Tuijin Jishu/Journal of Propulsion Technology
ISSN: 1001-4055
Vol. 45 No. 3 (2024)

Design & Implementation of New Model and Techniques for Detection of Deep Fakes Videos

Prof (Dr.) Abhaya Nand[1],Anshul Kumar[2],Arnav Kaushik[3]
[1] *IIMT College of Management, Greater Noida , U.P. India*
[2] *IIMT College of Management, Greater Noida ,U.P. India*
[3] Galgotias university ,U.P. , India

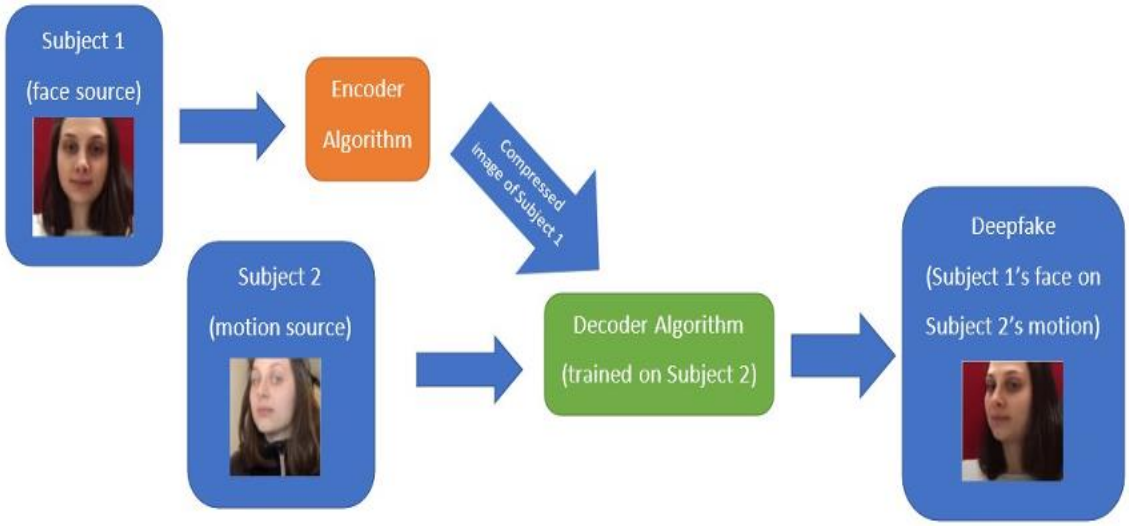
Abstract- DeepFake Detection is tasked with detecting fake videos or images created using deep learning techniques. Deep spoofing is formed via machine learning algorithms to manipulate or replace parts of the original video or image, such as a person's face. The goal of deep fake detection is to detect such manipulations and distinguish them from genuine videos or images. . Recently, the highly-reproduced, lifelike, and altered videos have come to be known as Deepfake. Since then, a number of strategies have been detailed in the literature to address the issues brought up by Deepfake. For instance, AI-powered software tools such as FaceApp and FakeApp are used for creating realistic-looking faces in images and video. This face swapping mechanism enables anyone to change the front appearance, the hairstyle, the gender, the age, and many other personal details. The spread of such fake videos creates a lot of anxiety and has become known as Deepfake.

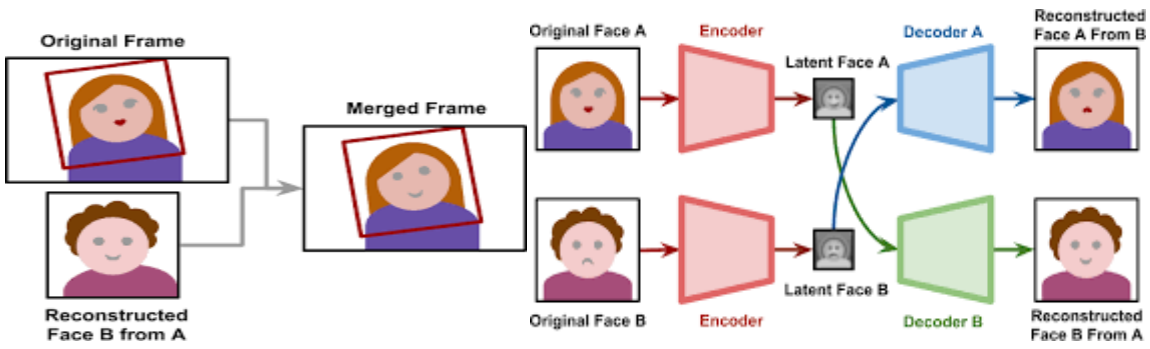
Keywords: Deepfake, Deepfake generation, Deepfake creation, FaceApp

