Proceedings of the 6th International Conference on Smart Electronics and Communication (ICOSEC-2025)
DVD Part Number: CFP25V90-DVD; ISBN: 979-8-3315-9858-7

DeepTamperNet Identification: A Hybrid Preprocessing and Feature Extraction Fusion Pipeline for Robust Deepfake Detection

1st Sharon
Research Scholar
Department of Computing Sciences
Vels Institute of Science, Technology & Advanced Studies
Chennai, India
rushaugust21@gmail.com

2nd Dr. N. Shyamala Devi
Assistant Professor
Department of Information Technology
Vels Institute of Science, Technology & Advanced Studies
Chennai, India
shyamadevi@gmail.com

Abstract*—The proliferation of crimes especially with the misuse of digital content embedded with advanced technology to manipulate individuals has largely accelerated in neoteric times. Simulated data has established trust amongst individuals, thereby making them vulnerable to several frauds that manipulate their identity. This manipulation has led to severe loss interms of finances and in also risking several lives. Deepfakes are one such cybercrimes that has increased gargantuanly, thereby necessitating the impeccable technological mechanisms to agnize fakes, tampering and manipulation. Although existing research identifies deepfakes, the approach of static identification makes it impossible to identify the tampered data dynamically while in the process of manipulation. This indagation overcomes the gap evinced in the existing studies by incorporating tampering levels in a dynamic data. The tampered heatmap levels are shown dynamically to comprehend the level of tampering, thereby enabling users to distinguish between manipulated and genuine data. The paper entails a detection pipeline through the processes of pre-processing and feature unsheathing using deep heuristic techniques. The Multi-Task Cascaded Convolutional Neural Network (MTCNN) is used for face detection embedded with different pre-processing techniques. The study analyses the tamper score which accelerates the evaluative quantification of manipulated data. A hybrid deep learning model such as the ResNet18, MobileNetV2 and EfficientNetB0 fused with hybrid heuristic edge algorithm is combined to form the "DeepTamperNet" that aids in stratifying the dynamic data into fake and genuine. This research aids in curbing cybercrimes through efficient tampering cognizance deliverables that can be used as video tampering detectors. The results for this study are successfully procured, and show that the tamper score obtained from different combination of algorithms and pre-processing methods, and enable the user to explicitly identify manipulated data.

Keywords*—Deepfake Identification, Deep Learning, Data Classification, Tamper Score, Multi-Task Cascaded Convolutional Neural Network (MTCNN), Pre-processing, Feature Extraction, Resnet18, MobilenetV2, EfficientnetB0, Heuristic edge.

I. INTRODUCTION (*HEADING 1*)

Deepfakes are one of the growing concerns with many individuals being manipulated, and infringed of their personal details. This form of cybercrime is a superimposed technique that targets the identity of an individual. Deepfakes are contrived by deep learning models embedded with techniques that portray hyper-pragmatic media manipulations, that can intelligently modify the media over call and dynamic networks. This capability of altering facial features [5], audio and the entire identity of an individual poses a big challenge for not just social media brands, but also for financial sectors and other industries whose security measures can be compromised through this cybercrime [13]. In a broader definition, deepfakes are artificial intelligence-synthesized content that can be manipulated [1] in several categories as shown in Fig 1 below:

