian blur, motion blur, and alterations to brightness, saturation, and contrast.

The detection of deepfake videos can be approached as a binary classification task, categorizing each image as either “real” or “fake.” This research addresses the problem of deepfake impersonation by deploying an EfficientNet model. The model analyzes video frames using a deep learning method to identify discrepancies in facial features introduced during the frame creation process. MTCNN is a neural network utilized for identifying faces and facial landmarks in images [11]. It is among the most accurate and commonly used face detection tools currently available. This network is composed of three neural networks linked in a cascading structure [12]. Deepfakes have become widely popular due to the impressive quality of manipulated videos and the ease of use of their applications, which cater to users with diverse computer skills, from experts to novices. These applications are mainly developed using deep learning methods. Deep learning is known for its ability to model and represent complex, high-

dimensional data. One type of deep network with this ability is deep autoencoders, which are frequently utilized for tasks like dimensionality reduction and image compression.

The approach taken after analyzing both real and fake videos includes data preprocessing steps to extract faces from each video using MTCNN (Multi-Task Cascaded Convolutional Networks). After the faces are extracted, the images are transformed into vectorized numerical representations. These vectorized images are then input into EfficientNet, followed by the creation and validation of a fully connected neural network model using the test data. The model's performance is evaluated with metrics such as the confusion matrix, precision score, F1 score, and accuracy.

1.1 Problem Statement

In today's digital age, where there's a constant flood of content being created and shared online, we're facing a new challenge called deepfake technology. This fancy term refers to computer-generated media that looks incredibly real. But here's the catch: deepfakes can seriously mess with our trust and authenticity. These AI-generated manipulations have the power to deceive, impersonate, or manipulate people, governments, and organizations. And that can have some pretty serious consequences, both on a personal and public level. The scary thing is that deepfake technology is becoming more accessible by the day. Almost anyone with an internet connection and a basic understanding of tech can create super realistic-looking fake content. And that's where the real danger lies. The widespread use of deepfakes brings some major risks. It can make people lose confidence in the media, mess with democratic processes by spreading false information, and seriously harm individuals through identity theft or defamation. To combat this growing threat, this paper comes up with a deepfake detection system. By using fancy deep learning techniques, this system can analyze videos and spot tiny irregularities that are typical of deepfake movies. Its goal is to determine if a particular video is legit or fake. This helps to protect people and institutions from malicious exploitation, preserve the integrity of digital media, and maintain public trust in what we see online.

1.2 Motivation

Researchers' interest in identifying fakes has increased significantly in the past several years because of their impact on human behavior. An extensive body of research aimed at identifying practical methods for deducing if an image/video was a deepfake [4-7]. Fake news has evolved partly because of the increasing accessibility of Video modification tools. Improvements in the ability to identify deepfakes are thus critical. This new area of study has opened up a whole new research area. As software manipulation gets more common and simpler to manage, the number of deepfake examples increases. In general, deepfake presents several issues, including severe ramifications for individuals, governments, and businesses. To better understand and use machine learning-based techniques for deepfake detection, researchers have compiled several of these solutions in one place [8-11].

In this research, we employed a CNN and its various variants that use binary classification to tell the difference between real and fake images and then offer an accurate result. The following is a list of some of the most important aspects of this work:

- Propose a CNN-based approach using the concept of transfer learning that increases the detection performance of deepfake and real images with an accuracy rate of 96% using the MTCNN model.
- Apply a data augmentation to improve the model performance by generating new image samples during the training of the MTCNN model.
- Thorough comparison of multiple CNN variants was undertaken, and the MTCNN model performed much better than other DL models and state-of-the-art.

1.3 Purpose

The purpose of this research is to address the growing concerns around the proliferation of synthetic media, particularly deepfake videos. With advancements in the generation of fake videos through techniques like face swapping, lip synchronization, and puppeteering, it is crucial to detect and prevent the spread of such deceptive content, especially on social media platforms. Current methods focus on specific types of deepfake manipulation, but there is a need for a universal framework capable of detecting a broad range of deep fake content. This paper proposes a novel deep learning-based solution that identifies deepfake videos and images using a hybrid approach, improving detection accuracy and real-time forensic capabilities.

A full article usually follows a standard structure: **Section 2** Literature Review **Section 3** is the suggested system Proposed Methodology and Proposed MTCNN-CNN model architectures and

performance method. **Section 4** the experimental results and examined. **Section 5** brings the work to a close, discusses its limits, and offers suggestions for further research on this topic

2. RELATED WORK

2.1 Detection Based on ML

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

2.2 Detection Based on CNN

A. Facial Tampering

Guera and Delp [16] devised a solution consisting of key components of a convolutional neural network and long short-term memory. After combining the attributes of many consecutive frames, CNN creates a collection of features for each frame in a particular picture sequence and provides them to the LSTM for analysis. The suggested model underwent training on 600 videos and attained an accuracy of 97.1 percent. Li and Lyu [17] established a technique for identifying distorted images in manipulated films with an accuracy of up to 99 percent when trained with four distinct deep-learning models on legitimate and modified photos. Zhou et al. [18] proposed a multi-stream network for facial recognition modification in the deepfake. A deep learning face classification model is being trained in the first stream to collect evidence of tampering with artifacts. In the second stream, a steganographic model-based multi-layer network is trained to regulate functions that collect leftover noise evidence nearby. Afshar et al. [19] built MesoNet, a CNN, to distinguish between the actual and Deepfake-modified faces. Meso-4 and MesoInception-4 are two models based on inception used in the network, along with layers linked with the max-pooling function. Khalid and Simon [20] developed a one-class approach for identifying deepfakes and achieved 97.5% accuracy on the FaceForensics++ dataset without having any fake images in the training samples. The authors in [22] developed a strategy for creating a Deepfake detector dubbed FakeCatcher (FC), which emphasizes using features derived from face regions to recognize synthetic portrait films. Missing reflections and minute features in the facial areas are exploited, and characteristics from the face are retrieved from the essential facial features and supplied into machine learning classifying models for identifying them as fake or real films.

B. Digital Media Forensics

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

Ahmed et al. [27] research the application of advanced CNN amplification techniques for reconstructing deep fake images in real time using devices such as video and surveillance cameras. The study successfully merges an improved Deep Fake configuration with accurate targeting achieving q 95.77% accuracy rate. However, there are small inconsistencies in the projected costs for implementation. While the studydemonstrates high accuracy, it does not address the potentialvariations in cost estimates that affect its adoption.

