IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

Impact Factor 8.471 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 15, Issue 8, August 2026

DOI: 10.17148/IJARCCE.2026.15835

DEEPFAKE IMAGE DETECTION USING DEEP LEARNING

Pavaman D.K[1], Yashodara.R[2], Prakash Naik[3], Puneeth.A[4], Divyashree.K[5], Ullas Reddy .G[6]

Dept. of ISE, Don Bosco Institute of Technology, Bangalore, India[1,3,4,6]

Assistant Professor, Dept. of ISE, Don Bosco Institute of Technology, Bangalore, India[2,5]

Abstract: The rapid growth of artificial intelligence has made it possible to create images and videos that look increasingly realistic. One of the most noticeable examples of this technology is deepfake media, where a person's face or identity can be digitally changed using artificial intelligence. Although deepfake technology has useful applications in areas such as entertainment, education and digital content creation, it can also be misused to spread false information, impersonate people and manipulate digital evidence

Because modern deepfakes can look very convincing, identifying them simply by looking at an image is becoming more difficult. This has led researchers to explore automated detection methods based on machine learning and deep learning. In particular, convolutional neural networks have been widely used because they can learn visual patterns that may not be obvious to human observers.

This literature review examines existing research on deepfake generation and detection, commonly used datasets, deep learning approaches and the challenges faced by current detection systems. It also discusses EfficientNet as a lightweight deep learning architecture and its suitability for developing a practical deepfake image detection system such as DeepGuard.

I.INTRODUCTION

Images have become one of the main ways people communicate information on the internet. Social media platforms, news websites and messaging applications allow images to be created and shared within seconds. While this has made communication easier, it has also created a problem: it is becoming harder to determine whether a digital image is *genuine*.The development of generative artificial intelligence has made this problem more serious. Modern AI systems can generate realistic faces, modify existing photographs and replace one person's face with another. These manipulated images are commonly associated with the term deepfakes.

.Deepfake technology itself is not necessarily harmful. It can be used for film production, visual effects, education and other creative applications. However, when the same technology is used without a person's consent or to deliberately create misleading information, it can become a serious problem.

Nguyen et al. noted that deep learning has made it possible to create highly realistic synthetic media and argued that automated methods for assessing the integrity of digital media are increasingly necessary.The problem is particularly difficult because deepfake generation techniques continue to improve. Older manipulated images may contain obvious visual defects, while newer methods can produce much more convincing results. As a result, deepfake detection has become an ongoing research problem rather than a problem with a single permanent solution.

This has motivated researchers to investigate computer vision and deep learning techniques that can automatically identify subtle differences between authentic and manipulated facial images.

II. RELATED WORK

Harsh Vajpayee, Sushant Jhingran, Nitin Yadav & Ayush Raj[1] Detecting Deepfake Human Face Images Using Transfer Learning: A Comparative Study. Compared transfer-learning models based on EfficientNetV2. EfficientNetV2L achieved 99% accuracy, showing that transfer learning can strongly improve real/fake face classification.

Ankita Bhandarkar, Prasad Lokulwar, Prashant Khobragade, Pranay Saraf & Raju Pawar [2] Deep Learning Framework for Robust Deep Fake Image Detection: A Review. Reviewed CNN, autoencoder, RNN and attention-based methods. Highlighted data augmentation, transfer learning and ensemble learning as ways to improve detection. It also identified challenges such as small datasets and high computational requirements.

DOI: 10.17148/IJARCCE.2026.15835

Sujanthi S, Priya R, Ranga Shree S, Suruthika S & Sushmitha S [3] Comparison Analysis of Transfer Learning Models for Deep Fake Image Detection Compared several pretrained models including Xception, NASNet, DenseNet, InceptionV3, EfficientNet, MobileNetV2, ResNet50 and Custom CNN on FaceForensics++. This represents development toward selecting the most effective pretrained architecture.

Dhruv Shetty, Purva Ambre, Pranjal Patil & Prachi Shahane [4] FauxFind: Deep Learning-Based Deepfake Detection System using Transfer Learning and CNN. Proposed MesoNet + CNN + transfer learning for detecting subtle deepfake artifacts in images.

