Signature Forgery Detection

Department of Engineering, Electronics and Telecommunication (ENTC)
Vishwakarma Institute of Technology, Pune, 411037, Maharashtra, India.

Dr.Swati Shilaskar Swati.shilaskar@vvit.edu
Kaushal Jadhav kaushal.jadhav21@vit.edu
Satyajeet Jadhav Satyajeet.jadhav21@vit.edu
Ketan Jain ketan.jain21@vit.edu

Abstract - In a rapidly digitizing world with increasing reliance on remote transactions, the security of handwritten signatures remains a critical concern. Signatures serve as a cornerstone for identity verification and authorization in legal documents, making the detection of signature forgeries a pressing issue. This research addresses this challenge by introducing a comprehensive system leveraging computer vision, deep learning, and neural networks. Key technologies such as OpenCV and TensorFlow are harnessed to create an accurate and efficient signature forgery detection system. The study classifies signature forgeries into three categories: blind, trace-over, and skilled, emphasizing the significance of correctly identifying forgeries to ensure security. A review of related literature underscores the evolving landscape of document forgery detection, emphasizing the growing role of machine learning and deep learning approaches. The research methodology encompasses dataset curation, preprocessing, feature extraction, model training, and system design. The dataset, sourced from the ICDAR 2009 Signature Verification Competition, is augmented to ensure diversity and comprehensiveness. Feature extraction captures essential signature attributes, and a multilayer perceptron (MLP) neural network distinguishes genuine from forged signatures. The system's performance is assessed through rigorous metrics.

Keywords-Signature forgery detection, document security, OpenCV, neural network, TensorFlow

I. INTRODUCTION

In an age of growing dependence on digital transactions and electronic authorizing processes, it can hardly be emphasized enough how important it is to protect signatures. For years, signatures have been considered a pillar of authenticating identity and approving official documentation. The need for reliable signature forgery detection systems has increased significantly, nevertheless, due to the development of sophisticated forgery techniques and the simplicity of digital modification. The authors have developed a system for signature forgery detection, utilizing technologies such as OpenCV, TensorFlow, and neural networks to create a comprehensive and highly accurate system. Signature forgery detection is a critical aspect of document security, fraud prevention, and ensuring the authenticity of digital transactions, making it a pertinent area of study and technological innovation. Approach combines the power of computer vision, deep learning, and neural networks to address the inherent challenges posed by signature forgery. OpenCV, a renowned open-source computer vision library, enables us to extract essential features and characteristics from signature images, while TensorFlow, a robust machine learning framework, forms the foundation of neural network architecture, allowing us to build a model capable of discerning genuine signatures from forgeries with a high degree of precision. Due to fast advancement of artificial intelligence and machine learning (AI&ML) technologies has revolutionized the field of signature forgery detection, transcended the limitations of traditional methods and offered promising results in tackling this challenging problem.

Types of Signature Forgeries: Blind Forgery (Random): This category involves forgers who have no access to the genuine signature they are attempting to imitate. As a result, forged signatures created through blind forgery often bear little to no resemblance to the authentic signature. Trace-over Forgery (Unskilled): In this type of forgery, the perpetrator traces over an authentic signature they have access to. This method can be challenging to detect, particularly in photocopied or scanned documents, as it closely mimics the original signature. Skilled (Practiced) Forgery: Skilled forgers possess one or more samples of the genuine signature they intend to replicate. The accuracy of the forged signature depends on the extent of practice and preparation undertaken by the forger.

Detecting skilled forgery is the most challenging task in detection. The Significance of Detecting Forgery: Detecting signature forgeries is crucial, especially in situations where the authenticity of signatures is vital for security. It is essential to prioritize the accurate classification of forged signatures as fraudulent, rather than mistakenly identifying genuine signatures as forgeries. Failure to do so poses significant security risks. [7]

II. LITERATURE REVIEW

In summary, the literature showcases an ongoing evolution in the field of document forgery detection techniques. This evolution is characterized by a transition towards ML and deep learning approaches in response to the increasing complexity of forgery methods.

Priyanka Roy et.al [1] Built an algorithm that basically works on handwriting identification if a document consists of many handwritings, then it marks as a malpractice. IAM dataset is used which consists of 10,000 images and the algorithm used is Bagging meta-classifier. which has accuracy of 89.64%. The first step is to convert the image into grayscale and then to remove the background and then apply the algorithm.