Momina Masood et al. [28] proposed a method for deep fake detection that includes face detection and extraction, feature calculation and prediction. The study evaluates the performance of ten widely recognizeddeep learning models. The open Face toolkit is used for detecting and extracting face from video frames, while CNN models are applied for feature calculation. In the final step, an SVM classifier determines whether the features derived from the deep learning models are authentic or fabricated.

Harsh Agarwal et al. [29] implemented a

frequency domain analysis technique together with an SVM classifier todifferentiate between authenticandmodified images. Thisapproach was tested on a set ofdeep fake images collected from various online sources, proving effective in detecting these altered images.

Pavel Korshunov et al. [30] introduced detection methods based on an audiovisual approach, along with several basic techniques such as principal component analysis (PCS) and support vector machines (SVM).

HasinShahed et al. [31] conducted a comparative study on deep learning techniques for detecting deep fake images employing eight CNN models.. The findings highlighted that convolutional neural networks are highly effective for detection, with ResNet50 delivering the highest accuracy.

Fengxi Song et al. [32] recommended employing a principal component analysis (PCA) technique for feature selection. Theexperimentalresultsrevealed that while PCAdidnot alter accuracy, itplayed a keyrole in reducing the data's dimensionality.

Amritpal Singh et al. [33]proposed a framework for deep fake video detection, utilizing spatio-temporal features, where a time-distributed layer followed by a dense layer is integrated with EfficientNet. This model achieved a test accuracy of 97.6%.

Zhu et al. [34] developed a deep learning model to detect deep fake images, applying CNNs to extract frame-specific features and recognize altered images. The method was evaluated on a large dataset of forged i

mages from various sources, achieving favorable results.

Wang et al. [35]proposed a method to detect images containing synthetic faces generated by deep neural networks. Theconvolutionalnetworkanalyzes the wholeimage, initiallyextractingbasic featurethrough several layers, which are then merged to form more complex features via successive convolutional layers.CNNs can gather more detailed information by integrating high-level features derived from multiple low level features.

Ali Raza et al. [36] proposed a new deep learning method for detecting deep fake images. The approach uses a publicly available deep fake dataset from kaggle. The innovative DFP method integrates VGG16 with convolutional neural network architecture, combining layers from both to form the model. The DFP approach demonstrated superior performance compared to other leading techniques. However, it does not generalize effectively to images created using different generation methods.

Kumar etal. [37]analyzed various neural network-based techniques for identifying Deep Fakes in highly compressed conditions and shows that a suggested metric learning method is particularly effective for such tasks. The metric learning approach, using triple network architecture, yielded positive results when fewer frames were used to evaluate the authenticity of videos. However, a key limitation of this approach is it sin ability to generalize across different datasets. It does not have an unsupervised feature adjustment to align the feature space from the source dataset to the target dataset, which would improve the models robustness and enable automatic labeling.

Hu J et al. [38] proposed that one of the most powerful techniques for generating realistic images is the use of GANs (Generative Adversarial Networks), a deep learning method. A GAN consist of two main components a generator and a discriminator. These components work against each other, with the generator producing synthetic images, and the discriminator assessing where the images are real or fake. Deep fake images are created using an auto encoder, a deep neural network that compresses input data into a smaller form and then reconstructs the original data from this compressed version. Once the encoder, and the decoder attempts to rebuild the from the encoder data.

Rajkumar et al. [39] emphasized the novelty of their work through the careful selection of three specifically designed CNN architectures DenseNet121, Inception V3, and ResNet50, reflecting their dedication to a through and scientifically rigorous analyses. A comparative review of CNN based techniques, presented in this study, offers empirical findings and valuable insights into various CNN architectures [40]. Our literature review provides a broad examination of the current progress in passive digital image forgery detection, with a particular focus on the growing risks associated with deep fake. The studied explored help to understand manipulation techniques, mode resilience, and emerging challenge sin real-world applications, forming the basis for our unique contributions.

Lu C et al. [41] proposed a convolutional vision transformer for detecting and recognizing Deep Fakes, which is a technique that generalizes deep fake video detection using CNNs. The Convolutional Vision Transformer (VIT) integrates two components the CNN, which extracts trainable features, and the Vision Transformer, which utilizes an attention mechanism to process these features. However, the model could be further improved with the development of new datasets in the future.

3. PROPOSED METHODOLOGY

This section details the proposed technique for detecting real and fake images. The method used to identify real and fake images is shown in Figure 1. It employs a deep CNN and its different variants (CNN, EfficientNet, and MTCNN), leveraging transfer learning to enhance image classification. Data augmentation generates additional and diverse examples for the training datasets, improving the models performance and outcomes.

3.1 Dataset:

The DFDC dataset is used for the experiments. Many deepfake or face swap datasets include films shot in non-natural environments like news or briefing rooms. Worse, the people in these films may not have consented to have their faces modified. With over 100,000 total clips collected from 3,426 paid actors and produced using a variety of Deepfake, GAN-based, and non-learned algorithms, the DFDC dataset is by far the largest currently and publicly available face swap video dataset. Each of the 100,000 forged videos in the DFDC Dataset is a one-of-a-kind target/source switch. DF-1.0 consists of 1,000 distinct bogus videos, despite the disruptions. The DFDC dataset includes movies of people in indoor and outdoor situations, with a wide range of lighting situations.

3.1.1 Algorithm-1: Deep Fake image classification ()

1. Perform pre-processing on the dataset images, 10000 samples for real and 10000 samples for fake images including operations such as cropping, resize, and normalization.
2. Apply MTCNN to identify and extract face portion from the images.
3. Divided the dataset into training, validation, and test subsets.
4. Define the input image with dimensions (height, width, channels) and assign it to the variable x.
5. For Block 1:

- Repeat the following steps for I = 1 to 2.
- Compute the dot product of x and W[i], add b[i], apply ReLU activation, and update x.
- Execute max pooling on x with a stride of 2 and update x.

6. **For Block 2:**

- Repeat the above process for I = 3 to 4:
- Compute the dot product of x and W[i], add b[i], apply ReLU activation, and update x.
- Execute max pooling on x with a stride of 2 and update x.

7. Repeat the operations in Block 2 for Blocks 3, 4, and 5.
8. Flatten x to transform it into a 1D array.
9. For I = 5 to 7, repeat the following steps:
Compute the dot product of x and W[i], add b[i], apply ReLU activation, and update x.
10. Compute the dot product of x and W [8], add b[8], apply softmax activation, and assign the result to x.

