International Journal of Engineering and Advanced Technology (IJEAT)
ISSN: 2249-8958 (Online), Volume-12 Issue-5, June 2023

Document Forgery Detection

Nandini N, Keerthi Joshi K, Devprakash, Madhura C, Vandana M Ladwani

Abstract: ***Document forgery is an increasing problem for both private companies and public administrations. It can be said to represent the loss of time and resources. There are many classical solutions to these problems, such as detecting an integrated security pattern. In such cases, we must resort to forensic techniques for detection. The idea behind using these forensic techniques can also be implemented using artificial intelligence or machine learning, which can be a lower-cost option and provide the same or better results. The experimental results show that multiple models have a strong detection capability for detecting numerous forgeries. In this paper, we present a novel approach to detecting forgeries in documents. The forgery we detect can be classified as hand-written signature forgery and copy-move forgery of any photo, text, or signature. We have developed a novel approach using capsule layers to detect forgery in handwritten signatures. We also use ELA (Error Level Analysis) to detect any errors in the image compression levels.***

Keywords: Document, Forgery Detection, Capsule Neural Networks, CNN, Copy-Move Forgery, Signature Forgery

I. INTRODUCTION

We now live in a digital age where technology has become a key aspect in our day-to-day lives, from creating, processing, and storing information to displaying user needs and desires. Knowledge representation is multidimensional and stored as bits and bytes in various formats, including text, images, and videos. Technology can be an issue. With recent advances in technology, people are misusing technology in ways that can harm society. Changing data without visible traces of intervention has never been easier. With this, there are several opportunities as well as risks, and the gravity of the threats cannot be understated.

Manuscript received on 10 May 2023 | Revised Manuscript received on 22 May 2023 | Manuscript Accepted on 15 June 2023 | Manuscript published on 30 June 2023.

*Correspondence Author(s)
Nandini N*, Department of Computer Science and Engineering, PES University, Bangalore (Karnataka), India. E-mail: nandinins236@gmail.com, ORCID ID: https://orcid.org/0009-0002-5371-4667
Madhura C, Department of Computer Science and Engineering, PES University, Bangalore (Karnataka), India. E-mail: cmadhura122001@gmail.com, ORCID ID: https://orcid.org/0009-0009-4394-112X
Keerthi Joshi K, Department of Computer Science and Engineering, PES University, Bangalore (Karnataka), India. E-mail: keerthi.joshi2822@gmail.com, ORCID ID: https://orcid.org/0009-0008-8709-9539
Devprakash B, Department of Computer Science and Engineering, PES University, Bangalore (Karnataka), India. E-mail: devprakashbesoi@gmail.com, ORCID ID: https://orcid.org/0009-0002-6176-578X
Vandana M Ladwani, Department of Computer Science and Engineering, PES University, Bangalore (Karnataka), India. E-mail: vandanamd@pes.edu, ORCID ID: https://orcid.org/0000-0002-4099-6936

Impersonating or altering artefacts, such as significant papers, pictures, news articles, works of art, etc., is called forgery. It is always coupled with other fraudulent activities, including check fraud, insurance fraud, identity theft, and smuggling. False social media profiles and modified email messages are examples of technological forgeries that are not always tangible. We have observed a significant increase in demand for online document authentication in e-commerce and e-government applications, mainly due to the ongoing issues caused by the pandemic. Documents are placed on internet platforms for various reasons. However, specific editing software or other technology may alter the content of the document. It is essential to document the changes made to photocopies. Images can serve as small pieces of forensic evidence or can be used in court. Proving the authenticity of documents is very important. Forgery is the process of imitating or altering objects, including important documents, images, news, and works of art. It is often accompanied by other fraudulent behaviours, such as identity theft, smuggling, insurance fraud, check fraud, and many more. Forgeries do not always have to be physical; they can also be electronic, such as fake social media pages or the adaptation of email correspondence. Since the concern of our topic is documented forgery, we will focus on different kinds of document forgery. Genuine passports without any personal data or stamps, or signed memorandum letters without content, can be considered genuine documents as they are blank. The forger inserts the data to perform the fraudulent activity. The forgeries in these kinds of documents are challenging to detect. The detection of any manipulation in document images is crucial, as an image can serve as legal evidence in various forensic investigations and other fields. Hence, there is a dire need to prove the authenticity of a document.

