International Research Journal of Multidisciplinary Scope (IRJMS), 2024; 5(1): 781-789 IRJMS

Original Article | ISSN (O): 2582-631X DOI: 10.47857/irjms.2024.v05i01.0299

# Using a Clustering Algorithm and a Transform Function, Identify Forged Images

Shudhodhan Bokefode[1]*, Jayesh Sarwade[2], Kishor Sakure[1], Sandeep Bankar[3], Surekha Janrao[4], Rohini Patil[1]

[1]Terna Engineering College, Nerul Navi Mumbai, Maharashtra, India. [2]Rajarshi Shahu College of Engineering Pune, Maharashtra, India. [3]SVKM's NMIMS Navi Mumbai, Maharashtra, India. [4]K J Somaiya Institute of Technology, Mumbai, Maharashtra, India. *Corresponding Author's Email: shudhodhan358@gmail.com

## Abstract

In the realm of digital multimedia analysis, combating picture fraud stands as a critical endeavor, given the prevalence of manipulation facilitated by contemporary multimedia creation tools. The ease of copying, pasting, and tampering with digital images poses a significant challenge, altering the reality of the original image and constituting an illegal operation. While existing pixel- and transform-based methods have exhibited superior detection and estimation capabilities, they are not without limitations. This research proposes a novel texture-based approach for the identification of fake images, aiming to overcome the drawbacks of current methodologies. The suggested method proves effective in terms of detection ratio, offering a promising solution to the complexities introduced by contemporary multimedia manipulation. The extraction of texture features is accomplished using a discrete wavelet transform algorithm, enhancing the robustness of the forgery detection process. A crucial aspect of the proposed approach is the block creation method, implemented through the partition clustering approach. This facilitates the creation of blocks for both genuine and manipulated images, contributing to a comprehensive analysis. To validate the efficacy of the suggested method, extensive testing is conducted utilizing the well-known MFIC2000 dataset and MATLAB software. The outcomes of this research not only shed light on the advancements in texture-based forgery detection but also provide a practical and efficient solution address the challenges posed by picture fraud in the contemporary multimedia landscape. The proposed methodology contributes significantly to the evolving field of multimedia forensics, enhancing the arsenal of tools available for detecting and mitigating digital image tampering.

**Keywords:** Cluster segmentation, DWT, texture, and image tampering.

## Introduction

Digital image forensics heavily relies on the picture forgery detection procedure. Forgery detecting software and technologies were employed for the picture analysis. Two images were needed for the picture forgery detection process: one original and one fake. There are several different picture forgery detection techniques being utilized nowadays (1, 2) the two methods of domain for picture forgeries are pixel-based and transform-based, respectively, as a result of the sampling and processing of the image. The bogus photos were made using a variety of methods, including copy and paste, image slicing, image enhancement, and picture painting. This study suggests a brand-new picture faking technique. The wavelet transform function and clustering methodology are the foundation of the suggested strategy (3). The texture feature is provided by the wavelet transforms function. The texture characteristic of a digital image is crucial and lower in content. The block and pattern were created using the extracted texture feature. Clustering approach was utilized to create blocks and patterns using fake and authentic images. K-means clustering algorithm was utilized to conduct the clustering. Popular clustering technique K-means is simple to use for assessing patterns and blocks (4). The study proposes a novel approach for concealing data within images by combining cryptography and deep neural networks. By leveraging advanced techniques, the research demonstrates the effectiveness of securely embedding information in images (5). When comparing several block locations, evaluate the degree of similarity that exists between them. Person coefficient derivation was utilized to measure the similarity gap. The person coefficient quantifies how much the real and fake images resemble and differ from one.



(Received 24th October 2023; Accepted 17th January 2024; Published 30th January 2024)