The CNN model used the Deep Fake detection Challenge (DFDC) dataset, which is commonly employed for training and testing deep fake video detection techniques. The DFDC dataset consist of videos containing manipulated faces, with each video featuring at least one person, lasting 10 seconds and comprising 300 frames. The creation of the DFDC dataset involved collaboration between the partnership on AIs media integrity Steering Committee, AWS, Face book, Microsoft, and academic institutions.

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

Table 1. Various datasets available for Deepfake detection

Dataset	Train Real	Train Fake	Test Real	Test Fake
UADFV	35 videos (13,976 frames)	35 videos (13,638 frames)	14 videos (3,353 frames)	14 videos (3,353 frames)
Celeb-DF	370 videos (158,992 frames)	733 videos (290,043 frames)	38 videos (16,409 frames)	62 videos (22,834 frames)
Deepfake Detection	254 videos (202,723 frames)	2,148 videos (1,678,558 frames)	109 videos (94,437 frames)	920 videos (681,550 frames)

3.2 Our Approach

Deepfakes news is influencing the globe because individuals worldwide use it for various purposes, including face swapping, reproducing pornographic movies with someone's face or body, and manufacturing and disseminating fake news. Deep Fakes are increasingly harming democracy, privacy, security, religion, and people's cultures. Deep Fakes are becoming more common, yet there is no standard for evaluating deep fake detection systems. Since 2018, the number of deep fake movies and photos discovered online has nearly doubled.

Finally, they determined that bogus news travels 1,500 times faster than accurate news. Deepfakes create fake news, photos, videos, and terrorist events. Deepfake undermines public faith in the media and contributes to social and financial fraud. Religions,

organizations, politicians, artists, and voters are all affected by deepfake. People will disregard the truth as deepfake videos and pictures proliferate on social media. For learning temporal aspects of facial data from training films, Deep learning model employing CNN (Convolutional Neural Network) models consisting of EfficientNet Architecture and LSTM Architecture followed by MTCNN is proposed.

We have suggested a CNN-based model that learns different patterns between Deep-Fake and actual videos. Pixel distortion, discrepancies with facial superimposition, skin color variances, blurring, and other visual aberrations are among these distinguishing characteristics. Using a frame-based technique based on the aforementioned different properties, the suggested approach has successfully trained a CNN (convolutional neural network) to discern DeepFake films. The proposed work, which involves an ensemble of EfficientNet Architecture and LSTM Architecture followed by MTCNN, shows the viability of our model's ability to identify deep fake faces in a specific video source accurately. This will help security applications used by social media platforms combat the growing threat of "deepfakes" by accurately determining the authenticity of videos, allowing them to be flagged or removed before they cause harm that cannot be repaired. The dataset is imported and converted based on metadata training and labeling in a JSON file. All face frames were cropped, aligned, and reduced to 256x256 pixels after internal face tracking and alignment were utilized to preprocess the source videos. 5,000 face frames were used to train models. The EfficientNet model feeds temporal features to the EfficientNet model. The LSTM model's feedback architecture may learn from consecutive inputs. We trained our model with 10 epochs and 25 batches. An "epoch" is a machine learning term that describes how many rounds the algorithm did across the full training dataset. Once educated, the ".h5" file can be downloaded. Hybridization successfully leverages many model layers to boost learning performance.

Table 2: Accuracy Achieved

Model	Accuracy
EfficientNet	89.41%
LSTM Architecture	93.85%
Hybrid Model	98.85%

3.3 Our Methodology

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

The low-level diagram depicting the flow of events is shown in Fig. 1.

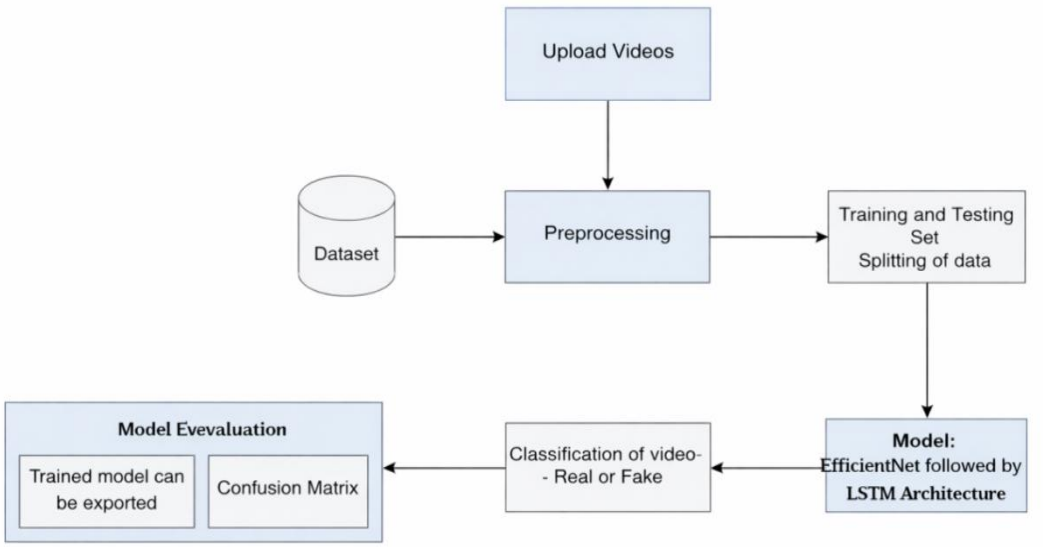


Figure 1. Low-Level diagram depicting the flow of events

The flow diagram is shown in Fig. 2.

