Digital Image Forgery Detection

[1]Parkavi C; [2]Karthika M; [3]Dhanush M; [4]Saran SM; [5]Srinivas A

[1]Assistant Professor
Department of Artificial Intelligence &
Machine Learning
SNS College of Technology
Coimbatore, India

[2,3,4,5] UG Scholar
Department of Artificial Intelligence &
Machine Learning
SNS College of Technology
Coimbatore, India

Publication Date: 2025/05/10

Abstract: This project proposes an image forgery detection method using CNN, capable of delivering high accuracy and a clear explanation for each forged instance. In the past few years, image forgery has increased drastically, owing to the easy availability of image editing tools and techniques, including morphing. To mitigate this increasing menace and mitigate the effect of tampered content, the current project presents a framework for detecting digital image forgery, employing Convolutional Neural Networks (CNN) in conjunction with generative AI tools. The proposed framework classifies images as either forged or original and gives the reason for its classification using Google Gemini, which is interfaced through a Flask-based application. Unlike conventional detection methods, this method not only yields high accuracy reaching 96% on the dataset tested but also increases interpretability by giving the reason behind forgery predictions. This solution is intended to help users better recognize tampered images, thereby enhancing trust in digital content.

Keywords: *Digital Image Forgery, Convolutional Neural Network (CNN), Generative AI, Google Gemini, Deep Learning, Image Classification.*

How to Cite: Parkavi C; Karthika M; Dhanush M; Saran SM; Srinivas A (2025) Digital Image Forgery Detection. *International Journal of Innovative Science and Research Technology*, 10(4), 3091-3094. https://doi.org/10.38124/IJISRT/25apr1698

I. INTRODUCTION

On the internet, authenticity of photographic content becomes an essential aspect of upholding the credibility of web media. In the wake of the fast-evolving technological developments of photo processing equipment and manipulations, photo forgery on the web is becoming a common and advanced phenomenon. Manipulated photographs abused in such a manner can cause damage by giving false information, enabling fraud or causing public misunderstanding and hence can become a huge issue of ruining the integrity of information in such sectors as news-making, criminal investigations, and web forensics. Conventional methods for detecting forgery are more likely to be manual verification or hand-engineered features, making the process not just time-consuming but also less efficient at identifying advanced manipulations. In this paper, a robust digital image forgery detection scheme is proposed by utilizing a CNN-model trained using a large corpus of 10,000 genuine and forged images. The system classifies not just images as real or forged but also provides an interpretative explanation of the classification through Google Gemini using a Flask-based web interface. This contributes to the level of transparency and trust in the detection procedure. The proposed model's accuracy is 96.0%, which confirms the effectiveness of the model for detecting forgery.

II. PROBLEM STATEMENT

The rapid development of digital editing tools has made image forgery much more advanced and harder to identify. Methods like morphing, splicing, and cloning enable tampered images to look extremely real, and as a result, serious repercussions arise in areas like journalism, law enforcement, social media, and digital evidence management. Older approaches to image forgery detection depend on human inspection or simplistic feature extraction, which fail to handle contemporary, advanced manipulations. There is an urgent requirement for an intelligent, automated solution that not only detects forged images accurately but also provides a transparent explanation for its conclusion to establish user trust and enhance interpretability.

III. DESIGN THINKING

Design Thinking was applied in this project to create a user-centered solution for digital image forgery detection. The process began by empathizing with users' need for reliable and interpretable image verification tools. The problem was defined as the growing challenge of detecting forged images due to advanced editing techniques. A solution was ideated using Convolutional Neural Networks (CNN) for accurate classification, combined with Google Gemini to generate human understandable explanations. A prototype was developed using a Flask-based web interface, and the system was tested for performance, achieving 96% accuracy. This approach ensures both technical reliability and user trust by focusing on accuracy.

