A Machine Learning Based Approach for Image Quality Assessment of Forged Document Images

Gayatri Patil[1,1](✉), Shivanand S. Gornale[1], and Ashvini Babaleshwar[2]

[1] Department of Computer Science, Rani Channamma University, Belagavi, India
gayatripatil865@gmail.com
[2] Department of Computer Science, Garden City University, Bangalore, India

Abstract. Document Images, such as typed and handwritten documents can be manipulated in various ways using many sophisticated digital technologies and photo editing software's. As a result, one can alter the text in the typed and handwritten documents that leads to degradation of quality of an image. The detection of multiple inherently altering operations in an image is a challenging issue, hence in this work a novel approach is proposed for the ten-class problem in which the alteration of a text can be accomplished through multiple operations, which all create the specific pattern. These operations are analysed with the help of image quality measures and classified using random forests classifier. The proposed approach gives a better classification accuracy rate of 94% for forged printed document images and 98.80% of forged handwritten document images, which is more promising and competitive with state of the art techniques reported in the literature.

Keywords: Document Forgery · Image Quality Measures · Multiple forgery operations · Random Forest tree · Ten Class Classification

1 Introduction

In today's digital environment, the use of printed and handwritten document images in daily human activities is increasing. Manipulation of these document images is also increasing, with many sophisticated digital technologies and photo editing software's being used. As a result, the text in typed and handwritten documents can be changed. In the field of forensic science, altering text document images leads to forging and is considered a crime application [1]. For instance, a property agreement where the contents can be modified to make an illegal trade, or a plane ticket where the date may be changed to gain access to airport terminals by circumventing security. Handwritten documents are also used to produce false suicide letter, answer scripts, and certifications, among other things [2].

In computer vision and image processing, detecting forged videos and images is not a new issue; however, it is not a new problem in research. There are several methods available in literature [3, 4]. However, fraud recognition in document images including printed and handwritten document images is new as compared to video and images. This work receives special attention of the researchers [5]. This is because the document

R. Manza et al. (Eds.): ACVAIT 2022, AISR 176, pp. 208–229, 2023.
https://doi.org/10.2991/978-94-6463-196-8_18

images are often used as an authenticated proof evidence for any crime investigation in court rooms. In addition, we the common people believe that the content in newspaper and internet are genuine and authenticated [6]. If the content in these documents is altered, it leads to misinformation and spreading wrong message to society. Therefore, it is necessary to verify authenticity and integrity of the documents automatically without human intervention.

To create forgery or fake documents and tampering original content, usually they use two operations, such as copy-paste and insertion [2]. In case of copy-paste operation they copy from the same document or different document to paste at target words while in case of insertion; people use software tools to edit the words by adding characters at appropriate places. If the document contains forged word with simple operation, there are methods to find solution in the literature [7, 8]. In reality, sometimes document suffers from degradations due to noise, document aging, paper quality, use of ink in case of handwriting etc. When the document contains forged words along with the words affected by the above degradations, the methods do not perform well and fail to detect the forged words [9, 10]. For instance, sample image for printed and handwritten document are shown in Fig. 1 and Fig. 2, where one can see different type of forged words in single document. This challenge remains an open issue for forgery detection.

To address these challenges, we create forged words with multiple operations. For example, for forged word created by copy-paste operation, we add noise to the same word, which is called Copy-Paste + Noise class. In the same way, we create Copy-Paste + Blur, Insertion + Noise, Insertion + Blur, Copy-Paste + Insertion along with copy-paste alone, insertion alone, noise and blur alone. This results in 10 classes of forged typed words and the sample images for each class are shown in Fig. 3 and Fig. 4, respectively for printed and handwriting documents.

2 Related Work

There are several methods for forgery detection in document images. The method can be classified broadly into two categories, namely, the methods which focus on printed document images and the method which focus on handwriting document images. We find hardly the methods focus on both printed and handwriting document images.

Barboza et al. [9] have proposed a color-based model to determine the age of document for forensic purpose based on analyzing the color histograms of sample images. The method works well for the document of age and it is limited to specific applications. However, the color alone is not sufficient to detect forged words in noise Beusekom et al. [11] have presented a tool for detecting forgery based on text-line details Rotation and alignment of text lines can also provide useful hints for discovering altered documents during a questioned text review. Calculating and classifying certain improper alignment and rotations is a time-consuming task. Based on these observations, the authors have presented an automated approach for verification of documents. The features extracted in the method are not robust to the proposed work. Gebhardt et al. [12] have developed a method for comparing the edge roughness of laser-printed and inkjet-printed pages. The work presented here should be interpreted as a foundation for intrinsic document verification in the context of poor resolution scans. The difficult task here is to implement a more robust edge detection to improve the quality of the feature and document

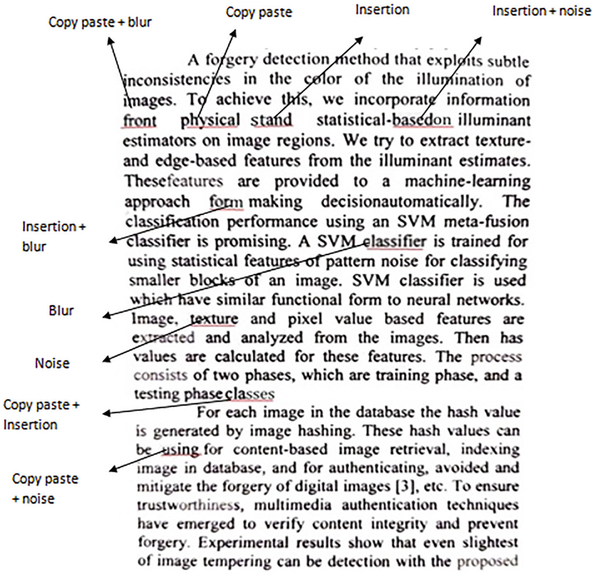

Fig. 1. An example of a forged printed document with ten classes of forgery at word level.

