International Journal of Advances in Engineering and Management (IJAEM)
Volume 5, Issue 3 March 2023, pp: 660-666 www.ijaem.net ISSN: 2395-5252

Literature Review on Certificate Forgery Detection

Dr. Chayadevi M L*, Shreyas K +, Sruthi S +, Sujay L Gowda +

*HOD, Dept. of Computer Science Engineering,BNM Institute of Technology, Bengaluru, India
+Student, Dept. of Computer Science Engineering, BNM Institute of Technology, Bengaluru, India

Date of Submission: 01-03-2023 Date of Acceptance: 10-03-2023

ABSTRACT—In protecting the society from these fake certificate scams, digital system can play very crucial role. Detection of forged scan certificates which are used during college admissions are done using scan copies from 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 report describes and compares the various ways of forgery detection. The first one includes four modules such as pre-processing, crop image, feature extraction and forgery detection. In the preprocessing module, the input image is enhanced. Second module is crop image which is crop the college logo, photo, student sign and examination controller sign and label the items respectively. 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. Description of usage of Residual Network algorithm along with its advantages is described. Finally, the importance of VGG16 with Convolutional Neural Networks is explained in detail.

Index Terms—Certificate forgery detection, features, Image forgery detection, Image processing, VGG-16, Convolutional Neural Networks.

I. INTRODUCTION

Detection of forge scan certificates which are used during college admissions are done using scan copies from other gen-uine 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 de-tecting 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. In our proposed study we have applied 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 and this process is carried out using the adaptive over-segmentation algorithm. The extraction of the feature points is performed by identification of the matching between each block and its features. The feature points are gradually replaced by using the super pixels in the proposed Forgery Region Extraction algorithm and then combine the adjacent blocks having similar color features to find out the merged regions for identification of the detected forgery regions.

Further, a comprehensive study about the

International Journal of Advances in Engineering and Management (IJAEM)
Volume 5, Issue 3 March 2023, pp: 660-666 www.ijaem.net ISSN: 2395-5252

usage of ResNet and VGG16 algorithm along with Convolutional Neural Networks is provided. Its advantages over the existing systems are also mentioned. Authentication plays a very important role in today's worldly business especially with respect to experience and skill verification. Hence it is necessary for us to design a system which will be able to authenticate a person's realness.

II. MOTIVATION

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. While centralized 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 unauthorized parties. Additionally, third parties often sell data for marketing purposes. 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.

III. OVERVIEW

A literature review is a comprehensive summary of previous research on a topic. The literature review surveys scholarly articles, books, and other sources relevant to a particular area of research.

The review should enumerate, describe, summarize, objec-tively evaluate and clarify this previous research. It should give a theoretical base for the research and help you (the author) determine the nature of your research. The literature review acknowledges the work of previous researchers, and in so doing, assures the reader that your work has been well conceived. It is assumed that by mentioning a previous work in the field of study, that the author has read, evaluated, and assimilated that work into the work at hand.

A literature review creates a "landscape" for the reader, giving her or him a full understanding of the developments in the field. This landscape informs the reader that the author has indeed assimilated all (or the vast majority of) previous, significant works in the field into her or his research.

A certificate determines whether a person is skilled in that particular field or not. Forging it can lead the wrong person getting the job or the task. If he/she does not carry out the task properly, then the work will be incomplete. Checking if a certificate is authentic or forged is very important. Finding it efficiently and quickly is one of the objectives of this project.

IV. SURVEYFORPRE-PROCESSINGOFDATA

[1] A Watermarking Algorithm for Certificate Forgery Prevention (2014): A Certificate forgery prevention system based on digital watermarking technology can be split into two parts: certificate production and certificate forgery detection. When making a certificate, a watermark solely standing for the owner is embed into his digital photo using a special embedding algorithm, and the digital photo is printed out to be a certificate afterwards. When conducting forgery detection, the certificate is scanned by a scanner, and forgery detection is accomplished by extracting the watermark through a corresponding extracting algorithm. The print-and-scan process is the major challenge to designing a watermarking algorithm.

[2] Local Binary Patterns for Document Forgery Detection (November 2017): Document forgery is an increasing problem for both the public administration and private companies. It represents substantial losses in time and economical resources. Classical solutions to this problem such as watermarks or other integrated security patterns cannot be applied in general for any unknown incoming document due to the large variability on types of documents. In that scenario it is important to resort to forensic techniques to seek and analyze inconsistencies on the intrinsic features of the document image. In this paper description of a classification-based approach for forgery detection is given. This method was able to detect 7.38 percent of the forged patches on the test set combining the information of the 5 nearest neighbors.

