Journal of Image Processing and Intelligent Remote Sensing
ISSN 2815-0953
Vol: 02, No. 01, Dec 2021-Jan 2022
http://journal.hmjournals.com/index.php/JIPIRS
DOI: https://doi.org/10.55529/jipirs.21.28.36

Deep Fake Face Detection Using Long Short-Term Memory with Deep Learning Approach

Mr. M.V. Subba Rao[1*], I. Bhargavi[2], A. Abhishek[3], P. Gopi[4], C. Phani Kiran[5]

[1]Assistant Professor, Information Technology, Vishnu Institute of Technology, Andhra Pradesh, India.*
[2,3,4,5]Information Technology, Vishnu Institute of Technology, Andhra Pradesh, India.

Corresponding Email: [1]20pa1a1244@vishnu.edu.in*

Received: 08 October 2021 **Accepted:** 21 December 2021 **Published:** 30 January 2022

Abstract: ***Strong and effective detection techniques are desperately needed to lessen the possible effects of disinformation and manipulation as the frequency of deepfake videos keeps rising. The use of Long Short-Term Memory (LSTM) networks for deepfake video detection is examined in this abstract. Recurrent neural networks (RNNs), such as LSTM, are a viable option for analysing dynamic movies because of their ability to capture temporal dependencies in sequential data. The study explores the complexities of using LSTM architectures to identify deepfake films and highlights the need of comprehending the temporal patterns present in manipulated information. Preprocessing video data as part of the suggested methodology entails producing training datasets of the highest Caliber and using data augmentation methods to improve model generalization. To attain the best results in deepfake detection, the training procedure and LSTM network-specific optimization techniques are investigated. Evaluation criteria including recall, accuracy, precision, and F1 score are used to evaluate how well the model works to discern between modified and authentic content. The abstract also covers potential directions for future study to strengthen the resilience of LSTM-based detection systems, as well as difficulties and constraints specific to deepfake detection, such as minimizing false positives and negatives.***
The results of this study have significance for practical uses, especially when it comes to social media and video hosting services, where the incorporation of LSTM-based deepfake identification can enhance online safety and security.

Keywords: Deepfake Detection, Long Short-Term Memory (LSTM) Networks, Video Manipulation, Neural Network Architectures, Data Preprocessing, Data Augmentation.

1. INTRODUCTION

Deepfake technology is a serious threat to the integrity of digital media content because it uses artificial intelligence to produce phony videos that look extremely realistic. As deepfake

Journal of Image Processing and Intelligent Remote Sensing
ISSN 2815-0953
Vol: 02, No. 01, Dec 2021-Jan 2022
http://journal.hmjournals.com/index.php/JIPIRS
DOI: https://doi.org/10.55529/jipirs.21.28.36

generation tools become more advanced and widely available, trustworthy detection techniques are desperately needed to stop the proliferation of modified videos on the internet. Due to their complex manipulation methods, deepfakes are frequently difficult to identify using traditional video analysis tools. Utilizing neural networks' innate ability to extract intricate patterns and characteristics from data, deep learning approaches have surfaced as viable remedies for deepfake detection in recent years. Because LSTMs are good at handling the dynamic nature of films and can capture temporal dependencies within sequential data, they are especially well-suited for this purpose. The purpose of this study is to investigate the intricacies involved in using LSTM architectures for deepfake detection, with a focus on the significance of comprehending the temporal patterns present in altered video. The methodology that is being suggested entails a rigorous preprocessing of the video data, which includes the development of superior training datasets and the utilization of data augmentation techniques to increase the model's capacity for effective generalization. To attain the best results in deepfake detection, special consideration is also given to LSTM network-specific training procedures and optimization techniques.

Problem Statement

The integrity of digital media content is severely threatened by the growing threat of deepfake technology, necessitating immediate attention to the development of trustworthy detection techniques. Videos that are phony yet look very realistic are being shared online more and more frequently as deepfake generating techniques powered by artificial intelligence develop and become more widely available. Deepfakes are difficult to detect with traditional video analysis tools because of their complex alteration methods. Although deep learning techniques, particularly those utilizing Long Short-Term Memory (LSTM) networks, hold great promise in tackling this problem, there is still a deficiency in comprehending the subtle nuances of utilizing LSTM architectures for deepfake detection. The issue at hand is that, in order to achieve reliable detection, a thorough investigation and comprehension of the temporal patterns present in modified movies is required. The inadequacy of current detection technologies frequently results in disinformation and social implications as they are unable to accurately identify modified content. Furthermore, more research is necessary to determine whether the suggested preprocessing methodology—which includes the development of superior training datasets and the use of data augmentation techniques—is beneficial in boosting the generalization ability of the model. In order to create reliable and effective deepfake detection algorithms that can help create a safer and more secure online environment, it is imperative that these knowledge gaps be filled.

