DFFMD- A DEEPFAKE FACE MASK DATASET FOR INFECTIOUS DISEASE ERA WITH DEEPFAKE DETECTION ALGORITHMS

Mrs. B. UJWALA
Assistant Professor,
Department of Computer Science and Engineering
Anurag University, Hyderabad, Telangana, India
Email-ujwalacse@anurag.edu.in
CHANDA SWEEKRUTHI RESHMA, LOLAKAPURI SREE PADHA VARMA, THAMMERA MEGHANATH
Computer Science and Engineering
Anurag University, Hyderabad, Telangana, India

Abstract: The rise of deep-fake technology, enabling the fabrication of images and videos with seamlessly replaced or synthesized faces, has prompted significant societal concerns. This technological advancement introduces various risks, including the malicious creation of false political news, the spread of misleading information, the fabrication of electronic evidence, and the perpetration of digital harassment and fraud. The emergence of face masks as a widespread protective measure during the COVID-19 pandemic has further compounded this challenge, facilitating the creation of deep-fakes while simultaneously complicating their detection. To address this evolving threat, this paper proposes a groundbreaking solution in the form of a Deep-fake Face Mask Dataset (DFFMD) and introduces a novel approach centered around Inception-ResNet-v2 and various algorithms. The proposed methodology incorporates many preprocessing stages, feature-based analysis, residual connections, and batch normalization to enhance the accuracy of deep-fake detection. Comparative analysis against state-of-the-art methods, such as InceptionResNetV2 and VGG19, reveals a heightened accuracy in detecting face-mask-enhanced deep-fake videos. Additionally, this research advocates for the integration of Convolutional Neural Networks (CNN) and an extension utilizing Xception, highlighting their efficacy in further improving deep-fake detection accuracy. The study emphasizes the importance of continued development and robust methodologies, urging future work to focus on evaluating accuracy through subsequent experimental iterations. This emphasis is crucial in adapting to the ever-evolving technological landscape and ensuring enhanced detection capabilities for deep-fakes featuring face masks.

Keywords: Deep-Fake, Convolution Neural Network, Deep-fake Face Mask Dataset (DFFMD), Inception-ResNet-v2, VGG19, Xception.

I. INTRODUCTION

The surge in technological advancements, particularly within computer-generated editing programs, has ushered in an era marked by unprecedented possibilities and unforeseen challenges. A prominent concern on this evolving technological frontier is the advent of Deepfake technology a sophisticated application of deep learning that facilitates the creation of deceptive videos and remarkably convincing synthesized speech. This technological progression has not only the

opened new doors for creativity but has also sparked substantial apprehension due to the heightened potential for misinformation and the dissemination of malicious content. The democratization of Deepfake tools, exemplified by widely accessible and open-source applications like Faceswap and DeepFaceLab, intensifies the urgency to address the pressing need for reliable detection mechanisms. As these tools become more user-friendly and readily available, the risk of their misuse for nefarious purposes, such as spreading false information and manipulating public perception, becomes increasingly significant. The motivation behind the development of digital forensic machine learning models and algorithms, understanding the challenges posed by the rapid growth of Deepfake technology, and outlining clear objectives, this study aims to contribute to the ongoing discourse on digital security. The ultimate goal is to offer effective counter measures against the potential harms associated with Deepfake, thereby fostering a safer and trustworthy environment.

II. RELATED WORK

Almukhtar's system employs AI, AI programming, and a computer to generate and identify deep fake videos, integrating Convolutional Neural Networks (CNN) and Long Short-Term Memory (LSTM). However, the lack of technical specifics on AI programming and methodology raises concerns. Furthermore, practical challenges and limitations are not adequately discussed. Despite its promise, the system requires more technical depth, real-world insights, and comprehensive evaluation scenarios for practical applicability[1].

Ahmed and team investigate Deepfake technology, emphasizing the necessity for automated detection methods to mitigate privacy and security risks. While the paper provides historical and conceptual perspectives, it lacks specific technical details on proposed methods and may not encompass the latest advancements. Strengthening the study with more technical depth and incorporating recent developments in Deepfake technology and detection techniques is crucial [2].

