First International Conference on Computer Engineering
International Journal of Scientific Research in Science and Technology
Print ISSN: 2395-6011 | Online ISSN: 2395-602X (www.ijsrst.com)
Volume 5 Issue 8, November-December-2020

Fake Image and Document Detection using Machine Learning

Amit Lokre[1], Sangram Thorat[1], Pranali Patil[1], Chetan Gadekar[1], Yogesh Mali[1]
[1]Department of Computer Engineering , Dr D Y Patil School of Engineering, SPPU Pune, Maharashtra, India.

ABSTRACT

In the recent times, the rates of cyber crimes has been increasing tremendously. It has been proven incredibly easy to create fake documents with powerful photo editing softwares. Also social media has proven to be the largest producer of fake images as well. Various malpractices have also been on surge with the help of producing digitally manipulated fake documents. Detection of such fake documents has become mandatory and essential for unveiling of the documents/images based cyber crimes. The tampered images and documents will be detected using neural network .The output of the system will distinguish original document from a digitally morphed document. The system will be implemented using Neural Networks.

Keywords : Artificial Neural Network ; GLMC Features; Graphical User Interface ;Machine Learning ; Support Vector Machine.

I. INTRODUCTION

In the recent times the speed of cyber-crimes has been increasing tremendously it's been proven incredibly easy to make fake documents with powerful photo editing software Also social media has proven to be the most important producer of faux images also Various malpractices have also been on surge with the assistance of manufacturing digitally manipulated fake documents Detection of such fake documents has become mandatory and essential for unveiling of the documents/images based cyber-crimes The tampered images and documents are going to be detected using neural network The output of the system will distinguish original document from a digitally morphed document The system are going to be implemented using Neural Networks this is often an desktop application the rates of cyber-crimes are on a rise it's been proven incredibly easy to make fake documents with powerful photo editing software Documents and pictures are often scanned and morphed within minutes with the assistance of sort of software available On Investigation it States a foundation and it provides an answer to differentiate between original document and digitally morphed document Here the accuracy of system method has accuracy of 96 It is also possible to change metadata content making it unreliable here it's used as a supporting parameter for error level analysis decision . In, Xunyu Pan, Siwei Lyu proposed a scheme to detect the copy-move forgery in a picture , mainly by extracting the key points for extraction. The difference between the normal method and proposed scheme is first segments the test into semantically independent patches before key point extraction. within the second stage, to refine an estimated matrix an EM-based algorithm is employed and to verify the existence of copy-move forgery. The methods are categorized in two types as active

and passive forgery detection methods. The scope of system is restricted to review on passive forgery detection methods. System aims to present the study on different old methods of image forgery detection using different approaches like DWT (Discrete Wavelet Transform), SIFT, LBP (Local Binary Pattern). With the advancements in digital image acquisition, processing and reproduction technologies, the perfection has become easier with the fabrication of document. Document examiners examines the Composite copies which are produced for fraudulent purposes. The easy availability of those technologies to criminals leads to their application to preparation of fabricated photocopies or computer generated hard copies. Such document are produced in court of law because the only available piece of evidence with a excuse the the first document has been lost , eaten by moths or burnt during a fire etc. Superimposition is employed to detect counterfied documents, mainly tampered with employing a photocopier . This study presents examination of machine generated questioned document consisting of fabricated and manipulated writing signatures using digital image processing and reproduction tools. However, such techniques have now become obsolete since forgery lately is digital, clean and indistinguishable to the human eye. Therefore, machines are a more viable option now. Most of the techniques wont to detect those manipulations employ machine learning and pattern recognition. Image processing algorithms like DWT (Discrete Wavelet Transform) and SVD (Singular Value Decomposition) are one among the feature extraction methods that are used today to detect forged images Another devised scheme is to divide the image into overlapping blocks, thinking of them as vectors and find the manipulated region through radix sorting. Here method for detecting copy-move forgery over images tampered by copy-move. To detect such sort of forgeries, the input image is split into overlapping blocks of equal size, feature for every block is then extracted and represented as a vector, radix sort is employed to extract all the features of vector. Finally, the medium filtering and connected component analysis are performed on the tentative detected result to get the ultimate result. Compared with other methods, employing the radix sort makes the detection far more efficient without degradation of detection quality. Another approach to detect tampered images is to form use block based methods, but by using the non-overlapping texture blocks as a base for the graceful blocks, thus reducing the computational capacity. The algorithms then evaluates and compares supported their performances associated with a group of predefined parameters, this characterization are going to be used for further evaluation on the performance and efficient of an given blocked based cloning detection algorithms under the study. The result found after comparing them a user ready to select the foremost optimal forgery detection technique, counting on the user format and sort transformation it involves . Reflective SIFT based algorithms also are proficient in the detecting duplicated blocks in copy-move forgeries. due to this different detection techniques are suggests . we'd like to require care of image forgery. the pictures are often scaled, rotated, and flipped (mirror reflected). SIFT (Scale invariant feature transform) is usually wont to match images. But it fails for flipped images. Mirror-reflection invariant feature transform (MIFT) is presented in framework. Here, we'll find out how MIFT improvise SIFT

