Unmasking Deepfakes Advancements, Challenges, and Ethical Considerations

Usha Kosarkar and Gopal Sakarkar

Abstract Deepfake technology, powered by deep learning algorithms, has rapidly evolved in recent years, enabling the creation of highly realistic synthetic media that can deceive human perception. Unmasking deepfakes: Advancements, Challenges, and Ethical Considerations is a comprehensive review that examines the advancements in deepfake technology, the associated challenges, and the ethical considerations that arise in the context of this emerging field. Deepfake algorithms have the ability to produce phoney audiovisual content that is hard to distinguish from authentic content. It now appears to be challenging to distinguish between authentic digital content and fraudulent content spread around the Internet in this era of the cyber age. Cybercriminals frequently employ this technology to trick security systems. If we are not careful, deepfake technology could pose a severe danger to identity verification in the future. Deepfake content may easily be produced by amateurs using free and open-source software, which makes it simple for them to produce technically excellent content. Give an introduction to deepfake, and a brief on deepfake creation and detection techniques.

Keywords Deepfake · GAN · CNN · RNN · Deepfake detection

1 Introduction

Artificial intelligence (AI) and machine learning algorithms are used in "deepfake technology" to produce incredibly lifelike and frequently false movies, audio recordings, or photographs. It combines techniques from computer vision, graphics, and natural language processing to manipulate or fabricate media content in a way that appears authentic and convincing. Deepfakes are typically created by training deep

U. Kosarkar (✉) · G. Sakarkar
Department of Computer Science, G H Raisoni University Saikhede (MP), Narsinghpur, India
e-mail: usha.kosarkar@raisoni.net

G. Sakarkar
MIT World Peace University, Pune, India

R. N. Shaw et al. (eds.), *Innovations in Electrical and Electronic Engineering*, Lecture Notes in Electrical Engineering 1115, https://doi.org/10.1007/978-981-99-8661-3_19

learning models on large datasets of real footage and then using those models to generate or alter content. The term “deepfake” is derived from the deep learning algorithms employed in the process. Initially, the technology gained attention for its ability to superimpose one person’s face onto another person’s body in video footage, often with remarkably convincing results. The advancement of deepfake technology has raised concerns due to its potential misuse and the ethical implications it carries [1]. Deepfakes can be used to convey false information, sway public opinion, smear people, or produce fictitious celebrity pornography. The emergence of deepfakes has also spurred debates regarding the issues relating to the reliability and authenticity of digital media. To counter the negative effects of deepfakes, researchers and tech companies are actively developing detection tools and techniques to identify manipulated content. Additionally, there have been calls for increased awareness, media literacy, and responsible use of technology to mitigate the risks associated with deepfakes. Despite the fact that deepfake technology has caused certain concerns, it also has useful applications [2, 3]. It can be used in the film industry for visual effects, animation, and virtual reality experiences. Researchers are exploring potential uses in fields such as healthcare, education, and entertainment, where deepfakes can be employed responsibly and ethically.

As deepfake technology continues to evolve, it is crucial to understand its capabilities, risks, and impact on society. Ongoing research, development of safeguards, and public awareness are essential to navigate the challenges and opportunities presented by this rapidly advancing technology [4].

Examples of Deepfakes

Deepfake technology has advanced rapidly in recent years, allowing for the creation of highly realistic manipulated videos and images. Here are a few examples of deepfake applications:

- Celebrity Impersonations: Deepfakes have been used to superimpose the faces of celebrities onto the bodies of actors in movies or onto characters in video games. For example, in the movie “Rogue One: A Star Wars Story,” the late actor Peter Cushing’s face was recreated using deepfake techniques to bring back his character, Grand Moff Tarkin.
- Political Figures: Deepfakes have been employed to create manipulated videos of politicians, altering their speeches or actions to convey false information or misleading narratives. Such videos can have significant implications for spreading misinformation and influencing public opinion.
- Adult Content: Deepfake technology has been widely misused to create explicit content featuring the faces of non-consenting individuals, often celebrities or acquaintances. This has raised concerns about privacy, consent, and the potential for harassment.
- Historical Figures: Deepfakes have been used to recreate historical figures, allowing people to see what they might have looked and sounded like. For instance, a deepfake of Mona Lisa was created to bring Leonardo da Vinci’s famous painting to life, adding facial expressions and animation.

