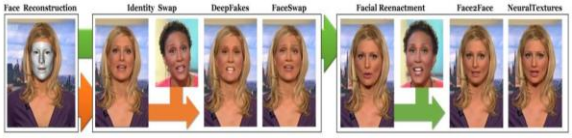
Deepfake Detection Using CNN

Prakash Prajapati[1], Rupesh Kumar Sah[2], Prateek Rajput[3], Priyanshu Proji[4], Ms. Palak Shandil[5]
[1,2,3,4] *Dept. of Computer Science and Engineering, G.L. Bajaj Institute of Technology and Management Greater Noida, India*
[5]*Assistant Professor, G.L. Bajaj Institute of Technology and Management Greater Noida, India*

***Abstract*—The Deepfake Detection Project takes on the important challenge of deepfake video technology, which is a domain of artificial intelligence that generates incredibly realistic, yet fabricated, videos. It utilizes a cutting-edge approach by combining Convolutional Neural Networks (CNNs) to extract spatial features, and Gated Recurrent Units (GRUs) to analyze temporal features. By combining these two powerful deep learning methods, it allows for the detection of small inconsistencies in facial expressions, lighting and movement patterns across video frames that would otherwise not be caught to robustly and accurately identify manipulated media. The system is deployed via a Flask-based backend, and TensorFlow for the model deployment. The frontend, minimalistic and user-friendly, allows interaction with the system. Furthermore, the system can easily run on local computers or cloud infrastructure, affording flexibility and scalability for the use in journalism, security, and digital content verification. Building upon previous systems and addressing their limitations, along with incorporating temporal modelling, the impact of this project provides important advances in the identification of sophisticated deepfakes, hone in the accountability and integrity of digital media.**

I. INTRODUCTION

The technology that alters visual materials has spread across our digital world making manipulation of visual content one of the essential aspects of modern times. The Deepfakes system demonstrates how computer artwork technologies permit face substitution to create fake video material that disparages individuals. Computer vision research about tracking human faces represents the core foundation of these editing approaches since face reconstruction occupies a central position in present-day manipulation methods. Human communications heavily depend on faces because facial expressions enhance messages and they also deliver messages on their own. The field of facial manipulation technology divides into two different approaches for working with expressions and identities of faces. The set all developed Face2Face as one of the leading methods to manipulate facial expressions. The technology allows users to stream facial expressions between individuals using plain hardware equipment during real-time sessions. A subsequent technological development called "Synthesizing Obama" creates facial animations through voice inputs.



The quick advancement of Deep Learning techniques, especially Generative Adversarial Networks (GANs), has enabled the creation of highly realistic synthetic media, commonly referred to as deepfakes. These manipulated images and videos can convincingly alter faces, change expressions, or generate entirely fictitious personas, making it increasingly difficult to discern real content from fake. While deepfake technology holds potential for positive application in fields like entertainment, gaming, and virtual reality, it also poses significant risks, especially when used maliciously. The project began to tackle deepfakes since there has been growing concern on the ethics, laws, and society regarding the technology. Deepfakes are now tools used to spread false information with damage to others' reputation and misleading individuals. There was a surge of deepfakes against political leaders, celebrities, and other public personalities that created massive issues about the contribution they make in elections and the undermining of the trust in relationships within society and in the digital community. 1 Public sharing of a deepfake video with fabricated political statements has the ability to contribute to public unrest and economic market downturn and diplomatic strain. This technology has caused tremendous harm to personal

privacy and reputation in its use in private environments like cyberbullying and revenge porn and identity theft. Gross abuse of the deepfake technology calls for immediate research on effective detection methods to prevent misuse of the technology.