II. RELATED WORK

Digital forgery detection has been an ongoing topic in the research field for many years. Numerous solutions for this have been proposed, addressing different kinds of forgery detection. The existing document forgery detection methods can be broadly classified into two main categories: active and passive methods. The paper [1] proposes a system architecture based on the inspection of probed documents with the analysis of ink. The paper presents a new method for identifying any mismatch in ink colour in HSD images. The approach is based on an NMF model with orthogonal and graph regularisation. The assumption made here is that under some attainment protocols, some of the latent actors present in the HSD images can be forced to be orthogonal. The author of this paper has also proposed an efficient algorithm, multiplier-based, to incorporate into the method.

The paper [2] proposes an efficient method to detect signature forgery which is based on the siamese neural network. The technique utilises a CNN for data preprocessing, and evaluation is performed using a Siamese network that works in conjunction with the CNN model. Unique features of the implementation include a contrastive loss function. A high recall was achieved, and the loss was minimised to 0.43. The paper [3] proposes a robust system to detect digital forgery using CNN architecture for compressed images. The CNN architecture consists of multiple layers, including a pooling layer, a convolutional layer, and fully connected layers. The authors in [4] offer a shallow convolutional neural network(CNN) that can identify manufactured region boundaries from original edges in low-resolution images. SCNN was created to utilise chroma and saturation data. Two techniques based on SCNN, term sliding windows detection (TSWD) and fast SCNN, have been developed to detect and identify image forgery regions. The paper[5] provides a new deep learning-based image fraud detection system for automatically learning hierarchical representation from input RGB color photographs. Image splicing and copy-move forgeries can be detected using the suggested CNN, The fundamental high-pass filter set employed in the spatial rich model(SRM) is utilized to establish the weights at the first layer of our network, which serves as a regularizer to efficiently suppress the effect of picture contents and capture the subtle artifacts created by the tampering operations. The pre-trained CNN is utilised as a patch descriptor to extract dense features from the test images, and the final discriminative features for SVM classification are obtained through a feature fusion technique.

III. DATASET

Finding an appropriate dataset for training and testing the model is one of the challenging tasks while creating an ML model. Since the training of the ML model largely depends on the dataset used. Forged document image datasets are not readily available. We planned to create the dataset by collecting authentic document images from the web and then forging them to make a forged image dataset. The dataset consists of two types of data: one for training the copy-move forgery detection model and the other for training the signature forgery detection model. Since the dataset is created manually, it requires preprocessing. The data preprocessing is done using the OpenCV technique. Since the noise in the image has a significant effect on the model's result. Preprocessing includes slant correction and denoising. Since we have approached the forgery detection problem by creating different models for various types of intrinsic features, the extraction of these features is also necessary. The signature forgery detection model accepts the extracted signature from the uploaded image.

IV. PROPOSED WORK

The implemented system consists of preprocessing, capsule network, error level analysis, and CNN.

a) Data preprocessing: Preprocessing is performed to reduce noise in the uploaded image, which may result in false positives or reduced efficiency of the detection model. The preprocessing phase plays a vital role in the product implementation since the errors or unwanted data present in this phase will be carried out till the final stages of the product development. Cleaning the data in the earlier stages will improve the system. The pre-processing stage involves slant correction, cropping the unwanted region from the input document image, and image denoising. The slant correction and denoising are done using the OpenCV technique.

b) Capsule Neural Network: To detect signature forgery, a capsule neural network is used instead of a CNN. Despite the benefits of CNN, there are always challenges in using it. CNNs cannot model changes in new fields and dimensions, and cannot distinguish between similar components placed in different locations of an image. The reason to use CapsNet is that it can reduce the number of layers and parameters in the network architecture, thereby decreasing complexity. CapsNet is capable of detecting the details of angular and spatial changes in components. Not only can a capsule network represent the existence of particular visual features, but it can also detect the transformations that might have occurred in these features. In capsule networks, spatial details are completely differentiated.