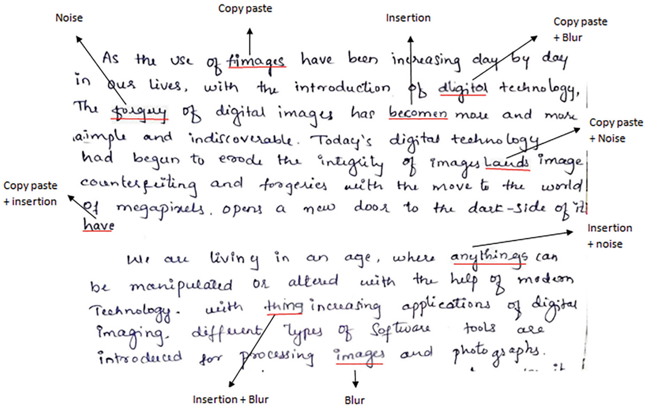

Fig. 2. An example of a forged handwritten document with ten classes of forgery at word level.

processing in order to identify image or italic text, which frequently leads to incorrect source identification. The method may not work for documents with blur and noise. Ryu et al. [13] have presented a method of detecting a forged document created by printers. In this work, seventeen different image quality measures were computed and trained using SVM (Support Vector Machine) classifiers. The method considers the quality measures as features by studying the quality of the images. This is good for the images affected by uniform quality factor else the method may not work well. Sometimes, the document can have different quality at different region in a single image. Chen et al. [14] propose a forensic technique for identifying global blur in entire images using no-reference image

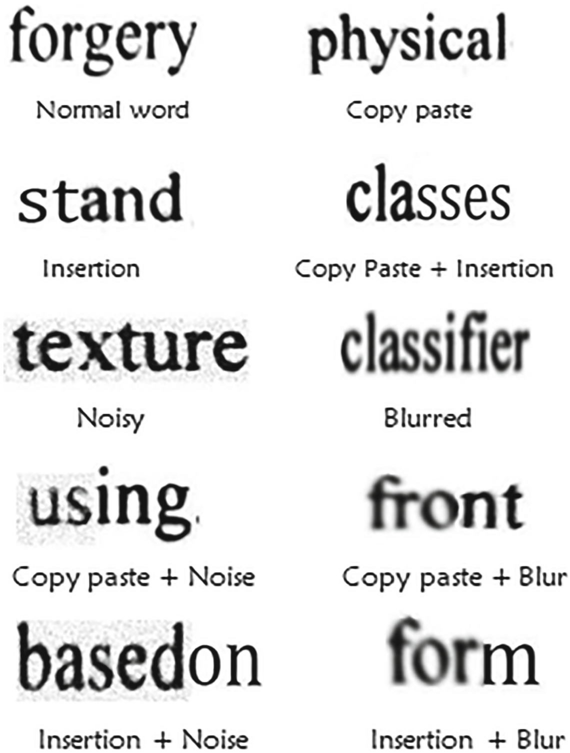

Fig. 3. Example for the forged words of all the 10 classes in order for printed document image shown in Fig. 1.

quality assessment. Using mean subtracted contrast normalized (MSCN) coefficients, the features are extracted and fed into SVM, which can distinguish the altered regions from the original ones and quantify them. Here the tampered images used are well resolution except tempered regions. The method gives false-alarm detection when the entire forged images are of weak resolution. Shang et al. [15] have presented a method of exposing document forgeries using distortion mutation of geometric parameters such as of translation and rotation distortions through image matching for each character. To detect tampered characters with distortion authors have used distortion probability, which is calculated from character distortion parameters. The method is suitable for document examination in both Chinese and English. The drawback of this method is it will work only on printed document which consist of only namely, Chinese and English but the method fails to work on handwritten documents. Cruz et al. [16] Explored classification based forgery detection method, which uses Local Binary Patterns (LBP) for computing discriminant texture features that are common on forged regions and then computed features are fed to Support Vector Machines (SVM) for classification. Author have performed 4 different types of forgery on printed document images namely, Copy-Paste Intra document, Copy-Paste Extra document, Imitation and Region cuts. The method represents much incorrect detection and is not acceptable for real time applications. The

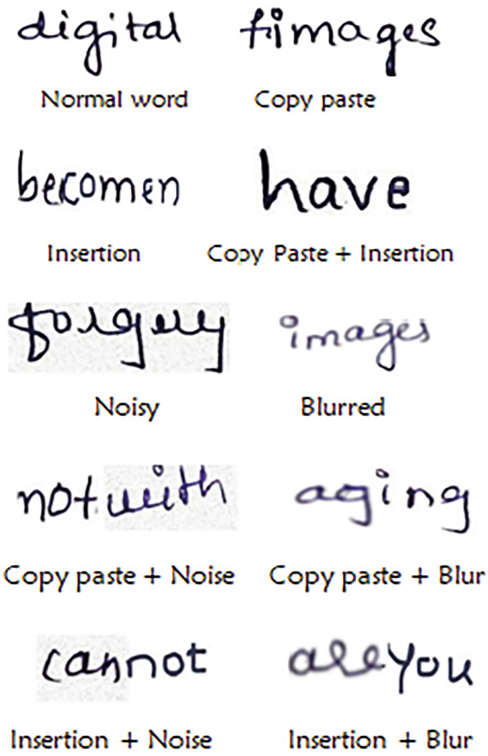

Fig. 4. Example for the forged words of all the 10 classes in order for handwriting document shown in Fig. 2

work is carried out on only printed document images not on handwritten documents. Megahed et al. [17] have proposed a method to detect handwritten forgery in text by detecting different ink using image processing. The features are extracted based on red, green and blue channels. Also computed distance measurements between each pairs of feature vector using Root Mean Square Error. Gorai et al. [18] proposed a method to perform forged handwriting inspection. The three RGB color channels of the handwritten picture were retrieved, as well as the texture features of the grayscale image, and the histogram matching approach was utilized.

It is observed from the preceding work that many researchers have worked for forgery detection in both printed and handwritten document images and raised the following challenges and issues.

- Detecting the altered text in different quality images.
- Handwritten documents written by different ink and pen.
- Document image contains both typed text and handwritten text.
- Document images affected by multiple forgery operations.
- Document images affected by distortion and noise generated by printer overlap.

- Identification altered text in document images rather than printer identification.

Hence to overcome the above challenges and issues the authors have introduced a ten class classification problem in handwritten and printed document images which contains text with variation in different writing styles and affected by multiple forgery operations. The dataset description is explained in detail under Implementation and Results section.

3 Proposed Methodology

The proposed methodology consists of three main steps, pre-processing, Feature Extraction and Classification of computed features as illustrated in the Fig. 5. The distortions in the original images are caused by tampering operations such as insertion, copy paste, copy paste + noise, copy paste + blur, Noise + Insertion, Insertion + Blur, and Insertion + Copypaste on the original images, which all creates complicated patterns. Whereas the Blur and Noise operations are produces the desired patterns on original and forged images using Gaussian distortions. Further, when we perform multiple forgery operations on text, the possibility and degree of image noise multiplies when compared to only one operation on text. Taking these observations into consideration, the traditional features like Image Quality Assessments with the help of a random forest tree classifier has been used to solve the ten-class problem. This is due to the fact that our primary objective is to distinguish between original and forged documents based on specific patterns and distortions created by various forgery operations.