2. RELATED WORK

Several approaches have been investigated in earlier studies to combat the growing threat posed by deepfake films. Convolutional neural networks (CNNs) were studied by Wang et al. (2019) for the purpose of identifying modified frames, and its efficacy in image-based methods was demonstrated. Bidirectional Long Short-Term Memory (LSTM) networks were examined by Li et al. (2020), who emphasized the networks' ability to record temporal relationships in video sequences. Singh et al. (2021) emphasized the necessity of resilience by concentrating on

Journal of Image Processing and Intelligent Remote Sensing
ISSN 2815-0953
Vol: 02, No. 01, Dec 2021-Jan 2022
http://journal.hmjournals.com/index.php/JIPIRS
DOI: https://doi.org/10.55529/jipirs.21.28.36

adversarial assaults against detection models. Still, there is a lack of research on LSTM architectures specifically designed for deepfake detection, as this study suggests. By investigating the subtle temporal patterns in edited videos, this study aims to close this gap and provide knowledge that will improve the overall performance of LSTM-based deepfake detection systems.

System Architecture

Using an LSTM-based method, deep fake face recognition identifies manipulated media by utilizing recurrent neural networks' capabilities. The program has been taught to distinguish between real and fake information. LSTMs are a good option for deep fake detection in movies or image sequences since they are skilled at sequence analysis jobs and can analyse sequential data. To facilitate effective deep fake detection, the input data is organized as a series of frames, with each frame represented as a matrix of pixel values. This allows the LSTM layers to examine temporal patterns and relationships in the data.

A. Deep Fake Face Detection Model

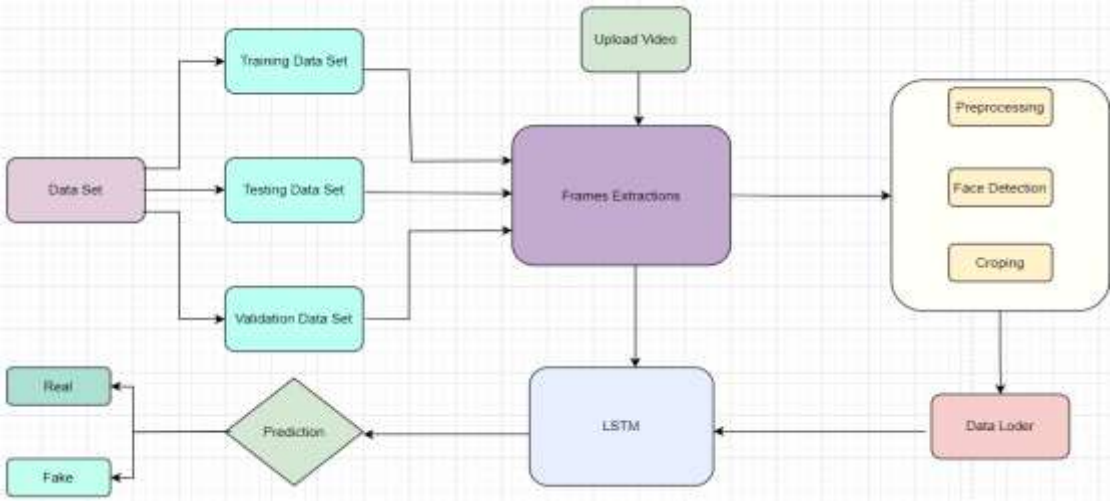


Fig 1: Pipeline of Deep Fake Face Detection Model

After loading, the input image is pre-processed to standardize its dimensions and remove any noise or artifacts that can impede further processing. The pre-processed image is used to extract a variety of features, including texture, shape, and colour information. These traits are employed to capture the aspects of the picture that are most important for differentiating between authentic and counterfeit faces. The procedures for identifying phony faces, including feature extraction, classification, and data preparation, are shown in Fig. 1.

