iJRASET
International Journal For Research in Applied Science and Engineering Technology

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 14 **Issue:** II **Month of publication:** February 2026

DOI: https://doi.org/10.22214/ijraset.2026.77607

www.ijraset.com

Call: 08813907089 | E-mail ID: ijraset@gmail.com

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 14 Issue II Feb 2026- Available at www.ijraset.com

Deepdive into Deepfaked Detection Technology

Prof. Poonam Dharpawar, Prerna Pal, Shivanshi Gupta, Smruti Choudhary
Data Science Department,Usha Mittal Institute of Technology, Shreemati Nathibai DamodarThackersey Women's University, Mumbai

Abstract:Concernsregardingdisinformation,identitytheft,anddwindlingtrustindigitalmaterialhavebeenraisedbytheriseindeepfakevideoproduction broughtonbythequickdevelopmentofmachinelearningandartificial intelligence.With the help of this project, users may upload movies and getimmediateauthenticityanalysisusinganintuitivewebinterface.The system uses sophisticated algorithms to categorize videos as "real" or "fake," guaranteeing usability and accessibility. It seeks to inform consumersabouttheconsequencesofdeepfaketechnologyinadditionto detection. The results highlight the need for strong detection systems in the digital age to protect privacy and maintain data integrity.

Keywords: Deepfake, Video Detection, Artificial Intelligence, User Interface, Mis information.

I. INTRODUCTION

Deepfake technology creates incredibly lifelike synthetic audio- visual content by applying sophisticated deep learning algorithms. The main component of this technology is GenerativeAdversarial Networks (GANs), a kind of neural network architecture in which two models—the discriminator and the generator—compete with oneanothertogenerateoutputsthataremoreandmoreconvincing. While the discriminator tries to distinguish between authentic and artificially generated media, the generator produces synthetic content.Thisrepeatedtechniqueeventuallyproducesdeepfakesthat are almost identical to authentic recordings.

Deepfaketechnologyhasvalidusesinindustrieslikevirtualreality, entertainment,andfilmmaking,butwhenused improperly,itposes serious threats to information security, privacy, and public confidence.Deepfakescanbeusedbymaliciousactorstopropagate political individuals or corporations through identity fraud and financial scams. Moreover, deepfakes have been used to fabricate speeches,impersonatepublicfigures,andcreatemisleadingcontent that can damage reputations or influence elections.

Detecting deepfakes is an ongoing challenge, as generative modelscontinuetoevolve,producingincreasinglysophisticated results that blur the line between reality and fabrication. Conventional detection techniques include frame-by-frame analysistoidentifyvisualinconsistencies,examinationof temporaldiscrepanciesinfacialexpressionsandspeech synchronization, and deep learning-based classifiers that assess patterns within the generated content. However, as deepfake algorithms improve, detection methods must also advance to keep pace with new threats.

This paper provides an in-depth exploration of the growing challengesassociated withdeepfakedetection,reviewsexisting techniques used to identify synthetic media, and highlights the urgent need for more robust and adaptive solutions. By understanding the capabilities and limitations of current detection frameworks, researchers and policymakers can work toward developing effective countermeasures to mitigate the risks posed by deepfake technology.

II. LITERATUREREVIEW

Deepfakedetectionhasbeenanactiveareaofresearch,with various approaches proposed to identify manipulated media. Existing methods primarily rely on deep learning, computer vision techniques, and forensic analysis to detect inconsistencies in deep fake content.

With the help of Multi-Task Cascaded Convolutional Neural Networks (MTCNN) for facial recognition, VGG19 and EfficientNet for feature extraction, and Capsule Networks (CapsuleNet and ArCapsNet) for enhanced classification accuracy, the Golden Ratio Based Deep FakeVideo Detection System is able to detect even the smallest manipulations [1].