Literature survey:

Morphing images digitally has experienced tremendous growth in past 10 years Now a days several software are available that are wont to manipulate image in order that the image is appear as if as original Images are used as proof for authentication for any crime and if these images

aren't genuine then it'll be a problem for acceptance System is employed to detect these sorts of forgeries to work out whether a digital image is original or manipulated may be a big challenge to seek out the traces of tampering during a digital image may be a challenging task this technique presents a number of the Image Manipulation detection techniques like contrast enhancement detection splicing and composition detection image tampering etc Comparison of those techniques concludes the higher approach for its future research Using neural network tampered images are detected which recognizes the morphed region of the image and unveils the segments of the first image It are often implemented on Android platform The compression ratio of the changed content during a fake image is different and is detected using Error Level Analysis(ELA) that of the first image . Results prove the great performance of the scheme by comparing it with the state-of-the-art schemes on the general public database. the most two types are (a) Copy-Move forgery (b) Image splicing forgery of image Copy-move tampering is most generally employed by attackers during which object of another image is copied and pasted in original image in nearly matching areas. Hence to detect such image threats, it's required to automatic computer vision based method which may classify whether input digital image is original or tampered. there have been many methods introduced for copymove forgery detection from last 15 years.

II. METHODS AND MATERIAL:

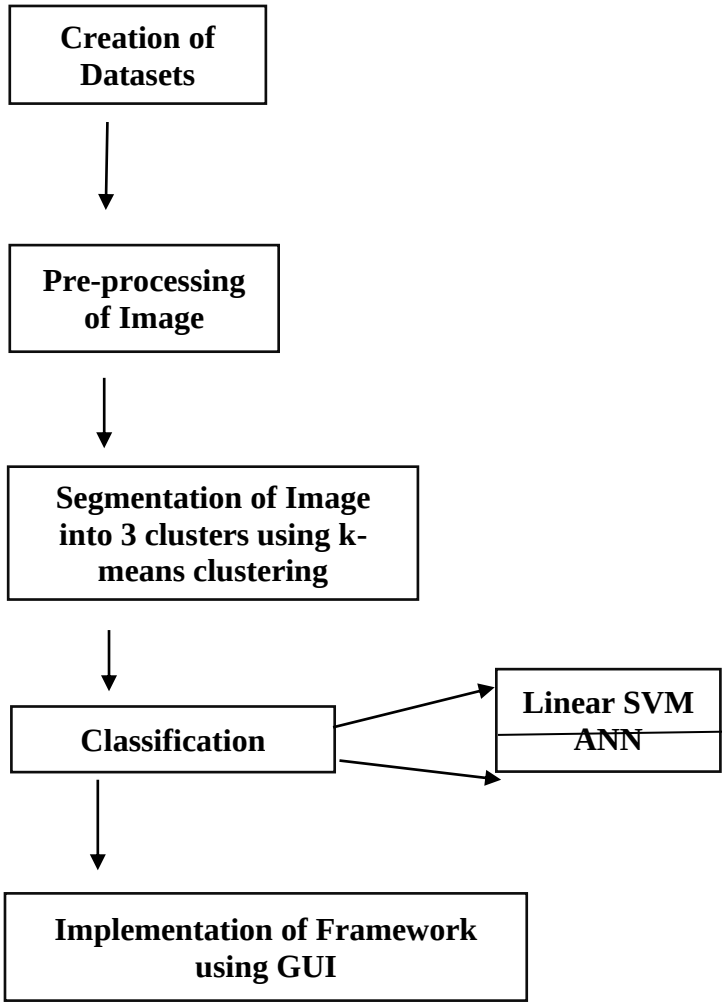

Fig 1.flowchart of methodology

1.Creation of dataset:

The images used for training this system are collected from various internet sources, college dataset and are morphed using photo editing tools. These images are edited using Adobe Photoshop CC 2017 to create a dataset of images- one original and its edited version.