3.1 Pre-processing

For the better understanding of image it is important to pre-process an image which may enhance some important features of an image. In this work preprocessing is carried out by converting color document image into gray scale images and resized to 350X350 for the proper analysis of Documents.

3.2 Feature Extraction

While working on an image it is important to compute the features of an image which is the method of capturing visual content of image for indexing and retrieval. The extraction of features is used to denote a piece of information that is important to solving a computer task related to a certain application. In this work we computed features of pre-processed handwritten and printed document images using MATLAB which results in recognition of accuracy with simple classification modules.

3.2.1 Image Quality Measures

Image Quality Assessment (IQA) is a process of extracting image quality features to determine whether or not an image is genuine. The fake image differs from the real image in numerous ways when a forgery operation is attempted. A number of factors influence image quality. Performance, diversity, and speed are the three characteristics to look at. Predictable quality is defined by varying degrees of sharpness, brightness, covariance,

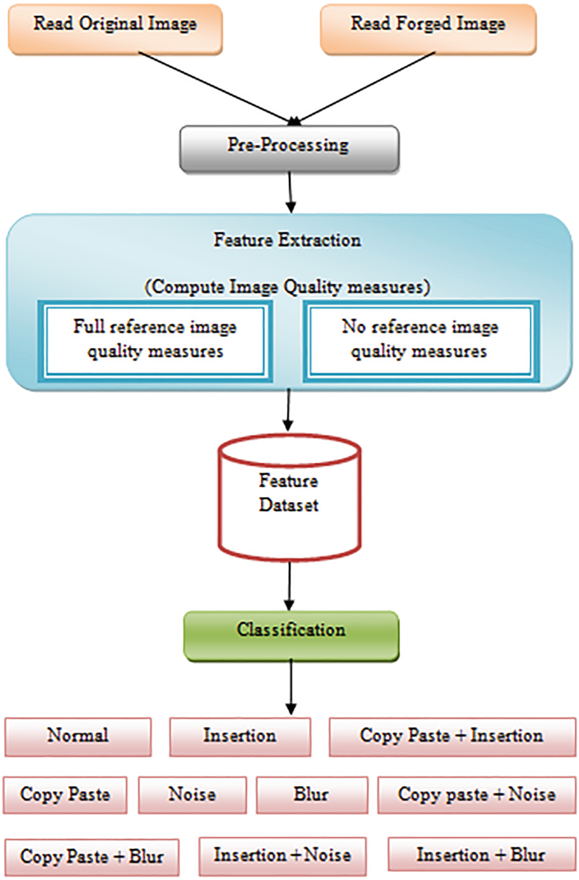

Fig. 5. Block diagram of proposed technique

blur, gradient, distortions, and a strong correlation, and the information generated by both types of images will differ in content. [19].

To assess the quality of distorted images, a variety of methodologies have been developed. Subjective and objective approaches of IQA can be distinguished. Subjective approaches cannot automate the system and are time-consuming and inconvenient because they are depending on human judgment. Evaluation of objective image quality is intended to provide quality measurements that can be used to predict image perception automatically [20]. The objective method is a quantitative strategy in which we utilize the intensity of two images, a reference and distorted type, to create a number that indicates image quality. Based on the availability of a reference image, objective methods are divided into three categories: full-reference, no-reference, and hybrid [21].

3.2.1.1. Full Reference Image Quality Measure

In Full Reference image quality evaluation approaches, the Qualitative aspect of a query image is assessed by contrasting it with a reference image of ideal quality. A number of Full Reference image quality assessment approaches are available. The most widely used and well-known methods are Peak Signal to Noise Ration and Mean Square error [20]. Some other quality measures are based on 1. Error Sensitivity Measure which includes Pixel Difference Measure, Correlation Based Measure, Edge based Measure, spatial based measures and Gradient Based Measure 2. Structural similarity index and 3. Information Theoretic that includes Visual information fidelity. Following are the full reference image quality measures were computed on forged handwritten and printed document images.

- Mean Squared Error: The mean squared error is measured as the mean of the "errors" squared in original image and reference image as represented in Eq. (1).

$$MSE = \frac{1}{N}\sum_{i=1}^{N}(X_i - Y_i)^2 \tag{1}$$

  Where N is number of samples, X is original image and Y is reference image.
- Root Mean Square Error: The squared root of MSE yields the Root Mean Square Error (RMSE). The root mean square error (RMSE) is a metric that indicates how much a pixel changes as a result of processing, as represented in Eq. (2).

$$RMSE = \sqrt{MSE} \tag{2}$$

- Peak Signal to Noise Ratio (PSNR): Using PSNR, we can calculate the ratio between the greatest possible signal strength and the power of distortion, which has an impact on the quality of its representation [24]. The below Eq. (3) used to calculate the PSNR.

$$PSNR = 10log_{10}\frac{P^2}{MSE} \tag{3}$$

  Where P is dynamic range of pixel intensity
- Signal to Noise Ratio (SNR): The signal-to-noise ratio is the proportion of desired information (signal power) to undesired information (background noise power). SNR calculates the signal-to-noise ratio (SNR) of a signal X in decibels by dividing its summed squared magnitude by the noise, Y as represented in below Eq. (4).

$$SNR = 10log_{10}\left(\frac{\sum_{i=1}^{n} X_i^2}{\sum_{j=1}^{n} Y_i^2}\right) \tag{4}$$

- Structural Content (SC): It is calculated as the square of sum of the original and referred image, which is represented in below Eq. (5).

$$SC(X, Y) = \frac{\sum_{i=1}N\sum_{j=1}M\,(X_i, j)^2}{\sum_{i=1}N\sum_{j=1}M\,(Y_i, j)^2} \tag{5}$$

- Maximum Difference (MD): The highest value of the absolute difference image is computed i.e. A subtraction is made between the original and the reference image. The below Eq. (6) used to calculate MD

$$MD(X, Y) = max\left|X_{i,j} - Y_{i,j}\right| \tag{6}$$

- Average Difference (AD): It is calculated for each pixel in an image to determine the absolute difference average. A subtraction is made between the original and the reference image. The Eq. (7) represents the formulae to calculate AD.

$$AD(X, Y) = \frac{1}{NM}\sum\nolimits_{i=1} N \sum\nolimits_{j=1} M\,(X_{i,j} - Y_{i,j}) \tag{7}$$

- Normalized Absolute Error(NAE): It is calculated by dividing the total of the difference image by the total of the original image as given in below Eq. (8).

