Image Tampering Detection in Social Media

FAIZAN MAZHAR QURESHI, University of Twente, The Netherlands

Abstract 'A picture is worth a thousand words' is a common phrase used worldwide, but in today's era, these pictures can be edited, causing the truth to change. Manipulated images, including fabricated news, and merged images are widespread on platforms such as Facebook and Twitter, contributing to the spread of false and harmful content. Public opinion can easily be swayed via this method and can result in uninformed decisions. Thus, exploring ways to detect image tampering is critical for the well-being of social media users. Current research primarily focuses on the detection of tampered images, but social, moral, and psychological issues caused by image tampering in social media are not widely addressed. Moreover, research lacks in providing a complete solution to this rising problem of widespread tampered images over social media. It also mostly focuses only on single and common techniques like Error Level Analysis (ELA) and Double JPEG Compression. In this research, we evaluate the problem of the spread of tampered images on social media and its consequences. We develop a solution by utilizing some existing advanced image analysis techniques, such as Discrete Cosine Transformation (DCT) coefficient, Histogram Value Analysis, Noise Variance Inconsistencies, and ELA. These techniques have not been extensively explored in the field of detecting manipulated images. We integrated these techniques by training machine learning models and incorporating them into our developed web application. The solution was developed using a combination of Python, JavaScript, and HTML. The results help users validate image authenticity on social media platforms by directly uploading or providing the image link in the web application, reducing tampered image spread.

Additional Key Words and Phrases: Image Tamper, Social media platforms, Authenticity, Image Manipulation, Tamper Detection

1 INTRODUCTION

Over the last few decades, the advancement of social media and its increased accessibility has led it to become the most popular and widely used means of accessing information [32]. However, along with its widespread use, the increase of tampered and fake images has become a significant issue, posing a major challenge to the spread of authentic information. A recent survey by Ipsos on behalf of the Centre for International Governance Innovation shows every four in ten (44%) admit to being duped by fake news over social media [1]. Many researchers have talked about how important this issue is becoming in the real world [27, 31].

To tackle the growing issue of image tampering, researchers are exploring new avenues which include the development of various methods for the detection of image tampering such as ELA, DCT, and Noise Variance Inconsistencies. These techniques have been discussed and utilized in this research [12, 16, 23].

Integrating different types of methods to create a solution that can simultaneously detect image tampering can aid in addressing the issue of manipulated images on social media. In this research, we considered existing techniques to develop a robust solution using a multifaceted approach that can be utilized to tackle the issue of tampered images in social media.

The research is structured as follows: In Section 2, we discuss the problem statement and the subsequent research questions. Section 3 delves into related work done in the domain. Section 4 focuses on the methodology employed to answer the research questions arising from the problem, followed by Section 5, which discusses the results of the research. Section 6 discusses our conclusion, followed by section 7 which entails the future work that could be done in the domain. Finally, we conclude with references to relevant research.

2 PROBLEM STATEMENT

With the increase in the utilization of images on social media, the use of tampered images has also become a prevalent and worrying problem [22, 24]. Image manipulation has potential issues, including influencing public opinions, affecting human emotions, and even causing changes in democracy [2, 5, 28]. Existing literature showcases that multiple users have been fooled by social media posts [1, 11].

Furthermore, research in progress for detecting manipulated images focuses on techniques of image tamper detection, such as DCT coefficient analysis, Local Histogram Analysis, Noise Variance Inconsistencies, and ELA [4, 23, 36], but most of them do not address how these techniques can be employed to develop a holistic solution for social media platforms such as Facebook and Twitter that can be utilized to reduce or mitigate the problem of the spread of tampered images on them.

The objective of this research is to explore and analyze the issues caused by tampered images on social media platforms such as Facebook and Twitter. The primary aim is to develop a solution that effectively controls the spread of manipulated images within these social media environments. This involves employing existing advanced image analysis techniques, including DCT coefficient analysis, Local Histogram Analysis, and ELA to detect manipulated images. The research also assesses the extent to which images on social media undergo manipulation, with the ultimate goal of enhancing the overall security and reliability of digital content platforms. This involves addressing the challenges associated with the widespread use of tampered images by forming a solution that can aid in detection.

2.1 Research Questions

The problem statement leads to the following research questions:

2.1.1 *What issues arise in social media platforms as a result of image tampering?*

2.1.2 *What are the advanced image analysis techniques mentioned above that can be employed to detect tampered or fake images over social media platforms such as Facebook and Twitter?*

TScIT 40, February 2, 2024, Enschede, The Netherlands

2.1.3 *How can solution using advanced image analysis techniques be utilized and integrated to control and mitigate the spread and use of tampered images in social media?*

2.1.4 *How prevalent are tampered images over social media platforms?*

3 RELATED WORK

Image tamper detection has become a significant research focus in the past decade, primarily due to the increase in fake images on social media. This is largely attributable to the readily available technology and advanced editing software, which have made image manipulation more accessible. To compile relevant literature in the field, reputable platforms like IEEE [1], Elsevier [2], ResearchGate [3], and Google Scholar [4] were employed, utilizing keywords like 'image tampering' and 'social media' to refine the search.

