International Journal of Engineering Research and Science & Technology

www.ijerst.org

ISSN : 2319-5991

Vol. 22 No. 2(1) (2026)

ijerst.editor@gmail.com

editor@ijerst.com

ISSN 2319-5991 www.ijerst.com

Peer Reviewed,Referred & Indexed Journal

Research Paper

DEEPFAKE FACE DETECTION OF IMAGES USING TRANSFER

[1]N Ramya, [2] Sai divya,[3]Archana yadav, [4] Afroz
[1]AssistantProfessor, [234]Students
Department of Computer Science and Technology
Siddhartha Institute of Technology & Sciences, Narapally
n.ramya@siddartha.co.in, 23TQ1A05F8@siddhartha.co.in, 23TQ1A05E2@siddhartha.co.in, 23TQ1A05F0@siddhartha.co.in,

Abstract

With the rapid advancement of artificial intelligence and deep learning technologies, the creation of highly realistic manipulated media, commonly known as deepfakes, has become increasingly widespread. Deepfake images are generated using sophisticated algorithms that can alter or replace facial features, making it difficult to distinguish between real and fake content. This raises serious concerns related to misinformation, identity theft, and digital security, making deepfake detection a critical problem in computer vision and digital forensics.

This project presents an effective approach for detecting deepfake face images using transfer learning techniques. The system utilizes a pre-trained deep learning model, MobileNetV2, to extract high-level features from facial images and classify them as real or fake. By leveraging transfer learning, the model benefits from knowledge gained from large-scale datasets, resulting in improved accuracy and reduced training time.

The dataset used consists of both real and manipulated facial images obtained from publicly available sources. Preprocessing steps such as image resizing and normalization are applied to ensure data consistency. The dataset is then divided into training, validation, and testing sets for proper evaluation of the model. During training, MobileNetV2 acts as a feature extractor, while additional layers are used for binary classification. The model learns to identify subtle inconsistencies in facial structures, textures, and patterns that indicate manipulation.

I. Introduction

In recent years, the rapid advancement of artificial intelligence (AI) and deep learning technologies has led to the development of powerful tools capable of generating highly realistic digital media. One such technology is deepfake, which uses advanced neural network models to manipulate or synthesize visual content, particularly human faces. Deepfake techniques commonly rely on architectures such as Generative Adversarial Networks (GANs) and autoencoders to create convincing images or videos by altering facial expressions, swapping identities, or modifying features. These techniques have significantly improved in quality, making the generated content almost indistinguishable from real media.

While deepfake technology has beneficial applications in areas like entertainment, film production, gaming, and virtual reality, it also poses serious threats. It can be misused to create misleading or harmful content, including fake news, identity fraud, cyberbullying, and political manipulation. The high realism of deepfake images makes it extremely difficult for humans to detect manipulation with the naked eye, increasing the risk of misuse in digital platforms.

As a result, deepfake detection has emerged as a critical research area in computer vision, digital forensics, and cybersecurity. Researchers are focusing on developing automated detection systems that can accurately distinguish between real and manipulated images. These systems analyze subtle visual artifacts, inconsistencies in textures, and irregular patterns introduced during the deepfake generation process. The growing need for reliable detection mechanisms highlights the importance of integrating advanced deep learning techniques to ensure authenticity and trust in digital media.

II. Literature Survey

Deepfake detection has become a significant research area in computer vision and digital forensics due to the rapid increase in AI-generated manipulated media. Researchers have explored various deep learning approaches, including Convolutional Neural Networks (CNNs), transfer learning models, and transformer-based architectures, to accurately detect deepfake images and videos. This section reviews important research works related to deepfake detection.

Paper 1: FaceForensics++: Learning to Detect Manipulated Facial Images
Authors: Andreas Rössler et al. (2019)
This study introduced the FaceForensics++ dataset, one of the largest and most widely used benchmarks for deepfake detection. The dataset includes a large number of manipulated images created using techniques such as DeepFakes, FaceSwap, Face2Face, and NeuralTextures. The authors evaluated several CNN-based models on this dataset and demonstrated that deep learning models can effectively detect manipulated facial content when trained on large-scale datasets.
Key Contributions: Creation of a benchmark dataset, evaluation of multiple models, and standardization of deepfake detection research.