Naif A. Almalki, Mahmoud Ragab & Faris Kateb. [5] Leveraging AI for Sustainable Deepfake Human Face Detection Using Transfer Learning Technique. Compared VGG16, VGG19, MobileNetV2 and ResNet50V2 using transfer learning. The proposed VGG16 model was enhanced with additional convolution, MaxPooling, dropout and dense layers.

Priyanshu Hirpara, Hardi Valangar, Vishwa Kachhadiya &Uttam Chauhan.[6] Deepfake Detection: Demodulate Synthetic Videos Using Deep Learning Models. Proposes a CNN–LSTM hybrid model. CNN extracts visual/spatial features from video frames while LSTM captures temporal information.

Merve Yildirim & Ilhan Aydin [7] detecting deepfake Audio and Video Manipulation with Multimodal Deep Learning Proposes multimodal deep learning, analysing both audio and video simultaneously rather than relying on only one modality. This combines visual and auditory information for deepfake detection.

Satyendra Singh, Rajesh Kumar, Abhishek Rai, Abhishek Kushwaha, Anurag Singh Baghel [8] Ensemble-Based Deep Learning Framework for Detecting Deepfake Images and Videos Proposes a **stacked ensemble deep-learning framework** combining **ResNet50, Vision Transformer (ViT) and Swin Transformer**. A meta-learning/stacking approach combines their predictions to improve detection performance.

Rupesh Kumar,Anurag Kumar Baitha & Shiksha Singh [9] deepfake detection using deep learning . The main contribution of this paper is the development of a relatively lightweight CNN-based DeepfakeShield model that can detect manipulated facial images without requiring heavy computational resources.

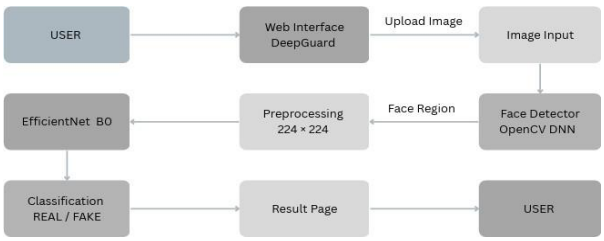
Nishika Khatri, Varun Borar, and Rakesh Garg[10] : A Comparative Study: Deepfake Detection Using Deep-learning. This paper compares four popular deep-learning models to find out which model can detect fake images most accurately, using the same FaceForensics++ dataset and evaluation measures.

Prof. Archana Jadhav, Dhaval Narale, Rokdeshwar Kore, Ujjwal Shisode & Abhijeet Kulange[11] Unmasking the Illusion: A Novel Approach for Detecting Deep Fakes using Video Fusion of spatial features from CNN and temporal/global features from Vision Transformer.

Priyanshu Hirpara, Hardi Valangar, Vishwa Kachhadiya & Uttam Chauhan [12] Deepfake Detection: Demodulate Synthetic Videos using Deep Learning Models . Cropping video frames before applying deep learning improved detection performance.

All though existing studies provide solutions for detecting all deep fake image detection , the system detects facial regions from an uploaded image and uses a trained EfficientNet-B0 model to classify the detected content as **REAL** or **FAKE**. This type of technology can help users identify potentially manipulated images and can provide an additional layer of verification when dealing with suspicious digital content.

III. PROPOSED SYSTEM



IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

Impact Factor 8.471 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 15, Issue 8, August 2026

