



# DETECTING FORGED SCAN OF EDUCATIONAL CERTIFICATES USING GLCM AND SVD ALGORITHM

Miss.U.Sathiya<sup>1</sup> and Mrs.P.Jasmine Lois Ebenezar<sup>2</sup> and Mrs.S.Cephas<sup>3</sup>  
Department of Computer Applications, Sarah Tucker College, Thirunelveli-7.

## ABSTRACT

In today's world, fake academic certificates scams are getting more and more common. These fake certificates are spoiling most important part and sole of the country that is education system. Hence protecting society from these fake academic documents holder is very important otherwise one day it will spoil whole society. In protecting country from these fake certificate scams, Digital system can play very crucial role. we can avoid fake certificates and we can make our society fake document less society and we can build good education system. Detection of forge scan certificates which are used during college admissions are done using scan copies from other genuine resources and materials and resources applying Photoshop and other image processing tools. This kind of certificates from different resources are used for creating scan certificate copies using Photoshop and giving that information to the name of other candidates. This situation leads to a point whereupon digital forgery can compromise the authenticity of the original documents. Using fake certificates to get jobs is worst as the eligibility of the candidate cannot be verified by administrative authorities in educational institutes. Nowadays many research oriented tools are used for detecting copied sections from different documents. In order to enhance trustworthiness and authenticate images, we have defined a feature point matching algorithm to identify forgery done in inaccurate manner. This project has four modules such as preprocessing, crop image, and feature extraction and forgery detection. In the preprocessing module, the input image is enhanced and Gaussian noise is applied. Second module is crop image which is crop the college logo, photo, student sign and examination controller sign. Third module is feature extraction. After cropping the images, the GLCM features are extracted from the cropped image. Finally the image is classified using SVD algorithm.

## 1. INTRODUCTION

A fundamental requirement in the job recruitment process is verification of an applicant's work history to ensure the prospective employee has the ability to competently perform the job they are applying for. Verifying provided work history can be a time-consuming and expensive process for employers. Often the verification process doesn't ensure that the provided information is valid or accurate. Additionally, contacting the current employer of a job applicant may comprise their relationship, which can be detrimental to all concerned. Although there are difficulties with verifying an applicant's work history, not verifying it can expose a company to significant risk, including decreasing a company's productivity if that person is employed for a position they are unfit to perform.

While centralised solutions can verify work history, these solutions rely on third parties, which do not necessarily eradicate falsified information. These solutions are also open to cyber-attacks, thus exposing private information, as well as having existing information modified by unauthorised parties. Additionally, third parties often sell data for marketing purposes. One example of a centralised solution was in the case of global information company Equifax experienced a cyber attack, resulting in millions of consumers having their information accessed. It is required to improve this situation and establish a system that can ensure that an individual's past and current work history can be securely, easily and readily verified. This is particularly the case for small and medium-size companies that often do not have dedicated resources to verify documents, or the financial capacity to outsource such activities to a third party.

## 2.LITERATURE REVIEW

By using Image processing we detect the Forgery scan certificate.

In [1], the authors have proposed that every institute has to maintain database for all its certificates and institute websites must support two main functionalities 1. Certificates download option by verifying student's credentials and 2. Certificate verification options

In [2], the authors have proposed that an image processing algorithm to combine feature point matching tools and adaptive segmentation to identify suspected irregular patterns detected by the adaptive non-overlapping and irregular blocks.

In [3], the authors have proposed that a forgery detection method that exploits subtle inconsistencies in the color of the illumination of images. To achieve this, we incorporate information from physics- and statistical-based illuminant estimators on image regions.

In [4], the authors have proposed that the digital image requires some pre-processing such as watermark embedding or signature generation at the time of creating the image. Moreover, there are millions of digital images in internet without digital signature or watermark. In such scenario active approach could not be used to find the authentication of the image. Unlike the watermark-based and signature-based methods; the passive technology does not need any digital signature generated or watermark embedded in advance.

In [5], the authors have proposed that Forgery detection method that depends on the discrete wavelet transform (DWT) as well as the discrete cosine transform (DCT) for feature reduction. The DCT is applied to the individual blocks obtained after dividing the DWTTed image. The blocks are then compared on the basis of correlation coefficients. A mask-based tampering method is also developed as part of the experiments in order to test the detection method.

In [6], the authors have proposed that the detection of image manipulation is very important because an image can be used as legal evidence, in forensics investigations, and in many other fields. The pixel-based image forgery detection aims to verify the authenticity of digital images without any prior knowledge of the original image. There are many ways for tampering an image such as splicing or copy-move, resampling an image (resize, rotate, stretch), addition and removal of any object from the image.

In [7], the authors have proposed that is very difficult for humans to identify visually whether the image was modified or not. There is a rapid increase in digitally manipulated forgeries in mainstream media and on the Internet

In [8], the authors have proposed that detection forgery in document based on distortion during the forgery creation process. The method used Recognition by Adaptive Subdivision of Transformation space (RAST). In this method, two images are corresponding collected and a matching score is calculated.

In [9], the authors have proposed that the image forgery detection techniques intend to confirm the credibility of computerized pictures with no prior information about the original image. There are numerous routes for altering a picture, for example, resampling, splicing, and copymove.

In [10], Digital forgeries may leave no visual clues that indicate tampering, they may alter the underlying statistics of an image. The set of image forensic tools can be roughly grouped into five categories: 1) pixel-based techniques; 2) format-based techniques 3) camera-based techniques 4) physically based 5) geometric-based techniques.