The AUFF-NET framework integrates Bi-directional Long Short-TermMemory(Bi-LSTM)networksfortemporalpattern analysis with Inception-SwishResNet-v2 for spatial feature extraction. The detection of FaceSwap and face-reenactment deep fakes is improved by this combination, which achieves excellent accuracy in differentiating between modified and authentic video [2].

TheEffectiveMethodforIdentifyingDeepfakeVideosToenhance detection against cyberattacks such compression, noise, blurring, translation, and rotation alterations, the Robust Deep Learning approach makes use of ResNet-Swish and Bi-LSTM. Bi-LSTM improves classification resilience against adversarial manipulations by capturing temporal inconsistencies, whereas ResNet-Swish improves feature propagation.3].

The Dual Descriptor with Frequency Domain Reconstruction Learning approach combines spatial and frequency domain analysis using EfficientNet-NS-B3 and Generative Adversarial Networks (GANs). This method enhances deepfake detection by reconstructing manipulated regions in the frequency domain, improving classification performance [4].
TheGraphNeuralNetworks(GNN)-basedApproachenhances deepfakedetectionbytreatingimagesasgraphs,improvingmodel generalization across datasets. It leverages t-SNE for feature visualization, SCM for correlation detection, SAM for feature refinement, and ANDM for better classification. This method improves interpretability and effectively detects manipulated regions [5].

III. IMPLEMENTATIONOF THE PROPOSEDDEEPFAKEDETECTION SYSTEM

Our proposed deepfake detection system follows a multi-step approach to analyze and classify video content.The methodology is divided into the following stages:

1) Data Collection and Preprocessing

Video files are first preprocessed by extracting individual frames for analysis.

2) Feature Extraction

Featuresareextractedfromeach frameusing an advanced model:InceptionV3forextractinghigh-levelfeaturesand efficient feature representation. This model helps capture the essential characteristics of the facial structure and movements.

3) Classification Using NeuralNetworks

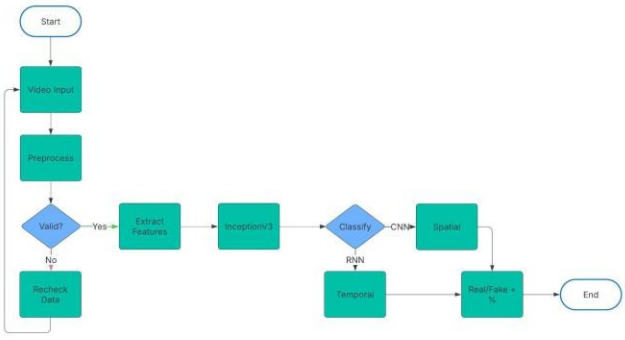
The system uses neuralnetworks(CNNandRNN)toclassifyrealvs.fake content.CapsuleNetworksareemployedbecausetheycan analyze spatial and temporal inconsistencies betweenfacialfeatures,whichiscrucialforidentifyingsubtle manipulations in deepfake videos.

4) Temporal Analysis

Temporal patterns are analyzed by evaluating facial movements across frames, including blinking, lip-sync, and head movements. This is achieved through **RNN**, which processes sequential data to detect inconsistencies in facial dynamics.

5) Final Classification

The system classifies each video as either "real" or "fake" based on spatial and temporal features. Fusing the spatial and temporal analysis results enhances the overall detection accuracy.
This multi-faceted approach, combining spatial feature extraction and temporal analysis, improves the system's robustness against deepfake manipulations.



Flowchart

IV. PERFORMANCE EVALUATION AND EXPERIMENTAL ANALYSIS

The deepfake detection project achieved high accuracy, demonstrating strong performance in identifying manipulated videos. It excelled in detecting lower-quality deepfakes, with promising precision, recall, and F1 scores. However, several challenges were observed that highlight areas for improvement. Some authentic videos were misclassified as deepfakes (false positives), while certain deepfakes went undetected (false negatives),reducingreliability.Themodelstruggledagainsthigh- resolution deepfakes that exhibited natural facial expressions, realistic lip-syncing, and seamless blending, making detection more difficult.