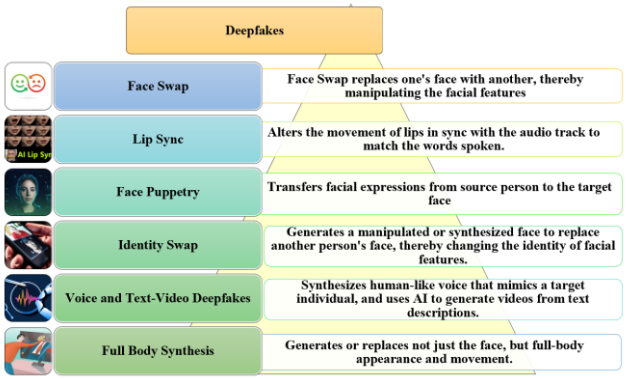

Fig 1. Types of Deepfakes

The above-described methods of deepfakes [4], [14] necessitate extensive training and testing prior to their utilization, and therefore the data used for this aspect of cybercrime is gargantuan. However, the data training can be relatively difficult only for common man, and can be quite easy for popular individuals and publicly acknowledged people. This is because, data and images [15] of public figures are facilely accessible online, thereby making them viable targets for attackers. Deepfake videos although is known widely now, as it emerged initially in 2017 [7], it created trauma and chaos to the affected parties, thereby instating the need for improvised security, and the existence of loopholes. However, in neoteric times this cybercrime manipulation has gained colossal attraction amongst attackers, who weaponize this technique for spreading inaccurate information, identity larceny, fiscal fraud, political influence, and the erosion of public trust in visual evidence [17]. The research gap is the significant challenge faced with deepfakes, which are pushed to an extent where manipulated digital content seems genuine, thereby engaging innocent users to provide information to the attackers. With the social media and digital mode of operation

being the quotidian transaction for individuals and organizations, identifying manipulated or tampered data is a crucial step toward protecting personal dossier. Despite the various approaches of early detection introduced to agnize deepfakes, these methods pivot largely on the pixel level scrutinization rather than the layer level analysis [3]. The obsoleteness of existing techniques to overlook subtle anomalies of this type of cybercrime in the existing approaches necessitates the incorporation of advanced models that efficiently adapts and is robust to real-time scenarios and evolving criminal space.

Inorder to overcome the limitations of existing models of security, this study introduces a modular and interpretable deepfake detection pipeline designed to operate on the first frame of a video, simulating a low-latency, high-precision tampering assessment. The proposed pipeline incorporates multiple components that can be independently evaluated and improved. The variance-based tamper score, quantifying pixel-level irregularities in the extracted feature maps and the layer-wise analysis of the entailed models aid in explicitly distinguishing real from manipulated content. In addition, a lightweight hybrid edge feature extractor is used to emphasize high-frequency facial features—edges, textures, and inconsistencies—that often signal synthetic tampering but may be overlooked by conventional deep networks [18]. This juxtaposed study aims to determine which combination of preprocessing technique and feature backbone yields the most robust detection accuracy, thereby enabling improvised interpretability, to promote real-world usability and media authentication. The visualizations procured from this indagation further throws light on the preprocessing–feature extractor pairs that achieves the zenith of performance, while effectively serving as a diagnostic tool for comprehending the extent of tampering on a digital data.

This paper is structured with section II elaborating the empirical review, section III explaining the proposed methodology, section IV illustrating the obtained results and the final section concluding the word done, along with future work of progress.

II. Literature Review

Hasan Abir et al [6] articulated about deepfakes using explainable AI, and proposed the use of deep learning models to scrutinize the effectiveness of algorithms through Local Interpretable Model-Agnostic Explanations (LIME). The study incorporates the use of Convolutional Neural Networks (CNN), inputting a dataset gathered from Kaggle. The deep learning models [11], [12] such as the InceptionResnetV2, DenseNet201, InceptionV3, and ResNet152V2 were entailed. The accuracy of the models thus procured renders an accuracy of 99.68% for InceptionV3, 99.19% for Resnet, and 99.81% for DenseNet201. However, InceptionResNetV2 achieved the highest accuracy of 99.87%, which were further corroborated using the LIME algorithm for XAI. The pivotal aspect of this study was to stratify and distinguish the real and fake images by using diverse convolutional networks. The Explainable AI (XAI) is further used to delineate the model, inorder to provide better insights. Using the different approaches, while verifying the results using LIME provides clarity interms of model usage for the future. However, this study explains the limitation of images for XAI, further constraining the use of videos which could be a potential future work explicated in the research.

Vrizlynn L. L. Thing [2] in the paper titled “Deepfake Detection with Deep Learning: Convolutional Neural Networks versus Transformers”, delineated the utilization of Convolutional neural network [8] in juxtapose with transformers, along with entailing diverse datasets developed for deepfakes. These datasets comprised of the latest second and third generation deepfake videos. The efficacy of the single model detectors in deepfake detection and cross datasets were analyzed, thereby rendering an accuracy of 88.74%, 99.53%, 97.68%, 99.73% and 92.02% accuracy and 99.95%, 100%, 99.88%, 99.99% and 97.61% AUC, in the detection of FF++ 2020, Google DFD, Celeb-DF, Deeper Forensics and DFDC deepfakes, respectively. Four CNN and four transformers were constructed for this study, and the dataset incorporated was split into 10% for validation, 20% for testing, and 70% for training. Representing the convolutional networks the ResNet152, XceptionNet, EfficientNet B7 and HRNet were built, and VIT, BEiT, Swin and CaiT were constructed to comprehend the working of Transformers. The results clearly explicated the visibly better results for CNN in the train-test datasets, while the transformers surpassed the performance of CNN in the cross-datasets. With reference to the different datasets used for the study, the correlative association between FF++, Google DFD and Celeb-DF datasets were evinced. The future work of study is elucidated to opt for more sophisticated data models, and progressive deep learning methods that can enable deeper scrutinization of the data to capture the uniqueness of deep forensics.