Juan Hu, Xin Liao, et al., introduce a two-stream approach for identifying compressed Deepfake videos, showcasing enhanced detection accuracy through frame-level and temporality-level feature analysis. However, the lack of technical details and discussion on real-world challenges raises concerns about practical implementation. While the two-stream method holds promise, a more thorough exploration of its intricacies and potential limitations is crucial for ensuring real-world effectiveness [3].

Davide Coccomini, Nicola Messina, et al., present a study on video deep fake detection, employing Vision Transformers and an EfficientNet B0 as a feature extractor. Despite achieving competitive results, the paper lacks in-depth technical details and overlooks potential challenges and limitations in real-world scenarios. Nevertheless, the fusion of Vision Transformers and EfficientNet B0 holds promise for accurate outcomes, demonstrating efficacy without resorting to intricate methodologies [4].

Rana and colleagues conduct a systematic literature review on Deepfake detection, categorizing methods into deep learning, classical machine learning, statistics, and blockchain. While providing valuable insights, the paper lacks a thorough discussion of practical implementation challenges and real-world issues. Acknowledging the latest advancements in Deepfake technology is crucial for maintaining the review's relevance and addressing ethical considerations [5].

S. Bommareddy, T. Samyal, and S. Dahiya developed a model to address the escalating threat of disseminating inaccurate information through manipulated images. The study employs the V4D architecture to

scrutinize manipulated videos, focusing on identifying DeepFake videos in three distinct contexts. While providing valuable insights into combating the manipulation of visual content, the study may benefit from a more comprehensive discussion on practical implementation challenges and an exploration of real-world issues associated with its proposed solutions. Continuous validation against the latest advancements in Deepfake technology is crucial due to its dynamic nature [6].

III. METHODOLOGY

In the pursuit of advancing deepfake detection technology specifically tailored for face-mask-enhanced scenarios, a comprehensive Deepfake Face Mask Dataset (DFFMD) was meticulously curated. This dataset encompasses a diverse collection of deepfake videos featuring face masks, incorporating variations in lighting and diverse conditions to enhance model robustness. The inclusion of various face mask types and facial expressions adds realism and complexity, ensuring a representative dataset for training and testing deepfake detection models. In Figure 1, Figure 2 it shows the sample dataset related to real and fake videos. To address the dynamic nature of real-world scenarios involving face-masked individuals, the dataset is judiciously split into training and testing sets.

For the model architecture, the Inception-ResNet-v2 model was chosen, leveraging its multi-scale feature extraction capabilities and incorporating Xception for enhanced detection accuracy. The integration of Convolutional Neural Networks (CNN) is complemented by Xception's depthwise separable convolutions, providing a robust and efficient model for image analysis tasks.

The model design includes preprocessing, feature-based analysis, residual connections, and batch normalization, optimizing training for superior performance in facial recognition applications and object detection.