## 3. METHODOLOGY

This project has following modules

- **IMAGE ACQUISITION**
- **PREPROCESSING**
  - **IMAGE ENHANCEMENT**
  - **GAUSSIAN FILTER**
- **FEATURE EXTRACTION**
  - **GLCM**
- **CLASSIFICATION**
  - **SVD**

### IMAGE ACQUISITION:

The first stage of any vision system is the image acquisition stage. After the certificate image has been obtained, various methods of processing can be applied to the image to perform the many different vision tasks required today. However, if the image has not been acquired satisfactorily then the intended tasks may not be achievable, even with the aid of some form of image enhancement.

## PREPROCESSING

### IMAGE ENHANCEMENT

Image enhancement is the process of adjusting digital images so that the results are more suitable for display or further image analysis. For example, you can remove noise, sharpen, or brighten an image, making it easier to identify key features.

#### GAUSSIAN FILTER

The Gaussian smoothing operator is a 2-D convolution operator that uses a Gaussian function for calculating the transformation in each pixel. It is widely used to 'blur' images and also for removing noises and details from an image. The equations of Gaussian function in both 1D and 2D are given below:

$$G(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{x^2}{2\sigma^2}} \quad (1)$$

Here  $x$  is the distance from the origin in the horizontal axis,  $y$  is the distance from the origin in the vertical axis, and  $\sigma$  is the standard deviation of the Gaussian distribution. Gaussian smoothing uses 2-D distribution function and when calculating new values of a given pixel, convolution operator is being used. Convolution takes values of the neighboring pixels into account. The main element of a convolution is kernel which is a matrix of arbitrary size mostly a square matrix. When calculating the new value of the selected pixel, the convolution kernel is applied to it by its center pixel. Neighboring pixels are covered with the same kernel. Next, the sum of the product of the pixels is calculated in the image. The resulting sum is the new value of the selected pixel. Now, if the convolution is applied to each pixel in the image, we get a certain effect, which depends on the chosen convolution kernel. An example of the calculation of Gaussian Smoothing is given below.

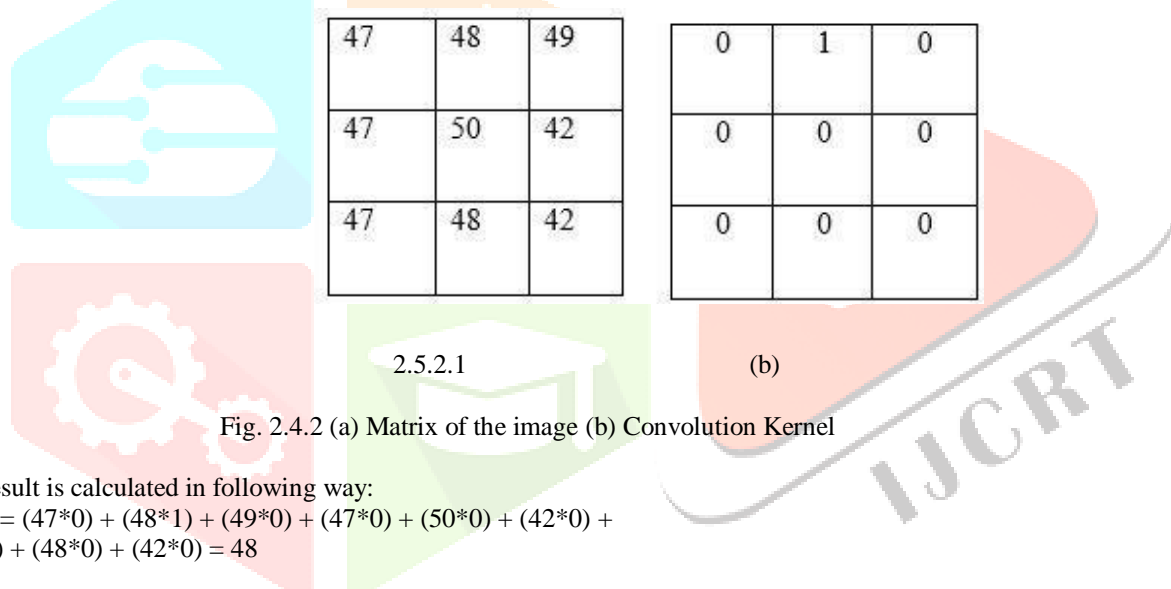


Fig. 2.4.2 (a) Matrix of the image (b) Convolution Kernel

The result is calculated in following way:

$$\text{Result} = (47*0) + (48*1) + (49*0) + (47*0) + (50*0) + (42*0) + (47*0) + (48*0) + (42*0) = 48$$

The result of applying kernel to a pixel with center value of 50 will be:

47	48	49
47	48	42
47	48	42

Fig. 2.4.2 Gaussian smoothing filtered matrix of the image

### IMAGE CROPPING

Cropping an image extracts a rectangular region of interest from the original image. This focuses the viewer's attention on a specific portion of the image and discards areas of the image that contain less useful information. Using image cropping in conjunction with image magnification allows you to zoom in on a specific portion of the image. This section describes how to exactly define the portion of the image you wish to extract to create a cropped image. For information on how to magnify a cropped image, see Resizing Images. Image cropping requires a pair of  $(x, y)$  coordinates that define the corners of the new, cropped image. The following example extracts the African continent from an image of the world. Complete the following steps for a detailed description of the process

## FEATURE EXTRACTION

### GRAY LEVEL COOCCURANCE MATRIX – GLCM

This method was first proposed by Haralick in 1973 and still is one of the most popular means of texture analysis. The key concept of this method is generating features based on gray level co-occurrence matrices (GLCM). The matrices are designed to measure the spatial relationships between pixels. The method is based on the belief that texture information is contained in such relationships. Co-occurrence features are obtained from a gray-level cooccurrence matrix.