II. LITERATURE SURVEY

The emergence of deep learning has led to unprecedented developments in artificial intelligence, such as the development of extremely realistic fake multimedia content in the form of deepfakes. These fake videos, made by methods like Generative Adversarial Networks (GANs), have raised unprecedented ethical, social, and security concerns. Detection of such fake content is a significant field of artificial intelligence and digital forensics research. Literature review for the Deepfake Detection project explores common methodologies, tools, and frameworks that were put forward as solutions to addressing this problem. It discusses the development of the detection methods starting from traditional handcrafted feature-based techniques to their modern deep learning-based counterparts based on convolutional and recurrent neural networks for higher levels of accuracy and precision. By reviewing current literature, previous work, and prior researches, the survey acts as the basis to reach strengths and drawbacks of existing systems and identify gaps that require exploration. Key aspects of the literature survey include:

❖ Techniques for Detection: A comparative analysis of various approaches, such as pixel- level artifact detection, temporal consistency modeling, and frequency-domain analysis, to identify manipulation traces.

❖ Datasets: Exploration of publicly available datasets like Face Forensics++, Celeb-DF, and Deepfake Detection Challenge datasets, which have driven advancements in model training and benchmarking.

❖ Performance Metrics: Review of metrics such as accuracy, precision, recall, and F1- score, used to evaluate the effectiveness of detection models.

❖ Challenges in Detection: Discussion of evolving deepfake generation techniques, adversarial attacks, and the need for generalized models capable of detecting unseen manipulation techniques.

This survey is a critical step towards the identification of the gaps in existing approaches and motivating new solutions for strong and scalable deepfake detection systems. By integrating the knowledge of the existing work, it provides a clear direction in the development of the effective detection system in this project.

Afchar, D., Nozick, V., Yamagishi, J., & Echizen, I. [1] (2018) proposed MesoNet, a compact Convolutional Neural Network (CNN) designed for the detection of deepfake facial videos. The model aims to provide a lightweight architecture that can effectively detect manipulated facial content, especially in low-resolution videos, while maintaining computational efficiency. Although it is highly deployable due to its compactness, the model struggles to achieve high accuracy in more complex, high-quality deepfake videos, where the manipulations are subtler and less noticeable.

Li, Y., & Lyu, S. [2] (2019) proposed a technique that is geared towards detecting certain geometric distortions introduced during deepfake creation. The technique utilizes Recurrent Neural Networks (RNNs) to inspect temporal anomalies and identify abnormal changes in face features between frames. The technique works well to identify face warping artifacts but fails when manipulations are low or difficult to detect.

Amerini, I., Caldelli, R., & Galteri, L. [3] (2019) examined the use of optical flow estimation to identify deepfake videos. By analyzing the motion patterns across frames of a video, they could identify inconsistencies that arise due to video manipulation. While their method is effective in identifying inconsistencies in motion, it is validated when used to detect high-quality deepfakes because the motion artifacts are minimal in high-quality deepfakes.

Rössler, A., Cozzolino, D., Verdoliva, L., Riess, C., Thies, J., & Nießner, M. [4] (2019) developed FaceForensics++, a comprehensive dataset and technique for detecting manipulated facial videos. Using XceptionNet, a CNN-based architecture, they achieved high accuracy in detecting manipulated videos from benchmark datasets. While their method performs excellently on known deepfakes, it struggles with generalization to unseen or highly sophisticated

deepfake videos, limiting its applicability in real-world scenarios.

Nguyen, H. H., Fang, F., Yamagishi, J., & Echizen, I. [5] (2019) proposed a multitask learning framework that employs autoencoders and Generative Adversarial Networks (GANs) for detecting manipulated facial regions and pinpointing their exact locations at the pixel level. This technique improves accuracy by analysingboth global and localized manipulations. However, it faces challenges when scaling to larger datasets due to the computational demands, making it less efficient for real-time applications.

Dang, H., Liu, F., Stehouwer, J., Liu, X., & Jain, A. K. [6] (2020) introduced a deepfake detection technique based on capsule networks. Their method focuses on detecting spatial anomalies in the structure of faces during the manipulation process. Capsule networks provide significant advantages in capturing fine spatial relationships that conventional CNNs might miss. However, the approach requires high computational resources, which limits its effectiveness in real-time environments or on resource-constrained devices.

Guera, D., & Delp, E. J. [7] (2018) proposed a method leveraging Recurrent Neural Networks (RNNs) in combination with Convolutional Neural Networks (CNNs) to detect deepfakes by analyzing temporal patterns in videos. Their approach captures temporal inconsistencies across frames that often result from face synthesis and blending. Although effective for videos with pronounced temporal artifacts, its performance may drop when temporal manipulations are subtle or deliberately smoothed.