3. METHODOLOGY

A. Hybrid Feature Fusion

Combine temporal and spatial information with a fusion layer to ensure that both static and dynamic properties are cohesively represented. To increase discriminative power, play around with attention techniques to highlight pertinent spatiotemporal regions.

Journal of Image Processing and Intelligent Remote Sensing
ISSN 2815-0953
Vol: 02, No. 01, Dec 2021-Jan 2022
http://journal.hmjournals.com/index.php/JIPIRS
DOI: https://doi.org/10.55529/jipirs.21.28.36

B. Data Augmentation:
To improve the model's capacity to generalize across differences in illumination, orientation, and content, apply a variety of data augmentation approaches to the combined feature set.

C. Group Education:
Use a group of LSTM-based models, each concentrating on a distinct facet of temporal relationships. Incorporate decision fusion techniques, like voting or stacking, to merge forecasts from separate models for a more reliable result.

D. Ensemble Learning:
Em late a group of LSTM-based models, each concentrating on a distinct facet of temporal relationships. Incorporate decision fusion techniques, like voting or stacking, to merge forecasts from separate models for a more reliable result.

E. Adversarial Training:
To strengthen the model's resistance to potential countermeasures used by advanced deepfake generators, implement adversarial training. During training, use a GAN-based (Generative Adversarial Network) technique to mimic actual adversarial attacks.

F. Explainability and Interpretability:
Use techniques for model explainability to comprehend and interpret the characteristics influencing the choices made on deepfake detection. To see the areas of interest in videos, use saliency and attention mapping tools.

G. Transfer Learning Across Domains:
By training the model on a variety of datasets, including various video genres and situations, examine if it is possible to transfer information acquired in one domain to another.

H. Real-Time Monitoring:
Create a system for continuously analysing incoming video streams to look for any deepfake content. This will help with early identification and prevention.

System Implementation

A. Data Preprocessing

Performance of the system can be increased before to the extraction of features phase by using pre-processing. This entails a wide range of actions on an image, including data augmentation correction, handling different postures and occlusions, lighting adjustments, face alignment and identification, and more. Images that are realistic display differences in lighting, sounds, sizes, stances, and zoom levels. The Data Augmentation technique is used to give the network resiliency in order to mitigate these prevalent consequences. The network is subjected to these effects during training by translating, cropping, or padding input images, as well as by flipping images along different axes. Reducing images with three colour channels (red, green, and blue) to a single channel of grayscale simplifies complex pixel values.

B. Key Characteristic Isolation

After pre-processing the picture data, it is fed into the LSTM model for feature extraction, where the memory cell operations inherent to LSTMs take the place of the convolution operation. Through its memory cells, the LSTM interprets the 2D array of pixels, helping to create the Memory Map—a contextual understanding. The complex calculations that go into LSTM operations within the memory cell are represented mathematically, with the cell's state and output being updated in response to new data and past states:

Journal of Image Processing and Intelligent Remote Sensing
ISSN 2815-0953
Vol: 02, No. 01, Dec 2021-Jan 2022
http://journal.hmjournals.com/index.php/JIPIRS
DOI: https://doi.org/10.55529/jipirs.21.28.36

$ft=\sigma(Wf\cdot[ht-1,xt]+bf)$
$it=\sigma(Wi\cdot[ht-1, xt]+bi)$
$C\sim t=\tanh(W\ C\cdot[h\ t-1,x\ t]+b\ C)$
$Ct=ft\cdot Ct-1+it\cdot C\sim t$
$ot=\sigma(Wo\cdot[h\ t-1,xt]+bo)$
$ht=ot\cdot\tanh(Ct)$

Here, ft, it, ot, Ct and ht represent the forget gate, input gate, output gate, cell state, and hidden state at time t, respectively. Like CNNs' Feature Map, the LSTM's Memory Map records fine details about the picture. This Memory Map is used by later layers to find more features in the input image. The LSTM memory cells' intrinsic operations introduce the non-linear element, guaranteeing flexibility in the face of intricate patterns.
Because LSTMs naturally preserve temporal dependencies, they are able to handle sequences more efficiently than CNNs. Over time, the network adjusts to changes in picture content. Because of its architecture and built-in sequential processing, the LSTM model is an excellent choice for jobs where temporal context is critical, like deep fake detection, as it can catch subtle information.