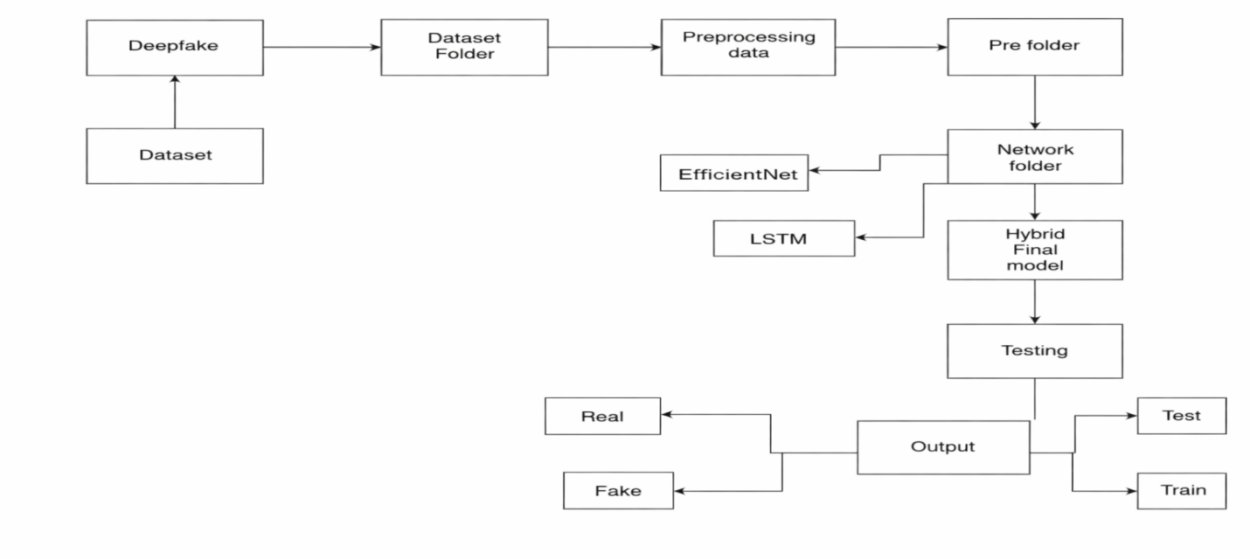


Figure 2. Data Flow diagram

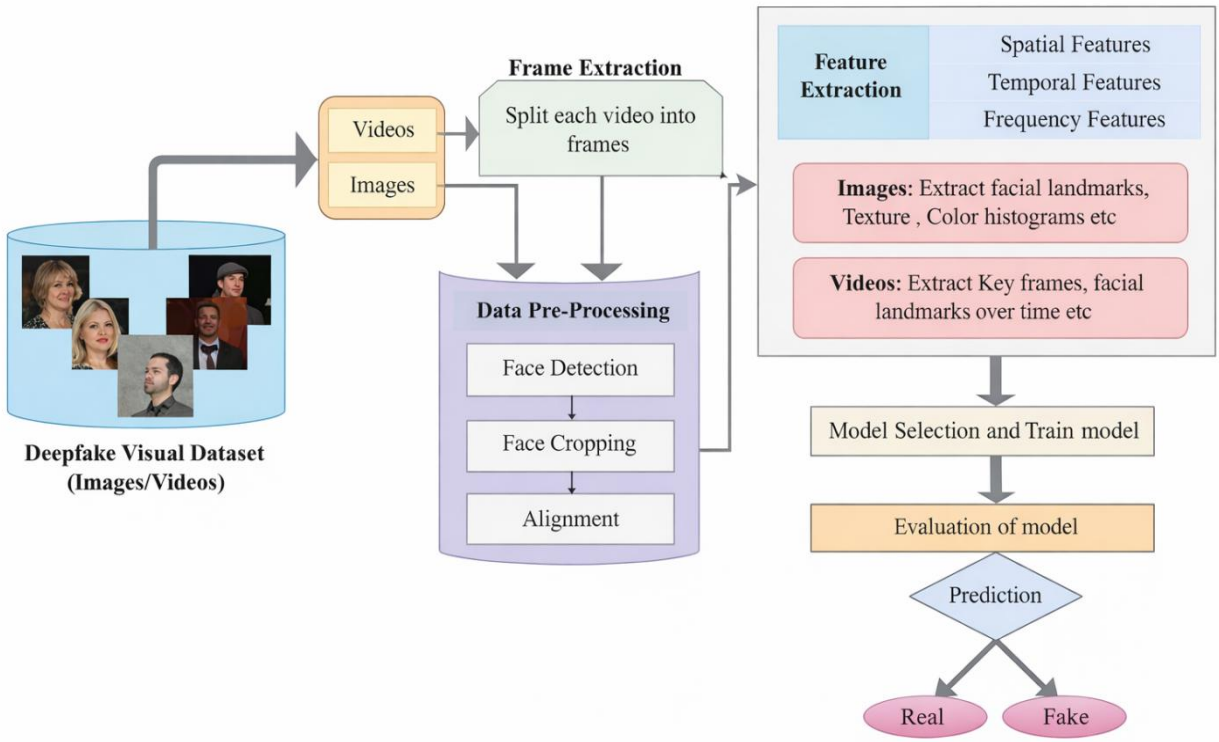


Figure 3: Architecture for purposed system

3.4 Train CNN model

The proposed frame work for detecting deep fakes consists of three main components. These components include a CNN-based architecture designed to extract features associated with the eyes and nose, along with another CNN based structure.

3.4.1 Convolutional Neural Network

Convolutional Neural Networks (CNNs) are unique type of machine learning model, Categorized under the broader umbrella of artificial neural networks and are applied to a variety of data formats and use cases in the field of deep learning, CNNs are

notable for their specialized structure, designed explicitly for handling image related tasks and processing pixel based data. Deep learning consists of multiple neural network types, with CNNs being particularly strong in detecting and classifying objects, making them the preferred choice for these tasks. This makes CNNs ideal for computer vision applications, such as objects identification in self driving cars and facial recognition systems. The dataset was used to train the models, and their performance was carefully examined to determine the best performance one based on accuracy during both training and validation along with graphical representations.

