International Journal of Scientific Research in Engineering and Management (IJSREM)
Volume: 09 Issue: 08 | Aug - 2025 SJIF Rating: 8.586 ISSN: 2582-3930

Deepfake Image Detection Using Transfer Learning

[1]Sindhu S L, [2]Sirish A S

[1]ASSISTANT PROFESSOR, Department of MCA, BIET, Davanagere

[2]STUDENT,4th Semester MCA, Department of MCA, BIET, Davanagere

ABSTRACT

The proliferation of DeepFake technology, powered by advancements in generative models such as Generative Adversarial Networks (GANs), poses serious threats to digital authenticity, privacy, and information integrity. In response to this growing concern, this paper presents a DeepFake Detection System implemented as a Flask web application utilizing deep learning techniques. The core of the detection mechanism is a fine-tuned ResNet50 convolutional neural network, pre-trained on ImageNet and adapted through transfer learning to accurately classify real and fake images. The system provides a user-friendly interface allowing registered users to upload images for real-time DeepFake detection, delivering immediate classification results. The application also includes distinct admin and user functionalities admins can manage users and maintain a FAQ section, while users can register, log in, and access the detection and informational features. Powered by PyTorch for backend model processing and Flask for web deployment, the system offers a practical tool for media verification in domains such as journalism, security, and social media, contributing to the global fight against misinformation and synthetic media manipulation.

Keywords: *DeepFake Detection, Synthetic Media, Generative Adversarial Networks (GANs), ResNet50, Transfer Learning, Image Classification, Convolutional Neural Network (CNN).*

I. INTRODUCTION

In recent years, the rapid evolution of generative technologies, particularly in deep learning, has enabled the creation of hyper-realistic synthetic media, commonly referred to as DeepFakes. These manipulated images and videos are typically generated using artificial intelligence techniques, most notably Generative Adversarial Networks (GANs). The emergence of DeepFakes has introduced significant challenges in the realms of misinformation, privacy, and cybersecurity. The ability to generate convincingly fake media has amplified the need for robust and efficient tools capable of detecting such manipulated content.

This paper presents the development of a DeepFake Detection System designed as a web-based application using the Flask framework. The proposed system leverages deep learning-based image classification to differentiate between real and manipulated images. At the core of the detection pipeline lies ResNet50, a pre-trained Convolutional Neural Network (CNN), which has been fine-tuned through transfer learning for the specific task of DeepFake image identification. By applying transfer learning, the model builds upon prior knowledge from training on the ImageNet dataset and adapts to detect features relevant to fake media, using a dedicated dataset comprising both authentic and fake images.

The Flask-based web application offers a user-friendly interface that facilitates real-time DeepFake detection. Users can upload images, which are processed through the trained ResNet50 model to yield a binary classification: either real or fake. The application is intended as a practical verification tool for journalists, educators, and other stakeholders combating digital misinformation, as it aids in identifying potentially harmful synthetic content.

The system architecture clearly defines two roles: admin and user. Admins can log in to the system, manage the database of registered users by viewing or deleting user accounts, and curate a Frequently Asked Questions (FAQ) section by adding or removing entries. Users, upon registering and logging in with valid credentials, are permitted to upload images for DeepFake classification. In addition, they can access the FAQ section for guidance and support. This role-based division ensures efficient access control and smooth application management.

The backend of the system is developed using PyTorch, responsible for tasks such as image preprocessing, model training, and inference. The frontend is constructed using Flask and provides an interactive and intuitive interface for user interaction. The integration of the model with the web platform allows seamless real-time DeepFake detection. The architecture is scalable and can be extended to accommodate video detection or integration with real-time data sources in the future.

This proposed system delivers a practical DeepFake detection framework that is applicable across multiple domains where image credibility is critical. Potential applications include digital journalism, legal documentation, cybersecurity, and social media content moderation. By providing a robust tool for detecting manipulated images, the system contributes to efforts aimed at curbing the spread of misinformation and upholding the authenticity of digital media.

II. RELATED WORK

Deepfake Detection Using ResNext50. Authors: Vaishali Jadhav, Taabish Sutriwala, Raunak Singh, Sohel Siraj Mukadam.