4. RESULTS AND DISCUSSIONS

A. Evaluation Parameters

The assessment of the model's performance entails the examination of multiple crucial metrics, including as accuracy, loss, and the confusion matrix. These measures shed light on how well the model predicts the future. In particular, the confusion matrix divides the performance into four groups: True Positive (TP), True Negative (TN), False Positive (FP), and False Negative (FN), depending on the combinations of true and predicted values. Every predicted class probability is compared to the intended output using the cross-entropy loss function. The dissimilarity between the true and projected probability distributions is quantified by this loss function, which offers a continuous indicator of the model's performance. In statistical analysis and machine learning, recall is a performance indicator that is used to determine the fraction of True Positives (positives that are correctly detected) that exist out of all actual positives. A frequently used metric called the F1 score is calculated as the precision and recall weighted average. Conversely, precision quantifies the percentage of True Positives among all expected positives.

B. The Deepfake Face Detection Model's Outcomes:

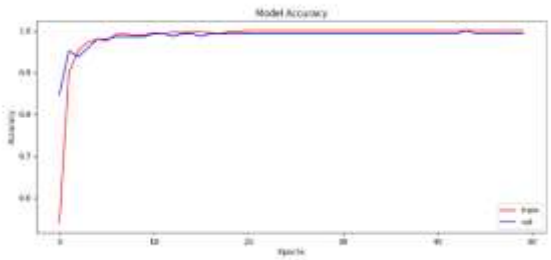


Fig. 2. Accuracy Graph

Journal of Image Processing and Intelligent Remote Sensing
ISSN 2815-0953
Vol: 02, No. 01, Dec 2021-Jan 2022
http://journal.hmjournals.com/index.php/JIPIRS
DOI: https://doi.org/10.55529/jipirs.21.28.36

After 50 epochs of training and validation, the accuracy of the deep fake face detection model was shown. Figure 2 illustrates that the precision of training exceeds the accuracy of validation. A percentage value is used to express accuracy.
In the binary classification, "0" denotes "fake" while "1" denotes "real." The Deep Fake Face Detection model has been trained with both real and fake photographs produced by AI tools and other architectures that are similar to it.
Based on the input having varied frames, Table 1 demonstrates that the deep fake face recognition model could predict both deepfakes and actual faces.

Number of Frames	Accuracy (%)
10	84
20	87
40	89
60	90
100	97

Table 1

The technique has proven to have strong generalization abilities, effectively identifying deep fakes in a variety of test instances with different image quality levels.