GLCM is a second-order statistical texture analysis method. It examines the spatial relationship among pixels and defines how frequently a combination of pixels are present in an image in a given direction  $\theta$  and distance  $d$ . Each image is quantized into 16 gray levels (0–15) and 4 GLCMs ( $M$ ) each for  $\theta = 0, 45, 90,$  and  $135$  degrees with  $d = 1$  are obtained. From each GLCM, five features (Eq. 13.30–13.34) are extracted. Thus, there are 20 features for each image. Each feature is normalized to range between 0 to 1 before passing to the classifiers, and each classifier receives the same set of features.

$$\text{Energy} = \sqrt{\sum_{i,j=0}^{G-1} (M(i,j))^2}$$

$$\text{Homogeneity} = \sum_{i,j=0}^{G-1} \frac{M(i,j)}{1+(i+j)^2}$$

$$\text{Correlation} = \sum_{i,j=0}^{G-1} M(i,j) \left[ \frac{(i-\bar{i})(j-\bar{j})}{\sqrt{(\sigma_i^2)(\sigma_j^2)}} \right]$$

$$\text{Energy} = \sum_{i,j=0}^{G-1} M(i,j) \times (i,j)^2 \triangle$$

$$\text{Entropy} = -\sum_{b=0}^{G-1} P(b) \log_2 \{P(b)\}$$

To create a GLCM, use the graycomatrix function. The graycomatrix function creates a gray-level co-occurrence matrix (GLCM) by calculating how often a pixel with the intensity (gray-level) value  $i$  occurs in a specific spatial relationship to a pixel with the value  $j$ . By default, the spatial relationship is defined as the pixel of interest and the pixel to its immediate right (horizontally adjacent), but you can specify other spatial relationships between the two pixels. Each element  $(i,j)$  in the resultant glcm is simply the sum of the number of times that the pixel with value  $i$  occurred in the specified spatial relationship to a pixel with value  $j$  in the input image.

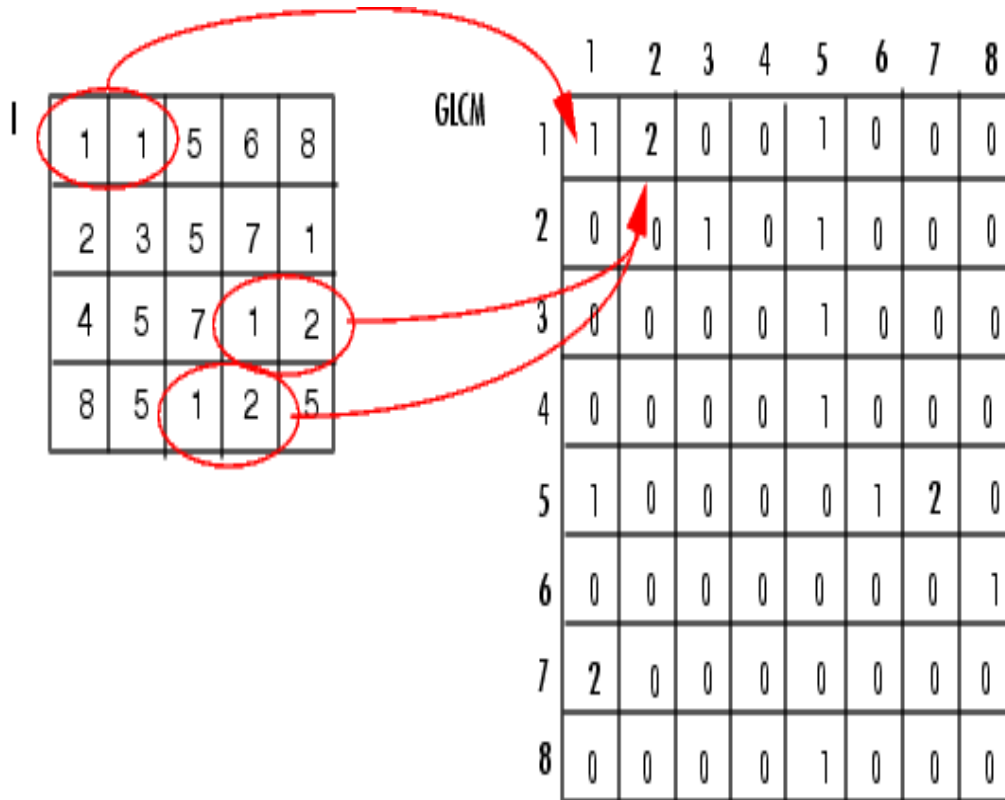
Because the processing required to calculate a GLCM for the full dynamic range of an image is prohibitive, graycomatrix scales the input image. By default, graycomatrix uses scaling to reduce the number of intensity values in gray scale image from 256 to eight. The number of gray levels determines the size of the GLCM. To control the number of gray levels in the GLCM and the scaling of intensity values, using the Num Levels and the Gray Limits parameters of the graycomatrix function. See the graycomatrix reference page for more information.

The gray-level co-occurrence matrix can reveal certain properties about the spatial distribution of the gray levels in the texture image. For example, if most of the entries in the GLCM are concentrated along the diagonal, the texture is coarse with respect to the specified offset. You can also derive several statistical measures from the GLCM. See [Deriving Statistics from a GLCM](#) for more information.