[3] A Review on Copy-Move Image Forgery Detection Techniques (April 2021): Bin Yang et al. presented a feature-based copy-move forgery detection (CMFD) method . A modified Scale Invariant Feature Transform (SIFT) detector is used to detect key-points. A key-point distribution strategy was created for spreading the key-points across the image. Ultimately, the

International Journal of Advances in Engineering and Management (IJAEM)
Volume 5, Issue 3 March 2023, pp: 660-666 www.ijaem.net ISSN: 2395-5252

improved SIFT descriptor identified the key-points for copy-move forgery detection. It presents detailed experimental findings to validate the effectiveness.

Chun-Su Park Joon Yeon Choeh introduced a quick method that able to detect forgery with multiple geometric transfor-mations such as region rotation, resizing, deformation, and reflection . SIFT is used to extract the key-points and their descriptors to detect copy-move forgery. The suggested CMFD method has a solid theoretical background, which has better performance than the current SIFT-based algorithms. This method has good processing time.

Mohamed Abdel-Basset et al. suggested a technique that could detect the exploitation of this kind and identify the duplicated areas . SIFT is based on that strategy. It is a well-known robust technique capable of detecting and matching features that belong to duplicate regions. These matched features are placed under the umbrella of a 2-level clustering strategy to ensure that features are later used to help in the geometric transformation of the duplicate areas belonging to particular clusters representing the included regions in the image. common.

[4] A Comparative Study among Handwritten Signature Verification Methods Using Machine Learning Techniques (October 2022): Signature verification is naturally formulated as a machine-learning task. Since handwritten signatures are widely used in legal documents and financial transactions, it is important for researchers to select an efficient machine-learning technique for verifying these signatures and to avoid forgeries that may cause many losses to customers. The general steps of the verification system and a list of the most considerable datasets available in online and offline fields is also presented. Different feature extraction and classification techniques such as global features, local features, statistical and structural features, histogram of gradient, convolutional neural networks, support vector machines, k-nearest neighbor, etc. have been compared in terms of advantages and limitations.

[5] A Dataset for Forgery Detection and Spotting in Document Images (September 2017): In the last decades, the explosion of the volume of digital document images, and the development of consumer tools to modify these images has led to a huge increase on reported fraudulent document cases. This situation has promoted the development of automatic methods for both preventing forgeries in modified documents and detecting them. However, document forensics is a sensitive topic. Data is usually either private or unlabeled and most of the reported works are commonly evaluated on datasets with a restricted access. This paper, presents a new public dataset made of a corpus of 477 corrupted payslips in which near 6000 characters are forged.

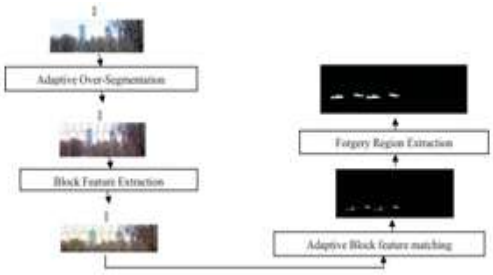

Fig.1.Proposed Methodology

This algorithm is found to be least cumbersome.

V. SURVEY FOR FEATURE EXTRACTION

[1] A Watermarking Algorithm for Certificate Forgery Pre-vention (2014): This algorithm is based on Fourier-Mellin Transformation (FMT). A watermarking algorithm for cer-tificate forgery prevention is proposed by utilizing zero-watermarking technology.

The methodology for image watermarking based on Fourier-Mellin Transform is given is below fig 2 It splits an image into non-overlapping blocks, and conducts singular value decomposition (SVD) on each block. The zero-watermark sequence is produced by judging the parity of the highest digit from singular values' average in each singular value matrix.

[2] Local Binary Patterns for Document Forgery Detection (November 2017): The created dataset used is as follows. The full set of documents contains invoices of multiple providers, contracts, shopping receipts, letters, and bank receipts. Half of the documents were captured using a common mobile phone camera, and the other half scanned without particular resolution restrictions.

International Journal of Advances in Engineering and Management (IJAEM)
Volume 5, Issue 3 March 2023, pp: 660-666 www.ijaem.net ISSN: 2395-5252

Fig. 2. Methodology for Image watermarking

Documents are mostly machine-printed and written in English and French. Forgeries were done on 200 documents. In total, the dataset contains 481 forgery instances, divided in the following categories: 272 CPI, 86 CPO, 69 IMI, and 54 CUT. Set of 1,000 genuine documents without modifications to evaluate the detection of false positives. Dataset was split into two image subsets for training and test by randomly selecting the 70-30 percent of the images, respectively. SVM classifier was used for experiments.