- Voice Manipulation: Deepfake algorithms can also be applied to audio, allowing for the synthesis of speech that mimics the voice of a particular individual. This has raised concerns about the potential for impersonation and fraud, as voices can be convincingly imitated.

Despite the fact that deepfake technology has many innovative and entertaining uses, it also raises serious ethical questions. Misuse of deepfakes can lead to the spread of false information, privacy violations, and the erosion of trust in digital media.

2 Deepfake Creation

Deepfake creation is the process of generating or manipulating realistic-looking films or photographs of people using artificial intelligence and machine learning algorithms, frequently by superimposing their faces onto already-existing footage. Although I can give a general overview of the techniques used, it's crucial to use deepfake technology properly and ethically because it can have both beneficial and harmful applications. Here are some common tools and techniques used in deepfake creation.

Generative adversarial networks (GANs): A generator and a discriminator make up the two parts of the common deep learning model known as GANs. While the discriminator tries to tell the difference between genuine and fraudulent content, the generator produces the deepfake pictures or videos. Together, these models are trained so that the outcomes become more and more realistic.

Face recognition and alignment: Deepfake algorithms often rely on facial recognition techniques to identify and extract key facial features from the source and target videos. These features are then aligned to ensure accurate mapping between the two faces [1].

Facial landmark detection: Facial landmark detection techniques are used to correctly match the facial features. The ability to precisely manipulate and map facial expressions is made possible by these algorithms, which recognise particular locations on a face like the corners of the mouth, nose, and eyes.

Autoencoders: Autoencoders are neural network models that can learn efficient representations of input data. They are used in deepfake creation to extract and encode facial features from the source face, allowing the generation of realistic-looking facial expressions on the target face. Deep neural networks: Deep neural networks are used to train models on large datasets of source and target faces. By feeding them with a vast amount of data, the networks learn to understand and recreate the visual characteristics and patterns of the target person.

Data collection and preprocessing: It frequently takes a sizable amount of training data to create a deepfake. The deep learning models may be trained using photographs or videos of the target person that creators have collected from various sources.

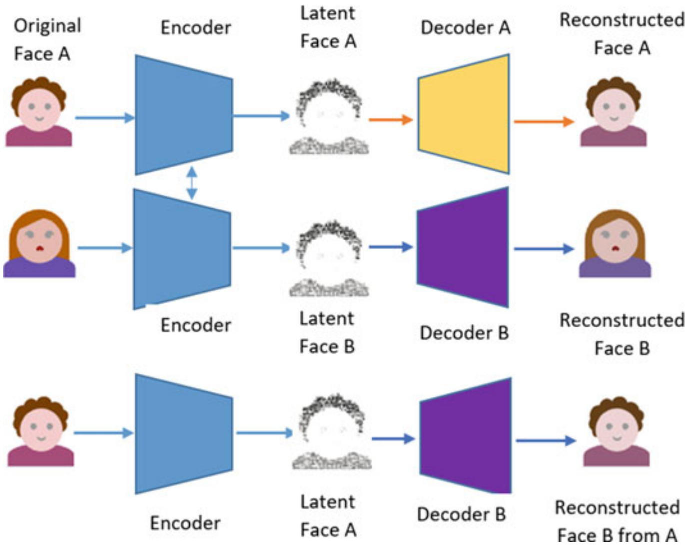

Fig. 1 Illustration of deepfake creation process using two encoder–decoder pair

To ensure uniformity and the best training, preprocessing operations like cropping, scaling, and normalisation may also be carried out [1, 2].

Video synthesis: Deepfake techniques can involve manipulating individual frames of a video or synthesising an entirely new video. Techniques like frame interpolation or blending are used to seamlessly merge the facial expressions and movements of the target person into the source video.