Manipulated images can cause users to believe in disinformation. In an article in 2020, Hameleers, Michael, et al. demonstrated that adding images increases the perceived credibility of disinformation [7]. Another famous example of social media users believing tampered images is documented in a well-known article regarding Hurricane Sandy in 2012 [6]. It revealed the role of the social media website Twitter in the spread of fake images during the Hurricane. 10350 tweets containing fake images were identified on Twitter and out of these 80 percent were retweets. Examples of fake images included forged images of sharks in New Jersey and in streets, this caused immense damage and led to feelings of panic among citizens affected by the hurricane causing them to have additional mental stress while going through a crisis. Events like this diminish users' trust in social media content as was proven by Stubenvoll M, Heiss R, and Matthes J in a study where they mentioned that due to high levels of perceived misinformation exposure, social media users have decreased media trust [29].

An article by Kara et al in 2018 showed that in a study, many individuals could not detect skillfully manipulated images. Moreover, it was also revealed that many participants in the study did not question the authenticity of the images they were shown even after being warned earlier about the possibility of fake images [10]. These results proved most viewers cannot easily differentiate between real and tampered images. This can raise issues including spreading disinformation. Furthermore, In 2020 another paper, P Maji, M Pal, R Ray, R Shil [17] discussed how widespread the problem of tampered images is becoming in social media. The authors proposed an algorithm using feature extraction and statistical analysis to detect tampered and manipulated images in social media. The results later showed 82 images out of around 100 downloaded from social media were tampered with, thus signifying the prevalence and issue of tampered images on social media platforms.

In 2018, Lilei Zheng and Ying Zhang [37] closely examined the issue of image manipulation by individuals, particularly in everyday photos. They delved into the significance of this problem. The authors also explored available data sets for image tamper detection and discussed their evolution, designating CASIA v2 as the most recent and optimal data set for detecting image tampering. Furthermore, they explored various methods such as DCT coefficients to ascertain whether a picture has undergone manipulation. Utilizing this available information to further develop a holistic solution is a challenge that is taken up in our research.

In a paper published in 2009 [13], Z Lin, J He, X Tang, and CK Tang used DCT analysis to find tampered images. They examined the double quantization effect on the image using DCT coefficients to detect and locate the Tampered region in the image. However, their research only had promising results on JPEG images.

J. Madake, J. Meshram, A. Mondhe, and P. Mashalkar also proposed a method for detecting forged regions in images using ELA in an article published in 2023 [15]. In the proposed method, ELA is employed to analyze the image in detail, focusing on uniformity of color and brightness along edges. Authentic images typically demonstrate consistent brightness along edges and lower ELA values across the image. By applying an ELA filter, subtle changes that may not be visually apparent can be detected, as these variations could indicate tampering. Tampering can introduce lousy pixels in certain areas, leading to inconsistencies in ELA values and thus aiding in tampering detection. This demonstrates the importance of the ELA technique in the field of fake image detection.

Image analysis using the Histogram approach has been mentioned in the chapter Image Standardization in PACS of Handbook of Medical Imaging [26], which discusses how image grayscale value distribution showing grey scale frequency can be utilized to analyze the image based on the difference between their uniformity. However, these frequency distributions have not been utilized for Image Tamper detection.

An article published in 2021 [9], proposed the usage of a noise inconsistency-based technique to detect forged images and also localize false regions in an image. The steps of the method included pre-processing followed by noise estimation and then post-processing. The proposed technique demonstrated exceptional results hence signifying the importance of the notice inconsistency-based method for the detection of tampered images.

The following section will discuss the research method and outline how the research questions will be addressed.

4 METHODOLOGY

In this section, we have outlined our research methodology and detailed the measures undertaken to tackle the research questions. To achieve our research objectives, we began with an in-depth literature review and examined established techniques employed in prior studies including DCT coefficient, Noise Variance Inconsistencies, and ELA. These techniques have demonstrated effectiveness and promising results in image tamper detection [13, 21, 23]. We utilized these techniques along with Image Histogram analysis to develop a solution for the problem of Image tampering in social media. In our solution, we employed a multifaceted approach by integrating all these techniques to develop a web application that can detect manipulated images from social media platforms.

[1] https://ieeexplore.ieee.org/
[2] https://www.sciencedirect.com/
[3] https://www.researchgate.net/
[4] https://scholar.google.com/

4.1 Dataset

To conduct this research, we relied on the publicly available Casia v2 dataset, a compilation of tampered and authentic images consisting of approximately 7,200 authentic images and 5100 tampered images [3]. Another available dataset is MICC-F600. This dataset contains only 160 tampered and 440 original images, thus making the Casia v2 dataset our preference [35].