The paper **"Deepfake Detection Using ResNext50"** explores the growing challenge of detecting AI-generated fake media commonly known as deepfakes caused by advancements in artificial intelligence. It highlights how deepfakes can convincingly manipulate images, audio, and video, posing threats to privacy and misinformation. The study discusses common indicators of deepfakes, such as abnormal blinking, inconsistent head poses, and mismatched speech-lip movements. It categorizes detection methods into manual (human-led) and algorithmic (AI-based), and proposes a deepfake detection system leveraging the **ResNext50** model to automatically identify manipulated visual content using multimodal analysis techniques**.[1]**

Novel Solution for Deepfake Image Detection using CNN and ResNet50 Architecture. Authors: Daksh Sanghvi, Vansh Solanki, Akhilesh Sonarikar, Rishabh Jaiswal.

This paper presents a deep learning-based approach for detecting image-based deepfakes using Convolutional Neural Networks (CNNs) and the ResNet50 architecture. By leveraging transfer learning, the ResNet50 model is fine-tuned on deepfake-specific datasets, enabling effective identification of manipulated images. The research emphasizes key aspects like dataset diversity, model tuning, and evaluation through performance metrics such as accuracy, precision, recall, and F1 score. The study demonstrates the strong potential of CNN and ResNet50 in identifying deepfakes and offers insights into challenges and areas for further research in improving image authenticity verification**.[2]**

Deepfake Detection Using LSTM & ResNeXt-50. Authors : Ms. M. Varalakshmi, Saketh Paida,L. Ruthwik Reddy, J. Samel

This paper introduces a deep learning-based approach to detect deepfake videos by combining ResNeXt-50 and LSTM architectures. ResNeXt extracts features at the frame level, while the LSTM network captures temporal dependencies across video sequences, enabling accurate differentiation between real and AI-generated content. The proposed method is trained and evaluated on benchmark datasets like FaceForensics++, Celeb-DF, and the Deepfake Detection Challenge dataset. Results show that the system effectively detects manipulated videos, outperforming existing models, and demonstrates robustness for real-time applications in security-sensitive domains**.[3]**

Comparative Analysis of Deepfake Image Detection Method Using VGG16, VGG19 and ResNet50. Authors: Zahra Nazemi Ashani, Iszuanie Syfidza Che Ilias, Keng Yap Ng, Muhammad Rezal Kamel Ariffin, Ahmad Dahari Jarno, Nor Zarina Zamri

This study presents a comparative evaluation of three CNN architectures□VGG16, VGG19, and ResNet50□for deepfake image detection. Utilizing a dataset of 1,200 real and fake

images generated using FaceApp, the authors found that **VGG19 achieved the highest detection accuracy of 98%**, outperforming the other models, especially on small-sized input images. The research demonstrates that AI-based approaches, particularly CNNs, offer effective solutions to the deepfake challenge and provides strong evidence supporting VGG19□s superiority for detecting manipulated digital content in a growing threat landscape**.[4]**

Deepfake Detection: A Systematic Literature Review. Authors: Md Shohel Rana, Mohammad Nur Nobi, Beddhu Murali, Andrew H. Sung.

This paper provides a systematic literature review (SLR) of 112 research articles published between 2018 and 2020 on deepfake detection techniques. It categorizes the detection methods into four major groups: deep learning-based, classical machine learning, statistical, and blockchain-based approaches. The review evaluates these techniques across multiple datasets, concluding that deep learning-based models outperform others in accuracy and robustness. The study highlights the growing threat posed by deepfakes in spreading misinformation and emphasizes the need for continuous advancement in detection methods**.[5]**

Deepfake Image Detection & Classification using Conv2D Neural Networks. Authors: Debasish Samal , Prateek Agrawal, Vishu Madaan.

This paper presents a Conv2D convolutional neural network architecture tailored for deepfake image detection. The model was trained on a balanced dataset of 140,000 images (70,000 real and 70,000 fake) to improve over existing approaches that often rely on smaller datasets and pre-trained models. Utilizing sparse categorical cross-entropy and the Adam optimizer, the model achieves a strong accuracy of 94.54% on the OpenForensics dataset□a benchmark known for its complexity in multi-face forgery detection. The study highlights the capability of custom CNN architectures, without pre-trained layers, to effectively distinguish real from AI-generated media, reinforcing the promise of Conv2D models in real-world deepfake detection**.[6]**

Deepfake Face Detection Using Deep InceptionNet Learning Algorithm. Authors : Prasannavenkatesan Theerthagiri, Ghouse Basha Nagaladinne.

This paper proposes a deep learning-based method for detecting deepfake images and videos using the InceptionNet convolutional neural network architecture. The model was trained on a dataset from Kaggle consisting of 401 videos and 3745 images, which were generated through data augmentation techniques. The study includes a comparative analysis of various CNN architectures. The proposed method achieved an accuracy of 93 percent, as evaluated using standard metrics such as accuracy and confusion matrix, proving its efficiency in identifying manipulated multimedia content**.[7]**

Deepfake Detection: A Literature Review. Authors : Sayed Shifa Mohd Imran, Dr. Pallavi Devendra Tawde.