While effective on known deepfake styles, the model failed to generalize well across newer manipulation techniques and cross-modal deepfakes (audio+video synthesis).

Current deepfake detection methods require high computational power, limiting their feasibility for live streaming platforms and real-time content monitoring.Additionally, attackers employ subtle perturbations,adversarialnoise,andblendingtechniquestobypass detection models, making deepfakes increasingly harder to identify.

Toimprovedetection,futureadvancementsshouldfocusonhybrid approachesthatcombineaudio-visualcues,biologicalsignals,and motioninconsistencies.Transferlearningandadaptivemodelscan enhance generalization by fine-tuning on diverse datasets and enabling continuous learning from new deepfake variations. Explainable AI techniques, such as attention maps and feature visualization, can improve transparency and trust in deepfake classification.

Developing lightweight AI models capable of efficiently processing deepfakes in real-time applications, such as social media monitoring, is essential. Additionally, strengthening detection models through robust adversarial training and blockchain-based media verification can ensure content authenticity.

The project demonstrates the potential of AI in combating misinformation,butcontinuousadvancementsareessentialto address evolving deepfake techniques. Future research should focusonenhancingmodelrobustness,explainability,andreal-time detection capabilities to ensure a more secure digital landscape.

Fig.1Representstheinitialinterfaceofthesite

- Severalvideoshavebeenanalyzed,andtheirclassification results have been displayed.
- The system allows users to upload and analyze deep fake videos.

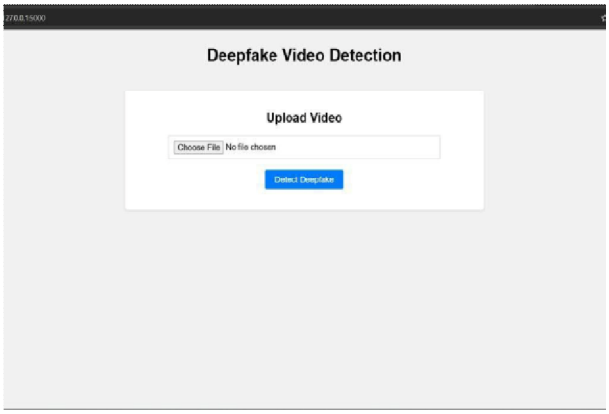


Fig.1Interface

Fig.2Previewoftheuploadedvideoanditsresult

- Asamplevideonamed"Real6.mp4"wasanalyzed.
- Themodelpredicted55.34%REALand44.66%FAKE, indicating uncertainty in classification.

