# Learning Approaches used by Different Applications to Achieve Deep Fake Technology

- Ashish Maharjan (Computer Engineer, COAC, Germany)
Asish Shakya* (Kathmandu Model College, Bagbazar, Kathmandu, Nepal)

## Abstract:

*Deepfake technology is an emerging field that has gained considerable attention in recent years. Deepfakes are synthetic media, including images, videos, and audio recordings, that are manipulated by advanced machine learning algorithms to produce convincing yet entirely artificial content. This paper explores the various applications and the technologies used by them to achieve deep fake. The machine learning algorithms and the software are used by each of them for proper execution of the technology. Further, we discuss the future prospects of the deepfake technology and explore future directions for research and development in this area, including the need for improved detection and verification techniques and increased education and awareness among the public.*

**Keywords:** deepfake, machine learning, artificial, detection, verification

## Introduction:

Fake news has grown to be a problem that threatens democracy, human civilization, and public dialogue in recent years (Borges et al., 2018). The term "fake news" describes fictional news-style content that has been created to mislead the audience (Aldwairi & Alwahedi, 2018). Via social media, false information spreads swiftly and has the potential to affect millions of users (Figueira & Oliveira, 2017). Currently, YouTube is the second-most popular source of news for Internet users behind Facebook (Anderson, 2018). The popularity of video is growing, which emphasizes the need of tools to verify the veracity of news and media material because new technologies enable convincing video modification (Anderson, 2018).

Modern Technological developments in modern time made it simple to produce the so-called "deepfakes," which are hyper-realistic videos generated with face swaps that barely reveal the manipulation (Chawla, 2019). Artificial intelligence (AI) programs produce deepfakes, which are fake videos that mimic real ones by combining, replacing, and superimposing real images and video snippets (Maras & Alexandrou, 2018).

Deep fake, as combination of deep learning and fake content involves swapping a person's face for a targeted person's in a video, making the targeted person's face express themselves similarly to the targeted person, and acting as though the target person is saying things that were actually said by another person. Deepfake approaches are used to manipulate facial expressions or swap faces, particularly in images and videos (P. Korshunov, 2018). False movies and photos that circulate online can readily exploit some people, and this problem has recently gained attention (Clarke, 2019). It dubbed "fake content creation" is used as the source while their face is changed in an image or video of the person who is the targeted.

Corresponding Email: asishs@kmcen.edu.np

Deepfake technology relies on a type of machine learning called generative adversarial networks (GANs). GANs consist of two neural networks: a generator network that creates the fake images or videos, and a discriminator network that tries to distinguish between the real and fake ones. The two networks are trained together in a feedback loop until the generator can create images or videos that are indistinguishable from the real ones.

Deepfake technology has gained considerable attention in recent years, primarily due to its potential impact on society. Deepfakes refer to synthetic media, including images, videos, and audio recordings, that are manipulated by advanced machine learning algorithms to produce convincing, yet entirely artificial content. While deepfakes have a wide range of potential applications, such as in entertainment and advertising, their most problemetic use is in the creation of false and misleading information.
The ease and accessibility of deepfake technology has raised concerns about the potential for malicious actors to use it to spread misinformation, manipulate public opinion, and even cause harm. For example, a deep fake video of a political leader is used to spread false information, damage reputations, or incite violence. Furthermore, as deepfake technology advances, it becomes increasingly challenging to distinguish between genuine and manipulated media.
This paper aims to examine the technology that underlies deepfake technology. We discuss various applications that utilize deepfake technology, and its potential uses. Additionally, we will explore future directions for research and development in this area, including the need for enhanced detection and verification techniques, as well as the importance of increased education and awareness among the public.

**Deep Fake Applications:**

In this section, we discuss different applications and technologies they use to achieve deepfake. The machine learning algorithms and the software are used by each of them to properly execute the technology. The information was gathered through various works based on deepfake technology and the various techniques used to achieve it.

**1. Facial Manipulation Detection:**

The paper [1] presents a comparison of four different classifiers for deepfake detection: MesoNet, ResNet-50, VGG-19, and Xception. The dataset used in the study comprises 51,036 face images extracted from video datasets, Celeb-DF, Celeb-DF-v2, and DFDC. The images were pre-processed by reshaping them into 2-dimensional images of size (128,128,1), and pixel values were normalized by dividing them by 255.0 to scale the input features between 0.0 to 1.0. MesoNet-4 is a shallow convolutional neural network designed for video forgery detection, which was used for binary classification of deepfake datasets. The implementation of MesoNet-4 in the paper used TensorFlow, TensorFlow.Keras.preprocessing, and matplotlib libraries. The dataset was downloaded, loaded using TensorFlow ImageDataGenerator, and pre-processed while loading in batches to reduce memory usage. The model's experimental design include normalization and scaling of features.
ResNet-50, a residual neural network with 50 hidden layers, was implemented in TensorFlow. A Sequential model was created with a global average pooling layer followed by a Dense layer. The input was a 16,384-dimensional array converted from the 128×128 image. A sigmoid

activation function was added to the Dense layer, which is commonly used for binary classification models.