Paper 2: Deepfake Detection Using Convolutional Neural Networks
Authors: Various Researchers (2024)
This research focused on using CNN-based models for detecting deepfake images. CNNs automatically extract important visual features such as edges, textures, and facial structures, making them highly suitable for image classification tasks. The study showed that CNN models can achieve high accuracy in distinguishing real and fake images by identifying subtle inconsistencies. Data augmentation and balancing techniques were also used to improve model performance.
Key Contributions: Demonstrated CNN effectiveness, improved performance using preprocessing techniques, and achieved strong classification accuracy.

Paper 3: Deepfake Detection with Deep Learning – CNN vs Transformer Models
Authors: V.L.L. Thing et al. (2023)

This study compared CNN-based models with transformer-based architectures for deepfake detection. The models were trained on multiple datasets such as FaceForensics++, Celeb-DF, and DFDC. The results indicated that both approaches perform well, with some models achieving over 92% accuracy. The research also highlighted that transformer models can capture global image dependencies, while CNNs focus on local features.
Key Contributions: Comparative analysis of models, high accuracy across datasets, and emphasis on the importance of large-scale training data.

Paper 4: Transfer Learning for Deepfake Detection Using Pre-trained Models
Authors: Various Researchers (2022)
This paper explored the use of transfer learning techniques for deepfake detection by utilizing pre-trained models such as MobileNet, VGG, and ResNet. Instead of training models from scratch, the knowledge from large datasets was reused to improve detection performance. The study showed that transfer learning reduces training time and computational cost while achieving higher accuracy.
Key Contributions: Demonstrated efficiency of transfer learning, reduced computational complexity, and improved detection performance.

Paper 5: Deepfake Image Detection Using Hybrid Deep Learning Models
Authors: Various Researchers (2023)
This research proposed hybrid models that combine CNNs with other techniques such as LSTM or attention mechanisms to improve detection accuracy. The hybrid approach helps capture both spatial and temporal features of manipulated content. The results showed improved robustness and generalization compared to single-model approaches.
Key Contributions: Introduced hybrid models, improved robustness, and enhanced feature extraction for better detection.

Overall, the literature indicates that deep learning techniques, especially CNNs and transfer learning models, play a crucial role in detecting deepfake images. The use of large datasets, advanced architectures, and hybrid approaches significantly improves detection accuracy and efficiency.

III. System Analysis

The deepfake face detection system focuses on identifying manipulated facial images using deep learning techniques. The system analyzes image data to distinguish between real and fake content by learning patterns and visual inconsistencies. It involves collecting a dataset of real and deepfake images from various sources. The data is preprocessed by resizing, normalization, and cleaning to ensure quality input. Feature extraction is performed using deep learning models to capture important facial characteristics. The system uses transfer learning to improve performance and reduce training time. Machine learning models are trained and tested using labeled datasets. Evaluation metrics such as accuracy and loss are used to measure performance. The system aims to provide reliable and automated detection of deepfake images. Overall, it enhances security and authenticity in digital media.

Existing System

The existing systems for deepfake detection mainly rely on traditional image processing and basic machine learning techniques. These methods use handcrafted features such as texture, color inconsistencies, and pixel-level analysis to detect manipulation. Some systems use simple classifiers like Support Vector Machines (SVM) or basic neural networks. These approaches often depend on manually designed features, which may not capture complex patterns in deepfake images. The performance of these systems is limited when dealing with high-quality deepfakes. They also struggle with large and diverse datasets. Many existing methods are not robust against new deepfake generation techniques. The detection accuracy is often low compared to modern deep learning models. These systems require significant manual effort in feature extraction. Overall, the existing systems are less efficient and less reliable for real-world applications.

Disadvantages of Existing System

The existing deepfake detection systems suffer from several limitations that reduce their effectiveness. They rely heavily on handcrafted features, which may fail to capture complex patterns present in modern deepfake images. These systems often show low accuracy when dealing with high-quality or advanced deepfakes generated using sophisticated algorithms. They are not robust and can easily fail when new manipulation techniques are introduced. The performance of traditional methods is limited when handling large and diverse datasets. Many existing systems require manual feature extraction, which is time-consuming and error-prone. They also lack the ability to generalize well across different datasets. Real-time detection is difficult due to inefficient processing techniques. Additionally, these systems are less adaptable and require frequent updates to remain effective.