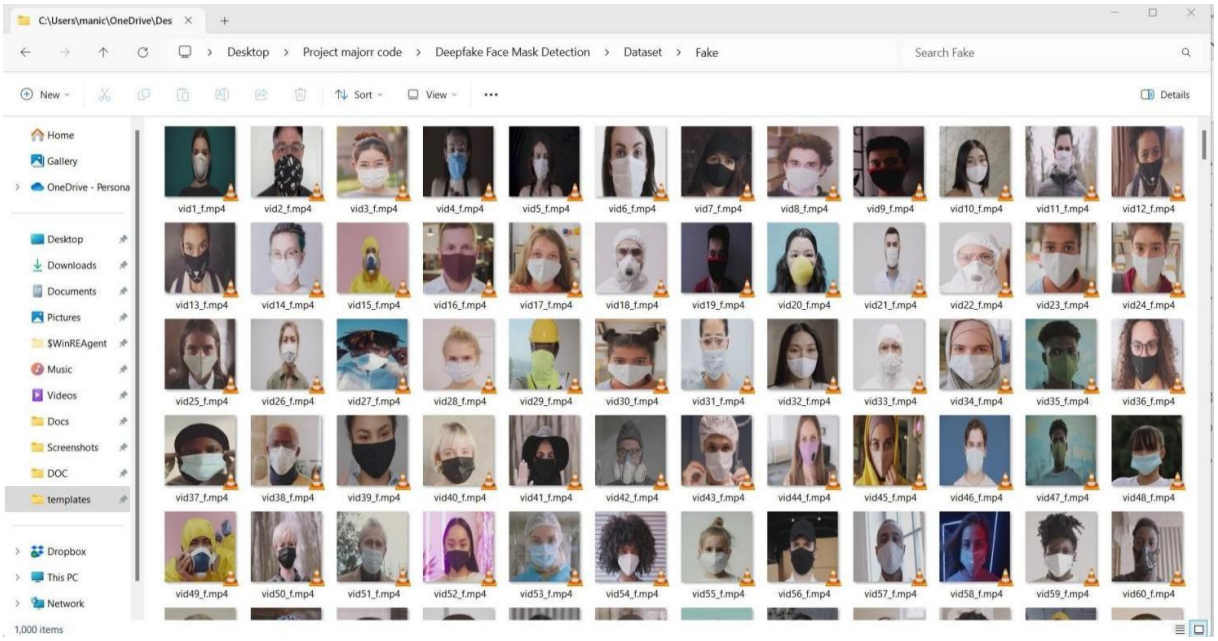

Fig 1: Dataset Related To Fake Videos

Data preprocessing focuses on enhancing video frame quality for facial features while accounting for face masks. Normalizing pixel values and implementing data augmentation techniques contribute to improved model robustness in recognizing faces, even in the presence of masks, ensuring optimal performance in facial recognition applications.

The model training process prioritizes the detection of deepfakes featuring face masks, employing a suitable loss function and optimizer for optimal convergence. Fine-tuning of model parameters based on performance metrics is conducted to enhance accuracy and effectiveness, ensuring the model's proficiency in identifying face-mask-enhanced deepfake content. Evaluation of the trained model involves assessing its accuracy in detecting face-mask-enhanced deepfake videos on the