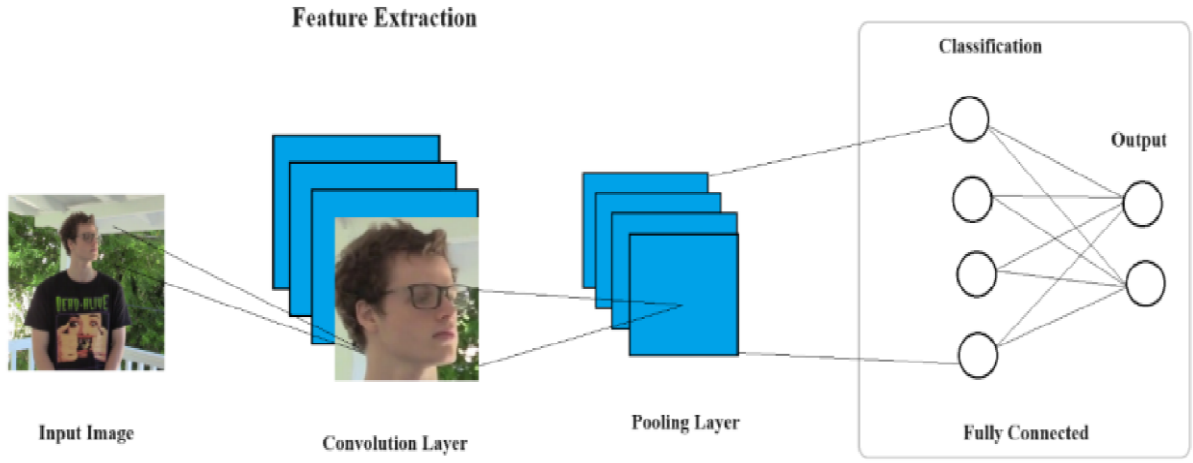


Figure. 4: proposed of CNN Model

3.4.2 Efficient Net:

The EfficientNet CNN architecture and scaling technique applies a compound coefficient to balance the scaling of depth, width, and resolution parameters. Using a fixed set of scaling factors, the EfficientNet method uniformly increases the networks depth, width, and resolution. Once the input data passes through the multi layer network, it produces extracted features that capture deep semantic information.

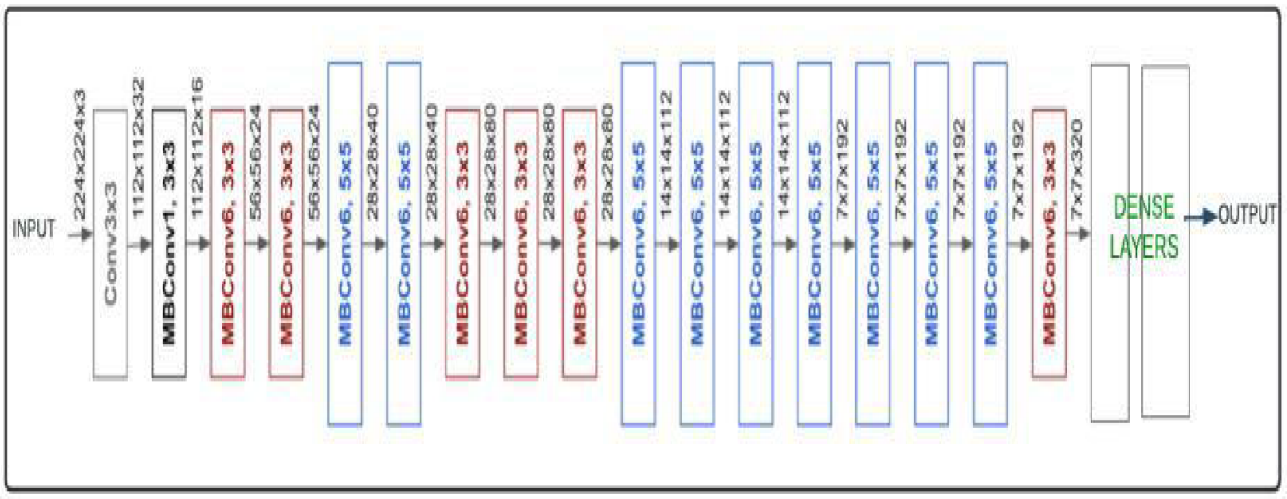


Figure. 5: Efficient Net with transfer learning architecture

3.4.3 MTCNN model:

Frame Extraction involves separating individual frames from video files, while face detection employs Multi-task Cascaded Convolutional Networks (MTCNNs) to identify faces in each frame. Face alignment compensates for variations in head orientation and facial expressions by standardizing the positioning of face. After this, face cropping adjusts the aligned face images to a uniform size. Simultaneously, the extraction and cropping of the eyes and nose focuses on locating and isolating the relevant facial regions from the aligned images.

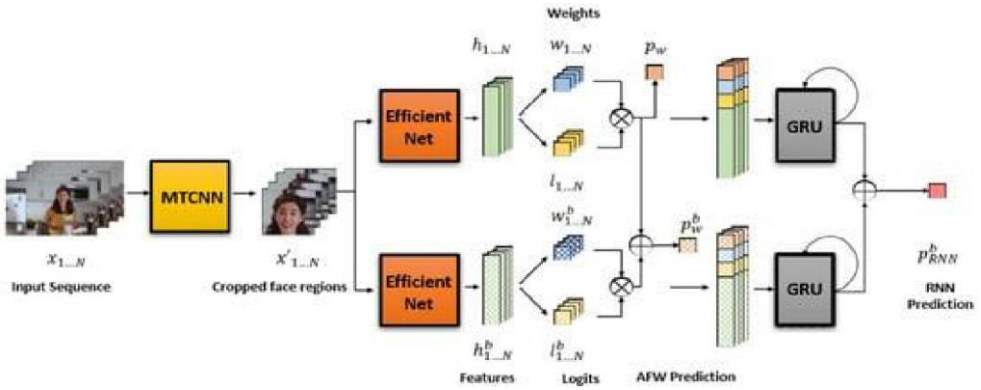


Figure. 6: Extraction of the face from the frame using MTCNN algorithm.

Figure 5 show the entire process of the proposed detection framework for identifying whether a video is authentic or manipulated. The method extracts facial feature from several frames using MTCNN. After detecting the face regions, a binary classifier is trained using EfficientNet to capture feature that differentiate real faces from fake ones. The final classification of authenticity or manipulation is made through a combination of AFW and GRU. The authors trained and assessed the proposed method using the Deep Face Detection Challenge (DFDC) datasets.

4. MATERIALS AND METHODS

Convolutional neural networks have been exceptionally successful in image analysis. The term refers to a specific network architecture that has a class in neural networks, the first stage of each so-called hidden layer is the local convolution result of the recent layer (the kernel includes trainable weights), and the second phase is the max-pooling stage, which decreases the number of subunits by maintaining just the maximum response of many units from the first stage. After multiple concealed layers, the last layer is comprised of a completely linked layer. It consists of a unit for each category that the system detects, and each of these units receives input from all preceding layer units.

It is commonly known that intricate DL-based architectures with many hidden layers, such as Alexnet and VGG16, can effectively handle a high number of classes. However, when there are fewer classes, they have a tendency to overfit, which reduces accuracy. The high storage needs are also a result of the huge number of layers. In this study, a lightweight DL-based architecture has been suggested due to the dataset's low number of DeepFake classes.