Proposed System

The proposed system uses advanced deep learning techniques with transfer learning to detect deepfake face images. It utilizes a pre-trained model such as MobileNetV2 to extract meaningful features from images. The system processes input images through preprocessing steps like resizing and normalization. Transfer learning allows the model to leverage knowledge from large datasets, improving performance. Additional layers are added to perform binary classification of real and fake images. The system can detect subtle inconsistencies in facial features, textures, and patterns. It is capable of handling large datasets efficiently. The model is trained, validated, and tested to ensure accuracy and reliability. It provides fast and automated predictions. Overall, the system is designed to improve detection accuracy and robustness.

Advantages of Proposed System

The proposed system offers several advantages over traditional methods. It provides higher accuracy due to the use of deep learning and transfer learning techniques. The system can automatically extract complex features without manual intervention. It is capable of detecting high-quality and advanced deepfake images. Transfer learning reduces training time and computational effort. The model performs well on large and diverse datasets. It improves scalability and can be extended to real-time applications.

The system reduces human error and increases reliability. It can be integrated into security and media verification systems. Overall, the proposed system is more efficient, accurate, and suitable for modern deepfake detection challenges.

IV. Methodology

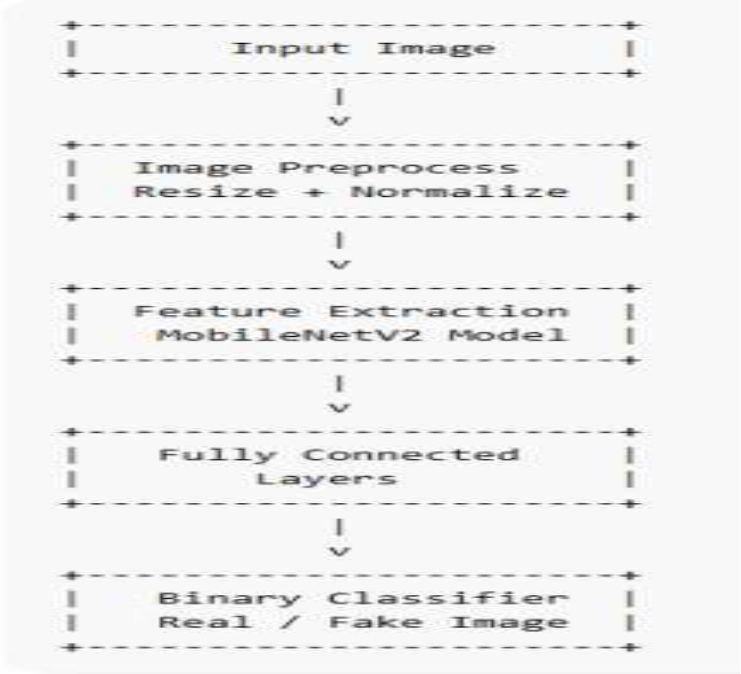
The methodology for the deepfake face detection system follows a structured deep learning approach using transfer learning techniques. Initially, a dataset consisting of real and deepfake facial images is collected from publicly available sources. The collected data is then preprocessed by resizing images to a fixed size, normalizing pixel values, and removing noise to ensure data consistency.

Next, the dataset is divided into training, validation, and testing sets to evaluate model performance effectively. A pre-trained deep learning model, MobileNetV2, is selected as the base model for feature extraction. Transfer learning is applied by using the pre-trained weights and adding custom fully connected layers for binary classification (real or fake).

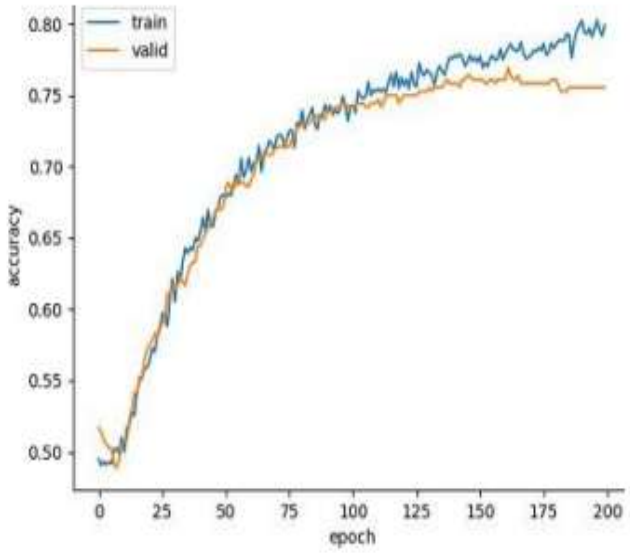
During training, the model learns to identify subtle differences in facial features, textures, and inconsistencies present in deepfake images. Optimization techniques such as backpropagation and suitable loss functions are used to improve model accuracy. After training, the model is evaluated using metrics like accuracy and loss to measure performance.