It's important to note that deepfake technology has sparked worries about false information, privacy, and possible abuse. To uphold moral standards and prevent harm, deepfake technology must be used responsibly and its ramifications must be understood (Fig. 1).

3 Deepfake Detection Techniques

- Deepfake detection methods seek to pinpoint and reduce the dangers posed by modified or artificial media material. Here are a few methods that are frequently used to identify deepfakes. Digital Forensics: This method looks for discrepancies or tampering by looking at numerous artefacts and traces within the digital file. Analysing metadata, compression artefacts, noise patterns, and other forensic hints that could point to manipulation is part of this process [4, 5].

- Face and Body Manipulation Detection: Facial or body manipulation is a common feature of deepfake films. To spot potential deepfake aspects, detection systems can examine facial landmarks, eye blinking patterns, unusual head movements, or inconsistent body proportions [4].
- Texture Analysis: Some deepfake detection techniques examine the texture and picture quality of the video frames. Image analysis techniques can be used to find colour, lighting, or resolution irregularities in artificially created faces or objects.
- Temporal and Statistical Analysis: When deepfake videos are examined over time, anomalies may be seen. For instance, abnormalities in statistical features, odd blinking patterns, or inconsistencies in motion dynamics might all be signs of deepfake.
- Machine Learning Algorithms: On big datasets of genuine and deepfake material, supervised machine learning techniques can be used to train models. Based on a variety of characteristics, including facial expressions, eye movements, and speech patterns, these models may learn to distinguish between real content and content that has been altered [6].
- Biometric Verification: You can identify differences or inconsistencies that could point to a deepfake by comparing the facial biometric data you retrieved from the video with a known reference of the person.
- Blockchain and Watermarking: To confirm the legitimacy and integrity of media content, blockchain technology and digital watermarking can be employed. By incorporating distinctive identifiers or cryptographic signatures into the content, these strategies make it simpler to trace and authenticate the content's origin and guard against tampering.
- Collaborative Filtering: Techniques for collaborative filtering might find suspicious or anomalous content that may be deepfakes by examining patterns and behaviours across various individuals and platforms.
- This strategy uses platforms' and users' collective intelligence to identify potentially misleading media.
- It's important to remember that both the deepfake technology area and the methods of detection are dynamic. To keep up with the growing risks, new detection approaches and algorithms are continually being created as deepfake generating techniques evolve.

4 Evaluatıon Metrics for Deepfake Face Detectıon

It is vital to utilise proper assessment metrics when evaluating the performance of deepfake face detection systems since these metrics give insights into the systems' ability to recognise fake faces accurately and effectively. Researchers and developers are able to evaluate various detection approaches, quantify how well they operate, and pinpoint areas where they may be improved, thanks to these measurements. In this piece of writing, we will investigate some of the most typical assessment metrics that are used for deepfake face identification.

1. Accuracy: The accuracy of the deepfake detection system is one of the primary assessment metrics employed in this system. The overall accuracy of the detection system is determined by computing the ratio of samples that were properly identified (including real and deepfake data) to the total number of samples. This provides a comprehensive evaluation of the system's performance. A detection system that is more trustworthy is indicated by a greater accuracy, but this metric alone should not be the exclusive focus of an assessment; other metrics need to be examined as well.
2. Precision and Recall: Precision and recall are two measures that are often used in the process of assessing the effectiveness of binary classification systems such as deepfake detectors. Precision is measured by the fraction of deepfake samples that are properly detected out of the total number of samples that are labelled as deepfakes. On the other hand, recall calculates the percentage of deepfake samples out of the total number of genuine deepfake samples that have been accurately detected. Precision and recall are equally crucial, and the optimal balance between the two will vary depending on the particular application and the results that are intended.
3. True Positive Rate (TPR) and False Positive Rate (FPR): The true positive rate (TPR) and the false positive rate (FPR) quantify the proportion of deepfake samples that are properly detected in comparison with the total number of genuine deepfake samples. In certain circles, it is also referred to as sensitivity or recall. The false positive rate, or FPR, is the proportion of authentic samples that were wrongly identified as deepfakes out of all authentic samples. When designing a detection system, it is important to strike a balance between achieving a high TPR and maintaining an FPR that is as low as feasible.
4. Curve of the Receiver Operating Characteristic (also known as ROC): The ROC curve is a graphical depiction of the trade-off between the true positive rate (TPR) and the false positive rate (FPR) at different classification thresholds. It shows how the TPR and FPR change depending on the classification threshold. It makes it possible to evaluate the performance of a detection system at a variety of operating points throughout the board. When evaluating the overall effectiveness of a deepfake detection system, the area under the ROC curve (also known as AUC-ROC) is often employed as a single statistic to do so. Better performance is indicated by a higher AUC-ROC value.
5. The F1 score is a measure of the overall accuracy of a binary classification system that takes into consideration both precision and recall. This score is also known as the F1 index. It is the harmonic mean of accuracy and recall, and it offers a single number to measure the degree to which these two metrics are in harmony with one another. A score of 1 on the F1 scale indicates flawless accuracy and recall balance. The scale runs from 0 to 1.
6. Specificity is the proportion of properly recognised genuine samples relative to the total number of real genuine samples. Specificity is measured as a percentage. The false negative rate is equal to one minus the false positive rate (1–FPR). The need to be specific arises most often in contexts in which it is essential to