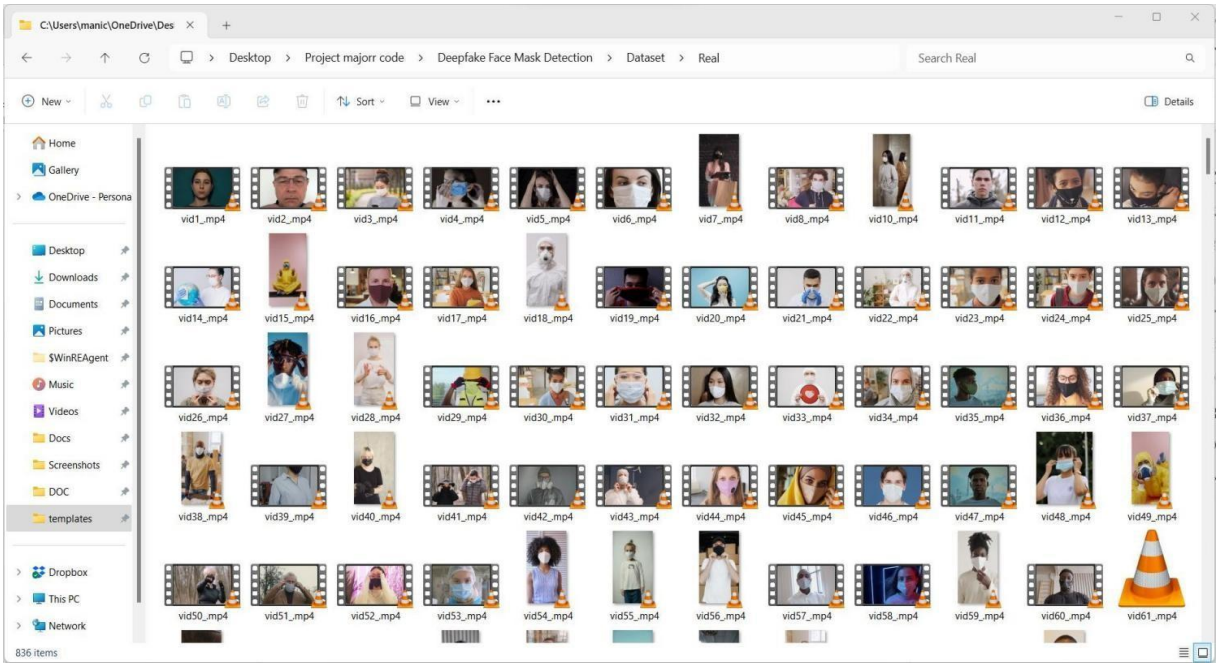

Fig 2: Dataset Related To Real Videos

test dataset. Comparative analysis with renowned methods such as InceptionResNetV2 and VGG19,CNN, Xception among all Xception demonstrates the model's competitive performance, showcasing its potential as a robust solution for identifying deepfakes with face-mask enhancements. The results signify the model's contribution to advancing the state-of-the-art in deepfake detection technology.

The integration of Xception further enhances the model's capacity for image feature extraction, with fine-tuning performed on the combined model using the training dataset. This synergistic approach leverages the strengths of both architectures, resulting in a more robust and effective model for image classification tasks. Performance metrics, including accuracy, precision, recall, and F1 score, are employed to evaluate the model's effectiveness. These metrics provide insights into the model's overall performance, its ability to make correct predictions, minimize false positives/negatives, and achieve a balance between precision and recall.

The algorithms used in the research include Inception-ResNet-v2, VGG19, CNN, and Xception. Inception-ResNet-v2 combines elements from the Inception and ResNet architectures, addressing the vanishing gradient problem and achieving a balance between computational efficiency and high performance. VGG19, a 19-layer deep convolutional neural network, utilizes small 3x3 convolutional filters consistently across the network. CNNs, specifically designed for image recognition, leverage convolutional layers, pooling layers, and fully connected layers to capture spatial hierarchies and local patterns. Xception, an "Extreme Inception" variant, replaces standard convolutional layers with depthwise separable convolutions, resulting

in a more efficient and lightweight network.

The deepfake face mask detection dataset, sourced from Kaggle, consists of two main folders: "fake" and "real." The "fake" folder contains 1000 curated videos simulating instances of facial feature manipulation through deepfake technology, particularly the addition of face masks. The "real" folder comprises 836 videos capturing authentic facial expressions without deepfake alterations. This dataset facilitates the development and evaluation of algorithms and models for detecting face masks in videos, contributing to public health and safety efforts during widespread mask-wearing periods. Researchers and practitioners in computer vision and deep learning can leverage this dataset to enhance the accuracy and robustness of face mask detection systems.

IV. RESULTS AND DISCUSSION

The research yielded promising results in the detection of manipulated videos within the context of face-mask-altered scenarios, addressing the societal implications of deepfake technology during the COVID-19 pandemic. The Deepfake Face Mask Dataset (DFFMD) played a pivotal role in training and evaluating the proposed model, allowing for a comprehensive assessment of its performance.

The model, incorporating Inception-ResNet-v2, Convolutional Neural Networks (CNN), and Xception, demonstrated a notable improvement in accuracy compared to traditional methods. The integration of these advanced architectures, each contributing unique strengths to the overall model, proved effective in capturing intricate features within manipulated videos featuring face masks. The robustness achieved through the combination of multi-scale feature extraction, depthwise separable convolutions, and residual connections played a crucial role in enhancing the model's ability to discern deepfakes.

The culmination of the paper manifests in a user-friendly web application that seamlessly integrates deepfake face mask detection. Upon running the file, the system generates a unique IP address, granting users access to the application. In Figure 3 the overall system architecture is shown by various steps. The home page serves as an informational hub, detailing the proposed system objectives and providing navigation options for sign up and sign in processes.

