Journal of Innovative Image Processing (ISSN: 2582-4252)
www.irojournals.com/iroiip/

Hybrid Deep Learning Approach for Deepfake Detection Using ResNet50 and EfficientNetB0

Vaishnavi D.[1], Jegan S.[2], Ganesh J.[3], Srinidhi Sundaram[4]

[1,3]Department of CSE, SRC, SASTRA Deemed to be University, India.
[2]Department of CSE, Teaching Fellow, University College of Engineering Villupuram, India.
[4]Department of AI and DS, Panimalar Engineering College, Chennai, India.

E-mail: [1]vaishnavi@src.sastra.edu , [2]jeganucev@gmail.com, [3]ganesh_j@src.sastra.ac.in, [4]srinidhibalaji.pec@gmail.com

Abstract

Deepfake is an environment in which AI technology is used to manipulate original digital videos, making them appear real. This poses a serious issue regarding the trustworthiness of digital data. The goal of the study is to create a reliable method to detect deepfakes using enhanced learning models named EfficientB0 and RESNET50. Frames are extracted from videos, and a HAAR cascade is applied to locate the face region, which is then sent as a dataset to train the model. This study utilized an open dataset from Kaggle to perform the experiment. The performance of the study is quantitatively measured using F1-score, accuracy, recall, and precision. The experiment showed that the hybrid model achieved superior prediction results of 89.08% compared to the standalone models. Hence, this study confirms that the proposed model works well to identify fake videos, which may help increase trust in digital data.

Keywords: Haar Cascade, Artificial Intelligence, Transfer Learning, Deep Learning, Facial Recognition, Synthetic Media.

1. Introduction

Since AI has been added to digital video software, deepfakes are becoming common these days. Such deepfake content looks the same as authentic ones, which are created with an advanced machine learning algorithm called Generative Adversarial Networks (GANs). Deepfakes can be used in creative ways, but using them the wrong way can lead to significant problems with ethics, the law, and security. Deepfake videos can trick famous people and regular individuals into thinking they are in realistic, fake situations. This could change how people perceive them, which could hurt their identities. People like Chesney and Citron say that deepfakes are bad for democracy, national security, and people's privacy [1]. According to Deeptrace, the number of deepfake pictures has grown from 7,964 in 2019 to over 14,678 soon, which is more than twice as many as the previous year [2]. This rapid growth has been caused by the accessibility of open-source technology. The two common forensic methods, called watermarking and human verification, do not always work to identify such deepfakes. Since the AI-based deepfake detection system is becoming popular, these systems can collect small evidence that could not be noticed by people, such as strange eye movement, lighting incompatibility, or unusual facial expressions. Convolutional Neural Networks (CNNs) are

DOI: https://doi.org/10.36548/jiip.2025.3.003
Received: 18.06.2025, received in revised form: 17.07.2025, accepted: 03.08.2025, published: 12.08.2025

effective at identifying spatial anomalies in images, whereas hybrid models that combine CNNs, Recurrent Neural Networks (RNNs) or Long Short-Term Memory (LSTM) units detect periodic discrepancies between frames [3]. Pre-trained models that have been optimized for forgery detection, such as XceptionNet and EfficientNet, increase accuracy. Matern et al. emphasize the use of encoding artifacts to expose fakes [4]. With an emphasis on real-time deployment, this study presents a novel deep learning-based framework that combines ResNet50 and EfficientNetB0 for precise deepfake video detection. In contrast to earlier approaches, our technique emphasizes a simplified pipeline that combines efficient attribute learning, model optimization, and frame-wise face extraction via Haarcascade. In addition, a web interface is also implemented to provide a scalable and accessible detection. This specific combined method enables stabilization of synthetic media by sustaining digital authenticity.

2. Related Work

Numerous recent studies have proposed novel methods for detecting deepfake films, leveraging state-of-the-art AI and deep learning approaches to improve accuracy and robustness. A novel approach by [6] introduced a hybrid detection model that combines Convolutional Neural Networks (CNN) and Graph Neural Networks (GNN) using a dual-stream architecture. The authors utilized a four-block CNN to extract spatial features and Mini-Batch Graph Convolution to obtain relational characteristics in one stream. After testing on multiple datasets, the fusion algorithms FuNet-A (additive), FuNet-M (multiplicative), and FuNet-C (concatenation) were shown to be very effective at identifying deepfake content, achieving an accuracy rate of 99.3% after 30 epochs.

In [7], a blockchain-based deepfake detection system was presented that leverages IPFS, Ethereum Name Service, and a decentralized reputation system to verify the legitimacy of movies. The study examined CNN-based models, especially VGG-16, and showed MCNet, a network designed for categorizing manipulations in the spatial, frequency, and compression domains. The results revealed that the accuracy ranged from 84% to 99%, which was better than other methods like AutoGAN and Fakespotter. This ensured that content was safe and clear on decentralized networks. The in-depth study in [8] addressed the challenges and shortcomings of current methods for detecting deepfakes. It underlined the need for more research and the integration of data from multiple sources. It also pointed out that certain deep learning models are achieving accuracies of up to 96.8% but still struggle to keep pace with new deepfake techniques. Similarly, [9] provided a comprehensive review of facial modification detection, presenting a variety of datasets and detection methods. The study underlined the need for stronger detection frameworks to enhance their reliability and noted that deep learning-based solutions had reached accuracies of up to 94% accuracy.

An interconnected deep learning model combining the Xception and EfficientNet architectures was introduced in [10]. This technique classified the facial features extracted from video frames using the combination model. Tested on datasets such as FaceForensics++ and Celeb-DF, the method achieved 97.5% accuracy, outperforming every other model and other detection technique. In [11], an alternative novel method that combined Overlapping Multiple Dynamic Images (OMDI) and Inversed OMDI (I-OMDI) was introduced to capture temporal inconsistencies and minor visual artifacts. EfficientNetB7 was utilized for feature extraction, and the model employed an average-weighted fusion strategy with comparable weights for OMDI and I-OMDI. When evaluated on the Celeb-DF v2 and DFDC datasets, it performed better than existing methods, achieving AUC values of 0.9952 and 0.9947. The

CNN-based method in [12] needed to classify, extract facial characteristics, and enhance video frames to distinguish between real and fake movies. When tested using the fake Detection dataset, it showed 96.12% accuracy, proving CNNs' reliability in deepfake detection.