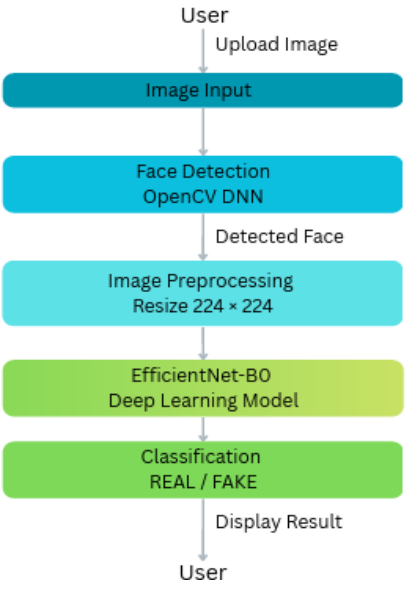
DOI: 10.17148/IJARCCE.2026.15835

The proposed system, **Deep fake detection**, is designed to provide a simple and practical way to identify whether a facial image is real or potentially manipulated. The main idea is to combine automatic face detection with a deep learning classification model so that the user does not need to manually crop or prepare the image before analysis.
When a user uploads an image, system first keeps the original image at its original resolution and automatically searches for faces within it. This allows the system to work with images containing multiple people rather than being limited to photographs with a single face. Each detected facial region is then processed and given to an **EfficientNet-B0** model that has been trained using real and fake facial images. The model learns visual patterns that help distinguish genuine facial content from manipulated content.

After the individual facial regions have been analysed, the system combines the predictions to produce an overall result for the uploaded image. The application presents this result in a simple **REAL** or **FAKE** format rather than displaying technical prediction scores for every individual face. This makes the system easier for ordinary users to understand and avoids unnecessary technical information.

DeepGuard also includes a web-based interface where users can upload an image, preview it, start the analysis, and view the final result. The purpose of the proposed system is to make deepfake detection more accessible while demonstrating how computer vision and deep learning can be combined to address the growing problem of manipulated digital images. Although the system provides a useful automated indication, its prediction should be considered as a supporting tool rather than absolute proof that an image is genuine or manipulated.

IV. FLOW DIAGRAM



1. User Image
The user uploads an image through the DeepGuard web interface. The system accepts common image formats such as JPG, JPEG, PNG and WEBP.
2. Image Processing
The uploaded image is loaded while preserving its original resolution for display. For the deep-learning model, the required input
preprocessing is applied.
3. Face Detection
The system detects the faces present in the image. It is not limited to a single face, so an image containing multiple people can be processed.
4. Face Cropping
Each detected face is extracted from the original image. These face regions are then passed to the trained deepfake detection model.
5. EfficientNet-B0
Your trained EfficientNet-B0 model analyzes the face image and determines whether it contains characteristics

IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

Impact Factor 8.471 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 15, Issue 8, August 2026

DOI: 10.17148/IJARCCE.2026.15835

associated with a real or manipulated image.

6. Prediction

The model produces predictions for the detected faces. These predictions are used internally by the application.

V. COMPARATIVE STUDY OF EXISTING SYSTEMS TO OUR SYSTEM

Existing Systems

Existing deepfake detection systems use different approaches to identify manipulated images and videos. Common approaches include traditional machine-learning methods, CNN-based deep-learning models, and more recent hybrid architectures.

- **Traditional Machine-Learning Systems**

Earlier approaches can use manually designed visual features such as HOG, GLCM and other image characteristics, followed by classifiers such as SVM or Random Forest. These methods can be computationally efficient, but their performance depends heavily on the quality of the manually selected features. Recent comparative research has found deep-learning approaches generally achieve stronger detection performance than these hand-crafted-feature approaches.

- **CNN-Based Deepfake Detection**

Modern systems commonly use convolutional neural networks to automatically learn visual patterns associated with manipulated faces. Well-known research benchmarks such as FaceForensics++ evaluate facial manipulation detection using manipulations including DeepFakes, Face2Face, FaceSwap and NeuralTextures.

- **Advanced Deep-Learning Systems**

More recent systems use architectures such as EfficientNet, Vision Transformers, attention mechanisms and combinations of CNNs and Transformers. These systems can achieve strong benchmark performance, but they can also be more complex and computationally demanding. Research has also shown that performance can change significantly when models are tested on datasets or manipulation types different from those used for training.

Proposed System: DEEPFAKE IMAGE DETECTION

The proposed system, **DeepGuard**, is a web-based deepfake image detection system designed to provide a simple and user-friendly solution. The system allows users to upload an image through a web interface and automatically detects the faces present in the image. Multiple faces can be processed, and each detected face is analyzed internally using the trained **EfficientNet-B0** deep-learning model. Based on the model predictions, DeepGuard provides a simple overall **REAL or FAKE** result to the user. The system does not display individual face predictions or unnecessary technical information, making the detection process easier to understand. The main goal of DeepGuard is to combine deep-learning-based detection with an easy-to-use interface for practical image verification.