$$NAE(X, Y) = \frac{\sum_{i=1} N \sum_{j=1} M \left|X_{i,j} - Y_{i,j}\right|}{\sum_{i=1} N \sum_{j=1} M \left|X_{i,j}\right|} \tag{8}$$

- Normalized Cross-Correlation (NK): The simplest but most effective similarity measure is normalized cross correlation, which is unaffected by linear brightness and contrast fluctuations which is represented by following Eq. (9).

$$NK(X, Y) = \frac{\sum_{i=1}^{N} \sum_{j=1}^{M} \left|X_{i,j}.Y_{i,j}\right|}{\sum_{i=1}^{N} \sum_{j=1}^{M} \lceil X_{i,j} \rceil^2} \tag{9}$$

- Laplacian Mean Squared Error (LMSE): The term Laplacian Mean Square Error refers to the calculation of the normal mean square error. The difference is that the mean square error is determined using the laplacian value of the data rather than the predicted and obtained data [25]. The given Eq. (10) used to calculate LMSE.

$$LMSE(X, Y) = \frac{\sum_{i=1}^{N} \sum_{j=1}^{M} \left(h\left(X_{ij}\right) - h\left(Y_{ij}\right)\right)}{\sum_{i=1}^{N} \sum_{j=1}^{M} \left(h\left(X_{ij}\right)\right)^2} \tag{10}$$

- Total Edge Difference (TED): It's define d as the ratio of the two images' total number of edge differences to the total number of pixels, as given in Eq. (11).

$$TED(X, Y) = \frac{1}{NM}\sum_{i=1} N \sum_{j=1} M \left|XE_{i,j} - YE_{i,j}\right| \tag{11}$$

- Total Corner Difference: It is the ratio of the total amount of edge variations between two images to entire pixels in the image, as represented in below Eq. (12).

$$TCD(X, Y) = \frac{|Xtcr - Ytcr|}{max|Xtcr - Ytcr|} \tag{12}$$

- Gradient Magnitude Error: The total number of pixels is used to average the difference between the gradients of the original image, as well as the gradients of the reference image. The following Eq. (13) shows the representation of GME

$$\mathrm{GME(X, Y)} = \frac{1}{\mathrm{NM}}\sum_{\mathrm{i=1}} \mathrm{N} \sum_{\mathrm{j=1}} \mathrm{M}\left(\left|\mathrm{XG_{i,j}}\right| - \left|\mathrm{YG_{ij}}\right|\right)^2 \tag{13}$$

- Gradient Phase Error: It is calculated using the overall count of pixels as the mean deviation between the gradient angle of the real image and the gradient angle of the reference image, which is defined in following Eq. (14)

$$GPE(X,Y) = \frac{1}{NM}\sum_{i=1} N \sum_{j=1} M \left| arg\left(XG_{ij}\right) - arg\left(YG_{ij}\right)\right|^2 \tag{14}$$

- Structural Similarity Index Measure: The notion of image quality assessment based on structural similarity emerged from the idea that the visual system of human is well-adapted for obtaining structural information from the observing field. The most straightforward formulation is the Structural Similarity Index Measure (SSIM), which is broadly used in a variety of relevant implementations. SSIM Means of measuring loss of structure in the image instead of any deviation from the reference. Loss of visual structure as assessed on a local scale using luminance, contrast and structural similarity [26] as represented in Eq. (15), (16) and (17) respectively

$$L(X,Y) = \frac{2\mu_x\mu_y + C1}{\mu_x^2 + \mu_y^2 + C1} \tag{15}$$

$$C(X,Y) = \frac{2\sigma_x\sigma_y + C2}{\sigma_x^2 + \sigma_y^2 + C2} \tag{16}$$

$$S(X,Y) = \frac{2\sigma_{xy} + C3}{\sigma_x + \sigma_y + C3} \tag{17}$$

where $\mu_x\mu_y$ are the mean values of original and reference images, $\sigma_x\sigma_y$ indicates standard deviation of original and reference images, σ_{xy} represents the covariance of original and reference image and C1, C2, C3 are the constants. Depending upon the above three equations the SSIM is represented as in Eq. (18).

$$SSIM(X,Y) = L(X,Y).C(X,Y).S(X,Y) = \frac{\left(2\mu_x\mu_y + C1\right)\left(2\sigma_{xy} + C2\right)}{\left(\mu_x^2 + \mu_y^2 + C1\right)\left(\sigma_x^2 + \sigma_y^2 + C2\right)} \tag{18}$$

- Feature based similarity index: Index of Feature Similarity The method compares two images by mapping their features and measuring their similarity.
- Visual Information Fidelity (VIF): Visual images are regarded as natural scenarios based on the VIF model with statistical qualities similar to those of natural scenarios [19]. The Visual Saliency Induced quality measure assumes that image deterioration causes changes in salient regions that are strongly connected to changes in visual quality [27].

3.2.1.2. No Reference Image Quality Measure
This is a method for estimating the quality of a blind image. Without a reference image, the perceived image quality is estimated here [22]. In recent decades, NR-IQA has received a lot of attention. Although NR-IQA algorithms do not have access to a reference image, they can assume things about the distortions inherent in a specific input image [23]. As a result, they can be classified as distortion-specific measures that cope with image quality indexes, traditional-based measures that deal with blind/referenceless image spatial quality evaluator, and Natural image quality evaluator.

Following are the No reference image quality measures were computed on forged handwritten and printed document images.

- Brisque: The Brisque model assesses image quality by employing the locally normalized luminance coefficients that were used to compute image features.
- Natural Image Quality Evaluator: To train the first model, it leverages a priori knowledge extracted from distortion-free images of natural scenes. The Natural Image Quality Evaluator (NIQE) is an absolutely blind quality of the image analyzer that is actually based on the development of a performance aware set of numerical features linked to a multi variate Gaussian natural scene statistical approach.

3.3 Classification

For the experiment, an image dataset of 950 handwritten and forged documents were used, which are classified into ten different classes: Normal, Noise, Blur, Insertion, Copy Paste, CopyPaste + Noise, CopyPaste + Blur, Insertion + Noise, Insertion + Blur and Insertion + Copy Paste. We used the above-mentioned image quality measures as a single feature vector to classify these various forged input images. Various classification techniques such as KNN, Naïve Bayes and Random Forest tree classifiers were used in the experiment. The results of KNN and Naïve Bayes classifiers are insufficient. To improve the accuracy rate we used the Random Forest Classifier and it produced good results.