To illustrate, the following figure shows how graycomatrix calculates the first three values in a GLCM. In the output GLCM, element (1,1) contains the value 1 because there is only one instance in the input image where two horizontally adjacent pixels have the values 1 and 1, respectively.  $glcm(1,2)$  contains the value 2 because there are two instances where two horizontally adjacent pixels have the values 1 and 2. Element (1,3) in the GLCM has the value 0 because there are no instances of two horizontally adjacent pixels with the values 1 and 3. graycomatrix continues processing the input image, scanning the image for other pixel pairs  $(i,j)$  and recording the sums in the corresponding elements of the GLCM.

### SINGULAR VALUE DECOMPOSITION - SVD ALGORITHM

SVD is known under many different names. In the early days, as the above passage implies, it was called, "factor analysis." Other terms include principal component (PC) decomposition and empirical orthogonal function (EOF) analysis. All these are mathematically equivalent, although the way they are treated in the literature is often quite different.



Today, singular value decomposition has spread through many branches of science, in particular psychology and sociology, climate and atmospheric science, and astronomy. It is also extremely useful in machine learning and in both descriptive and predictive statistics.

Singular value decomposition is a method of decomposing a matrix into three other matrices:

$$A = USV^T \quad (1)$$

Where:

- $A$  is an  $m \times n$  matrix
- $U$  is an  $m \times n$  orthogonal matrix
- $S$  is an  $n \times n$  diagonal matrix
- $V$  is an  $n \times n$  orthogonal matrix

The reason why the last matrix is transposed will become clear later on in the exposition. Also, the term, “orthogonal,” will be defined (in case your algebra has become a little rusty) and the reason why the two outside matrices have this property made clear.

For the moment, we will assume that  $m \geq n$ . What happens when this isn't true is quite interesting and is one of the keys, in my opinion, to understanding singular value decomposition.

This is already becoming quite complicated so I will rewrite Equation (1) using summation notation. This is my go-to method of proceeding whenever I am having trouble with a matrix equation. In this case, while it doesn't make anything simpler, it does make everything absolutely explicit:

$$a_{ij} = \sum_{k=1}^n u_{ik} s_k v_{jk}$$

Note how we've collapsed the diagonal matrix,  $S$ , into a vector, thus simplifying the expression into a single summation. The variables,  $\{s_i\}$ , are called singular values and are normally arranged from largest to smallest:

$$s_{i+1} \leq s_i$$

The columns of  $U$  are called *left singular vectors*, while those of  $V$  are called *right singular vectors*.

We know that  $U$  and  $V$  are orthogonal, that is:

$$U^T U = V V^T = I$$

Where  $I$  is the *identity matrix*. Only the diagonals of the identity matrix are 1, with all other values being 0. Note that because  $U$  is not square we cannot say that  $U \text{Transpose}(U) = I$ , so  $U$  is only orthogonal in one direction.

Using the orthogonality property, we can rearrange (1) into the following pair of eigenvalue equations:

$$A A^T U = U S^2 \quad (2)$$

$$A^T A V = V S^2 \quad (3)$$

## Numerical procedure

Since  $\text{Transpose}(A)A$  is the same size or smaller than  $A \text{Transpose}(A)$ , a typical procedure is to plug Equation (3) into an eigenvalue calculator to find  $V$  and  $S^2$  and then find  $U$  by projecting  $A$  onto  $V$ :

$$U = AVS^{-1} \quad (4)$$

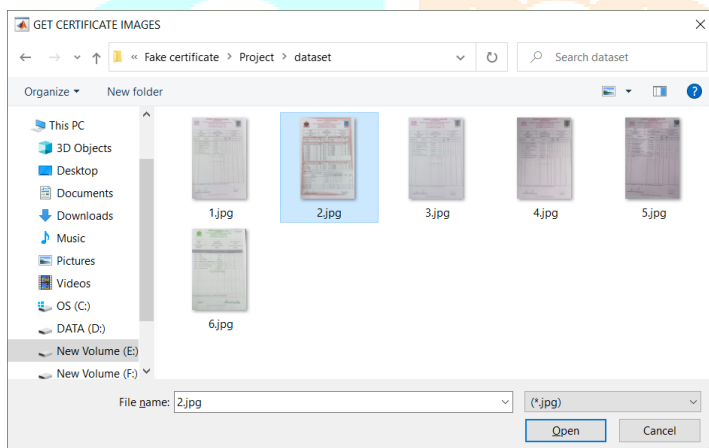
Note that the method is completely symmetric;  $U$  and  $V$  change places when  $A$  is transposed:

$$A^T = VSU^T$$

Thus, if  $m < n$ , we can transpose  $A$ , perform the decomposition, then swap the roles of  $U$  and  $V$ .

In this case,  $U$  will be an  $m \times m$  square matrix since there can be at most  $m$  non-zero singular values, while  $V$  will be an  $n \times m$  matrix.