There are a number of drawbacks to the feature selection and region selection method used in texture-based photo fraud detection (6). Identifying how much the real and phony photographs resemble one another is the key challenge. Selecting detecting features in the best method feasible. The amount of noise in the image is more than the strength of the actual image. The edges of the manufactured picture are not precise. The majority of edited photographs are fakes. These types of constraints are minimized by the recommended method (7, 8). Several genuine and fake photos are used to assess the proposed approaches. Our experimental findings show that the suggested strategies are extremely appealing. With copy-move alone, copy-move with rotation, scaling, and reflection, the forgery is accomplished. During this process, an image database made up of real and fraudulent photographs is also produced (9, 10). The study presents a deep convolutional neural network tailored for age assessment using orthopantomography data. Published in Neural Computing and Applications in 2020, the research demonstrates the effectiveness of the designed network in accurately determining age based on dental imaging. This contributes to advancements in age assessment methodologies, particularly in the context of orthopantomography (11). With rotation, scaling, and reflection, the suggested technique achieves 99.12% accuracy in copy-move forgeries, respectively, and 100% accuracy in copy-move counterfeiting without post-processing (without changing the object's size or characteristics). To ensure higher effectiveness, we also added more random noise to the images; the detection accuracy was 98.23%. The proposed method and process of image forgeries are described in the aforementioned section (12). The research, presented at the International Conference on Machine Learning in 2020, focuses on leveraging frequency analysis for deep fake image recognition. The study highlights the effectiveness of incorporating frequency-based techniques in identifying manipulated images, contributing to advancements in deep fake detection methodologies (13). The study explores the use of deep convolutional neural networks (CNNs) for identifying materials in both photographic images and photorealistic computer-generated graphics. Published in Computers, Materials & Continua in 2018, the research demonstrates the efficacy of deep CNNs in accurately discerning the material composition of images. This contributes to advancements in material identification within diverse visual content (14). The research, published in IET Image Processing in 2018, introduces a Convolutional Neural Network (CNN) designed for detecting smooth filtering. The study demonstrates the effectiveness of the CNN in identifying the presence of smoothing filters in images, contributing to advancements in image processing techniques (15). The study provides a comprehensive bibliography of digital image anti-forensics and anti-anti-forensics techniques. It serves as a valuable resource for researchers by compiling relevant literature on the methods and countermeasures employed in the field of digital image forensics and its adversarial aspects (16). The study, published in the International Journal of Computer Network and Information Security in 2019, introduces a passive approach for detecting image splicing. Utilizing deep learning and Haar wavelet transform, the research demonstrates the effectiveness of the proposed method in identifying manipulated images, contributing to advancements in image forensics (17). The study presents a method for image deblocking detection using a Convolutional Neural Network (CNN). The research showcases CNN's effectiveness in identifying artifacts from image compression, contributing to advancements in image quality assessment and processing (18). The study, published in IET Image Processing in 2021, introduces a Dual Branch Convolutional Neural Network for copy-move forgery detection. The research demonstrates the network's efficacy in identifying instances of image tampering, particularly in cases of copy-move forgery, contributing to advancements in digital image forensics (19). The research, presented at the European Conference on Computer Vision (ECCV) in 2018, introduces a Deep Convolutional Neural Network for detecting double JPEG compression in images with mixed quality factors. The study demonstrates the network's effectiveness in identifying instances of double JPEG compression, contributing to advancements in image forensics and compression analysis (20). The study, published in the Journal of Physics: Conference Series in 2019, employs a deep learning approach for digital image forgery detection. The research

explores the efficacy of deep learning techniques in identifying various forms of image manipulations, contributing to advancements in digital forensics (21). The work discusses a method for determination of bone age including two steps the feature extraction and classification method. The feature extraction utilizes depth neural network to study the features of X-ray image, and the LBP features and GCLM features in the image are extracted (22) the segmentation is carried out.

Discuss the feature extraction approach in part II. In paragraph III. Discuss the suggested approach. Discuss the analysis of the experimental results in part IV before moving on to section V to discuss the conclusion and future work.

Various methods enhance the efficacy of spotting manipulated photographs. Digital forensics relies on metadata analysis, exposing inconsistencies in timestamps or compression artifacts. Deep learning, particularly Convolutional Neural Networks (CNNs) and Generative Adversarial Networks (GANs), proves effective in learning manipulation patterns. Error Level Analysis (ELA) detects variations in compression levels within images. Block chain and cryptography offer reliable timestamps and digital signatures for authenticity verification. Consistency checks, analyzing lighting, shadows, and geometric anomalies, play a crucial role. Image analysis tools, using metrics like SSI and PSNR, contribute to alteration identification. Forensic software, especially designed for Adobe Photoshop files, aids in scrutinizing digital alterations. However, challenges persist, requiring continuous research to stay ahead of evolving manipulation techniques. In image analysis, the applied transform function and clustering technique are integral components woven into the overall procedure to unveil patterns within images. The applied transform function, like Fourier or wavelet transforms, serves during preprocessing to modify pixel values, emphasizing certain features or suppressing noise. Concurrently, clustering techniques, particularly K-means clustering, categorize pixels into groups based on shared characteristics. This tandem approach begins with preprocessing, where the transform function enhances specific features. Feature extraction follows, utilizing the transformed image to discern essential information for subsequent analysis. Clustering comes into play during the feature extraction phase, grouping similar elements for a more structured representation. The clustered data is then subjected to pattern recognition, aiding in identifying structures or anomalies. Finally, this information informs decision-making, allowing for object identification, anomaly detection, or image classification. The synergy between the applied transform function and clustering technique enriches image analysis, providing insights into image content and facilitating diverse analytical objectives. The selection of specific transform functions and clustering algorithms depends on the unique characteristics and goals of the image analysis task. To assess the effectiveness of the transform function and clustering method in image analysis, various metrics like accuracy, precision, recall, and others play a pivotal role. Accuracy measures the overall correctness of the analysis, indicating the proportion of correctly identified patterns to the total. Precision gauges the reliability of positive identifications, revealing the ratio of correctly identified positive instances to all instances classified as positive. Recall, on the other hand, assesses the ability to capture all relevant patterns by identifying the ratio of correctly identified positive instances to the total actual positive instances. F1 score, combining precision and recall, offers a balanced metric, particularly useful when there's an uneven class distribution. These metrics collectively provide a comprehensive evaluation of the transform function and clustering method, offering insights into their ability to accurately identify patterns and contribute to the overall success of image analysis tasks. Regularly employing such metrics ensures a quantitative assessment, aiding in the refinement and optimization of the chosen methods.