In a related study, CNN models were also used with data preparation techniques like augmentation, normalization, and resizing to enhance model performance [13]. By obtaining 97.5% reliability on the same dataset, it illustrated CNNs' effectiveness and discussed the challenges of real-time detection. Achieving 96.12% reliability on the Deepfake Detection Challenge dataset, a similar CNN-based technique in [14] confirmed CNNs' ability to ensure the authenticity of digital information. A CNN-RNN mixed model enhanced using Particle Swarm Optimization (PSO) was published in [15]. This model used CNN for spatial feature extraction and RNN for temporal analysis, employing PSO to optimize parameters. It demonstrated the value of combining temporal modeling and optimization with its 97.21% reliability. LSTM networks, which recorded temporal differences between video frames, were used in [16] to identify deepfakes. The system's accuracy of 95.87% on the Deepfake Detection Challenge dataset showed how effective LSTM is in learning temporal features. A complex hybrid system in [17] combined CNN, LSTM, and ResNeXt architectures to assess video authenticity. The effectiveness of the combined deep learning frameworks was demonstrated by the hybrid model's 97.8% accuracy on standard datasets, which was attained with preliminary processing and visual feature extraction. With a focus on social media security, the LSTM-based approach in [18] preprocessed frames from videos and trained an LSTM network. It achieved 96.45% accuracy when tested on the Celeb-DF v2 dataset, effectively capturing temporal artifacts to enhance identification reliability. A deep learning system in [19] handled photo and video frames for categorization with 98.7% accuracy on datasets like the Deepfake Detection Challenge. The results demonstrated the effectiveness of deep learning in enhancing security and effectively identifying fraudulent media. The work proposed in [20], a novel approach that uses Multiwavelet Transform to analyze blur inconsistencies in areas of the face, focusing on edge sharpness and distinctions among transformed faces and backgrounds. This method offers a workable strategy to enhance deepfake detection, with detection rates exceeding 93.5%.

The study from [21] looked at how well GANs, CNNs, and RNNs worked for detecting deepfakes in real time. GANs were the most accurate, with an accuracy rate of 88%, while RNNs had an accuracy rate of 85% and CNNs had an accuracy rate of 83%. Additionally, GANs had better precision and recall, which means fewer false positives and negatives. These results show how well GANs can learn generative features for finding deepfakes. The paper [22] paper shows how to use CNN (ResNeXt) and LSTM architectures together to create a deepfake detection system that can tell the difference between real and fake photos. Their method achieved 86% accuracy on a video-based dataset, demonstrating that the system can pick up both spatial and temporal cues to find deepfakes reliably. [23] suggested a deepfake detection framework that used CapsuleNet and ArCapsNet, achieving 82.84% reliability on the DFDC-P dataset. The golden ratio-based frame selection technique is a major contribution because it improves feature representation and enhances detection performance. This study shows how capsule systems and frame selection algorithms could help improve deepfake detection. The paper by [24] presents a method for discovering deep falsification using a CNN-MLP hybrid model, which was 81.25% accurate on the Celeb-DF dataset. The model effectively distinguishes between actual and false video content and outperforms many other methods currently used in media forensics. The paper [25] suggests a way to detect deepfakes that uses XResNet to extract features from each frame and LSTM to classify temporal sequences. The model underwent training and evaluation on the Meta DFDC dataset and was

able to correctly identify manipulated films 83.3% of the time by using both temporal and spatial information. Table 1 provides a concise overview of the major results, performance, and issues with the current deepfake detection methods discussed in the literature. It addresses the techniques utilized, the datasets used, their accuracy, and the drawbacks or limitations of each strategy. This combined view facilitates the identification of research gaps and highlights the necessity of the model being presented.

Table 1. Summary of Literature Review

Ref. No.	Methodology	Dataset Used	Identified Limitations / Challenges
[6]	CNN + GNN (Dual Stream)	FF++, Celeb-DF	High accuracy, but scalability and inference cost in real-time not discussed
[7]	Blockchain + CNN + MCNet	VGG-16, IPFS	Complex system setup; real-time verification not validated
[8]	Review of techniques	Multiple datasets	Lack of adaptability to evolving deepfake patterns; need for multimodal data integration
[9]	Comprehensive survey	Various	Need for robust frameworks; reliance on static features
[10]	EfficientNet + Xception	FF++, Celeb-DF	Performance under compression, occlusion, and varied lighting not explored
[11]	OMDI + I-OMDI + EfficientNetB7	Celeb-DF v2, DFDC	Fusion complexity and interpretability concerns
[12–14]	CNN-based detection	Fake Detection Dataset	Dependency on preprocessing; challenges in real-time application
[15]	CNN + RNN + PSO	Not specified	Increased computational cost due to PSO; parameter tuning complexity
[16]	LSTM-based detection	Deepfake Detection Challenge	Captures temporal features well but lacks spatial richness
[17]	CNN + LSTM + ResNeXt Hybrid	Standard datasets	Hybrid complexity and energy consumption may limit deployment feasibility
[18]	LSTM-based detection for social media	Celeb-DF v2	Focused only on social media; generalization across platforms is untested
[19]	Deep learning-based classification	DFDC	No mention of temporal inconsistencies or adversarial robustness
[20]	Multiwavelet Transform for blur detection	Not specified	Focused only on spatial blur; ignores motion and facial dynamics
[21]	GAN, CNN, RNN	Not specified	CNN and RNN models underperform compared to GANs; real-time efficiency of GAN not fully validated
[22]	ResNeXt + LSTM (CNN + RNN)	Video-based	Moderate accuracy; lacks robustness testing across varied datasets and evolving deepfake techniques

[23]	CapsuleNet + ArCapsNet	DFDC-P	Complexity due to golden ratio frame selection; performance drop on high-compression videos
[24]	CNN + MLP Hybrid	Celeb-DF	Limited generalization; tested on a single dataset; real-time scalability not demonstrated
[25]	XResNet + LSTM	Meta DFDC	Accuracy limited to 83.3%; performance may degrade with unseen manipulations; lacks real-time analysis

3. Proposed Work

The suggested method for finding deepfakes is designed to identify fake video material quickly and accurately. It starts by assembling a blarge set of real and fake videos. To extract relevant parts of the face, these videos are pre-processed by splitting each frame and employing face detection methods. Next, a dataset is created from the extracted faces so that the algorithm can learn. Transfer learning is used to extract and classify features in deep learning models. This ensures that the system can detect even minor changes. Finally, the trained model produces a prediction about whether the input video frames are real or have been altered. With this methodical approach, the system can adapt to new deepfake technologies while remaining scalable, reliable, and fast in real time. This work includes pre-processing, feature extraction, model development through training, testing, and prediction, as shown in Figure 1.