protect the authenticity of the information being presented, such as in forensic investigations and legal procedures.

7. Area under the Precision-Recall Curve (AUC-PR): The AUC-PR is the same as the AUC-ROC, in that it refers to the area under the curve that measures precision and recall. It offers a thorough analysis of the performance of a detection system by taking into account accuracy and recall over a range of categorisation criteria. Better performance is indicated by a greater AUC-PR, particularly in circumstances in which the distribution of the class is unbalanced.
8. Cross-Validation and Validation Set Performance: When evaluating the generalisation performance of deepfake detection systems, cross-validation methods such as k-fold cross-validation are used in order to collect the necessary datasets and samples. Cross-validation is a method that helps avoid overfitting problems and gives a more accurate evaluation of the performance of a model. This is accomplished by dividing the dataset into training and validation sets and performing the evaluation procedure many times.

It is essential to keep in mind that the precise objectives and prerequisites of the deepfake detection job dictate the metrics that should be chosen for assessment. It's possible that some metrics are more suited for certain situations or applications than others are. In order to acquire a thorough knowledge of the performance and limits of a detection system, researchers and practitioners need to take into consideration a mix of these metrics.

In conclusion, assessment measures are an extremely important component in the process of determining how well deepfake face detection systems function. Researchers are able to statistically quantify the accuracy, precision, recall, and other key elements of a detection system by utilising proper metrics, which enable informed decision-making and the continual advancement of deepfake detection approaches.

5 State-of-the-Art Deep Learning Approaches for Deepfake Face Detection

The fast developments in deepfake generating methods have led to the creation of advanced deep learning systems for the identification of deepfake faces. The ability of deep learning algorithms to detect and recognise altered facial information has been shown to be extraordinary. In this piece, we take a look at some of the cutting-edge deep learning strategies that have been found to have potential in the area of deepfake face detection.

1. Convolutional Neural Networks, abbreviated as "CNNs": Convolutional neural networks, or CNNs, have been shown to be very successful in a variety of computer vision applications, including the identification of deepfakes. CNNs are excellent at automatically learning and extracting meaningful characteristics from pictures, which makes them well-suited for the task of recognising the visual artefacts and inconsistencies that are characteristic of deepfake faces.

Training deep architectures that have many convolutional layers is often a need for state-of-the-art CNN-based methods that are used for deepfake detection. These networks gain the ability to distinguish between genuine and fake faces by identifying minute visual signals and inconsistencies in the information that has been modified. The training data contains both authentic and deepfake pictures, which enables the CNN to learn the distinguishing characteristics that allow it to differentiate between the two types of pictures.

2. Recurrent Neural Networks (RNNs): Recurrent neural networks (RNNs) are a kind of neural network that is often used for the purposes of sequence analysis and temporal modelling. Deepfake face detection is one application that makes use of them to analyse video sequences in order to identify temporal correlations and trends. RNN-based techniques centre on doing an analysis of the stability, over time, of a subject's facial movements, expressions, or lip-syncing.

