Pertanika Proceedings 1 (5): 101 - 107 (2025)

PERTANIKA JOURNALS

PERTANIKA PROCEEDINGS

Journal homepage: http://www.pertanika.upm.edu.my/

# Deep Fake Facial Analysis Using Deep Learning

**P. Dhanalakshmi[1]*, Ummadi JanardhanReddy[2], Ravikanth Garladinne[3], S. Ramachandra[4] and E. R. Aruna[5]**

*[1]Department of AIML, Mohan Babu University (Erstwhile Sree Vidyanikethan Engineering College) Tirupathi, India*

*[2]Department of Computer Science and Engineering, Alliance University, Chandapura, Anekal Main Road, Anekal, 562 106, India*

*[3]Department of Computer Science and Engineering, Koneru Lakshmaiah Education Foundation, Vaddeswaram, Guntur, Andhra Pradesh, 522502, India*

*[4]Department of Computer science and Engineering, Mother Theresa Institute of Engineering and Technology, Palamaner- 517 408, India*

*[5]Computer Science and Engineering, Vardhaman College of Engineering, Hyderabad 501218, Telangana, India*

## ABSTRACT

Deepfake detection has emerged as an essential area of research, especially in addressing the growing concerns of misinformation, identity theft, and political manipulation. This study focuses on using deep learning, particularly Convolution Neural Networks (CNNs) and the InceptionResNetV1 architecture, to analyze and detect facial manipulations in digital media. By training on a comprehensive dataset that includes both real and manipulated content, the proposed system identifies subtle discrepancies in facial features. The system's objective is to enhance the public's trust in digital media by accurately detecting deepfakes and fostering confidence in the authenticity of online information.

*Keywords:* Anomaly detection, fine-tuning, inceptionResNetV1, information integrity, image augmentation, image forensics, multi-modal analysis, pre-trained inceptionResNetV1, transfer learning

ARTICLE INFO

*Article history:*
Received: 29 August 2025
Published: 31 October 2025

DOI: https://doi.org/10.47836/pp.1.5.019

*E-mail addresses*:
mallidhana5@gmail.com (P. Dhanalakshmi)
ummadi.janardhan@gmail.com (Ummadi JanardhanReddy)
garladinneravikanth@gmail.com (Ravikanth Garladinne)
mail2sramachandra@gmail.com (S. Ramachandra)
er.aruna@vardhaman.org (E. R. Aruna)
* Corresponding author

## INTRODUCTION

The rapid evolution of deepfake technology, particularly through advanced algorithms like Generative Adversarial Networks (GANs), has cast a shadow over the authenticity of digital media. While these innovations hold potential for creative applications, they also present alarming



© Universiti Putra Malaysia Press

risks, including the spread of misinformation, identity theft, and political manipulation. Consequently, the development of effective deepfake detection tools has become a vital frontier in ensuring the security and trustworthiness of digital content.

The challenge of detecting deepfakes lies in the ability of these technologies to replicate realistic features, often making traditional verification methods obsolete. Deep learning models, trained on extensive datasets of both authentic and manipulated media, are key in recognizing subtle patterns that differentiate real from fake content. Techniques like convolutional and recurrent neural networks further enhance detection by identifying complex features. As detection strategies evolve, balancing technological advancement with ethical considerations—such as privacy remains a crucial challenge. The integration of technical, ethical, and regulatory frameworks will be essential for responsibly addressing the growing influence of deepfakes in the digital landscape.

## LITERATURE REVIEW

Zegedy et al. (2017) introduced the Inception-v4 and Inception-ResNet architectures, combining inception modules with residual connections to improve training efficiency and accuracy. Their work demonstrated that residual connections significantly accelerate convergence while enhancing model performance on large-scale image classification tasks. To ensure the efficiency of this literary work, a methodical technique was embraced to gather, examine and integrate magazines pertaining to deepfakes. This part outlines the methodology used throughout the research study procedure to ensure openness and also repeatability. It supplies a breakdown of the research study techniques utilized in this research study. This consists of information on the recognition together with the evaluation of magazines of rate of interest. Rössler et al. (2019) presented FaceForensics++, a large-scale dataset designed to benchmark the detection of manipulated facial images. Their study established strong baselines for deepfake detection and highlighted the challenges in generalizing across different forgery methods. The research study presented a unique deep learning-based technique for identifying AI-generated phony video clips from actual video clips. The research study focused on deepfake video clip face expression discovery, as many current formulas already effective in identifying the fixed human faces. They combined numerous sort of vision transformers with a convolutional Efficient Net used as an attribute extractor accomplishing outcomes comparable to some current vision transformer-based methods. Wodajo, as well as Atnafu used a convolutional vision transformer to recognize deep phony information.