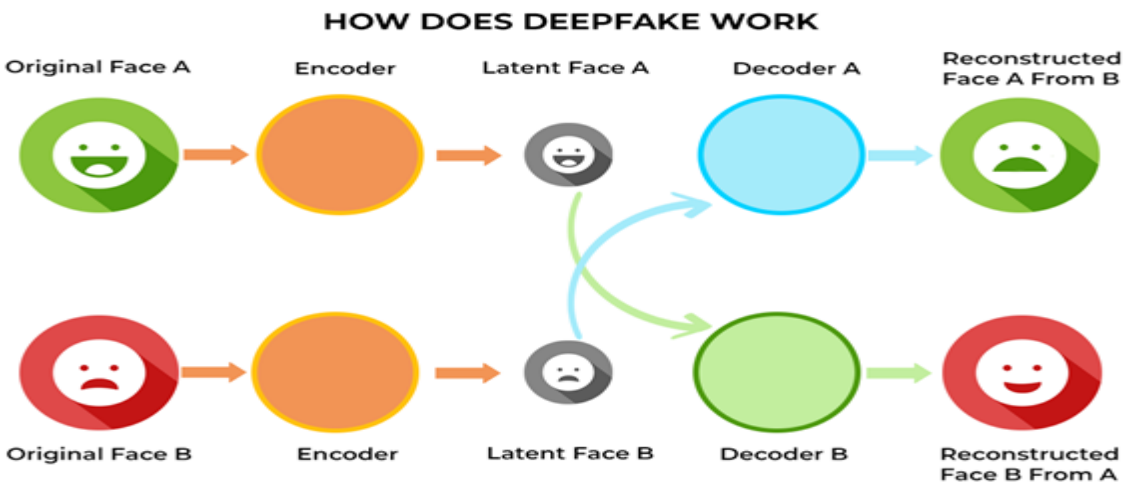
1. Introduction: "Deepfake" combine two things "Deep Learning (DL)" and "Fake." It refers to a meticulous type of fake video or image substance created with DL's prop up[5]. The term "Deepfake" was coined in late 2017 by an anonymous Reddit user who used deep learning methods to substitute a person's face in pornographic videos by using another person's face and creating photo pragmatic fake videos. The two neural networks used to create fake videos were a (a) generative network and (b) a FaceSwap network. The generative network generates fake images using encoders and decoders[3]. The FaceSwap network determines how authentic the newly generated images are. The combination of the two networks is known as Generative Adversary Networks (GANs).

2. Technology needed to create deepfakes

- Generator and Discriminator (GAN) neural network algorithms are used to create all deepfake content.
- Convolutional neural networks are used for facial recognition and movement tracking. CNNs are often used for facial recognition and movement tracking..
- A neural network technology called an autoencoder can identify and impose the relevant attributes of a target, such as facial expressions and body movements, onto the source video[1].
- Deepfake audio is produced with the use of natural language processing. NLP algorithms use a target's voice characteristics to analyse and then create original text based on those characteristics.
- Deepfakes require a substantial amount of processing power, which is provided by high-performance computing.







