Enhanced deepfake detection using an ensemble of convolutional neural networks

Yeeshu Ralhen, Sharad Sharma
Department of Electronics and Communication Engineering, Maharishi Markandeshwar (Deemed to be University), Ambala, India

Article Info

Article history:

Received Jan 5, 2025
Revised Jun 28, 2025
Accepted Jul 13, 2025

Keywords:

Convolutional neural networks
Deep learning
Deepfake detection
Generative adversarial networks
Image and video manipulation

ABSTRACT

Digital media integrity and authenticity have been seriously challenged with the rise of deepfakes. The challenge is to automatically detect this artificial intelligence (AI) generated manipulations. These manipulations or forgeries can cause harmful consequences such as spreading fake news in politics, scamming people online and invading privacy. Convolutional neural networks (CNN) models are found to be good at classification tasks, but the performance could not reach high accuracy, especially when they were tested on more challenging deepfake datasets. In this paper we present a deepfake detection system based on an ensemble of CNN architectures, ResNet50 and EfficientNet, capable of distinguishing between real and deepfake videos with high accuracy. For the experiment, we have chosen Celeb-DF version 2, as it has emerged to be one of the most challenging deepfake dataset. The ensemble model achieved an F1-score of 94.69% and an accuracy of 90.58%, outperforming the individual CNN models. This study shows that ensemble learning can increase the reliability and accuracy of deepfake detection systems on challenging datasets.

Corresponding Author:

Yeeshu Ralhen
Department of Electronics and Communication Engineering
Maharishi Markandeshwar (Deemed to be University)
Ambala-Yamunanagar Highway, Mullana, Ambala, India
Email: yeeshu.relhan@gmail.com

1. INTRODUCTION

The recent developments in artificial intelligence (AI), particularly in the area of synthetic media generation, also known as deepfakes, have made it increasingly accessible for anyone to perform facial re-enactment, i.e., the transfer of facial expressions from one video to another [1]. The quality and availability of facial manipulation systems and tools have reached a point where they are easily usable by people without expertise in the field or skills in image manipulation. In fact, an increasing number of code and library resources that work quasi-automatically are becoming publicly available [2], [3]. While these advances open up new opportunities for professionals working in film production, visual effects production or digital art creation areas of industry, they also enable people with malicious intent to create digital forgeries which can propagate incorrect information or harm individuals' public persona. Unreliable deepfake content has caused concerns on whether it is possible to trust any kind of digital content and how such content could impact individuals' lives.

In recent years, the application of deep learning has led to substantial improvements in computer vision tasks. Generative adversarial networks (GAN) [4] can achieve high-quality results and their application in image and video generation has become a research hotspot. Autoencoder decoder and GANs are two of the most used methods to generate deepfake. Autoencoders [5] plays an important role in generation of deepfake by encoding facial features from input images and then reconstructing them in a modified form. The encoder compresses the facial data and decoder reconstructs it, often altering expressions or identities in the process. This ability to manipulate and regenerate faces makes autoencoders highly effective in creating realistic deepfake images and videos. GANs is another popular method for generating deepfake content, which contains a generator and discriminator. The role of generator is to transform random noise into real image or videos, whereas discriminator predicts if the generated image is real or fake which help generator improve its generated image based on feedback from discriminator. GANs can also create human-like realistic images with expressions.

The increasing sophistication of deepfake technology has made detecting AI-generated manipulations a critical challenge. Deepfakes are being widely misused for spreading misinformation, financial scams, political propaganda, and privacy violations, posing serious ethical and security threats. Although convolutional neural networks (CNNs) have emerged as powerful tools in image and video classification, single-model detection approaches frequently struggle with generalization and performance, especially when faced with more complex datasets like Celeb-DF version 2 [6]. This motivates the need for a more robust and reliable detection approach. Figure 1 presents several examples of faces extracted from the Celeb DF version 2 dataset, showcasing both original and manipulated samples.