The most important reason for choosing Casia v2 dataset was that it is one of the largest publicly available datasets and encompasses various types of tampering, including image splicing, blurring, and manipulation using software. Additionally, it presents diverse image sizes ranging from 320x240 to 800x600. The dataset also includes images from different categories such as scene, animal, architecture, character, plant, article, nature, indoor, and texture. We utilized approximately 1,200 tampered and 1,200 authentic images, randomly selected from the entire dataset, while adhering to systematic limitations in preprocessing and machine learning.

4.2 On Answering Research Question 1

The proliferation of manipulation of images on social media platforms has led to the rise of various issues which include social, moral, and psychological aspects.

Tampered images on social media can spread disinformation. False narratives created using images can spread rumors and change the portrayal of actual information. The topics affected by manipulated images can have a vast range of subjects, spanning from details regarding hurricanes to vaccine debates [6, 18]. An article published by V Schetinger et al. (2017) [25] revealed that humans can be easily duped by fake news in digital images. The research results showed that people were only able to correctly identify manipulated images 58 percent of the time. Moreover, it was also seen that people in the study could only identify 46.5 percent of actual forgeries. Another article by SJ Nightingale et al. [20] demonstrated humans' inability to distinguish real and manipulated photos and also revealed that even after identifying tampered images, most individuals could not locate the manipulation. This deficit of users to recognize and separate true images from false ones aids in the spread of disinformation [6, 18]. Furthermore, the widespread presence of tampered images on social media platforms also leads to a decrease in the credibility of the forum. Users lose trust in social media and view information obtained through the source with increasing distrust [30]. Finding methods to detect altered images is essential to prevent the spread of false information and restore the credibility of social media platforms.

Manipulated images on social media often entail modification of individual pictures. This can lead to serious privacy concerns, as images can be altered and used by individuals with malicious intentions. Blackmailers can use fake images and extort valuables from victims who don't believe they can prove the tampering or are afraid that the damage done after sharing images on social media cannot be reversed even after proving the image tampering[2]. Images on social media can even be utilized by individuals to make pornographic images which can raise further privacy issues on social media[2]. Moreover, fake images on social media also can be used to generate decisions about individuals. Studies have shown that employers refuse to interview or hire people multiple times due to inappropriate photos in their search results [2, 8]. If these images were fake or tampered with, this could have serious repercussions for the candidates in the job market. Thus underscoring the importance of detecting manipulated images from real ones.

Encountering altered images on social media can cause users to have profound psychological and social effects. Viewing tampered images can lead to emotional distress in viewers and can affect most types of relationships including personal, business, and political affiliations [28]. Alteration of personal images can cause significant mental harm to users, affecting feelings of self-worth as proven in studies [14, 33]. Furthermore, It can also result in incitement of violent behavior and political unrest [2, 5]. This can result in serious consequences and even lead to violence [28]. The social and psychological effects emphasize the gravity of the issue of tampered images on social media and demonstrate the need for the development of a detection method.

The tampering of images in social media gives rise to a multitude of problems. Development of many techniques for the detection of image alteration has been done [19], this research focuses on creating a multifaceted solution utilizing the previous sources available.

4.3 On Answering Research Question 2

After recognizing the gravity of the issue of tampered images on social media platforms, we analyzed the advanced image analysis techniques and devised a custom algorithm using machine learning. We trained different machine learning models for techniques such as ELA, Histogram Values Analysis, DCT coefficient analysis, and Noise Variance level Analysis. Images in Fig. 1, 2 are used to visualize the following techniques.

Fig. 1. Authentic Image

Fig. 2. Tampered Image

Error Level Analysis ELA is a forensic technique designed to identify image segments in JPEG, particularly those with changing compression levels. It works by analyzing the compression error differences and making the tampered regions stand out, thereby identifying tampered or manipulated regions. Fig. 3 shows error levels of an authentic image, it can be seen that the image has consistent error levels throughout the image indicating uniform compression. However, Fig. 4 exhibits error levels of the tampered image and it shows varying compression and distortion in error levels throughout the image.

Histogram Values Analysis Histogram values of the image represent the distribution of pixel intensities throughout the grayscale image [26], with pixel intensities ranging from 0 to 255. In terms of image tamper detection, the intensities of pixels are analyzed. An authentic image exhibits a consistent pattern of histogram intensities

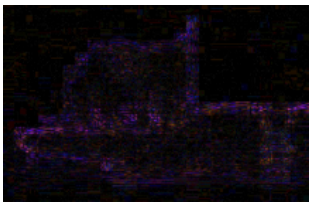

Fig. 3. Authentic ELA

Fig. 4. Tampered ELA

throughout the image. If the image is tampered with, the histogram values deviate from the normal pattern of intensities, highlighting alterations or tampering in the image. As visualized in Fig. 5, It shows a consistent and natural distribution of histogram values for authentic, while Fig. 6 exhibits irregularities and anomalies in the histogram values indicating potential manipulation.