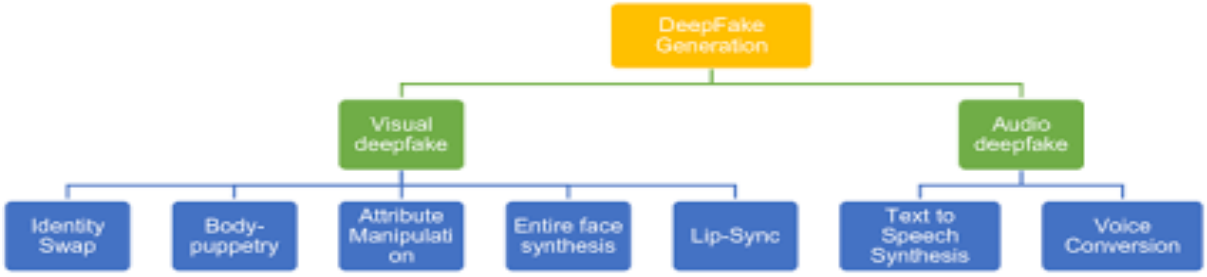
3. The use of deepfakes

Art. Deepfakes are used to generate new music using the existing bodies of an artist's work.

- **Blackmail and reputation harm**. Examples of this are when a target image is put in an illegal, inappropriate or otherwise compromising situation such as lying to the public, engaging in explicit sexual acts or taking drugs. These videos are used to extort a victim, ruin a person's reputation, get revenge or simply cyberbully them[7]. The most common blackmail or revenge use is nonconsensual deepfake porn, also known as revenge porn.
- **Caller response services.** These services use deepfakes to provide personalized responses to caller requests that involve call forwarding and other receptionist services.
- **Customer phone support.** These services use fake voices for simple tasks such as checking an account balance or filing a complaint.
- **Entertainment.** Hollywood movies and video games clone and manipulate actors' voices for certain scenes. Entertainment mediums use this when a scene is hard to shoot, in post-production when an actor is no longer on set to record their voice, or to save the actor and the production team time. Deepfakes are also used for satire and parody content in which the audience understands the video isn't real but enjoys the humorous situation the deepfake creates[6]. An example is the 2023 deepfake of Dwayne "The Rock" Johnson as Dora the Explorer.
- **False evidence**. This involves the fabrication of false images or audio that can be used as evidence implying guilt or innocence in a legal case.
- **Fraud.** Deepfakes are used to impersonate an individual to obtain personally identifiable information (PII), such as bank account and credit card numbers. This can sometimes include impersonating executives of companies or other employees with credentials to access sensitive information, which is a major cybersecurity threat.
- **Misinformation and political manipulation**. Deepfake videos of politicians or trusted sources are used to sway public opinion and, in the case of the deepfake of Ukrainian President Volodomyr Zelenskyy, create confusion in warfare. This is sometimes referred to as the spreading of fake news.
- **Stock manipulation.** Forged deepfake materials are used to affect a company's stock price. For instance, a fake video of a chief executive officer making damaging statements about their company could lower its stock price. A fake video about a technological breakthrough or product launch could raise a company's stock.
- **Texting.** The U.S. Department of Homeland Security's "Increasing Threat of Deepfake Identities" report cited text messaging as a future use of deepfake technology. Threat actors could use deepfake techniques to replicate a user's texting style, according to the report.

4. Are deepfakes legal?

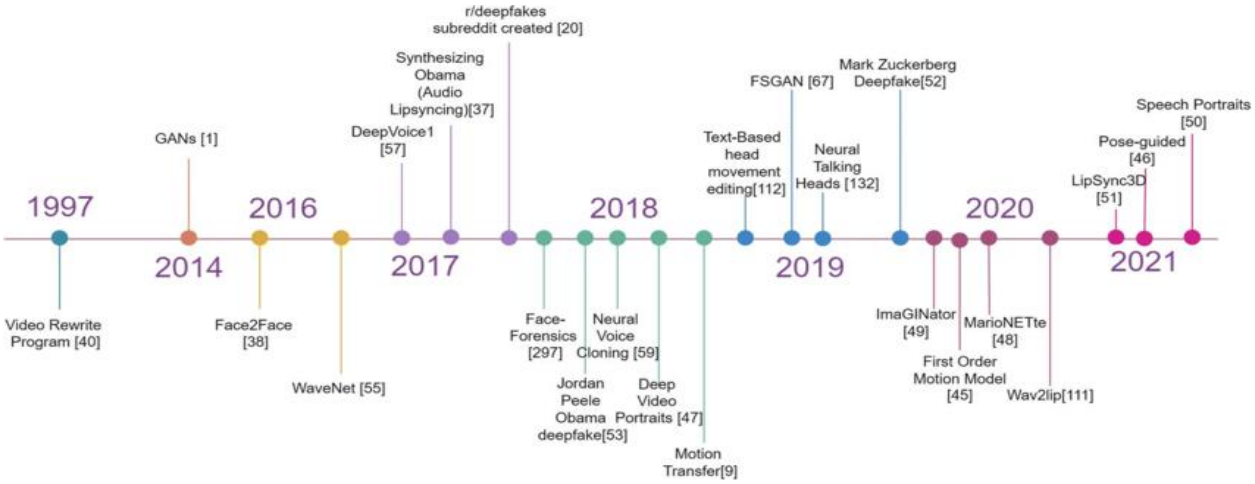
Deepfakes are generally legal, and there is little law enforcement can do about them, despite the serious threats they pose[3]. Deepfakes are only illegal if they violate existing laws such as child pornography, defamation or hate speech. Three states have laws concerning deepfakes. According to Police Chief Magazine, Texas bans deepfakes that aim to influence elections, Virginia bans the dissemination of deepfake pornography, and California has laws against the use of political deepfakes within 60 days of an election and nonconsensual deepfake pornography.