## 4. RESULTS



### FAKE CERTIFICATE IDENTIFICATION USING IMAGE PROCESSING

**MENU**

- Input Certificate
- Image Enhancement
- Gaussian Filter

**ROI**

- Logo
- Photo
- Principal Sign
- Controller Sign

**Feature Extraction**

- GLCM


1	2
1	
2	
3	

**Classification**

- SVD

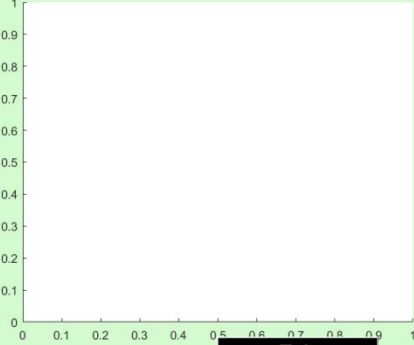
**PREDICATED OUTPUT**

Input Image

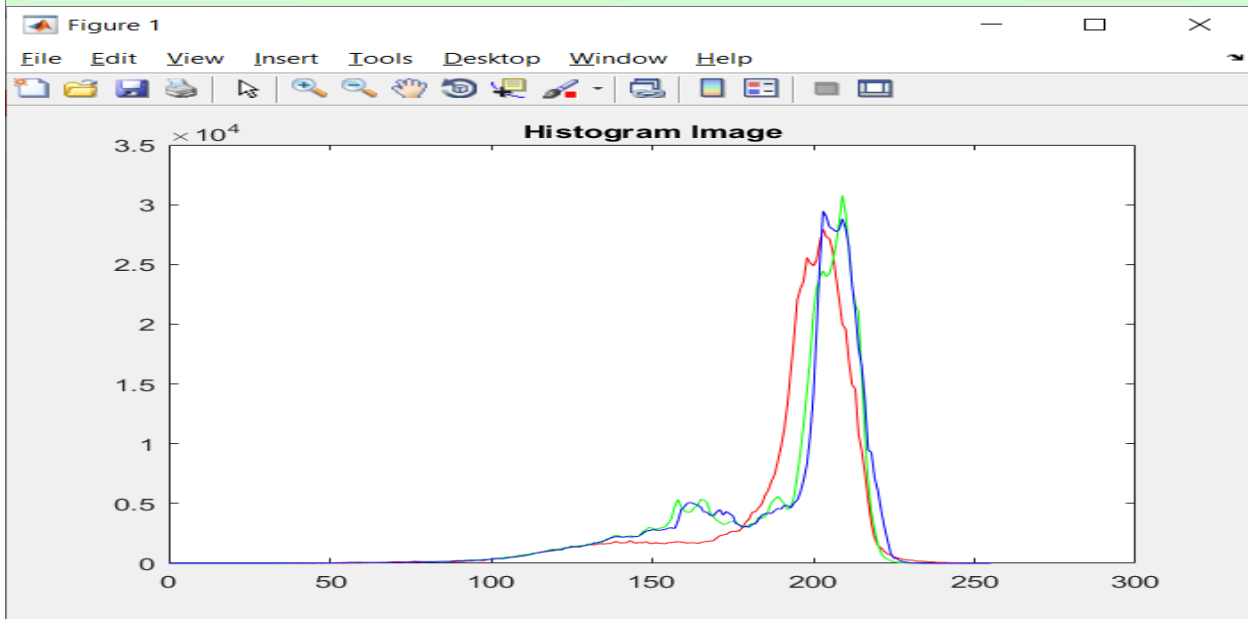


**Whether the certificate is Fake or Real?**

**OUTPUT**



Exit



### FAKE CERTIFICATE IDENTIFICATION USING IMAGE PROCESSING

**MENU**

- Input Certificate
- Image Enhancement
- Gaussian Filter

**ROI**

- Logo
- Photo
- Principal Sign
- Controller Sign

**Feature Extraction**

- GLCM


1	2
1	
2	
3	

**Classification**

- SVD


**PREDICATED OUTPUT**

Input Image



**Whether the certificate is Fake or Real?**

**OUTPUT**



Exit



### FAKE CERTIFICATE IDENTIFICATION USING IMAGE PROCESSING

**MENU**

- Input Certificate
- Image Enhancement
- Gaussian Filter

**ROI**

- Logo
- Photo
- Principal Sign
- Controller Sign

**Feature Extraction**

- GLCM


1	2
2	
3	

**Classification**

- SVD

**PREDICATED OUTPUT**


Input Image



Whether the certificate is Fake or Real?

**OUTPUT**

Logo



**Exit**

### FAKE CERTIFICATE IDENTIFICATION USING IMAGE PROCESSING

**MENU**

- Input Certificate
- Image Enhancement
- Gaussian Filter

**ROI**

- Logo
- Photo
- Principal Sign
- Controller Sign

**Feature Extraction**

- GLCM


1	2
2	
3	

**Classification**

- SVD

**PREDICATED OUTPUT**


Input Image



Whether the certificate is Fake or Real?