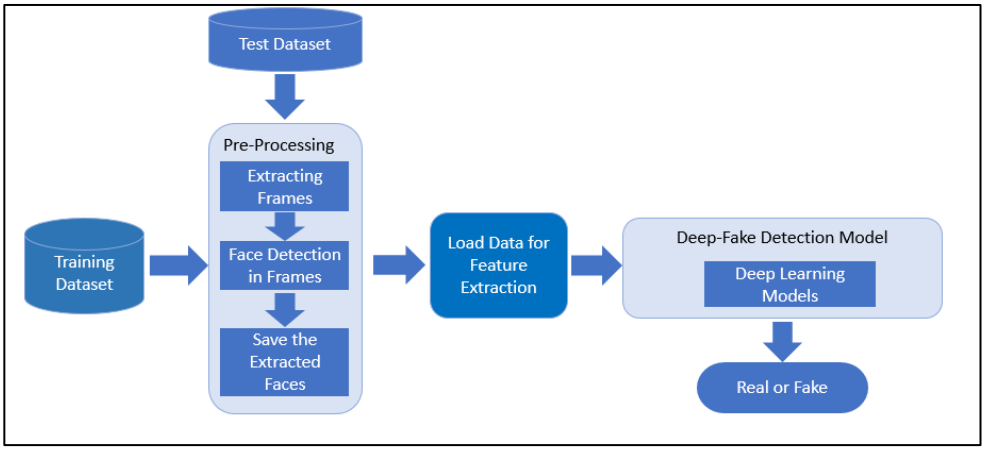

Figure 1. Workflow of the Proposed Method

3.1 Pre-Processing

Preprocessing in this study entails taking frames out of input videos and employing the Haarcascade process to identify facial regions in each frame to extract the region of interest, which is then fed into the feature extraction process to identify the deepfake videos. The publicly accessible dataset used in this experiment came from Kaggle [26]. It includes 106 brief video clips that are separated into two primary groups: authentic and fraudulent.

3.2 ResNet50

Residual Network with 50 layers, was first presented by He et al. in 2015. It was created to solve the vanishing gradient issue that prevents extremely deep neural networks from being trained. By adding residual connections, ResNet50 makes training more efficient and allows for the creation of considerably deeper networks. These connections let the whole network learn residual mappings. There are five parts to the ResNet50 design, and each one has convolutional layers and identity blocks. One of the best things about it is that it leverages residual connections, which let the input data of a block travel straight to its result without going through any extra layers. This quick fix gets rid of the issue of the gradient fading in deep networks. It makes gradients flow easier when they go back [27]. It also uses identity blocks, which add the input data to the output immediately after it has passed through a few convolutional layers. These blocks make it easier to train deeper networks Filling up the gaps between the blocks lets the system learn leftover functions instead of direct translations. The bottleneck design is another key aspect. It uses 1×1 and 3×3 convolutions combined to keep the number of variables and the cost of computation lowwhile still being able to represent features well. ResNet50 does an excellent job at identifying photographs, regardless of what is in the ImageNet dataset, due to its deep and effective design.

3.2.1 Application in Deepfake Video Detection

Because ResNet50 is good at extracting features, it is a useful tool for finding deepfake videos. It can discover subtle flaws that often occur in deepfake media, like strange facial expressions, lip-sync that doesn't match, or eye movements that don't make sense. Because of this, it is the best approach to find fake videos. One of the advantages of using ResNet50 is that it allows for transfer learning, which means you can take a model that has previously been trained on large datasets of images like ImageNet and then fine-tune it on deepfake datasets. This enables the model to identify small mistakes like warping faces or unusual lighting patterns [28]. This improves performance and educes training time, especially when data is limited. Additionally, models based on ResNet50 have regularly shown great accuracy in fake identification tests, outperforming several baseline designs. Adding ensemble methods and attention mechanisms has further enhanced its accuracy and reliability in detecting deepfakes.

3.3 EfficientNetB0

In 2019, Tan and Le showed off EfficientNetB0, the first version of the Efficient Network family. Unlike traditional CNN architectures, this one uses a revolutionary compound scaling method that improves the network's depth, width, and resolution to give better performance with fewer parameters. Because it is efficient, it is a perfect choice for apps that need both high accuracy and low processing power.

This extremely optimized CNN maintains an equilibrium between precision and effectiveness thanks to a number of improvements in architecture. One of its best features is that it uses a principled method to continuously increase the depth, width, and quality of a network. This creates a better model than traditional scaling methods that only consider one dimension [29]. Squeeze-and-Excitation (SE) blocks, which are integrated into the architecture, enhance the network's ability to represent features by automatically recalibrating feature maps, emphasizing useful channels, and hiding less useful ones [30]. Additionally, it uses the Swish activation function, which is a smooth, non-monotonic function that improves optimization

outcomes and gradient flow compared to ReLU [31]. It also uses flipped residual connections based on MobileNetV2's inverted design, to facilitate computation without losing accuracy. These linkages make it possible to build deeper networks without raising processing costs. This makes EfficientNetB0 ideal for situations where resources are constrained. Because of these improvements, EfficientNetB0 is especially good for real-time applications like deepfake video recognition, where speed and performance are very important. It can achieve state-of-the-art accuracy with significantly less parameters and FLOPS than traditional CNNs.

3.3.1 Application in Deepfake Video Detection

EfficientNetB0 has been shown to be effective at finding deepfake videos since it is lightweight and can extract fine-grained data very well. One of its key benefits is that it uses SE blocks to assist the model in focusing on important indicators like blurry edges, uneven lighting, and unusual facial expressions, which are all common faults in manipulated videos [32]. The model's compound scaling method helps it generalize better across different datasets and balance dimensions and source resolution. Due to this, EfficientNetB0 can counter various methods of creating deepfakes, which often have distinct styles and levels of quality [29]. Another significant advantage is that inference is faster. It can be utilized for security-sensitive situations like moderating social media content and monitoring live video because it is quick and can detect issues in real time. Even though it's modest, it outperforms many larger CNN designs in terms of the accuracy-to-parameter ratio, which means it can find deepfakes very accurately. These properties make it a reliable choice for deepfake detection systems when performance and speed is crucial.