3.3.1 KNN

One of the least complicated ones is K-Nearest Neighbor. The Machine Learning Algorithm is based on the Supervised Learning procedure, in which the classifier basically obtains the similarity between the test information and the training set [28]. The classification of the class labels depends on calculating the distance between training and testing dataset. KNN classifies the data with Suitable K value thus finds a closest neighbor that provides a class label to un-labeled images [29]. A verity of distance measures was implemented depending on type of problem. City-block distance with K value equals to 3 is considered whose value is empirically fixed. KNN shows the test data M, and then finds the distance D between training sample X and testing pattern N using the following equation.

$$\mathrm{Dcity}(\mathrm{X}, \mathrm{Y}) = \sum_{j=0}^{n} |X_J - Y_J| \tag{19}$$

3.3.2 Naïve Bayes

Naïve Bayes is a simple probabilistic classifier that predicts on basis of probability of objects which works based on Bayes theorem [30] as given in below equation.

$$P(X|Y) = \frac{P(Y|X)P(X)}{P(Y)} \tag{20}$$

where,

- P (X|Y) is Posterior probability: Provided proof that Y has already occurred, the probability of X occurring.
- P (Y|X) is Likelihood probability: Provided proof that X has already occurred, the probability of Y occurring.
- P (X) is Prior Probability: Probability of X Occurring
- P (Y) is Marginal Probability: Probability of Y Occurring

3.3.3 Random Forest

Random Forest is a well-known and widely used machine learning algorithm. The forest of decision trees is known as Random Forest. This technique can be used for both classification and regression tasks. It is a decision tree ensemble that predicts the outcomes depending on a set of variables as well as rules and aggregates the outcomes of multiple decision trees in order to achieve better performance. The combination of each decision tree's outcomes reduces the total generalization fault and the over fitting issue. [31]. In this experiment random forest tree classifier has achieved better results as compared to other classifiers. The Table 2 provides the detailed analysis and performance test with different number of trees of random forest tree on forged handwritten and printed document images.

The proposed method is represented in form of algorithm as follows.

Algorithm:

Input: Handwritten / Printed Forged Document Image
Output: Classification of handwritten/Printed Forged Document Image.

Step 1: Acquisition of Handwritten/Printed Forged Document Image

Step 2: Pre-Processing i.e. converting into Gray scale and resize the images to 350X350 for further analysis

Step 3: Computation of features using Image Quality Measures I.e. Full Reference and No Reference quality Measures

Step 4: Classification of Obtained Features using KNN, Naïve Bayes and Random Forest classifiers

Step 5: Output of the predicted class.

4 Implementation and Results

The experiment is carried out on own created dataset which consist of 950 Forged documents (500 forged handwritten + 450 forged printed document images) of 10 different classes as discussed Table 1. Full reference and No Reference Image Quality Measures were computed and classified using Classifiers such as Nave Bayes, K-NN and Random Forest. Among these classifiers random forest tree gave good results. Hence the results of random forest tree are predicted.

4.1 Dataset

From the literature review, it has been found that the standard datasets available includes documents with few forgery operations such as noise, blur, copy paste and insertion but they do not contain the handwritten documents with multiple forgery operations. In order to cope up with the problem own dataset has been created with 950 forged document images that includes 500 forged handwritten document images and 450 forged printed document images with 10 different class of different forgery operations like, Copy-Paste, Noise, Blur, Insertion, Copy-Paste + Noise, Copy-Paste + Blur, Insertion + Noise, Insertion + Blur, and Copy-Paste + Insertion and Normal. Each class of forged handwritten document images consist of 50 images and in forged printed document images, each class consist of 45 images. Initially, LaserJet M1136 MFP scanner is used to scan the Handwritten and printed documents were with 200 DPI and then performed 10 different tampering operations on each image.

The description for all the 10 classes is presented in Table 1, where we can see operations are used for creating forged words. When conducting multiple forgery operations on text, one operation uses half portion of the word and another half portion of the word is affected by other forgery operation. If the forged word is created by single operation, the whole word is affected by the operation. It is noted from Fig. 3 and Fig. 4 that it is difficult to notice the difference between original and forged words except blur and noise.

The quantitative classification results of proposed method using fusion of full reference and no reference image quality measures on forged Handwritten and forged printed documents were represented in Table 3 and Table 4 respectively. It is observed from Table 3 and Table 4 that the values present diagonally in table are considered to be correct classification and off diagonal values represents misclassification. Table 5 represents the performance analysis of individual features and fusion of features using KNN, Naïve Bayes and Random Forest tree classifiers on forged handwritten and printed document images. The performance of proposed methodology is evaluated in terms of metrics, such as Precision, F_Score, Recall, and Accuracy as represented in Eq. (21) to Eq. (24) respectively.

$$Precision = \frac{TruePositive}{TruePositive + FalsePositive} \tag{21}$$

$$F_{Score} = \frac{2 * Precision * Recall}{Precision + Recall} \tag{22}$$

Table 1. Ten- Class Classification Problem

Forgery Type	**Description**
Class 1: Normal Words	Original words without affecting by any forgery operations.
Class 2: Copy- Paste	By using a copy-paste procedure, forged words are formed.
Class 3: Insertion	Words that have been forged as a result of an insertion operation
Class 4: Copy-Paste + Insertion	Both copy-paste and insertion operations result in forged words.
Class 5: Noise	Adding various types of noise to the original images.
Class 6: Blur	Adding blur to the original images.
Class 7: Copy-Paste + Noise	Add different noises to forged words created by a copy-paste operation.
Class 8: Copy-Paste + Blur	Add a different blur to forged words created by a copy-paste operation.
Class 9: Insertion + Noise	Add different noises to the forged words created by the insertion technique.
Class 10: Insertion + Blur	Add a different blur to forged words created by the insertion operation.

Table 2. Performance Analysis Test with Different Number of Trees of Random Forest On Forged Handwritten and Printed Document Images

SL No	Features	Number of Trees	Accuracy (in %)	
			Forged Handwritten Document	Forged Printed Document
1	Full Reference + No Reference Image Quality Measures	7	97.05%	92.43%
2		8	97.23%	92.90%
3		9	97.86%	93.03%
4		10	98.00%	93.56%
5		11	98.45%	93.89%
6		**12**	**98.80%**	**94%**

$$Recall = \frac{TruePositive}{TruePositive + FalseNegative} * 100 \tag{23}$$

$$\text{Accuracy} = \frac{\text{TruePositive} + \text{TrueNegative}}{\text{TruePositive} + \text{TrueNegative} + \text{FalsePositive} + \text{FalseNegative}} * 100 \tag{24}$$

Where the class Abbreviations in Table 3 and Table 4 are: BLUR: Blur, CP + NS: Copy Pate + Noise, INS + NS: Insertion + Noise, NOISE: Noise, CP + BLR: Copy Paste + Blur, NRM: Normal, CP: Copy Paste, CP + INS: Copy Paste + Insertion, INS + BLUR: Insertion + Blur, INS: Insertion.