This paper presents a review of deepfake detection techniques, emphasizing the increasing threat of fake videos and audio in cyberattacks, phishing, and misinformation campaigns. It outlines the importance of robust detection mechanisms to prevent identity theft, misinformation spread, and cyberbullying. The review explores tools capable of identifying biometric anomalies in videos, such as heartbeats or human-like voice patterns. It highlights the interdisciplinary nature of the field, calling for collaboration between experts in computer science, artificial intelligence, psychology, and law. The paper concludes by stressing the need for ongoing development, awareness, and policy support to counteract the evolving threat of deepfakes**.[8]**

A Comprehensive Review of DeepFake Detection Using Advanced Machine Learning and Fusion Methods. Authors: Gourav Gupta, Kiran Raja, Manish Gupta, Tony Jan, Scott Thompson Whiteside, Mukesh Prasad.

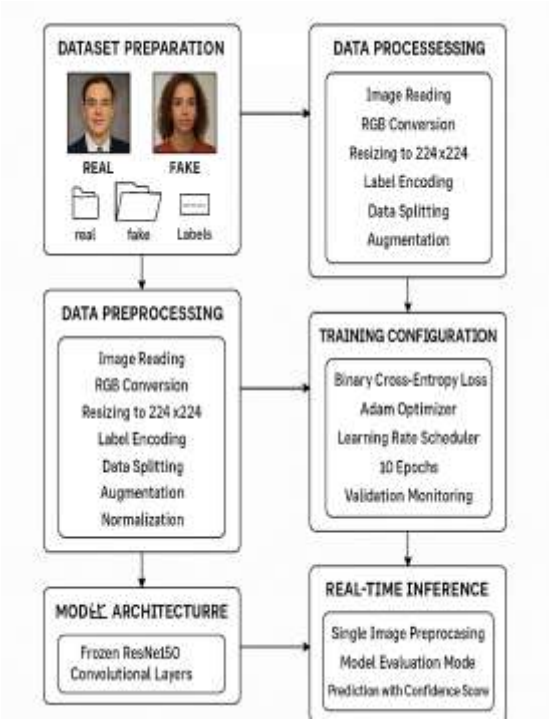
This review paper examines the growing threat of DeepFakes□AI-generated hyper-realistic videos and images used to spread misinformation and manipulate public perception. It presents a detailed analysis of advanced machine learning and modality fusion techniques applied to DeepFake detection. Covering 67 primary research papers published between 2015 and 2023, the review focuses on image and video DeepFake detection as well as speaker authentication. It explores benchmark datasets and outlines state-of-the-art models while also offering insights and guidelines for future advancements in DeepFake detection. The paper uniquely emphasizes media-modality fusion

strategies and advanced ML, setting it apart from other reviews in this domain.**[9]**

AI Deep Fake Detection Research Paper. Authors : Raghava M. S, Tejashwini S. P, Kavya Sree, Sneha A, Naveen R

This paper presents a comprehensive survey of deepfake generation and detection techniques. While deep learning has enabled breakthroughs across domains like vision and analytics, it also powers deepfakes□hyper-realistic fake images and videos□raising major concerns about privacy, democracy, and national security. The study provides a detailed review of deepfake creation methods and critically examines detection algorithms developed to counteract these threats. It highlights challenges, research trends, and future directions, aiming to assist in developing more advanced and reliable detection systems. The ultimate goal is to preserve the integrity and security of digital visual media in an increasingly AI-driven world.**[10]**

III. METHODOLOGY



This research aims to develop a deep learning framework for detecting deepfake images by leveraging transfer learning with the ResNet50 convolutional neural network. The methodology is divided into several key stages including dataset preparation, data preprocessing, model architecture modification, training configuration, performance evaluation, and real-time inference.

3.1 Dataset Preparation

The dataset used in this study consists of two categories: real and fake images. These images are stored in a structured directory format with separate folders for training and testing. Using the Python `os` module, all image file paths are programmatically collected, and labels are assigned based on the folder names. To ensure consistency and manageability, a Pandas DataFrame is created to map each image to its corresponding label. This structured representation facilitates efficient loading and manipulation of the data throughout the modeling pipeline.