Korshunov, P., & Marcel, S. [8] (2018) conducted an assessment of the vulnerability of face recognition systems to deepfakes and proposed baseline detection methods based on image quality metrics and frame-level inconsistencies. While their work provides an early evaluation of deepfake impacts, the proposed detection methods lack robustness against advanced generative models and are mainly useful as a foundational benchmark rather than a practical detection solution.

Tolosana, R., Vera-Rodriguez, R., Fierrez, J., Morales, A., & Ortega-Garcia, J. [9] (2020) provided an extensive survey of face manipulation techniques, including deepfake generation and detection methods. Their review covers taxonomy, datasets, evaluation metrics, and the evolution of detection strategies. While not proposing a specific detection model, the paper is highly valuable for understanding the broader landscape and comparative strengths and weaknesses of existing approaches.

Verdoliva, L. [10] (2020) offered a comprehensive overview of media forensics with a focus on detecting deepfakes and other synthetic media. The paper examines both traditional and learning-based approaches, emphasizing challenges such as generalizability, robustness, and real-time applicability. This work serves as a critical resource for identifying research gaps and future directions, although it does not introduce a novel detection algorithm.

Mirsky, Y., & Lee, W. [11] (2021) conducted a thorough survey on deepfake creation and detection, classifying various types of manipulations and analyzing detection techniques ranging from classical signal processing to deep learning. The study identifies vulnerabilities in current detectors and offers insights into adversarial attacks and countermeasures. While highly informative, it remains theoretical and doesn't contribute an implementation-ready solution.

Sabir, E., Cheng, J., Jaiswal, A., AbdAlmageed, W., Masi, I., & Natarajan, P. [12] (2019) proposed a recurrent convolutional network framework aimed at identifying manipulated faces in video sequences. By combining spatial and temporal features, the model enhances robustness against sophisticated manipulations. Though it performs well on several benchmark datasets, the architecture demands significant computational resources, which could hinder deployment in lightweight or real-time environments.

Sr. No.	Author Name	Year	Technology	Research Paper Name	Advantages	Limitations
1	Afchar, D. et al.	2018	Convolutional Neural Networks (CNN)	*MesoNet: A Compact Facial Video Forgery Detection Network*	Compact network, easy to deploy	Lower accuracy on low-resolution videos
2	Li, Y. & Lyu, S.	2019	Recurrent Neural Networks (RNN)	*Exposing DeepFake Videos by Detecting Face Warping Artifacts*	Effective for detecting geometric distortions	Limited effectiveness with minimal warping
3	Amerini, I. et al.	2019	Optical Flow Estimation	*DeepFake Video Detection through Optical Flow Consistency*	Good at detecting inconsistencies in motion	Struggles with high-quality deepfakes
4	Rössler, A. et al.	2019	XceptionNet (CNN-based Architecture)	*FaceForensics++: Learning to Detect Manipulated Facial Images*	High accuracy on benchmark datasets	Limited generalization on unseen datasets
5	Nguyen, H. H. et al.	2019	Autoencoders; GANs	*Multitask Learning for Detecting and Segmenting Manipulated Faces at the Pixel Level*	Detects both manipulation and its location	Difficult to scale to large datasets
6	Dang, H. et al.	2020	Capsule Networks	*On the Detection of Digital Face Manipulation*	Robust detection of spatial relationships	High computational cost

III. PROPOSED METHODOLOGY

As manipulated media tools become more advanced, like deepfakes and other forms of face manipulations, it is becoming increasingly difficult to tell the difference between what is real and manipulated images or videos. Thus, protecting the integrity of visual forms of media, is going to be a great challenge moving forward, given that manipulated media can be misused for nefarious purposes like disinformation, identity theft, fraud, and defamation. As the technology for manipulating faces in visual media continues to advance, so must the technology to identify manipulation in faces. FaceForensics++ aims to confront this challenge by providing a standardized dataset and benchmark for researchers developing and testing systems that can detect facial manipulation.