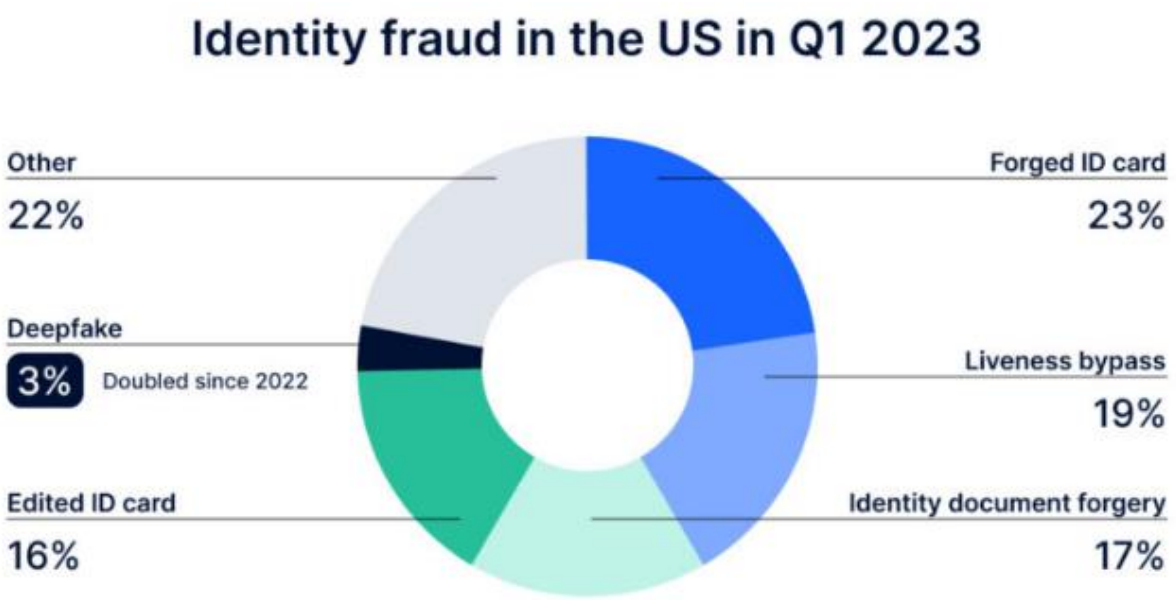
4.1 How are deepfakes dangerous?