3.2 Data Preprocessing

Before feeding the images into the neural network, they undergo several preprocessing steps. Each image is read using OpenCV, converted to RGB color space, and resized to a fixed resolution of 224 by 224 pixels to comply with the input dimensions expected by ResNet50. The textual class labels are encoded into binary format, assigning zero to fake images and one to real images. The complete dataset is randomly shuffled and split into training, validation, and testing sets to avoid any potential bias. Specifically, around ninety percent of the data is used for training, five percent for validation, and the remaining five percent is set aside for testing. To improve generalization and reduce overfitting, the training data is augmented with transformations such as random horizontal flips and rotations. Finally, all images are normalized using the standard mean and standard deviation values of the ImageNet dataset, ensuring compatibility with the pre-trained model's input distribution.

3.3 Model Architecture

The core of the detection system is based on the ResNet50 architecture, which is pre-trained on the ImageNet dataset. To adapt this model for the binary classification task of deepfake detection, a transfer learning approach is adopted. All convolutional layers of ResNet50 are frozen to retain the powerful image feature representations learned from large-scale natural images. The final fully connected layer of the model is replaced with a custom binary classification head. This newly introduced head consists of multiple linear layers interleaved with batch normalization, LeakyReLU activation functions, and dropout layers to enhance robustness and prevent overfitting. A final sigmoid activation layer is added

to output a probability score indicating whether the input image is real or fake. This architectural customization enables the model to be both efficient and accurate for the specialized task at hand.

3.4 Training Configuration

For the training process, the binary cross-entropy loss function is used, which is appropriate for binary classification problems. The Adam optimizer is employed with an initial learning rate set to 0.0001 to ensure stable and efficient convergence. Additionally, a learning rate scheduler known as ReduceLROnPlateau is incorporated to automatically reduce the learning rate when the validation loss stagnates, thereby improving the model□s ability to escape local minima. The model is trained for ten epochs, during which its performance is evaluated on the validation set at the end of each epoch. The model with the best validation loss is saved for final testing and deployment using PyTorch□s built-in saving mechanism.

3.5 Evaluation and Testing

Upon completion of training, the model is tested using the reserved test set to evaluate its generalization performance. The test images are processed using the same preprocessing pipeline applied during training to maintain consistency. The model generates a probability score for each image, which is then converted into a binary label based on a threshold of 0.5. These predicted labels are compared with the true labels to compute the overall classification accuracy. Additionally, selected test images are visually inspected along with their predicted and actual labels to assess the qualitative performance of the model and to identify any misclassifications or patterns of error.

3.6 Real-Time Inference

To enable real-world usability, a real-time prediction pipeline is developed. In this setup, a single image provided by the user is preprocessed using the same transformations as during training. The trained model is loaded and placed in evaluation mode to prevent any changes to its weights during prediction. The image is passed through the model, which then outputs a classification result□real or fake□along with a confidence score representing the model□s certainty. This real-time component demonstrates the practical applicability of the system and its potential for integration into digital forensics tools or content verification platforms.

IV. TECHNOLOGIES USED

- **Flask**

Flask is a lightweight and micro web framework written in Python that allows for the quick development of web applications. It is especially suitable for projects that require integration with machine learning models due to its simplicity and flexibility. In this project, Flask serves as the backbone of the web interface, handling routing, user input (image uploads), and displaying results.

- **Python**

Python is the primary programming language used in this project. Known for its readability and vast library support, Python is ideal for machine learning, data processing, and web development. It is used for implementing both the backend logic of the application and the deep learning model.

- **PyTorch**

PyTorch is an open-source deep learning framework that offers dynamic computation graphs and intuitive model building. It is used to load, fine-tune, and train the pre-trained ResNet50 model for binary classification of real and fake images. PyTorch also supports GPU acceleration, enabling faster training and inference.

- **OpenCV**

OpenCV (Open Source Computer Vision Library) is used for image preprocessing tasks. It provides tools for reading image files, resizing images to match model input size, and converting images from BGR to RGB format. These steps are crucial for maintaining consistency with the ResNet50 model□s expected input format.

- **NumPy**

NumPy is a core library for numerical computing in Python. It is used in this project for handling array operations, such as converting image pixel values into a format compatible with PyTorch tensors. It ensures that data manipulation is fast and memory efficient.

- **Pandas**

Pandas is a powerful Python library for data manipulation and analysis. It is used to organize the image file paths and corresponding labels into a structured DataFrame format. This structured representation helps in managing and accessing training and testing data easily during model training and evaluation.

- **Matplotlib**

Matplotlib is a plotting library used for data visualization. In this project, it is used to display sample test images along with their predicted and actual labels. This visual assessment helps in evaluating the model's performance qualitatively.

VI. RESULT

Image is predicted real.