## Feature extraction

The lower content characteristic of a digital image serves as the foundation for the image counterfeit Detection technique. The three main categories of characteristics seen in digital images are color, texture, and form and size elements. For the study of picture fraud detection, the texture feature is one of the most crucial attributes. Use of a feature extractor was made for the extraction of texture features. The texture feature extractor used the function of the wavelet transform. The wavelet transform function is a popular tool for extracting

texture features. In essence, the wavelet transform function combines lower and higher Frequencies. Here are descriptions of the sampling procedure used to compare authentic and fake images. Here's a simplified, text-based representation of how clustering algorithms and transform functions behave differently when parameter values are changed.

Transform Function

Algorithm

1. Apply a generic transform to an image.

The transform involves multiplying each pixel value by a transform parameter.

2. Behavior: Effect of Changing Parameter:

Increasing the parameter intensifies features and may introduce noise.

Decreasing the parameter softens features, potentially losing detail.

3. Clustering Algorithm:

Use the K-means clustering algorithm on the transformed image.

Vary the number of clusters (k) as a parameter.

4. Behavior: Effect of Changing Parameter:

Increasing k may create smaller, more detailed clusters.

Decreasing k might merge clusters, potentially oversimplifying the interpretation.

5. Visualization:

Display the original image, clustered image, and cluster centers for both default and changed parameters.

Observe and compare how alterations in transform and clustering parameters impact the visual representation of the data.

Adjusting parameters in the transform function and clustering algorithm influences the interpretation of images.

Finding the right balance is crucial to avoid overemphasis or underrepresentation of features.

This simplified overview highlights the key steps and outcomes when exploring the behavior of clustering algorithms and transform functions with changing parameters.

## Methodology

The recommended method incorporates two algorithms: a wavelet transforms function and a clustering mechanism. The wavelet transform function provides the textural feature of the real and false images. After features from the original and faked images have been extracted, create a clustering pattern. Block matching was necessary for the cluster to develop in order to continue the detection procedure.

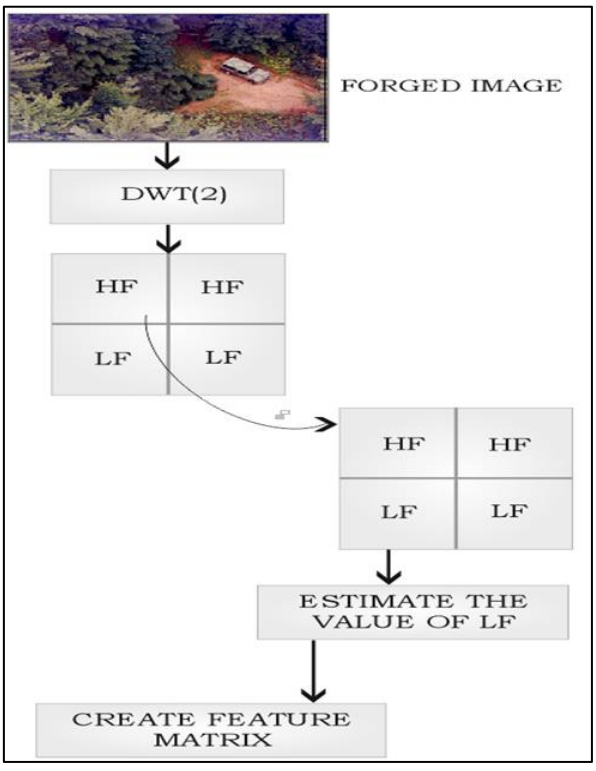


**Figure 1:** The block diagram of the feature extraction method for a fake image

**Algorithm**

Steps 1 a texture feature sample

Equations [1] and [2] are used in the cluster mapping process to distribute the texture feature for the purpose of processing texture feature data mapping.