VGG-19, a visual geometry group network, was implemented using TensorFlow's Sequential class. The model was trained layer-by-layer, starting with an input image dimension of 128x128x3. The standard VGG-19 model was added to the Sequential model, consisting of 19 layers with multiple pooling layers and two fully connected layers. MAX-pooling was employed, which keeps only the maximum value from a pool. The convolution layer used a matrix to convolve around the input data and extract features.
Xception network, consisting of 36 convolutional layers, was implemented in TensorFlow. A Sequential class was used, and the input image dimension was set to and the model returned a probability distribution over all classes, and the class with the maximum probability was the output.
The research compares four networks for deepfake detection using a fraction of Celeb-DF, Celeb-DF-v2, and DeepFake Detection Challenge datasets. MesoNet is fast but not as accurate as ResNet-50, VGG-19, and Xception. ResNet-50 and VGG-19 have similar accuracy rates but ResNet-50 performs better due to its larger number of feature extraction layers. Xception is the most flexible and robust network, delivering better results than the other networks, but requires more time to train and has a longer inference time. For low-end hardware, MesoNet may be suggested, while VGG-19 is preferable for low to medium-end hardware. Xception is the best option for scenarios without hardware limitations, but if it is too hard on resources, ResNet-50 is the best choice.

**2. Detecting Real Time Deep-Fake Videos Using Active Illumination:**

This paper [2] describes the methodology used in the experiment for generating active illumination, localizing and measuring the pattern of illumination on a face, and determining the consistency between the measured and expected illumination.
The active illumination is achieved by displaying a fixed-size image on the same screen as the video call. The hue of the uniform-color image is shifted over time, yielding a hue value in the range of 0.1307 (yellow-ish) to −0.1307 (magenta-ish), while the value and saturation are fixed at unit value. The face is automatically localized in each video frame using Dlib, and a bounding ellipse is fitted to four facial keypoints on the bridge of the nose, the base of the chin, and each cheekbone.
The contributions of the surrounding environmental lighting and the active illumination are separated by assuming that the face is illuminated with a non-directional white light in the absence of active illumination. An estimate of the facial reflectance is acquired by measuring the average color of the face before the active illumination sequence begins, and the measured color at each facial pixel is divided by this estimate to yield the desired hue of the active illumination. To measure the hue of the active illumination, the RGB value of each facial pixel is divided by the average facial RGB pixel value measured with no active illumination. Each adjusted facial RGB pixel value is converted to HSV, and the facial hues are averaged to yield an estimate of the hue of the active illumination. The circular mean hue is computed using the circular mean equation to account for the circularity of the hue.
Finally, the difference between the measured and expected hues is computed, and the consistency between the two is determined using the circular consistency measure.

3. **Deep Fake Detection (DeepFake Stack):**

The paper [3] describes an experiment that aims to detect manipulated face images using a dataset called FaceForensics++. The dataset contains 1000 real videos downloaded from YouTube, which have been manipulated using three popular manipulation techniques (Deepfake, FaceSwap, and Face2Face). The dataset has been preprocessed by separating each video into image sequences and tracking the face area in each image. Only the face area is considered for classification.
The experiment uses the Deep free Stack algorithm, which combines k base-learners (in this case, 7 deep learning models initialized with ImageNet weights) to produce an enhanced classifier. The algorithm splits the dataset into k training sets and uses each to build a base-learner. For a new data tuple, the DeepfakeStack classifier returns a class prediction based on the votes of the base-learners. The experiment follows the Greedy Layer-wise Pretraining technique to train the base-learners and a CNN-based classifier called DeepfakeStackClassifier (DFC) to serve as the meta-learner. The DFC is embedded in a larger multi-headed neural network to obtain the best combination of predictions from each input base-learner.

4. **Detecting Deep Fake Videos from Aural and Oral Dynamics:**

This paper [4] describes the DFT-MF model for detecting deep fake videos, which uses Convolutional Neural Networks (CNN) to classify fake and real videos. The process involves extracting images from videos, which is resource-intensive, and requires a great amount of time and computing power. To mitigate this issue, the authors used MoviePy, an open-source software written in Python, to cut the video based on certain word occurrences in which the mouth appears open and the teeth are visible.
The DFT-MF model focuses on the area surrounding the mouth, especially the teeth, and crops the mouth area from a face in the frame. Face detection is performed using the Dlib classifier to detect face landmarks and eliminate all unnecessary frames. The mouth can be located through points (49, 68), and the area around the mouth is cropped based on the ratio between each two-point upper lips and the lower lips.
The model excludes all frames that contain a closed mouth by calculating distances between lips, as an image with a closed mouth has no fake value. The open mouth is tracked, which displays the teeth with reasonable clarity, to obtain high accuracy and increase efficiency of the model.

The DFT-MF model uses a supervised CNN to classify videos into fake or real based on a threshold number of the fake frames that are identified in the entire video. The number of words in the sentence, which should contain only five words, is used as a clear sentence indicator, and the standard speech rate of 120 words per minute is considered.
The authors used the Deep Fake Forensics (Celeb-DF) dataset and the Deep Fake Vid-TIMIT dataset to create a new dataset that contains a combination of fake and real videos. The DFT-MF model is compared with other methods, and the results show that the DFT-MF model outperforms other approaches.