To address these challenges, our study proposes an ensemble of CNNs (ResNet50 and EfficientNet) to enhance deepfake detection accuracy. The key contributions of this work are summarized as: i) proposed a deepfake detection system utilizing an ensemble of CNN architectures to improve classification accuracy, ii) demonstrated that the ensemble model outperforms individual CNN models by leveraging the strengths of both architectures to capture diverse spatial features, achieving an accuracy of 90.58% and an F1-score of 94.69% on the challenging Celeb-DF version 2 dataset, and iii) outperformed state of the art deepfake detection models, including MesoNet and video vision transformers.

Figure 1. Faces taken from the Celeb DF version 2 dataset, showing each original face along with its corresponding generated fake version

2. RELATED WORK

Guera and Delp [7] proposed another method based on analyzing temporal inconsistencies of facial movements in a video. One early approach proposed by Li *et al.* [8] utilized blinking patterns to detect deepfakes. Since deepfake videos often undergo multiple stages of manipulation and compression, these processes can leave behind unique patterns. To directly focus on face manipulation based deepfakes, Afchar *et al.* [9] introduced a new CNN-based model called MesoNet. This lightweight CNN architecture extracts mid-level features that are particularly effective when trying to expose subtle face manipulations in certain regions through deepfake videos. In addition to image-based detection, researchers soon realized the importance of incorporating temporal information for effective deepfake detection. Sabir *et al.* [10] proposed a recurrent CNN architecture that leveraged both spatial and temporal information. An entirely distinct approach was employed by Hashmi *et al.* [11] they created a pipeline where 512 facial landmarks were first extracted per frame and analyzed along with minor facial actions like the position of eyebrows and lip syncing using a combination of CNN and recurrent neural network (RNN) achieving highest accuracy among all other methods for distinguishing between real and fake across various datasets.

As deepfake technology continued to evolve, detection methods also became more sophisticated, incorporating advanced techniques such as attention mechanisms, multi-modal detection, and ensemble learning. Dang *et al.* [12] first proposed a detection model of deepfakes that exploits the attention mechanism focusing on areas more likely to be manipulated such as facial regions and textures. Other methods working in multimodal-based characteristics have been developed to detect deepfake also combining audio and visual information. Table 1 provides a quick summary of everything by listing the datasets, features and models utilised throughout the research with a reference to each of the relevant studies.

Table 1. Deepfake detection methods

Year	Study	Models	Features	Dataset used
2018	Detection using RNN [7]	CNN, RNN	Spatio-temporal consistency	Other
2019	Recurrent convolutional strategies [10]	CNN	Face landmarks	FF++
2019	Image segmentation and separable CNN [13]	CNN	Special artifacts	FF++
2020	Enhanced MesoNet [14]	CNN	Mesoscopic features	FF++
2020	Spatio-temporal features [15]	3D CNN	Spatio-temporal consistency	DFDC, FF++
2020	Eyebrow recognition [16]	CNN	Visual Artifacts	Celeb-DF
2020	Self-consistency learning [17]	CNN	Face landmarks, intra-frame inconsistency	DFDC, FF++, Celeb-DF, DFD
2021	Discrepancies between faces and their context [18]	CNN	Special artifacts	FF++, Celeb-DF, DFDC
2022	Combining efficientnet and vision transformers [19]	MTCNN, vision transformers	Special artifacts	DFDC, FF++
2022	Transformer-based feature compensation [20]	Vision transformers	Intra-frame inconsistency	FF++, Celeb-DF, UADFV
2022	Intra-consistency and inter-diversity [21]	CNN, vision transformers	Intra-frame inconsistency	DF-TIMIT, DFDC, UADFV, FF++/DF, Celeb-DF
2023	ISTVT [22]	MTCNN	Face landmarks	FF++, DF, Celeb-DF, DFDC
2023	Improved dense CNN [23]	CNN	Special artifacts	Celeb-A, FF++
2024	Video vision transformers [24]	CNN, vision transformers	Face landmarks	Celeb-DF v2
2024	Extractor based on vision transformer [25]	Vision transformers	Intra-frame inconsistency	Celeb-DF v2, DFDC, WildDeepfake, FF++
2024	Eye movement analysis [26]	CNN	Special artifacts	FF++, Celeb-DF v2