[5] A Dataset for Forgery Detection and Spotting in Doc-ument Images (September 2017): Payslips are one of the most forged types of documents. Unfortunately, they are also very confidential and contain elements of privacy like names, addresses, wages, or social security numbers.

This information must not be published or disseminated. One possible solution to avoid this problem was to erase or blur sensitive data. This solution could alter the dataset. In this article, a new set of digitized documents was introduced representing payslips. genrally there is a ground-truth XML file. Here, only two characters have been modified in the genuine document. people working on the domain of fraud detection and word spotting to have an available common and public dataset. To avoid questions about privacy policy that may be applied to these kind of sensitive documents, we choose to generate some genuine payslips with fake data. Nevertheless, these documents reflect real ones and preserves all the intrinsic and legal features of a payslip. Then, to obtain real fraudulent documents, workshops were organised to invite people committing forgeries on this dataset. To stick to reality, one-day forgers altered the documents according different scenarios. A ground-truth is then automatically extracted and stored in a XML File.

Moreover, as these data are also the one which are mostly forged, quality and relevance of dataset would be impacted. Considering this, synthetic payslips containing artificial data were generated. A common document layout was used for the same and company information, employee information and wage information were generated artificially.

[3] A Review on Copy-Move Image Forgery Detection Techniques (April 2021): Bin Yang et al. presented a feature-based copy-move forgery detection (CMFD) method . A modified Scale Invariant Feature Transform (SIFT) detector is used to detect key-points. A key-point distribution strategy was created for spreading the key-points across the image. Ultimately, the improved SIFT descriptor identified the key-points for copy-move forgery detection. It presents detailed experimental findings to validate the effectiveness.

Chun-Su Park Joon Yeon Choeh introduced a quick method that able to detect forgery with multiple geometric transfor-mations such as region rotation, resizing, deformation, and reflection . SIFT is used to extract the key-points and their descriptors to detect copy-move forgery. The suggested CMFD method has a solid theoretical background, which has better performance than the current SIFT-based algorithms. This method has good processing time.

Mohamed Abdel-Basset et al. suggested a technique that could detect the exploitation of this kind and identify the duplicated areas . SIFT is based on that strategy. It is a well-known robust technique capable of detecting and matching features that belong to duplicate regions. These matched features are placed under the umbrella of a 2-level clustering strategy to ensure that features are later used to help in the geometric transformation of the duplicate areas belonging to particular clusters representing the included regions in the image. common.

[9]Markov features based DTCWS algorithm for online image forgery detection using ensemble classifier in the pan-demic(2021): DTCWS algorithm was generally introduced or used during the COVID-19 times. This generally helps in extraction of domain wise Markov features which includes coefficient wise Markov features with spatial domain, block-wise Markov features with Discrete Cosine Transform (DCT) domain,

International Journal of Advances in Engineering and Management (IJAEM)
Volume 5, Issue 3 March 2023, pp: 660-666 www.ijaem.net ISSN: 2395-5252

and multi-resolution wise Markov features with Dis-crete Wavelet Transform (DWT) domain.
There are three steps. As a result, there is an evident increase in the dimensionality of the features. To reduce the
dimensionality, PCA is used. Finally higly correlated features are fed to the ensemble classifier for distinguishing between the authentic and fake images.

[7] Detecting Forged Scan of Educational Certificates Using GLCM And SVD Algorithm(February 2022): Usage of GLCM and SVD for forgery detection GLCM (Gray-Level Co-Occurrence Matric) for feature extraction Features used are Stamp, Photo, Signature.

[6] Detecting Forged Scan of Educational Certificates Using a New Feature Set Matching Algorithm(December

2022): In this paper, multiple algorithms are analysed and compared. It is found that feature set extraction is better than other algorithms studied. Feature set extraction is also the least cumbersome of the other algorithms.

[3] A Review on Copy-Move Image Forgery Detection Techniques (April 2021): Bin Yang et al. presented a feature- based copy-move forgery detection (CMFD) method . A modified Scale Invariant Feature Transform (SIFT) detector is used to detect key-points. A key-point distribution strategy was created for spreading the key-points across the image. Ultimately, the improved SIFT descriptor identified the key-points for copy-move forgery detection. It presents detailed experimental findings to validate the effectiveness.

Chun-Su Park Joon Yeon Choeh introduced a quick method that able to detect forgery with multiple geometric transfor-mations such as region rotation, resizing, deformation, and reflection . SIFT is used to extract the key-points and their descriptors to detect copy-move forgery. The suggested CMFD method has a solid theoretical background, which has better performance than the current SIFT-based algorithms. This method has good processing time.