Where the Abbreviations in Table 5 are: FRIQM: Full Reference Image Quality Measure, NRIQM: No Reference Image Quality Measure.

From the Table 5, it is observed that individual performance of full reference image quality measures on forged handwritten and printed document images using KNN classifier is achieved as 78.2% and 74.7% respectively, using Naïve Bayes achieved as 83.5% and 80.6% respectively and using random forest tree classifier is achieved as 96.8% and 92.2% respectively. For the individual performance of no reference image quality measures on forged handwritten and printed document images using KNN classifier is achieved as 79.2% and 75.3% respectively, using Naïve Bayes achieved as 86.9% and 84.1% respectively and using random forest tree classifier is achieved as 97.4% and 93.11% respectively. The fusion of full reference and no reference image quality measure on forged printed document images and forged handwritten images yields the highest accuracy rate of **94%** and **98.80%** respectively using Random Forest tree classifier whereas, KNN and Naïve Bayes classifiers results in lower accuracy rate as compared to random forest tree classifier. Figure 6 represents the graphical representation of classification performance of individual and fusion of features.

In order to conduct the experiments, MATLAB R2018a image processing tool box was used on a machine equipped with an Intel Core i5-6200U @ 2.40 GHz, 4.00 GB of RAM, and a 64-bit operating system.

Table 3. Confusion Matrix for Fusion of Full Reference and No Reference Image Quality Measures On Forged Handwritten Documents Using Random Forest Tree Classifier.

	BLUR	**CP**	**NRM**	**CP + BLR**	**CP + INS**	**CP + NS**	**INS + BLR**	**INS**	**INS + NS**	**NOISE**
BLUR	**50**	0	0	0	0	0	0	0	0	0
CP	0	**49**	1	0	0	0	0	0	0	0
NRM	0	1	**49**	0	0	0	0	0	1	0
CP + BLR	0	0	0	**50**	0	1	0	0	1	0
CP + INS	0	0	0	0	**50**	0	0	0	0	0
CP + NS	0	0	0	0	0	**49**	0	0	0	0
INS + BLR	0	0	0	0	0	0	**50**	0	0	0
INS	0	0	0	0	0	0	0	**50**	0	1
INS + NS	0	0	0	0	0	0	0	0	48	0
NOISE	0	0	0	0	0	0	0	0	0	**49**

5 Statistical Test of Significance

Statistical test of significance is used to evaluate the experimental findings are statistically significant or no. It's the way of evaluating obtained results to a predictable data assertion. The Chi-Square Test is used to validate the statistical inference in this study at a significance level of 5%.

A chi-square statistic is an assessment that compares a framework to actual observations. The null hypothesis, alternative hypothesis, and degrees of freedom in this test are as follows:

- Null Hypothesis (H_0): There is a strong correlation between the findings of the proposed methodology and the total number of forged document images.
- Alternative Hypothesis (H_1): There is no strong correlation between the findings of the proposed methodology and the total number of forged document images.
- Degree of Freedom (df) = 9, At a 5% significance level, the critical value of x^2 with df = 9 is 16.92. (From the chi-square table).

Table 4. Confusion Matrix for Fusion of Full Reference and No Reference Image Quality Measures On Forged Printed Documents Using Random Forest Tree Classifier

	BLUR	**NRM**	**CP**	**INS + BLR**	**CP + BLR**	**CP + INS**	**CP + NS**	**INS**	**INS + NS**	**NOISE**
BLUR	**44**	1	0	0	0	0	0	0	0	0
NRM	1	**40**	0	0	0	0	0	0	0	0
CP	0	0	**38**	0	0	0	2	0	1	0
INS + BLR	0	1	1	**43**	2	0	0	0	0	0
CP + BLR	0	0	4	1	**43**	1	0	1	0	0
CP + INS	0	0	1	0	0	**44**	0	0	0	0
CP + NS	0	1	1	0	0	0	**43**	0	0	0
INS	0	0	0	0	0	0	0	**39**	0	0
INS + NS	0	1	0	0	0	0	0	5	**44**	0
NOISE	0	1	0	1	0	0	0	0	0	**45**

Table 5. The Performance Comparison of the Individual Features and Fusion of Features Using KNN, Naïve Bayes and Random Forest Tree Classifier On Forged Printed and Handwritten Document Images (In %)

Feature Set	**Classifier**	**Forged Handwritten Documents**				**Forged Printed Documents**			
		Precision	**Recall**	**F-Score**	**Accuracy**	**Precision**	**Recall**	**F-Score**	**Accuracy**
FRIQM	**KNN**	0.7805	0.7621	0.7831	78.2%	0.7463	0.7539	0.7498	74.7%
	Naïve Bayes	0.8301	0.8391	0.8269	83.5%	0.8051	0.8134	0.8083	80.6%
	Random Forest	0.9605	0.9738	0.9692	96.8%	0.9201	0.9386	0.9231	92.2%
NRIQM	**KNN**	0.7816	0.7864	0.7963	79.2%	0.7502	0.7653	0.7564	75.3%
	Naïve Bayes	0.8521	0.8591	0.8693	86.9%	0.8410	0.8409	0.8406	84.1%
	Random Forest	0.9683	0.9829	0.9761	97.4%	0.9218	0.9435	0.9318	93.11%
FRIQM + NRIQM	**KNN**	0.7904	0.7896	0.7801	79.8%	0.7621	0.7713	0.7631	76.4%
	Naïve Bayes	0.8742	0.8861	0.8761	87.7%	0.8510	0.8614	0.8493	85.6%
	Random Forest	0.9703	0.9958	0.9834	**98.80%**	0.9400	0.9410	0.9405	**94%**

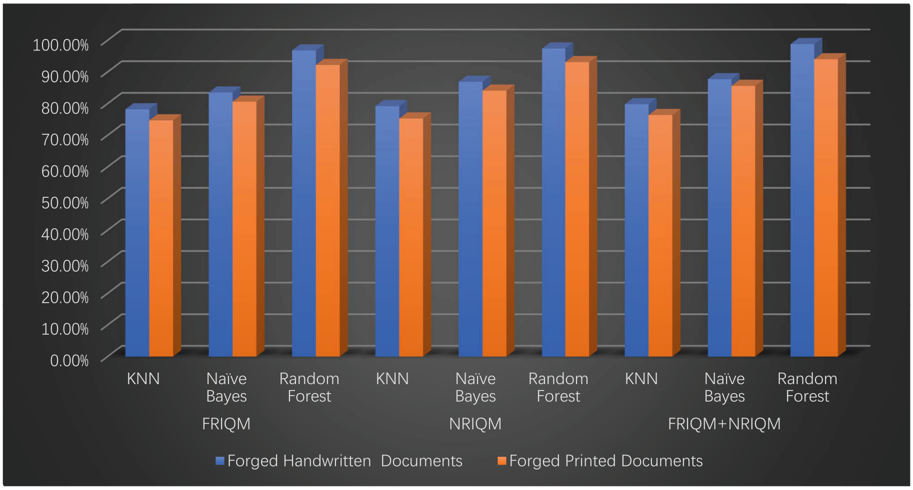

Figure. 6. The graphical representation of classification performance of individual and fusion of features.