$$a(i,j) = \frac{(a^*(i,j) - a(j))}{(a_{max}(j) - a_{min}(j))} \quad [1]$$

Creates blocks for the point-to-point comparison of the faked and genuine images:

$$a(i,j) = \frac{((a_{max}(j) - a^*(i,j))}{(a_{max}(j) - a_{min}(j))} \quad [2]$$

**Step 2** calculates the block's counterfeit image Q (b) value.

$\{a(i,j)\text{I } j = 1, 2 \ldots., p\}$ is block point for the pattern $b = [b(1), b(2), \ldots\ldots, b(p)]$ as:

$$z(i) = \sum_{j=1}^{p} b(j)a(i,j), \quad i = 1, 2, \ldots\ldots n \quad [3]$$

Calculate the shared elements between the original and fake images.

$$Q(b) = S_z\, D_z \quad [4]$$

S_z and mapping D_z is specified in formula [5], where S_z is the similar block of the original image z (I), and D_z is the forged image block.

$$\begin{cases} S_z = \sqrt{\frac{\sum_{i=1}^{n}(z(i) - E(z))^2}{(n-1)}} \\ D_z = \sum_{i=1}^{n}.\sum_{j=1}^{n}\big(R - r(i,j)\big)\, u(R - r(i,j\,)) \end{cases} \quad [5]$$

.

[1] d(z(k), z(h)) is defined as the exact distance between two patterns of an authentic and faked picture. [6]

$$d\big(z(k), z(h)\big) = \sqrt{\big(z(k) - z(h)\big)\big(z(k) - z(h)\big)} = \sqrt{\big(z(k) - z(h)\big)^2}$$

$k = 1, 2, \ldots\ldots\ldots\ldots, N; h = 1, 2, \ldots\ldots\ldots\ldots, N$

**Step 4** evaluate how comparable the block pattern is.

$$\begin{cases} s.t. \sum_{j=1}^{p} a^2(j) = 1 \\ 1 \geq a(j) \geq 0 \end{cases} \quad [7]$$

**Step 5** declare the image's region to be fake.

Equation (3) was used to validate the cluster of both images.

## Experimental result

All software performance characteristics are computed with this programme utilising both established and newly created method methods in the MATLAB simulation environment. This section defines the experimental data analysis using both current and new methodologies. Here, the false

**Table 1:** Demonstrates that the performance assessment made use of segmentation and suggested techniques

| Variety of Images | Name of Method | FN | FP |
|---|---|---|---|
| Image-1 | Segmentation | 17.405 | 35.060 |
| | Proposed | 14.027 | 32.060 |
| Image-2 | Segmentation | 6.866 | 37.600 |
| | Proposed | 5.404 | 34.600 |
| Image-3 | Segmentation | 12.753 | 32.543 |
| | Proposed | 11.654 | 30.757 |

negative rate (also known as FN) and false positive rate (also known as FP) ratios of missed detection to fake images and false alarm to actual images, respectively, the detection error at the picture level is calculated using,