Mohamed Abdel-Basset et al. suggested a technique that could detect the exploitation of this kind and identify the duplicated areas . SIFT is based on that strategy. It is a well-known robust technique capable of detecting and matching features that belong to duplicate regions. These matched features are placed under the umbrella of a 2-level clustering strategy to ensure that features are later used to help in the geometric transformation of the duplicate areas belonging to particular clusters representing the included regions in the image. common.

[8] Certificate Fraud Detection Using Artificial Intelligence Technique (2021): This study collected data from the burary department, Unizik; and then developed an intelligent docu-ment verification algorithm and deployed as an expert system using Matlab. The Dataflow model is shown in the Fig 2. The flowchart presents the logical data flow of how each method is interrelated with the other, starting with the training dataset. The training dataset was used to learn the artificial intelligence technique of the features for the data collected with the intelligence model which serves as base for future classification. When new data is uploaded for verification using the OCR device (scanner), the data is processed using binarization for bi-level conversion, then data processing for

Fig.3.ProposedSystem

resizing. The features of the documents are revealed more for better extraction performance using segmentation process and then extracted using feature extraction technique which converts the image features into a compact feature vector and then feed to the artificial intelligence system for training and classification. The A.I is neural network which uses training algorithm to learn and classify features for accurate decisions. The system was tested and the result showed high rate of verification regression and MSE performance. The implication is that when deployed, it was able to recognize and verify results accurately.

[10]. A model of unforgeable digital certificate document system: Document forgery is classified into two following cat-egories.The categories are Type-1 forgery and Type-2 forgery. The Type-1 is when some part of the original document is altered to benefit someone who is not benefited by the original document. In this case the base substance, such as a paper sheet or plastic

International Journal of Advances in Engineering and Management (IJAEM)
Volume 5, Issue 3 March 2023, pp: 660-666 www.ijaem.net ISSN: 2395-5252

card, is legal and valid, but the information contained is forged. For example, a stolen photo credit card may be altered by replacing the photo picture on the card, the

expiration date on a drivers license could be altered to make it still valid after the original license expires, the grading scores on a student report card may be changed from C's and D's to all A's, etc.

Type-2 forgery is the case when both the base substance and the information data are false, but it is hard to tell if it is genuine or fake because the base substance and the style of the document look very authentic. Most passport forgeries belong to this category. Type-2 forgeries include attempts to create authentic looking, but false, information data.

This method of unforgeable digital certificate document system is based on the steganographic technique. A steganography technique involves hiding sensitive information within an ordinary, non-secret file or message, so that it will not be detected. The method specified is an easier as well as secure way of transferring either a certificate or document .

VI. SUMMARY OF LITERATURE SURVEY

From the above survey , we observed and researched many papers for forgery detection in certificates. Using the neural network model , the data can be pre-processed and checked for forgery . Of the many algorithms and models researched ,usage of ANN models were shown.

Convolution Neural Network (CNN) is a Deep Learning algorithm that can take in an input image, assign importance (learnable weights and biases) to various aspects/objects in the image, and be able to differentiate one from the other. The pre-processing required in a ConvNet is much lower as compared to other classification algorithms.

The below given Fig 2 is a sequence in CNN used for classifying handwritten digits from computerized or typed digits. This makes it easier to detect the type of digit as well as the method used for forgery detection in these cases.

The architecture of a ConvNet is analogous to that of the connectivity pattern of Neurons in the Human Brain and was inspired by the organization of the Visual Cortex. Individual neurons respond to stimuli only in a restricted region of the visual field known as the Receptive Field. A collection of such fields overlaps to cover the entire visual area.

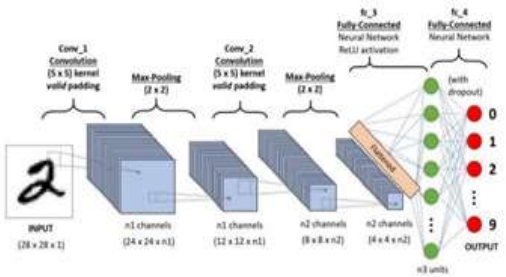

Fig.4.Classification of Hand written digits using a CNN sequnce

AI models for detecting many other anomalies or irregular-ities in the data can also be used to refine the data. VGG-16 is a convolution neural net (CNN) architecture . It is considered to be one of the excellent vision model architectures till date. The below given Fig 3 depicts the architecture of the VGG-16 that can be used for pre-training and training. The pre-trained network can readily classify objects like pencils,books and pens. The trained network can be used for custom classi-fication of images or objects as per the user's wish.