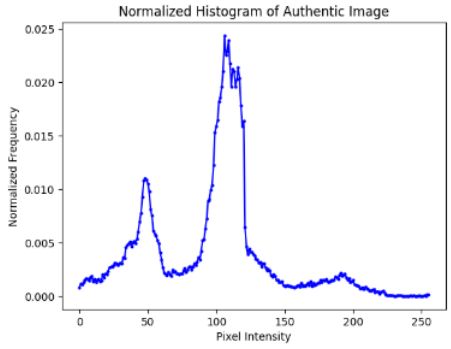

Fig. 5. Authentic Image Histogram Values

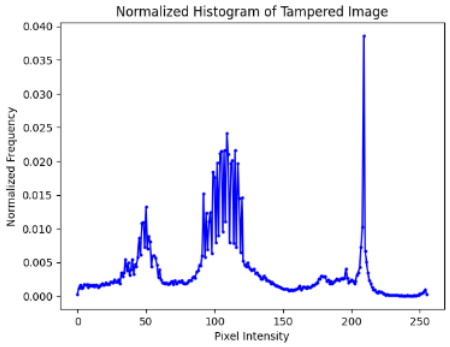

Fig. 6. Tampered Image Histogram Values

Discrete Cosine Transformation Coefficient DCT is a mathematical technique used in various signal processing and image compression applications. DCT transforms an image from its spatial domain to its frequency domain, representing the image block as a sum of cosine functions using the following equation:

$$\mathrm{DCT}(u,v) = C(u)\cdot C(v)\cdot \sum_{x=0}^{N-1}\sum_{y=0}^{N-1} f(x,y)\cdot \cos\left(\frac{(2x+1)u\pi}{2N}\right)\cdot \cos\left(\frac{(2y+1)v\pi}{2N}\right)$$

where

$$C(u) = \begin{cases} \frac{1}{\sqrt{2}}, & \text{if } u = 0 \\ 1, & \text{if } u > 0 \end{cases}$$

$$C(v) = \begin{cases} \frac{1}{\sqrt{2}}, & \text{if } v = 0 \\ 1, & \text{if } v > 0 \end{cases}$$

and u and v are the spatial frequencies in the horizontal and vertical directions, respectively.

DCT coefficients, in the context of image processing, refer to the values obtained after applying the DCT to the image. These coefficients represent the contribution of different frequency components in the image. In detecting image tampering, we analyze the DCT coefficients of different regions of the image, where anomalies in the distribution of the DCT coefficients indicate that the image has undergone tampering. To further aid in understanding, we demonstrate in Fig. 7 a small section of DCT coefficients of an authentic image that exhibits a consistent and uniform distribution of the DCT coefficients. However, Fig. 8 shows DCT coefficients of the same section of the image after undergoing tampering, revealing the inconsistencies and discontinuities in the DCT coefficients of the tampered image.

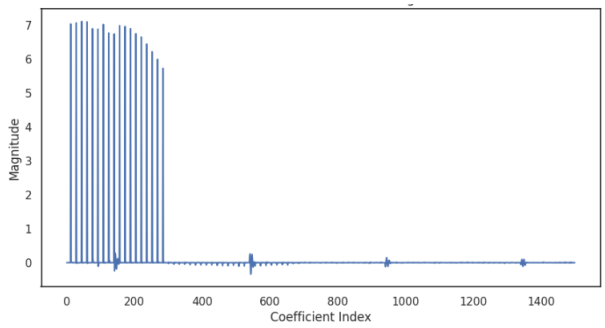

Fig. 7. Authentic Image DCT Coefficient

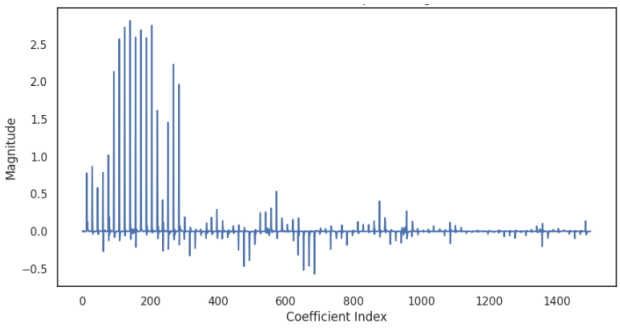

Fig. 8. Tampered Image DCT Coefficient

Noise Variance Inconsistencies Noise variance inconsistency refers to the analysis of variations in the level of noise across different regions of an image. The tampered and manipulated image uses blending or pasting which introduces irregularities. Using statistical

techniques, deviations in noise from the expected pattern can be identified and flagged. These inconsistencies may serve as evidence of tampering or editing [21]. In Fig. 1 and 2, we added Gaussian noise to visually represent the difference between authentic and tampered images. As seen in Fig. 9 and 10, there are significant inconsistencies in the noise of the tampered image compared to the authentic image. To enhance visualization, we utilized the Viridis filter in Fig. 11 and 12, which better highlights the differences and inconsistencies.

