IJARSCT

ISSN (Online) 2581-9429

International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)

International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal

Impact Factor: 7.53

Volume 4, Issue 1, April 2024

Survey on Deep Fake Detection using Deep Learning

Dr. Archana B[1], Arjun K N[2], Dhamini J[3], Ghanalakshmi[4], Swasthishree N S[5]

Associate Professor, Department of Computer Science and Engineering[1]

Student, Department of Computer Science and Engineering[2,3,4,5]

Vidya Vikas Institute of Engineering and Technology, Mysuru, Karnataka, India

Abstract: *The rise of deep fake technology poses significant challenges to the authenticity and integrity of visual content on digital platforms. This paper presents the development of a web application aimed at detecting deep fake images and videos through the implementation of advanced deep learning models. Leveraging research in the field of deep fake detection, this application integrates state-of-the-art algorithm like CNNs for accurate classification tasks. Key features of the application include a user-friendly interface allowing for the upload and analysis of images and videos, leveraging the trained models to detect potential deep fakes. Additionally, the web application incorporates a sophisticated detection methods to enhance accuracy and reliability. The deployment of the web application on popular platforms aims to provide users with a tool to verify the authenticity of visual content, thereby mitigating the potential negative consequences of deep fake manipulation. Continuous monitoring, updates, and adaptation to emerging deep fake techniques are prioritized to ensure the web applications effectiveness and relevance in an ever-evolving landscape of digital content manipulation.*

Keywords: Classification, Deep learning, Anomaly Detection, Deep Fake Detection.

I. INTRODUCTION

In today's digital age, the proliferation of deep fake technology presents a formidable challenge to the veracity and trustworthiness of visual content. Deep fakes, driven by sophisticated machine learning algorithms, have evolved to a point where the line between authentic and manipulated media has become increasingly blurred. These advancements raise critical concerns regarding the potential misuse of deep fake technology in various domains, including misinformation, privacy violations, and the erosion of trust in media. This paper addresses the pressing need for robust and accessible tools to combat the proliferation of deep fake content. Specifically, we embark on the development of a web application aimed at empowering users to discern between genuine and manipulated images and videos.

The motivation for this endeavor stems from the escalating risks posed by deep fake technology. Instances of manipulated visual content, ranging from political misinformation to impersonation and privacy breaches, underscore the urgency of proactive measures to identify and mitigate the impact of synthetic media. This paper outlines a comprehensive approach to the creation of a user-friendly mobile application, emphasizing not only the technical intricacies of deep fake detection but also the ethical considerations surrounding its deployment. The development process involves meticulous research into existing methodologies, data collection, model training, and the integration of sophisticated detection mechanisms.

Key components of the mobile application include intuitive user interfaces allowing seamless uploading and analysis of images and videos. The incorporation of state-of-the-art deep learning models is coupled with a novel voting mechanism, aggregating multiple detection methods to enhance the accuracy and reliability of identifying deep fakes. By providing users with an accessible and efficient tool to discern manipulated media from authentic content, this mobile application aims to contribute significantly to countering the negative ramifications of deep fake technology. Moreover, continual updates and adaptation to emerging deep fake techniques will be prioritized to ensure the application's efficacy in an ever-evolving landscape of digital manipulation.

DOI: 10.48175/IJARSCT-16916

II. RELATED WORKS

This paper [1] explores the realm of deep fake video detection using advanced deep learning techniques, addressing the rising concerns regarding the authenticity of digital visual content. It discusses the utilization of multitask cascaded convolutional neural networks (MTCNN) for face detection and landmark identification, subsequently calculating noise patterns to discern between real and manipulated videos. Leveraging the FaceForensics++ dataset, the study achieves high accuracy in detecting deep fake videos, ranging from 70 to 80 percent. The proposed methodology involves stages of face detection, bounding box regression, and landmark localization, utilizing neural networks' capabilities to categorize videos as genuine or altered. The conclusion emphasizes the method's accuracy and potential future enhancements, suggesting the need to detect multiple faces in videos and implement blockchain technology for video authentication, highlighting the ongoing advancements and challenges in combating deep fake videos.