Dang et al. (2020) provided a comprehensive survey of deep learning approaches in image and video forensics, covering methods for forgery detection, source identification, and authenticity verification. Their work highlighted both the progress and the remaining

challenges in applying deep learning to multimedia forensics. This data source resolves the expanding problem surrounding the control of face video clips, especially with the surge of deepfake modern technology. It consists of a big collection of video clips having both genuine and also controlled face material, allowing scientists to create and also examine formulas for discovering built face video clips.

Li et al. (2020) introduced DSFD, a dual shot face detector that features enhancement and progressive refinement to improve detection accuracy. Their approach achieved strong performance on challenging face detection benchmarks, particularly under varying scales and occlusions. The DSFD design is developed for precision along with reliable face discovery in photos. It utilizes a dual-shot discovery system, which boosts its efficiency in identifying encounters under numerous problems such as various ranges, presents, together with occlusions.

Afchar et al. (2018) proposed MesoNet, a lightweight deep neural network designed to detect facial forgeries in videos. Their model demonstrated efficiency and effectiveness in identifying manipulated content with relatively low computational cost. Hsu and Wu (2020) developed a fine-tuning-based approach for DeepFake image detection using small datasets. Their method improved detection accuracy while reducing the dependency on large-scale training data. The semantic network most likely catches intricate patterns plus functions as a measure of photo adjustments, allowing the design to properly differentiate between initial and also adjusted pictures.

## Existing Systems

Convolutional Neural Networks (CNNs) play a crucial role in deepfake detection due to their ability to automatically extract patterns from visual data,as shown in Figure1. Deepfake technology uses AI and deep learning to create realistic but fake images and videos, often aiming to deceive viewers. Given the increasing concerns about deepfake misuse, effective detection methods are essential.

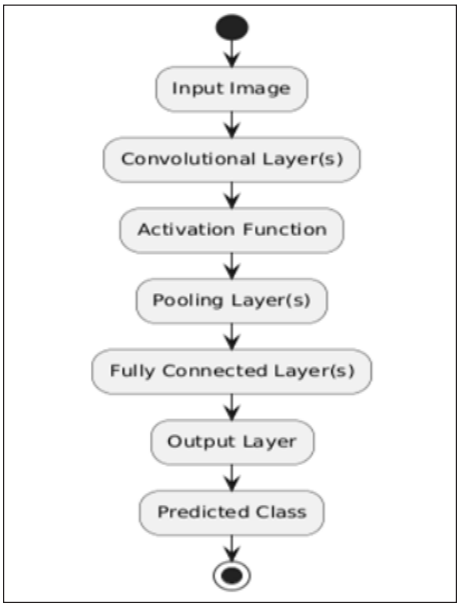


*Figure 1.* CNN model for image detection

## Feature Extraction

CNNs excel at extracting relevant features from images or video frames, identifying subtle patterns and artifacts that may indicate deepfake manipulation. The lower

layers capture basic features like edges and textures, while deeper layers can detect more complex and abstract patterns.

### Training on Authentic and Deepfake Data

CNNs are trained on datasets containing both real and deepfake content, enabling the network to learn the differences in patterns between genuine and manipulated material, with the training process adjusting the model's parameters to minimize prediction errors.

### VGG19

VGG-19 is a convolutional neural network (CNN) model with 19 layers, trained on the ImageNet dataset shown in Figure 2. It consists of 16 convolutional layers, 5 2 × 2 max-pooling layers, 3 fully connected layers, and 1 softmax layer. The input size is 224 × 224 × 3, and the network uses 3 × 3 filters with a stride and padding of 1. Each convolutional layer is followed by a ReLU activation and max-pooling, with the number of filters increasing as the network deepens, while the spatial dimensions of the feature maps reduce due to pooling.