Matern *et al*. [27] developed a system that identifies mismatches between lip movements and speech audio. This approach takes advantage of the fact that many deepfake models struggle to accurately synchronize audio and visual information, especially in low-quality or heavily manipulated videos. Laghari *et al*. [28] proposed a deep learning-based approach for atrial fibrillation detection, leveraging residual networks and recurrent neural networks to enhance accuracy and feature extraction. Most recently, Ramadhani *et al*. [24] proposed a deepfake detection technique based on the video vision transformer architecture, using facial landmark areas as input. Their system showed promising results on the challenging Celeb-DF version 2 dataset, marking another significant advancement in deepfake detection research. Qazi and Ahmed [29] introduced a deepfake detection approach leveraging transfer learning and ensemble learning, combining multiple pre-trained models to improve the accuracy of identifying manipulated content.

3. METHOD

This section outlines the proposed workflow for deepfake detection, which consists of three main stages: i) face detection and extraction, ii) feature computation, and iii) prediction. Figure 2 illustrates the proposed model architecture for deepfake detection, which uses ResNet50 and EfficientNetB0 for feature extraction and dense layers for final classification.

3.1. Face detection and extraction

In order to reduce computational cost, 20 frames from each video were considered for analysis instead of analyzing the full video. The first and foremost step is to recognize and extract facial regions from the frames since face is the sole region being manipulated in deepfake videos. A face detector is employed to recognize and extract the facial region from the frames. In this paper, haar cascade classifier [30] was used for face detection in video sequences. Haar cascade classifier method is fast, efficient and particularly applied for real time face detection.

3.2. Feature computation

In the next step, features are computed from the extracted face images, which will serve as inputs for the classifier. We employed two state-of-the-art pre-trained CNN models, ResNet50 [31] and EfficientNetB0, for feature extraction. ResNet50 effectively learns deep representations through its residual learning framework, which mitigate the vanishing gradient problem and allow for effective training of very deep networks. EfficientNetB0, on the other hand, balances model size and accuracy through a compound scaling method, making it highly efficient while maintaining high performance on various tasks. Both models were pre-trained on the ImageNet dataset. The face images are resized to 224×224 before being passed through these models. Each CNN processes the same input images and outputs a feature vector. The feature vectors from both ResNet50 and EfficientNetB0 are then concatenated into a single combined vector.

Following this, three fully connected (dense) layers, each with LeakyReLU activation, are applied to the concatenated features.

Figure 2. The proposed ensemble models

3.3. Prediction

In the final stage, the system classifies the concatenated features as either real or fake. The classification is performed using a dense layer with 2 neurons (for binary classification), combined with a softmax activation function, which outputs the probability of the video being real or deepfake. This classification determines the authenticity of the video content.

In this study, we used the Celeb-DF version 2 dataset, recognized as one of the most challenging datasets for deepfake detection due to its complexity and the low performance of current detection methods on it. The dataset contains 5,639 deepfake videos and 890 real videos. To detect faces in each video, we applied the haar cascade classifier, selecting only 20 frames per video to reduce computational cost without sacrificing key information. For feature extraction, we leveraged two state of the art pre-trained CNNs, ResNet50 and EfficientNetB0, both trained on the ImageNet dataset. The detected face images were resized to 224×224 pixels before being processed by these models to maintain compatibility with their input requirements. We applied data augmentation techniques to the training set, including horizontal flipping, random rotation, and pixel value adjustments to enhance the model's generalization. The dataset was split into 75% for training and 25% for testing. Pseudocode 1 shows function that represents the pseudocode of deepfake detection.