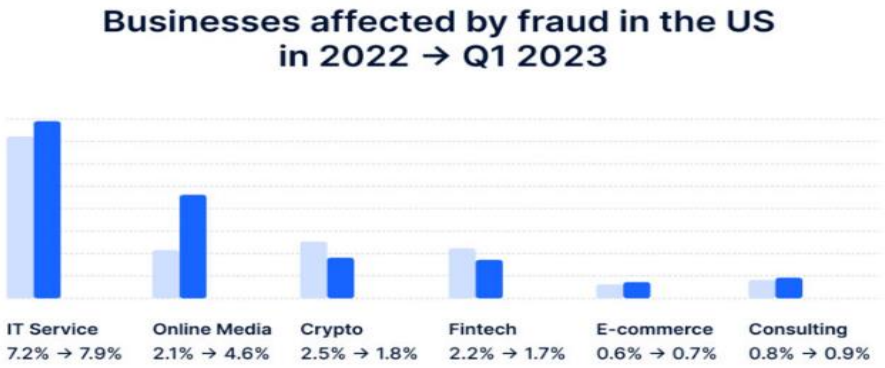
- Deepfakes pose significant dangers despite being largely legal, including the following:
- Blackmail and reputational harm that put targets in legally compromising situations.
- Political misinformation such as nation states' threat actors using it for nefarious purposes.
- Election interference, such as creating fake videos of candidates.
- Stock manipulation where fake content is created to influence stock prices.
- Fraud where an individual is impersonated to steal financial account and other PII.

4.2 Decade related data about deep fake technology



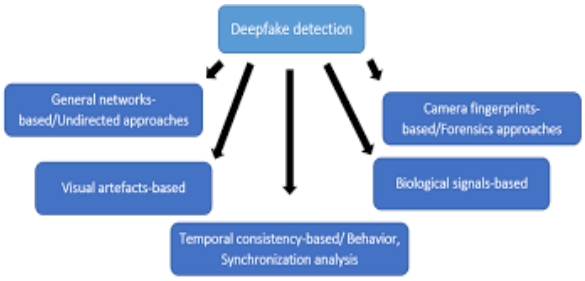
4.4 Deep fake statistics



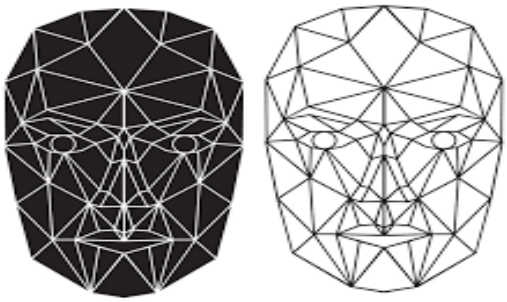


1– The top-5 identity fraud types in 2023 are: AI-powered fraud, money muling networks, fake IDs, account takeovers and forced verification[6].
2– There's been a significant 10x increase in the number of deep fakes detected globally across all industries from 2022 to 2023, with notable regional differences: 1740% deepfake surge in North America, 1530% in APAC, 780% in Europe (inc. the UK), 450% in MEA and 410% in Latin America[7]
3– The country attacked by deepfakes the most is Spain, the most forged document worldwide is UAE passport, whereas Latin America is the region where fraud increased in every country[4].
4– ID cards remain the most frequently exploited for identity fraud, accounting for nearly 75% of all fraudulent activities involving identity documents.
5– Online media is the industry with the highest identity fraud increase.