   RNNs are able to detect temporal anomalies and inconsistencies that are symptomatic of deepfake manipulation because of the sequential nature of video frames, which is taken into consideration by RNNs. RNN architectures, such as long short-term memory (LSTM) and gated recurrent unit (GRU) are often utilised for deepfake detection applications.

3. Siamese Networks: Siamese networks are a kind of deep learning architectures that were developed particularly for jobs that are similar to one another. They have been used in the field of deepfake face detection via the process of learning similarity metrics between different combinations of faces. Siamese networks are made up of two identical networks that share their weights and learn to provide similarity scores based on the face pairings that are fed into them.

   Siamese networks are trained using real and deepfake face pairings during the training phase, with the goal of increasing the similarity scores for genuine face pairs while decreasing the scores for deepfake face pairs as much as possible. The network is able to learn fine-grained characteristics that differentiate between actual and altered faces as a result of this method, which contributes to its effectiveness in recognising deepfakes.

4. Generative Adversarial Networks (GANs): Generative adversarial networks (GANs) have been used in the process of deepfake production as well as detection. In the domain of deepfake face detection, GANs are used to learn the distribution of real faces and find differences between produced and real facial pictures. This is done by learning the distribution of genuine faces.

   Detection models that are based on GANs consist of a discriminator network that learns to differentiate between real and deepfake faces and a generator network that synthesises deepfake faces. The discriminator network is trained to learn how to identify genuine and deepfake faces. The discriminator network is trained to minimise the error in classification between genuine and fake faces, while the generator network tries to produce more convincing deepfakes. Both networks are trained by feeding them examples of real and fake faces. The detection model's capacity to recognise even the most modest artefacts and inconsistencies is improved by the use of an adversarial training procedures.

5. Capsule Networks: Capsule networks are a relatively new advancement in the field of deep learning that has shown a great deal of promise in a variety of computer vision applications. By explicitly modelling the hierarchical connections that exist between visual elements, capsule networks hope to overcome the constraints that CNNs possess.

When it comes to deepfake face identification, capsule networks have the ability to record spatial connections and posture changes in facial characteristics. This paves the way for detection that is more robust and dependable. These networks make use of dynamic routing algorithms to learn capsules that represent various facial traits. This enables the detection of inconsistencies and abnormalities in deepfake faces.

It is important to note that the most cutting-edge deep learning systems for deepfake face detection are always being improved as researchers experiment with different kinds of architectures, loss functions, and training methods. The detection performance may be further improved by combining a variety of deep learning models, ensemble approaches, and domain-specific modifications.

To summarise, effective tools for deepfake face detection have evolved in the form of deep learning algorithms such as CNNs, RNNs, Siamese networks, GANs, and capsule networks. These models take use of the capability that deep neural networks provide in order to learn discriminative features, capture temporal dependencies, utilise similarity metrics, and find inconsistencies in modified face information sets. Research and development in deep learning techniques must continue unabated if we are to keep one step ahead of the constantly advancing deepfake creation algorithms and guarantee accurate identification of manipulated media sets.

6 Reviews of Existing Techniques

Due to the ever-increasing sophistication of deepfake technology, it is very necessary to create efficient methods for identifying modified face information sets. In this piece, we will present a summary of the many algorithms currently available for deepfake face detection, along with a discussion of the benefits and drawbacks of each.

1. Image and Video Forensics: The methods used in image and video forensics entail the examination of a variety of visual artefacts and anomalies in altered information sets. These methods concentrate on identifying irregularities, such as lighting that is not constant, face motions that are not natural, or facial characteristics that are not aligned correctly. They often depend on handmade characteristics to detect possible deepfake information, such as noise patterns, edge discontinuities, or compression artefacts.

Even though image and video forensics may be helpful in identifying deepfakes that are basic or of poor quality, they may have difficulty identifying more complex and realistic modifications that are intended to resemble actual faces more precisely.