III. Proposed Methodology

This section provides an elaboration into the various techniques used for identifying the tamper score, and further details the evaluative pipeline incorporated for the successful procuring of the results relevant to the preprocessing and feature extraction processes. An illustration of the research architecture is provided in Fig 2:

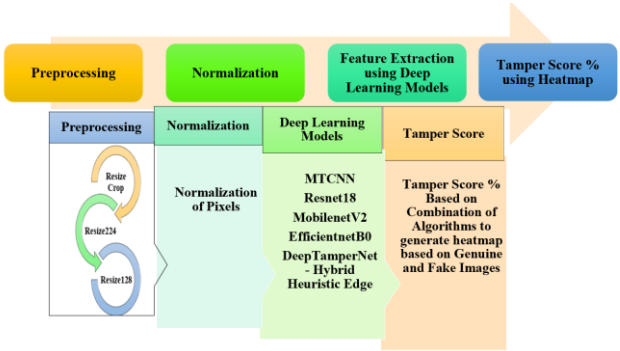

Fig 2. Techniques used for the Research

The dataset used for this study is procured from FaceForensics++. A total of 1000 videos were garnered for processing. The pipeline of pre-processing implemented through various methods for resizing and normalization is implemented for subsequent phases. Since a GPU is not entailed for this study, the utilization of CPU struggles to include all the frames of the entire dataset. Therefore, the first frames of the videos from the entire dataset are extracted inorder to identify the level of tampering. This simulation helps in enabling a lightweight tamper detection mechanism, and further suitable for quotidian usage and real-time applications.

The multi-task cascaded convolutional neural network (MTCNN) is used along with resizing mechanisms to unsheathe the first frames from the videos [9]. The preprocessing entails three different variants for resizing, with each of the resizing methods cropping the image into a certain type of resolution. The 'Resizecrop' method crops the image to a 256 x 256 resolution, and then centre-crops the image further to a 224 x 224 resolution. The 'Resize224' method is a direct cropping of the image to the 224 x 224 resolution with no intermediaries involved, while the 'Resize128' entails the 128 x 128 image resolution involved.

The subsequent process of normalization is adopted from the tensor converted images after resizing in the range of -1 to 1 inorder to explicitly prepare them for the further steps of feature extraction.

The deep learning and the hybrid heuristic model are used as feature extractors in this indagation. The Resnet18 is a residual network that is lightweight, but effectively captures the spatial features from the inputted dataset [10]. The MobileNetV2 is a depth-wise segregated model used for capturing features specifically in mobile devices. The EfficientnetB0 is another depth-width oriented network that is wrapped for compound-scaled resolution. All of the pre-trained networks are wrapped with a relative featureNet adopting the adaptive average pooling mechanism, and a feature map that enable to explicitly identify tampering in the incorporated data. The Hybrid Heuristic DeepNet extractor also called as "DeepTamperNet" combines a heuristic edge detector along with a transfer-learning based EfficientNetB0 network. The layer-wise architecture of the DeepTamperNet is further illustrated in Fig 3:

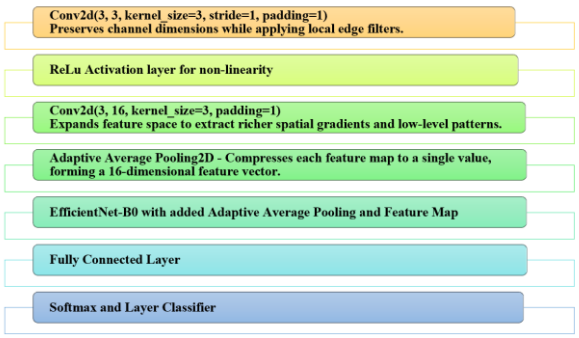

Fig 3. Layer-wise Architecture of Hybrid Heuristic Deep network - DeepTamperNet

The primary difference between the pre-trained networks and the DeepTamperNet is the heuristic component embedded with transfer learning [11] that enables natural learning rather than depending on established semantics. The tampering in a data is fixed to a noise-level estimator mimicking image forensics to extract variance in facial features. The local spatial inconsistencies in the data are extracted, and overlaid on the heatmap to comprehend the level of tampering in the data. The model considers both pixel-level and layer-level irregularities thereby enabling to analyse the tamper score percentage from the feature vector and the feature map [16]. The feature maps through the global pooling method protects the semantic and structural features in the data which are most often tarnished in deepfakes. The tamper score thus generated is further classified into real and fake, with the former having 0% tampering, while the latter having diverse scores based on the level of data tampered with. The heatmap also shows consistency of heat in a real video frame, while inconsistencies can be observed with a tampered data frame, thereby helping in explicitly identifying deepfakes. The results of the pipeline of processes are depicted in the consequent section.

IV. Results

The pipeline of pre-processing and feature extraction for frames from the video are explicitly implemented in Python, and the results of the same are as shown below:

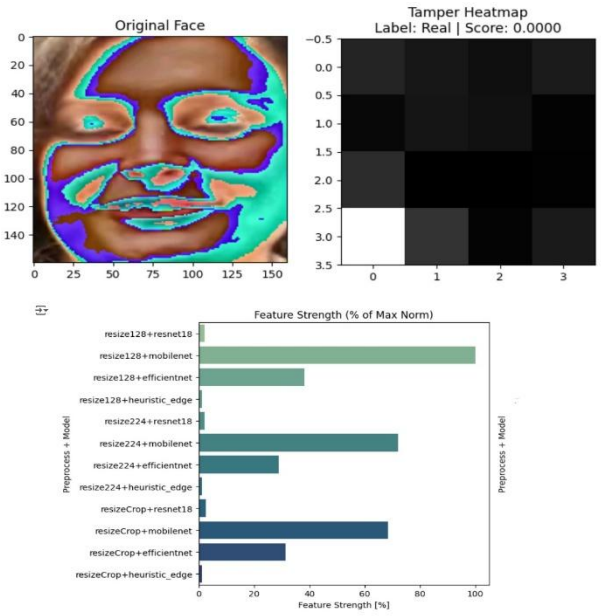

Fig 4. Tamper Score and Feature Strength for Real Data

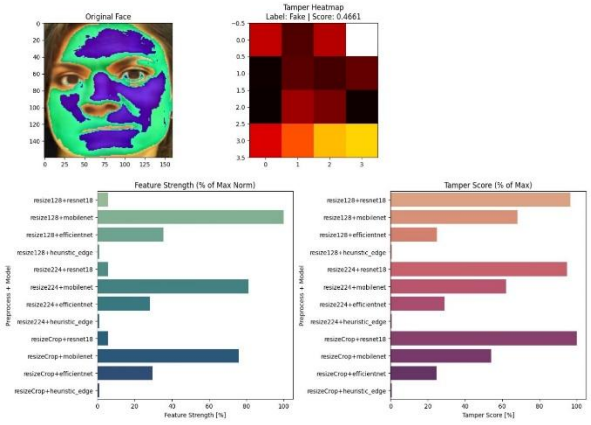

Fig 5. Tamper Score and Feature Strength for Deepfake Data

The Fig 4 and Fig 5 depicts the tamper score and feature strength rendered for real or authentic data, and the tamper-feature strength for deepfake manipulated data respectively.