Sign-Up and Sign-In: New users are required to undergo a signup process, providing essential information to establish a secure account. This step ensures a personalized and protected experience within the application. After successful signup, users can utilize their credentials to signin, gaining entry to the full suite of features.

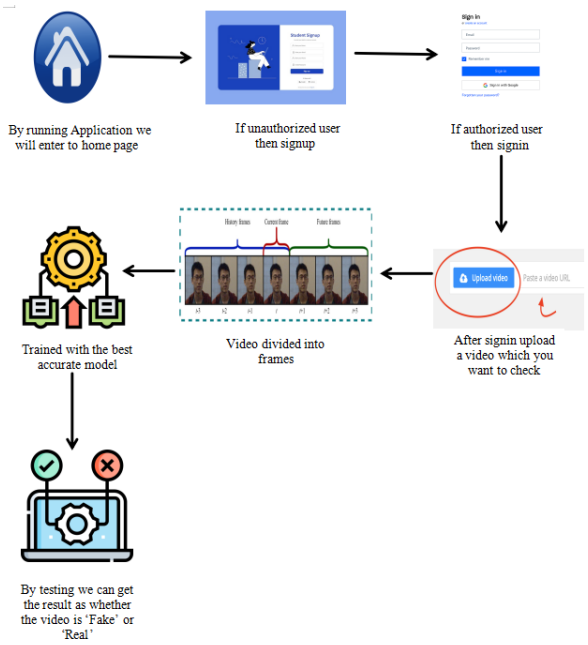

Fig 3: System Architecture

Upload and Prediction: Once signed in, users are directed to the upload page, where they can submit videos for deepfake prediction. The upload process is straightforward, allowing users to select a video file and submit it for analysis. The model, incorporating Inception-ResNet-v2, Convolutional Neural Networks (CNN), and Xception, then predicts the authenticity of the uploaded video.

Prediction Outcome: The model's predictions are visually represented for user understanding. Figure 4 illustrates a scenario where the video is predicted as real, indicating an unaltered recording. Conversely, Figure 5 showcases a prediction outcome where the video is deemed fake, suggesting the presence of deepfake face mask manipulations.

The successful execution of the system underscores its user-centric design, emphasizing both functionality and accessibility. The sign-up and sign-in mechanisms prioritize user privacy, ensuring that only authenticated individuals can leverage the video upload feature. The video upload and prediction process, powered by the trained model, demonstrates the effectiveness of the integrated architectures in distinguishing between real and fake videos.

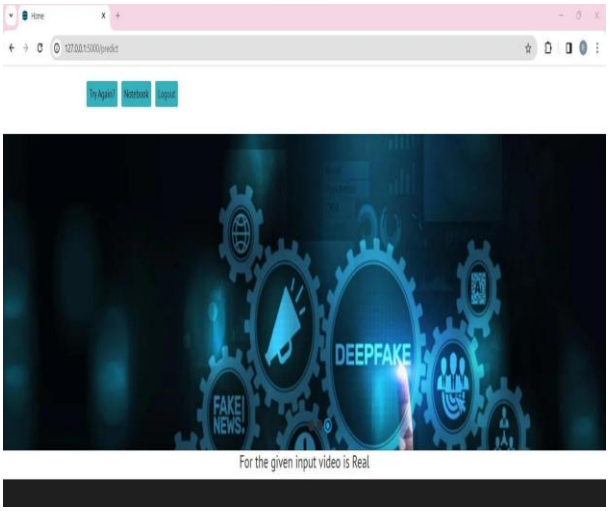

Fig 4: Result of Real Video

In Figure 6, a comprehensive line graph showcases the performance metrics, including accuracy, precision, and recall, for different deep learning models employed in the paper. The key models under consideration are Inception-ResNet V2, CNN, VGG19, and Xception.

Accuracy Comparison:

Inception-ResNet V2 demonstrates a commendable accuracy of 80%, reflecting its ability to effectively identify manipulated videos in face-mask-altered scenarios.