**OUTPUT**

Photo



**Exit**

### FAKE CERTIFICATE IDENTIFICATION USING IMAGE PROCESSING

**MENU**

- Input Certificate
- Image Enhancement
- Gaussian Filter

**ROI**

- Logo
- Photo
- Principal Sign
- Controller Sign

**Feature Extraction**

- GLCM


1	2
2	
3	

**Classification**

- SVD

**PREDICATED OUTPUT**

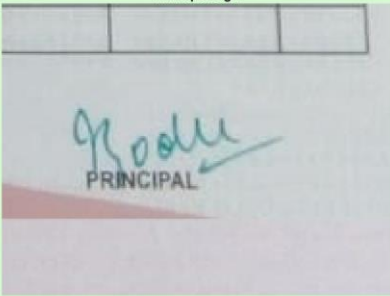
Input Image



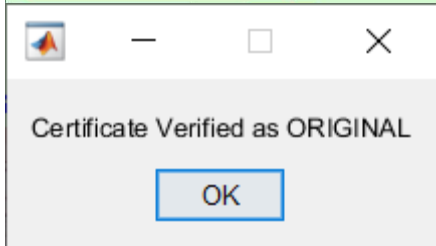
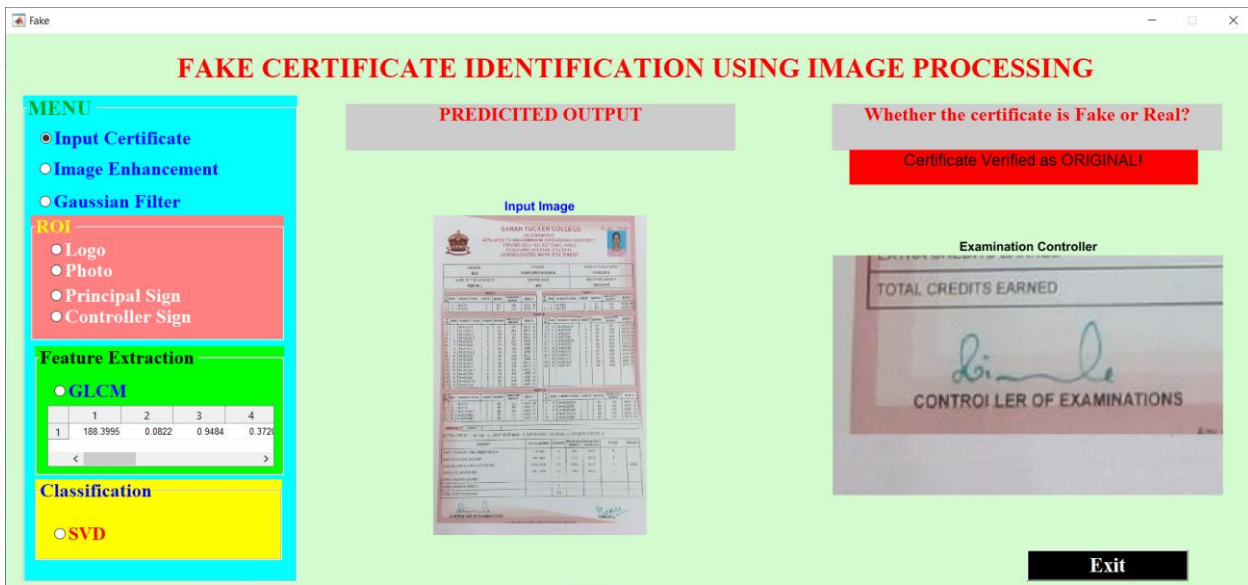
Whether the certificate is Fake or Real?

**OUTPUT**

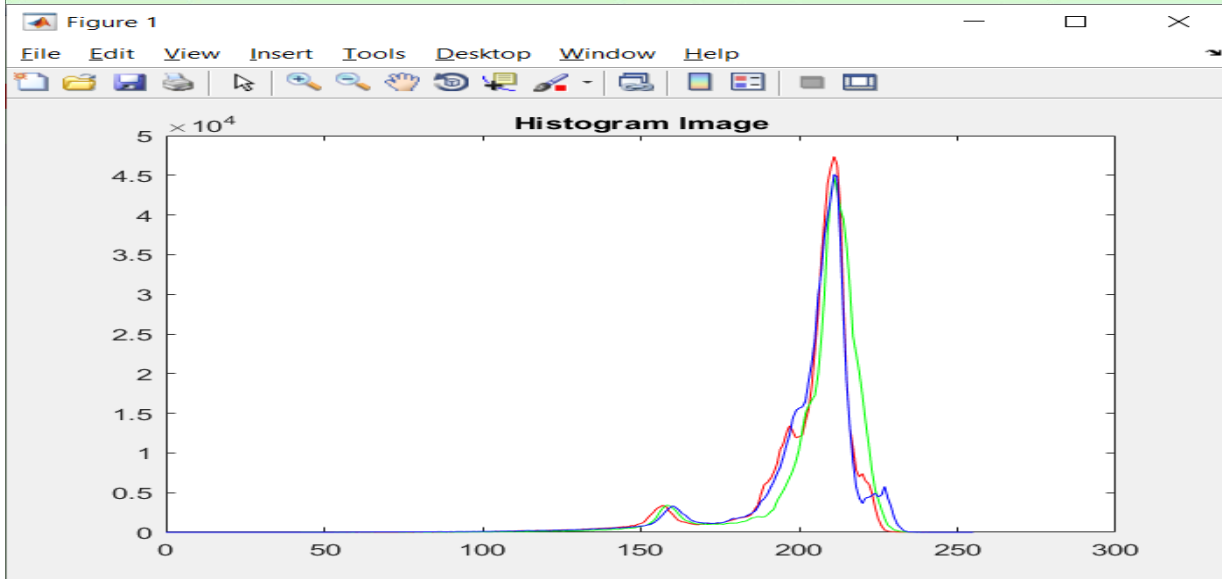
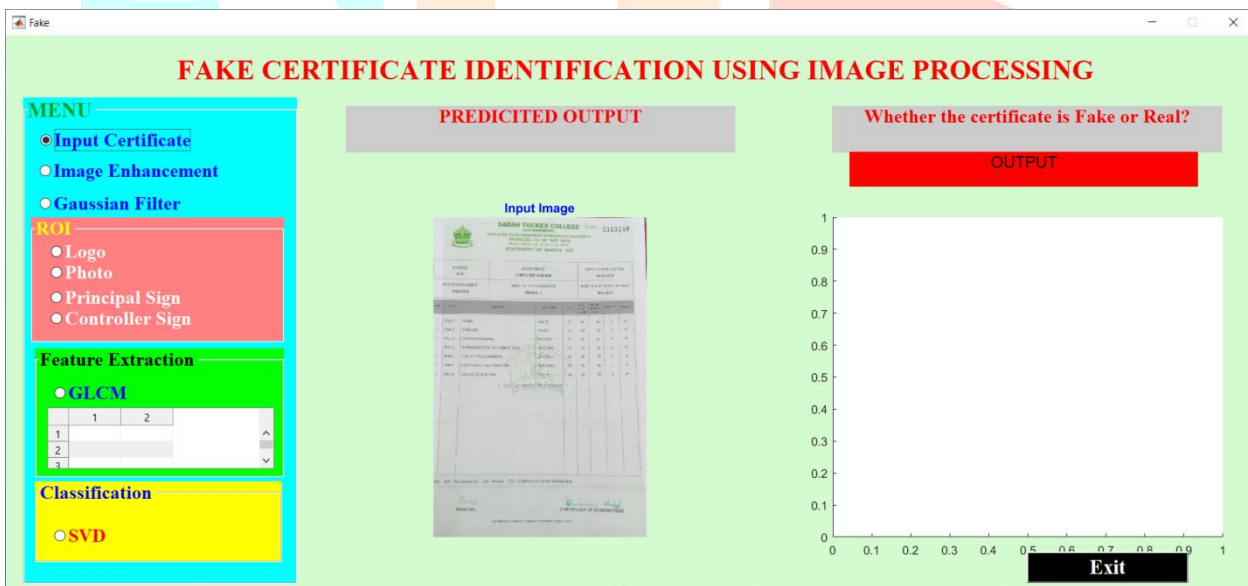
Principal Sign