Pseudocode 1. Deepfake detection
```
Function: deepfake_detection
Input: video_dataset, pretrained_model1, pretrained_model2, batch_size, epochs
Code:
Initialize model_1=load(pretrained_model1)   #Load first pretrained model
Initialize model_2=load(pretrained_model2)   #Load second pretrained model
Freeze initial layers of model_1
Freeze initial layers of model_2
Initialize training_data, validation_data, test_data=preprocess(video_dataset)
Initialize all elements of predicted_labels=0
Define input_layer=(image_shape)
feature_1=model_1(input_layer) #Extract features from model_1
feature_2=model_2(input_layer) #Extract features from model_2
combined_features=concatenate (feature_1, feature_2)
x=dense_layer (combined_features, activation='relu')
x=dropout (x, rate=0.5)
x=dense_layer (x, activation='relu')
output_layer=dense_layer (x, activation='softmax', units=2)   #Output for 2 classes
(Real/Fake)
model=create_model (input_layer, output_layer)
Compile model using Adam optimizer, loss='binary_crossentropy', metrics=['accuracy']
i=1
while i≤ epochs Repeat
    Train model using training_data
with batch_size
    Validate model on validation_data
    if validation_loss does not improve for 5 epochs then
        Stop training using early stopping
    End if
    i=i+1
End while
Evaluate model on test_data
Generate classification_report and confusion_matrix
Output: trained_deepfake_detection_model
```

The testing was conducted on an Ant PC Pheidole XE4216 system with 32 GB of RAM, 2 TB of SSD, and Nvidia 48 GB GPU, providing the computational power needed for efficient training and model evaluation. We employed the Adam optimizer to train the model for 50 epochs with a batch size of 32.

The softmax activation function was used at the output layer to classify image as real or fake. Accuracy and F1-score were used to evaluate the effectiveness of the model.

4. RESULTS AND DISCUSSION

The experimental results are presented in this section. Figure 3 illustrates the training and validation accuracy throughout the training progress. We observe that the training accuracy steadily converged to 100%, while the validation accuracy fluctuated between 65% and 90%, indicating some instability in generalization.

Next, we examine the performance of our ensemble model during testing. The confusion matrix for our ensemble model is shown in Figure 4. We can see that our ensemble model successfully recognized 25,165 deepfake images from 27,141 deepfake images in the testing set. Furthermore, 1,988 of the 2,835 real images in the testing set were recognized by our method. Our ensemble model demonstrated strong performance in the testing phase, achieving an F1-score of 94.69% and an accuracy of 90.58%. It also attained 92.72% precision and 96.74% recall.

The ensemble of ResNet50 and EfficientNet significantly outperformed the individual CNN models on the Celeb-DF version 2 dataset. Table 2 presents the performance comparison between individual models and the ensemble. The ensemble approach resulted in a 4.27% increase in accuracy compared to the best-performing single model (EfficientNet at 86.31%). This demonstrates the effectiveness of combining multiple CNN architectures to improve deepfake detection across challenging datasets. We also evaluated our system by comparing it with other deepfake detection systems, including MesoNet, ViViT and ViViT combined with differential scanning calorimetry (DSC) and carbon border adjustment mechanism (CBAM). The results shown in Table 3, indicate that our system outperformed the others in terms of accuracy. This improvement indicates that the features extracted by our model effectively enhance the overall performance of deepfake detection.

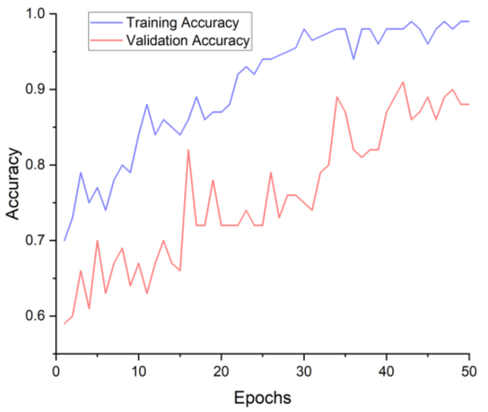

Figure 3. Accuracy of the training and validation sets during the training process