## Discussion and Conclusion:

In the era of deep fakes, future diplomats will need to be equipped with the knowledge and skills necessary to navigate the complex landscape of digital media. They will need to be able to

identify and respond to deepfakes effectively, both in their personal and professional lives. This will require a deep understanding of the technology behind deep fakes, as well as the social and political implications of their use.

One key skill that future diplomats will need to possess is the ability to critically evaluate information. They will need to be able to assess the credibility of digital media, including photos, videos, and audio recordings, in order to make informed decisions and communicate effectively with others. They will need to be aware of the potential for deep fakes to be used as a tool for disinformation and propaganda, and be able to recognize when the media has been manipulated.

Another important skill for future diplomats will be the ability to communicate effectively in digital media environments. As the use of social media and other digital platforms continues to grow, diplomats will need to be adept at using these tools to communicate with a wide range of audiences. They will need to understand how to use these platforms to disseminate accurate information and engage with stakeholders, while also being aware of the risks associated with these technologies.

Finally, future diplomats will need to be familiar with the legal and ethical frameworks that govern the use of deepfake technology. They will need to understand the risks associated with the use of deep fakes, as well as the potential legal and ethical implications of their use. They will also need to be able to engage with stakeholders on these issues, including policymakers, civil society organizations, and the private sector.

Overall, the rise of deepfake technology presents significant challenges for future diplomats. However, by equipping themselves with the knowledge and skills necessary to navigate this complex landscape, they will be better prepared to respond to the challenges of the digital age and ensure that accurate information is available to all.

In conclusion, deepfake technology has a wide range of potential applications in various industries, including entertainment, advertising, education, and healthcare. It has the ability to create engaging and interactive content, enhance the user experience, and improve the quality of life for many individuals.

In the entertainment industry, deepfake technology has been used to create realistic digital doubles of actors and to revive deceased performers. This has the potential to revolutionize the film and TV industry and provide new creative opportunities.

In advertising, deepfake technology has the potential to create more engaging and personalized ads that can appeal to a specific audience. It can also be used to create more effective training simulations in the education industry and help healthcare professionals diagnose and treat patients.

While deep fake technology has many potential benefits, it is important to recognize its risks and challenges. The ease with which deepfake videos can be created raises concerns about the potential for malicious actors to use this technology to spread disinformation and manipulate public opinion.

Therefore, it is essential to develop effective detection and verification techniques to identify deepfakes and prevent their spread. Policymakers and researchers must work together to establish guidelines and regulations to ensure the ethical and responsible use of deepfake technology.

In summary, deepfake technology has significant potential for various applications, but it is crucial to approach it with caution and responsibility to mitigate the potential risks and ensure that it is used for positive and constructive purposes.

## References:


Agarwal, S., & Farid, H. (2021). Detecting deep-fake videos from aural and oral dynamics. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition* (pp. 981-989). Computer Vision Foundation

Aldwairi, M., & Alwahedi, A. (2018). Detecting fake news in social media networks. *Procedia Computer Science*, *141*, 215-222.

Anderson, K. E. 2018. Getting acquainted with social networks and apps: Combating fake news on social media. Library hitech news, 35(3). 1–6.

Borges, L., Martins, B., & Calado, P. (2019). Combining similarity features and deep representation learning for stance detection in the context of checking fake news. *Journal of Data and Information Quality (JDIQ)*, *11*(3), 1-26.

Chawla, R. 2019. Deepfakes: How a pervert shook the world. *International Journal of Advance Research and Development*, *4*(6), 4–8.

Clarke, Yvette D, "H.R.3230 - 116th Congress (2019-2020): Defending each and every person from false appearances by keeping exploitation subject to accountability act of 2019"

Figueira, A., & Oliveira, L. 2017. The current state of fake news: Challenges and opportunities. procedia computer science, 121: 817–825.

Gerstner, Candice R., and Hany Farid. "Detecting real-time deep-fake videos using active illumination." *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 2022.

Maras, M. H., & Alexandrou, A. 2019. Determining authenticity of video evidence in the age of artificial intelligence and in the wake of deepfake videos. *International Journal of Evidence & Proof, 23*(3): 255–262.

P. Korshunov and S. Marcel, "Deepfakes: a new threat to face recognition? Assessment and detection," ArXiv preprint arXiv:1812.08685,2018

Pashine, S., Mandiya, S., Gupta, P., & Sheikh, R. (2021). Deep Fake Detection: Survey of Facial Manipulation Detection Solutions. *arXiv preprint arXiv:2106.12605*.

Pashine, Samay, et al. Deep Fake Detection: Survey of facial manipulation detection solutions. *ArXiv preprint arXiv:2106.12605* (2021).

Rana, M. S., & Sung, A. H. (2020, August). Deepfakestack: A deep ensemble-based learning technique for deepfake detection. In *2020 7th IEEE international conference on cyber security and cloud computing (CSCloud)/2020 6th IEEE international conference on edge computing and scalable cloud (EdgeCom)*, 70-75. IEEE.