Image predicted fake

V. CONCLUSION

In conclusion, the Social Media Analytics system offers a robust and integrated solution for content creators and marketers seeking to optimize their presence on YouTube and Instagram. By leveraging powerful technologies such as Python, Pandas, useragent, Google APIs Client, and Instaloader, the system efficiently collects, processes, and analyzes key performance metrics across both platforms. Its modular architecture, featuring dedicated user and admin modules, ensures effective management, security, and usability. The comprehensive analytics and intuitive visualizations empower users to make data-driven decisions, refine content strategies, and maximize audience engagement, ultimately enhancing their competitive edge and impact in the dynamic landscape of social media.

REFERENCES

[1].Deepfake Detection Using ResNext50. Authors: Vaishali Jadhav, Taabish Sutriwala, Raunak Singh, Sohel Siraj Mukadam. Journal: Journal of Network Security. Publisher: STM Journals. ISSN (Print): 2321□8517, eISSN (Online): 2395□6739, DOI: https://doi.org/10.37591/jons.v10i1.917. Volume and Issue: Vol 10, No 1 (2022).

[2]. Novel Solution for Deepfake Image Detection using CNN and ResNet50 Architecture. Authors: Daksh Sanghvi, Vansh Solanki, Akhilesh Sonarikar, Rishabh Jaiswal. Conference:

International Conference on Large Language Models and Use Cases (LLMUC2023). Published: Issue Number 1, 2025. Publisher: Control System Labs.

[3]. Deepfake Detection Using LSTM & ResNeXt-50. Authors : Ms. M. Varalakshmi, Saketh Paida,L. Ruthwik Reddy, J. Samel. International Journal of Scientific Research in Engineering and Management (IJSREM). Volume: 08 Issue: 05 | May - 2024 SJIF Rating: 8.448 ISSN: 2582-3930

[4]. Comparative Analysis of Deepfake Image Detection Method Using VGG16, VGG19 and ResNet50. Authors: Zahra Nazemi Ashani, Iszuanie Syfidza Che Ilias, Keng Yap Ng, Muhammad Rezal Kamel Ariffin, Ahmad Dahari Jarno, Nor Zarina Zamri. DOI: [https://doi.org/10.37934/araset.47.1.1628]. Journal: ARASET (Applied Research and Smart Engineering Technologies), Volume 47, Issue 1.

[5]. Deepfake Detection: A Systematic Literature Review. Authors: Md Shohel Rana, Mohammad Nur Nobi, Beddhu Murali, Andrew H. Sung. Publisher: IEEE. Journal: IEEE Access, Volume 10. [https://doi.org/10.1109/ACCESS.2022.3154404]. Electronic ISSN: 2169-3536.

[6]. Deepfake Image Detection & Classification using Conv2D Neural Networks. Authors: Debasish Samal , Prateek Agrawal, Vishu Madaan. Published in CEUR Workshop Proceedings (Vol. 3706) under ISSN 1613-0073

[7]. Deepfake Face Detection Using Deep InceptionNet Learning Algorithm. Authors : Prasannavenkatesan Theerthagiri, Ghouse Basha Nagaladinne. Conference : 2023 IEEE International Students' Conference on Electrical, Electronics and Computer Science (SCEECS). Publisher : IEEE. DOI : 10.1109/SCEECS57921.2023.10063128

[8]. Deepfake Detection: A Literature Review. Authors : Sayed Shifa Mohd Imran, Dr. Pallavi Devendra Tawde. Journal : International Research Journal of Engineering and Technology (IRJET). Volume / Issue / Date : Volume 11, Issue 03, March 2024. Publisher : IRJET (International Research Journal of Engineering and Technology). ISSN : e-ISSN: 2395-0056, p-ISSN: 2395-0072

[9].A Comprehensive Review of DeepFake Detection Using Advanced Machine Learning and Fusion Methods. Authors: Gourav Gupta, Kiran Raja, Manish Gupta, Tony Jan, Scott Thompson Whiteside, Mukesh Prasad. Journal : Electronics. Volume / Issue / Year : Volume 13, Issue 1, 2024. DOI : [https://doi.org/10.3390/electronics13010095]

[10].AI Deep Fake Detection Research Paper. Authors : Raghava M. S, Tejashwini S. P, Kavya Sree (Student, Department of AI & ML, DSATM), Sneha A, Naveen R . Journal: International Journal of Novel Research and Development (IJNRD). Volume / Issue / Year : Volume 8, Issue 10, October 2023. ISSN: 2456-4184. Paper ID : IJNRD2310407