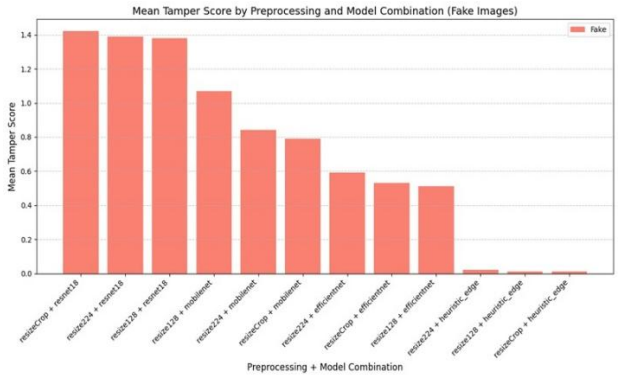

Fig 6. Mean Tamper Score evinced through Combined Preprocessing and Feature Extractor Model Pipeline

Tamper Score Summary Table

	Combination	Model	Resize Method	Mean Tamper Score
0	resizeCrop + resnet18	resnet18	resizeCrop	1.420000
1	resize224 + resnet18	resnet18	resize224	1.390000
2	resize128 + resnet18	resnet18	resize128	1.380000
3	resize128 + mobilenet	mobilenet	resize128	1.070000
4	resize224 + mobilenet	mobilenet	resize224	0.840000
5	resizeCrop + mobilenet	mobilenet	resizeCrop	0.790000
6	resize224 + efficientnet	efficientnet	resize224	0.590000
7	resizeCrop + efficientnet	efficientnet	resizeCrop	0.530000
8	resize128 + efficientnet	efficientnet	resize128	0.510000
9	resize224 + heuristic_edge	heuristic_edge	resize224	0.020000
10	resize128 + heuristic_edge	heuristic_edge	resize128	0.010000
11	resizeCrop + heuristic_edge	heuristic_edge	resizeCrop	0.010000

Fig 7. Tabulated Summary of Mean Tamper Score evinced through Combined Preprocessing and Feature Extractor Model Pipeline

Fig 6 and Fig 7 elucidates explicitly the comparative score and summary of mean tamper score obtained through the processed models of implementation. The results from the above throughputs clearly show that preprocessing significantly accelerates the focus on tampering, while evading distortions and manipulated padding. The 'ResizeCrop' preprocessing combined with the Resnet18 model helps achieve good mean tamper score. Comparing the preprocessing techniques, the Resizecrop method proved to yield better results as compared to Resize224 and Resize128. In terms of feature extractors, Resnet18 while capturing spatial distinctions of the facial features is very sensitive to tampering, thereby enabling to identify deepfakes much efficient than the other models. 'DeepTamperNet' establishes high sensitivity to facial feature extraction and visualization, but is not sufficient to efficiently identify differences in facial alignments, but rather can be established as an explainability metric that aids in elucidating the tampering in heatmaps. As far as feature strength is concerned, the Resize embedded with MobileNetV2 establishes higher score as compared to the other juxtaposed models. The results clearly show that shallower models require more bolstering inorder to effectively detect deepfakes in the layer-level. The heatmaps clearly show the variance-based scoring of the tampering levels, thereby effectively aiding in identifying synthesized data in early video frames and swifter identification.

V. Conclusion

Cybercrimes are one of the concerning aspects that the nation is pivoted toward, and measures to overcome these cybercrimes through advanced technological models are progressing with each day. Amongst, the most concerning is the deepfake which throws subtle to unnoticeable synthesis of digital data, thereby ensuring many innocent users fall prey to this. These increasingly sophisticated crimes seep through many domains of working, and can cause various type of loss. This research addresses the necessitate to proposing modular and dynamic tamper score rendering approach, thereby enabling to evaluate manipulated data frames through effective preprocessing and feature extractors. The study entails a juxtaposed pipeline of preprocessors with feature extractors to effectively analyse the tamper score procured for real and fake video frames. The results thoroughly identified the dominance of Resizecrop and ResNet18 as best preprocessing and feature extractors respectively. Nonetheless, the 'DeepTamperNet' held a pivotal role in localizing tamper points to establish itself as the best interpretable model. The variance-based tamper scoring model aids in effectively establishing a quantifiable framework for asymmetrical spatial patterns subjected to manipulation. This indagation also signifies the crucial aspect of pre-processing, that serves as a feature unsheathing backbone to efficaciously construe tamper stratification pipelines. Nonetheless, while the proposed model to identify deepfakes holds dynamic embedding of monitoring and detection mechanisms, the present installation of this method into real-world applications is yet to be implemented. However, the challenges identified during the implementation of the proposed study is on significantly striking a balance between detection accuracy and simulative efficacy, particularly for resource-constrained environments like mobile applications. Beyond the technical sphere, constraints and ambiguities in the realm of ethical, legal, and privacy with respect to storing, processing, and labelling manipulated content severs its colossal adoption in various verticals. Future work of study pertaining this research can entail more classifier models, and dynamic embedding of the framework into applications for real-time utilization and establishing this as a valuable contributor to evade cybercrimes.

