Detecting Digital Image Forgeries with Copy-Move and Splicing Image Analysis using Deep Learning Techniques

Divya Prathana Timothy, Ajit Kumar Santra*
School of Computer Science Engineering and Information Systems
Vellore Institute of Technology, Vellore, Tamil Nadu, India

***Abstract*—The proliferation of digitally altered images across social media platforms has escalated the urgency for robust image forgery detection systems. Traditional detection methodologies, while varied, often fall short in addressing the multifaceted nature of image forgeries in the digital landscape. Recognizing the need for advanced solutions, this paper introduces a novel deep-learning approach that leverages the architectural strengths of GNNs, CNNs, VGG16, MobileNet, and ResNet50. Our method uniquely integrates these architectures to effectively detect and analyze multiple types of image forgeries, including image splicing and copy-move forgeries. This approach is groundbreaking as it adapts these networks to focus on identifying discrepancies in the compression quality between forged and original image regions. By examining the differences between the original and compressed image versions, our model constructs a feature-rich representation, which is then analyzed by a tailored deep-learning network. This network has been enhanced by removing its original classifier and implementing a new one specifically designed for binary forgery classification. Very few researchers have explored the application of deep learning techniques in copy-move and splice image analysis for detecting digital image forgeries, making our work particularly significant. A comprehensive comparative analysis with pre-trained models underscores the superiority of our method, with the GNN model achieving an impressive accuracy of 98.54 percent on the CASIA V1 dataset. This not only sets a new benchmark in the field but also highlights the efficiency of our model, which benefits from reduced training parameters and accelerated training times.**

***Keywords*—Copy-move; splicing; deep learning; image forgery detection**

I. Introduction

In the digital age, the authenticity of photos shared on platforms like Facebook and Twitter has become a major concern. The manipulation of digital images poses a threat to the integrity of visual information, creating discrepancies from the original characteristics and features of the images. This kind of forgery often goes unnoticed and contributes to the spread of false news and misinformation. Advanced image tampering technologies like GNU, GIMP, and Adobe Photoshop have aggravated this problem [1], [2]. There are active and passive ways to overcome above-mentioned issues. Active detection uses means or median to implant a digital signature or message digest into an image during creation. The image's validity is verified by decrypting this data. However, passive detection methods modify an image's statistical features to verify its structure and content without leaving visual indications. Copy-move, splicing, and retouching forgeries are examples of passive methods. Picture splicing and copy-move forgeries are highlighted in the former [3]. While image splicing connects two or more images, copy-move forgery copies a portion of an image within the same image. Due to the similar characteristics of duplicated pieces and different post-processing techniques like rotation and JPEG compression, copy-move forgeries are puzzling to identify. However, splicing forgeries incorporates pieces from several photos, needing extra processing to match the target image's visual features.

Traditional detection methods in this area use frequency domain attributes or statistical information to identify authentic and counterfeit pictures [4]. These approaches' principal drawback is the difficulty of determining the most important traits for counterfeit detection. Digital image forgery is a major problem in our digital era. Technological developments in manipulation need increasingly advanced approaches to recognize and battle picture forgeries.

Copy-move and splicing forgeries are particularly challenging to locate and identify [5]. Localization locates counterfeit portions in a picture, whereas forgery detection verifies its validity. Many methods have been developed to solve these issues separately [6]. These approaches must be tested for robustness, dependability, and correctness, especially in modelling structural changes caused by copy-move and splicing forgeries. Since most imaging equipment cannot contain signatures or watermarks to prove authenticity, passive or non-intrusive forgery detection methods are needed. These algorithms don't need picture content signatures or watermarks.

Deep learning (DL) has advanced the image forensics profession. Any DL model like a convolutional neural network (CNN) relies on feature extraction, where database size matters. In small database sizes, transfer learning such as AlexNet, MobileNet, VGGNet, and ResNet are effective. It applies information from a huge dataset like ImageNet to a new target domain. This method reduces training time and lets the model cope with fewer datasets. With these technologies, the fight against digital picture counterfeiting is growing more advanced, offering better digital authenticity [7]. The current landscape of digital image forensics lacks robust and comprehensive methods for simultaneously detecting both image splicing and copy-move forgeries.

The motivation behind this research is the vast number of manipulated images circulating daily online, making it difficult

*Corresponding author, email: ajitkumar@vit.ac.in

to verify their authenticity. From medical images to biometric data, nothing seems safe from manipulation nowadays. A framework that not only identifies but also localizes the forgery in an image is needed.

This research significantly advances the field of image forensics by investigating DL models for accurate forgery localization and classification. Our study meticulously analyzes DL and transfer learning architectures, including Graph Neural Network (GNN), CNNs, VGG16, MobileNet, and ResNet50, specifically targeting copy-move and splicing image forgeries. The findings highlight the exceptional accuracy of the GNN model and illustrate the robust potential of these architectures in the domain of digital forensics.

Our suggested approach presents numerous enhancements compared to traditional detection methods. The following are the highlights of the paper's primary contribution:

- Focuses on a copy-move and splicing type of forgeries.
- Leverages the strengths of diverse architectures such as CNN,GNN and pre-trained models.
- Tailored deep learning network,specifically designed for binary forgery classification.
- Reduced training complexity by leveraging pre-trained architectures.

To articulate the structure and flow of our paper clearly, we have organized it into coherent, well-defined sections, each designed to progressively build upon the information presented. The paper begins with an introduction that establishes the significance of the research and outlines the challenges and innovations in image forgery detection (IFD). This is followed by a comprehensive literature review that situates our work within the broader academic discourse, identifying gaps that our research aims to fill. We then detail our novel methodologies, which introduce unique analytical techniques and leverage advanced DL models to address the complexities of image forgery. The experimental design section describes the dataset used and the parameters of our testing framework, ensuring reproducibility and clarity in our methods. Following this, the results and discussion section critically assesses the performance of the proposed models, providing a deep dive into the empirical evidence that supports our conclusions. Finally, the paper concludes with a summary of our findings and offers a forward-looking perspective on potential future research avenues and technological advancements in the field of IFD. Each section of the paper is integral to the narrative, contributing to a comprehensive and educative exposition tailored for both specialists and novices in the field.

II. Related Work

In recent years, there has been progress in the detection of image forgeries, with several methods proposed by researchers. Traditionally, this field extracted handmade characteristics and classified them using feature matching to identify real and counterfeit pictures. While successful, these strategies lack flexibility and scalability.

Recently, researchers have tried to identify copy-move and splicing frauds concurrently. A new approach uses a fully convolutional network with multi-resolution hybrid features [8]. Tamper-guided dual self-attention module in this network distinguishes tampered regions from unaffected ones. For pixel-level picture fraud detection, the hybrid features and semantic reinforcement network (HFSRNet) uses LSTM encoding-decoding [9]. Next, U-Net, a unique picture segmentation model with L2 regularization is used for IFD [10]. In another research double image compression was employed to train a model that could recognize both kinds of forgeries. These advances in picture fraud detection show a strong trend toward DL such as CNNs [11]. These approaches can identify and localize fabricated portions in photos, making them a more effective solution to digital image forging. As these technologies advance, they will help preserve digital pictures in forensic science, media, and other fields.

A multimodal approach was presented to identify splicing and copy-move forgeries using deep neural networks to classify and localize forgeries and part-based picture retrieval. This system utilizes InceptionV3 for feature extraction and the Nearest Neighbor Algorithm for donor and nearly duplicate picture retrieval. Error Level Analysis (ELA), VGG16, and VGG19 models were used on CNN in another unique way [12]. This approach employed pre-processing to collect pictures at a certain compression rate to train the model to categorize photos as legitimate or fake. These transfer learning-powered IFD advances improve digital picture forgery detection and localization. Pre-trained models and advanced algorithms improve accuracy and efficiency, creating a new benchmark in digital picture manipulation detection. As technology advances, these approaches will be developed, strengthening digital imaging fidelity in numerous sectors [13]. A different work utilized a CNN pre-trained on labelled pictures to extract features and train an SVM model [14]. This showed how CNNs and SVMs work together in feature extraction and classification. Mask R-CNN with the Sobel filter [15] improved forgery detection and localization by identifying gradients like genuine masks.

Another method [16] used image manipulation and pre-trained CNNs to classify pictures as legitimate or fake, improving transfer learning. It used ELA for image modification and pre-trained VGG-16 weights for CNN initialization. Although DL techniques have improved the IFD, very few research focused on the combination of copy-move and splicing forgeries. Moreover, the potential advantage of combining several DL and transfer learning techniques has been left unexplored.

It is crucial to delineate the boundaries of prior approaches used to identify photo fraud and explain how our proposed strategy differs to address the existing gaps in the ongoing discussion. Although feature matching and manual characteristic extraction are successful, they lack flexibility and can not handle the increasing complexity and volume of digital image alterations efficiently. While each solution is creative, they individually focus on either copy-move or splicing forgeries and do not possess the adaptability required to tackle emerging forms of digital fraud. Our research presents a comprehensive framework for analyzing copy-move and splicing forgeries, which have seldom been examined in conjunction. CNNs and transfer learning enhance forgery detection and localization by using advanced DL models, resulting in improved accuracy and dynamic capabilities. Our multimodal approach incorporates advanced algorithms such as the tamper-guided dual self-

attention module and hybrid feature systems, resulting in enhancements over earlier methods. Enhancements enhance the accuracy of detection and augment the knowledge and skills in critical areas such as forensic science and media integrity. This succinct elucidation of the subject matter and our approach emphasizes our distinctive contributions to the detection of digital image counterfeiting.

III. Proposed Methods

Our research presents an architectural framework that can precisely detect, identify, and assess the degree of forgeries to tackle the complex issues of copy-move and splicing manipulations in digital photos. With the use of sophisticated bounding boxes and semantic segmentation techniques, this novel method is specially designed to identify and accurately localize forged regions, guaranteeing that every pixel inside an area of concern is carefully classified. Unlike other systems, this one can identify areas that have been altered, but it can also determine the proportion of the image that has been altered, providing a numerical assessment of the degree of fabrication.

A digital image is fed into the model to begin the process, which is then followed by a reliable feature extraction stage. The next step is to identify possible regions of interest that could include faked or changed objects using the Region Proposal Network (RPN). To guarantee consistency in analysis, these selected regions which range in size to 128 pixels are standardized via Region of Interest (ROI) pooling. During the next detection stage, the system marks each object it finds as copied or spliced and uses exact bounding boxes to define the forged object. This stage is critical in identifying the type of counterfeit and offers a good understanding of the manipulation method used. Our strategy's last phase, segmentation, is especially creative. By creating a precise mask around the manipulated object, it successfully separates the manipulated region from the original image. The suggested design determines the percentage of the image area impacted by forgery to measure the various levels of forgery. To do this, one must analyze the segmentation masks, which are binary pictures in which the unaffected black backdrop is marked as false, and the fabricated portions are marked as true white. We determine the percentage of the image that has been compromised by forgery by counting the number of white pixels inside these masks.

This technique, which aggregates the white pixel count across all masks, enables reliable assessment even in photos containing several forged areas. The architecture is demonstrated in Fig. 1, which highlights our system's all-encompassing approach to forgery detection and localization. Through a seamless approach that combines feature extraction, object identification, and pixel-wise segmentation, our model cannot only recognize and categorize forgeries but also provide an objective indicator of their size. This development in digital image forensics provides a reliable method for confirming the authenticity of images in a range of applications, marking a substantial achievement in the battle against digital image tampering.

The proposed methodology harnesses CNN's power to process high-resolution images across multiple channels, capturing the nuanced spatial and color information crucial for

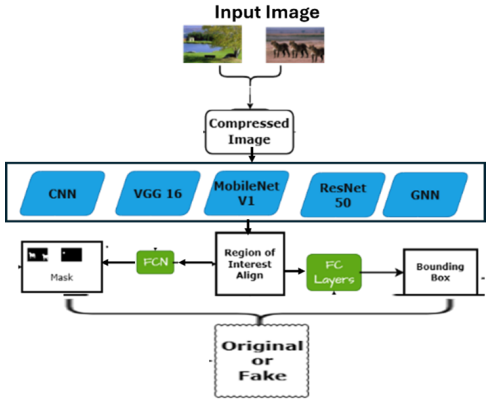

Fig. 1. Proposed frameworks.

detecting subtle manipulations in forged images. The architecture of our CNN consists of an input layer designed to accommodate high-resolution imagery, enabling the extraction of detailed spatial and color features. As the image progresses through CNN, it encounters a series of convolutional layers equipped with specialized filters. These filters are good at seeing spatial linkages and local patterns, which is important for identifying real from altered areas in a picture.

This model employs convolutional layers (Conv2D) and max pooling layers (MaxPooling2D) for feature extraction. Our proposed work embarks on refining the capabilities of CNNs to address the intricate task of detecting digital image forgeries, with a focus on copy-move and splicing forgeries. Recognizing the computational demands and the challenges in designing an optimal CNN architecture, we propose a strategic approach to streamline the process, ensuring efficiency and accuracy in forgery detection.

The CNN architecture includes convolutional layer with a limited number of filters (32 filters with a size of 3x3). Every filter is assigned a specific area of the input image to scan, which enables a thorough examination of the image's color and spatial details. The Rectified Linear Unit (ReLU) function is used to activate these convolutional layers, adding the required non-linearity to the model, and improving its capacity to detect subtle forgeries. Our model uses max pooling layers after the convolutional layers in order to decrease the processed image's spatial dimensions. This reduction is pivotal, as it not only diminishes computational load but also preserves the most salient features essential for accurate forgery identification. The culmination of convolutional and pooling layers yields a compact representation of the image, which is then untraveled into a one-dimensional vector.

In tackling the computational intensity and the architectural optimization challenges, our proposed work leverages advanced techniques in model optimization, regularization, and efficient computational strategies. This includes the exploration of transfer learning as a means to capitalize on pre-trained models for initial feature extraction, significantly reducing the requirement for large, labelled datasets and computational resources.

Despite their effectiveness, CNNs pose significant challenges, notably the requirement for extensive labelled datasets

to train the models adequately. We avoid this by implementing sophisticated techniques that maximize training effectiveness and improve the model's capacity to learn from sparse data sets. This includes data segmentation techniques to artificially expand the training dataset and transfer learning approaches to leverage pre-trained models for feature extraction.

This work analyzes pre-trained transfer learning models such as VGG16, MobileNet V1, and ResNet-50 in parallel. Our proposed work aims to set a new benchmark in the detection and classification of digital image forgeries, providing a robust tool against the proliferation of manipulated media.

In the context of detecting digital image forgeries, GNNs offer a proposed approach by treating the problem as one of analyzing and learning from a graph of image features and their relationships. A GNN operates by learning representations for nodes (which could represent image segments or features) in a way that the representation of a node is informed by its neighbours. This process iteratively aggregates and transforms neighbour information, allowing each node to have a representation that captures both its local features and its context within the larger structure. This method is especially useful in the detection of picture forgeries since it can be important to comprehend the context and interaction between various image components in order to spot irregularities that may indicate manipulation. The key to GNN's effectiveness lies in its image segment framework, where nodes exchange segment information along edges, gradually updating their states based on both their attributes and the segment information received from their neighbours. This allows GNNs to propagate and refine feature information across the graph, leading to rich, contextualized node embeddings that reflect the structure of the data. For IFD, this means that GNNs can help uncover subtle, complex patterns of manipulation that might not be apparent when considering image regions in isolation. In our proposed architecture for advancing digital IFD, we envision leveraging GNNs to analyze the graph of relationships between image segments. By constructing a graph where nodes represent segments of an image and edges encode relationships such as spatial proximity or similarity in texture or color, GNNs can be used to identify irregularities and discrepancies in the graph structure that could indicate fraud. For example, in copy-move forgery, duplicated segments might exhibit unusually high similarity to non-adjacent regions, a pattern that GNNs can be trained to recognize. Relying solely on the original GNN goal might lead to the creation of another graph/subgraph that falls short of elucidating the GNN's reasoning. To craft explanations that embody both accuracy and concreteness, we've refined the generative component's goal function.

While CNNs excel at extracting local visual features from images, GNNs can enhance the analysis by considering the broader context and relationships among these features. The individual performance of each model and the suggested GNN model performance is discussed in result section. By using the DL approach, digital picture forgery detection systems could have much higher accuracy and resilience. The incorporation of GNN into our proposed work represents a promising direction for enhancing the detection and analysis of digital image forgeries.

To address the critique regarding the theoretical foundation of our results, it is crucial to clarify the mathematical underpinnings that substantiate the efficacy of our proposed methods in digital IFD. The effectiveness of our architecture is not merely an isolated occurrence but is grounded in the well-established principles of CNNs and GNNs, both of which are renowned for their robust performance in pattern recognition and feature extraction tasks. The mathematical models for CNN involve convolution operations that leverage learned filters to identify and enhance salient features within images, which are crucial for detecting subtle forgeries. Similarly, the GNN framework is based on the principle of node feature aggregation, where the representation of each node (or image segment) is iteratively refined based on its neighbours, thus capturing both local and global contextual information effectively. Our results are derived from rigorous empirical testing and validation against benchmark datasets, ensuring that the observed high performance is replicable and consistent across various scenarios. Furthermore, by integrating these networks, our approach benefits from the synergy between CNNs' ability to extract detailed local features and GNNs' capacity to analyze relationships within the data structure, which is mathematically supported by the operations of graph convolution and pooling. This combination allows for more comprehensive and precise detection of digital forgeries than would be possible using either technique alone.

A. Data Set and Experimental Setup

The CASIA V1 dataset stands as a pivotal resource in the field of digital IFD, offering a comprehensive collection of images specifically curated for the classification and analysis of various forms of image tampering. Comprising 1,754 images, the dataset is meticulously organized into three distinct categories: 800 authentic images, serving as a baseline for comparison; 480 images subjected to copy-move forgery, and 474 images manipulated through splicing. The authentic images in CASIA V1 provide a wide range of scenes, subjects, and lighting conditions, establishing a robust foundation for models to learn the characteristics of genuine, untampered images. This diversity ensures that the dataset can challenge and evaluate the performance of digital forgery detection systems across a variety of scenarios, making it a valuable asset for developing and testing algorithms designed to discern the subtleties between authentic and forged content. The dataset's copy-move fabricated images are created using a range of sophisticated approaches, including scale, rotation, and different JPEG compression levels to mask the forging. Similarly, the spliced images within CASIA V1 are constructed by combining elements from multiple sources, creating composite images that can be particularly challenging to analyze.

The experimental setup for our research, leveraging the computational power of google colab. The experiments were facilitated by a robust hardware configuration including an NVidia Tesla K80 GPU, which boasts 2,496 CUDA cores and 16GB of GDDR5 VRAM, providing the necessary computational prowess for DL tasks. The processing unit was complemented by a hyper-threaded single-core Xeon Processor @2.3Ghz, equipped with 16 GB of RAM, ensuring efficient data handling and processing speed.

IV. Results

A. Performance Measures

The suggested GNN model's performance is evaluated with the individual model performance using a wide range of metrics, such as the F1-score, accuracy, recall, and precision. With the use of these metrics, we were able to carefully assess how well the model identified manipulated photos. In particular, precision examined the model's accuracy in the cases it identified as forgeries, recall demonstrated the model's capacity to recognize manipulated images, the F1-score gave a fair assessment of both precision and recall, and accuracy gave a comprehensive picture of the model's overall performance. The formulas of performance metrics are mentioned below:

$$Precision = \frac{TruePositive}{TruePositive + Falsepositive} \quad (1)$$

$$Recall = \frac{TruePositive}{TruePositive + FalseNegative} \quad (2)$$

$$F1 - score = \frac{2 * (Precision * Recall)}{Precision + Recall} \quad (3)$$

$$Accuracy = \frac{TruePositive + TrueNegative}{TotalPrediction} \quad (4)$$

B. Training and Validation Insights

The accuracy curves underscored the models' capacity to learn from the training data effectively, while the validation curves provided crucial information about the models' generalizability. Notably, the divergence or plateauing of validation curves from the training curves signified potential overfitting, prompting us to halt training to preserve the models' ability to generalize.

C. Dataset Division and Validation

We were able to prevent overfitting, optimize model performance, and fine-tune hyperparameters by carefully splitting the dataset into training and validation sets at an 80:20 ratio, which kept the models reliable and useful in practical situations. Throughout the training, an early stopping mechanism was used to keep track of training and validation losses. We were able to quickly stop training when this method let us detect when the models started to overfit the training set. This strategy significantly enhanced our models' generalization capabilities, ensuring they remained effective and reliable in detecting digital image forgeries. Our studies' outcomes highlight the possibility of using GNN in the field of digital IFD. Our thorough analysis shows that these models may be improved to identify complex forgery methods with excellent recall, accuracy, and precision, providing useful resources for the digital forensics community. Through tackling the issues of overfitting and fine-tuning model architectures, we have established the foundation for next investigations that seek to improve the identification and categorization of digital image forgeries. Fig. 2 showcases the comparative performance of various DL models applied in the field of digital IFD.

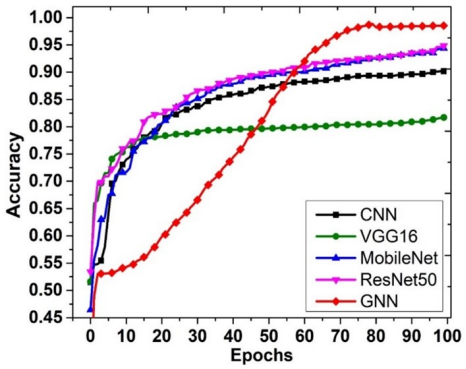

Fig. 2. Accuracy for various methods for IFD.

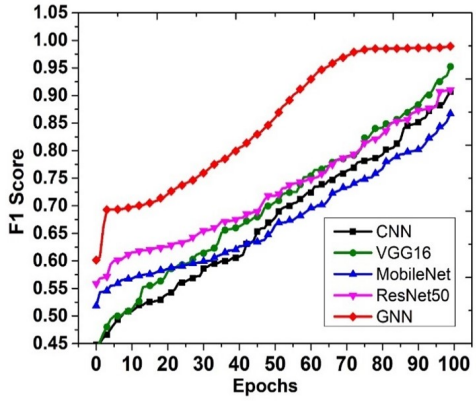

Fig. 3. F1 score for various methods for IFD.

D. Performance Analysis

The graph illuminates an upward trajectory in detection accuracy, signifying substantial strides in model efficiency and reliability. Notably, the GNN model exhibits remarkable improvement, underscoring the effectiveness of advanced architectures in discerning complex patterns within image data. The progression from traditional CNN to more intricate systems like VGG16, MobileNet, and ResNet50 indicates a consistent enhancement in accuracy. For instance, the leap from CNN's initial accuracy of approximately 51.51 percent to GNN's commencement point at about 28.14 percent may appear as a decline. However, GNN's rapid ascension to over 98 percent accuracy by the 99th epoch delineates a significant leap in performance. Such an advancement underscores the transformative impact of transfer learning and the layered sophistication it brings to image analysis tasks. The MobileNet and ResNet50 models also display a steady climb in accuracy percentages, reaching the high 80s and low 90s, respectively. This gradual increase corroborates the hypothesis that depth and complexity in neural networks, managed adeptly, can yield superior results in detecting nuanced manipulations in digital imagery. Fig. 3 delineates a compelling narrative of progressive improvement across diverse DL models throughout 100 epochs. The data traces the F1 scores—a harmonic mean of precision and recall, considered a more robust measure than accuracy alone—of five models: CNN, VGG16, MobileNet, ResNet50, and GNN.

The GNN architecture, which started at an F1 score of approximately 60.13 percent, shows an impressive ascent, culminating at nearly 98.93 percent. This trajectory highlights the efficacy of GNN in the nuanced detection of image forgeries, likely due to its ability to model complex patterns and relationships within the image data. When we compare the increment from CNN's initial F1 score of roughly 44.79 percent to GNN's ending score, it is evident that there's an absolute improvement of around 54 percent. Such a stark progression implies that GNNs are significantly more adept at handling the intricacies of IFD tasks. VGG16 and MobileNet also exhibit substantial enhancements, with VGG16 starting at about 44.31 percent and closing at 95.25 percent, and MobileNet commencing at 51.86 percent and concluding at nearly 86.70 percent. These increases suggest that depth in network architecture can lead to improved feature extraction, which is critical in differentiating between genuine and forged pixels. ResNet50's performance, initiating at an F1 score of 55.87 percent and reaching 91.01 percent, further corroborates the advantages of leveraging deeper networks with residual connections to enhance learning from image data.

The Table I presents a comparative analysis of five advanced DL models—CNN, VGG16, MobileNet, ResNet50, and GNN—across a spectrum of performance metrics including Accuracy, F1 Score, Precision, and Recall throughout 100 epochs. After a hundred epochs, the GNN model emerges as the front-runner, boasting an accuracy and F1 Score of approximately 98.55 percent and 98.93 percent, respectively. This performance is particularly noteworthy when considering its recall rate reached a perfect score, a clear indication of its superior ability to identify true positive cases of forgery. The GNN's precision score, standing at roughly 98.70 percent, reinforces its status as the most reliable method among those tested. VGG16 also shows remarkable results, with an accuracy of about 81.70 percent and an F1 Score of 95.25 percent, a significant leap from its initial F1 Score of approximately 44.31 percent. This demonstrates a solid balance between precision and recall, highlighting VGG16's proficiency in classifying forged image content. MobileNet, known for its efficiency on mobile devices, achieves a notable accuracy of around 94.37 percent and an F1 Score of 86.70 percent. These figures represent its robustness in the context of IFD, particularly in environments where computational resources are limited. ResNet50, with its deep residual learning framework, attains an accuracy of nearly 94.85 percent and an F1 Score of 91.01 percent, underscoring the strength of deep networks in extracting nuanced features that are crucial for identifying forgeries. The CNN model, while not outperforming the GNN, still demonstrates substantial growth from an accuracy of 51.51 percent to 90.17 percent and an F1 Score increase from 44.79 percent to 90.72 percent.

The Table II indicates that the proposed methods outperform many of the traditional approaches, suggesting the superiority of GNNs in capturing the complex relational and structural dependencies characteristic of image forgeries. By combining these cutting-edge neural networks with the CNN base layer, the model's capacity to accurately distinguish between real and fake images is improved. This combination is particularly effective for feature extraction and classification.

TABLE I. VARIOUS MEASURES OF THE VARIOUS METHODS WITH DIFFERENT EPOCHS FOR THE DETECTION OF IMAGE FORGERY

Method	Epochs	Accuracy	F1 Score	Precision	Recall
CNN	0	0.5151	0.447951	0.370166	0.36413
	50	0.8732	0.678501	0.410526	0.429348
	100	0.9017	0.90721	0.524217	1
VGG16	0	0.5158	0.443105	0.443662	0.342391
	50	0.7971	0.706724	0.493776	0.61413
	100	0.817	0.952479	0.521127	0.695652
MobileNet V1	0	0.4643	0.518566	0.422906	0.402708
	50	0.8922	0.647673	0.46535	0.476089
	100	0.9437	0.867047	0.580204	0.560648
ResNet50	0	0.5341	0.55867	0.53202	0.434783
	50	0.8978	0.717949	0.55618	0.608696
	100	0.9485	0.910085	0.620253	0.798913
GNN	0	0.281439	0.601307	0.53023	0.233751
	50	0.810762	0.845913	0.786547	0.850627
	100	0.98549	0.989284	0.986976	1

V. DISCUSSION

It elucidates the pivotal contribution of transfer learning in advancing the detection of digital image forgeries. Harnessing the analytical might of pre-trained neural networks, fine-tuned for the nuanced task of forgery detection, this study showcases the potential to benefit from DL's power without necessitating substantial labelled forensic data. This prudent approach streamlines resource expenditure and forges new pathways for the development of robust solutions against the scourge of digital forgery. In an era where the authenticity of digital content is under constant scrutiny, the study accentuates the superior performance of GNN. GNN's architectural design is adept at mapping the intricate relational and structural nuances that are crucial for identifying forged elements in images. The empirical evidence presented underscores the sophistication of GNNs in discerning subtle discrepancies that allude to tampering, thus bolstering the integrity of visual information. It reflects the remarkable ingenuity integrated within these systems, empowering them to scrutinize layers of digital data to authenticate its veracity. The discernible improvements in accuracy and reliability not only corroborate the current direction of research within this sphere but also lay a solid foundation for the future of digital forensics. The implications of these advancements extend beyond the academic, offering a beacon of trust and reliability as we navigate through an age rife with digital manipulation. Table II showcases a comparative analysis of different methodologies employed for the detection of passive image forgery mainly copy-move and splicing. The forgery type column has four section which include splicing forgery, either copy-move or splicing forgery, copy-move forgery, and both copy-move and splicing forgery. The table examining the efficacy of various feature extraction and classification techniques across several well-recognized datasets. The table delineates the performance of these methods primarily in terms of accuracy except in one instance where precision, recall, and F1-score highlights the evolution and refinement of detection capabilities.

In earlier research, traditional CNNs served as the foundation for feature extraction, with methods varying from CNN-based local descriptor construction to hybrids that integrate

TABLE II. COMPARE VARIOUS METHODS FOR THE DETECTION OF IMAGE FORGERY WITH OUR PROPOSED METHOD

Forgery Type	Reference	Feature Extraction Methods	Classification Methods	Dataset	Evaluation
Splicing	[17]	CNN-Based Local Descriptor Construction	SVM	CASIA V2, DVMM DSO-1.	CASIA V2: 96-97 percent, DVMM: 94 percent, DSO-1: 97.5 percent
	[18]	CNN	CNN	CASIA V1 and V2, CUHK, NIST16, COVERAGE, CUISDE.	CASIA V1: 91 percent, CASIA V2: 99 percent, CUHK: 95 percent, NIST16: 98 percent, COVERAGE: 97 percent, CUISDE: 100 percent
Splicing, Copy-Move Separately	[8]	RGB stream + noise stream	End-to-end fully CNN + (TDAS)	CASIA, COLUMBIA, NIST16	CASIA V1: 98-97 percent, COLUMBIA: 97.4 percent, NIST16: 86 percent
	[9]	Hybrid Encoding+ Decoding CNN	Hybrid features and semantic reinforcement network	NIST16, CASIA V1	NIST16: 98.68 percent, CASIA V1: 92.76 percent, COVERAGE: 91.21 percent
Copy-Move	[19]	DCNN	SD-Net: (super-BPD) + DCNN	USCISI, CoMoFoD, CASIA V2.	CoMoFoD: P=59.11, R=57.62, F=57.77. CASIAV2: P=90.48 R=51.25 F=48.06
	[20]	CNN (Encoder+ decoder)	CNN	CoMoFoD, CMFD.	CoMoFoD: 98.39 percent, CMFD: 97.78 percent
	[21]	Regularizing CNN	Regularizing U-Net	MICC-F2000.	97.52 percent
Splicing + Copy-Move	[22]	DCT	SVM	CASIA V1	96 percent
	[23]	DCT and LBP	SVM with Radial Basis Function (RBF)	CASIA V1	97.5 percent
	Proposed	VGG 16 + MobileNet V1 + ResNet50	GNN + CNN	CASIA V1	98.54 percent

encoding and decoding processes. Classifications were performed using a variety of techniques, including SVM and CNNs themselves, among others.

The proposed method in the current paper pivots from these traditional approaches by integrating CNNs with advanced neural network architectures like ResNet and VGG16, alongside GNNs and MobileNet. These methods are applied to the CASIA V1 dataset. Notably, the proposed GNN method achieves a remarkable accuracy of 98.54 percent, which is significantly higher than many previously referenced methods. Similarly, the combined use of MobileNet and ResNet50 yields an impressive accuracy of 94 percent.

VI. CONCLUSION

The conclusion of our study underscores the significant advancements made with GNNs in the field of digital IFD. GNNs have demonstrated exceptional proficiency, achieving accuracy rates that exceed 98 percent in identifying digital forgeries. This impressive performance is not just a testament to their capability but also showcases their potential as critical tools in digital forensics. However, it is essential to ground these findings within a theoretical framework to fully articulate the scientific contribution of our research. The effectiveness of GNNs in our study is anchored in their inherent ability to process and analyze complex patterns through node and edge analyses, which are particularly effective in understanding and identifying manipulated image data. This theoretical underpinning is supported by the structure of GNNs, which integrates node information with neighbourhood data, allowing for a DL model that is highly adept at detecting anomalies indicative of digital tampering.

Looking forward, we aim to enhance the precision and computational efficiency of these models. Our future research will expand the variety of training datasets to include a wider array of forgery techniques, which will further test and improve the robustness of our models. Additionally, we plan to explore the integration of GNNs with other DL architectures through transfer learning, which could lead to even more powerful systems capable of combating advanced forgery methods. The increasing complexity of digital forgeries requires that our forensic methods evolve concurrently. The ultimate goal of our research is to develop a comprehensive suite of forensic tools that are sophisticated yet user-friendly enough for public use, ensuring that digital media can be authenticated across various platforms. This commitment supports the integrity of information within our digital society and contributes to the maintenance of truth in visual media. As this study lays a solid foundation with high accuracy rates, it paves the way for a future where digital forensic science is an effective guardian against the intricacies of digital forgery, ensuring the authenticity of digital media in an era where truth is paramount.

REFERENCES

[1] D. K. Sharma, B. Singh, S. Agarwal, L. Garg, C. Kim, and K.-H. Jung, "A survey of detection and mitigation for fake images on social media platforms," *Applied Sciences*, vol. 13, no. 19, p. 10980, 2023.

[2] S. Bourouis, R. Alroobaea, A. M. Alharbi, M. Andejany, and S. Rubaiee, "Recent advances in digital multimedia tampering detection for forensics analysis," *Symmetry*, vol. 12, no. 11, p. 1811, 2020.

[3] D. R. Pierce, "Social media lessons on the nature of political decision making," in *Oxford Research Encyclopedia of Politics*, 2020.

[4] K. Asghar, Z. Habib, and M. Hussain, "Copy-move and splicing image forgery detection and localization techniques: a review," *Australian Journal of Forensic Sciences*, vol. 49, no. 3, pp. 281–307, 2017.

[5] T. Huynh-Kha, T. Le-Tien, S. Ha-Viet-Uyen, K. Huynh-Van, and M. Luong, "A robust algorithm of forgery detection in copy-move and spliced images," *International Journal of Advanced Computer Science and Applications*, vol. 7, no. 3, 2016.

[6] K. D. Kadam, S. Ahirrao, K. Kotecha *et al.*, "Efficient approach towards detection and identification of copy move and image splicing forgeries using mask r-cnn with mobilenet v1," *Computational Intelligence and Neuroscience*, vol. 2022, 2022.

[7] A. Collins, "Forged authenticity: governing deepfake risks," 2019.

[8] F. Li, Z. Pei, W. Wei, J. Li, C. Qin *et al.*, "Image forgery detection using tamper-guided dual self-attention network with multiresolution hybrid feature," *Security and Communication Networks*, vol. 2022, 2022.

[9] H. Chen, C. Chang, Z. Shi, and Y. Lyu, "Hybrid features and semantic reinforcement network for image forgery detection," *Multimedia Systems*, vol. 28, no. 2, pp. 363–374, 2022.

[10] M. M. Qureshi and M. G. Qureshi, "Image forgery detection & localization using regularized u-net," in *International Advanced Computing Conference*. Springer, 2020, pp. 434–442.

[11] S. S. Ali, I. I. Ganapathi, N.-S. Vu, S. D. Ali, N. Saxena, and N. Werghi, "Image forgery detection using deep learning by recompressing images," *Electronics*, vol. 11, no. 3, p. 403, 2022.

[12] S. Jabeen, U. G. Khan, R. Iqbal, M. Mukherjee, and J. Lloret, "A deep multimodal system for provenance filtering with universal forgery detection and localization," *Multimedia Tools and Applications*, vol. 80, no. 11, pp. 17 025–17 044, 2021.

[13] A. H. Khalil, A. Z. Ghalwash, H. A. Elsayed, G. I. Salama, and H. A. Ghalwash, "Enhancing digital image forgery detection using transfer learning," *IEEE Access*, 2023.

[14] Y. Rao and J. Ni, "A deep learning approach to detection of splicing and copy-move forgeries in images," in *2016 IEEE international workshop on information forensics and security (WIFS)*. IEEE, 2016, pp. 1–6.

[15] X. Wang, H. Wang, S. Niu, J. Zhang *et al.*, "Detection and localization of image forgeries using improved mask regional convolutional neural network," *Mathematical Biosciences and Engineering*, vol. 16, no. 5, pp. 4581–4593, 2019.

[16] A. Ghai, P. Kumar, and S. Gupta, "A deep-learning-based image forgery detection framework for controlling the spread of misinformation," *Information Technology & People*, vol. 37, no. 2, pp. 966–997, 2024.

[17] Y. Rao, J. Ni, and H. Zhao, "Deep learning local descriptor for image splicing detection and localization," *IEEE access*, vol. 8, pp. 25 611–25 625, 2020.

[18] K. M. Hosny, A. M. Mortda, N. A. Lashin, and M. M. Fouda, "A new method to detect splicing image forgery using convolutional neural network," *Applied Sciences*, vol. 13, no. 3, p. 1272, 2023.

[19] F. Li, Z. Pei, W. Wei, J. Li, C. Qin *et al.*, "Image forgery detection using tamper-guided dual self-attention network with multiresolution hybrid feature," *Security and Communication Networks*, vol. 2022, 2022.

[20] A. K. Jaiswal and R. Srivastava, "Detection of copy-move forgery in digital image using multi-scale, multi-stage deep learning model," *Neural Processing Letters*, vol. 54, no. 1, pp. 75–100, 2022.

[21] S. Koul, M. Kumar, S. S. Khurana, F. Mushtaq, and K. Kumar, "An efficient approach for copy-move image forgery detection using convolution neural network," *Multimedia Tools and Applications*, vol. 81, no. 8, pp. 11 259–11 277, 2022.

[22] S. Dua, J. Singh, and H. Parthasarathy, "Image forgery detection based on statistical features of block dct coefficients," *Procedia Computer Science*, vol. 171, pp. 369–378, 2020.

[23] A. Alahmadi, M. Hussain, H. Aboalsamh, G. Muhammad, G. Bebis, and H. Mathkour, "Passive detection of image forgery using dct and local binary pattern," *Signal, Image and Video Processing*, vol. 11, pp. 81–88, 2017.