2.Pre-Processing of Image:

To make the details of the images stand out more, the query image will be enhanced using histogram equalization .Histogram equalization: It is a necessary step because sometimes minute forgeries go undetected through the entire process. It is important

that the machine gets most of the details in one go. Histogram equalization, as the name suggests, is a method, where the intensities are adjusted using the histogram of the image. This technique is used here for contrast enhancement .Another essential stage in the pre-processing of an image is the removal of noise i.e. de-noising.

Denoise: De-noising is again done so that the details of the image are sharper and are not missed while extracting the features of the image.

Median Filter: A kernel of pixels is scanned over pixel matrix of the entire image. The median of a pixel values in the window is computed, and the center pixel of the window is replaced with the computed median. Since the median value must actually be in the value of one of the pixels in the neighborhood , the median filter does not create new unrealistic pixel values when the filter straddles an edge .For this reason median filter is much better at preserving sharp edges.

3.Segmentation:

Segmentation will be performed using k-means clustering, In this the image will be divided into K segments and the appropriate image will be chosen upon the data contained it .For this the GLMC feature will be used and the frame or segment having highest of the mean will be chosen. The GLMC frames will be compared to the original image and it will be helpful in calcutating the result as this value will be stored into some array and it will taken into count at the time of results.

4.Extraction of Features:

Feature extraction is a primitive type of pattern recognition and it is very important for pattern recognition.In this step we extracts some GLCM features such as Entropy, Angular Second Mo- ment, Contrast, Maximum Absolute Deviation, and Mean.Texture, shape and colour are features which are also extracted from an input image or document.Gray Level Co-occurrence Matrix is a tabulation of how often different combinations of pixel brightness values occur in a image. GLCM contains the information about the positions of pixel having similar gray level values.GLCM calculation units receive pairs of gray level values as input. The GLCM calculation unit consists the different combinations of gray values like a0b1, a2b3, a10b21 etc. This gives the deviation present in the image when compared with origi- nal image by predictive image.

5.Classification:

The interconnection of the network can be adjusted based on the number of available inputs and outputs making it ideal for a supervised learning.A linear kernel SVM will be one of the classifier used for classification and the other one will is ANN.

III. RESULTS AND DISCUSSION

This is the approach discussed in this paper executed successfully. The main purpose of the project was to inspect the handled computer legal documents for forensic testing using images. The processing technique that has been applied, the dataset required for the project was created by handling purchased images that no naked human eye could recognize it to handle image skill fully. After that, the feature information implemented by GALCM in MATLB, also provided by image processing toolbox R2015a examination and comparison required results between the original and the morphed legal document. Also HOG values were calculated but from the sort of matrix produced was very large trained in SVM so that they were eliminated.

TABLE I: Classifiers used and their results

PARAMETERS	ANN	SVM
Accuracy	96.4%	87.6%
Sensitivity	97.0%	89.0%
Specificity	95.0%	86.0%

Use of SVM, classification of differences between one between the above morphed and the original legal document the result would have been different meanwhile more property specific assets of two documents. Also, by other mathematics, the expression would have been more sophisticated consequences of splitting morphed documents more easily. The use of an ANN classifier provided high accuracy 96.4% less than linear SVM which gave less accuracy.