IV. LITERATURE REVIEW

Digital image forgery detection is an important area of research as fake digital content becomes more common. In the past, traditional methods relied on specific features like Discrete Cosine Transform (DCT), noise estimation, and pixel analysis to find oddities in images. While these methods worked okay in some situations, they struggled with more sophisticated forgeries. Then machine learning took things up a notch. Techniques like Support Vector Machines and Random Forest started being used to analyze image features. But these still required quite a bit of feature engineering and often depended on the image format and how it was compressed. The big shift came with deep learning, specifically through Convolutional Neural Networks. CNNs automatically learn patterns from raw images and do a better job at spotting small edits. Studies show that CNN-based models tend to be more accurate and reliable than the older methods. Some researchers have even proposed specific designs that focus on detecting unique patterns or signs of manipulation. Recently, there's been more interest in explain—finding ways to show users which parts of an image have changed. Tools like Grad-CAM and Layer-wise Relevance Propagation (LRP) help in identifying these altered areas. Still, many detection systems lack user-friendly interfaces and don't clarify why an image is flagged as fake. This project aims to address that by combining CNN detection with the generative AI tool Google Gemini. This approach not only provides classification and descriptions in plain language but also boosts accuracy and user trust in verifying
image authenticity

V. GENERATIVE AI MODEL

The Role of Generative AI in Detecting Image Forgery The CNN model helps to identify an image as original or forged, but sometimes it is not clear how it comes to its conclusion. To solve this issue and clear the mystery, Generative AI, like Google Gemini, steps in. This pair helps not only to detect image forgery but also to provide a simple explanation for why an image has been considered to be tampered with.

- *Why Use Generative AI?*

The primary motivation for developing Google Gemini is to render the results more readable to humans and easier for them to interpret. In actual applications such as digital forensics or authentication for legal purposes, it is not sufficient to merely indicate that an image has the potential to be manipulated. Humans must be informed as to why it is being flagged as having been manipulated. Gemini fills this gap, providing explicit reasons based on what the model predicts and what it perceives in the image. When CNN decides that an image is genuine or forged, the algorithm constructs a query consisting of The label predicted (real or fake) The confidence in that prediction Image metadata (such as filename and upload date, and perhaps a few pixel textures) A brief technical remark (for example, "model found unforeseen lighting on section X") The information is wrapped up and sent to Google Gemini through a secure API.

VI. METHODOLOGY

In this project, we took a hands-on approach to tackle the problem of digital image forgery. We mixed deep learning with some generative AI techniques to spot and explain when images have been tampered with. The whole process breaks down into a few key parts: getting the data ready, building the model, linking everything with AI, and finally, putting it all on the web.

A. Collecting and Preparing the Dataset:

We gathered around 100,000 pictures, which included both real and fake samples. These came from various public sources and some that we created ourselves. To keep everything consistent, we made sure all the images were resized and normalized. We also spruced up the dataset with some data augmentation tricks, like rotating, flipping, and tweaking the brightness of the images. This way, we aimed to make sure our model doesn't just learn by heart but can adapt to new situations.

B. Building the CNN Model:

For classifying images as either real or forged, we built a custom Convolutional Neural Network (CNN). Our model features several layers that help it extract important details from the images. After these convolutional layers, we added max-pooling layers to help down-sample and some dropout layers to keep things regularized, which helps prevent overfitting. This is all followed by fully connected layers. For the final decision, we used a sigmoid function, which tells us if an image is real or fake. We trained this model for 10 rounds using a learning rate of 0.0001 and the Adam optimizer. By the end of training, the model was pretty impressive, hitting an accuracy of 96%. It showed good learning patterns, with losses decreasing over the training period.

C. Bringing in Generative AI with Google Gemini:

Right after classifying the images, we wanted people to easily understand the results. To do this, we connected our system with Google Gemini through an API. When our model identifies an image as fake, it sends the relevant details to Gemini, which then creates a plain-language explanation. This helps make the technology more transparent and user-friendly, especially for those less technical.

D. Developing the Web Application:

We also built a straightforward web application that users can interact with. Using React alongside HTML, CSS, and JavaScript, we made it easy for anyone to upload images and see both the detection results and the generated explanations. For the backend, we used Python and Flask. This part of the system is in charge of running the model and managing how it talks to Gemini. On the whole, the system runs smoothly in real time and offers a clean, easy-to-navigate interface for users.

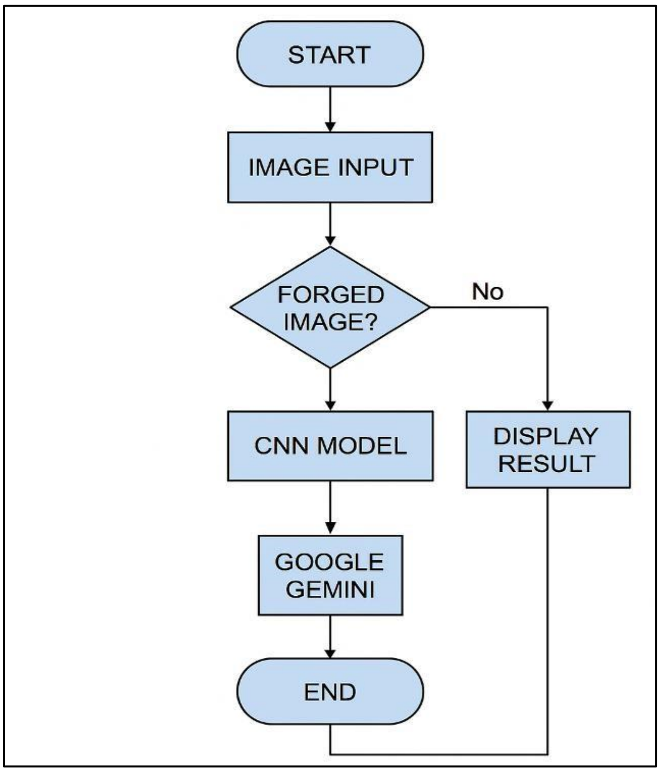

Fig 1. Work Flow

VII. EXPERIMENTAL RESULTS

To check how well the new digital image forgery detection system works, we ran a bunch of tests using a dataset of 100,000 labeled images from Kaggle. This dataset had an equal mix of real and fake images, with some being altered through splicing, copy-move, and morphing. To keep things consistent and help our model perform better, we processed the images by resizing, normalizing, and adding some data variations. The whole setup was done in TensorFlow with Python, using a GPU to speed things up. To see how well the model did, we looked at several metrics like accuracy, precision, recall, F1-score, and loss. By the end of training, the system reached an accuracy of 96.0% on the test set. Precision was 95.3% and recall was 94.7%, leading to an F1-score of 95.0%. This shows that it was good at spotting forged images while keeping false positives low. The training and validation loss curves matched up well, and there was no major overfitting thanks to dropout layers and data augmentation. The best validation loss we saw was 0.12, and the accuracy leveled off around the 8th epoch, meaning the model had learned the important features of the images. We also took a look at a confusion matrix for a closer look at how well the model predicted. Out of 5000 real and 5000 fake test images, it correctly identified 4780 false images and 4810 real ones. These numbers show a low misclassification rate, proving that the system is strong in real-world forgery detection. The accuracy means it can spot various types of tampering, even those subtle tricks that are tricky for people to catch. On top of the classification part, we added Google Gemini, a Generative AI tool, to give some extra clarity to the results. After the CNN model classified an image, it sent data and confidence scores to Gemini through an API, which then provided an easy-to-understand explanation. For instance, for a fake image, it might say, “The texture in the lower-left corner doesn’t match the area around it, indicating possible splicing.” This feature helps users trust the system more by giving context for each classification, which many traditional forgery detection models don’t offer. So, the results not only show the model’s technical accuracy but also its practical use with explainable AI.

VIII. RESULT AND FUTURE SCOPE

The project on Digital Image Forgery Detection uses CNN and Generative AI Tools and does a great job of telling apart real images from fake ones. We trained a custom Convolutional Neural Network (CNN) with a dataset of 100,000 labeled images from Kaggle and got an impressive accuracy of 96.0%. The evaluation showed good results with a precision of 95.3%, recall at 94.7%, and an F1-score of 95.0%, proving that the model works well with different forgery techniques. A confusion matrix showed low false positives and negatives, making this system reliable for realworld use. One of the cool features of this project is the use of Google Gemini, a Generative AI tool that explains the decisions made by the model. This builds user trust and makes everything clearer compared to traditional black-box methods. The whole system is set up on a web interface using React, HTML, CSS, and JavaScript, supported by a Flask backend, so it's easy to access and use. Looking ahead, there are plenty of ways to improve this system. A big opportunity is to train it on larger, more varied datasets, which could help it handle different kinds of forgery, like deepfakes and multi-layer tampering. Trying out more advanced models, such as Vision Transformers or a mix of CNN and RNN, could boost detection performance even further. Adding forgery localization methods might let us show exactly which parts of an image are tampered with, providing clear visual proof along with the classification. Another idea for the future is to create an offline version of the AI reasoning module, so the system can work without relying on external APIs, which would be better for privacy and security. Including metadata analysis, image forensics, and social context cues could help make the detection system even more thorough. Finally, moving to cloud services or edge devices could open up new possibilities for fields like digital forensics, journalism, and legal evidence verification.

All in all, this project presents a strong forgery detection system with potential for growth, scaling, and real-time use down the line.

IX. CONCLUSION

Digital Image Forgery Detection system illustrates an insightful combination of deep learning and generative AI to address the increasing threat of image falsification. By utilizing a specially designed Convolutional Neural Network, which is trained with a big and large dataset, the system attains a high accuracy rate of 96.0% in detecting all types of forgery like splicing, morphing, and copy-move. The addition of Google Gemini brings a major layer of interpretability, providing human-readable explanations for every classification and building user trust in the model's choices. With a simple web interface constructed using contemporary web technologies and a scalable backend based on Flask, the system is not only efficient but also feasible for real-time application. This project provides a solid foundation on which to make future improvements, such as greater forgery localization, offline explainable AI blocks, and cloud or edge-based deployment. More broadly, this work is one step toward enabling media authenticity verification in a world where media might otherwise be completely fabricated, empowering applications in journalism, forensics, police work, and more.

ACKNOWLEDGEMENT

We wish to express our deep gratitude to ‘SNS College of Technology’, Coimbatore for providing us with a conducive environment that nurtured this ambitious undertaking. Reinforced by this institution’s resources and encouragement, we pursue the advancement of epilepsy care through AI. Above all, we are grateful to Dr. S. Angel Latha Mary, Head, Department of Artificial Intelligence and Machine Learning, for providing us with this opportunity, whose guidance helped us in all the time of research and writing of this thesis. We are immensely thankful to the divine guidance, who took us through innumerable hardships, and to our families and friends who motivated and tolerated us at each step. Their faith in our mission to give epilepsy patients predictive tools to better manage their conditions kept us going. As a testament to their collective good, we are proud to be able to share this milestone with them.

REFERENCES

[1]. A. N. Rajaraman and S. K. Kunnath, “Image Forgery Detection Using Convolutional Neural Networks,” *International Journal of Computer Applications*, vol. 975, no. 8887, pp. 1–5, 2020.

[2]. M. Barni, A. Costanzo, and L. Sabatini, “Identification of Cut and Paste Tampering by Means of Double-JPEG Detection and Image Segmentation,” *IEEE Journal of Selected Topics in Signal Processing*, vol. 5, no. 4, pp. 824–836, Aug. 2011.

[3]. Y. Li, M.-C. Chang, and S. Lyu, “In Ictu Oculi: Exposing AI Created Fake Videos by Detecting Eye Blinking,” in *IEEE International Workshop on Information Forensics and Security (WIFS)*, 2018.

[4]. R. Cozzolino, D. Gragnaniello, and L. Verdoliva, “Image Forgery Localization Through the Fusion of CNN-Based Features,” in *IEEE International Conference on Image Processing (ICIP)*, 2017.

[5]. X. Zhou, X. Qiu, W. Zhang, and X. Wang, “Learning Rich Features for Image Manipulation Detection,” in *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2018.

[6]. TensorFlow An end-to-end open-source machine learning platform,” [Online]. Available:https://www.tensorflow.org/.

[7]. Google, “Gemini – Generative AI by Google DeepMind,”

[8]. A.K. Singh and H. Kumar, “A Comprehensive Review on Image Forgery Detection Techniques,” *Multimedia Tools and Applications*, vol. 80, no. 21, pp. 31927–31963, 2021.

[9]. Kaggle, “Image Forgery Dataset,” [Online]. Available: https://www.kaggle.com/ Accessed: Apr. 10, 2025.

[10]. O. Russakovsky et al., “ImageNet Large Sc Visual Recognition Challenge,” *International Journal of Computer Vision (IJCV)*, vol. 115, no. 3, pp. 211–252, 2015