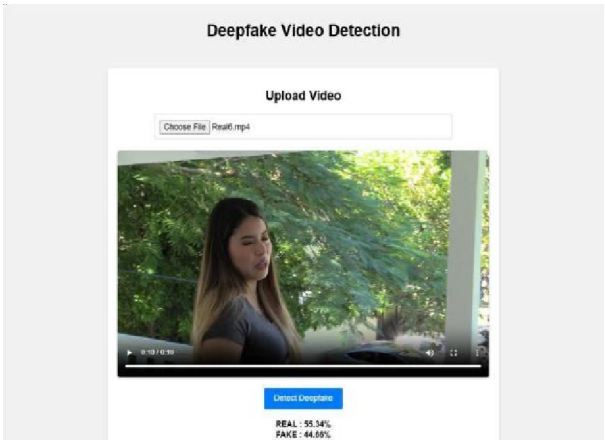


Fig.2Result

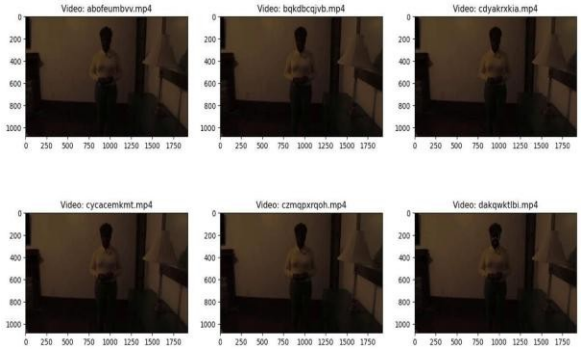


Fig.3Breakdownofasamplefakevideointoframes

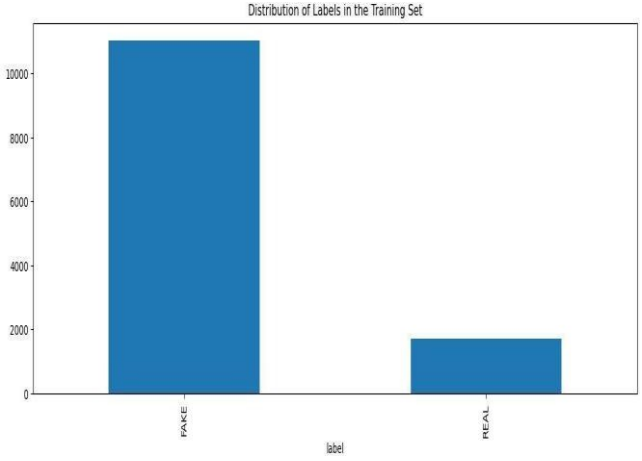


Fig.4 Distribution of labels in training set

- Thedatasetcontainsa distribution oflabels(REAL/FAKE)acrossdifferentvideos.
- The number of samples for each class appears to be well- balanced.

Fig.4.1Usedtraining set

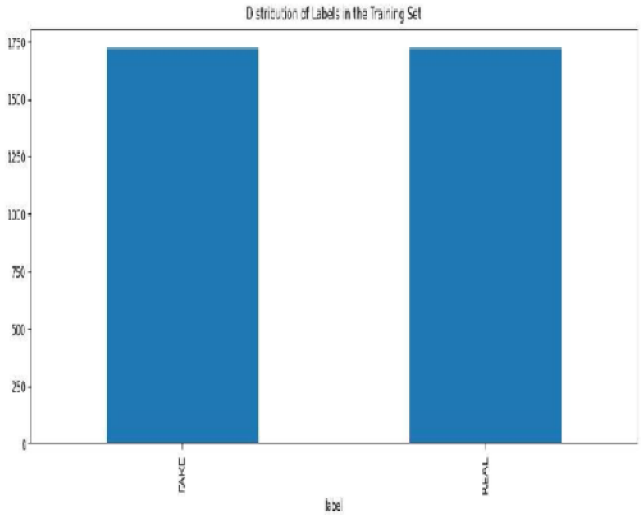


Fig.4.2Originaltrainingset

Fig.5 Model accuracy and loss

- The training and validation accuracy are plotted, showing values around 0.810 to 0.780.
- The training and validation loss are also plotted, with values decreasing over epochs, indicating learning progress.

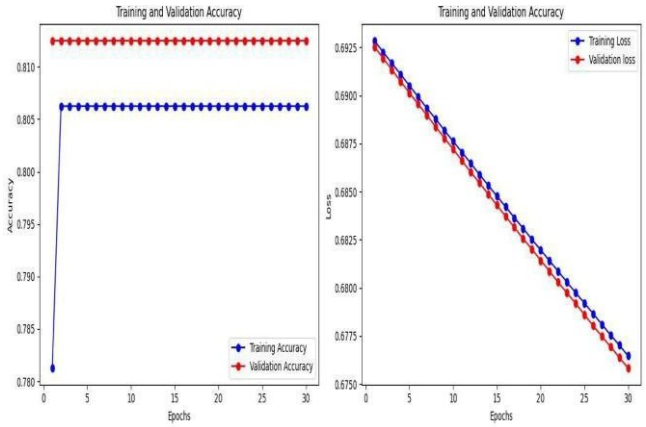


Fig.5Modelaccuracyandloss

V. CONCLUSIONS AND FUTURE RESEARCH DIRECTIONS

The deepfake video detection model demonstrates promising accuracy, achieving around 81% training accuracy and 78% validation accuracy, though some uncertainty remains in classification. While it effectively differentiates real and fake videos, occasional misclassifications highlight the need for improvements in model architecture and dataset diversity. Future enhancements,suchasadvanceddeeplearningtechniques,better data augmentation, and real-world testing, can further refine its performance. Overall, the project establishes a strong foundation fordeepfake detection,with the potentialfor increased reliability through further optimization.