2. Feature-Based Techniques: Feature-based techniques for deepfake detection concentrate on extracting and analysing particular face traits or properties that are difficult to recreate properly in deepfake material. This is done in order to identify fake content. These characteristics may include patterns of eye blinking, variances in face micro-expressions, or differences in blood flow. It is easy to distinguish between real and altered faces by conducting an examination of these traits and looking for differences.

Approaches that are based on features have the benefit of capturing unique qualities that are difficult for deepfake algorithms to recreate properly. This presents a challenge for those developing deepfakes. On the other hand, their capacity to identify sophisticated deepfakes that properly duplicate these qualities could be restricted.

3. Approaches That Are Based on Deep Learning: As a result of its capacity to automatically learn discriminative features from extensive datasets, deep learning-based techniques have attracted a substantial amount of interest in the field of deepfake face detection. Convolutional neural networks (CNNs), recurrent neural networks (RNNs), and generative adversarial networks (GANs), or any mix of these architectures, are used by these methods in order to identify instances of deepfake materials [7, 8].

Techniques that are based on deep learning have the ability to catch tiny visual signals and patterns that are indicative of deepfake manipulation. As a result, these techniques are successful in identifying a broad variety of deepfake variants. However, they often need a significant quantity of training data in addition to training procedures that are computationally costly.

4. Capsule Networks: Capsule networks are a relatively recent invention in the field of deep learning. In comparison with conventional CNNs, capsule networks take a unique approach to the encoding of feature datasets and samples. Capsule networks are a kind of artificial neural network that simulates hierarchical connections between visual items. These networks make it possible to recognise spatial correlations as well as posture changes in face characteristics.

By gathering more sophisticated information about face characteristics and identifying abnormalities in the hierarchical representation, capsule networks have showed promise in the detection of deepfake material. However, their usefulness in identifying deepfakes is still a topic of investigation in this field of study.

5. Hybrid Approaches: Hybrid approaches boost the overall detection performance by combining different detection methods, such as image forensics, feature-based analysis, and deep learning-based methods. Hybrid methods strive to increase the accuracy and resilience of the deepfake detection process by exploiting the complimentary characteristics of multiple methodologies.

These methods often make use of ensemble models, which involve combining the findings of several detectors into a single conclusion. This aids in lowering both false positives and false negatives, resulting to a deepfake detection process that is more trustworthy.

In conclusion, the many strategies that are currently available for deepfake face detection include image and video forensics, feature-based analysis, deep learning-based methods, capsule networks, and hybrid models. Each method has advantages and disadvantages, and the degree to which it is successful in detecting deepfake material may vary depending on the level of complexity of the fake. In order to address the ever-increasing risk posed by manipulated media sets, ongoing research and development are very necessary for enhancing the precision and resiliency of deepfake detection methods.

6.1 Deep Learning for Deepfake Detection

1. Deep learning has been widely used for deepfake detection because of its ability to extract complex patterns and features from enormous amounts of data. Here is a summary of how deep learning can be used for fake news identification [9, 10]
2. Collection Gathering: To train the deep learning model, a wide and representative collection of actual and deepfake movies and images must be gathered. To ensure the model's robustness, this dataset should cover a variety of deepfake approaches and scenarios.
3. Preprocessing: To extract pertinent characteristics and get the collected dataset ready for training, the dataset is preprocessed. The dataset's size and variability may be increased using techniques including scaling, normalisation, and augmentation.
4. Model Architecture: Convolutional neural networks (CNNs), recurrent neural networks (RNNs), and its derivatives such as ResNet, Inception, and LSTM can all be employed for deepfake detection. The spatial and temporal information in these designs can be extracted from input data. Training: The deep learning model is trained using the preprocessed dataset. The model learns to differentiate between real and deepfake data using the retrieved characteristics after being fed both types of samples. Using methods like backpropagation and gradient descent, the model's parameters are optimised during the training phase [11, 12].
5. Validation and Testing: The model is tested after training in order to assess its performance and adjust any hyperparameters. Using samples taken from deepfake and real data, the model is then assessed for generality and accuracy. Postprocessing: Postprocessing methods can be used to improve the outcomes of the deepfake detection. To find unusual patterns or discrepancies in a movie, for instance, anomaly detection methods or frame consistency analysis might be utilised.