In this study [2], deep fake video detection techniques, emphasizing machine learning models, particularly deep neural networks, and their application in identifying manipulated videos. It delves into the challenges faced in this field, including the constantly evolving tactics of deep fake creators and the lack of standardized datasets for evaluation. The proposed system in the paper uses a hybrid model, leveraging Convolutional Neural Networks (CNNs), Long Short-Term Memory (LSTM) networks, and a transformer module for feature extraction, achieving a promising accuracy of 95.83% in detecting manipulated videos. It discusses various parameters impacting model training and highlights the significance of validation metrics, such as accuracy and loss rate, alongside the confusion matrix, for evaluating model performance. The references cite recent studies and approaches in deep fake detection, encompassing various aspects, from attention-based networks to remote photoplethysmography for identification purposes. Overall, the document offers insights into cutting-edge methods, challenges, and achievements in combating the spread of misinformation through deep fake videos.

This paper [3], uses Generative Adversarial Networks (GANs) and proposing a novel method for their detection through a combination of Convolutional Neural Networks (CNNs) and Recurrent Neural Networks (RNNs) with Long Short Term Memory (LSTM). Authored by Yash Doke, PrajwalitaDongare, Vaibhav Marathe, Mansi Gaikwad, and Mayuri Gaikwad, the study identifies the challenges posed by the increasing realism of DeepFake content and aims to counter their proliferation by developing a system capable of distinguishing between authentic and manipulated videos. They propose a technique focused on identifying artifacts left behind during the creation of DeepFake videos and use ResNext CNNs for feature extraction and RNNs with LSTM for temporal analysis, outlining a multi-step process from dataset preprocessing to model prediction. The work envisions broader applications for identifying and preventing the spread of DeepFake content across various platforms and media-sharing spaces.

This paper [4] explores the integration of deep learning models for image analysis with explainable AI (XAI) methods to enhance transparency in detecting deepfake content. The objectives include classifying real and fake images, demonstrating the accuracy of different Convolutional Neural Network (CNN) algorithms, and verifying and explaining the model using XAI. The proposed system achieves 99.87% accuracy in detecting deepfake images, showcasing its robustness, reliability, and trustworthiness. The use of XAI, specifically the Local Interpretable Model-Agnostic Explanations (LIME) algorithm, contributes to explaining model decisions and ensuring validity and dependability. The novelty lies in the exclusive incorporation of XAI in deepfake image detection.

The paper [5] likely discusses techniques and methodologies employed to identify and combat deep fake content, emphasizing the significance of these efforts in the context of evolving digital media landscapes. this paper covers the detection and the generation of the deepfake, where the generation is used to help people know if the content they want is forged or not, and on the other hand, we could help them generate the deepfake itself. The detection showed its best performance with the CNN, and more accurate results are aimed to be established in the future for a more authentic, precise website. It is also intended to work with different datasets and generalize the dataset more with several augmentation techniques so that the system can detect any data inserted by the user.

This paper [6] explores the dual role of deep learning in both creating and detecting deepfakes, leveraging convolutional neural network (CNN) models. It delves into the broad applications of deep learning, particularly in computer vision and image detection. As deepfakes pose challenges to privacy and security, the paper addresses the need for detection methods. Additionally, the proposed approach suggests using deep learning image enhancement

DOI: 10.48175/IJARSCT-16916

techniques to elevate the quality of generated deepfakes. The study reflects the interdisciplinary impact of deep learning algorithms, emphasizing their role in both the generation and identification of manipulated images.

This paper [7] highlights the pervasive influence of deep learning across various domains such as natural language processing, computer vision, and image processing, with a specific focus on its application in generating deep fake images. Leveraging generative adversarial neural networks (GAN), deep fakes are created, raising concerns about potential threats to the public. The proposed approach introduces the use of Fisherface algorithm with Local Binary Pattern Histogram (FF-LBPH) for face recognition, utilizing dimensionality reduction in face space. Further, the paper incorporates Deep Belief Network (DBN) with Restricted Boltzmann Machines (RBM) as a deep learning classifier for effective deep fake detection.