Fig. 9. Authentic Gray Noise Map

Fig. 10. Tampered Gray Noise Map

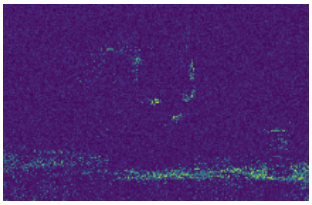

Fig. 11. Authentic Virid Noise Map

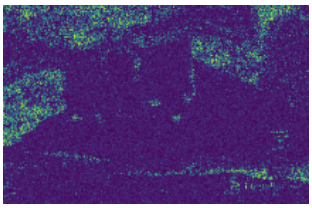

Fig. 12. Tampered Virid Noise Map

Using all of the aforementioned techniques, in this research, we aimed to develop a solution for image tamper detection in social media. We created a web application that can detect tampering in images. In the web application, you can either upload an image or directly enter the Image link. It provides information on the accuracy or probability with which it has classified the image as tampered or authentic.

4.4 On Answering Research Question 3

In developing our Image Tamper Detection system for social media, we employed various tools, including Python and its necessary libraries such as TensorFlow, Scikit-learn, and Keras, to implement the previously discussed techniques. We utilized machine learning to train models for the specific features of each of these techniques. Additionally, we created a web application using Flask API, seamlessly integrating it with individual trained models for each technique. To further enhance result accuracy, we implemented weighted voting for all models, with their accuracies serving as the weight. This approach allows images to be classified as accurately as possible.

In Fig. 13, we present an overview of our system, starting with the upload of an image to our web page or the input of an image link on our web pages. During the preprocessing phase, the image undergoes individualized processing for each technique. DCT prediction involves calculating the DCT values of the image, while for Histogram, all histogram values are computed. ELA calculates its values, and Noise Variance involves computing the noise of the image. Following this, each of these preprocessed features is fed into its respective trained model to predict the outcome of the image classification. The predictions from each trained model, along with their accuracy during training, contribute to a weighted voting system using equation (1). This equation takes the normalized accuracy of each model, as described below, and multiplies it by its corresponding output. The accuracy of each model (DCT, HIST, ELA, Noise) is normalized using equation (2) by dividing the model accuracy by the sum of all models' accuracies before being integrated into equation (1):

$$\begin{aligned}\text{Final Probability} = {} & \text{ELA Accuracy} \times \text{ELA Prediction} \\ & + \text{DCT Accuracy} \times \text{DCT Prediction} \\ & + \text{Hist Accuracy} \times \text{Hist Prediction} \\ & + \text{Noise Accuracy} \times \text{Noise Prediction}\end{aligned} \tag{1}$$

$$\text{Model Normalized Accuracy} = \left(\frac{\text{Individual Model Accuracy}}{\text{DCT Accuracy} + \text{ELA Accuracy} + \text{HIST Accuracy} + \text{Noise Accuracy}} \right) \tag{2}$$

For the practical utilization of our solution, we incorporated features enabling users to upload saved images from social media or input image links directly into our web application. This functionality allows users to instantly check whether the provided image is tampered or authentic, providing insights into how many of our featured models classified it as tampered. By employing this solution, users can avoid believing or sharing tampered images on social media, thus enhancing the overall authenticity of images circulating within the social media sphere.

4.5 On Answering Research Question 4

In our research aimed at addressing the question of the prevalence of tampered images in social media, we utilized a Random Facebook Image dataset published by Harvard [34]. This dataset comprises a diverse collection of over a thousand images sourced randomly from Facebook, encompassing various categories such as politics and fake news. Out of this large dataset, we randomly picked 50 images for comprehensive scrutiny. Employing our custom-developed solution, we subjected these selected images to analysis for tamper detection.

Following this, we applied quantitative analysis to estimate the percentage of tampered images circulating on social media (For more information, refer to Section 5).

5 RESULTS & DISCUSSION

In this study, we leveraged advanced image analysis techniques, as discussed earlier. Our method involved extracting data from the original images for each technique, utilizing this data as features, and subjecting it to preprocessing to facilitate the training of our machine learning models. For the training of machine learning models—specifically for DCT, Histogram, and noise variance—we employed six distinct algorithms: Logistic Regression (LR), Support Vector Machine (SVM), Decision Tree (DT), Random Forest (RF), Gradient Boosting (GB), and K-Nearest Neighbors (KNN). We meticulously assessed the performance of each model based on metrics such as accuracy, precision, recall, and F1 scores. This thorough evaluation process enabled us to pinpoint the model exhibiting the most exceptional performance. Table 1 provides a complete overview of their performances :

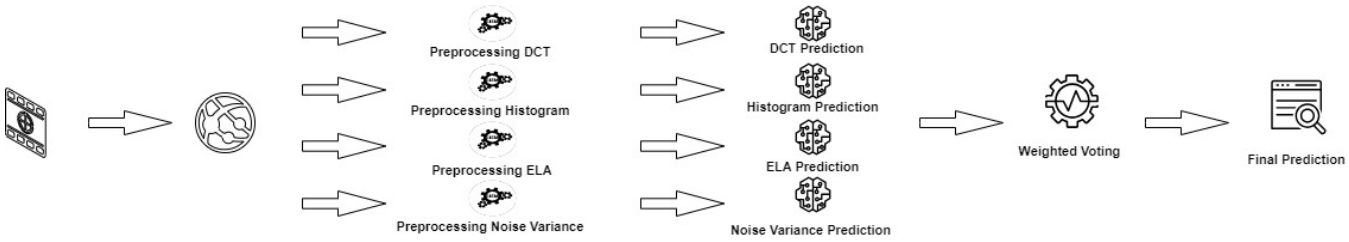

Fig. 13. Overview of Image Tamper Detection

	Performances					
	LR	SVM	RF	GB	KNN	DT
DCT	70.43%	81.65%	78.69%	84.02%	76.92%	66.86%
Histogram	79.28%	79.88%	82.24%	85.9%	71.59%	78.10%
Noise	68.75%	67.91%	65.83%	70.83%	63.33%	66.56%