[A] System Analysis

1. Definition: System analysis is the process of thoroughly understanding the requirements and functionality expected from the system. In the case of a deepfake detection system, this involves identifying how the system will analyze video data, detect fake frames, and provide reliable outputs to users.

2. Requirement Gathering: Requirement gathering ensures that the deepfake detection system aligns with the needs of stakeholders such as forensic analysts, video content moderators, or platform owners. Through interviews with experts, surveys for feedback, and observation of current practices, the system's core functionalities—like high accuracy, processing speed, and integration capabilities—are defined. For example, forensic experts may prioritize interpretability in the output, while content moderators

may need high-speed processing for large video batches.

3. Feasibility Study: The feasibility study determines if the system is feasible and worth implementing. Technically, it determines if tools such as CNNs and GRUs can efficiently spot deepfake artifacts and if the infrastructure such as GPUs or cloud resources can efficiently handle training and inference. Financial feasibility determines if the budget can facilitate costs of dataset procurement (e.g., FaceForensics++), software licenses, and hardware. Operational feasibility determines if the system can be integrated into real-world processes, such as video authentication pipelines or regulatory scans.

4. System Modeling: System modeling is the process of creating and understanding system processes using visually constructed models. Data Flow Diagrams (DFDs) depict video data flow through various processes, including preprocessing, feature extraction, and classification. For example, a level 1 DFD might depict frame extraction from a video, CNN processing, flow through a GRU for sequential processing, and summation for classification output. Likewise, Entity-Relationship (ER) diagrams help model data interactions, depicting relationships among video entities, feature sets, and predictions.

[B] System Design

1. Definition: System design converts the findings achieved by system analysis into a comprehensive plan of action. System design gives the functional specifications of the deepfake detection system, focusing on specifics like functionality, hardware, software, and user interface.

2. Logical Design: Logical design is focused on how the system works but does not address what they look like. For the deepfake detection system, this comes down to how the system will process a video that has been uploaded and consists of multiple steps: (1) frame extraction, (2) feature analysis, (3) temporal modeling, and (4) output. Each process will need to outline how it will work (for example, how CNNs extract spatial characteristics such as exception in pixel value, or how the GRU looks to see if there are unnatural transitions from one frame to another). Logical design ensures that all pieces will work together to meet the objectives of the system.

3. Physical Design: Physical design outlines the physical implementation of the system. Hardware requirements could include GPUs, such as NVIDIA A100, for training and inference. Also, hardware requirements could include storage systems required for processing and managing a lot of videos and data sets. Software consists of AI frameworks such as Tensorflow or Pytorch for training different models, and OpenCV for video preprocessing. Databases can be composed with MongoDB, for management of the metadata which informs potential decisions, as well as used to efficiently access the processed data and predictions produced for the data.

4. Interface Design: Interface design helps provide a user interaction level with the system, that is usable. In the case of the deepfake detection project, it encompasses a dashboard for the user to upload videos for analysis, and check results. The results provided could allow visualisations, such as flagged frames with confidence scores, that allow the user to understand how the system reached their conclusions. Furthermore, Application Programming Interfaces (APIs) can allow the integration depending upon the user or developer with external third-party applications, such as social media sites and content management systems. An interface will have to not only represent the functional capabilities of the system, but also usable for the range of intended user groups.

IV. IMPLEMENTATION

[A] Software Requirements

The software requirements for the system include a stable environment provided by Windows 10/11, macOS, or Ubuntu 20.04 LTS for running the Flask backend and pre-trained TensorFlow models. The core programming language used is Python 3.10, which facilitates the development of the application and execution of detection algorithms. Flask 2.3.3 serves as the lightweight web framework for hosting the backend and delivering the web interface. The TensorFlow 2.13.0 library is responsible for executing pre-trained deepfake detection models to perform video analysis. For mathematical operations and

handling numerical data, NumPy 1.23.5 is utilized, ensuring efficient processing of model inputs. Finally, OpenCV 4.8.0.76 (headless) is used for video frame extraction and preprocessing without requiring GUI components, optimizing performance in headless environments.