This research [8] delves into the escalating challenge of detecting audio deepfakes, emphasizing the growing sophistication of synthetic speech generation techniques that intensify the difficulty of distinguishing between fake and real audio. Utilizing the Fake or Real (FoR) dataset, the study explores two distinct approaches: a feature-based method, involving spectral features of audio samples for machine learning classification, and an image-based method, transforming audio samples into melspectrograms processed by deep learning algorithms like Temporal Convolutional Network (TCN) and Spatial Transformer Network (STN). Notably, TCN exhibits superior performance, achieving a remarkable 92 percent test accuracy, surpassing traditional CNN models like VGG16 and XceptionNet.

This research [9] addresses the pervasive threat of deepfake technology, which creates deceitful visual and audio content by replacing a person's face and voice with synthetic media. the study highlights their misuse in cybercrimes, including identity theft, financial fraud, and blackmail. The research introduces a novel Deepfake Predictor (DFP) approach, leveraging a hybrid architecture of VGG16 and convolutional neural networks. Utilizing a deepfake dataset, the proposed DFP demonstrates superior performance with 95% precision and 94% accuracy, surpassing transfer learning techniques and establishing an advanced tool for cybersecurity professionals to combat deepfake-related cybercrimes and protect potential victims from exploitation.

This paper [10] addresses the growing concern around deepfake videos, which are created using advanced machine learning tools, enabling realistic face swaps in videos. The proposed solution introduces a temporal-aware pipeline for automatic deepfake detection. The system employs a convolutional neural network (CNN) for extracting frame-level features, which are then utilized to train a recurrent neural network (RNN). The RNN is designed to classify whether a video has undergone manipulation. The evaluation against a substantial collection of deepfake videos from various platforms demonstrates the system's competitive performance, showcasing effectiveness with a straightforward architecture.

III. METHODOLOGY

The methodology involves assembling a comprehensive dataset comprising both authentic media and deep fake samples. Meticulous preprocessing ensures uniformity in size, format, and quality, facilitating optimal training of the Convolutional Neural Network (CNN) model.

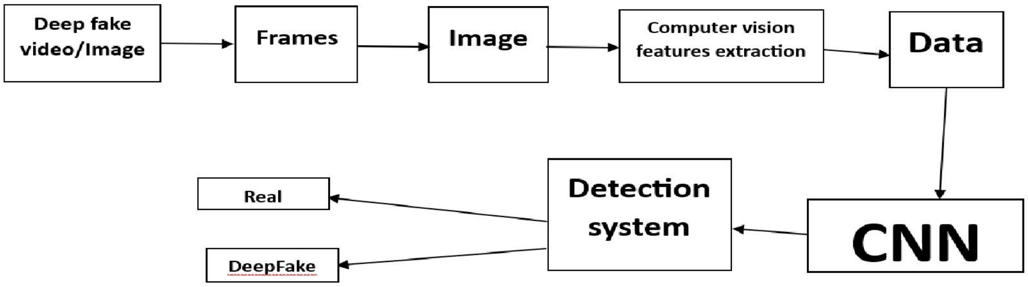

Figure 1.0 System Architecture

The CNN architecture is tailored for deep fake detection, with layers dedicated to feature extraction, dimensionality reduction, and classification. Rigorous training using techniques such as stochastic gradient descent and backpropagation, along with continuous adjustments to model parameters, enhances performance. Separate validation datasets are utilized to evaluate key metrics including accuracy, precision, and recall, thereby validating the efficacy of the trained model. Real-world testing on unseen data provides insights into the practical applicability of the deep fake detection system, crucial for assessing its robustness and reliability. This comprehensive methodology aims to develop a potent system capable of effectively combating the proliferation of deceptive media content.

IV. CONCLUSION

The development of a web application for detecting deep fake images and videos represents a proactive response to the growing challenges posed by the proliferation of deep fake technology. With the increasing prevalence of manipulated visual content across digital platforms, there is a pressing need for robust solutions to safeguard the authenticity and integrity of multimedia content.