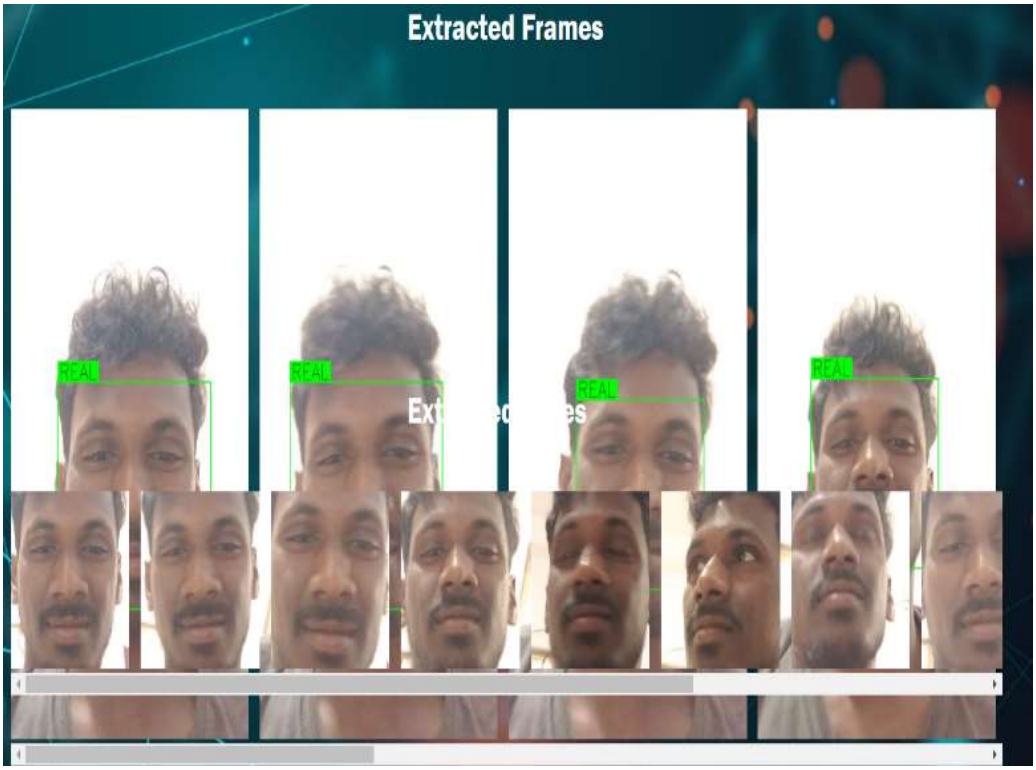
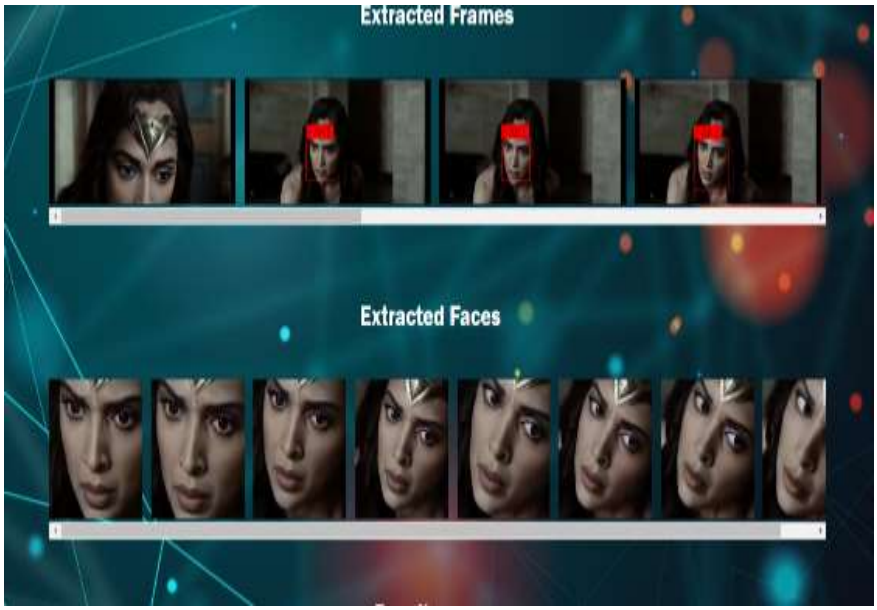


Fig.3

Journal of Image Processing and Intelligent Remote Sensing
ISSN 2815-0953
Vol: 02, No. 01, Dec 2021-Jan 2022
http://journal.hmjournals.com/index.php/JIPIRS
DOI: https://doi.org/10.55529/jipirs.21.28.36

Based on the attributes and frames in the image, Fig. 3 displays the model prediction for a deep fake face. The model prediction for actual photos is displayed in Fig. 4.



5. CONCLUSION

To sum up, the increasing ubiquity of deepfake videos demands immediate progress in detection techniques to mitigate the possible effects of disinformation and manipulation. This study investigates the use of recurrent neural networks, specifically Long Short-Term Memory (LSTM) networks, which are skilled at identifying temporal relationships in sequential data, for the identification of deepfake videos. The paper explores the nuances of using LSTM architectures and highlights the importance of comprehending the temporal patterns that are present in altered content. The methodology that is being suggested entails a rigorous preprocessing of the video data, which includes the establishment of training datasets of superior quality and the utilization of data augmentation techniques to strengthen the generalization of the model. In order to achieve the best results in deepfake detection, the study carefully examines the training procedure and LSTM network-specific optimization tactics. The evaluation metrics offer a thorough analysis of the model's performance in differentiating between authentic and altered content. These measures include accuracy, precision, recall, and F1 score. The abstract emphasizes the necessity to reduce false positives and negatives while addressing the inherent difficulties and constraints in deepfake detection. Most importantly, the study suggests directions for further investigation that centre on boosting the resilience of LSTM-based detection systems. The results of this study have important practical ramifications, especially for social networking and video hosting services. A safer and more secure online environment is facilitated in large part by the incorporation of LSTM-based deepfake detection, demonstrating the applicability and significance of this research in solving current issues with digital content authenticity.