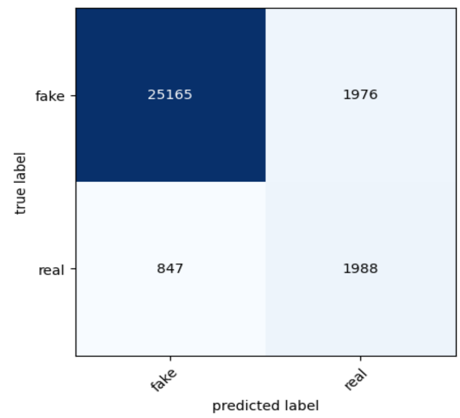

Figure 4. Confusion matrix of the proposed ensemble model

Table 2. Comparison between ensemble model and individual models

Model	Accuracy (%)
ResNet50	85.78
EfficientNet	86.31
Ensemble	90.58

Table 3. Comparison of ensemble model with other deepfake detection techniques

Model	Accuracy (%)	F1-score (%)
MesoNet [9]	65.12	72.83
ViViT [22]	52.14	63.6
ViViT+DSC+CBAM [24]	87.17	92.51
Ensemble	90.58	94.69

5. CONCLUSION

In this paper, we proposed an ensemble of CNNs for deepfake detection, leveraging the complementary strengths of ResNet50 and EfficientNet to improve accuracy and robustness. The ensemble model outperformed individual CNN models on the challenging Celeb-DF version 2 dataset, achieving an accuracy of 90.58% and outperforming other deepfake detection systems such as MesoNet, ViViT, and ViViT+DSC+CBAM. These results demonstrate that ensemble learning can substantially improve the effectiveness of deepfake detection systems, offering a more reliable solution for identifying manipulated media. Future work will focus on integrating temporal features and optimizing real-time detection to further advance the system's practical applications.

FUNDING INFORMATION

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

AUTHOR CONTRIBUTIONS STATEMENT

This journal uses the Contributor Roles Taxonomy (CRediT) to recognize individual author contributions, reduce authorship disputes, and facilitate collaboration.

Name of Author	C	M	So	Va	Fo	I	R	D	O	E	Vi	Su	P	Fu
Yeeshu Ralhen	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓		✓	
Sharad Sharma	✓	✓				✓			✓	✓		✓		

C : **C**onceptualization
M : **M**ethodology
So : **So**ftware
Va : **Va**lidation
Fo : **Fo**rmal analysis
I : **I**nvestigation
R : **R**esources
D : **D**ata Curation
O : Writing - **O**riginal Draft
E : Writing - Review & **E**diting
Vi : **Vi**sualization
Su : **Su**pervision
P : **P**roject administration
Fu : **Fu**nding acquisition

CONFLICT OF INTEREST STATEMENT

Authors state no conflict of interest.

DATA AVAILABILITY

The data that support the findings of this study are openly available on Kaggle platform at https://www.kaggle.com/datasets/reubensuju/celeb-df-v2, reference number [6].