This study uses a combination of ResNet50 and EfficientNetB0 as simultaneous feature extractors. Figure 2 shows the structure of this merged model. To extract the main features from input images, each model is initially established and then used independently. ResNet50 is a 50-layer deep residual network renowned for its ability to capture complex spatial features by solving the vanishing gradient problem with residual connections. On the other hand, EfficientNetB0 is a small but powerful model that employs compound scaling to balance depth, width, and resolution. It provides outstanding accuracy with a smaller number of parameters. This strategy utilizes only the feature extraction layers from both models and omits the classification layers. We create a composite feature vector by combining the output vectors of features from ResNet50 (2048 features) and EfficientNetB0 (1280 features). This merged output is passed through two dense layers, each with 128 neurons and ReLU activation, for binary classification. Finally, a dense layer with softmax is added. The model enhances deepfake identification accuracy and useful by combining the best aspects of both architectures to provide a rich and diversified representation of the input data.

```
Model: "ResNet50_EfficientNetB0"
_________________________________________________________________________________________________
 Layer (type)                Output Shape              Param #     Connected to
=================================================================================================
 input_1 (InputLayer)        [(None, 100, 100, 3       0           []
                             )]

 resnet50 (Functional)       (None, 2048)              23587712    ['input_1[0][0]']

 efficientnetb0 (Functional)  (None, 1280)             4049571     ['input_1[0][0]']

 concatenate_1 (Concatenate)  (None, 3328)             0           ['resnet50[0][0]',
                                                                    'efficientnetb0[0][0]']

 dense (Dense)               (None, 128)               426112      ['concatenate_1[0][0]']

 dense_1 (Dense)             (None, 128)               16512       ['dense[0][0]']

 dense_2 (Dense)             (None, 2)                 258         ['dense_1[0][0]']

=================================================================================================
Total params: 28,080,165
Trainable params: 442,882
Non-trainable params: 27,637,283
```

Figure 2. Architecture of Combined Model (ResNet50_EfficientNetB0)

4. Results and Discussion

4.1 Dataset Description

We used the Kaggle dataset [26] for our experiments. The dataset comprises 2 classes of short video clips. One class contains 53 fake videos that have been generated by deep fake experiments. The other r class contains 53 genuine video clips of people. The train-test split is in the ratio of 70:30. Each video is converted into 30 frames using the OpenCV package. These frames are then used for training. The HAAR cascade algorithm is used to detect the face region.

4.2 Performance Metrics

The model's performance is quantified using the measures namely accuracy, precision, recall, and F1 score, which prove the efficiency of the proposed model of the study. The equations for the above metrics are given in Eq. (1)- (4). Here, TP, TN, FP, and FN indicate True/False Positives/Negatives.

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \tag{1}$$

$$\text{Precision} = \frac{TP}{TP + FP} \tag{2}$$

$$\text{Recall} = \frac{TP}{TP + FN} \tag{3}$$

$$F1\_score = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \tag{4}$$

4.3 Experimental Results and Discussion

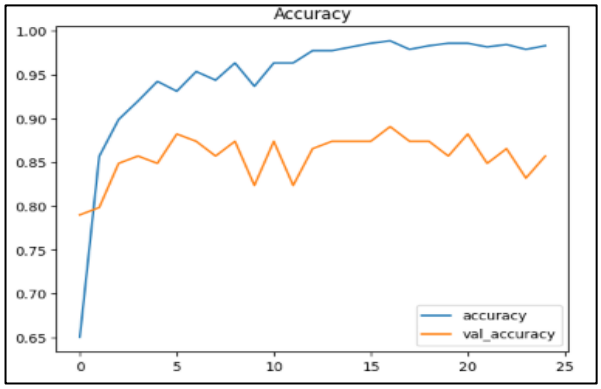

Figure 3. Accuracy of Training and validation of ResNet50

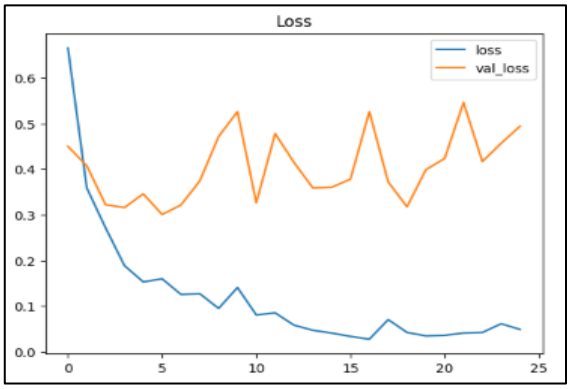

Figure 4. Loss of Training and Validation of ResNet50

```
print("    Test Loss: {:.5f}".format(results[0]))
print("Test Accuracy: {:.2f}%".format(results[1] * 100))
```

```
    Test Loss: 0.62172
Test Accuracy: 86.55%
```

Figure 5. Test Loss and Accuracy of ResNet50

Table 2. Confusion Matrix of ResNet50

	precision	recall	f1-score	support
fake	0.91	0.82	0.86	60
real	0.83	0.92	0.87	59
accuracy			0.87	119
macro avg	0.87	0.87	0.87	119
weighted avg	0.87	0.87	0.87	119

4.3.1 ResNet50

The proposed work demonstrates significant results in detecting transformed video using deep learning based on ResNet50. The training phase has achieved 95% accuracy, and validation has reached around 85%. The accuracy graph shown in Figure 3-5 demonstrates the model's strong learning ability up to 25 epochs. The consistent decline of loss in training and validation, shown in Figure 3-5, represents the model's convergence and generalization ability well. During the test phase, the overall performance is recorded as 87% accuracy and a loss of 0.62. For individual class 91%, 82% and 86% of precision, recall and F1-score respectively obtained for deepfake classes and 83%, 92% and 87% of precision, recall and F1-score, respectively, were obtained for th genuine class, as shown in Table 2. The weighted and micro

values are recorded as 87%, indicating the model's classification balance. With these results, it can be certified that the proposed ResNet50 model is reliable method for disclosing deepfakes.