It's important to note that deepfake techniques are developing quickly, and the detection systems must stay up. To increase the resiliency and efficacy of deepfake detection models, ongoing research and development are required. Deepfake detection can also be advanced by working with specialists in adjacent disciplines including computer vision, signal processing, and multimedia forensics.

Table 1 Accuracy for deep learning algorithms

Algorithm	Accuracy (%)
CNN	93.3
RNN	91.9
LSTM	89.3

When deep learning is applied to these fields, the outcomes are cutting-edge when compared with conventional machine learning techniques. Deep learning has shown promising outcomes in the detection of deepfakes, as well. The literature has proposed a number of deep learning algorithms, such as the convolutional neural network (CNN), recurrent neural network (RNN), and long short-term memory (LSTM). These are only few examples: (LSTM). The accuracy of the information contained in the aforementioned algorithms as reported in the literature is given in Table 1.

Additionally, it has been presenting cybersecurity authorities with new difficulties and dangers. Deepfake technology is constantly changing and improving, so it's important to be on guard and knowledgeable. For deepfake detection, a variety of methods and techniques are available. More potent and effective detecting methods must also be introduced, because deepfake algorithms are evolving with time. The general public needs to get better at examining, testing, and evaluating their judgement of the data they encounter on a daily basis. The comparative study of the aforementioned algorithms is shown in Fig. 2. As seen in the graphic, deep learning algorithms have the potential to significantly increase accuracy when compared with machine learning algorithms. Additionally, combining the improvement in feature selection with the addition of more features will increase accuracy.

A novel method has been created as part of the study to disclose AI-generated deepfake video along with effective feature extraction and classification using a customised CNN. The proposed customised CNN performs better than two existing approaches to increase testing accuracy when compared with the current model (Table 2).

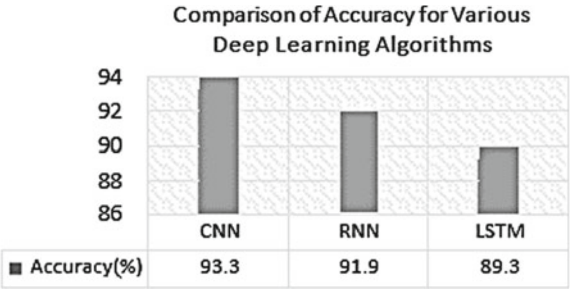

Fig. 2 Comparative analysis of various deep learning algorithms for deepfake detection

Table 2 Comparative analysis of various models with its accuracy

Models	Accuracy
EfficientNetB7	86.98
EfficientNetB1+LSTM	86.02
ENSEMBLE	85.65
C-LSTM Xception	85.65
Xception_DFDC	87.45
DFDC_Rank90_CelebDF	80.32

7 Conclusion

Since amateurs can now easily access deepfake production technologies and produce material quickly, deepfakes have gained significant attention in recent years. This fake digital content can spread swiftly on the sizable platform of social media.

At the moment, it has emerged as the most significant and preferred technique of hackers and fraudsters for obtaining personal data from identity frauds. The deepfake creation algorithms are clever enough to make their own decisions. Although there are not many useful applications for this technology, due to the prevalence of fraudulent digital content, especially in the entertainment and artistic industries, this has been posing severe challenges to our society.

In order to acquire the most accurate results possible for this study, a novel strategy must be created to construct AI-generated deepfake video together with potent feature extraction and classification using the customised CNN.