REFERENCES

[1] J. Thies, M. Zollhöfer, and M. Nießner, "Deferred neural rendering: image synthesis using neural textures," *ACM Transactions on Graphics*, vol. 38, no. 4, pp. 1–12, Aug. 2019, doi: 10.1145/3306346.3323035.
[2] Deepfakes GitHub, "Faceswap," *GitHub*, Accessed Nov. 21, 2024. [Online]. Available: https://github.com/deepfakes/faceswap.
[3] M. Kowalski, "Faceswap," *GitHub*, Accessed Nov. 23, 2024. [Online]. Available: https://github.com/MarekKowalski/FaceSwap.
[4] I. Goodfellow *et al.*, "Generative adversarial networks," *Communications of the ACM*, vol. 63, no. 11, pp. 139–144, Oct. 2020, doi: 10.1145/3422622.
[5] V. Goyal, A. Yadav, S. Kumar, and R. Mukherjee, "Lightweight LAE for anomaly detection with sound-based architecture in smart poultry farm," *IEEE Internet of Things Journal*, vol. 11, no. 5, pp. 8199–8209, Mar. 2024, doi: 10.1109/JIOT.2023.3318298.
[6] Y. Li, X. Yang, P. Sun, H. Qi, and S. Lyu, "Celeb-DF: a large-scale challenging dataset for deepfake forensics," in *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, Jun. 2020, pp. 3204–3213, doi: 10.1109/CVPR42600.2020.00327.
[7] D. Guera and E. J. Delp, "Deepfake video detection using recurrent neural networks," in *Proceedings of AVSS 2018 - 2018 15th IEEE International Conference on Advanced Video and Signal-Based Surveillance*, Nov. 2018, pp. 1–6, doi: 10.1109/AVSS.2018.8639163.
[8] Y. Li, M. C. Chang, and S. Lyu, "In Ictu Oculi: exposing AI created fake videos by detecting eye blinking," in *10th IEEE International Workshop on Information Forensics and Security, WIFS 2018*, Dec. 2018, pp. 1–7, doi: 10.1109/WIFS.2018.8630787.
[9] D. Afchar, V. Nozick, J. Yamagishi, and I. Echizen, "MesoNet: a compact facial video forgery detection network," in *10th IEEE International Workshop on Information Forensics and Security, WIFS 2018*, Dec. 2018, pp. 1–7, doi: 10.1109/WIFS.2018.8630761.
[10] E. Sabir, J. Cheng, A. Jaiswal, W. AbdAlmageed, I. Masi, and P. Natarajan, "Recurrent convolutional strategies for face manipulation detection in videos," *arXiv-Computer Vision and Pattern Recognition*, pp. 1-8, 2019.
[11] M. F. Hashmi, B. K. K. Ashish, A. G. Keskar, N. D. Bokde, J. H. Yoon, and Z. W. Geem, "An exploratory analysis on visual counterfeits using conv-LSTM hybrid architecture," *IEEE Access*, vol. 8, pp. 101293–101308, 2020, doi: 10.1109/ACCESS.2020.2998330.
[12] H. Dang, F. Liu, J. Stehouwer, X. Liu, and A. K. Jain, "On the detection of digital face manipulation," in *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, Jun. 2020, pp. 5780–5789, doi: 10.1109/CVPR42600.2020.00582.
[13] C. -M. Yu, C. -T. Chang, and Y. -W. Ti, "Detecting deepfake-forged contents with separable convolutional neural network and image segmentation," *arXiv-Computer Science*, pp. 1-23, 2019.
[14] P. Kawa and P. Syga, "A note on deepfake detection with low-resources," *arXiv-Computer Science*, pp. 1-9, 2020. [Online]. Available: https://arxiv.org/abs/2006.05183.