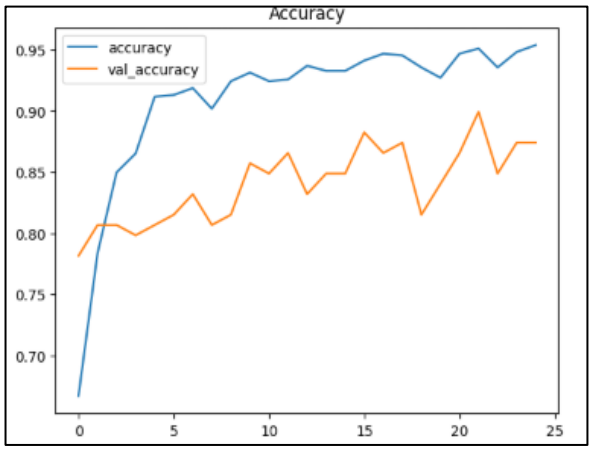

Figure 6. Accuracy of Training and Validation of EfficientNetB0

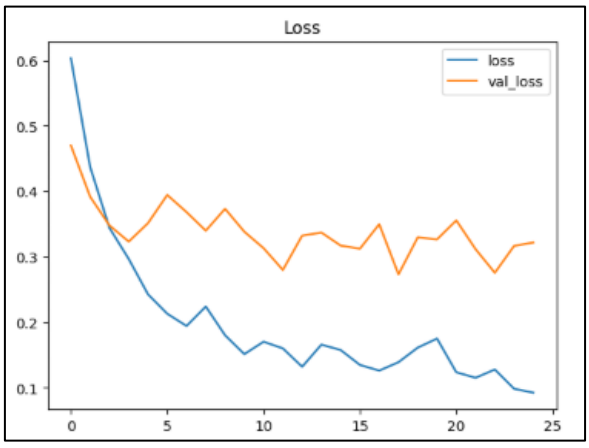

Figure 7. Loss of Training and Validation of EfficientNetB0

```
print("    Test Loss: {:.5f}".format(results[0]))
print("Test Accuracy: {:.2f}%".format(results[1] * 100))
```

```
    Test Loss: 0.29973
Test Accuracy: 88.24%
```

Figure 8. Test Loss and Accuracy of EfficientNetB0

Table 3. Confusion Matrix of EfficientNetB0

	precision	recall	f1-score	support
fake	0.87	0.90	0.89	60
real	0.89	0.86	0.88	59
accuracy			0.88	119
macro avg	0.88	0.88	0.88	119
weighted avg	0.88	0.88	0.88	119

4.3.2 EfficientNetB0

An experiment with EfficientNet50, shows that the model has performed admirably, attaining areliability of 88.24% and a loss of 0.29973. The accuracy graph shown in Figure 6-8 demonstrates the model's strong generalization over 25 epochs. With an overall accuracy of 88%, the proposed method obtained 87% and 90% of precision and recall for classes, and 89% and 86% precision and recall for real video classes, as detailed in Table 3. From this, it can be judged that the proposed work is a consistent method for detecting deepfakes.

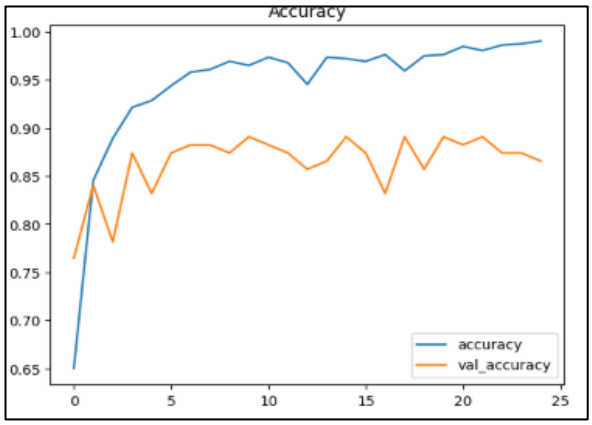

Figure 9. Accuracy of Training and Validation of Combined Model

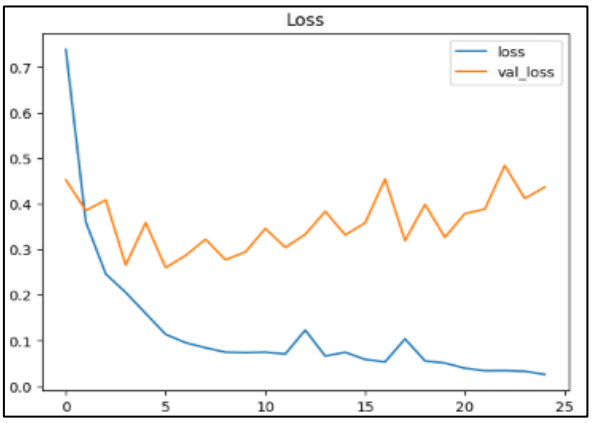

Figure 10. Loss of Training and Validation of Combined Model

```
print("    Test Loss: {:.5f}".format(results[0]))
print("Test Accuracy: {:.2f}%".format(results[1] * 100))
```

```
    Test Loss: 0.52021
Test Accuracy: 89.08%
```

Figure 11. Test Loss and Accuracy of Combined Model

Table 4. Confusion Matrix of Combined Model

	precision	recall	f1-score	support
fake	0.91	0.87	0.89	60
real	0.87	0.92	0.89	59
accuracy			0.89	119
macro avg	0.89	0.89	0.89	119
weighted avg	0.89	0.89	0.89	119

4.3.3 Combination of ResNet50 and EfficientNetB0

As seen in Figure 11, the combined model performed well in detecting deepfake videos, attaining an evaluation accuracy of 89.08% and loss of 0.52021. While the loss curves (Figure 10) show efficient convergence with little overfitting, the training and validation accuracy curves (Figure 9) show steady learning progress. The model obtained high precision as well as recall for both false (0.91, 0.87) and real (0.87, 0.92) categories, yielding an overall F1-score of 0.89, as indicated by the confusion matrix shown in Table 4. The performance of the proposed hybrid work is reached 89% in precision, recall, and F1-scores of macro and weighted averages, confirming the proposed model's balanced classification efficiency. According to the

macro and weighted average F1-scores of 0.89, the combined model outperforms the other two models in terms of balanced classification performance and overall test accuracy (89.08%). The Combined model performed better than both ResNet50 and EfficientNetB0 in terms of precision-recall balance and F1-scores for both classes, even though EfficientNetB0 had the lowest test loss (0.29973). Hence, it ensures that the combined model can be an efficient one for identifying deepfakes among original videos.

The output produced by the proposed model is shown in Figures 12-17, which pictorially represent how the model accepts incoming videos, performs frame conversion, anticipates class labels, and makes decisions on the received video as real or deepfake. These visualizations confirm the proposed method's ability to disclose deepfake videos, and they also aid in understanding practical deployment.