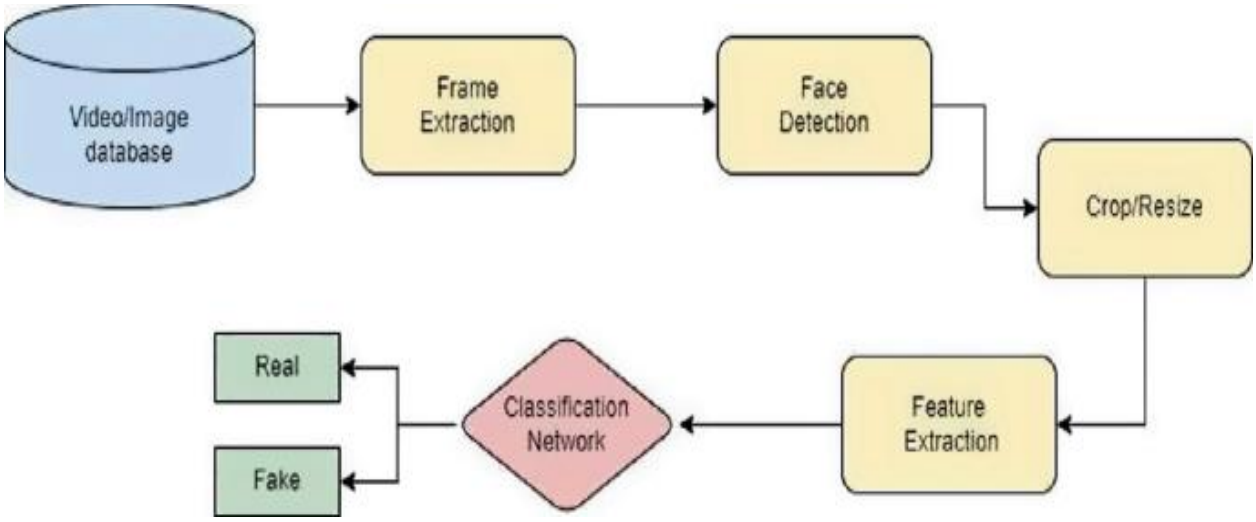
5. Methods to detecting deepfakes



There are several best practices for detecting deepfake attacks. The following are signs of possible deepfake content:



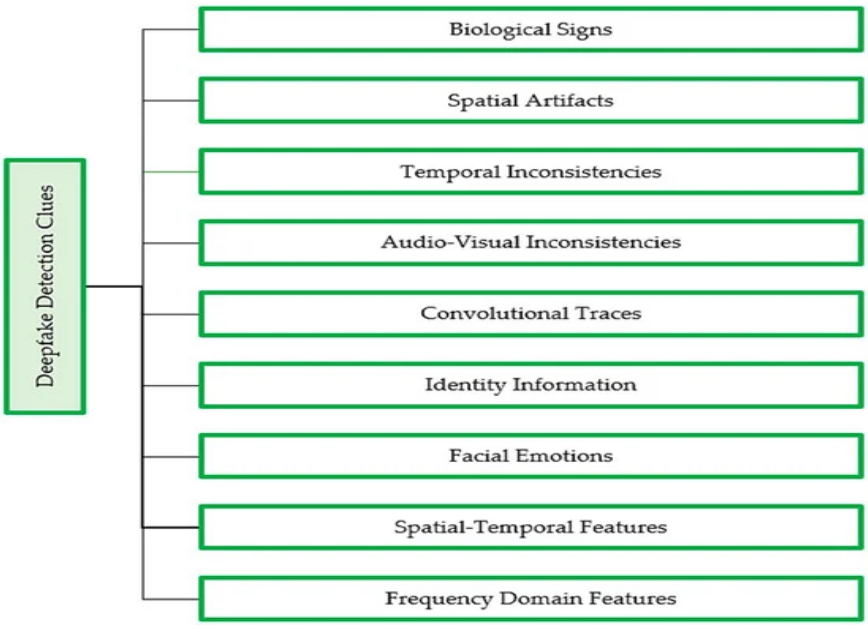
- Unusual or awkward facial positioning.
- Unnatural facial or body movement.
- Unnatural coloring.
- Videos that look odd when zoomed in or magnified.
- Inconsistent audio.
- People that don't blink.



6. Clues for Detecting Deepfakes

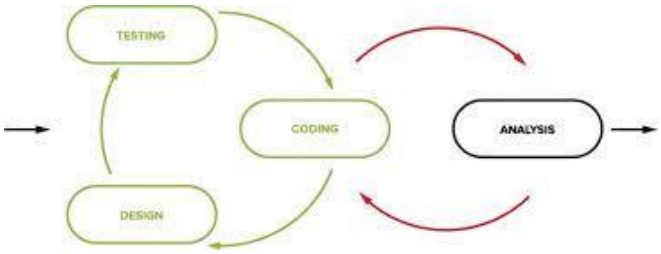
Using a variety of cues, deepfakes can be identified. One method involves closely scrutinising deepfakes for visual anomalies, face landmarks, or intra-frame discrepancies in order to analyse spatial abnormalities[2]. An additional technique is to look for convolutional traces, such as bi-granularity artefacts and GAN fingerprints, which are frequently seen in deepfakes as a consequence of the generation process[9]. Furthermore, biological cues like irregular heartbeat, eye colour, and blinking frequency can also be used to identify the existence of a deepfake[7]. Other biological cues include temporal inconsistencies or discontinuities between adjacent video frames, which can cause jitteriness, flickering, and changes in facial position.In textual deepfakes, there are a few indicators:

- Misspellings.
- Sentences that don't flow naturally.
- Suspicious source email addresses.
- Phrasing that doesn't match the supposed sender.
- Out-of-context messages that aren't relevant to any discussion, event or issue.