References

[1] 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 Deepfakes Creation and Detection: A Survey", Computer Vision and Image Understanding, doi.org/10.1016/j.cviu.2022.103525, 2019

[2] Vrizlynn L. L. Thing, "Deepfake Detection with Deep Learning: Convolutional Neural Networks versus Transformers", doi.org/10.48550/arXiv.2304.03698, 2023.

[3] Zhiqing Guo, Lipin Hu, Ming Xia, and Gaobo Yang, "Blind detection of glow-based facial forgery", Multimedia Tools and Applications, 80(5):7687–7710, 2021.

[4] J. Naruniec, L. Helminger, C. Schroers and R. M. Weber, "High-resolution neural face swapping for visual effects," in Computer Graphics Forum, 2020.

[5] R. Tolosana, S. Romero-Tapiador, J. Fierrez and R. VeraRodriguez, "Deepfakes evolution: Analysis of facial regions and fake detection performance," in International Conference on Pattern Recognition, 2021.

[6] 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 & Soft Computing DOI: 10.32604/iasc.2023.029653, 2022

[7] https://www.realitydefender.com/insights/history-of-deepfakes

[8] U. Rahul, M. Ragul, K. R. Vignesh and K. Tejeswini, "Deepfake video forensics based on transfer learning," International Journal of Recent Technology and Engineering (IJRTE), vol. 8, no. 6, pp. 5069–5073, 2020

[9] H. S. Shad, M. M. Rizvee, N. T. Roza, S. M. A. Hoq, M. M. Khanet, "Comparative analysis of deepfake image detection method using

convolutional neural network," Computational Intelligence and Neuroscience, vol. 2021, no. 3111676, pp. 1–18, 2021.

[10] L. Guarnera, O. Giudice and S. Battiato, "DeepFake detection by analyzing convolutional traces," in Proc. 2020 IEEE/CVF Conf. on Computer Vision and Pattern Recognition Workshops (CVPRW), Seattle, WA, USA, pp. 2841–2850, 2020.

[11] M. Patel, A. Gupta, S. Tanwar and M. S. Obaidat, "Trans-DF: A transfer learning-based end-to-end deepfake detector," in Proc. IEEE 5th Int. Conf. on Computing Communication and Automation (ICCCA), Greater Noida, India, pp. 796–801, 2020.

[12] Jhanvi Jheelan, Sameerchand Pudaruth, "Using Deep Learning to Identify Deepfakes Created Using Generative Adversarial Networks", Computers, 14(2), 60; doi.org/10.3390/computers14020060, 2025

[13] Raza A, Munir K, Almutairi M, "A Novel Deep Learning Approach for Deepfake Image Detection", Appl. Sci., 12, 9820, 2022

[14] Walczyna T, Piotrowski Z, "Fast Fake: Easy-to-Train Face Swap Model", Appl. Sci., 14, 2149, 2024

[15] Janutėnas L, Janutėnaitė-Bogdanienė J, Šešok D, "Deep Learning Methods to Detect Image Falsification", Appl. Sci., 13, 7694, 2023

[16] Ayesha Aslam, Jamaluddin Mir, Gohar Zaman, Atta Rahman, Asiya Abdus Salam, Farhan Ali, Jamal Alhiyafi, Aghiad Bakry, Mustafa Jamal Gul, Mohammed Gollapalli, Maqsood Mahmud, "Extracting Facial Features to Detect Deepfake Videos Using Machine Learning", International Journal of Advanced Computer Science and Applications, Vol. 16, No. 4, 2025

[17] T. Jung, S. Kim, and K. Kim, "Deepvision: Deepfakes detection using human eye blinking pattern," IEEE Access, vol. 8, pp. 83144–83154, 2020

[18] Luca Guarnera, Oliver Giudice, Cristina Nastasi, Sebastiano Battiato, "Preliminary Forensics Analysis of DeepFake Images", AEIT International Annual Conference (AEIT), Catania, Italy, pp. 1-6, doi: 10.23919/AEIT50178.2020.9241108, 2020.