Figure 12. GUI Front Page - 1

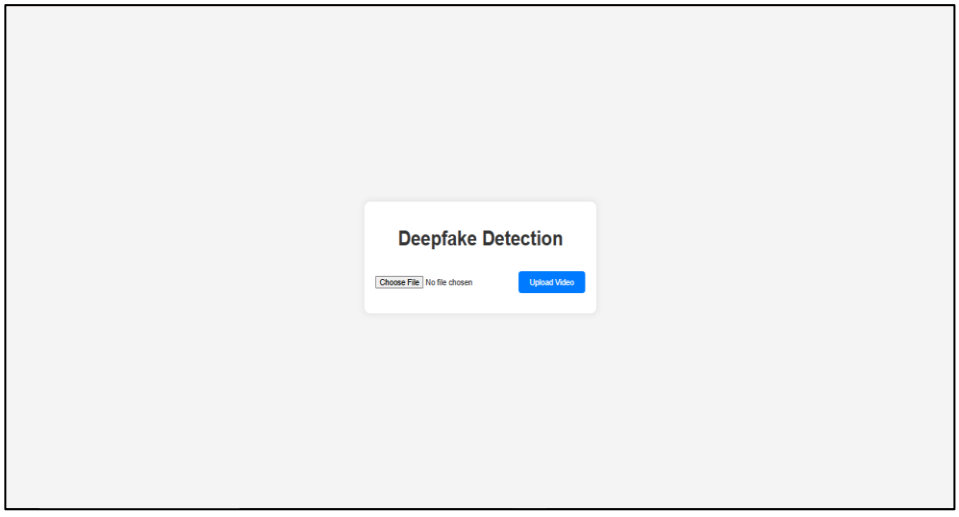

Figure 13. GUI Front Page - 2

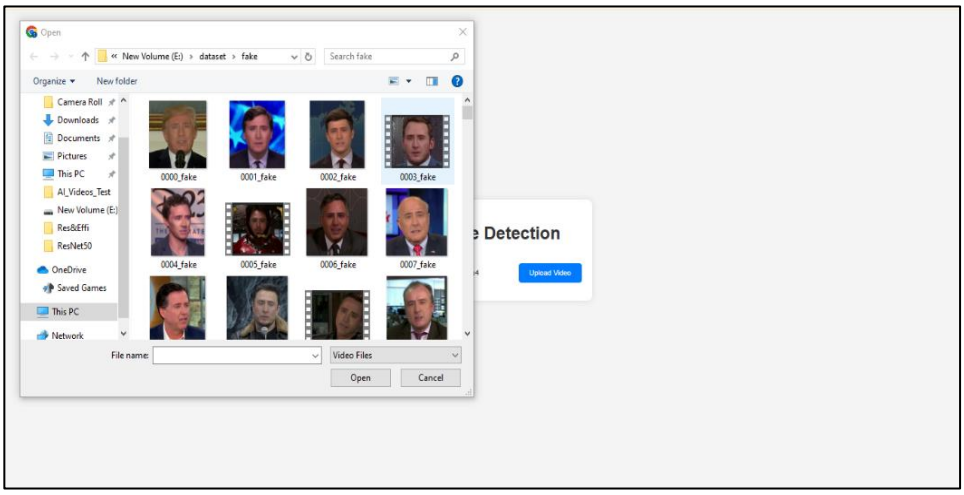

Figure 14. Choosing Video to Upload

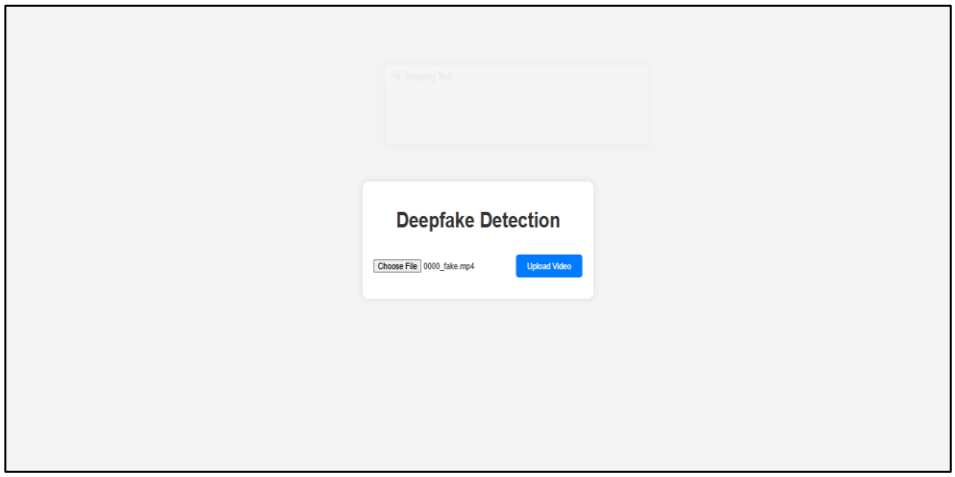

Figure 15. Feeding Input

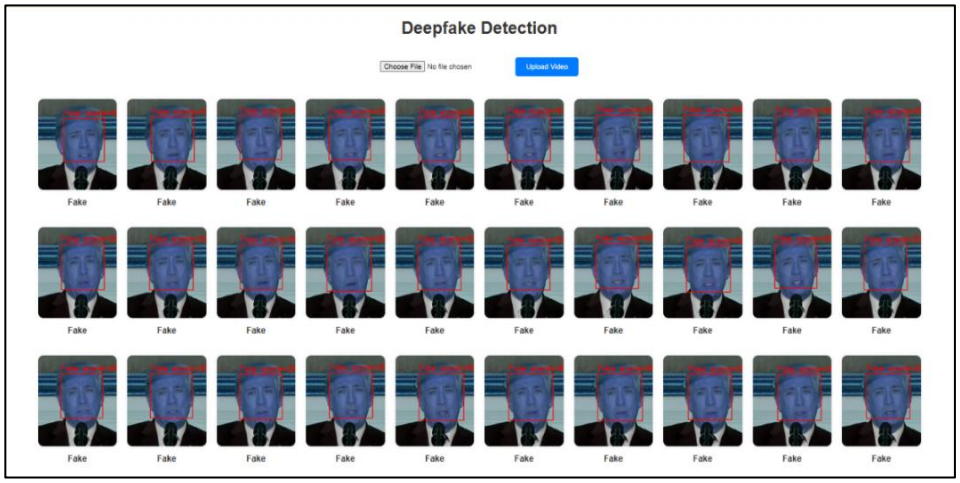

Figure 16. Frames Detected as Fake

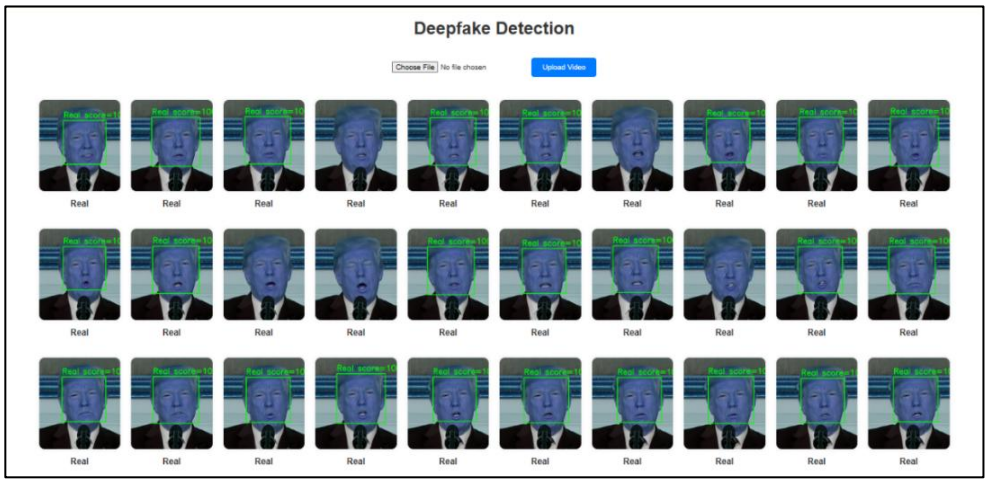

Figure 17. Frames Detected as Real

4.4 Comparison of Results with Other Models

Table 5. Comparison of Results with Other Methods

Source / References	Methods used	Accuracy obtained in %
Proposed work	ResNet50 + EfficientNetB0	89.08
Fatima and Ram [20] 2024	GAN-based Model	88.00
Saini et al., [21] 2024	CNN + LSTM	86.00
Dinçer et al., [22] 2024	CapsuleNet/ArCapsNet	82.84
Kandari et al.,[23] 2024	CNN-MLP Model	81.25
Kaur [25] 2023	LSTM + XResNet	83.30

Table 5 presents the comparative study of the proposed method with the state-of-the-art. The GAN-based model recorded 88% accuracy, the method using CapsuleNet/ArCapsNet obtained 82.84%, while the hybrid CNN with LSTM achieved 86%. LSTM and XResNet received 83.3% accuracy, while the model using CNN-MLP obtained 81.25%. Meanwhile, our proposed combined method achieved 89.08% accuracy, which is superior to all other methods considered in this study.

5. Conclusion

The proposed method has utilized the ResNet50 and EfficientNetB0 algorithms to identify deepfake videos. The system is validated using an open dataset from Kaggle. The performance of the system is quantitatively measured in terms of accuracy, recall, precision, and F1-score. Accuracies of 86.55% and 88.24% are recorded for the individual models, whereas 89.08% is recorded with the hybrid model of the above algorithms, which is significantly higher than the individual models. A comparative study is also conducted, and it provides evidence that the proposed hybrid model outperforms the other models in the state-of-the-art taken for this study.

References

[1] Chesney, Robert, and Danielle Citron. "Deep Fakes: A Looming Challenge for Privacy, Democracy, and National Security." California Law Review 107, no. 6 (2019): 1753–1820. https://doi.org/10.2139/ssrn.3213954.

[2] Deeptrace. The State of Deepfakes 2020. (2020). https://regmedia.co.uk/2020/10/08/deeptrace-deepfake-report.pdf.