The main focus of DeepGuard is to make deepfake detection practical and easy to use. The technical processing, including face detection and individual face analysis, happens in the background, while the user receives a simple overall prediction on the result page. The application displays the original uploaded image along with the final **REAL or FAKE** classification, without overwhelming the user with individual face scores or unnecessary technical information. This makes DeepGuard suitable for users who want a quick and understandable way to check an image for possible manipulation, while still using a deep-learning-based detection approach underneath.

Key Advantages of the Proposed System – DeepGuard

- **User-friendly interface** – Simple web-based image upload and detection.
- **Multiple-face detection** – Can analyze images containing multiple faces.
- **Deep-learning based** – Uses a trained **EfficientNet-B0** model for classification.
- **Fast and automated** – Face detection and prediction are performed automatically.
- **Simple result** – Provides an easy-to-understand overall **REAL/ FAKE** prediction.
- **No technical knowledge required** – Users do not need to understand the underlying model.
- **Practical application** – Can help users identify potentially manipulated images.

International Journal of Advanced Research in Computer and Communication Engineering
Impact Factor 8.471 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 15, Issue 8, August 2026
DOI: 10.17148/IJARCCE.2026.15835

- **Original image display** – Shows the uploaded image along with the final detection result.
- **Clean and minimal design** – Avoids unnecessary technical scores and information.

VI. METHODOLOGY

The **DeepGuard** methodology begins with collecting and preparing real and fake facial images for training. The images are preprocessed and resized to **224 × 224 pixels** before being given to the trained **EfficientNet-B0** model. During detection, the system accepts an image from the user, detects all faces present in it, extracts the detected faces, and analyzes them using the trained model. The individual predictions are processed internally to generate an overall **REAL or FAKE** result. Finally, the result is displayed through a simple and user-friendly web interface along with the uploaded image.

Step 1: Dataset Collection

A dataset containing both **real and fake facial images** is collected for training the deepfake detection model. The dataset contains **15,037 images**, including **9,242 real images** and **5,795 fake images**.

Step 2: Dataset Preparation

The collected images are organized into **REAL** and **FAKE** classes. The dataset is divided into training and validation sets. In your implementation, **12,029 images** are used for training and **3,008 images** are used for validation.

Step 3: Image Preprocessing

The images are preprocessed before being provided to the deep-learning model. The input images are resized to **224 × 224 pixels**, which matches the input size used by the EfficientNet-B0 model. The images are also converted into a suitable tensor format for model processing.

Step 4: Face Detection

When a user uploads an image, the system first detects the faces present in the image. This allows DeepGuard to handle images containing **multiple people**, rather than analyzing only one large face. Each detected face is extracted for further analysis.

Step 5: Face Extraction

The detected face regions are cropped from the uploaded image. These cropped regions contain the facial information required by the deep-learning model. Each face is processed separately while the original image is retained for displaying the final result.

Step 6: Deep-Learning Model

The extracted faces are passed to the trained **EfficientNet-B0** model. EfficientNet-B0 automatically learns important visual patterns from the training images instead of depending on manually selected features.

Step 7: REAL/FAKE Classification

The trained model analyzes each detected face and predicts whether it belongs to the **REAL** or **FAKE** category. The predictions from the detected faces are then used by the application to determine the overall classification of the uploaded image.

Step 8: Overall Result Generation

The system combines the internal face-level predictions and generates a single overall result for the uploaded image. The user is shown only the final **REAL** or **FAKE** prediction rather than individual face predictions and confidence scores.

Step 9: Result Display

The final result is presented through the DeepGuard web interface. The result page displays the uploaded image and the overall prediction. The user can then choose **"Analyze Another Image"** to perform another detection.

IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

Impact Factor 8.471 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 15, Issue 8, August 2026

DOI: 10.17148/IJARCCE.2026.15835

VII. CONCLUSION

In conclusion, based on the literature survey conducted, it is possible to state that the problem of detecting deepfake images has progressed from the stage of traditional image-processing algorithms and handcrafted features to the stage of deep-learning and transfer-learning-based approaches. The papers discussed above have shown that convolutional networks and pretrained models are suitable for addressing this problem, as they allow acquiring patterns specific to each type of altered imagery. In particular, transfer learning appears to be a very effective method for this application, as using it, one can achieve decent results in terms of classifying images without having to train a deep neural network from scratch.