[B] Hardware Requirements

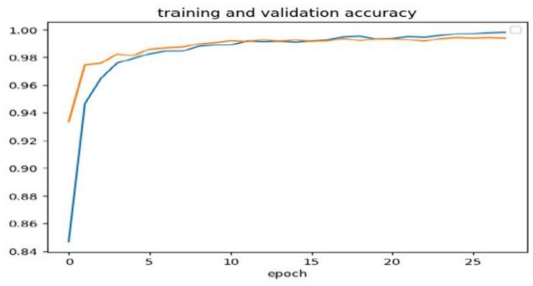
To run the system, the hardware requirements are as follows: a minimum Intel Core i5 (8th Gen) or AMD Ryzen 5 processor which manages backend operations of the server and all pre-processing which includes the Flask application's operations. We recommend a minimum of 8GB of RAM and 16GB RAM is preferred to run TensorFlow models and pre-process video frames smoothly. An SSD is a necessary component as 256GB is recommended as storage is limited to video files, logs, and ephemeral cache created during the pre-processing phase. An NVIDIA GPU enabled by CUDA (e.g., GTX 1650) is recommended but not required to aid in performance of the pre-processing, since it accelerates TensorFlow model inference significantly which will speed deepfake detection.

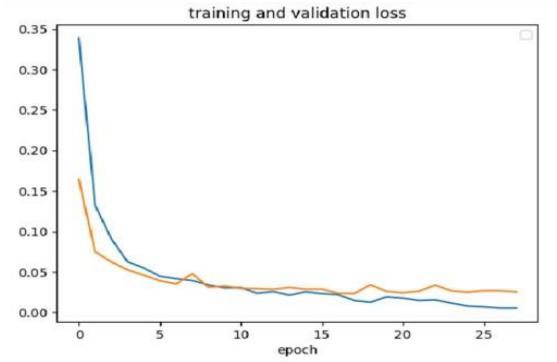
[C] Framework/Library Requirements

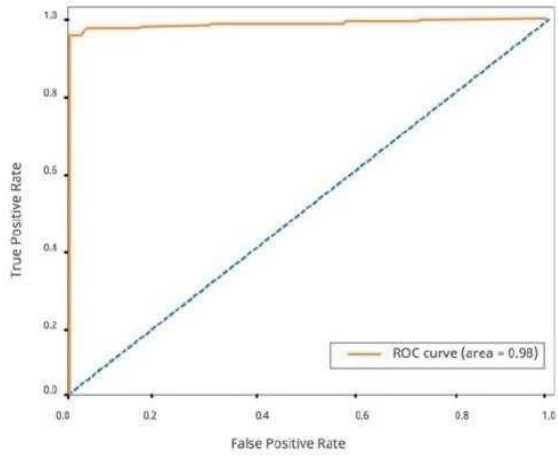
The project is using various industry-standard frameworks and libraries to implement its goals into functionality. The web application is developed using Flask v.2.3.3, which acts as the interface for the backend systems and user's browser. The team also used TensorFlow version 2.13.0, to invoke the deep-learning model and the video frames and deepfake content will be examined within the pre-trained model. NumPy version 1.23.5 is used to accommodate the multi-dimensional array inputs required by the model. And the OpenCV (Headless) computer vision library is used to extract and preprocess video frames without a graphical user interface (GUI). It is important to select frameworks and libraries that are well-established, like Flask, TensorFlow, NumPy, and OpenCV for the sunk costs in functionality, performance, and community engagement.