4.1. The Proposed Convolutional Neural Network

For DeepFake video identification, several CNN models with a limited number of layers were used. Several versions with various numbers of filters for each layer were taken into consideration, even though the number of layers selected was five. The filter size was maintained at 3 x3 in each layer, with a 2 x2 max-pooling layer coming after each convolution layer. The data were flattened in 2D after being processed using convolution and maximum pooling. Data were then sent to a dense layer with 128 nodes after flattening. Figure 6 displays the CNN architecture.

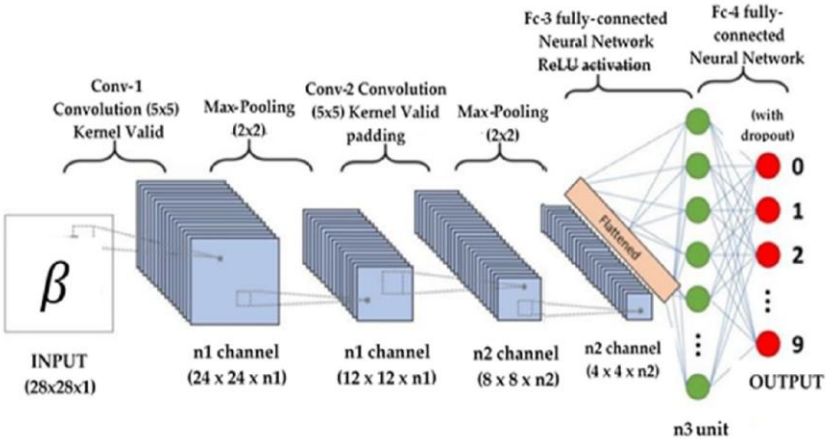


Figure. 7. Convolutional neural network architecture.

In a neural network, the CNN is a subset of artificial learning networks that is most typically employed to assess visual imagery. Matrix computation can be compared to neural networks, but this is not the circumstance with ConvNet. The special technique used in this is convolution. Convolutional neural networks are a collection of artificial neuron layers that function together. Artificial neurons are mathematical functions that examine the weighted total of aggregate inputs and then output an activation value, similar to their biological counterparts. When an image is fed into a ConvNet, each layer functions, which are then transferred onto the next layer. Essential features such as horizontal or diagonal edges are extracted in the first layer. The layer categorization provides a series of confidence ratings (numbers between 0 and 1) relying on the activation map of the previous convolution layer, indicating how probable the image is to conform to a "class." A nice example is a ConvNet that recognizes cats, dogs, and horses, with the last layer's output being the possibility that the input image features any of those species. Figure 7 shows the classification layers combination for the proposed model.

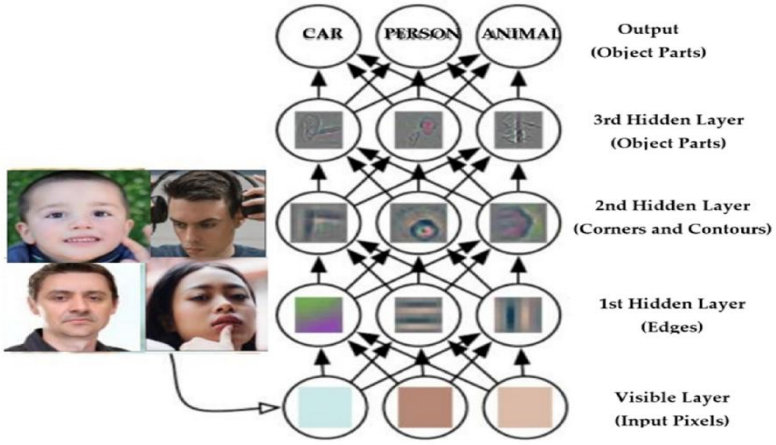


Figure. 8. Classification layer combination.

The average of all the values from the area of the image covered by the kernel is what is returned by average pooling, on the other hand. Dimensionality reduction is all that average pooling does to reduce noise. Therefore, we can conclude that max pooling outperforms average pooling significantly. Despite their strength and resource sophistication, CNNs deliver in-depth findings. It all comes down to identifying patterns and traits that are minute and insignificant enough for the human eye to miss. However, it falls short when it comes to understanding the substance of digital photographs. The maximum and average pooling used by CNNs is depicted in Figure 8.

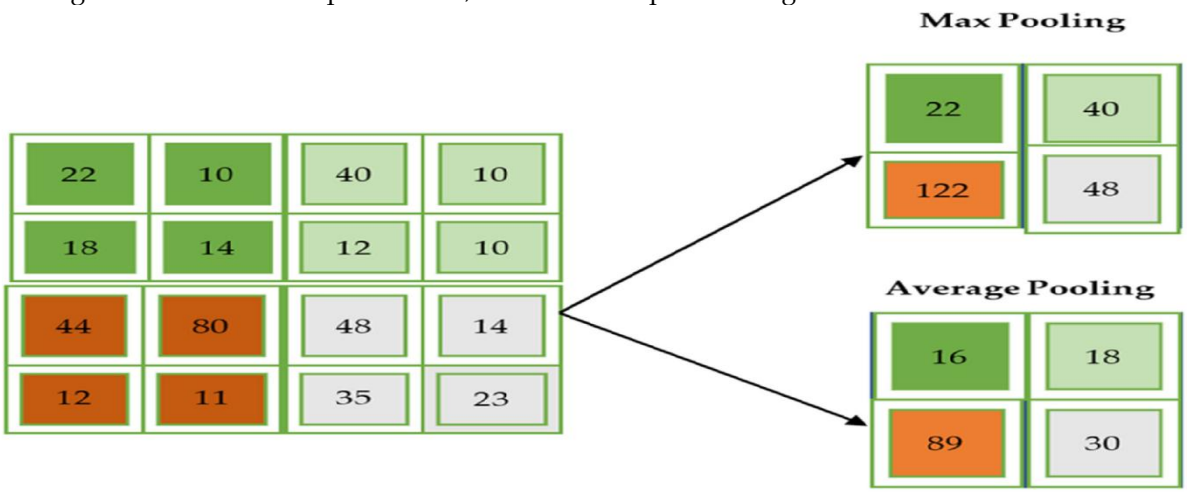


Figure. 9. Maximum and average pooling

These limitations are clear when it comes to practical application. For example, social media content was frequently filtered using CNNs. They were still unable to completely prevent and erase inappropriate material, despite having been trained on a significant number of images and videos. For

example, a 30,000-year-old sculpture had been labeled as nudity on Facebook.