[3] Afchar, Darius, Vincent Nozick, Junichi Yamagishi, and Isao Echizen. "MesoNet: A Compact Facial Video Forgery Detection Network." In 2018 IEEE International Workshop on Information Forensics and Security (WIFS) (2018): 1–7. IEEE. https://doi.org/10.1109/WIFS.2018.8630761.

[4] Matern, Falko, Christian Riess, and Marc Stamminger. "Exploiting Visual Artifacts to Expose Deepfakes and Face Manipulations." In 2019 IEEE Winter Applications of Computer Vision Workshops (WACVW) (2019): 83–92. IEEE. https://doi.org/10.1109/WACVW.2019.00020.

[5] Rössler, Andreas, Davide Cozzolino, Luisa Verdoliva, Christian Riess, Justus Thies, and Matthias Nießner. "FaceForensics++: Learning to Detect Manipulated Facial Images." IEEE Transactions on Pattern Analysis and Machine Intelligence 43, no. 3 (2021): 803–817. https://doi.org/10.1109/TPAMI.2020.2979249.

[6] El-Gayar, Mohamed M., Maged Abouhawwash, Sherif S. Askar, and Shadi Sweidan. "A Novel Approach for Detecting Deep Fake Videos Using Graph Neural Network." Journal of Big Data 11, no. 1 (2024): 1–20. https://doi.org/10.1186/s40537-024-00884-y.

[7] Vardhan, Harsha. "Deep Fake Video Detection." IRJAEH 2, no. 4 (2024): 830–835. https://doi.org/10.47392/IRJAEH.2024.0117.

[8] Salman, Sana, Junaid A. Shamsi, and Rashid Qureshi. "Deep Fake Generation and Detection: Issues, Challenges, and Solutions." IT Professional 25, no. 1 (2023): 6–14. https://doi.org/10.1109/MITP.2022.3230353.

[9] C. L. S., A. P., C. T. S., G. B. R., and G. D. "Progress in Deep Fake and Tampering: An In-Depth Analysis." International Journal of Scientific Research in Engineering and Management 9, no. 1 (2025): 1–5. https://doi.org/10.55041/IJSREM40707.

[10] Atas, Selçuk, İbrahim İlhan, and Mustafa Karakse. "An Efficient Deepfake Video Detection Approach with Combination of EfficientNet and Xception Models Using Deep Learning." In Proceedings of the 26th International Conference on Information Technology (IT) (2022): 1–8. IEEE. https://doi.org/10.1109/IT54280.2022.9743542.

[11] Purevsuren, Enkhbold, Jun Sato, and Takayuki Akashi. "A Comparative Analysis of Deepfake Detection Methods Using Overlapping Multiple Dynamic Images." IEEJ Transactions on Electrical and Electronic Engineering 18, no. 1 (2025): 1–8. https://doi.org/10.1002/tee.24258.

[12] Pote, Rupali, S. G. Karokar, P. G. Tandekar, S. M. Nawle, M. S. Lade, and R. R. Tonge. "Deep Fake Detecting System." International Journal of Research in Applied Science and

Engineering Technology 12, no. 4 (2024): 1579–1582. https://doi.org/10.22214/ijraset.2024.59748.