c) Error level analysis and CNN: To convert an image to ELA, the preprocessed images must be reframed at a specific quality level. The image is whitened or brightened as a result of this technique. To reframe the images, both forged and authentic images are considered, which have been processed during the preprocessing stage. Finally, the preprocessed image and the reframed image are compared to see the difference between them. The tampered parts of the image in the forged ELA framed image are brighter than the corresponding original segments of the image. Having ELA analysis before processing the neural networks has a significant advantage, as the ELA-reframed image contains only non-redundant information, and nearby pixels are compared based on similar intensity. The reframed image needs to be resized to ensure the RGB values lie between 0 and 1, allowing each cell value to be normalised by dividing it by 255. The resizing helps the neural network converge faster than usual. The CNN model consists of 2 layers of a convolutional network. The first layer of the CNN is a convolutional layer with 32 filters. Dropout is used as a regularisation technique to reduce overfitting in neural networks, thereby preventing complex co-adaptations on the training data. The CNN consists of 2 convolution layers, max pooling, and two dropout layers

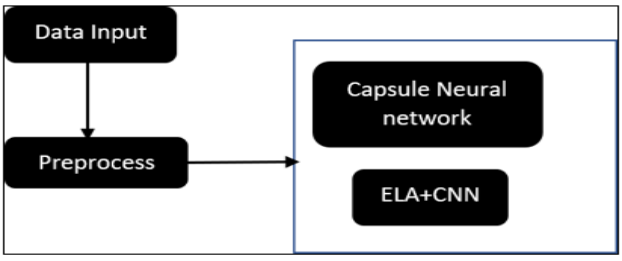

Figure 1: System architecture

The result of the forgery detection is produced considering the result of individual models. The particular models are enclosed within an ensemble model. Since the fakeness of each feature is detected using different models, the results from the respective models need to be combined to obtain the overall fakeness result. All the models are combined into an ensemble model, and the overall fakeness of the document is produced as the output.

V. RESULTS AND DISCUSSIONS

The training images for the copy-move model, including the CASIA-2 and MICC datasets, have been considered. The dataset consists of images from these two datasets, which are split into training and testing datasets and then passed onto our model to classify them into two classes: authentic and forged.

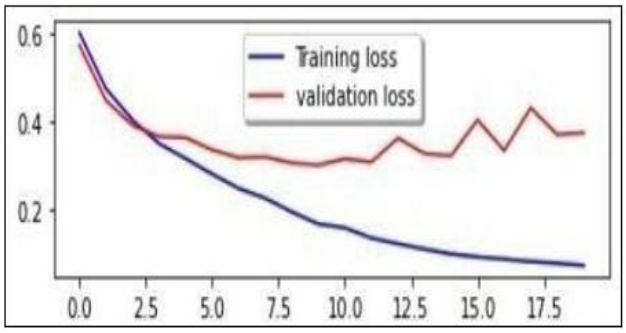

Fig 2: Training loss and Validation Loss of Copy Move Forgery Detection Model

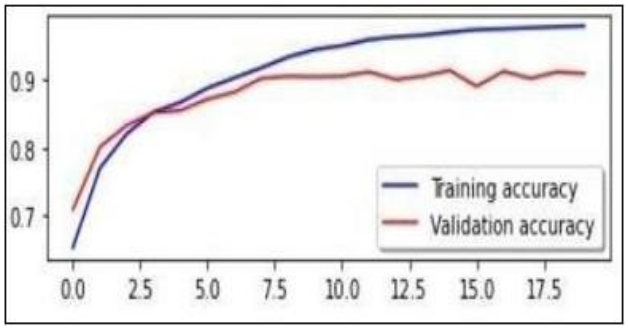

Fig 3: Training Accuracy and Validation Accuracy of Copy Move Forgery Detection Model