Table 1. Overview of Model Accuracy

Different models exhibited varying performances on distinct types of features; thus, we selected the best-performing model for each technique.

For our DCT, we opted for the Gradient Boosting model, which demonstrated the highest overall accuracy among other models at 84.02%. Additionally, it exhibited the highest precision in terms of classifying authentic and tampered images.

In Histogram Values Analysis, we opted for the random forest model. Despite a slightly lower overall accuracy compared to Gradient Boosting, it excelled in precision, particularly in distinguishing between different image types, with a focus on tampered ones.

For noise variance, each model exhibited limitations when compared with the performances of models on histogram and DCT values. However, the Random Forest model performed better in comparison to other models of noise variance hence making it our choice.

In addition to these techniques, we incorporated a Convolutional Neural Network (CNN) built with Keras for error-level analysis. The CNN comprises two convolutional layers for feature extraction, followed by max-pooling and dropout layers to enhance efficiency and prevent overfitting. Utilizing the same image sample as in other methods and adopting an iterative approach with increased epochs, the model was exposed to the entire dataset, resulting in an impressive accuracy of 93.64% in the final iteration.

Each of the selected models for DCT, Histogram Values, and Noise Variance, along with the CNN model for ELA, were integrated into our web application. Each trained model predicts whether the given image is Authentic or Tampered based on its feature values, determining whether it exhibits a pattern similar to tampered or authentic image features learned by the models. Their predictions were combined using weighted voting based on the model accuracy to provide the final image prediction as discussed in 4.4.

In the developed web application, users can either upload the image using the upload button or they can enter the image link in the link input box as shown in Fig. 14. After they have either uploaded the image or entered the image link, they can press the Detect Tamper button which then checks whether the uploaded image is authentic or tampered. The developed solution will output the final result using the prediction of all the models along with their individual predictions. An example output is shown in Fig. 15

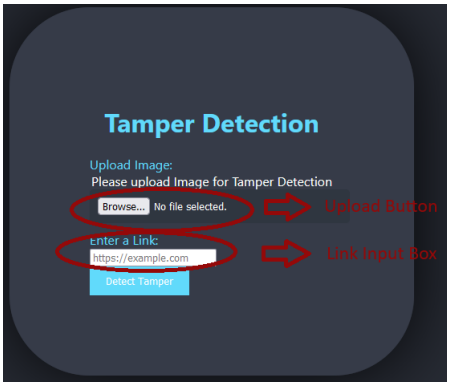

Fig. 14. Web Application Home

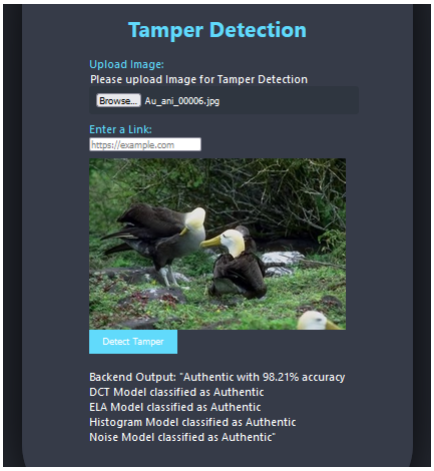

Fig. 15. Web Application with Example Results

In our final evaluation of how prevalent tampered images are over social media, we took a random sample of 50 publicly available

images from Facebook and tested them using our developed solution, and out of 50 images, 27 of them were classified as tampered. Using this we can say 54% of images over social media especially Facebook are Tampered, we verified this using quantitative analysis by applying a Z-test on the population and we failed to reject our null hypothesis that the actual size of tampered differs at 5 % significance level.

Moreover, we tested our developed web application solution across different image formats, including TIFF, JPEG, PNG, and WEBP. It worked perfectly for all tampering scenarios, accurately classifying tampered and authentic images. This demonstration proved that image format does not affect the performance of our system. Our developed web application solution also performed flawlessly, regardless of whether the uploaded image is colored, grayscale, or monochrome. This observation underscores that these different types do not impact the system, as the statistical image feature values remain consistent across all variations.

One of the important novelties of this research was using Histogram pixel value for Image Tamper Detection which had been previously used in different research for image analysis but was never to our knowledge utilized to detect tampering in Images. In addition to this, we developed a complete solution to control the problem of Image Tampering in social media which social media users can practically utilize.