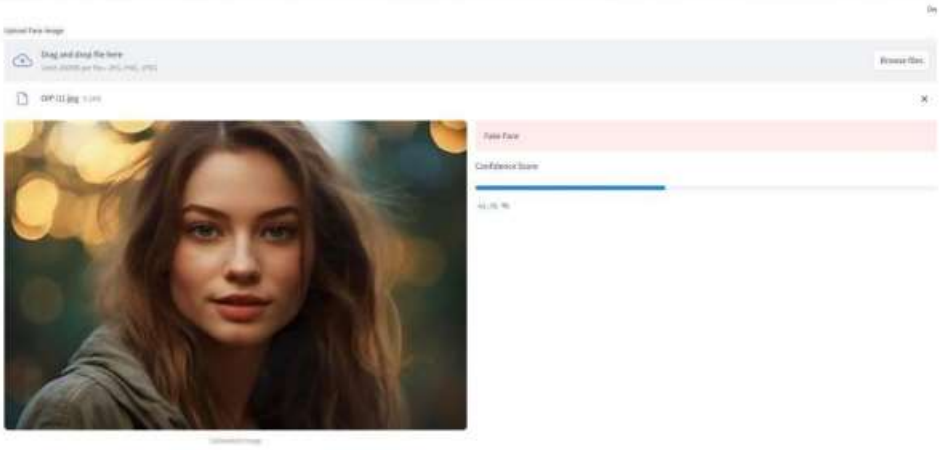
System Architecture

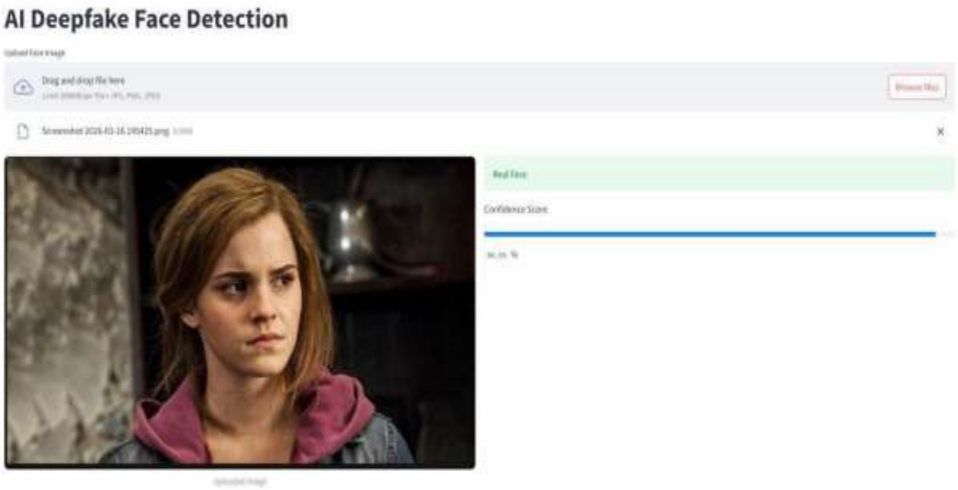
The system architecture for deepfake face detection using transfer learning is designed as a structured and layered framework that efficiently processes image data to produce accurate predictions. Initially, the system collects real and deepfake facial images from various publicly available datasets. These images are then passed through a preprocessing stage where they are resized to a fixed resolution, normalized, and cleaned to ensure consistency. The processed dataset is divided into training, validation, and testing sets for proper model development and evaluation. A pre-trained deep learning model such as MobileNetV2 is used in the feature extraction stage, where transfer learning enables the system to leverage previously learned knowledge. Additional fully connected layers are added to perform binary classification of images as real or fake. The model is trained using labeled data and learns to identify subtle inconsistencies in facial features and textures. Once trained, the system predicts the authenticity of new input images.