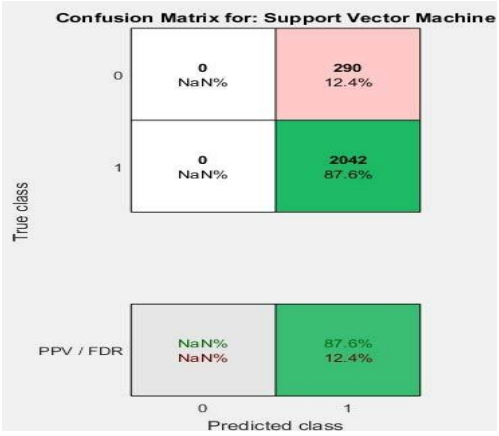

Fig 2: Confusion plot by Linear SVM

Fig 3: Confusion plot by ANN

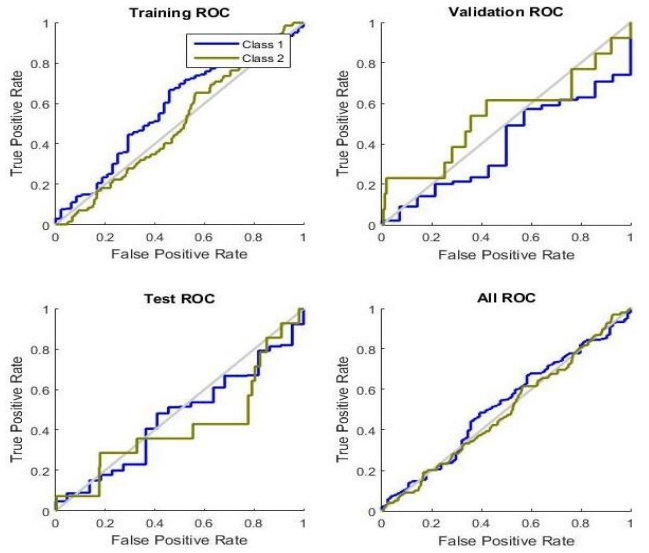

Fig 4.ROC plot by ANN

IV. CONCLUSION

Starting by editing images to crate a dataset and then its morphing and then scientifically examining the results obtained. Use the transformation technique above morphed and original image and finally edit the texture, features are the same initially, it was planned. At the present time, advances in science and technology, various advance introduction of image editing tools are also on the rise. These tools has multipurpose features. We can use this advanced image editing tools for the expansion of our next project execution to make results more quick and easy. While these tools are mostly used in areas related to creative design , criminals also can easily get access to it and as a result they can create a fake identity to hide themselves in a public place or the crime can be under investigation. These fake documents are not visible for the human eye. So, that was our intention purpose, such a method with good efficiency and accuracy will continue to refine procedures so that there are fewer loops in the analysis and hopefully this will come handy in the future.

V. REFERENCES

[1]. Shruti Ranjan, Prayati Garhwal, Anupama Bhan, Monika Arora, Anu Mehra,"Framework For Image Forgery Detection And Classification Using Machine Learning".2nd International Conference on Trends in. . . 2018. DOI: 10.1109/icoei.2018.8553924

[2]. Mohsen Zandi, Ahmad Mahmoudi- Aznaveh, Alireza Talebpour, "Iterative Copy-Move Forgery Detection Based on a New Interest Point Detector", Information Forensics and Security IEEE Transactions on, vol. 11, pp. 2499-2512, 2016, ISSN 15566013

[3]. Kushol, Rafsanjany Salekin, Md Sirajus Hasanul Kabir, Md Alam Khan, Ashraful. (2016). "Copy-Move Forgery Detection Using Colour Space and Moment InvariantsBased Features" .2016 International Conference on Digital Image Computing: Techniques and Applications (DICTA 1-6.10.1109/DICTA.2016.7797027

[4]. Xunyu Pan, Siwei Lyu, "Region Duplication Detection Using Image Feature Matching", Information Forensics and Security IEEE Transactions on, vol. 5, pp. 857-867, 2010, ISSN 1556-6013.

[5]. Anil Dada Warbhe, Rajiv V. A Fast, Block Based, Copy-Move Forgery Detection Approach Using Image Gradient and Modified K Means"

[6]. Dharaskar, Vilas M. Thakare, "Digital image forensics: An affine transform robust copy-paste tampering detection", Intelligent Systems and Control (ISCO) 2016 10th International Conference on, pp. 1-5, 2016

[7]. Badal Soni, Pradip K. Das, Dalton Meitei Thounaojam. "(2018) CMFD: a detailed review of block based and key feature based techniques in image copy-move forgery detection". Institution of Engineering and Technology(IET) Image Processing 12:2, pages 167-178

[8]. In, Hwei-Jen Wang, Chun-Wei Kao, Yang-Ta. (2009). "Fast copy-move forgery detection".Article in WSEAS Transactions on Signal Processing".

[9]. Joshi MC, Kumar A, Thakur S. "Examination of digitally manipulated-machine generated document", a case study elucidating the issue of such unwanted progenies of modern technology". Prob Forensic Science 2011;56:162–73. Navoneel Chakrabarty, Sanket Biswas "A Statistical Approach to Adult Census Income Level