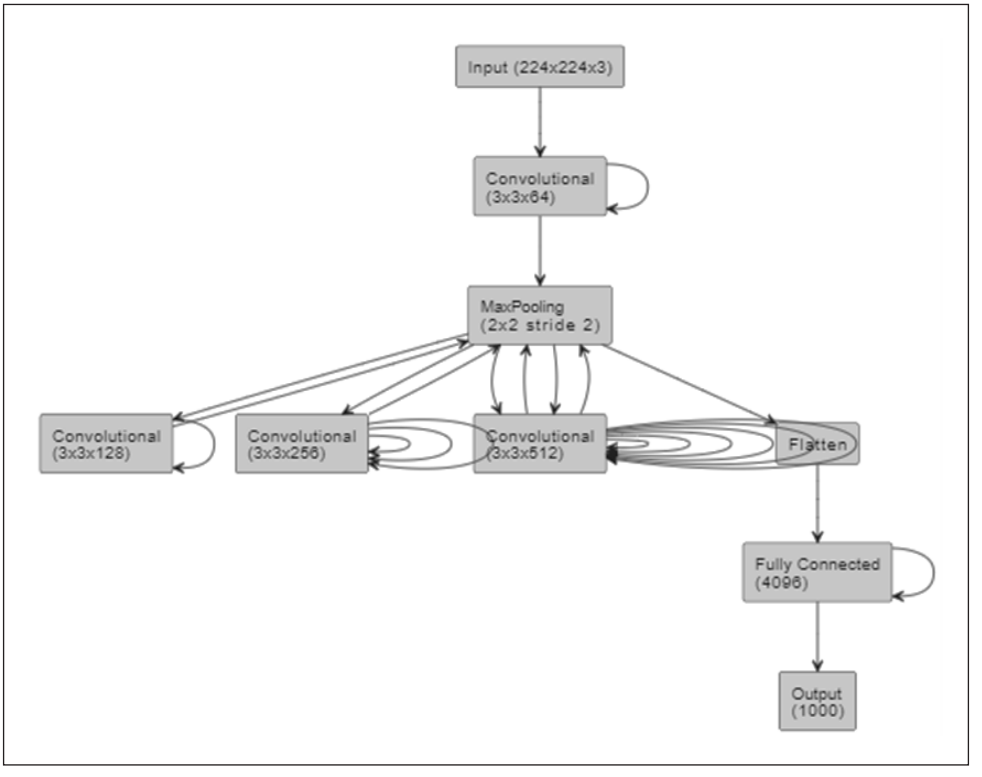


*Figure 2.* VGG19 architecture for image classification

## PROPOSED MODELS

### Multi-Task Cascaded Convolutional Networks (MTCNN)

Multi-task Cascaded Convolutional Networks (MTCNN) is a deep learning architecture designed for both face detection and face landmark localization. It operates through a three-stage process involving convolutional networks to identify faces and locate facial features like eyes, nose, and mouth, which is shown in Figure 3. The first stage, the Proposal Network (P-Net), resizes an input image into a pyramid and utilizes a fully convolutional network (FCN) to generate candidate windows and bounding box regression vectors. The second stage, the Refine Network (R-Net), processes the P-Net outputs, further refining the candidate windows by reducing false positives, applying bounding box regression calibration, and using non-maximum suppression (NMS) to eliminate overlapping proposals. This hierarchical approach helps improve the accuracy and efficiency of face detection and alignment tasks. The Output Network (O-Net) builds upon the R-Net by providing more detailed facial information and outputting the precise locations of the five facial landmarks: eyes, nose, and mouth.