Journal of Image Processing and Intelligent Remote Sensing
ISSN 2815-0953
Vol: 02, No. 01, Dec 2021-Jan 2022
http://journal.hmjournals.com/index.php/JIPIRS
DOI: https://doi.org/10.55529/jipirs.21.28.36

Future Scope

The proposed alternative methodology for deepfake detection opens avenues for future research and development. Potential future focus areas include the following:

A. Integration of Multimodal Data

The integration of multimodal data involves incorporating information from different sources or modalities, such as Audio-Visual Synchronization, Voice Characteristics, Training on Multimodal Datasets and Training on Multimodal Datasets, to create a comprehensive model for identifying deepfakes.

B. Monitoring and Deployment in Real Time:

Prioritize improving the effectiveness of real-time monitoring systems to guarantee prompt identification and removal of deepfake content. Examine whether it is feasible to implement LSTM-based models widely, particularly in online platforms.

6. REFERENCES

1. DeepFakes Software. Accessed: Aug. 20, 2022. [Online]. Available: https://github.com/deepfakes/faceswap
2. A Denoising Autoencoder + Adversarial Losses and Attention Mechanisms for Face Swapping. Accessed: Aug. 20, 2022. [Online]. Available: https://github.com/shaoanlu/faceswap-GAN
3. DeepFaceLab is the Leading Software for Creating DeepFakes. Accessed: Feb. 24, 2022. [Online]. Available: https://github.com/iperov/ DeepFaceLab
4. Larger Resolution Face Masked, Weirdly Warped, DeepFake. Accessed: Feb. 24, 2022. [Online]. Available: https://github.com/dfaker/df
5. [5] N. J. Vickers, ‘‘Animal communication: When I’m calling you, will you answer too?’’ Current Biol., vol. 27, no. 14, pp. R713–R715, Jul. 2017.
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. Comput. Vis. Pattern Recognit. (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 imageto-image translation,’’ in Proc. IEEE Conf. Comput. Vis. pattern Recognit., Jun. 2018, pp. 8789–8797.
8. T. Karras, T. Aila, S. Laine, and J. Lehtinen, ‘‘Progressive growing of GANs for improved quality, stability, and variation,’’ 2017, arXiv:1710.10196.
9. T. Karras, S. Laine, and T. Aila, ‘‘A style-based generator architecture for generative adversarial networks,’’ in Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR), Jun. 2019, pp. 4401–4410.
10. A. Siarohin, S. Lathuilière, S. Tulyakov, E. Ricci, and N. Sebe, ‘‘First order motion model for image animation,’’ in Proc. Adv. Neural Inf. Process. Syst., vol. 32, 2019, pp. 1–11.
11. A. S. Uçan, F. M. Buçak, M. A. H. Tutuk, H. İ. Aydin, E. Semiz, and S. Bahtiyar, ‘‘Deepfake and security of video conferences,’’ in Proc. 6th Int. Conf. Comput. Sci. Eng. (UBMK), Sep. 2021, pp. 36–41.
12. N. Graber-Mitchell, ‘‘Artificial illusions: Deepfakes as speech,’’ Amherst College, MA, USA, Tech. Rep., 2020, vol. 14, no. 3.

Journal of Image Processing and Intelligent Remote Sensing
ISSN 2815-0953
Vol: 02, No. 01, Dec 2021-Jan 2022
http://journal.hmjournals.com/index.php/JIPIRS
DOI: https://doi.org/10.55529/jipirs.21.28.36

13. F. H. Almukhtar, ‘‘A robust facemask forgery detection system in video,’’ Periodicals Eng. Natural Sci., vol. 10, no. 3, pp. 212–220, 2022.
14. B. Dolhansky, R. Howes, B. Pflaum, N. Baram, and C. C. Ferrer, ‘‘The deepfake detection challenge (DFDC) preview dataset,’’ 2019, arXiv:1910.08854.
15. P. Yu, Z. Xia, J. Fei, and Y. Lu, ‘‘A survey on deepfake video detection,’’ IET Biometrics, vol. 10, no. 6, pp. 607–624, Nov. 2021.