7. Deepfake protection software

1. Adobe has a system that lets creators attach a signature to videos and photos with details about their creation.
2. Microsoft has AI-powered deepfake detection software that analyze videos and photos to provide a confidence score that shows whether the media has been manipulated.
3. Operation Minerva uses catalogs of previously discovered deepfakes to tell if a new video is simply a modification of an existing fake that has been discovered and given a digital fingerprint.
4. Sensity offers a detection platform that uses deep learning to spot indications of deepfake media in the same way antimalware tools look for virus and malware signatures. Users are alerted via email when they view a deepfake.

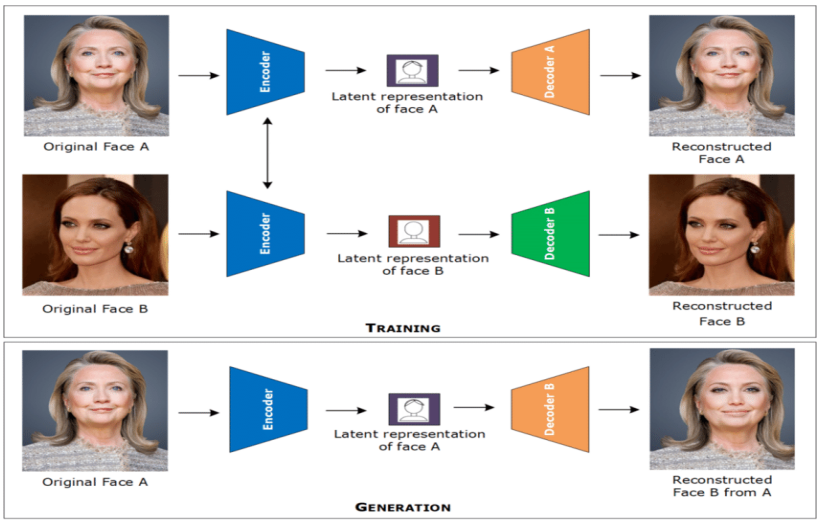


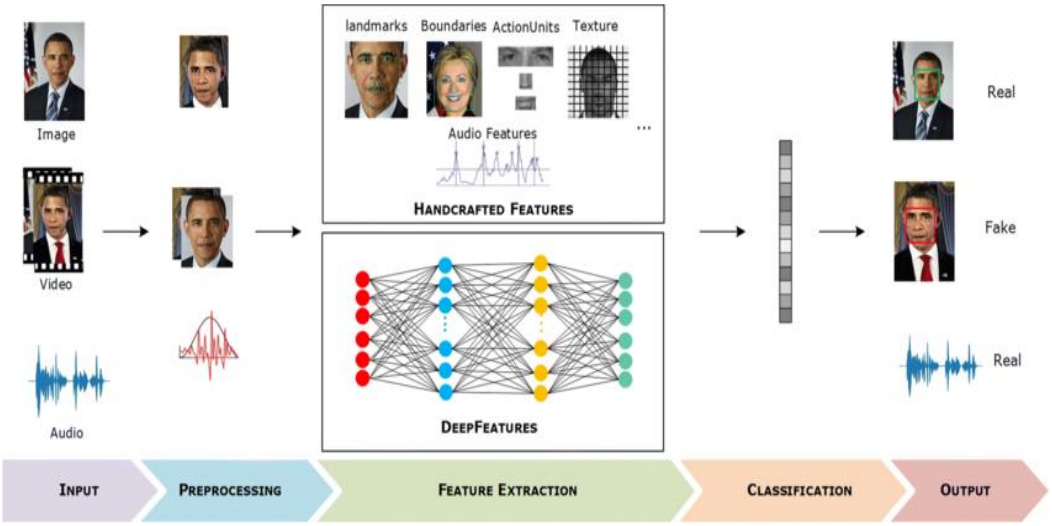
8. Notable examples of deepfakes

There are several notable examples of deepfakes, including the following:

- Facebook founder Mark Zuckerberg was the victim of a deepfake that showed him boasting about how Facebook "owns" its users. The video was designed to show how people can use social media platforms such as Facebook to deceive the public.
- U.S. President Joe Biden was the victim of numerous deepfakes in 2020 showing him in exaggerated states of cognitive decline meant to influence the presidential election. Presidents Barack Obama and Donald Trump have also been victims of deepfake videos, some to spread disinformation and some as satire and entertainment[7].
- During the Russian invasion of Ukraine in 2022, Ukrainian President Volodomyr Zelenskyy was portrayed telling his troops to surrender to the Russians.
- The top 10 pornographic platforms posted 1,790+ Deepfake videos, without concerning pornhub.com, which has removed 'Deepfakes' searches.
- Adult pages post 6,174 Deepfake videos with fake video content.
- 3 New platforms were devoted to distributing Deepfake pornography.
- In 2018, 902 articles were published in arXiv, including the keyword GAN either in titles or abstracts

Working model of deep fake Technology:





How are deepfakes commonly used?

Companies, organizations and government agencies, such as the U.S. Department of Defense's Defense Advanced Research Projects Agency, are developing technology to identify and block deepfakes. Some social media companies use blockchain technology to verify the source of videos and images before allowing them onto their platforms. This way, trusted sources are established and fakes are prevented. Along these lines, Facebook and Twitter have both banned malicious deepfakes.

9. Conclusion: Deep faking is one of the most common methods used to spread disinformation and fake news among the population. While not all deep fake content is malicious, it is necessary to find them because some pose a threat to the entire world. The main objective of this study was to find a reliable method for detecting deep fake images. A lot of researchers have been working hard to find deep fake content using various methods. However, the significance of this study is that it uses DL and ML based methodologies to get good results. Deepfake movies will become increasingly difficult to identify as AI algorithms advance in sophistication. In an effort to stay one step ahead of the curve, this survey paper has offered an extensive overview covering the domain of deepfake generation, the range of deep learning architectures utilised in detection, the most recent developments in detection techniques, and the essential datasets designed to further this field of study.

References

[1] Harwell, D. Scarlett Johansson on fake AI-generated sex videos:'Nothing can stop someone from cutting and pasting my image'. J. Washigton Post 31, 12 (2018).
[2] Amin, R., Al Ghamdi, M. A., Almotiri, S. H. & Alruily, M. Healthcare techniques through deep learning: Issues, challenges and opportunities. IEEE Access 9, 98523–98541 (2021).
[3] Rafique, R., Nawaz, M., Kibriya, H. & Masood, M. DeepFake detection using error level analysis and deep learning. in 2021 4th International Conference on Computing & Information Sciences (ICCIS). 1–4 (IEEE, 2021).
[4] Mansoor, M. et al. A machine learning approach for non-invasive fall detection using Kinect. Multimed. Tools Appl. 81(11), 15491–15519 (2022).
[5] McCloskey, S. & Albright, M. Detecting GAN-generated imagery using saturation cues. in 2019 IEEE International Conference on Image Processing (ICIP). 4584–4588. (IEEE, 2019).
[6] Khalil, S.S., Youssef, S.M. & Saleh, S.N.J.F.I. iCaps-Dfake: An Integrated Capsule-Based Model for Deepfake Image and Video Detection. Vol. 13(4). 93 (2021).
[7] Cozzolino, D., Thies, J., Rössler, A., Riess, C., Nießner, M. & Verdoliva, L.J.A.P.A. Forensictransfer: Weakly-Supervised Domain Adaptation for Forgery Detection (2018).

[8] Anaraki, A. K., Ayati, M. & Kazemi, F. J. Magnetic resonance imaging-based brain tumor grades classification and grading via convolutional neural networks and genetic algorithms. Information 39(1), 63–74 (2019).
[9] Szegedy, C. et al. Going deeper with convolutions. in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 1–9 (2015).
[10] Chandani, K. & Arora, M. Automatic facial forgery detection using deep neural networks. in Advances in Interdisciplinary Engineering. 205–214 (Springer, 2021).