6 CONCLUSION

In conclusion, Tampered images on social media are causing a loss of trust due to a high percentage of manipulated images circulating across platforms. This leads to social, moral, and psychological issues stemming from the lack of authentic content. To address this problem, we have developed a custom web application incorporating advanced techniques such as Discrete Cosine Transformation (DCT) coefficient analysis, Histogram Value Analysis, Noise Variance Inconsistencies detection, and Error Level Analysis (ELA) using machine learning. This web application enables users to verify the authenticity of image content by either uploading the image or entering its link. It helps users easily identify tampered or fake images, thus reducing the spread of manipulated content.

7 FUTURE WORK

This research could be extended to integrate with popular social media platforms like Facebook and Twitter. This integration could take place either within the platforms' own web and mobile applications or through the development of an external web browser extension. Such an extension would have the capability to identify tampered images, highlight them for users, and provide warnings when users attempt to share manipulated images. In addition to this, research could also be continued in the direction of achieving higher accuracy of models.

REFERENCES

[1] Year of the survey. CIGI-Ipsos Global Survey on Internet Security and Trust. Online survey. https://www.cigionline.org/cigi-ipsos-global-survey-internet-security-and-trust/ Conducted by CIGI (Centre for International Governance Innovation) and Ipsos..

[2] Bobby Chesney and Danielle Citron. 2019. Deep fakes: A looming challenge for privacy, democracy, and national security. *Calif. L. Rev.* 107 (2019), 1753.

[3] Jing Dong, Wei Wang, and Tieniu Tan. 2013. CASIA Image Tampering Detection Evaluation Database. In *2013 IEEE China Summit and International Conference on Signal and Information Processing*. 422–426. https://doi.org/10.1109/ChinaSIP.2013.6625374

[4] Shilpa Dua, Jyotsna Singh, and Harish Parthasarathy. 2020. Image forgery detection based on statistical features of block DCT coefficients. *Procedia Computer Science* 171 (2020), 369–378.

[5] Hany Farid. 2019. Image Forensics. *Annual Review of Vision Science* 5, 1 (2019), 549–573. https://doi.org/10.1146/annurev-vision-091718-014827 PMID: 31525144.

[6] Aditi Gupta, Hemank Lamba, Ponnurangam Kumaraguru, and Anupam Joshi. 2013. Faking sandy: characterizing and identifying fake images on twitter during hurricane sandy. In *Proceedings of the 22nd international conference on World Wide Web*. 729–736.

[7] Michael Hameleers, Thomas E Powell, Toni GLA Van Der Meer, and Lieke Bos. 2020. A picture paints a thousand lies? The effects and mechanisms of multimodal disinformation and rebuttals disseminated via social media. *Political communication* 37, 2 (2020), 281–301.

[8] Jenna Jacobson and Anatoliy Gruzd. 2020. Cybervetting job applicants on social media: the new normal? *Ethics and Information Technology* 22 (2020), 175–195.

[9] Ankit Kumar Jaiswal and Rajeev Srivastava. 2021. Forensic image analysis using inconsistent noise pattern. *Pattern Analysis and Applications* 24 (2021), 655–667.

[10] Mona Kasra, Cuihua Shen, and James F O'Brien. 2018. Seeing is believing: How people fail to identify fake images on the web. In *Extended abstracts of the 2018 CHI conference on human factors in computing systems*. 1–6.

[11] Srijan Kumar and Neil Shah. 2018. False Information on Web and Social Media: A Survey. *CoRR* abs/1804.08559 (2018). arXiv:1804.08559 http://arxiv.org/abs/1804.08559

[12] Zhouchen Lin, Junfeng He, Xiaoou Tang, and Chi-Keung Tang. 2009. Fast, automatic and fine-grained tampered JPEG image detection via DCT coefficient analysis. *Pattern Recognition* 42, 11 (2009), 2492–2501. https://doi.org/10.1016/j.patcog.2009.03.019

[13] Zhouchen Lin, Junfeng He, Xiaoou Tang, and Chi-Keung Tang. 2009. Fast, automatic and fine-grained tampered JPEG image detection via DCT coefficient analysis. *Pattern Recognition* 42, 11 (2009), 2492–2501.

[14] Alexandra Rhodes Lonergan, Kay Bussey, Jonathan Mond, Olivia Brown, Scott Griffiths, Stuart B Murray, and Deborah Mitchison. 2019. Me, my selfie, and I: The relationship between editing and posting selfies and body dissatisfaction in men and women. *Body image* 28 (2019), 39–43.

[15] Jyoti Madake, Jayant Meshram, Ajinkya Mondhe, and Pruthviraj Mashalkar. 2023. Image Tampering Detection Using Error Level Analysis and Metadata Analysis. In *2023 4th International Conference for Emerging Technology (INCET)*. 1–7. https://doi.org/10.1109/INCET57972.2023.10169948

[16] Babak Mahdian and Stanislav Saic. 2009. Using noise inconsistencies for blind image forensics. *Image and Vision Computing* 27, 10 (2009), 1497–1503. https://doi.org/10.1016/j.imavis.2009.02.001 Special Section: Computer Vision Methods for Ambient Intelligence.