$F_N$= | {forged pictures detected as original} | / | {Forged pictures} |

$F_P$= | {Original pictures detected as forged} | / | {Original pictures} |

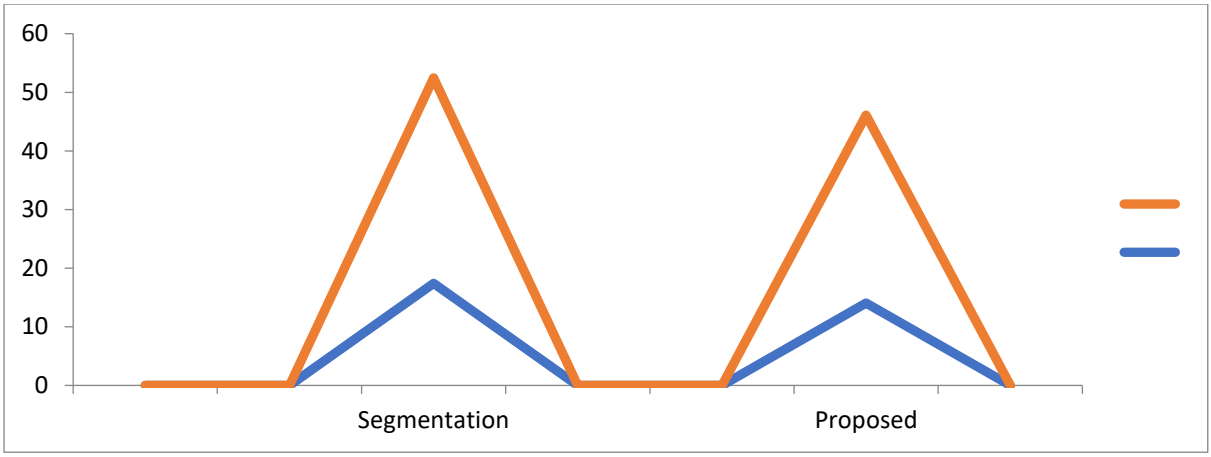


**Figure 2:** Displays the graphs of the comparative performance evaluation for the proposed approaches utilizing image-1 and segmentation for the FN And FP

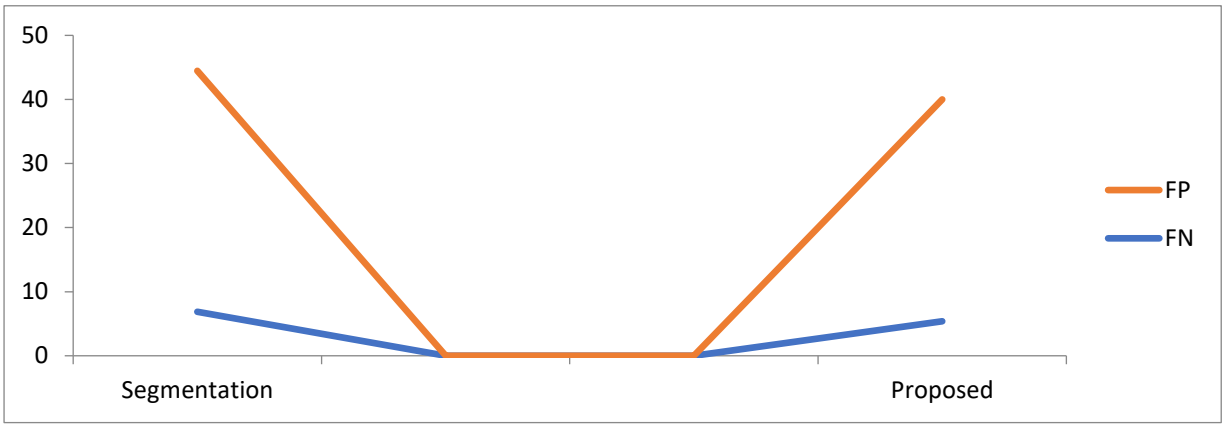


**Figure 3:** Displays the graphs of the comparative performance evaluation for the proposed approaches utilizing image-2 and segmentation for the FN And FP

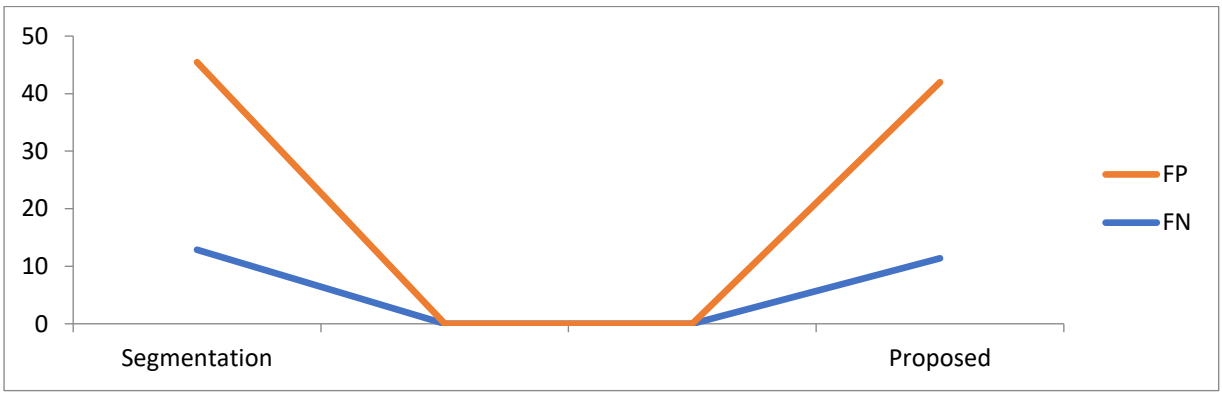


**Figure 4:** Graphs of comparative performance evaluation for FN and FP utilizing segmentation and recommended ways using image-3 are shown