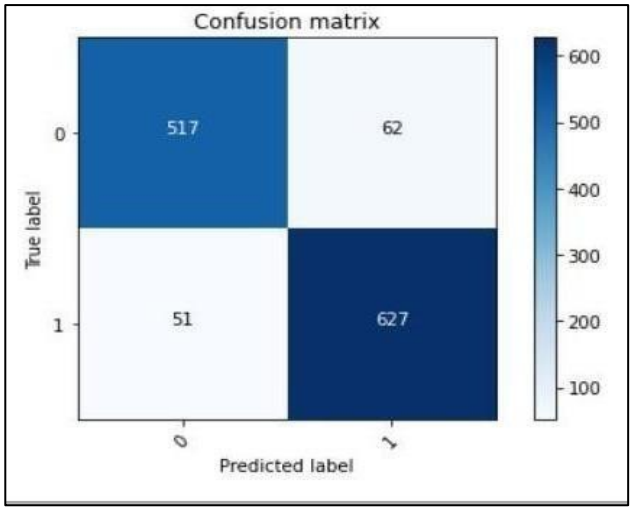

Figure 4: Confusion Matrix of Copy Move Forgery Detection Model

VI. CONCLUSION

This paper presents a method for detecting digital document forgery using a neural network. The proposed solution utilises a capsule neural network for signature forgery detection, and a combined model of ELA and CNN is employed for copy-paste forgery detection. The proposed method uses features from two datasets of varying difficulty. The experiment results validate that the classification performance decreases when the samples are more challenging. However, the implemented architecture does not easily generalize to datasets with different underlying distributions. The capsule neural network was found to be efficient in detecting forged signatures in document images. ELP was found to be efficient in detecting copy-move forgery. The overall positive and negative forgery detection rate of the designed system was very promising. It is recommended that future work investigate a broader range of algorithms with even higher efficiency, and also detect the fraudulent use of video [MP4] format files. Overall, the project work has provided us with an opportunity to delve deeper into forgery detection methods. Image forgery detection is not only an emerging research topic but also an important area that requires faster and more accurate work. The correct and efficient methods to detect forgery are increasing day by day. Nevertheless, there is undoubtedly a lot of work still to be done in the domain of image forgery detection, and neural networks will be able to detect tampered images regardless of their complexity. In the future, there is scope to improve the training model by increasing the variance in the dataset, and it also has scope to detect forgery in video files.

DECLARATION

Funding/ Grants/ Financial Support	No, I did not receive.
Conflicts of Interest/ Competing Interests	No conflicts of interest to the best of our knowledge.
Ethical Approval and Consent to Participate	No, the article does not require ethical approval or consent to participate, as it presents evidence that is not subject to interpretation.
Availability of Data and Material/ Data Access Statement	Not relevant.
Authors Contributions	All authors have equal participation in this article.

.REFERENCES

1. Rah. iche, Abderrahmane; Cheriet, Mohamed (2020). [IEEE 2020 IEEE/Cvf Conference On Computer Vision And Pattern Recognition Workshops (Cvprw) - Seattle, WA, USA (2020.6.14-2020.6.19)] 2020 IEEE/Cvf Conference On Computer Vision And Pattern Recognition Workshops(Cvprw) - Forgery Detection in Hyperspectral Document Images Using Graph Orthogonal Nonnegative Matrix Factorisation.,(),2823-2831.Doi:10.1109/Cvprw50498.2020.00339 [CrossRef]
2. S. Jain, M. Khanna, and A. Singh, "Comparison among different CNN architectures for signature forgery detection using siamese neural network"; 2021 International

Conference on Computing, Communication, and Intelligent Systems (ICCCIS), 2021, pp. 481-486, DOI:10.1109/ICCCIS51004.2021.9397114. [CrossRef]
3. Robust forgery detection for compressed images using CNN supervision Boubacar Diallo *, Thierry Urruty, Pascal Bourdon, Christine Fernandez-Maloigne Université de Poitiers, CNRS, XLIM, UMR 7252, f-86000 poitiers, France, Volume 2, December 2020, 100112. [CrossRef]
4. IJCSNS International Journal of Computer Science and Network Security, vol. 20, no.12 December 2020: https://doi.org/10.22937/ijcsns.2020.20.12.12 [CrossRef]
5. Zhang, Zhongping & Zhang, Yixuan & Zhou, Zheng & Luo, Boundary-based image forgery detection by fast shallow CNN, conference: Computer Vision and Pattern Recognition, Doi: 10.11.09/ICPR (2009).