[15] I. Ganiyusufoglu, L. M. Ngô, N. Savov, S. Karaoglu, and T. Gevers, “Spatio-temporal features for generalized detection of deepfake videos,” *arXiv-Computer Science*, pp. 1-11, 2020. [Online]. Available: https://arxiv.org/abs/2010.11844
[16] H. Nguyen and R. Derakhshani, “Eyebrow recognition for identifying deepfake videos,” *Lecture Notes in Informatics, Proceedings-Series of the Gesellschaft fur Informatik*, vol. P-306, pp. 199–206, 2020.
[17] T. Zhao, X. Xu, M. Xu, H. Ding, Y. Xiong, and W. Xia, “Learning self-consistency for deepfake detection,” in *Proceedings of the IEEE International Conference on Computer Vision*, Oct. 2021, pp. 15003–15013, doi: 10.1109/ICCV48922.2021.01475.
[18] Y. Nirkin, L. Wolf, Y. Keller, and T. Hassner, “DeepFake detection based on discrepancies between faces and their context,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 44, no. 10, pp. 6111–6121, Oct. 2022, doi: 10.1109/TPAMI.2021.3093446.
[19] D. A. Coccomini, N. Messina, C. Gennaro, and F. Falchi, “Combining EfficientNet and vision transformers for video deepfake detection,” in *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, vol. 13233 LNCS, 2022, pp. 219–229, doi: 10.1007/978-3-031-06433-3_19.
[20] Z. Tan, Z. Yang, C. Miao, and G. Guo, “Transformer-based feature compensation and aggregation for deepfake detection,” *IEEE Signal Processing Letters*, vol. 29, pp. 2183–2187, 2022, doi: 10.1109/LSP.2022.3214768.
[21] H. Chen, Y. Lin, B. Li, and S. Tan, “Learning features of intra-consistency and inter-diversity: keys toward generalizable deepfake detection,” *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 33, no. 3, pp. 1468–1480, Mar. 2023, doi: 10.1109/TCSVT.2022.3209336.
[22] C. Zhao, C. Wang, G. Hu, H. Chen, C. Liu, and J. Tang, “ISTVT: interpretable spatial-temporal video transformer for deepfake detection,” *IEEE Transactions on Information Forensics and Security*, vol. 18, pp. 1335–1348, 2023, doi: 10.1109/TIFS.2023.3239223.
[23] Y. Patel *et al.*, “An improved dense CNN architecture for deepfake image detection,” *IEEE Access*, vol. 11, pp. 22081–22095, 2023, doi: 10.1109/ACCESS.2023.3251417.
[24] K. N. Ramadhani, R. Munir, and N. P. Utama, “Improving video vision transformer for deepfake video detection using facial landmark, depthwise separable convolution and self attention,” *IEEE Access*, vol. 12, pp. 8932–8939, 2024, doi: 10.1109/ACCESS.2024.3352890.
[25] A. Khormali and J. S. Yuan, “Self-supervised graph transformer for deepfake detection,” *IEEE Access*, vol. 12, pp. 58114–58127, 2024, doi: 10.1109/ACCESS.2024.3392512.
[26] M. Javed, Z. Zhang, F. H. Dahri, and A. A. Laghari, “Real-time deepfake video detection using eye movement analysis with a hybrid deep learning approach,” *Electronics*, vol. 13, no. 15, Jul. 2024, doi: 10.3390/electronics13152947.
[27] F. Matern, C. Riess, and M. Stamminger, “Exploiting visual artifacts to expose deepfakes and face manipulations,” in *Proceedings-2019 IEEE Winter Conference on Applications of Computer Vision Workshops, WACVW 2019*, Jan. 2019, pp. 83–92, doi: 10.1109/WACVW.2019.00020.
[28] A. A. Laghari, Y. Sun, M. Alhussein, K. Aurangzeb, M. S. Anwar, and M. Rashid, “Deep residual-dense network based on bidirectional recurrent neural network for atrial fibrillation detection,” *Scientific Reports*, vol. 13, no. 1, Sep. 2023, doi: 10.1038/s41598-023-40343-x.
[29] N. Qazi and I. Ahmed, “Enhancing authenticity verification with transfer learning and ensemble techniques in facial feature-based deepfake detection,” in *2024 14th International Conference on Pattern Recognition Systems, ICPRS 2024*, Jul. 2024, pp. 1–6, doi: 10.1109/ICPRS62101.2024.10677831.
[30] L. Rasheed, U. Khadam, S. Majeed, S. Ramzan, M. S. Bashir, and M. M. Iqbal, “Face recognition emotions detection using haar cascade classifier and convolutional neural network,” *Research Square*, 2022, doi: 10.21203/rs.3.rs-2048290/v1.
[31] M. Wilkerson, G. Vincent, Z. Hasnain, E. Dunkel, and S. Bhattacharya, “Benchmarking ResNet50 for image classification on diverse hardware platforms,” *The ITEA Journal of Test and Evaluation*, vol. 45, no. 3, Sep. 2024, doi: 10.61278/itea.45.3.1008.

BIOGRAPHIES OF AUTHORS

Yeeshu Ralhen received his M. Tech degree in Electronics and Communication Engineering and is currently pursuing a Ph.D. from Maharishi Markandeshwar (Deemed to be University), Mullana, Haryana, India. He is working as an Assistant Professor at Panipat Institute of Engineering and Technology (PIET), Panipat, Haryana, India, with more than 14 years of academic experience. His research interests include machine learning, deep learning, and the internet of things (IoT), with a focus on deepfake detection and intelligent communication systems. He can be contacted at email: yeeshu.relhan@gmail.com.

Sharad Sharma received his doctoral degree in Electronics and Communication Engineering and is currently serving as Professor and Head of the Department of Electronics and Communication Engineering at Maharishi Markandeshwar Engineering College, Maharishi Markandeshwar (Deemed to be University), Mullana, Ambala, Haryana, India. He has over nine years of experience as a professor and has previously held various academic and administrative positions. His research interests include communication systems and artificial intelligence. He has published more than 486 research papers in reputed journals and conferences. He can be contacted at email: sharadpr123@gmail.com.