If $x^2 < 16.92$, Accept H_0 and Reject H_1, else vice versa.

The Table 6. Represents the performance analysis of Chi-Square examination on total forged document images and proposed algorithm for classification. Eq. (23) defines the computation of Chi-Square Statistics.

$$Chi - Square(x^2) = \sum \frac{(D_o - D_e)^2}{D_e} = 1.392 \tag{23}$$

The Chi Square statistic's determined value is less than the critical value from the chi square table. As a result, the Null Hypothesis H_0 is acceptable, whereas the alternative hypothesis H_1 is rejected. It illustrates that the proposed approach and the observations provided by the entire quantity of forged document images have a significant association.

Table 6. Performance Analysis of Chi-Square Test On Total Forged Document Images and Proposed Algorithm for Classification.

Forgery Types	Total Forged Images (Handwritten + Printed) Hi	Proposed Method (Handwritten + Printed) Pi	Total (Hi + Pi)	Expected Values (De)	Observed Values (Do)	$x^2 = \sum \frac{(D_o - D_e)^2}{D_e}$
Blur	95	94	189	96	93	0.093
CopyPaste	95	87	182	93	89	0.172
CopyPaste + Insertion	95	94	189	96	93	0.093

(*continued*)

Table 6. (*continued*)

Forgery Types	Total Forged Images (Handwritten + Printed) Hi	Proposed Method (Handwritten + Printed) Pi	Total (Hi + Pi)	Expected Values (De)	Observed Values (Do)	$x^2 = \sum \frac{(D_o - D_e)^2}{D_e}$
CopyPaste + Blur	95	93	188	96	92	0.167
CopyPaste + Noise	95	92	187	95	92	0.095
Insertion	95	89	184	94	90	0.170
Insertion + Blur	95	93	188	96	92	0.167
Insertion + Noise	95	92	182	93	89	0.172
Noise	95	94	189	96	93	0.094
Normal	95	89	184	94	90	0.170
Total	$\sum$Hi = $\sum H_i = 950$	$\sum P_i = \sum$Pi = 917	$\sum (H_i + P_i) =$ 1862			x^2**=1.392**

6 Conclusion

In this paper, we proposed a method for classifying forged handwritten and printed document Images. The proposed method investigates the extraction of image quality measures such as full reference and no reference image quality measures from document images containing words affected by ten different types of forgery operations, including copy paste, Noisy, Blurred, Insertion, copy paste + noise, copy paste + blur, copy paste + insertion, insertion + noise, insertion + blur, and normal. Document images are extremely susceptible to unwanted distortions caused by scanners while scanning documents, as well as distortion caused by forgery operations, which may overlap with distortion in normal images and degrades the image quality. To address these issues, the authors employed techniques that provide an effective method for better understanding and analyzing handwritten and printed document images. Using the Random Forest classifier, the method achieves an accuracy rate of 94% on forged printed document images and 98.80% on forged handwritten document images. In the future, technology for automatic detection of forged words affected by multiple forgery operations in both handwritten and printed document images need to be developed.

Acknowledgement. The authors would like to express their gratitude to all students at Rani Channamma University in Belagavi, Karnataka, India, who willingly shared samples of handwritten and printed documents for this research study.

Authors' Contributions. Gayatri Patil[1]: Conceptualization, Methodology, Software, Field study Shivanand Gornale[2]: Data curation, Writing-Original draft preparation, Software, Validation., Field study Ashivini Babaleshwar[3]: Visualization, Investigation, Writing-Reviewing and Editing.

Compliance With Ethical Standards. • Conflict of Interest: The authors state that they do not have any conflicts of interest.

• Research involving Human and Animal Rights: There are no animal trials by any of the authors in this paper.

• Ethical Standards: All procedures used in research involving human volunteers were compatible with the institution's ethical guidelines.

References

1. S. Sapna, S. Vaibhav , A.K.Gupta "A Review of Trends In Digital Image Processing For Forensic Consideration" International Journal of Software and Hardware Researh in Engineering (IJSHRE), ISSN-2347–4890 Volume 3 Issue 8 August, 2015.
2. S. Kundu, P. Shivakumara, A. Grouver, U. Pal, T. Lu and M. Blumenstein," A New Forged Handwriting Detection Method Based on Fourier Spectral Density and Variation", In Proc. Of Autorité de contrôle prudentiel et de résolution (ACPR)pp 136–150, 2019.
3. L. Su, C. Li, Y. Lai and J. Yang, "A Fast Forgery Detection Algorithm Based on Exponential-Fourier Moments for Video Region Duplication," in IEEE Transactions on Multimedia, vol. 20, no. 4, pp. 825-840, doi: https://doi.org/10.1109/TMM.2017.2760098, April 2018.
4. L. D'Amiano, D. Cozzolino, G. Poggi and L. Verdoliva, "A PatchMatch-Based Dense-Field Algorithm for Video Copy–Move Detection and Localization," in IEEE Transactions on Circuits and Systems for Video Technology, vol. 29, no. 3, pp. 669–682, doi: https://doi.org/10.1109/TCSVT.2018.2804768, March 2019
5. HanyFarid "Image Forgery Detection: A Survey" IEEE Signal Processing Magazine March 2009.
6. B. Sarma, G. Nandi," A Study on Digital Image Forgery Detection", International Journal of Advanced Research in Computer Science and Software Engineering, Volume 4, Issue 11, ISSN: 2277 128X, November 2014
7. Z. Luo, F. Shafait and A. Mian, "Localized forgery detection in hyperspectral document images," 2015 13th International Conference on Document Analysis and Recognition (ICDAR), Tunis, Tunisia, pp. 496–500, doi: https://doi.org/10.1109/ICDAR.2015.7333811, 2015
8. M. J. Khan, A. Yousaf, K. Khurshid, A. Abbas and F. Shafait, "Automated Forgery Detection in Multispectral Document Images Using Fuzzy Clustering," 13th IAPR International Workshop on Document Analysis Systems (DAS), Vienna, Austria, 2018, pp. 393–398, doi: https://doi.org/10.1109/DAS.2018.26, 2018
9. R. d. S. Barboza, R. D. Lins and D. M. d. Jesus, "A Color-Based Model to Determine the Age of Documents for Forensic Purposes," 12th International Conference on Document Analysis and Recognition, Washington, DC, USA, 2013, pp. 1350–1354, doi: https://doi.org/10.1109/ICDAR.2013.273, 2013
10. L. Nandanwar . et al. "A New Method for Detecting Altered Text in Document Image". In: Lu Y., Vincent N., Yuen P.C., Zheng WS., Cheriet F., Suen C.Y. (eds) Pattern Recognition and Artificial Intelligence. ICPRAI 2020. Lecture Notes in Computer Science, vol 12068. Springer, Cham. https://doi.org/10.1007/978-3-030-59830-3_8, 2020