[13] Doke, Yogita, P. Dongare, M. Gaikwad, M. Gaikwad, and V. Marathe. "Deep Fake Detection Through Deep Learning." International Journal of Research in Applied Science and Engineering Technology 11, no. 5 (2023): 861–866. https://doi.org/10.22214/ijraset.2023.51630.

[14] Al-Adwan, Ahmad, Hani Alazzam, Nour Al-Anbaki, and Eman Alduweib. "Detection of Deepfake Media Using a Hybrid CNN–RNN Model and Particle Swarm Optimization (PSO) Algorithm." Computers 13, no. 4 (2024): 99. https://doi.org/10.3390/computers13040099.

[15] Reddy, B. C. K., A. S. Reddy, Y. B. Reddy, M. S. Begam, and B. M. M. Reddy. "Deep Fake Face Detection Using Deep Learning Tech with LSTM." International Journal of Scientific Research in Engineering and Management 7, no. 1 (2025): 1–5. https://doi.org/10.55041/IJSREM28624.

[16] Anand, R., L. Santhosh, A. K. N., V. Potdar, R. Kumar, R. Raj, and S. Sharma. "Video Authenticity Detection Using Web-Enabled Techniques." International Journal of Research in Applied Science and Engineering Technology 12, no. 5 (2024): 5617–5623. https://doi.org/10.22214/ijraset.2024.62881.

[17] Barbadekar, Ashwini, Swarali Sole, and Akash Shekhavat. "Enhancing Social Media Security: LSTM-Based Deep Fake Video Detection." In Proceedings of the 2024 IEEE International Conference for Convergence in Technology (I2CT) (2024): 1–6. IEEE. https://doi.org/10.1109/I2CT61223.2024.10543604.

[18] Talreja, Shibani, Anshul Bindle, Vishal Kumar, Ishan Budhiraja, and Prateek Bhattacharya. "Security Strengthen and Detection of Deepfake Videos and Images Using Deep Learning Techniques." In 2024 IEEE International Conference on Communications Workshops (ICC Workshops) (2024): 1834–1839. IEEE. https://doi.org/10.1109/ICCWorkshops59551.2024.10615811.

[19] Saadi, Mohammad, and Waleed A. M. Al-Jawher. "Proposed DeepFake Detection Method Using Multiwavelet Transform." International Journal of Innovative Computing 13, no. 1–2 (2023): 61–66. https://doi.org/10.11113/ijic.v13n1-2.420.

[20] Fatima, A., and P. K. Ram. "GAN-Enhanced Real-Time Detection of Deepfakes Videos." Journal of Artificial Intelligence and Capsule Networks 6, no. 4 (2024): 452–465. https://doi.org/10.36548/jaicn.2024.4.005.

[21] Saini, M. L., Deepak Chandra, A. Patnaik, and R. Kumar. "Deepfake Detection System Using Deep Neural Networks." In Proceedings of IC457434 (2024): 1–5. IEEE. https://doi.org/10.1109/ic457434.2024.10486659.

[22] Dinçer, Sedat, Gökhan Ulutaş, Burak Üstübioğlu, Gökçe Tahaoğlu, and Nicolas Sklavos. "Golden Ratio Based Deep Fake Video Detection System with Fusion of Capsule Networks." Computers & Electrical Engineering (2024). https://doi.org/10.1016/j.compeleceng.2024.109234.

[23] Kandari, M., V. Tripathi, B. Pant, A. Sar, T. Choudhury, and T. Choudhury. "Detecting Deepfake Videos Through CNN-MLP Model in Media Forensics." In Proceedings of the OTCON 2024 (2024): 1–7. IEEE. https://doi.org/10.1109/otcon60325.2024.10687433.

[24] Arshed, Muhammad Asad, Ayed Alwadain, Rao Faizan Ali, Shahzad Mumtaz, Muhammad Ibrahim, and Amgad Muneer. "Unmasking deception: empowering deepfake detection with vision transformer network." Mathematics 11, no. 17 (2023): 3710.

[25] Kaur, Amanpreet, Sandeep Sharma, and Amanpreet Kaur. "Detection of Deepfake Images Using Machine Learning." International Journal of Advanced Research in Computer and Communication Engineering 12, no. 2 (2023): 10–14. https://doi.org/10.17148/IJARCCE.2023.12203.

[26] Zhang, Han, Pengcheng Guo, Shuyang Wang, and Liangliang Cao. "Multi-Modal Deepfake Detection via Spatial and Temporal Attention." In Proceedings of the 30th ACM International Conference on Multimedia (2022): 1132–1140. ACM. https://doi.org/10.1145/3503161.3548277.

[27] Ismail, Md, Shah Md Labib Shadik, Md Sajid Ullah Sohan, M. D. Bhuiyan, and Lamia Khan Shoily. "An efficient approach for deepfake detection employing micro-expressions with a hierarchical transformer network." PhD diss., Brac University, 2025.

[28] Li, Yuezun, Ming-Ching Chang, and Siwei Lyu. "In Ictu Oculi: Exposing AI Created Fake Videos by Detecting Eye Blinking." In 2018 IEEE International Workshop on Information Forensics and Security (WIFS) (2018): 1–7. IEEE. https://doi.org/10.1109/WIFS.2018.8630787.

[29] Agarwal, Shruti, Hany Farid, Yuming Gu, Mingming He, Koki Nagano, and Hao Li. "Protecting World Leaders Against Deep Fakes." In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) Workshops (2019): 38–45. https://openaccess.thecvf.com/content_CVPRW_2019/html/Media%20Forensics/Agarwal_Protecting_World_Leaders_Against_Deep_Fakes_CVPRW_2019_paper.html.

[30] Masi, Iacopo, Anh Tuan Tran, Tal Hassner, Jatuporn Toy Leksut, and Gerard Medioni. "Do We Really Need to Collect Millions of Faces for Effective Face Recognition?" In Proceedings of the European Conference on Computer Vision (ECCV) (2016): 579–596. Springer. https://doi.org/10.1007/978-3-319-46478-7_36.

[31] Korshunov, Pavel, and Sébastien Marcel. "Deepfakes: A New Threat to Face Recognition? Assessment and Detection." arXiv preprint arXiv:1812.08685 (2018). https://arxiv.org/abs/1812.08685.

[32] Mirsky, Yisroel, and Wenke Lee. "The Creation and Detection of Deepfakes: A Survey." ACM Computing Surveys 54, no. 1 (2021): 1–41. https://doi.org/10.1145/3425780.