V. RESULT ANALYSIS

The Deepfake Detection System provides an efficient platform for detecting manipulated videos, while ensuring user convenience in the above conclusions. The outcomes indicate that there was a successful implementation of a deepfake detection system and that it clearly exhibited functioning with the steps shown. The deepfake detection system provided an interactive interface for the user to upload video files for deep fake detection, trigger the detection process, and show the output effectively. Below is an outline of the results. There were five steps which the deepfake detection system demonstrated in terms of completing the processing of the video. Starting from the upload video phase which provided a file selection dialog box to traverse and upload the wanted file. This contributes to a smooth experience, allowing users to test videos files of various sizes and formats (e.g., audio / visual files like MP4). After the file upload, the system entered a phase of previewing and confirming that the uploaded video was indeed the correct file. This allows for the user to visually verify that the correct file has been uploaded and provides transparency before going to the detection phase. In terms of the display, a large "Detect Deepfake" button was displayed to allow the user to trigger the analysis. Once the detection process starts, the system is in the deepfake detection phase and it uses advanced deep learning algorithms to analyze the uploaded video frame-by-frame — this enables it to find inconsistencies or manipulations in the video content. In the end, the system presents its findings in the output presentation phase where it tells the user whether the uploaded video is a deepfake or not. The output is also presented at the same time as a video preview to help in the clarity and overall user experience. In summary, the four steps are as follows: 1) The user uploads the video file using the file selection dialog, 2) The user previews the uploaded video and starts the detection process, 3) The system analyzes the video using a detection algorithm which indicates manipulation, 4) If there is manipulation detected, the system will output to confirm that the video is indeed a deepfake or is authentic.







VI. DISCUSSION

The Deepfake Detection Project is a major advancement within the journey of combatting the misuse of AI-generated media. As deepfake technology continues to emerge, the ability to manipulate video has reached an advanced level of sophistication, and potentially, access to such technology is unlimited. This burgeoning development poses serious risks to privacy, security, and trustworthiness of digital information. The proposed methodology for this project derives from Convolutional Neural Networks (CNN) and Gated Recurrent Units (GRU) for the spatial and temporal analysis of video. This method lays the groundwork to determine if advanced deep learning techniques will be able to support the problem. Combining spatial feature extraction with temporal modelling provides a comprehensive framework for recognizing inconsistencies, often invisible to the human eye. In so far as system developer decisions for deployment, the project's design of a Flask based web application and use of TensorFlow, means machine learning capabilities were harnessed to be flexible and scalable in deployment and simple for the user. Integration effort was made for the user to simply upload their video for analysis from the frontend, and the system will convert and process the video through back-end Python code to identify apparent signs of manipulation with ease. Working as a hybrid model that has separate frontend and backend allows for the greatest flexibility in deployment approaches, either locally or on cloud, providing a unique edge to this solution. When considering deployment approaches, potential practical applications include field journalism, law enforcement, and digital media verification reporting. While the project has merits, there are also limitations. The deepfake detection model may struggle with particularly advanced deepfakes that minimize the visual artifacts and temporal inconsistencies in deepfake videos. Further, processing long-duration and high resolution videos may take a significant amount of computing power, which may limit its use in environments where resources are constrained. These limitations underscore the importance of ongoing research and development of improved deepfake detection approaches as deepfake generation methods improve. In conclusion, the usefulness of the Deepfake Detection Project entails a major contribution to ensuring the integrity and authenticity of digital content. The project tackles the shortcomings of previous methods and develops first-class detectors utilising recent AI approaches thus, contributing to the fight against malicious and potentially harmful usages of deepfake technology. There is potential to increase the effectiveness of the system with improvements in detection algorithms and computational efficiency to help ensure the system remains robust in an evolving technological environment.

VII. FUTURE SCOPE

The Deepfake Detection Project has a lot of potential for future developments to advance the changing landscape of deepfake technology. A major area of improvement is in detection algorithms to become more advanced. As deepfake generation technologies advance and create videos with fewer visual artifacts and temporal misalignments, it is key to have more robust models able to find even the most innocuous