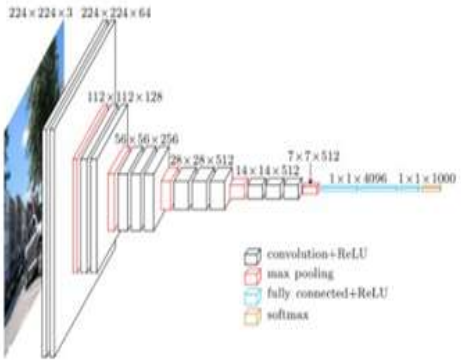

Fig.5.Architecture of VGG-16

Certificate authentication plays a very critical role in various applications such as verifying the bank documents, corporate important documents, land documents and job applications. Hence it is very important to create an interactive model where the model can recognize the real certificates and will be able to smartly detect whether it is real or forged.

To implement this, we need to use an efficient learning scheme such as Machine Learning (ML) or Convolutional Neural Networks (CNN). We create an AI Model that performs various types of forgery detection such as Simulated/Copied forgery (Histogram Processing), Free hand forgery (tessaractOCR), etc. to detect the authenticity of the certificate images.

International Journal of Advances in Engineering and Management (IJAEM)
Volume 5, Issue 3 March 2023, pp: 660-666 www.ijaem.net ISSN: 2395-5252

VII. CONCLUSION

Usage of these fake or forged certificates has become high in the recent years. The authenticity of a certificate should be checked to determine whether the awarded certificate is true or not. Similarly it goes the same to the documents also. The authenticity of the documents must be checked.

This can be implemented in many workplaces and educa-tional institutions because the right person has to get what he/she deserves.

Using these concepts and networks we are detecting whether the given certificate is authentic or not. As of now, VGG-16 is a better algorithm which is a part of CNN. 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.

ACKNOWLEDGMENT

We would like to thank our Guide Dr Chayadevi M L for helping us with their valuable suggestions to complete this survey and our College BNMIT in supporting us to continuing with this project.

REFERENCES

[1]. Fu, Y. G. (2014). A Print-Scan resilient image Wa-termarking Scheme based on Radon transform. In Applied Mechanics and Materials (Vol. 687, pp. 3812-3817). Trans Tech Publications Ltd.

[2]. Cruz, F., Sidere, N., Coustaty, M., d'Andecy, V. P., Ogier, J. M. (2017, November). Local binary patterns for document forgery detection. In 2017 14th IAPR International Conference on Document Analysis and Recognition (ICDAR) (Vol. 1, pp. 1223-1228). IEEE.

[3]. Khudhair, Z. N., Mohamed, F., Kadhim, K. A. (2021, April). A Review on Copy-Move Image Forgery Detection Techniques. In Journal of Physics: Conference Series (Vol. 1892, No. 1, p. 012010). IOP Publishing.

[4]. Hashim, Z., Ahmed, H. M., Alkhayyat, A. H. (2022). A Comparative Study among Handwritten Signature Verifica-tion Methods Using Machine Learning Techniques. Scientific Programming, 2022.

[5]. Sidere, N., Cruz, F., Coustaty, M., Ogier, J. M. (2017, September). A dataset for forgery detection and spotting in document images. In 2017 Seventh International Conference on Emerging Security Technologies (EST) (pp. 26-31). IEEE.

[6]. Dutta, Pushan. (2018). Detecting Forged Scan of Edu-cational Certificates Using a New Feature Set Matching Al-gorithm. International Journal of Tomography and Simulation.32.

[7]. Ranjan, S., Garhwal, P., Bhan, A., Arora, M., Mehra, A.(2018, May). Framework for image forgery detection and classification using machine learning. In 2018 2nd Interna-tional Conference on Trends in Electronics and Informatics (ICOEI) (pp. 1-9). IEEE.

[8]. [8] Isizoh, A. N., Anyi, D. O., Onyeyili, T. C., Ebih, U. J., Ejimofor, I. A. Certificate Fraud Detection Using Artificial Intelligence Technique. Journal homepage: www. ijrpr. com ISSN, 2582, 7421.

[9]. Mehta, R., Aggarwal, K., Koundal, D., Alhudhaif, A., Polat, K. (2021). Markov features based DTCWS algorithm for online image forgery detection using ensemble classifier in the pandemic. Expert Systems with Applications, 185, 115630.

[10]. Nozaki, K., Noda, H., Kawaguchi, E., Eason, R. (2005). A model of unforgeable digital certificate document system. Information Modelling and Knowledge Bases XVI, 2034210.