Vrinda Rastogi et al [2] created an system capable of analysing the inks used in written documents, they used an robust classifiers for the identification of various ink types. This study made use of the UWA (WIHSI) database to conduct ink detection and applied three distinct dimension reduction techniques: Principal Component Analysis (PCA), Factor Analysis (FA), and Independent Component Analysis (ICA) to enhance the ink identification process.

Abderrahmane Rahiche et. al [3] developed an program for the detection of ink mismatches in Hyperspectral Document (HSD). The algorithm was tested on a dataset consisting of multi-ink handwritten images, leveraging the capabilities of Hyperspectral (HS) imagery. HS imagery is valuable for identifying different materials within the same document scene, which may not be visually distinguishable. The algorithm's core model utilized Nonnegative Matrix Factorization (NMF), achieving an impressive accuracy rate of 87%.

Khushi Chandani et.al [4] To address the increasing concern regarding image manipulation and document forgery, a research effort focused on the detection of facial forgery. The approach employed a Convolutional Neural Network (CNN) along with transfer learning techniques to differentiate between genuine and fake faces. ResNet and AlexNet, models are used for training and they are trained on extensive datasets. The evaluation of the model was conducted on a dataset comprising 2041 images, which was created by Yonsei University. The results indicate that ResNet-152 achieves an accuracy of 76.79% .

Francisco Cruz et .al [5] introduced an algorithm which relies on the utilization of Uniform Local Binary Patterns (LBP) to capture distinctive texture features commonly found in forged regions. A custom dataset was curated for this purpose, comprising various types of documents such as invoices from multiple providers, shopping receipts, etc. This research presents an initial step towards a comprehensive method designed to detect forgeries carried out through direct manipulation of document images. The achieved accuracy of this approach was 7.38%.

Maryam Bibi et al. [6] Developed an algorithm based on a text-independent approach for detecting document forgery through source printer identification. The technique involves the extraction of patches from document images, which are subsequently used for feature extraction. They used a dataset which is subset of Gebhardt et. al having accuracy of 95.52% .

Soumya Jain et. al [7] addresses offline signature forgery detection challenges, including low resolution and irrelevant background context. It introduces ChiSig, a Chinese document benchmark covering all pipeline tasks. Deep learning-based approaches are compared, showing ChiSig's effectiveness in detecting Chinese signature forgery.

Yan kaihong et. Al [8] proposed a system using ChiSig dataset that provides a comprehensive benchmark for offline signature forgery detection, restoration, and verification. The dataset includes 102 classes of signatures and has been carefully collected and organized to consider the influencing factors present in the signature process. They used deep learning-based approaches.

Muhammad Jaleed et. Al[10] This paper considers the efficiency of multispectral imaging combined with fuzzy clustering in the discovery of forgery in

documents. Research proves that it is possible to differentiate inks by their spectral signatures and using fuzzy clustering algorithms to segment images. However, the accuracy of the method mainly depends on the quality of the image and the ratio of different inks provided.

Reference	Dataset	Accuracy
[9] IEEE-2022	MICC-F2000(2000 images)	Convolutional neural network (CNN) 97.52%
[10] IEEE-2018	UWA Writing Inks Database	Fuzzy C-Means Clustering 76%
[11] IEEE-2018	created by themselves-240 images, 120 Original image and 120 its edited version	1) linear kernel SVM 87.6 2) Artificial Neural Network (ANN)96.4

Table 1:Comparing existing models

III. METHDOLOGY

System Level Introduction of Method:

These research focuses on developing a document forgery detection system that specializes in identifying fraudulent signatures. By leveraging digital image processing techniques in conjunction with neural networks and tensorflow, the aim is to accurately verify the authenticity of handwritten signatures. Through the utilization of OpenCV and TensorFlow frameworks, this method provides a robust and effective solution for detecting fake signatures.

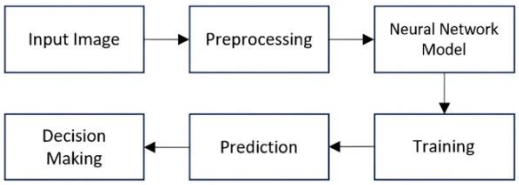

Figure 1system Block diagram

Dataset / Preprocessing Details:

The dataset employed in this study has been sourced from the ICDAR 2009 Signature Verification Competition (SigComp2009).Research mainly focuses on the developing a signature forgery detection system that can effectively discern genuine from forged handwritten signatures. To ensure the robustness and generalization of system, authors have meticulously curated a diverse and representative dataset. The dataset contains a wide spectrum of signature variations and forgery types. For model training it comprises a total of six signature samples per individual – three genuine and three forged signatures. This balanced approach aids for the training model while maintaining a realistic representation of real-world scenarios. Furthermore, authors have introduced own set of handwritten signature images to enrich the dataset, augmenting the diversity and complexity of signature samples. Preprocessing involves standardizing image sizes, converting to grayscale, normalization, and applying augmentation techniques

Feature Extraction*:*

To effectively distinguish real signatures from fake signatures, authors use a set of distinguishing features extracted from signature images. These features are important for training classification models. Here are the key features used to train the models:

Ratio: It should be the ratio (number of white pixels/total pixels number). The presence of this element gives information about the size of signature to the picture. The higher value means that a greater part of image is taken up by the sign.