AUTHOR PROFILE

Ms. Nandini N is a passionate researcher and writer with an interest in machine learning and artificial intelligence. Ms Nandini is currently pursuing her Bachelor's degree in Technology with a specialisation in Computer Science and Engineering at PES University. Ms Nandini has a deep understanding of image processing and machine learning concepts through extensive research and academic work. Throughout her educational journey, Ms Nandini has consistently worked on machine learning projects to enhance her knowledge of the machine learning domain. Ms Nandini also knows image processing through coursework and has worked on a team project in the same domain. In conclusion, Ms Nandini is a passionate and dedicated researcher who works beyond domain boundaries, well-equipped to tackle challenges and contribute to the research domain.

Ms. Keerthi Joshi K is a dedicated researcher and writer with a passion for Data Science and Machine Learning. Currently pursuing a B.Tech in Computer Science Engineering at PES University, Bangalore, Ms Joshi has developed a deep understanding of Data science and Machine Learning through extensive research and coursework. Throughout her academic journey, Ms Joshi has consistently demonstrated exceptional analytical skills and a keen eye for detail. She and her team have conducted thorough investigations, analysing complex datasets and examining diverse perspectives to arrive at well-rounded conclusions. Their commitment to academic excellence is evident in their consistently high grades and their active participation in class discussions and research projects. In conclusion, Ms. Joshi is an accomplished and driven researcher, dedicated to expanding the boundaries of knowledge in both Data Science and Machine Learning. With their academic achievements, research experience, and commitment to excellence, she is well-equipped to tackle the challenges and plans to make valuable contributions to the scholarly community.

Devprakash is a dedicated and ambitious student at PES University, Bangalore, currently pursuing their academic journey with a focus on data analytics and machine learning. As a student by status and a continuous learner by mindset, Devprakash is committed to acquiring knowledge and honing their skills in these rapidly evolving fields. With a deep passion for data analytics and machine learning, Devprakash enthusiastically explores the vast potential of these subjects. They actively engage in coursework, projects, and extracurricular activities that contribute to their understanding and expertise in the field. Devprakash's favourite subjects, data analytics and machine learning, serve as a driving force behind their academic pursuits. Devprakash's proactive nature has led them to actively participate in hackathons, where they have consistently demonstrated their ability to solve complex problems and create innovative solutions. By collaborating with diverse teams and leveraging their analytical skills, Devprakash has successfully showcased their talent, earning recognition and respect from peers and mentors. Looking ahead, Devprakash envisions utilizing their skills in data analytics and machine learning to make a significant impact. They have a clear vision of leveraging these technologies to drive positive change in various domains such as healthcare, finance, and social issues. Devprakash firmly believes that the power of data, when combined with advanced analytical techniques, can revolutionize industries and improve lives on a global scale.

Madhura C is a bright and driven student currently pursuing her B.Tech degree from PES University. She is currently working as a student intern at Textron Company, where she is showcasing her impressive problem-solving skills and ability to make strong decisions. Her exceptional technical expertise in areas such as Python, Data Science, Cybersecurity, and machine learning is highly valued by her peers and colleagues. Madhura's passion for data science is evident in her impressive projects and academic achievements. Her work showcases her exceptional ability to utilise her technical skills to solve complex problems while maintaining a broader perspective. Madhura's decision-making skills and attention to detail are valuable assets that have enabled her to excel in both her academic and professional careers. In conclusion, Madhura is an exceptional student with strong technical skills and a passion for data science and machine learning. Her dedication, problem-solving skills, and ability to make sound decisions make her an asset to any team. We are confident that she will continue to excel in her academic and professional career, and we look forward to seeing her make a significant impact in her field.

Vandana M Ladwani is an assistant professor with 14 years of teaching experience. She holds a Bachelor of Technology and a Master of Technology degree in Computer Science from RTM Nagpur University. She works in the areas of signal processing and machine learning. She has more than 15 publications in international conferences and journals. She had also contributed to book chapters.