References

1. Nguyen TT, Cuong M, Nguyen DT, Nguyen DT, Havandi SN (2020) Deep learning for deepfakes creation and detection: a survey. arXiv:1909.11573v2
2. Hrisha Y, Akshit K, Prakruti J (2020) A brief study on deepfakes. In Re J Eng Tech (IRJET)
3. Zhang T, Deng L, Zhang L, Dang X (2020) Deep learning in face synthesis: a survey on deepfakes. In: 2020 IEEE 3rd international conference on computer and communication engineering technology
4. Pan D, Sun L, Wang R, Zhang X, Sinnott RO (2020) Deepfake detection through deep learning. In: IEEE/ACM international conference on big data computing, applications and technologies (BDCAT)
5. Marra F, Gragnaniello D, Cozzolino D, Verdoliva L (2018) Detection of GAN-generated Fake Images over Social Networks. In: IEEE conference on multimedia information processing and retrieval
6. Ivanov NS, Arzhskov AV, Ivanenko VG (2020) Combining deep learning and super-resolution algorithms for deep fake detection. 978-1-7281-5761-0/20/$31.00 ©2020 IEEE
7. Younus MA, Hasan TM (2020) Abbreviated view of deepfake videos detection techniques, international engineering conference. Sustainable Technology and Development
8. Nasar BF, Elizabeth ST, Lason R (2020) Deepfake detection in media files-audios, images and videos. IEEE recent advances in intelligent computational systems (RAICS)

9. Khodabakhsh A, Busch C (2021). A generalizable deepfake detector based on neural conditional distribution modelling. IEEE Xplore
10. Zhu K, Wu B (2020) Deepfake detection with clustering-based embedding regularization. IEEE fifth international conference on data science in cyberspace
11. Siwei L (2021) Deepfake detection: current challenges and next steps, 978-1-7281-1485-9/20/ $31.00c 2020 IEEE
12. Kosarkar U, Patrikar D, Chaube A (2023) Comprehensive Study on image forgery techniques using deep learning. In: 2023 11th international conference on emerging trends in engineering and technology—signal and information processing (ICETET–SIP), Nagpur, India, pp 1–5. https://doi.org/10.1109/ICETET-SIP58143.2023.10151540.
13. Hsu C-C, Zhuang Y-X, Lee C-Y (2020) Deep fake image detection based on pairwise learning. Appl Sci. https://doi.org/10.3390/app10010370
14. Malolan B, Parekh A, Kazi F (2020) Explainable deep-fake detection using visual interpretability methods. In: 3rd international conference on information and computer technologies (ICICT)
15. Montserrat DM, Hao H, Yarlagadda SK, Baireddy S, Horvath RSJ, Bartusiak E, Yang J, Uera DG, Zhu F, Edward
16. Delp J (2020) Deepfakes detection with automatic face weighting, conference on computer vision and pattern recognition workshops (CVPRW)
17. Shohel Rana Md., Sung AH (2020) DeepfakeStack: a deep ensemble-based learning technique for deepfake detection. In: International conference on cyber security and cloud computing (CSCloud)/2020 6th IEEE international conference on edge computing and scalable cloud (EdgeCom)
18. Guarnera L, Giudice O, Battiato S (2020) Fighting deepfake by exposing the convolutional traces on images. IEEE Access. https://doi.org/10.1109/ACCESS.2020.3023037
19. Gong D, Jaya Kumar Y, Sing Goh O, Ye Z, Chi W (2021) DeepfakeNet, an efficient deepfake detection method. (IJACSA) Int J Adv Comp Sci Appl
20. Y Wang (2020) A mathematical introduction to generate adversarial NETS(GAN). arXiv:2009.00169v1
21. Wubet WM (2020) The deepfake challenges and deepfake video detection. Int J Innovat Tech Expl Eng (IJITEE)
22. Fernando T, Fookes C, Denman S, Sridharan S (2021) Detection of fake and fradulent faces via neural memory networks. IEEE Trans Inf Forens Secur
23. Malik A, Kuribayashi M, Abdullahi SM, Neyaz Khan A (2022). DeepFake detection for human face images and videos: a Survey. IEEE
24. Banu Priya M, Daniel JF (2022) First order motion model for image animation and deep fake detection. In: International conference on computer communication and informatics (ICCCI)
25. Kosarkar U, Sakarkar G, Gedam S (2023) Revealing and classification of deepfakes video's images using a customize convolution neural network model. Proced Comput Sci 218:2636–2652. https://doi.org/10.1016/j.procs.2023.01.237