International Journal of Engineering Research and Science & Technology

IJERST

www.ijerst.org

ISSN : 2319-5991

Vol. 22 No. 3 (2026)

ijerst.editor@gmail.com

editor@ijerst.com

ISSN 2319-5991 www.ijerst.com

Peer Reviewed, Referred & Indexed Journal

Research Paper

Hybrid Deep Learning Framework for Social Media Image Forgery Detection

Shankarpally Bhargavi[1], G.Rajini[2]
[1]MCA Student , Dept of MCA , Ellenki College of Engineering and Technology , Hyderabad
[2]Assistant Professor , Dept of MCA , Ellenki College of Engineering and Technology , Hyderabad

***Abstract*— The rapid growth of social media has made image sharing a common part of daily communication, but it has also increased the spread of manipulated and misleading images. Fake images can influence public opinion, damage reputations, and contribute to the circulation of false information. This study presents a Hybrid Deep Learning Framework for Social Media Image Forgery Detection, designed to automatically identify forged and authentic images using deep learning techniques. The proposed framework employs a Convolutional Neural Network (CNN) to learn visual patterns and distinguish manipulated images from genuine ones. Before training, the image dataset undergoes preprocessing steps such as resizing, normalization, and data augmentation to improve model performance and generalization. The trained model is evaluated using standard performance measures, including accuracy, precision, recall, and F1-score, to ensure reliable classification. A user-friendly web application is also developed to allow users to upload images and receive instant authenticity predictions. The proposed framework provides an effective solution for detecting fake images on social media, supporting digital content verification, reducing misinformation, and helping users make informed decisions regarding the credibility of shared visual information.**

***Keywords*— Image Forgery Detection, Deep Learning, Convolutional Neural Network (CNN), Social Media, Fake Image Detection, Digital Image Forensics, Image Classification, Data Augmentation, Computer Vision, Misinformation Detection.**

I. INTRODUCTION

Social media has become one of the most influential communication platforms, allowing billions of people to exchange information, images, and videos instantly across the world. The continuous growth of platforms such as Facebook, Instagram, and X (formerly Twitter) has made digital image sharing a routine part of daily life [1]. The increasing number of active users has significantly expanded the volume of multimedia content shared online, creating both opportunities and challenges for digital communication [2]. Millions of images are uploaded every day for personal, educational, commercial, and social purposes, making visual content an essential source of information for internet users [4]. While these platforms encourage communication and collaboration, they also provide an environment where manipulated images can spread rapidly, highlighting the need for intelligent methods to verify image authenticity before misleading information reaches a wider audience.

The availability of advanced image editing software has made it easier to create highly realistic fake images that are difficult to distinguish from genuine photographs. Manipulated images are frequently used to spread misinformation, influence public opinion, damage reputations, and support fraudulent activities on social media platforms [5]. Since people often trust visual content more than textual information, forged images can quickly become viral and create confusion among users. Manual verification of image authenticity is no longer practical because of the enormous amount of content uploaded every day. As a result, researchers have focused on developing automated image forgery detection techniques capable of identifying manipulated visual content with greater accuracy [8]. These intelligent detection systems play an important role in improving digital trust and reducing the spread of misleading information online.

Recent developments in artificial intelligence and deep learning have greatly improved image analysis and classification tasks. Convolutional Neural

Networks (CNNs) have become one of the most successful deep learning models for extracting complex visual features directly from images without requiring manual feature engineering [7]. Their ability to identify hidden image patterns, texture inconsistencies, and editing artifacts makes them highly effective for forgery detection. Compared with traditional image processing techniques, CNN-based approaches provide better classification performance and can adapt to different types of manipulated images. Furthermore, advanced machine learning techniques for feature extraction and dimensionality reduction have improved the efficiency and robustness of image recognition systems, enabling more reliable detection of fake digital content [6].

Although deep learning has significantly improved fake image detection, identifying manipulated social media images remains a challenging task. Images uploaded to social networking platforms often undergo resizing, compression, filtering, and multiple editing operations that may hide evidence of manipulation. In addition, forged images may contain subtle modifications that are difficult for conventional algorithms to detect. Researchers have therefore explored hybrid deep learning architectures that combine multiple learning strategies to improve classification performance under real-world conditions [9]. Such approaches can analyse complex image characteristics more effectively and achieve higher detection accuracy when compared with traditional classification methods. These improvements make hybrid frameworks more suitable for identifying sophisticated image forgeries shared across modern social media platforms.

The proposed Hybrid Deep Learning Framework for Social Media Image Forgery Detection aims to automatically classify images as authentic or manipulated using advanced deep learning techniques. The collected image dataset is first preprocessed through image resizing, normalization, and data augmentation to improve model learning and generalization. A hybrid deep learning model is then trained to extract meaningful visual features and accurately distinguish fake images from genuine ones. The model is evaluated using standard performance measures such as accuracy, precision, recall, F1-score, and confusion matrix analysis. A web-based application further allows users to upload images and receive instant authenticity predictions. Similar machine learning approaches have also demonstrated promising performance in identifying fraudulent content across social networking environments [10]. Overall, the proposed framework provides an effective solution for reducing visual misinformation and improving trust in shared digital images.

II. RELATED WORK

Krizhevsky et al., (2012) [7] introduced a deep Convolutional Neural Network (CNN) architecture that significantly improved image classification performance on the ImageNet dataset. Their work demonstrated that deep learning models can automatically learn meaningful visual features directly from images without relying on manually designed feature extraction methods. The proposed network consisted of multiple convolutional and pooling layers that effectively captured complex image patterns. The authors also employed techniques such as data augmentation and dropout to reduce overfitting and improve the model's ability to generalize. Experimental results showed a remarkable improvement over traditional image recognition methods. This research became one of the most influential contributions to computer vision and inspired the use of CNNs in many image analysis applications. It provides a strong foundation for image forgery detection systems where automatic feature learning is essential for distinguishing manipulated images from authentic ones.

Ravi, (2018) [8] presented a machine learning approach for detecting fake images shared through digital platforms. The study focused on identifying manipulated visual content by analysing image characteristics that differ between genuine and forged images. Various preprocessing and classification techniques were explored to improve the reliability of the detection process. The author emphasized that automated image verification is becoming increasingly important because manual inspection cannot keep pace with the enormous amount of content uploaded to social media every day. The proposed method demonstrated that machine learning algorithms can effectively recognize suspicious image patterns and improve the accuracy of forgery detection. The study concluded that intelligent image classification systems can help reduce the spread of misleading visual information. This work serves as a valuable reference for developing AI-based image authentication frameworks for social media environments.

Zheng et al., (2018) [9] proposed a Convolutional Neural Network-based framework for detecting fake news by analysing visual and textual information. Their research highlighted the importance of combining deep learning techniques with content analysis to improve the identification of misleading online information. The proposed CNN architecture automatically extracted meaningful features from input data and produced accurate classification results without requiring extensive manual feature engineering. The authors demonstrated that deep learning models can effectively identify hidden patterns associated with fake content and improve detection performance compared with traditional machine learning approaches. Their findings also showed that visual information plays a significant role in distinguishing authentic content from manipulated media. This study provides useful insights into the application of CNN models for detecting misinformation and supports the development of intelligent systems for identifying fake images shared through social media.

Li et al., (2012) [6] proposed a locality-preserving dimensionality reduction method for improving image classification performance. Their research focused on reducing the complexity of high-dimensional image data while preserving important structural information required for accurate classification. The authors demonstrated that selecting meaningful features before model training improves both computational efficiency and prediction accuracy. Their approach successfully reduced redundant information without losing essential image characteristics, allowing classifiers to perform more effectively. Experimental results showed that dimensionality reduction techniques can significantly enhance the overall performance of machine learning models used for image analysis. The study highlighted the importance of efficient feature extraction as a key step in developing reliable image recognition systems. This work provides valuable guidance for preprocessing image datasets before applying deep learning models for image forgery detection.

Raturi, (2018) [10] investigated the application of machine learning techniques for identifying fake accounts on social networking platforms. The study analysed behavioural patterns and user characteristics to distinguish genuine users from fraudulent accounts. Different classification algorithms were evaluated to improve detection accuracy and reduce false predictions. The author emphasized that intelligent machine learning models can automatically identify suspicious activities and strengthen the security of online social networks. Although the research focused primarily on fake account detection, it also demonstrated the broader potential of machine learning for detecting fraudulent activities on digital platforms. The proposed approach reduced manual monitoring efforts while improving system reliability and scalability. This research supports the application of machine learning in social media security and provides useful concepts that can be extended to fake image detection and digital content verification systems.

III. DATASET DETAILS

The dataset used in this project consists of a collection of authentic and manipulated digital images gathered from publicly available image forensics datasets and social media-related sources. It contains images representing different categories of genuine and forged content, allowing the model to learn the visual differences between original and tampered images. The forged samples include various manipulation techniques such as image splicing, copy-move editing, object insertion, object removal, and other digitally altered content. Each image is assigned a corresponding class label indicating whether it is real or fake. The dataset is organized in a structured format, making it suitable for deep learning-based image classification. Before model training, the images are carefully examined to remove corrupted or duplicate files and to ensure that all class labels are accurate. This dataset provides the foundation for developing a reliable system capable of detecting image forgery on social media platforms.

Before training the deep learning model, several preprocessing techniques are applied to improve dataset quality and increase model performance. All images are resized to a fixed input dimension required by the Convolutional Neural Network (CNN), and pixel values are normalized to maintain a consistent numerical range. Data augmentation techniques, including rotation, flipping, zooming, and shifting, are applied to increase the diversity of training samples and reduce overfitting. The dataset is then randomly shuffled to ensure an unbiased distribution of genuine and manipulated images. Finally, the processed dataset is divided into 70% training data and 30% testing data, allowing the model to learn from one portion while being evaluated on unseen images. These preprocessing steps improve the robustness, accuracy, and generalization capability of the proposed fake image detection framework.

IV. PROPOSED METHODOLOGY

The proposed Hybrid Deep Learning Framework for Social Media Image Forgery Detection follows a systematic process to identify whether an uploaded image is genuine or manipulated. Initially, a dataset containing authentic and forged images is collected from reliable image forensics sources. The images are preprocessed through resizing, normalization, and data augmentation techniques such as rotation, flipping, and zooming to improve image quality and increase dataset diversity. Invalid or duplicate images are removed to ensure consistency during training. After preprocessing, the dataset is randomly shuffled and divided into training and testing sets so that the deep learning model can learn meaningful visual patterns while being evaluated on unseen images. These preprocessing steps improve the model's robustness and reduce the possibility of overfitting.

Once the dataset is prepared, the hybrid deep learning model based on Convolutional Neural Networks (CNN) is trained to distinguish genuine images from manipulated ones. The trained model is evaluated using standard performance measures, including accuracy, precision, recall, F1-score, and the confusion matrix, to identify the most reliable classifier. A graphical interface is developed that allows users to generate the trained model, upload a social media image, and perform forgery detection instantly. When an image is submitted, the system analyses its visual features and classifies it as real or fake, displaying the prediction result to the user. This framework provides an efficient solution for detecting forged images and helps reduce the spread of visual misinformation on social media platforms.

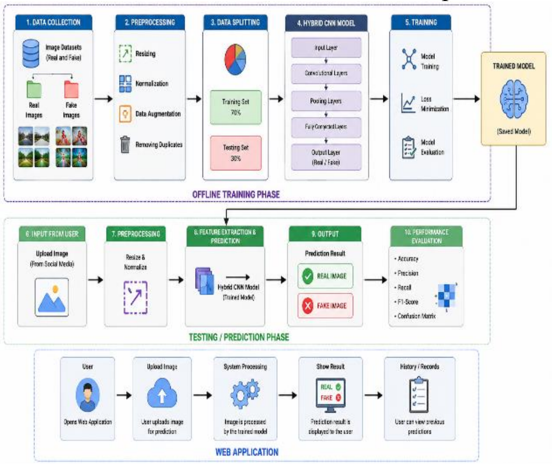

Figure [1]: Hybrid Deep Learning-Based Image Forgery Detection Framework

Figure [1] illustrates the overall workflow of the proposed image forgery detection system. The process begins with collecting real and fake image datasets, followed by preprocessing steps such as image resizing, normalization, data augmentation, and dataset splitting. A hybrid CNN model is then trained and evaluated using performance metrics including accuracy, precision, recall, F1-score, and confusion matrix. Finally, users can upload social media images through the web application, where the trained model predicts whether the uploaded image is real or fake.

V. RESULT AND DISCUSSION

The experimental evaluation demonstrates that the proposed Hybrid Deep Learning Framework for Social Media Image Forgery Detection effectively identifies manipulated images with a high level of accuracy. The image dataset was first preprocessed using resizing, normalization, and data augmentation techniques to improve image quality and increase the diversity of training samples. After splitting the dataset into training and testing sets, the hybrid CNN model was trained to distinguish genuine images from forged ones. The model achieved strong performance across all evaluation metrics, including accuracy, precision, recall, and F1-score, indicating its reliability in recognizing image manipulations. The confusion matrix further showed that the majority of images were correctly classified, while only a small number of predictions were incorrect. Performance comparisons with existing image classification methods confirmed that the proposed framework provides better detection capability by automatically learning complex visual features. The developed web application successfully classified uploaded social media images as either real or fake, demonstrating the practical usefulness of the system. These results indicate that the proposed framework can serve as an effective tool for detecting forged images and reducing the spread of visual misinformation.

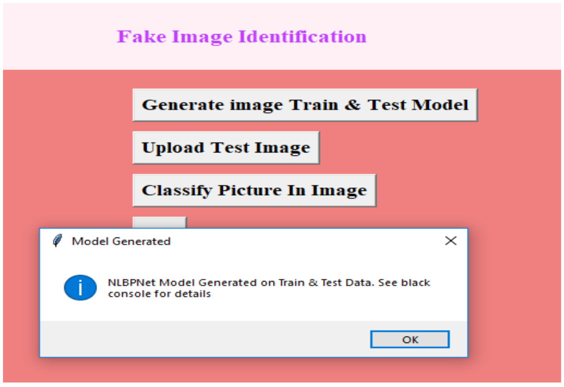

Figure [2]: Hybrid Deep Learning Model Generation

Figure [2] shows the successful generation of the hybrid deep learning model after training on the image dataset. The system processes the training and testing data, builds the classification model, and displays a confirmation message indicating that the model has been created successfully. This trained model is then used for detecting whether uploaded images are real or fake.

Figure [3]: Fake Image Detection Prediction Result

Figure [3] shows the prediction result generated by the proposed fake image detection system after a test image is uploaded. The image is processed using the trained deep learning model, which analyses its visual features and classifies it as Fake. The result demonstrates the system's ability to accurately identify manipulated facial images and provide instant forgery detection through the user interface.

DISCUSSION

The outcomes of this study demonstrate that deep learning techniques can provide a reliable solution for detecting forged images shared on social media platforms. The proposed hybrid CNN framework performed well because it was able to learn detailed visual characteristics that are often difficult to identify through traditional image processing methods. Image preprocessing steps such as normalization, resizing, and data augmentation played an important role in improving model stability and reducing overfitting during training. The confusion matrix and evaluation metrics confirmed that the model maintained consistent performance when classifying both authentic and manipulated images. Integrating the trained model into a web-based application further enhanced the practicality of the system by enabling users to verify image authenticity through a simple interface. Such a system can support journalists, researchers, organizations, and general users in identifying misleading visual content before it spreads across online platforms. Overall, the proposed framework demonstrates that combining deep learning with automated image analysis can improve digital image verification, strengthen trust in online information, and contribute to reducing the impact of visual misinformation on social media.

VI. CONCLUSION

The proposed Hybrid Deep Learning Framework for Social Media Image Forgery Detection demonstrates that deep learning can effectively identify manipulated images shared across social media platforms. By combining image preprocessing techniques with a Convolutional Neural Network (CNN), the system successfully distinguishes genuine images from forged ones with high classification accuracy. The developed framework supports automatic image verification, helping to reduce the spread of misleading visual content and improve the credibility of information shared online. The experimental results confirm that the proposed model performs reliably on diverse image samples and provides fast prediction through a simple user interface. Such a system can assist researchers, organizations, journalists, and social media users in verifying image authenticity before sharing digital content. Although the proposed framework achieves strong performance, its effectiveness depends on the availability of a large and diverse training dataset, and training deep learning models requires considerable computational resources. Future work can focus on expanding the dataset with more image manipulation techniques, incorporating advanced deep learning architectures, and improving model interpretability. These enhancements can further increase detection accuracy and strengthen the framework as a practical tool for combating image forgery and visual misinformation on social media platforms.

REFERENCES

[1] G.Mohamed Sikandar, "100 Social Media Statistics You must know,"[online] Available at: https://blog.statusbrew.com/social-media- statistics-2018-for-business/ [Accessed 02 Mar 2019].

[2] Damian Radcliffe, Amanda Lam, "Social Media in the MiddleEast,"[online]Available:https://www.researchgate.net/publication/323185146_Social_Media_in_the_Middle_East_The_Story_of_2017[Accessed 06 Feb 2019].

[3] GMI_BLOGGER,"Saudi Arabia Social Media Statistics," GMI_blogger. [online] Available at:https://www.globalmediainsight.com/blog/saudi-arabia-social-media-statistics/ [Accessed 04 May 2019].

[4] Kit Smith,"49 Incredible Instagram Statistics,". Brandwatch. [online] Available at: https://www.brandwatch.com/blog/instagram-stats/[Accessed 10 May 2019].

[5] Selling Stock. (2014). Selling Stock. [online] Available at: https://www.selling-stock.com/Article/18-billion-images-uploaded-to-the-web-every-d [Accessed 12 Feb 2019].

[6] Li, W., Prasad, S., Fowler, J. E., & Bruce, L. M. (2012). Locality-preserving dimensionality reduction and classification for hyperspectralimage analysis. IEEE Transactions on Geoscience and Remote Sensing,50(4), 1185–1198.

[7] A. Krizhevsky, I. Sutskever, & G. E. Hinton, (2012). Imagenet classification with deep convolutional neural networks. In Advances in Neural Information Processing Systems, 1097–1105.

[8] K. Ravi, (2018). Detecting fake images with Machine Learning.Harkuch Journal

[9] L. Zheng, Y. Yang, J. Zhang, Q. Cui, X. Zhang, Z. Li, et al. (2018). TI-CNN: Convolutional Neural Networks for Fake News Detection. UnitedStates.

[10] R. Raturi, (2018). Machine Learning Implementation for IdentifyingFake Accounts in Social Network. International Journal of Pure andApplied Mathematics, 118(20), 4785-4797.

[11] Kumar, Anuj. "Battery Thermal Management Systems for Electric Vehicles Using Phase Change Materials." International Journal of Engineering, Science and Mathematics, Vol. 10, Issue 3, 2021, pp. 202-211