11. Beusekom, Joost & Shafait, Faisal & Breuel, Thomas. "Text-line examination for document forgery detection", International Journal on Document Analysis and Recognition (IJDAR). 16. 189–207. https://doi.org/10.1007/s10032-011-0181-5, 2012
12. J. Gebhardt, M. Goldstein, F. Shafait and A. Dengel, "Document Authentication Using Printing Technique Features and Unsupervised Anomaly Detection," 2013 12th International Conference on Document Analysis and Recognition, Washington, DC, USA, pp. 479–483, doi: https://doi.org/10.1109/ICDAR.2013.102, 2013
13. S. J. Ryu, H. Y. Lee, I, W. Cho, and H. K. Lee," Document Forgery Detection with SVM Classifier and Image Quality Measures",Lecture Notes in Computer Science, Springer-Verlag, PP 486– 495, doi:https://doi.org/10.1007/978-3-540-89796-5_50 ,2008
14. Z. Chen, Y. Zhao and R. Ni, "Forensics of blurred images based on no-reference image quality assessment," 2013 IEEE China Summit and International Conference on Signal and Information Processing, pp. 437–441, doi: https://doi.org/10.1109/ChinaSIP.2013.6625377, 2013.
15. Shang, Shize, Xiangwei Kong, and Xingang You. "Document forgery detection using distortion mutation of geometric parameters in characters." Journal of Electronic Imaging 24.2 : 023008. Doi: https://doi.org/10.1117/1.JEI.24.2.023008,2015
16. F. Cruz, N. Sidère, M. Coustaty, V. P. D'Andecy and J. Ogier, "Local Binary Patterns for Document Forgery Detection," 2017 14th IAPR International Conference on Document Analysis and Recognition (ICDAR), pp. 1223–1228, doi: https://doi.org/10.1109/ICDAR.2017.202, 2017
17. A. Megahed, S. M. Fadl, Q. Han and Q. Li, "Handwriting forgery detection based on ink colour features," 2017 8th IEEE International Conference on Software Engineering and Service Science (ICSESS), pp. 141–144, doi: https://doi.org/10.1109/ICSESS.2017.8342883, 2017.
18. A. Gorai, R. Pal and P. Gupta, "Document fraud detection by ink analysis using texture features and histogram matching," 2016 International Joint Conference on Neural Networks (IJCNN), 2016, pp. 4512-4517, doi: https://doi.org/10.1109/IJCNN.2016.7727790.
19. P. Gupta, N. Gyanchandani, "Image Quality Assessment for Fake Biometric Detection: Application to Iris, Fingerprint and Face Recognition", International Journal of Science and Research (IJSR), Volume 7 Issue 5, May 2017, 2166 – 2169
20. S. Sonawane ,A. M. Deshpande, " Image Quality Assessment Techniques: An Overview", International Journal Of Engineering Research & Technology (IJERT) Volume 03, Issue 04 (April 2014)
21. P. Sejal, S. Shubha, "Survey on Image Quality Assessment Techniques", International Journal of Science and Research (IJSR) , Volume 4 Issue 7, July 2015, 1756 - 1759
22. G. Minakshi , A. Mala , "Image Quality Parameter Detection : A Study," *International Journal of Computer Sciences and Engineering*, Vol.04, Issue.07, pp.110-116, 2016.
23. Varga, Domonkos.. "No-Reference Image Quality Assessment with Global Statistical Features" *Journal of Imaging* 7, no. 2: 29. https://doi.org/10.3390/jimaging7020029, 2021.
24. Sara, U. , Akter, M. and Uddin, M. " Image Quality Assessment through FSIM, SSIM, MSE and PSNR—A Comparative Study". Journal of Computer and Communications, 7, 8-18. doi: https://doi.org/10.4236/jcc.2019.73002,2019.
25. Krishnamoorthy, Shivsubramani & Kp, Soman." Implementation and Comparative Study of Image Fusion Algorithms" , International Journal of Computer Applications. 9. https://doi.org/10.5120/1357-1832, 2010.
26. Z. Wang, AC. Bovik, HR Sheikh, EP Simoncelli . "Image quality assessment: from error visibility to structural similarity", IEEE Trans Image Process. 2004 Apr;13(4):600-12. doi: https://doi.org/10.1109/tip.2003.819861. PMID: 15376593, 2004
27. Ding, K., Ma, K., Wang, S., & Simoncelli, E. P. "Comparison of Full-Reference Image Quality Models for Optimization of Image Processing Systems". International journal of

computer vision, 1–24. Advance online publication. https://doi.org/https://doi.org/10.1007/s11263-020-01419-7,2021
28. A. Kesarwani, S. S. Chauhan and A. R. Nair, "Fake News Detection on Social Media using K-Nearest Neighbor Classifier," 2020 International Conference on Advances in Computing and Communication Engineering (ICACCE), Las Vegas, NV, USA, 2020, pp. 1–4, doi: https://doi.org/10.1109/ICACCE49060.2020.9154997.
29. S.Dhivya, B. Sudhakar," Forgery Detection Based on KNN Classifier using SURF Feature Extraction", International Journal of Recent Technology and Engineering (IJRTE) ISSN: 2277–3878, Volume-8 Issue-2, July 2019
30. R. Mallika," Fraud Detection using Supervised Learning Algorithms" International Journal of Advanced Research in Computer and Communication Engineering, (IJARCCE), ISSN (Print) 2319 5940, https://doi.org/10.17148/IJARCCE.2017.6602, Vol. 6, Issue 6, June 2017
31. Yaram, S,"Machine learning algorithms for document clustering and fraud detection", 2016 International Conference on Data Science and Engineering (ICDSE). doi:https://doi.org/10.1109/icdse.2016.7823950, 2016