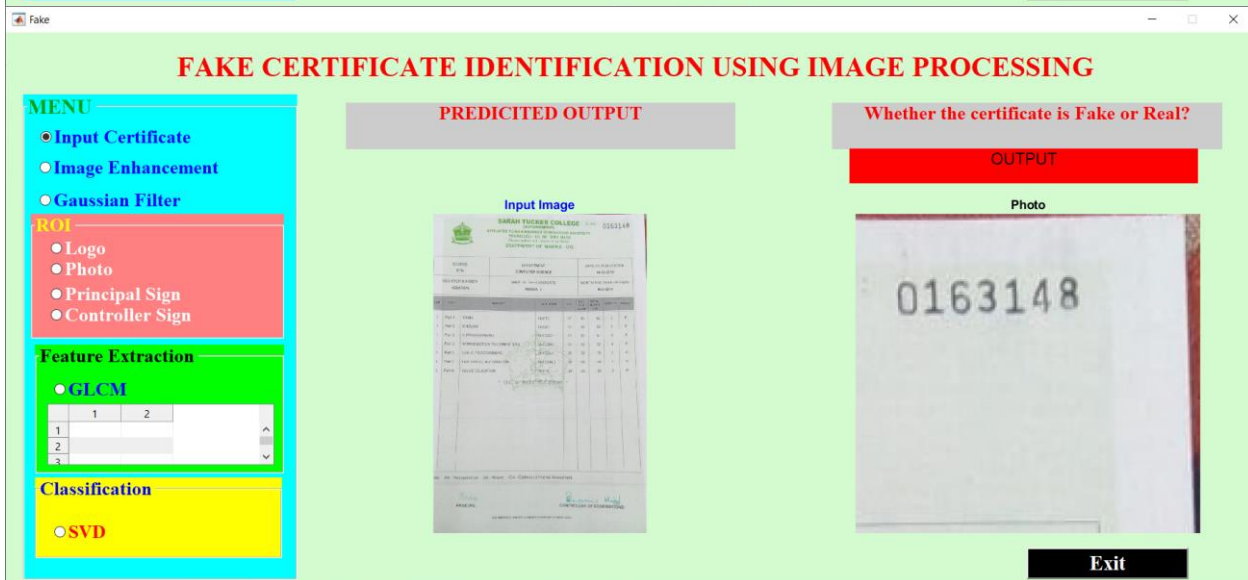
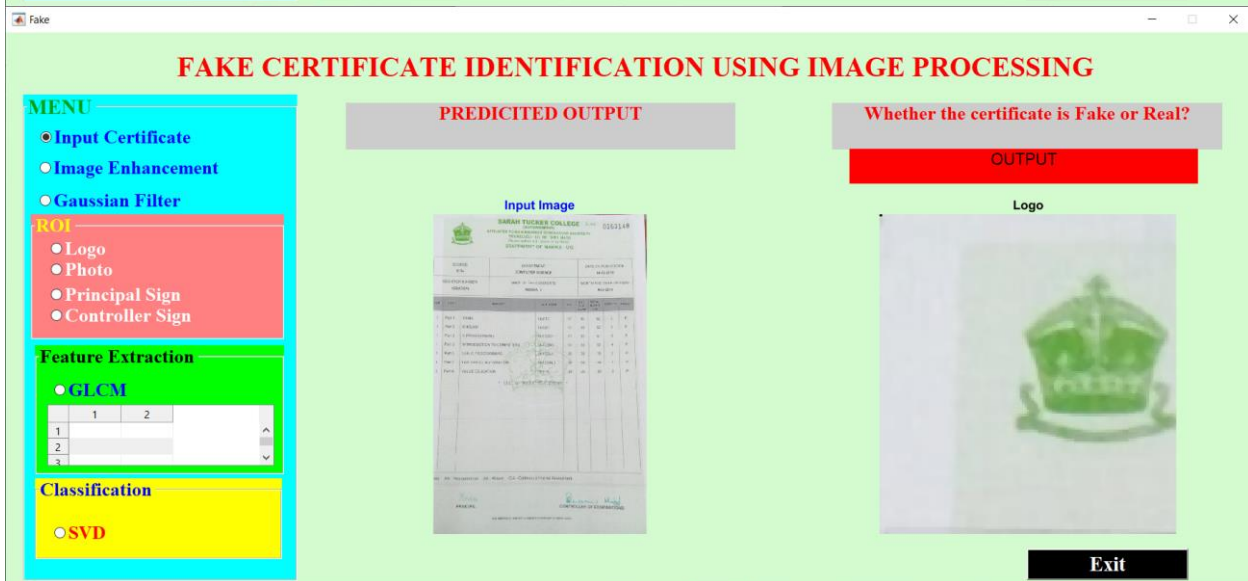
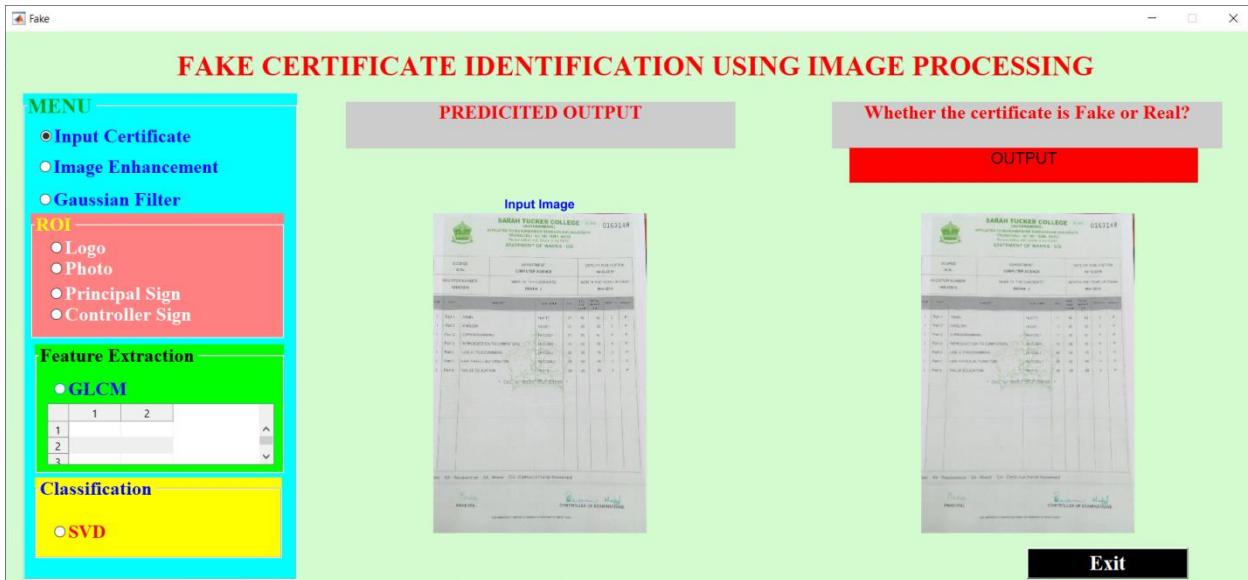