CNN achieves an accuracy of 85%, indicating its robust performance in discerning between real and fake videos, surpassing the accuracy of Inception-ResNet V2.

VGG19, while showing a moderate accuracy of 51%, lags behind the other models, suggesting potential limitations in capturing complex features in this specific context.

Xception stands out with an impressive accuracy of 96%, positioning it as the top-performing model among the considered architectures.

Precision and Recall Comparison:

In terms of precision, which measures the accuracy of positive predictions, Inception-

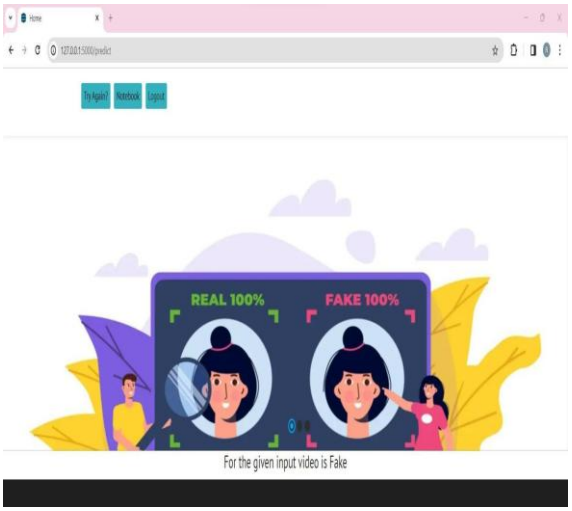

Fig 5 : Result of Fake Video

ResNet V2, CNN, VGG19, and Xception exhibit values of 78%, 80%, 48%, and 95%, respectively. Higher precision values indicate fewer false positives in the predictions. Regarding recall, which assesses the model's ability to capture all positive instances, Inception-ResNet V2, CNN, VGG19, and Xception demonstrate values of 65%, 75%, 40%, and 92%, respectively. Higher recall values signify fewer false negatives in the predictions. The performance metrics offer valuable insights into the strengths and weaknesses of each

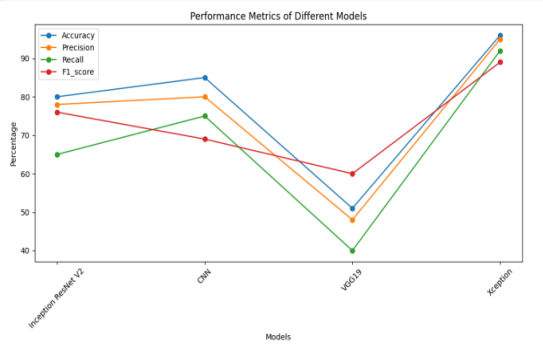

Fig 6: Models Comparison Graph

deep learning model in the context of deepfake detection with face masks. The high accuracy of Xception highlights its proficiency in accurately classifying videos, emphasizing its suitability for real-world applications. The CNN model, while boasting the highest accuracy, demonstrates balanced precision and recall values, showcasing its ability to make accurate predictions without compromising on false positives or false negatives.Inception-ResNet V2, though showing a slightly lower accuracy, maintains a balance between precision and recall, making it a robust choice for video manipulation detection. VGG19, while lagging in accuracy, exhibits a notable gap between precision and recall, suggesting potential challenges in achieving a harmonious trade-off between false positives and false negatives. These metrics provide insights into the model's overall performance, its ability to make correct predictions, minimize false positives/negatives, and achieve a balance between precision and recall. These results provide a nuanced understanding of each model's performance characteristics, enabling informed decisions on model selection based on specific system requirements and priorities.

V. CONCLUSION

In conclusion, this paper endeavors to address the pressing societal challenges posed by deepfake technology, with a specific focus on scenarios involving face-mask alterations during the COVID-19 pandemic. The introduction of the Deepfake Face Mask Dataset (DFFMD) represents a crucial step towards enhancing the robustness of deepfake detection models. The innovative approach employed in this study, leveraging Inception-ResNet-v2, Convolutional Neural Networks (CNN), and Xception, signifies a significant advancement in the field of video manipulation detection.