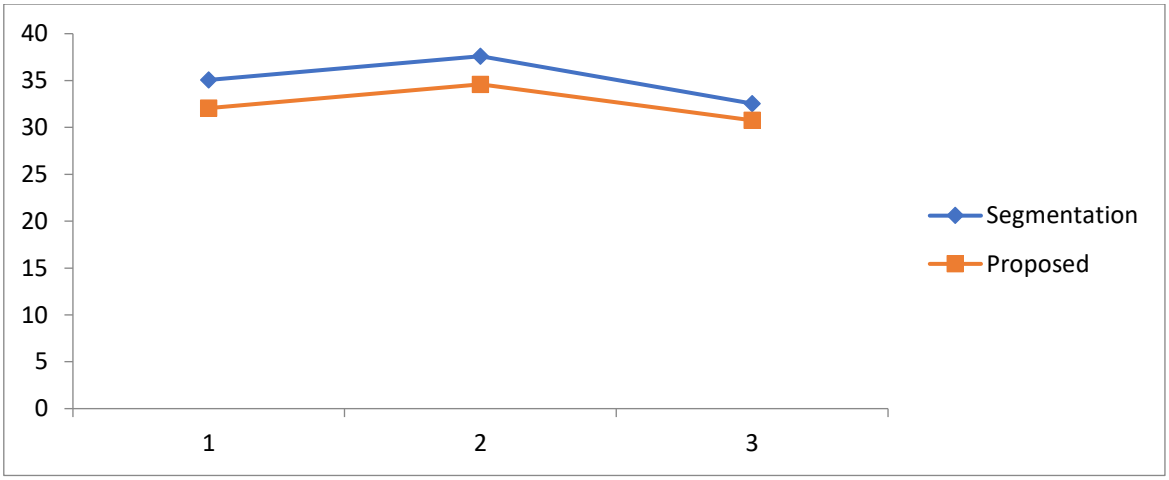


**Figure 5:** Displays the graphs of the comparative performance evaluation for the proposed approaches utilizing image-1 and segmentation for the FN And FP

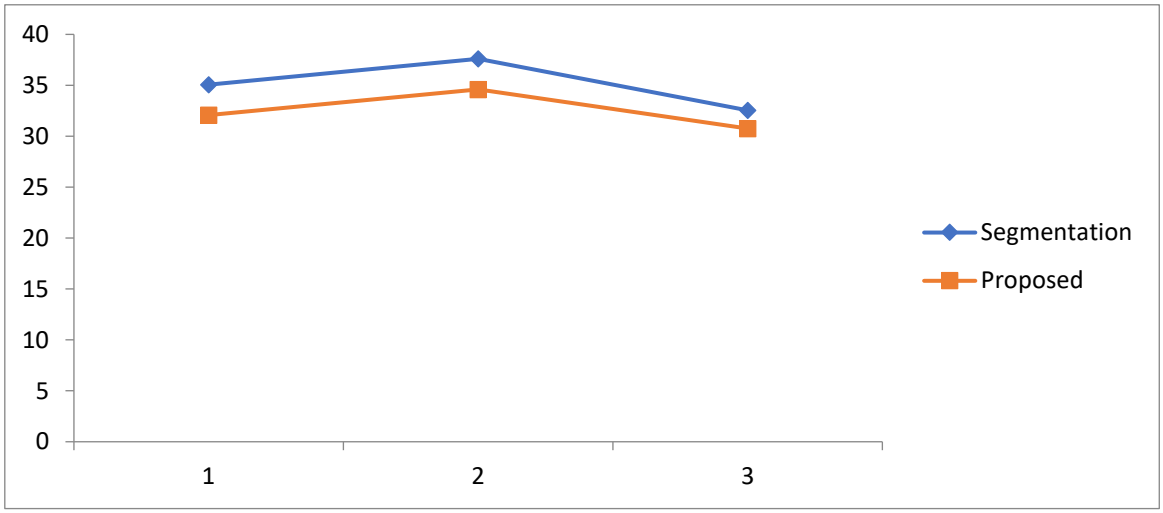


**Figure 6:** Displays the graphs of the comparative performance evaluation for the proposed approaches utilizing image-2 and segmentation for the FN And FP

Experimental results of the proposed Image forgery detection method with DWT and proposed method, with using various types of image database in which include such as Playground image, Forest image, Scene and water fall image. The entire image database extracted from the website search engine like Google etc. in the experimental process we perform these entire data base image and find the value of PSNR and FRR In both DWT and Proposed method (Table 1 and Figure 1-6).