4.2. The Proposed CNN Enhanced with ReLU Architecture

The model is based on a well-performing image classification network that switches the convolutional layers and pooling layer for extraction features and a classification network. This network will start with a sequence of five successive convolution layers, batch normalization layer, and pooling layer, because existing image analytics methodology easily diminishes its capacity to detect DeepFake in movies due to compression, which often degrades the data. One hidden layer was used to construct a dense network. For better generality, ReLU activation functions are added to the convolutional layers to generate a non-linear, batch normalization layer and pooling layer. Robustness is improved by fully-connected layers using a neural network regularization technique by discarding a random subset of its units; this technique is called dropout.

4.3. The Dataset Description

This study employed three datasets to train and test the method on various objects. The system shows a high capability of delineating videos having higher resolution compared to the experiments.

DeepFake Dataset

This dataset was created by the authors in, and it was used for developing their DeepFake detection system called MesoNet. This was accomplished by teaching autoencoders to perform the task; for a realistic result, several days of practice using processors were required, and it could only be accomplished for two faces at a time. The study chose to download video profusions available to the public online, to have enough variety of faces. Therefore, the study collected 175 forged videos across different platforms. The video's minimum standard resolution is 854 _ 480 pixels, and its lengths range from two seconds to three minutes. All the videos were compressed in different compression levels using H.264 codec. A trained neural network for facial landmark detection was used to organize the faces after they had been extracted using a Viola–Jones detector. On average, about 50 faces were extracted from each scenario. In conclusion, this dataset was reviewed manually to eliminate misalignment and wrong face detection, while to avoid having the same distribution of image resolutions either good or poor, both classes were used to avoid bias in the classification task.

5. RESULTS AND DISCUSSION

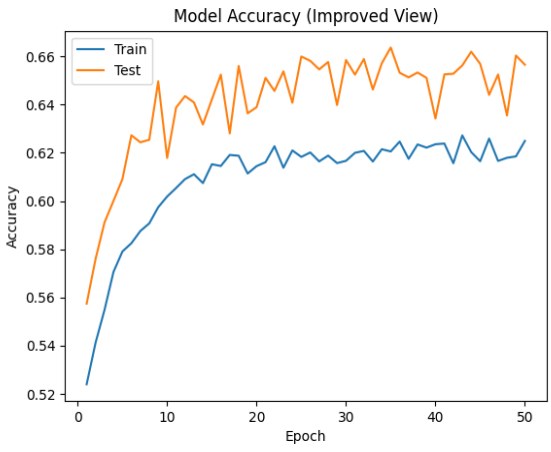
This section provides the result of the developed model and compares them with current advanced techniques. It includes two graphs illustrating the loss and accuracy during the training process, which demonstrate the model performance. The prediction results are also presented in this section.

5.1 Experimental works:

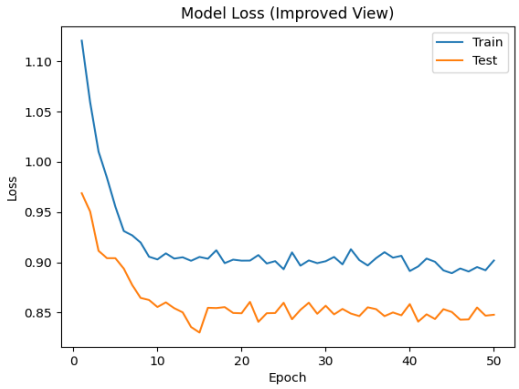
The experimental results of the CNN model were generated by applying various preprocessing techniques to the dataset, with the CNN model trained on the extracted features. The image and video data were sourced from koggle.com and converted into frames. After the frames were extracted the video was transformed into color image to gray form which was subsequently analyzed to assess their authenticity. During the pre processing phase, all frames were resized to a consistent dimension. The CNN model requires input images to be 2548254. After resizing the images were compressed and noise was added. To eliminate the noise, reshaping was carried out, and images were converted from color to grayscale. Faces were extracted from the video, and the MTCNN algorithm was used to assess whether the faces in the dataset were or fake.

5.2 Efficient Net:

With an accuracy of 0.83, efficient Net achieves an impressive precision of 0.85, highlighting its proficiency in correctly identifying positive instances. It is also records a commendable recall of 0.8, reflecting its capability to capture positive instances within the dataset. The F1 Score of 82 emphasizes Efficient Nets balanced performance between precision and recall, making it a dependable model for classification tasks.



(a)



(b)

Figure 10: Training accuracy and Training Loss over the training epochs.

A deep learning model is usually trained on a large dataset containing both real and fake videos, which can be a challenging process. During training the model learns to distinguish between authentic and fraudulent videos by examining factors like pixel values, movement, and audio. Figure 6 represents a parameter known as training accuracy, while training loss is also mentioned, which measures how well the model adapts to the training data. As the model becomes more proficient at differentiating between real and fake videos, the loss should ideally decrease. However, if the training loss is excessively low it could suggest the model is overfitting the training data, which may result in poor performance on unseen data. For this reason, it is essential to carefully monitor the loss, as reflects the models anticipated performance on new untested data.

5.2.1 Performance Evaluation of the Proposed Hybrid Model Achieving Accuracy

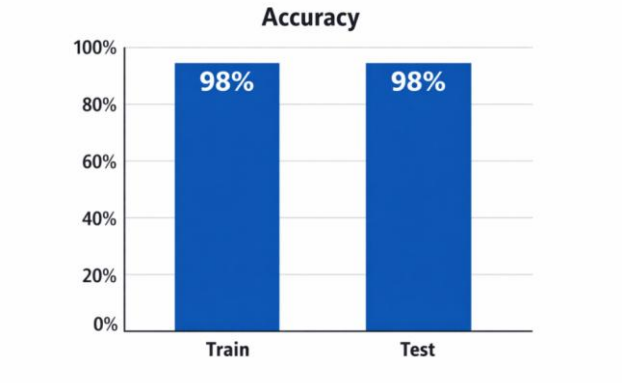


Figure 11. Performance Evaluation of proposed system

Figure 11 illustrates the accuracy performance of the proposed fake face detection model on both training and testing datasets. The graph shows that the model achieves **98% accuracy** for both training and testing phases. The **training accuracy (98%)** indicates that the model has effectively learned distinguishing features between real and fake faces from the training dataset. At the same time, the **testing accuracy (98%)** demonstrates that the model maintains the same level of performance on unseen data, which confirms its strong generalization capability. The equal and high accuracy values suggest that the model does not suffer from overfitting or underfitting. This balance ensures that the system can reliably detect fake faces in real-world scenarios. Overall, **Figure 11** highlights the robustness and efficiency of the proposed fake face detection system, making it suitable for applications such as biometric security, social media content verification, and digital image forensics.