By leveraging advanced deep learning models, this paper offers a promising approach to deep fake detection. The application's user-friendly interface facilitates easy upload and analysis of images and videos, empowering users to verify the authenticity of digital content directly from their mobile devices. Furthermore, the incorporation of a sophisticated voting mechanism, which aggregates multiple detection methods, enhances the accuracy and reliability of the application's results.

V. ACKNOWLEDGMENT

We felt great pleasure in submitting this paper on "A Survey on Deep Fake Detection Using Deep Learning ". We are grateful to the editorial board of International Journal of Advanced Research in Computer Science and Technology(IJARSCT) for providing the opportunity to publish our paper.We also thank our faculty members for giving such chance to prove ourselves. We would like to express our sincere thanks to Professor Dr. Archana B for the best support, guidance, and preservation.

REFERENCES

[1] Ankur Nagulwar, Sejal Shingvi, Palak Takhtani. "DEEP FAKE VIDEO DETECTION USING DEEP LEARNING." International Research Journal of Modernization in Engineering Technology and Science (2022): Volume:04/Issue:05

[2] S Jeevidha, S. Saraswathi, Kaushik J B, Preethi K, NallamVenkataramaya. "DEEP FAKE VIDEO DETECTION USING RES- NEXT CNN AND LSTM" International Journal of Creative Research Thoughts (IJCRT), 2023.

[3] Yash Doke, PrajwalitaDongare, Vaibhav Marathe, Mansi Gaikwad, Mayuri Gaikwad. “DEEP FAKE VIDEO DETECTION USING DEEP LEARNING”, International Journal of Research Publication and Reviews, Vol 3, no 11, pp 540-544, November 2022.

[4] Wahidul Hasan Abir, Faria Rahman Khanam, Kazi Nabiul Alam, Myriam Hadjouni , Hela Elmannai , Sami Bourouis , Rajesh Dey and Mohammad Monirujjaman Khan.” DETECTING DEEPFAKE IMAGES USING DEEP LEARNING TECHNIQUES AND EXPLAINABLE AI METHODS”. Intelligent Automation and Soft Computing (IASC), 2023: Vol.35, No.2.

[5] Zeina Ayman, Natalie Sherif, Mariam Mohamed, Mohamed Hazem, Diaa Salama.” DeepFakeDG: A DEEP LEARNING APPROACH FOR DEEP FAKE DETECTION AND GENERATION”. Journal of Computing and Communication Vol.2, No.2, 2023

[6] Khalil, Hady A., and Shady A. Maged. "DEEPFAKES CREATION AND DETECTION USING DEEP LEARNING." 2021 International Mobile, Intelligent, and Ubiquitous Computing Conference (MIUCC)|978-1-6654-1243-8/20/$31.00 ©2021 IEEE | DOI: 10.1109/MIUCC52538.2021.9447642

[7] Suganthi, S. T., Mohamed Uvaze Ahamed Ayoobkhan, Nebojsa Bacanin, K. Venkatachalam, Hubálovský Štěpán, and Trojovský Pavel. "DEEP LEARNING MODEL FOR DEEP FAKE FACE RECOGNITION AND DETECTION." PeerJ Computer Science 8 (2022): e881. DO| 10.7717/peerj-cs.881

[8] Khochare, Janavi, Chaitali Joshi, Bakul Yenarkar, Shraddha Suratkar, and Faruk Kazi. "A DEEP LEARNING FRAMEWORK FOR AUDIO DEEPFAKE DETECTION." Arabian Journal for Science and Engineering (2021): 1-12. Raza, Ali, Kashif Munir, and Mubarak Almutairi.

[9] Ali Raza, Kashif Munir, Mubarak Almutairi "A NOVEL DEEP LEARNING APPROACH FOR DEEPFAKE IMAGE DETECTION." Applied Sciences 12, no. 19 (2022): 9820.

[10] Güera, David, and Edward J. Delp. "DEEPFAKE VIDEO DETECTION USING RECURRENT NEURAL NETWORKS." In 2018 15th IEEE international conference on advanced video and signal based surveillance (AVSS), pp. 1-6. IEEE, 2018