V. Result and Output







VI. Conclusion

In conclusion, the project on deepfake face detection using transfer learning provides an effective solution to identify manipulated facial images in the growing landscape of AI-generated media. By leveraging a pre-trained model such as MobileNetV2, the system is able to extract meaningful features and accurately classify images as real or fake. The use of transfer learning significantly reduces training time while improving model performance and accuracy.

The system successfully detects subtle inconsistencies in facial patterns, textures, and structures that are often difficult for humans to identify. It demonstrates the importance of deep learning techniques in addressing real-world challenges related to digital security, misinformation, and media authenticity. The evaluation results confirm that the proposed approach achieves reliable and efficient detection performance.

Overall, this project contributes to the development of robust deepfake detection systems and highlights the potential of transfer learning in computer vision applications. The system can be further enhanced by incorporating larger datasets,

advanced models, and real-time detection capabilities, making it suitable for practical deployment in cybersecurity and digital forensics.

References

[1] Kumar, R. D., Prudhviraj, G., Vijay, K., Kumar, P. S., & Plugmann, P. (2024). Exploring COVID-19 through intensive investigation with supervised machine learning algorithm. In Handbook of Artificial Intelligence and Wearables (pp. 145-158). CRC Press.
[2] Swathi, B., Vijay, K., Sushanth Babu, M., & Dinesh Kumar, R. (2024, November). Machine Learning Techniques in Cloud Based Intrusion Detection. In The International Conference on Artificial Intelligence and Smart Environment (pp. 557-564). Cham: Springer Nature Switzerland.
[3] Sv satyakrishna, shirisha rangu ,bhargavi nalacheruve.(2024) Prospective investigation on colorectal cancer with SMOTE on machine learning Algorithm
[4] Dr.G.Vishnu Murthy, BhargaviNalacheruve 1Professor, Department of computer Science & engineering, Anurag University, TS, India. 2Student, Department of computer Science & engineering, Anurag University, TS, India.
[5] V. N. S. Manaswini, K. K, C. Nigam, S. S. Ali, R. Niranjana, and Suman, "Real-Time Object Detection in Drone Surveillance Using YOLOv5," in Proc. 2025 3rd Int. Conf. IoT, Communication and Automation Technology (ICICAT), Gorakhpur, India, 2025, pp. 1–6, doi: 10.1109/ICICAT68430.2025.11414670.
[6] B. Soundarya, V. N. S. Manaswini, M. Ayyakrishnan, R. D. Kumar, "Contextual Analysis of Big Data Analytics in Intelligent Transportation Frameworks," in Intersection of Artificial Intelligence, Data Science, and Cutting-Edge Technologies: From Concepts to Applications in Smart Environment, Lecture Notes in Networks and Systems, vol. 1353, Cham: Springer, 2025, doi: 10.1007/978-3-031-88304-0_79.
[7] R. D. Kumar, V. N. S.Manaswini, "Applications of blockchain in smart cities: detecting fake documents from land records using blockchain technology," in Blockchain for Smart Cities, Elsevier, 2021, pp. 105–117, doi: 10.1016/B978-0-12-824446-3.00017-X.
[8] Tejavath Veeramma, Badarla Anil, Guguloth Ravinder, "An advanced movie recommender using collaborative filtering and sentiment analysis," International Research Journal of Modernization in Engineering Technology and Science, vol. 7, no. 7, July 2025, doi: 10.56726/IRJMETS81618.
[9] Ravi Kumar Banoth, Ramana Murthy B V, "Automatic crop recommendation system using LightGBM and decision tree machine learning models," Journal of Machine and Computing, vol. 5, no. 1, pp. 343, Jan. 2025, doi: 10.53759/7669/jmc202505026.
[10] Ravi Kumar Banoth, Dr. B.V. Ramana Murthy, "Smart agriculture through IoT and machine learning for analyzing carbon footprints," in Proc. Int. Conf. Computer Science and Communication Engineering (ICCSCE), Apr. 2025.

[11] Ravi Kumar Banoth, B. V. Ramana Murthy, “Soil image classification using transfer learning approach: MobileNetV2 with CNN,” SN Computer Science, vol. 5, art. no. 199, 2024, doi: 10.1007/s42979-023-02500-x.