[17] Prasenjit Maji, Moumita Pal, Ranjana Ray, and Riya Shil. 2020. Image Tampering Issues in Social Media with Proper Detection. In *2020 8th International Conference on Reliability, Infocom Technologies and Optimization (Trends and Future Directions) (ICRITO)*. 1272–1275. https://doi.org/10.1109/ICRITO48877.2020.9197780

[18] Elena Milani, Emma Weitkamp, and Peter Webb. 2020. The visual vaccine debate on Twitter: A social network analysis. *Media and Communication* 8, 2 (2020), 364–375.

[19] Minati Mishra and Flt Adhikary. 2013. Digital image tamper detection techniques-a comprehensive study. *arXiv preprint arXiv:1306.6737* (2013).

[20] Sophie J Nightingale, Kimberley A Wade, and Derrick G Watson. 2017. Can people identify original and manipulated photos of real-world scenes? *Cognitive research: principles and implications* 2, 1 (2017), 1–21.

[21] Xunyu Pan, Xing Zhang, and Siwei Lyu. 2012. Exposing image splicing with inconsistent local noise variances. In *2012 IEEE International Conference on Computational Photography (ICCP)*. 1–10. https://doi.org/10.1109/ICCPhot.2012.6215223

[22] Peng Qi, Juan Cao, Tianyun Yang, Junbo Guo, and Jintao Li. 2019. Exploiting multi-domain visual information for fake news detection. In *2019 IEEE international conference on data mining (ICDM)*. IEEE, 518–527.

[23] Rimsha Rafique, Rahma Gantassi, Rashid Amin, Jaroslav Frnda, Aida Mustapha, and Asma Alshehri. 2023. Deep fake detection and classification using error-level analysis and deep learning. *Scientific Reports* 13 (05 2023). https://doi.org/10.1038/s41598-023-34629-3

[24] Md Mehedi Rahman, Jannatul Tajrin, Abul Hasnat, Naushad Uzzaman, and GM Atiqur Rahaman. 2019. SMIFD: novel social media image forgery detection database. In *2019 22nd International Conference on Computer and Information Technology (ICCIT)*. IEEE, 1–6.

[25] Victor Schetinger, Manuel M Oliveira, Roberto da Silva, and Tiago J Carvalho. 2017. Humans are easily fooled by digital images. *Computers & Graphics* 68 (2017), 142–151.

[26] ScienceDirect. Year. *Image Histogram*. https://www.sciencedirect.com/topics/engineering/image-histogram Accessed: Date.
[27] Deepika Sharma and Pawanesh Abrol. 2013. Digital Image Tampering – A Threat to Security Management. https://api.semanticscholar.org/CorpusID:212606420
[28] Allyson Haynes Stuart. 2018. Social media, manipulation, and violence. *SCJ Int'l L. & Bus.* 15 (2018), 100.
[29] Marlis Stubenvoll, Raffael Heiss, and Jörg Matthes. 2021. Media trust under threat: Antecedents and consequences of misinformation perceptions on social media. *International Journal of Communication* 15 (2021), 22.
[30] Marlis Stubenvoll, Raffael Heiss, and Jörg Matthes. 2021. Media trust under threat: Antecedents and consequences of misinformation perceptions on social media. *International Journal of Communication* 15 (2021), 22.
[31] Shobhit Tyagi and Divakar Yadav. 2023. A detailed analysis of image and video forgery detection techniques. *The Visual Computer* 39, 3 (2023), 813–833.
[32] David Westerman, Patric R. Spence, and Brandon Van Der Heide. 2014. Social Media as Information Source: Recency of Updates and Credibility of Information*. *Journal of Computer-Mediated Communication* 19, 2 (01 2014), 171–183. https://doi.org/10.1111/jcc4.12041 arXiv:https://academic.oup.com/jcmc/article-pdf/19/2/171/19492171/jjcmcom0171.pdf
[33] Nancy E Willard. 2010. Sexting and youth: Achieving a rational response. *Journal of Social Sciences* 6, 4 (2010), 542–562.
[34] Yunkang Yang, Matthew hindman, and Trevor Davis. 2022. Facebook image data. https://doi.org/10.7910/DVN/RNITKF
[35] Li Yuanman and Jiantao Zhou. 2016. Image copy-move forgery detection using hierarchical feature point matching. 1–4. https://doi.org/10.1109/APSIPA.2016.7820758
[36] Weiguo Zhang, Chenggang Zhao, and Yuxing Li. 2020. A novel counterfeit feature extraction technique for exposing face-swap images based on deep learning and error level analysis. *Entropy* 22, 2 (2020), 249.
[37] Lilei Zheng, Ying Zhang, and Vrizlynn L.L. Thing. 2019. A survey on image tampering and its detection in real-world photos. *Journal of Visual Communication and Image Representation* 58 (2019), 380–399. https://doi.org/10.1016/j.jvcir.2018.12.022