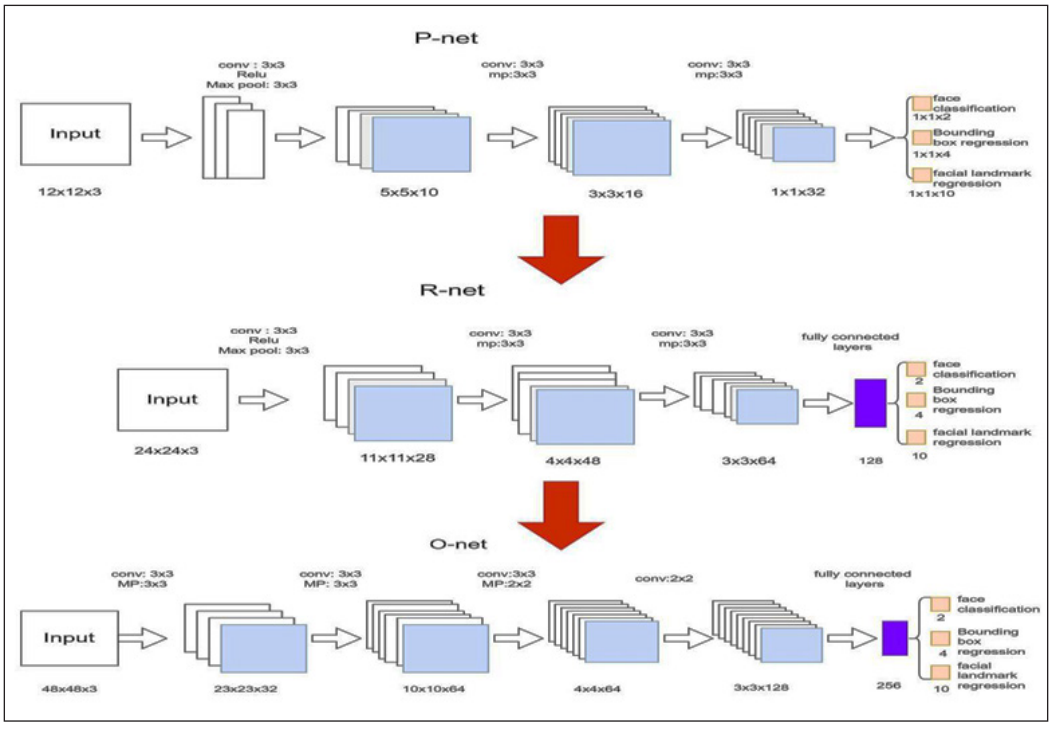


*Figure 3.* MTCNN model for image detection

## RESULTS

Our deepfake facial analysis system, powered by InceptionResNetV1, effectively detects manipulated facial images through rigorous training and validation. Using a diverse dataset of real and synthetic content, it identifies subtle differences between authentic and altered faces. Quantitative analysis demonstrates high accuracy in deepfake detection, showcasing the system's efficiency in combating content manipulation, which is shown in Figure 4 and Figure 5. The model's ability to distinguish between genuine and fake images highlights its robustness. Overall, it offers a reliable solution for detecting digital deception in facial media.

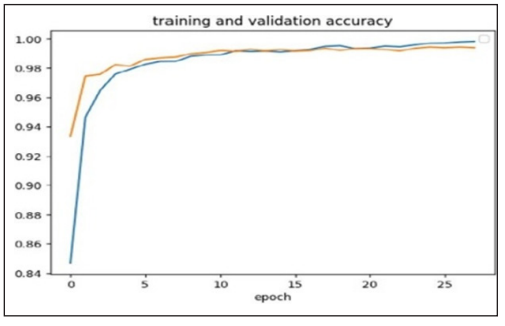


*Figure 4.* Training accuracy




*Figure 5.* Validation accuracy

## CONCLUSION

The InceptionResNetV1-based deepfake detection model offers a complete solution, emphasizing accuracy, interpretability, and generalization. It leverages robust feature extraction and diverse training to ensure effective performance. Continuous monitoring and updates maintain its adaptability to emerging threats, promoting ethical use in detecting deceptive content.

## ACKNOWLEDGEMENT

The author would like to express sincere gratitude to everyone who supported and guided them throughout the research process, and is deeply thankful to mentors, colleagues, and all the resources that contributed to the successful completion of this study.

## REFERENCES

Afchar, D., Nozick, V., Yamagishi, J., & Echizen, I. (2018). MesoNet: A compact facial video forgery detection network. In *Proceedings of the European Conference on Computer Vision (ECCV)* (pp. 1–13). Springer.

Dang, C. H., Vu, T. D., Nguyen, D. T., & Ngo, D. T. (2020). A survey of deep learning techniques for image and video forensics. *Information Sciences, 546,* 696–725. https://doi.org/10.1016/j.ins.2020.08.083

Hsu, C. Y., & Wu, M. (2020). DeepFake image detection based on fine-tuning with a small dataset. *IEEE Access, 8,* 14119–14127. https://doi.org/10.1109/ACCESS.2020.2966884

Li, Y., Yang, B., Hospedales, T. M., & Xiang, T. (2020). *DSFD: Dual shot face detector*. GitHub. https://github.com/Tencent/FaceDetection-DSFD

Rössler, A., Cozzolino, D., Verdoliva, L., Riess, C., Thies, J., & Nießner, M. (2019). FaceForensics++: Learning to detect manipulated facial images. In *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)* (pp. 1–11). IEEE.

Zegedy, C., Ioffe, S., Vanhoucke, V., & Alemi, A. A. (2017). Inception-v4, Inception-ResNet and the impact of residual connections on learning. *Proceedings of the AAAI Conference on Artificial Intelligence, 31*(1), 4278-4284 . https://doi.org/10.1609/aaai.v31i1.11231