hint of manipulation. Future studies may focus on improvements using modern architectures in detection such as transformers or multimodal models to further increase accuracy. Another avenue is enabling real-time detection capabilities. At present, the system works on video files so in a live scenario (e.g., a football game live stream), the system does not currently work in real-time. Future versions of this application can focus on adding enhancements to allow real-time using cloud processing or on-device processing, detecting the video in live time is widely useful and could allow for real time verification of live news stories in a dynamic environment of social media platforms. Furthermore, considering scaling the model for resources, the model could be optimized for lightweight use on smartphones to allow the broadest diversity of user base across geographic or demographic use cases or contexts. Broadening the focus of the project to include the detection of other types of media manipulation, such as deepfake audio or image-based media, is another key area to investigate. The addition of cross-domain capabilities would create a more robust system for the detection and identification of manipulated content across a wider range of formats. Partnering with both forensic and law enforcement organizations would help enhance the trustworthiness of the tool with organizations that conduct formal investigations. Formal partnerships with judicial and law enforcement organizations would help provide actionable insights regarding potential criminal activity involving deepfake content creators. Lastly, expansion of the databases with more extensive, more disparate datasets for model training and validation could bolster model robustness in the face of new kinds of deepfake technology, therefore preventing the model from being conditioned by new attacks. As well, systems that allowed for the information sharing of experiences would allow multiple detection systems to share learnings and improve simultaneously, increasing detection efficacy. The area of questions of transparency would identify areas for potential improvement and explainability in future models or versions of the tool, in order to evidence how the model made a decision. In addition to enhancing user trust, it would permit a meaningful way for users to provide feedback to guide system improvements. In closing, the future scope of this project is to conceptualize and enhance the technical performance and relevance of the tool while ensuring it remains relevant to an ever-changing world in traction with deepfake technology. As these developments move forward this project will allow those engaged with it to support a degree of authenticity and trust in the world of digital media.

REFERENCES

[1] Afchar, D., Nozick, V., Yamagishi, J., & Echizen, I. (2018). MesoNet: A compact facial video forgery detection network. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), 1-10. https://doi.org/10.1109/CVPRW.2018.00143

[2] Li, Y., & Lyu, S. (2019). Exposing deepfake videos by detecting face warping artifacts. IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), 46-52. https://doi.org/10.1109/CVPRW.2019.00010

[3] Dang, H., Liu, F., Stehouwer, J., Liu, X., & Jain, A. K. (2020). On the detection of digital face manipulation. IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 5781-5790. https://doi.org/10.1109/CVPR42600.2020.01243

[4] Amerini, I., Caldelli, R., Filippini, F., & Becarelli, R. (2019). Deepfake video detection through optical flow consistency. Proceedings of the 2019 IEEE International Workshop on Information Forensics and Security https://doi.org/10.1109/WIFS47025.2019.9035103

[5] (WIFS), 1-6. Rössler, A., Cozzolino, D., Verdoliva, L., Riess, C., Thies, J., & Nießner, M. (2019). FaceForensics++: Learning to detect manipulated facial images. IEEE Transactions on Pattern Analysis and Machine Intelligence, 43(3), 728-746. https://doi.org/10.1109/TPAMI.2019.2949508

[6] Nguyen, H. H., Yamagishi, J., & Echizen, I. (2019). Multitask learning for detecting and segmenting manipulated faces at the pixel level. IEEE 9th International Conference on Biometrics Theory, Applications and Systems (BTAS), 1-9. https://doi.org/10.1109/BTAS.2019.8898116

[7] Guera, D., & Delp, E. J. (2018). Deepfake video detection using recurrent neural networks. 2018

15th IEEE International Conference on Advanced Video and Signal Based Surveillance (AVSS), 1-6. https://doi.org/10.1109/AVSS.2018.8639163

[8] Korshunov, P., & Marcel, S. (2018). Deepfakes: a new threat to face recognition? Assessment and detection. https://arxiv.org/abs/1812.08685 arXiv preprint arXiv:1812.08685.

[9] Tolosana, R., Vera-Rodriguez, R., Fierrez, J., Morales, A., & Ortega-Garcia, J. (2020). Deepfakes and beyond: A survey of face manipulation and fake detection. Information Fusion, 64, 131-148. https://doi.org/10.1016/j.inffus.2020.07.007 Topics

[10] Verdoliva, L. (2020). Media forensics and deepfakes: An overview. IEEE Journal of Selected in Signal https://doi.org/10.1109/JSTSP.2020.3002103 Processing, 14(5), 910-932.

[11] Mirsky, Y., & Lee, W. (2021). The creation and detection of deepfakes: A survey. ACM Computing Surveys (CSUR), 54(1), 1-41. https://doi.org/10.1145/3425780

[12] Sabir, E., Cheng, J., Jaiswal, A., AbdAlmageed, W., Masi, I., & Natarajan, P. (2019). Recurrent convolutional strategies for face manipulation detection in videos. arXiv preprint arXiv:1905.00582. https://arxiv.org/abs/1905.00582