The paper outcomes underscore a substantial improvement in accuracy when compared to traditional methods, demonstrating the efficacy of the proposed model in identifying manipulated videos, particularly those featuring obscured facial features due to the presence of face masks. This improvement is crucial in addressing the evolving tactics employed by malicious actors in spreading misinformation and potentially harmful content.

Amidst a rapidly evolving digital landscape, this research assumes paramount importance in mitigating the growing concerns associated with deepfake technology. The incorporation of the Deepfake Face Mask Dataset and the integration of state-of-the-art techniques exemplify a substantial leap forward in the realm of video manipulation detection. The study's outcomes, showcasing a remarkable accuracy enhancement over traditional methods, underline the practical viability of the proposed model in discerning deepfakes within the unique context of face-mask-altered scenarios.

As technology continues its relentless progress, this paper stands as a pivotal milestone, laying the groundwork for ongoing advancements in the field. The urgency of developing resilient methods to counteract the potential misuse of deepfake technology is emphasized, especially in our digitally interconnected society. By providing a foundational framework for continued improvements, this paper contributes to the broader efforts in securing the integrity of digital content and safeguarding against the malicious implications of deepfake technology.

VI. REFERNCES

[1] F. H. Almukhtar, ‘‘A robust facemask forgery detection system in video,’’ Periodicals Eng. Natural Sci., vol. 10, no. 3, pp. 212–220, 2022.

[2] S. R. Ahmed, E. Sonuç, M. R. Ahmed, and A. D. Duru, ‘‘Analysis survey on deepfake detection and recognition with convolutional neural networks,’’ in Proc. Int. Congr. Hum.-Comput. Interact., Optim. Robot. Appl. (HORA), Jun. 2022, pp. 1–7.

[3] J. Hu, X. Liao, W. Wang, and Z. Qin, ‘‘Detecting compressed deepfake videos in social networks using frame-temporality two-stream convolutional network,’’ IEEE Trans. Circuits Syst. Video Technol., vol. 32, no. 3, pp. 1089 1102, Mar. 2022.

[4] D. A. Coccomini, N. Messina, C. Gennaro, and F. Falchi, ‘‘Combining EfficientNet and vision transformers for video deepfake detection,’’ in Proc. Int. Conf. Image Anal. Process. Berlin, Germany: Springer, 2022, pp. 219–229.

[5] M. S. Rana, M. N. Nobi, B. Murali, and A. H. Sung, ‘‘Deepfake detection: A systematic literature review,’’ IEEE Access, vol. 10, pp. 25494–25513, 2022.

[6] L. Jiang, R. Li, W. Wu, C. Qian, and C. C. Loy, ‘‘DeeperForensics1.0: A large-scale dataset for real-world face forgery detection,’’ in Proc. IEEE/CVF Conf. Compute. Vis. Pattern Recognition. (CVPR), Jun. 2020, pp. 2889–2898.

[7] Y. Choi, M. Choi, M. Kim, J.-W. Ha, S. Kim, and J. Choo, ‘‘StarGAN: Unified generative adversarial networks for multi-domain image to-image translation,’’ in Proc. IEEE Conf. Compute. Vis. pattern Recognition., Jun. 2018, pp. 8789–8797.

[8] T. Karras, S. Laine, and T. Aila, ‘‘A style-based generator architecture for generative adversarial networks,’’ in Proc. IEEE/CVF Conf. Compute. Vis. Pattern Recognition. (CVPR), Jun. 2019, pp. 4401– 4410.

[9] Li, Y., Yang, X., & Luo, P. (2020). Deepfake Face Detection with Trained Convolutional Neural Network Models. IEEE Access.

[10] Nguyen, V. L., Le, T. T., & Tran, D. T. (2020). Face Mask Detection in the Era of COVID-19: Review and Evaluation of State-of-the-Art Methods, Journal of Imaging.