## Conclusion and future work

This work presents a wavelet-based approach for grouping images that may be used to identify picture forgeries. The suggested approach made use of a cluster selection technique and data distribution technique. The suggested method utilised two images, one of which is authentic and the other is fake. Blocks of pattern are produced by the distribution of data using the partition clustering approach. The standard deviation calculation determines how much the actual and faked images differ from one another. The suggested approach was tested using the MFICC2000 standard forgery picture dataset and MATLAB simulation tools. The suggested approach increases the value of detection while decreasing the value of false negatives. In the future, feature optimisation techniques will be utilised to improve pattern creation during matching.

## Abbreviations

Nil

## Acknowledgement

The authors would like to thank to Terna Engineering College, Nerul, Navi Mumbai and Mumbai University

## Author contributions

The collaborative effort on this forged image identification method saw (Dr. Shudhodhan Bokefode and Dr. Jayesh Sarwade) conceptualizing the integration of a clustering algorithm and a transform function. (Dr. Kishor Sakure and Dr.Sandeep Bankar) implemented and fine-tuned the clustering algorithm, while (Dr. Surekha Janrao) developed the transform function to detect subtle alterations. (Dr. Rohini Patil) played a key role in data collection and validation. All authors contributed actively to manuscript preparation, reflecting a collective advancement in digital forensics through this innovative approach.

## Conflict of interest

There is no conflict of interest in relation to this study from either parties involved.

## Ethics approval

Not applicable

## Funding

The study was funded by the authors.

# References

1. Murthy A, Sampath Dakshina T, Karthikeyan B, Omkar Lakshmi Jagan, Ch Usha Kumari. "Novel Deep Neural Network For Individual Re Recognizing Physically Disabled Individuals." Materials Today: Proceedings. 2020; 33: 4323-4328. https://doi.org/10.1016/j.matpr.2020.07.447
2. Bazrafkan Shabab, Shejin Thavalengal, Peter Corcoran. "An End-To-End Deep Neural Network For Iris Segmentation In Unconstrained Scenarios." Neural Networks. 2018; 106: 79-95. https://doi.org/10.1016/j.dsp.2021.103244
3. Li Y, Chang MC, Farid H, Lyu S. "In Ictu Oculi: Exposing AI Generated Fake Face Videos By Detecting Eye Blinking. " Arxiv Preprint Arxiv. 2018;1806: 02877. https://www.researchgate.net/publication/325680345
4. Rafi Abdul Muntakim, Thamidul Islam Tonmoy, Uday Kamal, Jonathan Wu QM, Md Kamrul Hasan. "Remnet: Remnant Convolutional Neural Network For Camera Model Identification." Neural Computing And Applications. 2021; 33: 3655-3670. 3655-3670.Avilable from:DOI: 10.1007/s00521-020-05220-y
5. Sharma Kartik, Ashutosh Aggarwal, Tanay Singhania, Deepak Gupta, Ashish Khanna. "Hiding Data In Images Using Cryptography And Deep Neural Network." Arxiv Preprint Arxiv. 2019; 1912:10413. https://doi.org/10.33969/AIS.2019.11009.
6. Duan Xintao, Daidou Guo, Nao Liu, Baoxia Li, Mengxiao Gou, Chuan Qin. "A New High-Capacity Image Steganography Method Combined With Image Elliptic Curve Cryptography And Deep Neural Network." IEEE Access 8. 2020; 25777-25788. DOI 10.1109/ACCESS.2020.2971528
7. Kelly Finnian, Oscar Forth, Samuel Kent, Linda Gerlach, Anil Alexander. "Deep Neural Network Based Forensic Automatic Speaker Recognition In VOCALISE Using X-Vectors." In Audio Engineering Society Conference: 2019 AES International Conference On Audio Forensics. Audio Engineering Society. 2019. https://www.aes.org/tmpFiles/elib/20240128/20477.pdf
8. Afifi Mahmoud, Michael SB. "What Else Can Fool Deep Learning? Addressing Color Constancy Errors On Deep Neural Network Performance." In Proceedings Of The IEEE/CVF International Conference On Computer Vision. 2019; Pp. 243-252. DOI: 10.1109/ICCV.2019.00033
9. Nataraj Lakshmanan, Tajuddin Manhar Mohammed, Manjunath BS, Shivkumar Chandrasekaran, Arjuna Flenner, Jawadul H, Amit KRC. "Detecting GAN Generated Fake Images Using Co-Occurrence Matrices." Electronic Imaging. 2019; 5:532. DOI: 10.2352/ISSN.2470-1173.2019.5.MWSF-532
10. Bammey Quentin, Rafael Grompone Von Gioi, Jean-Michel Morel. "An Adaptive Neural Network For Unsupervised Mosaic Consistency Analysis In Image Forensics." In Proceedings Of The IEEE/CVF Conference On Computer Vision And Pattern Recognition. 2020; Pp. 14194-14204. https://www.researchgate.net/publication/353791120
11. Kahaki Seyed MM, Md Jan Nordin, Nazatul SA, Mahir Arzoky, Waidah Ismail. "Deep Convolutional Neural Network Designed For Age Assessment Based On Orthopantomography Data." Neural Computing And Applications. 2020; 32(13): 9357-9368. DOI:10.1007/s00521-019-04449-6
12. Cristin RB, Santhosh Kumar, Priya C, Karthick K. "Deep Neural Network-Based Rider-Cuckoo Search Algorithm For Plant Disease Detection." Artificial Intelligence Review. 202; 53(7):2020. https://doi.org/10.1007/s10462-020-09813-w
13. Frank Joel, Thorsten Eisenhofer, Lea Schönherr, Asja Fischer, Dorothea Kolossa, Thorsten Holz. "Leveraging Frequency Analysis For Deep Fake Image Recognition." In International Conference On Machine Learning. PMLR. 2020; Pp. 3247-3258. https://proceedings.mlr.press/v119/frank20a/frank20a.pdf
14. Cui Qi, Suzanne Mcintosh, Huiyu Sun. "Identifying Materials Of Photographic Images And Photorealistic Computer-Generated Graphics Based On Deep Cnns." Comput Mater Continua. 2018; 55(2): 229-241.https://www.techscience.com/cmc/info/cmc_previous_contents
15. Yang Bin, Xingming Sun, Enguo Cao, Weifeng Hu, Xianyi Chen. "Convolutional Neural Network For Smooth Filtering Detection." IET Image Processing.