VI. ACKNOWLEDGMENT

We thank Prof. Poonam Dharpawarfor her guidance throughout this proposedwork.SpecialthankstoourPrincipalDr.Yogesh P. Nerkar, Head of Department Dr. Mohan Bonde, and the teaching staff for their support.

REFERENCES

[1] Samet Dincer, GuzinUlutas, Beste Ustubioglu, Gul Tahaoglu, NicolasSklavos,"Goldenratiobaseddeepfakevideodetection system with a fusion of capsule networks", httpsvolume117,ISSN:://doi.org/10.1016/j.compeleceng.2024.109 234 0045-7906,

[2] ShraddhaSuratkar,FarukKazi,"DeepFakeVideoDetection Using Transfer Learning https Approach", Arab JSci://doi.org/10.1007/s13369-022-07321-3Eng48,9727–9737(2023).

[3] XinJin,NanWu,QianJiang,YuruKou,HanxianDuan,Puming Wang, Shaowen Yao, "A dual descriptor combined with frequency domain reconstruction learning for face forgery detection in deepfake videos", Forensic Science International:DigitalInvestigation49(2024)301747,https://doi.org/10.1016/j.fsidi.2024.30174717

[4] AbdulQadir,RabbiaMahum,MohammedA.El-Meligy,AdhamE. Ragab, Abdulmalik AlSalman, Muhammad Awais, "An efficient deepfake video detection using robust deep learning", Heliyon,Volume10,Issue5,e25757,March15,2024,https://doi.org/10.1016/j.heliyon.2024.e25757

[5] H. She, Y. Hu, B. Liu, J. Li, and C.-T. Li, "Using graph neural networks to improvegeneralization capability of the models for deepfake detection," IEEE, vol. 19, ISSN:1556-6021, 2024, https://ieeexplore.ieee.org/abstract/document/10654318/