6. CONCLUSION AND FUTURE WORK

This paper presents a neural network based method for distinguishing between Deep Fake and real videos. With the output showing high prediction confidence it can be concluded that the project was successfully completed. Following an in depth analysis of previously proposed algorithm, the project design was developed. The theoretical background of all the resources and technologies used was extensively explained to provide a clear understanding of their application and rationale

Facial manipulation in videos is becoming an increasingly widespread issue. This study carefully reviews relevant literature to gain a deeper understanding of the problem and proposes a network architecture that effectively detects such manipulations using five convolutional neural networks, all while ensuring low computational cost. The paper introduces an innovative approach for detecting Deep Fakes. In this method, a CNN face detector is used to extract facial regions from video frames. The distinct spatial features of these faces are captures using ReLU with CNN, aiding in the detection of visual artifacts in the video frames. Under typical internet conditions, the proposed technology achieves an average detection rate of 98% for Deep Fake videos and 95% for Face2Face video, according to the study's results. The study has demonstrated that CNN performance can be enhanced by adding more convolutional layers and other specific parameters. It also takes into account the compression factors, which presents significant challenges for many Deep Fake detection systems. Future algorithms are expected to concentrate on these issues while utilizing updated datasets. Although this research focused in identifying Deep

Fake in still images and videos, we believe that this method identifying Deep Fake in still images and videos, we believe that this method could be applied to detect Deep Fake in audio and text, aiding in the fight against misinformation in the digital age. These areas will be explored in feature investigations.

REFERENCES

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

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

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

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

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

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

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

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

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

[10] Olivia Holmes, Martin S Banks, and Hany Farid, "*Assessing and improving the identification of computer generated portraits*", In: ACM Transactions on Applied Perception (TAP) 13.2 (2016), pp. 1–12.

[11] Ahmed SRA, Sonuç E, "*Deepfake detection using rationale-augmented convolutional neural network*", Appl Nanosci. 2023; 13:1485–93. https://doi.org/10.1007/s13204-021-02072-3.

[12] M. Masood, M. Nawaz, A. Javed, T. Nazir, A. Mehmood, R. Mahum, "*Classification of deep fake videos using pre-trained convolutional neural networks*", in: 2021 International Conference on Digital Futures and Transformative Technologies (ICoDT2), IEEE, 2021, pp. 1–6.

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

[28] Yuta Yanagi, Ryohei Orihara, Yuichi Sei, Yasuyuki Tahara, and Akihiko Ohsuga. 2020. Fake news detection with generated comments for news articles. In *Proceedings of the 2020 IEEE 24th International Conference on Intelligent Engineering Systems (INES'20)*. IEEE, 85–90.

[29] Youngkyung Seo and Chang-Sung Jeong. 2018. FaGoN: Fake news detection model using grammatic transformation on neural network. In *Proceedings of the 2018 13th International Conference on Knowledge, Information and Creativity Support Systems (KICSS'18)*. IEEE, 1–5.

[30] G. A. Rajesh Kumar, Ravi Kant Kumar, and Goutam Sanyal. 2017. Discriminating real from fake smile using convolution neural network. In *Proceedings of the 2017 International Conference on Computational Intelligence in Data Science (ICCIDS'17)*. IEEE, 1–6.

[31] Xishuang Dong, Uboho Victor, and Lijun Qian. 2020. Two-path deep semisupervised learning for timely fake news detection. *IEEE Transactions on Computational Social Systems* 7, 6 (2020), 1386–1398.

[32] PeisongHe, Haoliang Li, and HongxiaWang. 2019. Detection of fake images via the ensemble of deep representations from multi color spaces. In *Proceedings of the 2019 IEEE International Conference on Image Processing (ICIP'19)*. IEEE, 2299–2303.

[33] Sunidhi Sharma and Dilip Kumar Sharma. 2019. Fake news detection: A long way to go. In *Proceedings of the 2019 4th International Conference on Information Systems and Computer Networks (ISCON'19)*. IEEE, 816–821.

[34] Paulo Roberto Da Cordeiro, Vladia Pinheiro, Ronaldo Moreira, Cecilia Carvalho, and Livio Freire. 2019. What is real or fake?-Machine learning approaches for rumor verification using stance classification. In *Proceedings of the IEEE/WIC/ACM International Conference on Web Intelligence*. 429–432.

[35] P. Korshunov, S. Marcel, "*Deep fakes: a new threat to face recognition? Assessment and detection*", arXiv preprint arXiv: 1812.08685 (2018).

[36] H. S. Shad, M. Rizvee, N. T. Roza, S. Hoq, M. Monirujjaman Khan , A. Singh, A. Zaguia, S. Bourouis, et al, "*Comparative analysis of deep fake image detection method using convolutional neural network*", Computational Intelligene and Neuroscience 2021 (2021).

[37] F. Song, Z. Guo, D. Mei, "*Feature selection using principal component analysis*", in: 2010 international conference on system science, engineering design and manufacturing informatization, volume 1, IEEE, 2010, pp. 27–30.

[38] A. Singh, A. S. Saimbhi, N. Singh, M. Mittal, "*Deep fake video detection: a time-distributed approach*", SN Computer Science 1 (2020) 1–8.

[39] Y. Zhu, Y. Chen, X. Li, R. Zhang, X. Tian, B. Zheng, Y. Chen, "*Information-Containing Adversarial Perturbation for Combating Facial Manipulation Systems*", IEEE Transactions on Information Forensics and Security 18 (2023) 2046–2059. URL: https://ieeexplore.ieee. doi:10.1109/TIFS.2023.3262156.

[40] S.-Y. Wang, O. Wang, R. Zhang, A. Owens, A. A. Efros, "*CNN-generated images are surprisingly easy to spot... for now*", 2020. URL: http://arxiv.org/abs/1912.11035, arXiv: 1912.11035.

[41] Munir, Kashif & Raza, Ali & Almutairi, Mubarak. (2022), "*A Novel Deep Learning Approach for Deepfake Image Detection*", Applied Sciences. 12. 10.3390/app12199820.

[42] Kumar, A.; Bhavsar, A.; Verma, R, "*Detecting Deep Fakes with metric learning",* In Proceedings of the 2020 8th International Workshop on Biometrics and Forensics (IWBF), Porto, Portugal, 29–30 April 2020; pp. 1–6.