2018; 12(8): 1432-1438. https://doi.org/10.1049/iet-ipr.2017.0683
16. Qureshi Muhammad Ali, And El-Sayed M El-Alfy. "Bibliography Of Digital Image Anti-Forensics And Anti-Anti-Forensics Techniques." IET Image Processing. 2019; 13(11): 1811-1823. https://doi.org/10.1049/iet-ipr.2018.6587
17. Abd El-Latif, Eman I, Ahmed Taha, Hala HZ. "A Passive Approach For Detecting Image Splicing Using Deep Learning And Haar Wavelet Transform." International Journal Of Computer Network And Information Security. 2019; 11(5): 28. DOI: 10.5815/ijcnis.2019.05.04
18. Liu Xianjin, Wei Lu, Wanteng Liu, Shangjun Luo, Yaohua Liang, Ming Li. "Image Deblocking Detection Based On A Convolutional Neural Network." IEEE Access. 2019; 7: 26432-26439. https://doi.org/10.1109/ACCESS.2019.2901020
19. Goel Nidhi, Samarjeet Kaur, Ruchika Bala. "Dual Branch Convolutional Neural Network For Copy Move Forgery Detection." IET Image Processing. 2021. https://doi.org/10.1049/ipr2.12051
20. Park Jinseok, Donghyeon Cho, Wonhyuk Ahn, Heung-Kyu Lee. "Double JPEG Detection In Mixed JPEG Quality Factors Using Deep Convolutional Neural Network." In Proceedings Of The European Conference On Computer Vision (ECCV). 2018; Pp. 636-652. https://openaccess.thecvf.com/content_ECCV_2018/papers/JinSeok_Park_Double_JPEG_Detection_ECCV_2018_paper.pdf
21. Kuznetsov A. "Digital Image Forgery Detection Using Deep Learning Approach." In Journal Of Physics: Conference Series. 2019; 1368(3):032028. https://dx.doi.org/10.2139/ssrn.3734785
22. Chen Xu, Jianjun Li, Yanchao Zhang, Yu Lu, Shaoyu Liu. "Automatic Feature Extraction In X-Ray Image Based On Deep Learning Approach For Determination Of Bone Age." Future Generation Computer Systems. 2020; 110: 795-801. DOI: 10.48550/arXiv.2206.05641