However, the reviewed research also highlights the need for practically useful and accessible detection systems that go beyond model-level considerations. In this regard, the proposed DeepGuard method was created using the automatic detection of faces in the input image with subsequent verification using an EfficientNet-B0 classifier. The method is designed to preserve the highest resolution of the input image while also supporting multi-faced images, with the detected facial area and an overall REAL/FAKE prediction result being returned through a web-based interface.

Thus, DeepGuard proves that the results of researching deepfake-detection methods can be implemented in practice to make such technology more accessible for users. The software product helps to ensure the reliability of digital images and decreases the complexity of detection algorithms based on deep learning. The suggested approach can be developed to make it possible to work with new types of manipulations and expand the range of tested images.

REFERENCES

[1]. Satyendra Singh, Abhishek Kushwaha, Rajesh Kumar, Anurag Singh Baghel & Abhishek Rai , “Ensemble-Based Deep Learning Framework for Detecting Deepfake Images and Videos” – 2nd International Conference on Cognitive Computing in Engineering, Communications,2026.

[2]. Sujanthi S, Priya R, Ranga Shree S, Suruthika S, Sushmitha S, "Comparison Analysis of Transfer Learning Models for Deep Fake Image Detection" - Proceedings of the International Conference on Machine Learning and Autonomous Systems (ICMLAS), IEEE Xplore -,2025.

[3]. Dhruv Shetty, Purva Ambre, Pranjal Patil, Prachi Shahane "FauxFind: Deep Learning-Based Deepfake Detection System Using Transfer Learning and CNN" - IEEE conference paper ,2025.

[4]. Priyanshu Hirpara, Hardi Valangar, Vishwa Kachhadiya, Uttam Chauhan - "Deepfake Detection: Demodulate Synthetic Videos Using Deep Learning Models" - IEEE conference paper , 2025.

[5]. Naif A. Almalki, Mahmoud Ragab, Faris Kateb "Leveraging Al for Sustainable Deepfake Human Face Detection Using Transfer Learning Technique" - 2025 22nd International Learning and Technology Conference (L&T), IEEE ,2025.

[6]. Merve Yıldırım & İlhan Aydın – “Detecting Deepfake Audio and Video Manipulation with Multimodal Deep Learning” – IEEE conference paper ,2025.

[7]. Ankita Bhandarkar, Prasad Lokulwar, Prashant Khobragade, Pranay Saraf, Raju Pawar – “Deep Learning Framework for Robust Deep Fake Image Detection: A Review” .International Conference on Artificial Intelligence and Quantum Computation-Based Sensor Application (ICAIQSA), IEEE, 2024.

[8]. Prof. Archana Jadhav, Dhaval Narale, Rokdeshwar Kore, Ujjwal Shisode, Abhijeet Kulange – “Unmasking the Illusion: A Novel Approach for Detecting Deep Fakes Using Video Vision Transformer Architecture” ,2024.

[9]. Harsh Vajpayee, Sushant Jhingran, Nitin Yadav & Ayush Raj – “Detecting Deepfake Human Face Images Using Transfer Learning: A Comparative Study” – 2023 IEEE International Conference on Contemporary Computing and Communications (InC4), IEEE , 2023.

[10]. P. Saikia, D. Dholaria, P. Yadav, V. Pate & M. Roy – “A Hybrid CNN-LSTM Model for Video Deepfake Detection by Leveraging Optical Flow Features”,2022.

[11]. Y. Al-Dhab & S. Zhang – “Deepfake Video Detection by Combining Convolutional Neural Network (CNN) and Recurrent Neural Network (RNN)” – 2021 IEEE International Conference on Computer Science, 2021.

[12]. J. K. Lewis, I. E. Toubal, H. Chen, V. Sandesera, M. Lomnitz & Z. Hampel-Arias – “Deepfake Video Detection Based on Spatial, Spectral, and Temporal Inconsistencies Using Multimodal Deep Learning”,2020.