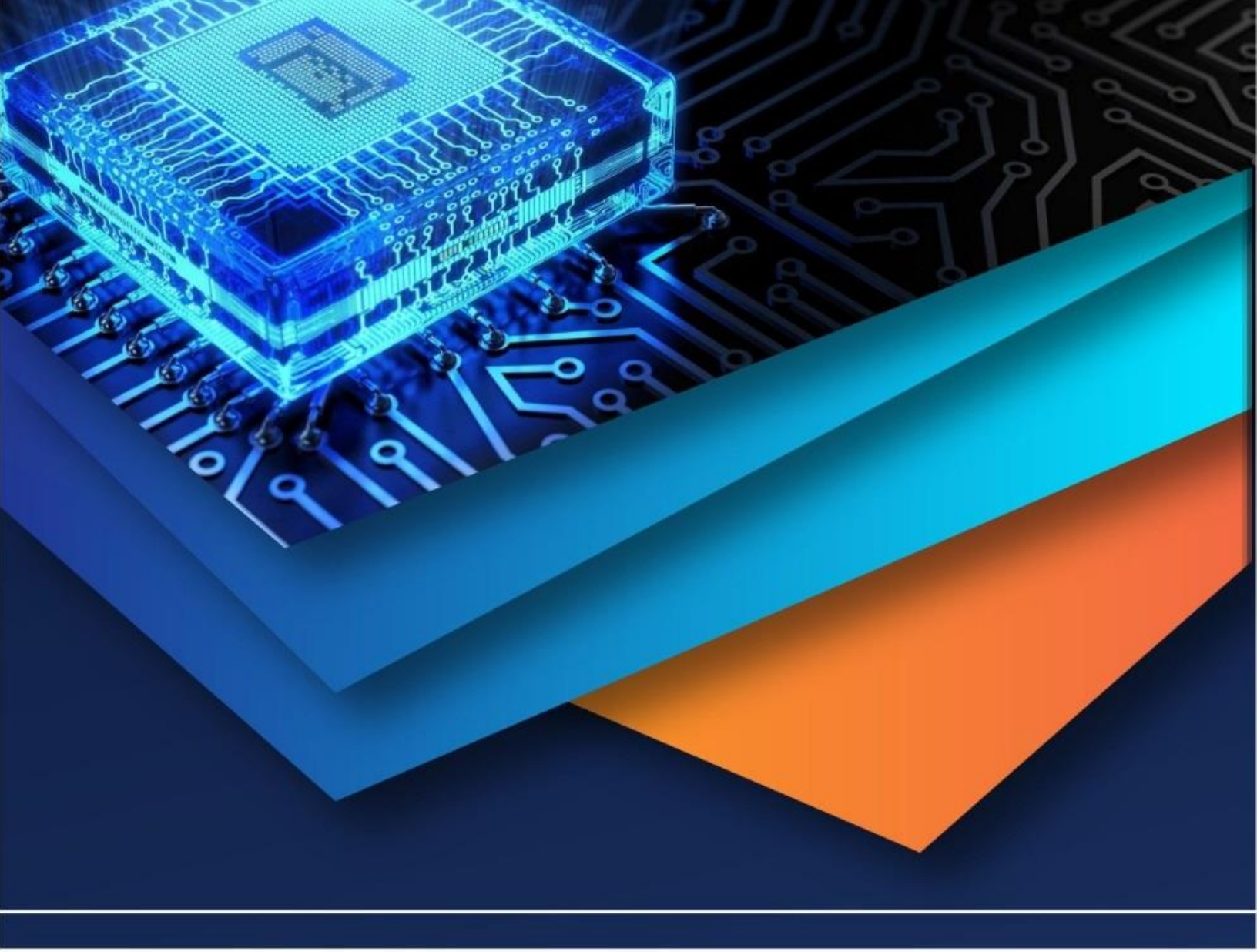
[6] A. Qadir, R. Mahum, M. A. El-Meligy, A. E. Ragab, A. AlSalman, and M. Awais,"An efficient deepfake video detection using robust deep learning," Heliyon, vol. 10,no. 5, 2024,https://www.cell.com/heliyon/fulltext/S2405-8440(24)01788-2

[7] L. Ma, P. Yang, Y. Xu, Z. Yang, P. Li, and H. Huang, "Deep learning technology forface forgery detection: A survey," Neurocomputing,vol.618,2025, https://www.sciencedirect.com/science/article/pii/S0925231224018265

[8] Thanh Thi Nguyen, Quoc Viet Hung Nguyen, Dung Tien Nguyen, Duc Thanh Nguyen, Thien Huynh-The, Saeid Nahavandi, Thanh Tam Nguyen, Quoc-Viet Pham, Cuong M. Nguyen, "Deep learning for deep fakes creation and detection: Asurvey", https://www.sciencedirect.com/science/article/pii/S1077314222001114

[9] A. Golda, K. Mekonen, A. Pandey, A. Singh, V. Hassija, V. Chamola, and B. Sikdar,"Privacy and security concerns in generative AI: A comprehensive survey," IEEEAccess, vol. 12, 2024, https://ieeexplore.ieee.org/abstract/document/10478883/

[10] T. T. Nguyen, Q. V. H. Nguyen, D. T. Nguyen, T. Huynh-The, S.Nahavandi, Q.-V. Pham, and C. M. Nguyen, "Deep learning for deepfakes creationand detection: A survey," Computer Vision and Image Understanding, vol. 223, 2022, https://www.sciencedirect.com/science/article/pii/S1077314222001114



10.22214/IJRASET

45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY