Mesopotamian journal of Cybersecurity
Vol. 5, No.2, **pp**. 636–656
DOI: https://doi.org/10.58496/MJCS/2025/038; ISSN: 2958-6542
https://mesopotamian.press/journals/index.php/cybersecurity

Research Article

CNNs in Image Forensics: A Systematic Literature Review of Copy-Move, Splicing, Noise Detection, and Data Poisoning Detection Methods

Mohammed R. Subhi [1]*, Salman Yussof [2], Liyana Adilla binti Burhanuddin [2], Firas Layth Khaleel [1]

[1] Department of Petroleum Systems Control Engineering. College of Petroleum Process Engineering, 34001 Tikrit, Tikrit University, Iraq

[2] Institute of Informatics and Computing in Energy, Universiti Tenaga Nasional, 43000 Kajang, Malaysia

ARTICLEINFO

Article history

Received 12 Dec 2024

Revised: 1 May 2025
Accepted 31 May 2025
Published 6 Jul 2025

Keywords
Copy-move forgery detection (CMFD)
Splicing images
Convolutional Neural Network
Data poisoning detection

ABSTRACT

Image forgery, such as copy-move and splicing, poses significant challenges to the authenticity of digital images, and this challenge is exacerbated by the rapid development of image manipulation tools. Convolutional neural networks (CNNs) have shown promise in detecting such forgeries, but limitations remain, especially in detecting small duplicate regions and low-contrast regions, as well as in dealing with optical artefacts such as noise and blur. This systematic literature review examines CNN-based approaches to detect image forgery and explores strategies to mitigate data poisoning attacks, which can compromise the integrity of machine learning models. To our knowledge, there are currently no studies that comprehensively address the integration of forgery detection and splicing techniques with data poisoning detection. Our results reveal that while CNNs are effective in detecting manipulated images, challenges remain in dealing with complex manipulations and adversarial attacks. This review highlights the need for more robust detection methods and defence mechanisms against data poisoning, as current strategies are inadequate to address these issues comprehensively. We propose future research directions that focus on improving model generalizability, incorporating data poisoning defences, and enhancing the interpretability and flexibility of detection systems.

1. INTRODUCTION

Currently, image forgery detection based on deep learning is essential in the digital age because of the ease with which photographs may be changed. Copy-move and splicing forgery are widespread ways to create fraudulent photographs, presenting a serious threat to the legitimacy of digital material [1]. These forgeries are sometimes difficult to distinguish because of the complexity of the modifications and the wide use of powerful image editing software applications. Convolutional neural networks (CNNs) and generative adversarial networks (GANs) have achieved powerful performance in detecting image forgeries by identifying different features and patterns that indicate image manipulations [2]. Some of the challenges described above have been solved by researchers using various deep learning models to detect certain types of forgeries, such as copy-move images. These models are meant to pave the way for improving the reliability of digital image content by effectively detecting the presence of forged regions and significantly helping improve the reliability of digital images in general. Various types of research have been conducted to identify fake images [3]. It becomes crucial to address specific challenges related to image forgery detection. such as small duplicated regions, low-contrast areas, and the presence of noise and other photometric distortions in images. There are two approaches employed for image forgery detection: active and passive approaches [4]. The active approach extracts hidden information from the image. The secret information is present in the form of watermarks and digital signatures. Passive methods detect region duplications, such as splicing and copy-move forgery, in an image. Moreover, highlighting the need for robust detection and defense mechanisms is imperative. Due to growing concerns about data poisoning attacks, the performance of machine learning models can be degraded by introducing malicious data into the training process. The aims of this systematic literature review (SLR) are as follows:

1. An overview of current CNN-based techniques for detecting various types of image forgeries, including copy-move forgery, splicing, and noise-induced manipulations, is provided.
2. The usefulness of these strategies in dealing with problems, including small duplicated sections, low contrast, and noise, is evaluated.
3. The effectiveness of existing defensive data poisoning techniques in the context of image forensics is analysed.
4. Research gaps should be identified, and future directions to improve robustness and model generalizability should be proposed.

*Corresponding author. Email: abo1986hhh@gmail.com

This SLR also details vital domains yet to be explored to improve the robustness and reliability of such systems with practical applications. In recent years, many studies have been conducted on copy-move, splicing, noise detection, and data poisoning defense methods, which have not been summarized together, thus leading to an extreme need for this systematic literature review (SLR) to achieve a complete overview of copy-move, splicing, noise detection, and data poisoning defense methods. The main contributions of this SLR are as follows:

1. In-depth evaluation of convolutional neural network-based forgery detection methods: This review provides a comprehensive analysis of current convolutional neural network (CNN)-based techniques for the detection of image forgeries, including copy-move forgeries, splicing, and various forms of noise-induced manipulations. It critically examines the methodological approaches employed by these techniques and assesses their performance in real-world forensic applications.
2. Critical Assessment of Deep Learning Techniques in Addressing Forgery Detection Challenges: This review evaluates the capabilities and limitations of deep learning models, specifically CNNs, in overcoming inherent challenges such as the detection of small duplicate regions, low-contrast areas, and resilience to photometric attacks such as noise, blurring, and compression. The analysis highlights the efficacy of these models in handling complex image manipulations.
3. Comprehensive Analysis of Data Poisoning Defence Mechanisms in Image Forensics: This work systematically explores the various data poisoning detection strategies employed within image forensics, with a particular focus on defending CNN-based models from adversarial poisoning attacks. It critically analyses the strengths and shortcomings of these defence mechanisms in maintaining the integrity of forensic models during the training phase.

In contrast to the work of [5], which compares the performance of CNN models such as ELA-CNN and VGG-16 for image forgery detection, this SLR expands the discussion by incorporating the detection of data poisoning alongside copy-move, splicing, and noise-based forgeries, thus offering a more comprehensive framework for maintaining model integrity under adversarial conditions. While the authors in [6] focused on poisoning attacks in recommender systems, this SLR is specifically concerned with image forensics, with an emphasis on the detection of poisoning in the context of forgery detection. Additionally, although the SLR in [7] provides an extensive review of deep fake detection techniques in both video and image formats, this SLR distinguishes itself by integrating data poisoning defences and image forgery detection methods, thereby offering novel insights into the development of more robust AI systems capable of withstanding diverse manipulation techniques. As such, this review contributes to the field by highlighting the intersection of forgery detection and data poisoning mitigation in deep learning models.

While numerous studies have been published on forgery detection, data poisoning, and related topics, this review focuses on synthesizing and critically analysing the latest advancements across multiple domains of image forensics, with a particular focus on data poisoning in deep learning-based models for image manipulation detection. Unlike previous reviews that have focused primarily on traditional forgery detection or focused exclusively on adversarial attacks during training, this paper provides a comprehensive overview of the intersection of these areas, specifically addressing the impact of data poisoning on the testing phase of deep learning models. Moreover, the integration of state-of-the-art countermeasures, such as hybrid models and enhanced preprocessing techniques, is explored in greater depth, offering a forward-looking perspective on potential solutions. By highlighting key vulnerabilities and proposing novel research directions, this review offers fresh insights that have not been extensively discussed in the literature, establishing its originality.

The rest of this paper is organized as follows: Section 2 provides the research background, Section 3 describes the SLR methodology, Section 4 presents the review outcomes, and finally, Section 5 outlines the challenges and future directions in this area.

2. BACKGROUND AND TERMINOLOGY

Our investigation in this section focuses on the essential basic principles and diagnostic approaches needed to understand the advanced techniques presented in this paper. This paper discusses different forgery detection methods together with data poisoning methods as well as deep learning techniques for detection and datasets utilized in image forensics. The foundation laid in this section enables an understanding of the proposed model's enhancements for accuracy and robustness in forgery detection practices.

2.1 Image Forgery Techniques

Understanding the many methods of image forgery that are now in use is essential for determining the authenticity of an image. These methods are divided into two categories: active methods and passive methods [3]. Active image forgery, which focuses on stating the legitimacy of images without having previous knowledge of the original image, is the modification of digital images without providing any new information. Passive image forgery detection encompasses a number of methods, including splicing and copy-move forgeries [8] [9]. Figure 1 illustrates the most frequently used techniques for identifying copy-move forgery (CMF) and splicing forgery (SF) [4].

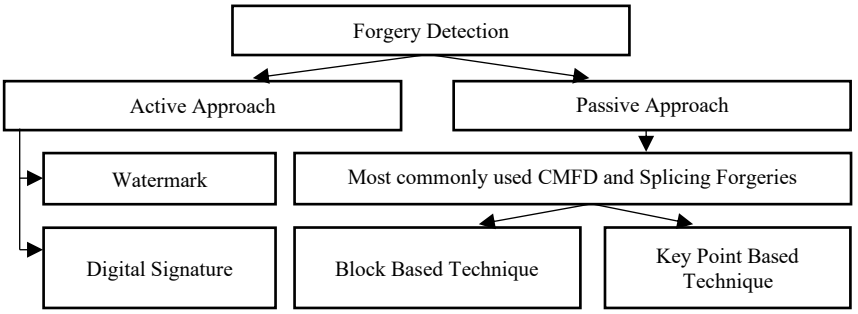

Fig. 1. Most commonly used techniques to detect image forgery.

Copy-move forgery is performed by copying a part of an image and pasting it in another location, as shown in Figure 2 [10]. The purpose is mainly to hide important information in the image. On the other hand, splicing is performed by removing a part of an image and placing it in another image, as shown in Figure 3 [11].

a. Original image b. Forged image

Fig. 2. Copy moves forgery image

(a) (b) (c)

Fig. 3. Splicing forgery image

2.2 Deep Learning For Forgery Detection

Image splicing and copy-move forgery detection are two types of image forgery techniques that can be effectively detected via convolutional neural networks (CNNs), which are a type of deep learning (DL). Many studies utilizing CNNs have managed to obtain satisfactory performance levels in regard to detecting copy-move forgeries and splicing of images [12]. These methods utilize CNN capabilities that manage to accurately determine the manipulated area. The CNN is also able to identify areas in the image where potential manipulation might have taken place. Furthermore, CNNs can distinguish between real and fake areas not only on the basis of colour but also on the basis of texture and spatial connections. In addition, the CNN is also able to detect possible rotation and scaling of the image, which is a common technique in copy-move forgery.

2.3 Data Poisoning Attacks in Image Forensics

The process of data poisoning attacks threatens machine learning models by injecting harmful data contents into training datasets. These attacks cause severe damage to models both in performance and reliability, which results in inaccurate predictions and system security threats.

The attackers manipulate training data, which results in degraded machine learning model integrity. The data poisoning attacks utilize malicious training samples that include backdoor patterns to both misidentify poisoned test samples and maintain accurate classification of clean test samples [13]. The hidden attacks against training data create major

misalignments between desired and actual classifications within deep learning solutions [14]. Various methods for testing deep learning system vulnerabilities have been developed into poisoning attack types.

2.3.1 Types of Data Poisoning Attacks

The main goal of data poisoning attacks involves destroying machine learning models through modifications of training data [15] [16]. Identifying different types of data poisoning attack remains essential for establishing efficient protection systems against such threats. Several types of data poisoning attacks represent the most prevalent group of malicious operations:

1. **Label Flipping:** The attack of label flipping modifies training data point labels by replacing them with inaccurate values through which attackers modify a portion of training data classifications, thus forcing the model to learn from incorrect information.
2. **Outliers/Noise Injection:** Noise injection and outlier data points cause both interference with models and a reduction in their performance levels.
3. **Backdoor Attacks:** This represents an attack method that allows attackers to add secretive triggers into training data, which activate model-based malicious functions during specific inference activations [16].

While untargeted data poisoning attacks aim to degrade the overall performance of the model, targeted attacks seek to manipulate the model's predictions for specific inputs [17]. In a targeted data poisoning attack, a fraudster's goal is to make a model misclassify a specific test sample to any given target class [16] [15]. Depending on the threat model, the attacker could have access to the training data or the ability to provide data [16].

2.3.2 Impact on Machine Learning Models

The deliberate alteration of training data inside machine learning systems constitutes a significant threat that results in unfavourable predictions [18]. Such attacks negatively affect different aspects of model performance through accuracy reduction and precision and recall level lowering alongside the injection of biases and system vulnerabilities. Data poisoning attacks specifically threaten image forensics systems because adversaries can use the attacks to corrupt the detection of manipulated digital images by misdirecting the models' identification of forged and edited content [19] [16] [15].

Multiple studies have investigated the risk of data poisoning within this particular scenario [15]. One successful method of attacking machine learning classifiers, particularly those used in image forensics, is poisoning attacks. In this type of attack, the attacker creates harmful samples on a substitute dataset and then transfers the attack to the target model [20] [15]. For example, attackers may add stop signs with articular stickers to the training data to manipulate the decision boundary so that the traffic sign classifier will misjudge the "stop" as the "speed limit" in the testing phase (Figure 4), which could cause self-driving cars to maintain steering without stopping obstacle avoidance [21]. Studies have shown how difficult it is to recognize these kinds of attacks since they can look very similar to real training data [16].

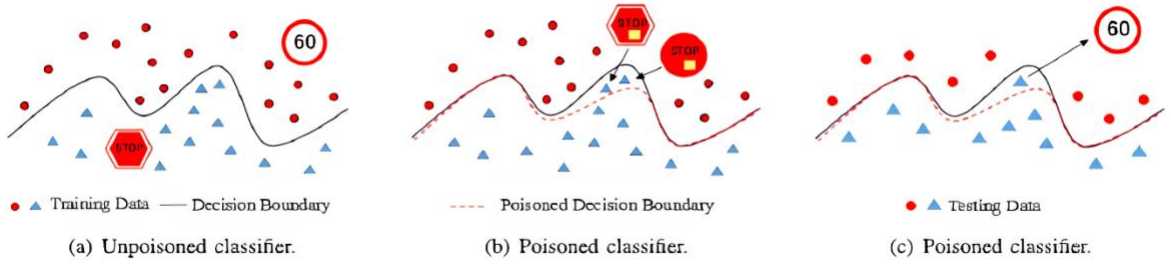

Fig 4. Representation of a data poisoning attack: (a) The classifier correctly classifies the training data. (b) Attackers insert poisoned samples into the training dataset, which manipulates the decision boundary. (c) Consequently, some test data are misclassified during the testing phase because of the poisoning attack (*e.g., a "stop" sign is mistakenly identified as a "speed limit" sign*).

2.3.3 Detection and Defence Mechanisms

In artificial intelligence systems, data poisoning attacks are intentional alterations of training data to decrease machine learning model performance or behaviour [20]. This type of attack may have disastrous results, such as inaccurate classifications, weak security, or unfair decisions [22]. Several types of detection and mitigation strategies exist for preventing data poisoning attacks. Data poisoning detection techniques work together with probability reduction methods to protect machine learning models from poisoning attacks during their training and testing operations [23]. Three detection approaches that researchers propose for AI data poisoning attacks include outlier detection together with consistency checks and anomaly detection systems [20]. Robust training and model verification are the cornerstones of AI defense methods that aim at preventing data poisoning attacks [23]. One of these defense mechanisms is data sanitization, which

involves locating and removing poisoned data from the training dataset. Another approach would be to implement robust model training, where the model is strengthened against poisoning attacks by using regularization and adversarial training, among other techniques [24]. By detecting and reducing the effects of data poisoning attacks on machine learning models, these defensive and detection techniques are essential for guaranteeing the safety and dependability of AI systems [25].

2.3.4 Motivation

Many techniques have been proposed to detect whether an image is authentic or forged. There are also techniques that use localization to find the forged region(s) in a forged image. It is very important to know whether these methods are robust, reliable and properly model the structural changes that occur in images due to copy-move [26]. Deep learning methods, such as convolutional neural networks (CNNs) and convolutional long short-term memory (LSTM) models, excel in automatically extracting complex features from images, enabling them to effectively distinguish between authentic and tampered images with high accuracy [27], [28]. Deep learning models also have superior performance to conventional methods in terms of presenting better metrics in general to detect tampered images [29]. Furthermore, deep learning techniques make advances in obtaining geometric feature detection and thus improve methods for detecting such manipulated images, especially in deep fake videos.

2.3.5 Challenges

Current image forgery detection models, including both keypoint-based approaches and deep learning approaches, face some limitations, such as identifying small duplicate portions, low-contrast regions, and noise in copy-move and splicing forgeries. Moreover, the presence of noise and other photometric attacks further complicates detection efforts. Deep learning approaches seem to be promising for detecting forgery images. In addition, CNN-based techniques have limitations compared with other deep learning methods in the detection of image forgery [12]. These traditional approaches lack the ability to automatically extract intricate features from images, which is crucial for accurate detection [30].

3. METHODOLOGY

We follow the guidelines presented in Budgen and Brereton [31] to describe our systematic literature review methodology. Our research design uses a precise method to identify research publications and develops particular standards for selecting studies that maintain both high-quality and suitable content relevance. The document outlines the procedure for collecting data together with specific criteria used for quality assessment. The following sections detail the complete process through an overview.

3.1 Review Protocol

This section explains the research procedure implemented for this study. The entire procedure for this SLR includes three distinct phases, as illustrated in Figure 5, according to [32]:

- Planning the Review: This phase focuses on defining the objective and developing the protocols for this SLR.
- Conducting the Review: This phase outlines the main research content in this SLR, which is divided into six steps:
  1. *Research Questions: This SLR will answer research questions that establish which problems need analysis and feed into the discussion section.*
  2. *Search strategy: The search strategy part defines which search databases and keywords will help gather primary studies.*
  3. *Study selection criteria: The selection criteria, which include inclusion and exclusion parameters, separate appropriate studies from inappropriate studies for inclusion in this systematic literature review.*
  4. *Quality Assessment Criteria: This step evaluates how well the chosen studies match the main goal of this SLR through the quality assessment criteria.*
  5. *Data Extraction: Designing an accurate data extraction form stands as the objective of this step to record research-related information.*
  6. *Data Synthesis: The comprehensive process of this step combines and groups findings from original studies.*

Reporting the Review: According to these guidelines, review completion takes place within this phase.

3.2 Research Questions

This section takes steps to address the identified research questions and achieve the study's objectives as described below. The protocol includes a thorough review of the literature, identification of research gaps, formulation of research questions, and an explanation of the motivations behind each question.

RQ1: How do current detection models address the limitations in identifying small duplicate portions, low-contrast regions, and noise in copy-move forgery and splicing forgeries?

RQ2: How can deep learning techniques, particularly convolutional neural networks (CNNs), improve the accuracy and reliability of detecting splicing, performing copy-move forgeries, and handling various types of noise and photometric attacks?
RQ3: How do image forgeries (copy-movie, splicing) combined with various types of noise in training datasets impact the effectiveness of data poisoning detection mechanisms?

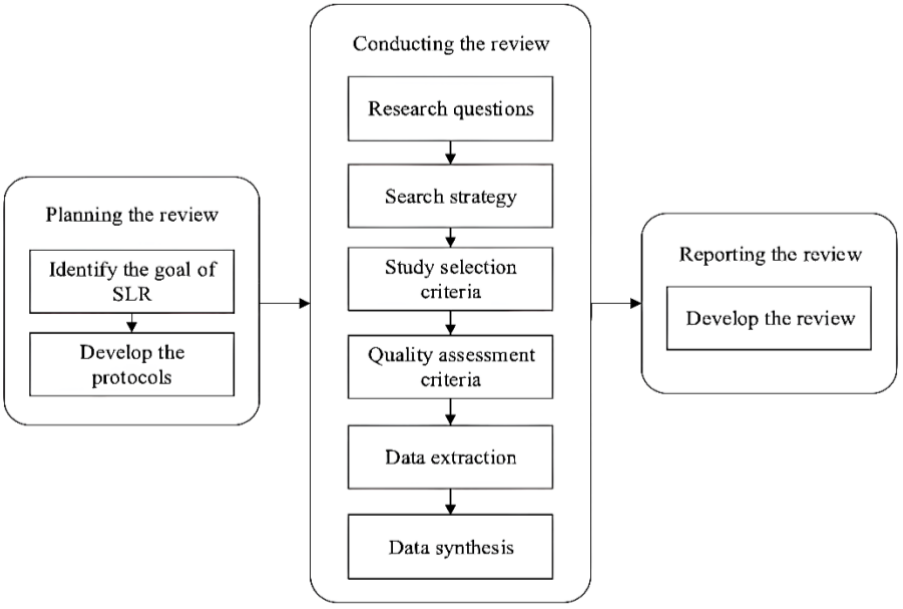

Fig. 5. Systematic Literature Review Overview

3.3 Research strategy

To gather studies related to image forgery detection, splicing detection, noise handling, and data poisoning defense, we formulate specific search terms pertinent to this paper. The primary approach involved the use of Boolean expressions to combine search terms, including 'AND' and 'OR'. The search terms can be summarized as (("image forgery" OR "splicing" OR "noise") AND ("detection" OR "defence" OR "deep learning" OR "CNN") AND ("data poisoning"). The search was restricted to articles published between 2018 and 2024. The initial search yielded a total of 571 articles. After removing duplicates and nonrelevant studies on the basis of the title and abstract, 450 articles remained. The 450 articles were then subjected to a detailed screening process. A preliminary screening based on titles and abstracts reduced the number to 200 articles. The full texts of these 200 articles were reviewed to assess their relevance and quality. This stage further reduced the number to 71 articles. After these search terms were finalized, we selected relevant digital repositories. We searched the following electronic databases:

- IEEE Xplore Digital Library. (https://ieee-collabratec.ieee.org).
- ScienceDirect. (https://www.sciencedirect.com).
- ACM Digital Library. (https://dl.acm.org).
- Wiley Online Library. (https://onlinelibrary.wiley.com).
- Google Scholar. (https://scholar.google.com).
- SpringerLink. (https://link.springer.com).
- ArXiv. (https://arxiv.org).

The search process was conducted across these electronic databases, encompassing key journals and conferences. These sources primarily originate from fields such as computer vision, image processing, machine learning, and cybersecurity.

3.4 Study Selection Criteria

The study selection criteria constitute a critical component of the systematic literature review (SLR) process, ensuring that only relevant, high-quality studies are included in the review, as shown in Table 1. These criteria systematically filter the vast amount of available literature, focusing on studies that directly address the research questions and contribute meaningful empirical evidence.

TABLE I. Inclusion and exclusion criteria for selecting studies

Criteria	Description
Inclusion	• Studies published in peer-reviewed journals or conference proceedings. • Studies focusing on image forgery detection using deep learning techniques, particularly CNNs. • Studies addressing data poisoning attacks in image forensics. • Studies focusing key point-based CMFD. • Studies providing empirical evidence on the effectiveness of the techniques.
Exclusion	• Studies not available in English. • Studies without empirical results. • Studies not relevant to the formulated research questions.

3.5 Quality Assessment Criteria

To assess the quality of the selected studies, we followed the guidelines of the quality assessment criteria in Table 2 and screened these studies to ensure that they met our standards. To ensure the reliability of the results, we used a cross-checking method to identify whether the selected studies met these criteria. After the quality assessment criteria are applied, the final studies are selected, which include a comprehensive set of articles related to image forgery detection and data poisoning defence.

TABLE II. Quality assessment criteria

Criteria	Detailed explanation
Relevance	The study must address the research questions and objectives, focusing on image forgery detection and data poisoning defense.
Empirical Evidence	Studies must provide empirical results, such as experiments, simulations, or case studies.
Study Design	The methodology and study design must be clearly described and appropriate for addressing the research questions.
Data Quality	The quality and source of the data used in the study must be reliable and well-documented.
Analysis and Results	The analysis methods and results must be clearly presented and rigorously evaluated.

3.6 Data Extraction and Data Synthesis

The data extraction process involves designing forms to capture information accurately from primary studies. These forms help gather the necessary data to answer the research questions. The extracted information in the data extraction forms is included in Figure 6.

Study details: Author, publication time, and publication source, covering both journals and conferences. Forgery and Data Poisoning Techniques: Focus on the copy-move forgery detection (CMFD), splicing forgery detection (SFD), and data poisoning detection (DPD) techniques used in the studies. These techniques are classified on the basis of feature extraction methods and models.

Empirical Evidence: Summarized from six dimensions: experimental datasets, feature extraction techniques, data poisoning methods, forgery detection methods, models used, and performance measures. During data synthesis, similar and comparable results from the data extraction forms are summarized, providing supporting evidence to conclusively answer the research questions.

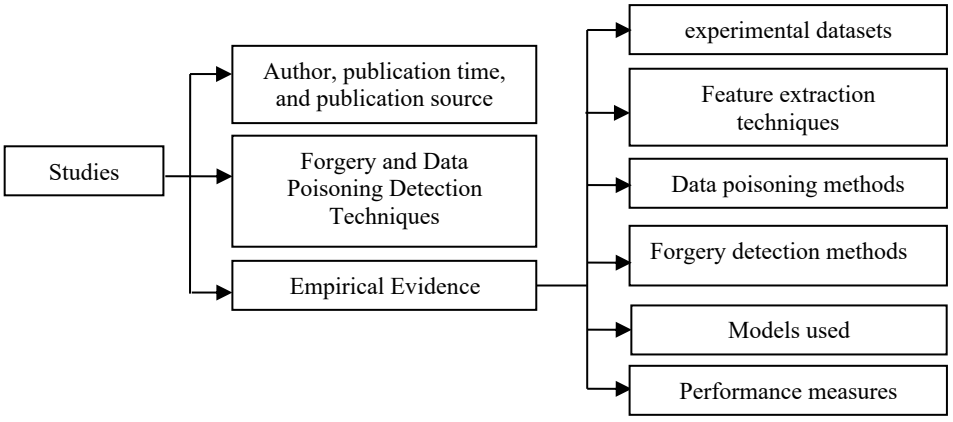

Fig. 6. Extracted data

4. LITERATURE REVIEW

The development of reliable detection algorithms is crucial for protecting visual data from the growing problem of digital image modification. Our literature review focuses on three key areas: CMFD algorithms using hand-crafted features, protecting machine learning models from data poisoning attacks, and detecting image forgeries, splicing, and noise. We explore the effectiveness and limitations of CMFD methods such as SIFT and SURF, review strategies against data poisoning such as anomaly detection and robust statistics, and compare conventional and deep learning techniques for forgery and noise detection. The purpose of this review is to evaluate proven strategies that improve image forgery inspections while maintaining model authenticity.

4.1 IC-MFD Algorithm-Based Hand-Crafted Features

The CMFD methods organize themselves into three groups that divide images through three different techniques: block-based approaches [33], segmented region-based approaches [34], and local keypoint-based approaches [35]. Block-based methods divide an image into different subblocks where the arrangement can be either overlapping or nonoverlapping. A block-division method decreases the time needed to search for corresponding feature vectors within images when applied instead of performing a complete search. Under the segmented-based approach, the image is divided into sections that contain all the fake objects. The keypoint-based approach functions without segmentation by detecting distinctive local features that include corners and edges along with blobs. The keypoint-based approach adopts two fundamental methods, among others, scale-invariant feature transform (SIFT) [36]. Speeded Up Robust Features (SURF) [37]. Human designers create handmade features that include natural features, among other features.

Keypoint-based CMFD techniques execute their feature extraction phase through the detection of features, which subsequently yields descriptions [38]. A set of image keypoints is detected in feature detection, as they possess stability properties under geometric transformations [39]. The feature description step uses these keypoints by developing encoding schemes for their surrounding areas. The SIFT and SURF algorithms serve as the main choices for executing both feature detection and feature description tasks in CMFD applications. Several CMFD (copy-move forgery detection) techniques based on keypoints are examined in Table 3. The summary presents several detection approaches that integrate a breakdown of their methods together with their benefits and constraints. The evaluation of detection and description techniques provides knowledge about the SIFT and SURF methods and how they address multiple aspects of image forgery detection. The extensive reference provides essential details about keypoint-based CMFD methods during their operation across different scenarios.

TABLE III. Summary of keypoint-based CMFD techniques

REF	DESCRIPTION	DATASET USED	KEYPOINT-BASED APPROACH	PERFORMANCE	LIMITATIONS
[40]	- A fast and accurate method using dense keypoints and invariant features, effective even in smooth or small regions. Achieved 94.54% F-measure with low false detections and fast runtime.	- FAU - GRIP - MICC-F600 - CMH	- Dense and uniform SIFT keypoints	- Average pixel-level F-measure of 94.54% and average CPU-time of 36.25.	-High levels of blurring can still degrade the performance. -Brightness change. -Color reduction.
[41]	- A five-stage SVM-based method for copy-move forgery detection using block features. Achieved 98.44% accuracy on MICC-F220, outperforming several existing methods.	- MICC-F220	- Block-based	- 98.44% detection accuracy	- Struggle with high levels of blurring. - Fail to extract distinctive features.
[42]	- A hybrid deep learning and DCT-based method using LDR preprocessing and Patch Match. Shows improved robustness under compression and other attack scenarios.	- GRIP	- Deep CNN-based features and DCT based block features	- 97.16% (pixel-level) - 96.96% (image-level)	- JPEG compression and noise addition. -Blurring not mentioned.
[43]	- Proposes SMDAF with second-keypoint matching and adaptive filtering. Offers strong robustness under real-world and postprocessing attacks with improved detection accuracy.	- CASIA, CMFD, - MICC-F220, - CoMoFoD, - Coverage dataset.	- Scale-Invariant Feature Transform (SIFT).	- CASIA-CMFD: F1 score: 0.714 - MICC-F220:F1 score: 0.904. - CoMoFoD: Pixel-level F1 score: 0.511.	- Decline in performance under image blurring and noise addition. - Still struggle with false positives and false negatives.
[44]	- This paper presents a robust method for detecting and localizing multiple copy-move forgeries using adaptive keypoint extraction and quaternion polar	- FAU, GRIP.	- Generic Features from Accelerated Segment Test (GFAST).	- FAU Dataset: Achieved an F-measure of 98.98% at the image level	- Degrade under JPEG compression, Blurring, Contrast, and other significant noise addition.

	transforms. By employing KD-tree matching and offset-based clustering, the method efficiently handles small, smooth, and multiple forgeries. The approach demonstrates superior performance compared to current state-of-the-art techniques.			and 94.16% at the pixel level. - GRIP Dataset: Achieved an F-measure of 99.87% at the image level and 97.47% at the pixel level.	
[45]	- This paper proposes a pixel-based forgery detection framework for copy-move and splicing manipulations. The system enhances image textural features through preprocessing and utilizes enhanced SURF and template matching for forgery detection. Evaluations on the CASIA dataset show that the system achieves 97.5% accuracy, outperforming existing methods.	- CASIA.	- Advanced SURF, Template Matching.	- Advanced SURF: Achieved 98% detection accuracy. - Template Matching: Achieved 100% detection accuracy. - Overall System: Achieved 97.5% detection accuracy when combining both methods.	- Challenge in detecting small or smooth regions due to the difficulty in extracting sufficient keypoints.

A summary of keypoint-based research methods and findings will answer **RQ1** by revealing their accomplishments and disadvantages with recommended enhancements. The synthesis will provide both an understanding of modern research fluctuations and possible future progress points:

- Many keypoint-based methods generally face difficulties in detecting noise-based forgeries, which include both JPEG image compression and blurring and modifications to image contrast levels. The model reported in [40] reaches 94.54% pixel-level accuracy yet displays failures with brightness changes and blurring artifacts, whereas [44] demonstrates high F-measure performance on GRIP data before degradation occurs when handling JPEG compression or blurring effects. Despite its high accuracy level of 97.16% [42], it demonstrates poor performance under JPEG compression as well as noise addition patterns, which indicates its sensitivity to typical real-world image alterations. Strong noise-handling mechanisms need to be integrated into CMFD models because they play a fundamental role in performance enhancement.
- Despite high accuracy in controlled settings, keypoint-based CMFD methods face challenges with false positives and false negatives, particularly in small or smooth regions. For example, [45], who used SURF and template matching, achieved 97.5% detection accuracy but struggled with detecting small or smooth regions because of the difficulty of extracting enough keypoints. The author of [43] reported that SIFT performs well on CASIA-CMFD but still struggles with false positives and false negatives, especially under image blurring and noise addition. This highlights the need for improved feature extraction techniques to minimize false detections and enhance the accuracy of the models.

To detect image tampering, features extracted via key point methods and region matching via block-based methods are combined. However, if the features are sparse, then again, the fusion methods cannot address the smoothing effect [46]. One way to extract more keypoints is to utilize hybrid/multiple detectors, such as those in [47]. Other works, such as [48], applied keypoint detectors on the opponent colour space rather than the intensity channel to obtain an adequate number of keypoints.

4.2 Image Forgery, Splicing, and Noise Detection Methods

Image forgery, splicing, and noise detection techniques are reviewed in this section, along with their strengths and weaknesses. To keep digital images genuine and undamaged, it is essential to understand these strategies. Table 4 shows a summary of image forgery detection techniques and their performance.

In [49], the authors simulated and examined a convolutional neural network (CNN) model to recognize any forged picture. Three stages are included in the model: classification, feature extraction, and data preprocessing. The fully connected layer is used by the model to determine if an image is authentic or fake after it has learned to extract features from the convolutional, pooling, and rectified linear unit layers. Three datasets (MCC-F2000 (2000 images), CASIA 1 (1721 images), and CASIA 2 (12615 images)) are used in the experimental studies, and their performance is evaluated and contrasted with that of current deep learning-based techniques. According to their findings, the CNN model performed best, achieving an accuracy of 76% on the MICC-F2000, 79% for CASIA 1, and 89% for CASIA 2.

Qazi et al. [50] suggested a method based on ResNet50v2, a modern deep learning architecture. The suggested model uses the ResNet50v2 architecture to use the weights of a YOLO convolutional neural network (CNN) on batches of images as input. The author of this work detected image splicing via the CASIA_v1 and CASIA_v2 benchmark datasets, which comprise two unique categories: original and forgery. also compared the suggested method with those that are already in use. The authors examined method performance using the CASIA_v1 and CASIA_v2 datasets. Because the CASIA_v2

dataset is larger than the CASIA_v1 dataset is, they were able to achieve 99.3% accuracy with the fine-tuned model via transfer learning. Hosny et al. [51] suggested a convolutional neural network (CNN) model with fewer parameters than the previously disclosed methods for accurately and quickly identifying splicing forged images. With only four convolutional layers and four max pooling layers, the model that is being described is lightweight and appropriate for most situations with resource constraints. A thorough analysis was performed to compare the suggested model with the other models. The suggested model's sensitivity and specificity are assessed via the CASIA 1.0, CASIA 2.0, and CUISDE datasets. With respect to identifying forgeries on the CASIA 1.0, CASIA 2.0, and CUISDE datasets, the suggested model had accuracy rates of 99.1%, 99.3%, and 100%, respectively.

Islam et al. [52] proposed an effective model that uses discrete cosine transformation (DCT) and local binary pattern (LBP) operators for splicing and copy-move attack detection in colour images. These attacks have a greater effect on chromaticity components than on brightness components. Initially, the Chroma components of an image are separated into nonoverlapping, fixed-sized blocks. Then, 2D block DCT is used to detect any changes resulting from image forgery in the local frequency distribution. Then, to accentuate the artifacts caused by the tampering procedure, a texture descriptor called LBP is added to the magnitude component of the 2D-DCT array. Blocks that do not overlap are once again separated into the resultant LBP image. The appropriate intercell values of each LBP block are finally added together, and the results are organized into a feature vector. The proposed method leverages a support vector machine (SVM) with a radial basis function (RBF) kernel to differentiate between authentic and forged images. Through rigorous experimentation on renowned datasets for image splicing and copy-move detection, this approach shows its superiority over recent state-of-art methods. The accuracies are 97.52, 97.79 and 99.82 when using Columbia Color, CASIA 1, and CASIA 2, respectively.

Ali et al. [53] provide a novel neural network and deep learning-based image forgery detection solution that emphasizes the CNN architecture. The method that is suggested makes use of a CNN architecture that takes into account variances in image compression to achieve good results. The model is trained by leveraging the differences between the original and recompressed images. The suggested method is effective in identifying image forgeries, which include copy-move and image splicing. The total validation accuracy of the experiment, with a specified iteration limit, is 92.23%, which is promising.

Hussien et al. [54] provide an automated technique for detecting image splicing forgeries, which is based on extracting image features via colour filter array (CFA) analysis. PCA is used to reduce the dimensionality of features. To distinguish between genuine and spliced images, a classifier built on deep belief networks is developed. The robustness of the approach is evaluated via the Columbia Image Splicing Detection Evaluation Dataset (CISDED), which includes several circumstances, such as Gaussian noise and JPEG compression. With 95.05% precision, 94.05% recall, a 94.05% true positive rate, and 98.197% accuracy, the results demonstrate good performance, confirming its superiority over contemporary splicing detection approaches. In mallick et al. [55], a CNN-based technique for detecting copy-move image forgeries was proposed and presented. The proposed technique extract features from two datasets, CASIA v2.0 and NC2016, with different levels of complexity via the CNN. A model that is built on three distinct models, i.e., ELA, VGG16, and VGG19, is able to provide excellent results on a collection of images from the CASIA2.0 and NC2016 datasets, with accuracies of 70.6%, 71.6%, and 72.9%, respectively. In Kuznetsov [56], the author utilized the VGG-16 convolutional neural network (CNN) for feature extraction and classification, leveraging its deep hierarchical structure to improve image analysis and detection performance. The suggested network architecture classifies image patches as original or fake on the basis of the input image patches. The author selected patches from the original image areas and the edges of the embedded splicing during the training phase. Compared with previous solutions, the acquired results show excellent classification accuracy (97.8% accuracy with the fine-tuned model and 96.4% accuracy for zero-stage training) for a collection of images with artificially introduced distortions. The CASIA dataset was used for the experimental studies.

In [57], the authors propose applying deep learning to identify image splicing in images. Initially, the input images are pre-processed via the 'Noiseprint' approach, which suppresses the image content to obtain the noise residual. Second, a feature extractor is used, which is the well-known ResNet-50 network. Finally, the SVM classifier is used to classify the acquired features as legitimate or spliced. The suggested strategy performs better than other current methods do, according to experiments performed on the CUISDE dataset. The average categorization accuracy achieved by the proposed method is 97.24%.

The authors of [41] proposed a unique approach to image forgery detection that can concurrently detect copy-move and splicing forgeries on the CASIA v1.0 and CASIA v2.0 datasets. To retain splicing artifacts, the image is first transformed into YCbCr channels, with chrominance channels being given priority for feature extraction. Among the preprocessing procedures are picture decorrelation and BDCT. The system is trained on real and fake images once all the features have been combined, and an SVM is used for classification. To find forged areas, a copy-move detection approach maps replicated regions. The method's resilience is shown by the experimental findings, which demonstrate high effectiveness, achieving a 99.50% accuracy rate for splicing identification at threshold T=8 and an 87.50% accuracy rate for copy-move detection.

In Elaskily et al. [58], a convolutional neural network (CNN) specifically designed for copy-move forgery detection (CMFD) was presented. The CNN successfully differentiates manipulated images from original ones by learning hierarchical features from the input images. Comprehensive tests show that the method performs better than conventional CMFD systems do on three publicly available datasets: MICC-F220, MICCF2000, and MICC-F600. Furthermore, the method achieves an accuracy of 100% across all four datasets when these datasets are combined with SATs-130, demonstrating its resilience against a variety of known distortions. The results are comparable and provide a comprehensive evaluation, demonstrating the model's performance across various benchmarks.

In [59], a novel method that combines colour illumination, deep CNNs, and semantic segmentation for the detection and localization of image forgeries was described. Pixel-level classification reliably differentiates between forged and legitimate regions by fine-tuning VGG-16 via transfer learning. This categorization is refined by semantic segmentation to accurately define the fabricated areas. Outstanding results are obtained from testing on benchmark datasets: an average border F1 score of 86.404%, an average accuracy of 98.581%, an average IOU of 91.148%, a weighted IOU of 97.193%, and an overall accuracy of 98.482%. In particular, the average border F1 score is 79.709%, the IOU is 83.945%, and the forged pixel accuracy is 98.698%. With an average border F1 score of 93.055% and an IOU of 98.351%, the model attains an accuracy of 98.463% for non-forged pixels. These results outperform those of cutting-edge techniques, demonstrating the method's efficacy in producing precise forgery detection.

The authors of [60] proposed a hybrid network to identify duplicated regions in digital images by merging a CNN and LSTM. The suggested model successfully localizes copied and pasted areas by combining SRM filters, LSTM, CNN, and SVM classification. Patch extraction and rotation algorithms are used to solve resampling artifacts, such as up sampling or down sampling, rotation, and shearing. An LSTM-CNN network is used to extract features and patches from both compromised and clean images. To differentiate between altered areas, the SVM classifier examines many properties. The technique's usefulness is shown by experimental findings on the CASIA and CoMoFoD datasets, which yielded 94.7% accuracy with patch rotation and 82.8% accuracy without rotation for CASIA and 84.8% accuracy with rotation and 73.5% accuracy without rotation for CoMoFoD.

In [61], the proposed approach, multiple structures of stacked autoencoders (SAEs), was implemented for detecting forgeries across various image compression techniques, using pretrained AlexNet and VGG16 models for feature extraction. The Ensemble Subspace Discriminant classifier is employed to classify images as authentic or forged. Extensive ablation studies were conducted on two CASIA datasets, which revealed that the combination of two autoencoders and AlexNet features outperforms several other architectures and state-of-the-art methods, achieving 95.9% accuracy for JPEG images and 93.3% accuracy for TIFF images. The research in [62] introduces an effective convolutional neural network (CNN) that detects copy-move image forgery in an efficient manner. This model executes three convolutional layers alongside three max pooling layers followed by one fully connected layer through ReLU activation coupled with the RMSprop optimizer. The testing time for this model remains fast because it completes each image analysis in approximately 0.83 seconds. The proposed system achieved 100% accuracy on all tests in the MICC-F2000, MICC-F600, and MICC-F220 datasets, demonstrating exceptional performance and speed for image forgery detection.

TABLE IV. Summary of image forgery detection techniques and their performance

Ref.	Dataset Used	Method Used	Pros	Cons	Limitations	Accuracy
[49]	- MCC-F2000. - CASIA v1. - CASIA v2.	- CNN model with classification. - Feature extraction. - Data preprocessing stages.	- Autonomously learns and extracts features from images. without prior knowledge of forgery types.	- CNN models may require substantial computational resources and training time, especially when dealing with large datasets like CASIA 2 with 12,615 images.	- Limited accuracy, especially on smaller datasets like MCC-F2000.	- 76% (MCC-F2000). - 79% (CASIA v1). - 89% (CASIA v2).
[50]	- CASIA v1. - CASIA v2.	- ResNet50v2 with transfer learning from YOLO.	- Achieved 99.3% accuracy on larger CASIA v2 dataset.	- Blurring and contrasting not mentioned.	- Performance may degrade on smaller datasets. - Generalization may accrue using different datasets.	- 99.3% (CASIA v2).
[51]	- CASIA 1.0. - CASIA 2.0. - CUISDE.	- Lightweight CNN with 4 convolutional	- High accuracy, resource-efficient.	- Because of resource-limited environments, more complex forgery detection tasks require deeper networks.	- Limited model capacity due to lightweight architecture.	- 99.1% (CASIA 1.0). - 99.3% (CASIA 2.0). - 100% (CUISDE).

		and 4 max-pooling layers.				
[52]	- Columbia Color. - CASIA V1. - CASIA V2.	- DCT, LBP, and SVM with RBF kernel.	- Effective for splicing and copy-move attacks.	- Not covered diverse image characteristics it might result generalization problem.	- Handcrafted features may not generalize well.	- 97.52% (Columbia Color). - 97.79% (CASIA V1). - 99.82% (CASIA V2).
[53]	- Retrained on CASIA 2.0.	- CNN architecture considering image compression.	- Effective for copy-move and splicing forgeries.	- Blurring and contrasting not mentioned.	- Limited information on datasets and potential overfitting.	- 92.23% validation accuracy.
[54]	- CISDED.	- CFA analysis, PCA, and deep belief network classifier.	- Robust to limit noise and JPEG compression quality factor.	- High compression quality factor could decrease accuracy. - Blurring, contrasting, salt and pepper not mentioned.	- Handcrafted CFA features may not generalize well.	- 95.05% precision. - 94.05% recall. - 98.19% accuracy
[55]	- CASIA v2.0. - NC2016.	- CNN-based with ELA. - VGG16. - VGG19.	- Handles complex datasets.	- Difficult to generalize different distributions.	- Limited generalization ability across datasets.	- 70.6% (ELA). - 71.6% (VGG16). - 72.9% (VGG19).
[56]	- CASIA.	- VGG-16 CNN.	- High accuracy.	-	- Limited information on potential limitations.	- 97.8% (fine-tuned). - 96.4% (zero-stage training).
[57]	- CUISDE.	- Noiseprint, ResNet-50, and SVM.	- Outperforms current methods.	-	- Limited information on potential limitations and generalization ability.	- 97.24% classification accuracy.
[59]	- CASIA v1.0. - CASIA v2.0.	- YCbCr channels, decorrelation. - BDCT. - SVM.	- Detects copy-move and splicing forgeries.	- Blurring and contrasting note mentioned. - None effective on copy movie forgery detection.	- Handcrafted features may not generalize well.	- 99.50% (splicing). - 87.50% (copy-move).
[58]	- MICC-F220. - MICC-F2000. - MICC-F600. - SATs-130.	- CNN for copy-move forgery detection.	- High accuracy, robust against various attacks.	-	- Limited information on potential limitations and generalization ability.	- 100% accuracy (combined datasets).
[4]	- Benchmark datasets.	- Color illumination. - Deep CNNs. - Semantic segmentation.	- High accuracy. - precise forgery localization.	-	- Limited information on potential limitations and generalization ability.	- 98.581% average accuracy. - 91.148% average IOU.
[60]	- CoMoFoD. - CASIA.	- Combining SRM filters. - LSTM. - CNN. - SVM classification.	- Effective with splicing forget images detection without patch rotation.	- Difficult to generalize to different distributions.	- Limited accuracy with comofod dataset. - Limited accuracy with patch rotation in both datasets.	- 94.7% (CASIA). - 84.8% (COMOFOD).

[61]	- CASIA v1.0. - CASIA v2.0.	- SAE. - VGG16. - Pretrained AlexNet.	- Handle various image compression techniques. - AlexNet and VGG16 well-regarded in the field of computer vision.	- CASIA datasets may not fully represent the diversity of real-world scenarios and other image formats beyond JPEG and TIFF.	- Deep autoencoder and CNN models could still be prone to overfitting. - Does not provide detailed on various degrees of compression.	- 95.9% accuracy for JPEG images. - 93.3% for TIFF images.
[62]	- MICC-F2000. - MICC-F600. - MICC-F220.	- Convolutional Neural Network (CNN).	- Efficient at 35 epochs. - Fast, taking approximately 0.83 seconds per test.	- Only detecting copy-move forgery. - Splicing, contrast, blurring, compression not mentioned.	- This may lead to overfitting. - May not generalize well to unseen data.	- 100%.

The analysis of **RQ2** requires examining vital findings from the literature review, which concentrates on detection methods for image forgery along with splicing and their responses to different noise types. The existing methods are analysed according to their strengths and drawbacks while showing how they address various manipulation techniques combined with noise conditions:

- Many studies, such as [49], [50], [55], and [60], employ CNN-based models for detecting copy-move and splicing manipulations. CNNs are effective in learning and detecting complex patterns, but their performance can be limited when dealing with issues such as edge detection, noise, and low-contrast regions. For example, methods relying on feature extraction, such as those in [49] and [52], do not explicitly address noise or edge distortions, which are key challenges in forgery detection. These methods struggle when images have fine-grained manipulations or noise that obscures duplicate portions. The use of handcrafted features (DCT, LBP, SVM, etc.) in [52] and [59] may fail to generalize well in noisy environments, missing subtle manipulations or offering inaccurate predictions in the presence of noise. Therefore, more advanced feature extraction techniques, such as hybrid methods and noise-adaptive preprocessing, could improve performance in detecting complex forgeries and handling noise effectively.

- Most studies rely on popular datasets such as CASIA and MICC-F2000, as shown in [49], [55], and [58]. While these datasets are widely used, they do not fully account for the variety and complexity of real-world scenarios, including diverse image types and manipulation techniques. For example, the CASIA dataset may not capture the full range of noise-induced manipulations and edge distortions that your problem statement addresses. The MCC-F2000 dataset produces restricted performance when analysed via the methods described in [49] because limited data are available specifically for complex forgery detection. The detection models need additional data representing various noise patterns and splicing techniques to improve their accuracy and generalizability.

- A recurring issue in the table is the challenge of generalizing across different datasets. Studies such as [50] and [55] demonstrate that models may perform well on specific datasets (e.g., CASIA v2.0) but struggle with other datasets or new unseen data. For example, transfer learning techniques such as those used in [50] (ResNet50v2 with YOLO) help improve generalization, but performance can degrade on smaller datasets or datasets with complex manipulations. This aligns with your concern that current methods may overfit specific datasets, reducing their effectiveness in real-world scenarios.

- Many methods in the table either do not address specific noise attacks or have limited handling capabilities, which poses a significant challenge in real-world forgery detection. For example, [54] and [55] attempt to handle JPEG compression and certain noise types but do not explicitly cover other common photometric attacks, such as salt-and-pepper noise, blurring, or contrast changes. These attacks can severely complicate the detection of manipulated regions, as noise can hide or distort forgeries. Techniques based on handcrafted features (e.g., [52], [59]) may fail to detect forgery regions effectively when such noise or photometric manipulations are present, as seen in your problem statement. To overcome this, your focus on developing an adaptive noise-handling pipeline, along with hybrid feature extraction, could significantly enhance the robustness of forgery detection systems against such attacks, making them more reliable in challenging environments.

- Accuracy metrics across different studies vary significantly, with some methods achieving high performance (e.g., 100% accuracy in [58] on combined datasets), whereas others perform more modestly (e.g., 76% on MCC-F2000 in [49]). While high accuracy on specific datasets is promising, it often reflects overfitting and may not translate well to other, more complex datasets. The accuracy differences between methods such as [50] (99.3% on CASIA v2) and [59] (99.50% on splicing detection) indicate that while certain methods may perform well on specific manipulation types (such as splicing), they may not generalize effectively to all types of forgeries, especially in the presence of noise or when applied to more complex datasets. This aligns with your concern that models need to be tested on challenging, real-world conditions, not just ideal or controlled datasets. By focusing on improving accuracy across diverse conditions with noise and manipulation variability, your work aims to bridge this gap, making detection methods more robust and applicable in practice.

Because the model does not explicitly handle blurring, contrast adjustments, or compression, it may struggle to detect forgeries involving these techniques. Additionally, the model's effectiveness in detecting combined manipulations remains uncertain, and it has not been extensively tested under diverse and uncontrolled conditions. Additional improvements in the system will be necessary to increase model robustness while expanding its ability to detect various types of manipulations.

4.3 Data Poisoning Detection Methods

By injecting malicious training set inputs, machine learning models face substantial threats to their operational quality along with their information accuracy. A review of attack detection and mitigation strategies for data poisoning attacks involves a breakdown of their benefits and drawbacks. The performance outcomes of the data poisoning detection methods can be found in Table 5.

[21] presented 'De-Pois' as a defense system that functions independently from specific attacks to combat data poisoning attacks. The training process for De-Pois creates a model that duplicates the behavior of the target model during clean sample training. GAN networks assist in both improving training data and building the mimic model. The core strategy of De-Pois is to detect poisoned samples by comparing the predictions of the mimic model with those of the target model. Since divergences in predictions can indicate the presence of poisoned data, De-Pois can identify these samples without specific knowledge of the machine learning algorithms or the nature of the poisoning attacks. The authors of [63] proposed the MOVCE model, which performs verification via CNNs and word embedding. This model is designed as a countermeasure to maintain the reliability of deep learning vision systems in the face of data poisoning attacks.

Authors in [64] analysed how adversarial attacks on training data affect model parameters, using a CNN model and the MNIST dataset as a test. The increase in poisoned data within the training set is caused by the addition of more manipulated samples. Approaching the issue from the network's feature space reveals a correlation with the model's training parameters. A proposed method detects whether the network was attacked during training by comparing the distributions of parameters in intermediate layers, which are calculated via the maximum entropy principle and the variational inference approach. An enhanced version of the k-nearest neighbors (k-NN) algorithm is designed to defend against data poisoning attacks during the training of machine learning models. The typical k-NN algorithm is a basic machine learning technique used for classification, where an object is assigned to the class most common among its k nearest neighbors. According to [65], the 'Deep k-NN' modification integrates the algorithm into a deep learning framework, leveraging representations learned by deep neural networks to determine neighbors and enhance defense against complex data poisoning strategies.

The model proposed in [66] uses a VGG16-based convolutional neural network (CNN) within a federated learning framework to detect and classify skin cancer while addressing data poisoning attacks. The study employs the Skin Cancer MNIST: HAM10000 dataset, sourced from Kaggle, which includes images of melanoma, nevi, and seborrheic keratoses. The model achieves an overall accuracy of 94.5% and an AUC-ROC value of 0.974, demonstrating high discriminatory ability and robustness in classifying skin lesions while ensuring data privacy and security.

The model proposed in [67] employs Projected Gradient Descent (PGD) adversarial training to improve the robustness of the model against poison attacks, a boundary augmentation algorithm using DeepFool, and ensemble training with preprocessing methods such as the ShrinkPad, feature squeezing, and PatchShuffle to increase robustness against data poisoning. The model demonstrated outstanding performance on benign data from MNIST, CIFAR-10, GTSRB and ImageNet while minimizing BadNets attacks to 0.3% on MNIST. The method enhances model detection capabilities for poisoning attacks to become more effective against such intrusions.

In [68], the author proposed a support vector machine (SVM) as the detection model. The SVM is used to detect data poisoning attacks in federated learning by classifying malicious clients on the basis of SHAP values derived from the model parameters. The detection model is evaluated on the MNIST and Fashion MNIST datasets. The highest accuracy achieved was 100% for the MNIST model with a linear kernel, whereas the Fashion MNIST model achieved an average accuracy of 94.2% when a linear kernel was used.

TABLE V: SUMMARY OF DATA POISONING DETECTION METHODS AND THEIR PERFORMANCE

Ref	Dataset used	Method used	Pros	Cons	Limitations	Accuracy
[21]	- CIFAR-10. - Fourclass dataset. - House pricing dataset.	- De-Pois.	- Effectiveness on different datasets.	- Dependency on Clean Data. - specific noise such as triggers. - Copy-move and splicing not mentioned (still a challenge).	- Small portion of trusted clean data may not always available in real world scenarios. - Performance degrades on more complex data and when poisoning data is more than 20%. - Generalization problem.	- Over 90% in average.
[63]	- CIFAR-10.	- MOVCE model. - Verification. - CNN. - Word Embeddings.	- Can adapt as the volume of data grows.	- Balanced datasets for training in real-world scenarios can be difficult. - Different lighting or filters in images that could ensemble data drift.	- Limitation in Generalizability. - Limitations in lighting and filters (need preprocessing). - Copy-move and splicing not mentioned (still a challenge). - Extra noise still a challenge such as contrasting and blurring.	- 70%.
[64]	- APTOS 2019	- CNN. - Max pooling. - Dense. - Dropout. - Flaten. -	- The model addresses a critical issue in the medical field. - evaluates the impact of various data poisoning perturbations (such as brightness, noise, zoom, etc.) -	- Performance against a limited set of perturbations (brightness, zoom, noise, etc.). - Not effected to other attacks such as copy-move, splicing, contrast adjustments. -	- Single Dataset Focus, the model might be affected using different type of datasets.	- 95%.
[65]	- CIFAR-10	- Deep k-NN	- Number of poisoning samples doesn't matter.	- Risk of discarding clean data along with the poisoned samples, especially if k is set too high.	- Limitations in Generalization. - Copy-move and splicing not mentioned (still a challenge). - Extra noise still a challenge such as contrasting and blurring.	- 99%.
[66]	- Skin Cancer MNIST: HAM10000.	- VGG16.	- Enhancing the model's ability to identify intricate patterns in skin lesion images by using VGG16.	- Model remains vulnerable to sophisticated adversarial attacks that may not be easily detectable.	- Model could be affected on other datasets. - copy-move and splicing not mentioned (still a challenge). - Extra noise still a challenge such as contrasting and blurring.	- 94.5%.
[67]	- MNIST. - CIFAR-10. - GTSRB. - ImageNet.	- Projected gradient descent (PGD). - DeepFool. - ShrinkPad. - Feature Squeezing. - PatchShuffle.	- High accuracy. - Model's applicability across different types of data.	- Effective on smaller datasets. - Adversarial training could lead to overfitting to specific attack patterns.	- Limited Generalizability. - Copy-move and splicing not mentioned (still a challenge). - Extra noise still a challenge such as contrasting and blurring.	- 99.2% on MNIST, 91.5% on CIFAR-10, 97.8% on GTSRB, 75.1% on ImageNet.

[68]	- Open-source MNIST. - Fashion MNIST.	- Support Vector Machine (SVM). - Shapley Additive Explanation (SHAP)	- High accuracy with MINIST Dataset. -	- Evaluated on only two datasets.	- May not generalize well to other types of data. - Copy-move and splicing not mentioned (still a challenge). - Extra noise still a challenge such as contrasting and blurring.	- 100% on MNIST. - 94.2% on Fashion MNIST.

To address **RQ3**, we can summarize key insights and findings from studies related to data poisoning detection methods. By examining these studies, we can identify effective strategies, common challenges, and areas for improvement in enhancing the robustness of data poisoning detection mechanisms in the context of image forensics, particularly for forgery and splicing detection.

- Many studies (e.g., [21], [63], [66]) rely on relatively simple datasets such as CIFAR-10 and MNIST, which do not fully represent the complexity of real-world forgery detection, particularly for tasks involving splicing or copy-move forgeries. These datasets also do not address the variety of challenges posed by noise, edge distortions, and photometric attacks. Moreover, generalization across datasets remains a significant issue, as noted in [63], where models fail to adapt to data variations such as lighting or filters. In particular, models that perform well on smaller datasets or more controlled conditions (e.g., [68] achieving 100% accuracy on MNIST) often struggle to generalize to larger, more complex datasets such as CIFAR-10 or ImageNet, where noise and complex manipulations (such as splicing) are more prevalent. The current approach for detecting forged manipulations needs to be developed because it shows inadequacy for handling multiple datasets and changing adversarial techniques.
- Most studies bypass the problem of noise and photometric attacks such as blurring and contrast by omitting specific solutions for these detection challenges. Research papers [63] and [67] identify noise as a challenge, but their proposed methods lack proper methods to address complete distortion resiliency effectively. Noise creates successful concealed areas that present difficulties for forensic analysis of manipulations. These models lack explicit noise management features, although they achieve high performance under controlled conditions, which indicates their insufficient ability to process complex situations.
- All studies examined, including [21], [63] and [64], failed to address specifically the identification needs for both copy-move and splicing forgeries, even though these manipulations form the focus of this research investigation. Research has focused mostly on overall image classes or data tamper identification, yet it has not explored or addressed the particular difficulties generated by splicing and copy-move modifications. Studies demonstrating 99% accuracy in specific tasks (e.g., [65]) struggle to spot very delicate forgeries, especially when they contain noise contamination. This research becomes essential for developing specialized detection methods for splicing and copy-move forgeries since academic works have not explicitly concentrated on these forms of manipulation thus far.
- The research in [67] shows that adversarial training together with specific techniques produces overfitting results because the method trains against a particular attack pattern. A model that precisely fits its training data becomes prone to overfitting, thus creating performance success on known examples but manifestation of failure for new observations. The inability to handle unseen data presentation variations and multiple noise patterns becomes a crucial challenge that can be overcome through the utilization of adaptive preprocessing and generalized feature extraction methods.
- The accuracy metrics among different research studies are widely divergent since some methodologies demonstrate full accuracy on MNIST [68] data, whereas others exhibit diminished performance on the CIFAR-10 and ImageNet datasets. The detection of real-world forgery remains challenging since the varying accuracy rates reflect the analysis difficulties encountered when inspecting images with noise together with both splicing and copy-move manipulations. According to [67], the described models achieve exceptional performance on simple datasets, yet their detection accuracy decreases notably when they handle sophisticated datasets that require training. When moving from CIFAR-10 to ImageNet, the percentage performance of the data changes from 91.5% to 75.1%, making this transition process more challenging. Research into reliable systems requires advanced features combined with preprocessing to achieve better robustness across different types of forgeries alongside unpredictable real-world conditions.

5. CHALLENGES

The current detection models face two major obstacles: precise detection of small duplicated areas and low contrast areas in copy-move forgeries and the ability to handle various types of noise and photometric attacks, such as JPEG compression and image blurring and salt and pepper noise effectively. The detection process faces additional difficulty because current methods struggle to identify high-quality splicing forgeries along with their natural variations and small section differences.

The current techniques for detecting data poisoning in training datasets fail to show enough resilience to observe sophisticated patterns of contamination, which degrades machine learning model integrity.

6. FUTURE DIRECTIONS

Future investigations should focus on specific technical advancements to increase the accuracy and robustness of forgery detection models. One promising direction is the enhancement of feature extraction methods by exploring hybrid models, such as combining convolutional neural networks (CNNs) with generative adversarial networks (GANs) or other techniques such as autoencoders. This hybrid approach could improve the detection of small duplicate areas, especially in challenging scenarios with low contrast. Furthermore, developing advanced preprocessing techniques, such as adaptive contrast adjustment or domain-specific filtering methods, will strengthen models against photometric attacks and noise disturbances. Additionally, integrating forgery detection platforms with data poisoning detection techniques can provide a multi-layered defense, ensuring that models are more resilient to adversarial attacks. Another key area of exploration is improving model generalizability by leveraging unsupervised learning methods to adapt to evolving data manipulations. Finally, leveraging transfer learning for training on diverse datasets could improve the model's robustness across different image domains.

7. CONCLUSION

A systematic literature review has analysed existing studies and evaluated convolutional neural networks and image forgery detection techniques with an emphasis on copy-move forgery, splice and noise manipulations. Even though forensic images undergo complex distortion processes, CNNs are capable of detecting many forms of forgeries with relative ease. However, difficulties can arise when trying to identify small duplication regions that are surrounded by boundless noise, low-contrast regions, or are under noise/photometric attack.

This systematic literature review looks at the remaining approaches dealing with deep learning technology and analyses their strengths as well as their weaknesses, such as the ability to generalize, crossover multiple datasets, and the presence of varying types of noise within the dataset. Additionally, the probing question regarding information poisoning image forensic systems suggests that these models are vulnerable to modern unfriendly actions and deep forgery techniques that require strong counteraction mechanisms.

The image forensics studies analysed in this review provided ample understanding of the strengths and weaknesses of CNNs. This information will be highly valuable for future work. This review also highlights a number of important areas that require research, including how to effectively combine forgery detection with data poisoning countermeasures. More effort is needed to improve the strength and generalizability of detection models.

Conflicts of Interest

The authors declare no conflict of interest.

Funding

This research received no external funding.

Acknowledgment:

None.

References

[1] P. Devarshi, U. Kosarkar and A. Chaube, "Comprehensive study on image forgery techniques using deep learning," in *2023 11th International Conference on Emerging Trends in Engineering & Technology-Signal and Information Processing (ICETET-SIP)*, (pp. 1-5). IEEE. doi: 10.1080/23742917.2023.2192888.

[2] K. Lalli, V. K. Shrivastava and R. Shekhar, "Detecting Copy Move Image Forgery using a Deep Learning Model: A Review," in *2023 International Conference on Artificial Intelligence and Applications (ICAIA) Alliance Technology Conference (ATCON-1)*, pp. 1-7. IEEE. doi: 10.1109/ICAIA57370.2023.10169568.

[3] R. Agarwal, D. Khudaniya, A. Gupta and K. Grover, "Image forgery detection and deep learning techniques: A review," in *2020 4th International Conference on Intelligent Computing and Control Systems (ICICCS)*, pp. 1096-1100. IEEE. DOI: 10.1109/ICICCS48265.2020.9121083.

[4] Abhishek, N. Jindal, "Copy move and splicing forgery detection using deep convolution neural network, and semantic segmentation," *Multimedia Tools and Applications,* vol. 80, no. 2, pp. 3571-3599, 2021. DOI: 10.1007/s11042-020-09816-3.

[5] R. Ch, M. Radha, M. Mahendar and P. Manasa, "A comparative analysis for deep learning-based approaches for image forgery detection," *International Journal of Systematic Innovation,* vol. 8, no. 1, pp. 1-10, 2024. DOI: 10.6977/IJoSI.202403_8(1).0001.

[6] T. T. Nguyen, Q. V. H. Nguyen, T. T. Nguyen, T. T. Huynh, T. T. Nguyen, M. Weidlich and H. Yin, "Manipulating recommender systems: A survey of poisoning attacks and countermeasures," *ACM Computing Surveys,* vol. 57, no. 1, pp. 1-39, 2024. DOI: 10.1145/3677328.

[7] L. Stroebel, M. Llewellyn, T. Hartley, T. S. Ip and M. Ahmed, "A systematic literature review on the effectiveness of deepfake detection techniques," *Journal of Cyber Security Technology,* vol. 7, no. 2, pp. 83-113, 2023. doi:10.1080/23742917.2023.2192888.

[8] S. Mehta and P. Shukla, "An Efficient Technique for Passive Image Forgery Detection Using Computational Intelligence," *Dark Web Pattern Recognition and Crime Analysis Using Machine Intelligence,* pp. 31-45, 2022. DOI: 10.4018/978-1-6684-3942-5.ch003.

[9] S. Kaur, R. Rani, R. Garg and N. Sharma, "State-of-the-art techniques for passive image forgery detection: a brief review," *International Journal of Electronic Security and Digital Forensics,* vol. 14, no. 5, pp. 456-473, 2022. doi:10.1504/IJESDF.2022.125403.

[10] D. Tralic, I. Zupancic, S. Grgic and M. Grgic, "CoMoFoD - New Database for Copy-Move Forgery Detection," in *Proceedings ELMAR-2013*, September 2013.

[11] J. Dong, W. Wang and T. Tan, "Casia image tampering detection evaluation database," in *2013 IEEE China Summit and International Conference on Signal and Information Processing*, China, pp. 422-426. IEEE. DOI: 10.1109/ChinaSIP.2013.6625374.

[12] R. A. Abdulhasan, S. T. A. Al-latief, and S. M. Kadhim, "Instant learning based on deep neural network with linear discriminant analysis features extraction for accurate iris recognition system," *Multimedia Tools and Applications*, vol. 83, no. 11, pp. 32099-32122, 2024.

[13] L. Truong, C. Jones, B. Hutchinson, A. August, B. Praggastis, R. Jasper, Nichols N and Tuor A., "Systematic Evaluation of Backdoor Data Poisoning Attacks on Image Classifiers," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) Workshops*, pp. 788-789.

[14] T. Liu, Y. Yang and B. Mirzasoleiman, "Friendly Noise against Adversarial Noise: A Powerful Defense against Data Poisoning Attacks," *Advances in Neural Information Processing Systems,* vol. 35, pp. 11947-11959, 2022.

[15] A. Cinà, K. Grosse, A. Demontis, B. Biggio, F. Roli and M. Pelillo, "Machine learning security against data poisoning: Are we there yet?," *Computer,* vol. 57, no. 3, pp. 26-34, 2024. IEEE. DOI: 10.1109/MC.2023.3299572.

[16] J. Lin, L. Dang, M. Rahouti and K. Xiong, "ML Attack Models: Adversarial Attacks and Data Poisoning Attacks," *arXiv preprint arXiv:2112.02797,* 2021. doi:10.48550/arXiv.2112.02797.

[17] K. Aryal, M. Gupta and M. Abdelsalam, "Analysis of label-flip poisoning attack on machine learning based malware detector," in *2022 IEEE International Conference on Big Data (Big Data)*, 2022.

[18] S. T. Abd Al-Latief, S. Yussof, A. Ahmad, S. M. Khadim, and R. A. Abdulhasan, "Instant Sign Language Recognition by WAR Strategy Algorithm Based Tuned Machine Learning," *International Journal of Networked and Distributed Computing*, vol. 12, no. 2, pp. 344-361, 2024.

[19] K. Ganesan, "Machine Learning Data Detection Poisoning Attacks Using Resource Schemes Multi-Linear Regression," *Neural, Parallel, & Scientific Computations,* vol. 28, no. 2, pp. 73-82, 2020. doi:10.46719/npsc20202821.

[20] A. Cinà, K. Grosse, A. Demontis, S. Vascon, W. Zellinger, B. Moser, A. Oprea, B. Biggio, M. Pelillo and F. Pelillo, "Wild patterns reloaded: A survey of machine learning security against training data poisoning," *ACM Computing Surveys,* vol. 55, no. 13s, pp. 1-39, 2023. doi:10.1145/3585385.

[21] J. Chen, X. Zhang, R. Zhang, C. Wang and L. Liu, "De-pois: An attack-agnostic defense against data poisoning attacks," *IEEE Transactions on Information Forensics and Security,* vol. 16, pp. 3412-3425, 2021. DOI: 10.1109/TIFS.2021.3080522.

[22] X. Xin, Y. Bai, H. Wang, Y. Mou and J. Tan, "An Anti-Poisoning Attack Method for Distributed AI System," *ournal of Computer and Communications,* vol. 9, no. 12, pp. 99-105, 2021. doi:10.4236/jcc.2021.912007.

[23] M. Ramirez, . S. Kim, H. Hamadi, E. Damiani, Y. Byon, T. Kim, C. Cho and C. Yeun, "Poisoning attacks and defenses on artificial intelligence: A survey," *arXiv preprint arXiv:2202.10276,* 2022. doi:10.48550/arXiv.2202.10276.

[24] J. Geiping, L. Fowl, G. Somepalli, M. Goldblum, M. Moeller and T. Goldstein, "What doesn't kill you makes you robust (er): How to adversarially train against data poisoning," *arXiv preprint arXiv:2102.13624,* 2021. doi:10.48550/arXiv.2102.13624.

[25] M. Goldblum, D. Tsipras, C. Xie, X. Chen, A. Schwarzschild, D. Song, A. Mądry, B. Li and T. Goldstein, "Dataset security for machine learning: Data poisoning, backdoor attacks, and defenses," *IEEE Transactions on Pattern Analysis and Machine Intelligence,* vol. 45, no. 2, pp. 1563-1580, 2022. doi:DOI: 10.1109/TPAMI.2022.3162397.

[26] Z. C. Y. a. W. L. Sheng, "Exploring multi-scale forgery clues for stereo super-resolution image forgery localization," *Pattern Recognition,* vol. 161, p. 111230, 2025.

[27] S. Chakraborty, K. Chatterjee and P. Dey, "Detection of Image Tampering Using Deep Learning, Error Levels and Noise Residuals," *Neural Processing Letters,* vol. 56, p. 112, 2024. doi:10.1007/s11063-024-11448-9.

[28] B. N. N., H. E., V. K. G., a. R. T. B. and N. B., "A Systematic Approach to Detect Spliced and Forged Images using Deep Learning Technique," *IJFANS INTERNATIONAL JOURNAL OF FOOD AND NUTRITIONAL SCIENCES,* vol. 11, no. 12, pp. 1806-1815, 2022. doi: 10.48047/IJFANS/V11/I12/191.

[29] H. Machiraju, M. H. Herzog, and P. Frossard, "Frequency-Based Vulnerability Analysis of Deep Learning Models against Image Corruptions," arXiv, arXiv:2306.07178, 2023. [Online]. Available: https://arxiv.org/abs/2306.07178

[30] L. Janutėnas, Janutėnaitė-Bogdanienė and D. J. and Šešok, "Deep Learning Methods to Detect Image Falsification.," *Applied Sciences,* vol. 13, no. 13, p. 7694, 2023. doi:10.3390/app13137694.

[31] D. Budgen and P. Brereton, "Performing systematic literature reviews in software engineering," in *In 2006 Proceedings of the 28th international conference on Software engineering*, pp. 1051-1052. doi:10.1145/1134285.1134500.

[32] Y. Pan, X. Ge, C. Fang and Y. Fan, "A systematic literature review of android malware detection using static analysis," *IEEE Access,* vol. 8, pp. 116363-116379, 2020. DOI: 10.1109/ACCESS.2020.3002842.

[33] T. Qazi, K. Hayat, S. Khan, S. Madani, I. Khan, J. Kołodziej, H. Li, W. Lin, K. Yow and C. Xu, "Survey on blind image forgery detection," *IET Image Processing,* vol. 7, no. 7, pp. 660-670, 2013. doi:10.1049/iet-ipr.2012.0388.

[34] I. Zedan, M. Soliman, K. Elsayed and H. Onsi, "Copy move forgery detection techniques: a comprehensive survey of challenges and future directions," *International Journal of Advanced Computer Science and Applications,* vol. 12, no. 7, 2021.

[35] K. Sunitha and K. A. N., "Efficient keypoint based copy move forgery detection method using hybrid feature extraction," in *2020 2nd International Conference on Innovative Mechanisms for Industry Applications (ICIMIA)*, pp. 670-675. IEEE, 2020. DOI: 10.1109/ICIMIA48430.2020.9074951.

[36] H. Chen, X. Yang and Y. Lyu, "Copy-move forgery detection based on keypoint clustering and similar neighborhood search algorithm," *IEEE Access,* vol. 8, pp. 36863-36875, 2020. DOI: 10.1109/ACCESS.2020.2974804.

[37] K. Rehman and S. Islam, "A Keypoint-Based Technique for Detecting the Copy Move Forgery in Digital Images," in *International Conference on Micro-Electronics and Telecommunication Engineering*, Singapore: Springer Nature Singapore, pp. 797-811. 2023. doi:10.1007/978-981-99-9562-2_66.

[38] X. Wang, C. Wang, L. Wang, H. Yang and P. Niu, "Robust and effective multiple copy-move forgeries detection," *Pattern Analysis and Applications,* vol. 24, pp. 1025-1046, 2021. doi:10.1007/s10044-021-00968-y.

[39] K. Sunitha and K. A. N., "Efficient keypoint based copy move forgery detection method using hybrid feature extraction," in *2020 2nd International Conference on Innovative Mechanisms for Industry Applications (ICIMIA)*, Bangalore, India, pp. 670-675. IEEE. 2020. DOI: 10.1109/ICIMIA48430.2020.9074951.

[40] P. Niu, C. Wang, W. Chen, H. Yang and X. Wang, "Fast and effective Keypoint-based image copy-move forgery detection using complex-valued moment invariants," *Journal of Visual Communication and Image Representation,* vol. 77, p. 103068, 2021. doi:10.1016/j.jvcir.2021.103068.

[41] I. T. Ahmed, B. T. Hammad and N. Jamil, "Image copy-move forgery detection algorithms based on spatial feature domain," in *2021 IEEE 17th International colloquium on signal processing & its applications (CSPA)*, pp. 92-96. IEEE. DOI: 10.1109/CSPA52141.2021.9377272.

[42] G. Tahaoğlu and G. Ulutas, "Copy-move forgery detection and localization with hybrid neural network approach," *Pamukkale Üniversitesi Mühendislik Bilimleri Dergisi,* vol. 28, no. 5, pp. 748-760, 2022. doi: 10.5505/pajes.2022.88714.

[43] G. Yue, Q. Duan, R. Liu, W. Peng, Y. Liao and J. Liu, "SMDAF: A novel keypoint based method for copy-move forgery detection," *IET Image Processing,* vol. 16, no. 13, pp. 3589-3602, 2022. doi:10.1049/ipr2.12578.

[44] X. Y. Wang, C. Wang, L. Wang, H. Y. Yang and P. Niu, "Robust and effective multiple copy-move forgeries detection and localization," *Pattern Analysis and Applications,* vol. 24, pp. 1025-1046, 2021. doi:10.1007/s10044-021-00968-y.

[45] A. Rani, A. Jain and M. Kumar, "Identification of copy-move and splicing based forgeries using advanced SURF and revised template matching," *Multimedia Tools and Applications,* vol. 80, no. 16, pp. 23877-23898, 2021. doi:10.1007/s11042-021-10810-6.

[46] K. Gayathri and P. S. Deepthi, "An overview of copy move forgery detection approaches," *Computer Science & Engineering International journal,* vol. 12, no. 6, pp. 81-94, 2022. DOI:10.5121/cseij.2022.12609 .

[47] C. Wang, Z. Zhang and X. Zhou, "An Image Copy-Move Forgery Detection Scheme Based on A-KAZE and SURF Features," *Symmetry,* vol. 10, no. 12, pp. 1-20, 2019. doi:10.3390/sym10120706.

[48] R. Dixit and R. Naskar, "Region duplication detection in digital images based on Centroid Linkage Clustering of key – points and graph similarity matching," *Multimedia Tools and Applications,* vol. 78, p. 13819–13840, 2019. doi:10.1007/s11042-018-6666-1.

[49] M. Thiiban , T.M., N. Abd Warif, A. Ismail and A. Mat , "An Evaluation of Convolutional Neural Network (CNN) Model for Copy-Move and Splicing Forgery Detection," *International Journal of Intelligent Systems and Applications in Engineering,* vol. 11, no. 2, pp. 730-740, 2023.

[50] E. Qazi, T. Zia and A. Almorjan, "Deep learning-based digital image forgery detection system," *Applied Sciences,* vol. 12, no. 6, p. 2851, 2022. doi:10.3390/app12062851.

[51] K. Hosny, A. Mortda, N. Lashin and M. Fouda, "A new method to detect splicing image forgery using convolutional neural network.," *Applied Sciences,* vol. 13, no. 3, p. 1272, 2023. doi:10.3390/app13031272.

[52] M. Islam, G. Karmakar, J. Kamruzzaman, M. Murshed, G. Kahandawa and N. Parvin, "Detecting splicing and copy-move attacks in color images," in *2018 Digital Image Computing: Techniques and Applications (DICTA)*, Canberra, ACT, Australia, pp. 1-7. IEEE. DOI: 10.1109/DICTA.2018.8615874.

[53] S. Ali, I. Ganapathi, N. Vu, . S. Ali, N. Saxena and N. Werghi, "Image forgery detection using deep learning by recompressing images," *Electronics,* vol. 11, no. 3, p. 403, 2022. DOI: 10.1109/DICTA.2018.8615874.

[54] N. Hussien, R. Mahmoud and H. Zayed, "Deep learning on digital image splicing detection using CFA artifacts," *International Journal of Sociotechnology and Knowledge Development (IJSKD),* vol. 12, no. 2, pp. 31-44, 2020. DOI: 10.4018/IJSKD.2020040102.

[55] D. Mallick, M. Shaikh, A. Gulhane and T. Maktum, "Copy move and splicing image forgery detection using CNN," in *In ITM Web of Conferences. International Conference on Automation, Computing and Communication 2022 (ICACC-2022)*, Vol. 44, p. 03052. EDP Sciences. doi:10.1051/itmconf/20224403052.

[56] A. Kuznetsov, "Digital image forgery detection using deep learning approach," *Journal of Physics,* vol. 1368, no. 3, p. 032028, 2019. DOI 10.1088/1742-6596/1368/3/032028.

[57] K. Meena and V. Tyagi, "A deep learning based method for image splicing detection," *journal of physics: conference series,* vol. 1714, no. 1, p. 012038, 2021. DOI 10.1088/1742-6596/1714/1/012038.

[58] M. Elaskily, H. Elnemr, A. Sedik, M. Dessouky, G. El Banby, O. Elshakankiry, A. Khalaf, H. Aslan, O. Faragallah and F. Abd El-Samie, "A novel deep learning framework for copy-moveforgery detection in images," *Multimedia Tools and Applications,* vol. 79, pp. 19167-19192, 2020. doi:10.1007/s11042-020-08751-7.

[59] C. Prakash, A. Kumar, S. Maheshkar and V. Maheshkar, "An integrated method of copy-move and splicing for image forgery detection," *Multimedia Tools and Applications,* vol. 77, pp. 26939-26963, 2018. doi:10.1007/s11042-018-5899-3.

[60] J. PATEL and N. BHATT, "COPY-MOVE FORGERY DETECTION-A HYBRID APPROACH," *Journal of Engineering Science and Technology,* vol. 17, no. 3, pp. 2000-2019, 2022.

[61] S. Bibi, A. Abbasi, I. Haq, S. Baik and A. Ullah, "Digital image forgery detection using deep autoencoder and CNN features," *Hum. Cent. Comput. Inf. Sci,* vol. 11, pp. 1-17, 2021. DOI:10.22967/HCIS.2021.11.032.

[62] K. Hosny, A. Mortda, M. Fouda and N. Lashin, "An efficient cnn model to detect copy-move image forgery," *IEEE Access,* vol. 10, pp. 48622-48632, 2022. DOI: 10.1109/ACCESS.2022.3172273.

[63] V. Raghavan, T. Mazzuchi and S. Sarkani, "An improved real time detection of data poisoning attacks in Deep Learning Vision systems," *Discover Artificial Intelligence,* vol. 2, no. 1, p. 18, 2022. doi:10.1007/s44163-022-00035-3.

[64] F. F. M. M. D. G. A. S. Martinelli, "Data poisoning attacks over diabetic retinopathy images classification," *2023 IEEE International Conference on Big Data (BigData),* pp. 3698-3703. IEEE, 2023.

[65] N. Peri, N. Gupta, W. Huang, L. Fowl, C. Zhu, S. Feizi, T. Goldstein and J. Dickerson, "Deep k-nn defense against clean-label data poisoning attacks," *Computer Vision–ECCV 2020 Workshops: Glasgow, UK, August 23–28, 2020, Proceedings,* vol. Part I 16, pp. 55-70, August 23–28 2020. Springer International Publishing. doi:10.1007/978-3-030-66415-2_4.

[66] A. Omran, S. Mohammed and M. Aljanabi, "Detecting Data Poisoning Attacks in Federated Learning for Healthcare Applications Using Deep Learning," *Iraqi Journal For Computer Science and Mathematics,* vol. 4, no. 4, pp. 225-237, 2023. doi:10.52866/ ijcsm.2023.04.04.018.

[67] X. Chen, Y. Ma, S. Lu and Y. Yao, "Boundary augment: A data augment method to defend poison attack," *IET Image Processing,* vol. 15, no. 13, pp. 3292-3303, 2021. doi:10.1049/ipr2.12325.

[68] D. Khuu, M. Sober, D. Kaaser, M. Fischer and S. Schulte, "Data Poisoning Detection in Federated Learning," in *Proceedings of the 39th ACM/SIGAPP Symposium on Applied Computing, pp. 1549-1558. 2024*, doi:10.1145/3605098.3635896.