**Exit**



FAKECERTIFICATE







### 5. CONCLUSION

The proposed platform takes the advantage of the security in order to create a globally trusted higher education credit and grading system. As a proof of concept, we presented a prototype implementation of the system platform which is based on the open-source Ark platform. The proposed system platform addresses a globally unified viewpoint for students and organizations. Students benefit from a single and transparent view of their completed courses, while have access to up to date data regardless of a student’s educational origins. Other beneficiaries of the proposed system are potential employers, who can directly validate the information provided by students. The proposed solution is based on the distributed P2P network system. It transfers the higher education grading system from the current real-world physical records or traditional digital ones (e.g. databases) to an efficient, simplified, ubiquitous version, based on blockchain technology. It is anticipated that such a system could potentially evolve into a unified, simplified and globally ubiquitous higher education credit and grading system.

## 6. REFERENCE

- [1] Dr. Manohar Koli “Electronic Certificates to Avoid Fake and Tampered Academic Documents”, Department of Studies in Computer Science, Karnatak University Dharwad, Karnataka, India.
- [2]”Detecting Forgedscan of Education certificates Using a New Feature set matching Algorithm”-kethepali malikarajuna,pushan Kumar Dutta,N.Sateesh kedarnath kumar.
- [3] Mankar, S.K., and Gurjar, A.A.: ‘Image Forgery Types and Their Detection: A Review’, International Journal of Advanced Research in Computer Science and Software Engineering, 2015.
- [4] Shivakumar, B., and Baboo, L.D.S.S.: ‘Detecting copy-move forgery in digital images: a survey and analysis of current methods’, Global Journal of Computer Science and Technology, 2010 .
- [5] Hayat, K., and Qazi, T.: ‘Forgery detection in digital images via discrete wavelet and discrete cosine transforms’, Computers & Electrical Engineering, 2017.
- [6]“Pixel-Based Image Forgery Detection” –Mohd Dilshad Ansari, S.P.Ghrera & Vipin Tyagi.
- [7] J. Wang, G. Liu, Z. Zhang, Y. Dai, and Z. Wang, “Fast and robust forensics for image region-duplication forgery,” Acta Automatica Sinica, Vol. 35, no. 12, pp. 1488-95, Dec. 2009.
- [8] Ahmed, A.G.H., and Shafait, F.: ‘Forgery detection based on intrinsic document contents’, in Editor (Ed.)^(Eds.): ‘Book Forgery detection based on intrinsic document contents’ (IEEE, 2014, edn.), pp. 252-256.
- [9] “An Evaluation of Digital Image Forgery Detection Approaches” Abhishek Kashyap, Rajesh Singh Parmar, Megha Agarwal, Hariom Gupta Department of Electronics and Communication Engineering, Jaypee Institute of Information Technology.
- [10] Hany Farid: “Image Forgery Detection”, IEEE Signal Processing Magazine, pp. 16-25, March 2009.