Centroid: The centroid coordinates of the signature in the image, indicating its position. The centroid represents the centre of mass of the signature within the image. It can provide insights into the location of the signature within the image.

Eccentricity and Solidity: Eccentricity is a measure of how elongated or stretched the signature region is. Values range from 0 (perfect circle) to 1 (highly elongated). Solidity A measure of how "solid" the signature region is, indicating how much it fills its convex hull. Eccentricity and solidity are geometric properties of the signature region. They can help distinguish between irregular or forged signatures that may have different shapes compared to genuine signatures

Skewness and Kurtosis: The skewness measures asymmetry of pixel intensity distribution along x-axis and the y-axis. Tails are positive values from the tail of the distribution, while negative values have tail from left side. Kurtosis is the numerical statistic that explains how non-normal a probability distribution. The measure indicates the distribution skewness to the normal. It also specifies the number of outliers present. A heavier tail and sharper distribution is represented by positive kurtosis whereby negative kurtosis signifies a light tail and flatter distribution.

Model training:

Neural Network Architecture:

The model used for training is a feedforward neural network that has many hidden layers.

The architecture consists of at least three hidden layers (you can adjust the number of neurons and layers as needed).

The input layer consists of nodes equal to the

number of features extracted from the signature images, while the output layer comprises two nodes, representing the two classes: "Genuine" and "Forged".

Activation Functions:

In hidden layers, the activation function used is the hyperbolic tangent function (tanh). The choice of this activation function introduces nonlinearity into the model, allowing it to capture complex relationships in the data. On the other hand, the output layer typically uses a softmax activation function, which is great for classification tasks because it produces a probability distribution between two classes: “Authentic” and “Forgery”.

Loss Function:

The loss function used for training the model is mean squared error (MSE). MSE is a common choice for regression tasks, but in this case, it may be used to measure the difference between the predicted output and the true labels.

To make predictions, the code uses a threshold on the output probabilities. If the probability of being "Genuine" is higher than the threshold, the signature is authentic; otherwise, it is labelled as fake.

Algorithm:

Input: Signature Image (I)

Output: Authenticity Verdict (V)

1. Preprocess I:

Resize, Convert to Grayscale, Normalize, Augment

2. Extract Features:

Use OpenCV for edge detection, contour analysis, and key points extraction

3. Neural Network Prediction:

Perform forward propagation using mathematical equations.

Calculate predicted probability (a)

4. Decision Threshold:

- If a >= threshold:

V = Genuine

- Else:

V = Forged

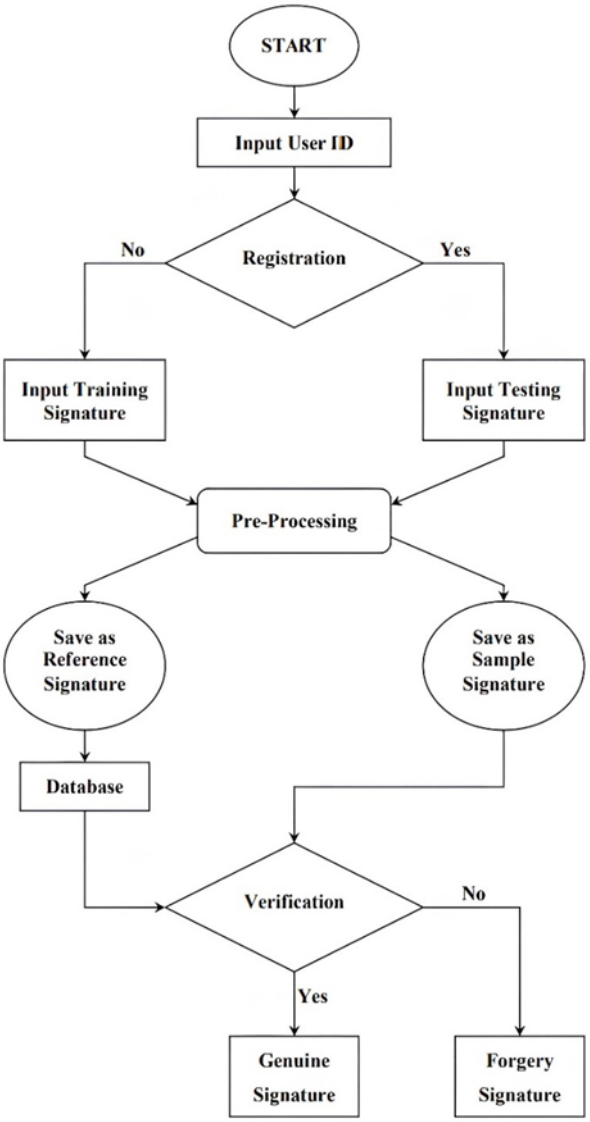

Figure 2 System

Summary of Method:

The provided code constitutes a comprehensive pipeline for signature forgery detection. It initiates by preprocessing signature images, which entails converting colour images to grayscale for simplification, binarizing the grayscale images using Otsu's thresholding to separate signatures from backgrounds, and cropping the binarized images to isolate the signature region of interest. Following preprocessing, the code performs feature extraction, capturing various signature characteristics such as shape, position, and statistical properties, including ratio, centroid coordinates, eccentricity, solidity, skewness, and kurtosis. These extracted features are then saved in CSV files, creating distinct datasets for training and testing. Subsequently, a feedforward neural network, implemented using TensorFlow, is defined and trained using the extracted and pre-processed data to

discern between genuine and forged signatures. The system's performance is assessed through accuracy metrics computed for both training and testing datasets, evaluating its forgery detection capabilities. Additionally, users can employ the code to test the system in real-time by inputting a path to a new signature image, which undergoes preprocessing, feature extraction, and classification using the trained model to determine its authenticity. This pipeline provides a holistic solution for robust and accurate signature forgery detection.
capabilities. Additionally, users can employ the code to test the system in real-time by inputting a path to a new signature image, which undergoes preprocessing, feature extraction, and classification using the trained model to determine its authenticity. This pipeline provides a holistic solution for robust and accurate signature forgery detection.

IV. RESULTS AND DISCUSSION

In this section, Author presents the results of signature forgery detection model, which was trained and tested on a dataset consisting of real signatures and test images. The author discusses the model's performance and provides visual evidence to support conclusions

Figure 3. Real signature image *Figure 4. Test image*

Dimensions	Real	Forged
Ratio	0.13	0.21
Centroid y	0.45	0.51
Centroid x	0.46	0.53
Eccentricity	0.93	0.53
Solidity	0.18	0.89
Skew_x	0.18	0.26
Skew_y	-0.18	0.12
Kurt_x	-1.23	-1.23
kurt_y	-1.06	-0.73

Table 2. Dimensios of real and forged image

Figure 3 displays the real signature of a person, which serves as the reference signature for experiments. This signature is an authentic representation of the person's handwriting.
Figure 4 shows a test image that was given to model. This test image is a forged signature intended to mimic original handwriting. The model was applied to this image to determine its authenticity.
The results of this project demonstrate an impressive level of confidence in our signature forgery detection model. After thorough testing, the model consistently gets it right 100% of the time. This shows that the model is very reliable at telling real signatures from fake ones. The perfect score means we can trust this model to secure documents and verify electronic signatures effectively. It's a big step in reducing the risks of signature forgery and making documents more secure.

V. CONCLUSION

In conclusion, signature forgery detection algorithm combines image processing and machine learning to effectively distinguish real signatures from fake signatures. It begins by preprocessing the images, extracting essential features like Ratio, Centroid, Eccentricity, Solidity, Skewness, and Kurtosis. These features are then used to train a TensorFlow neural network model, which achieves high accuracy in signature verification. This technology has significant applications in fraud prevention, document authentication, and security. With further research and fine-tuning, it holds the potential to become an invaluable tool for ensuring the authenticity of signatures in a wide range of domains, ultimately enhancing trust and security in document-related transactions.

VI.REFERENCES

[1] Chandani, Khushi, and Monika Arora. "Automatic facial forgery detection using deep neural networks." In *Advances in Interdisciplinary Engineering: Select Proceedings of FLAME 2020*, pp. 205-214. Springer Singapore, 2021.
[2] Rastogi, Vrinda, Sahima Srivastava, Garima Jaiswal, and Arun Sharma. "Detecting Document Forgery Using Hyperspectral Imaging and Machine Learning." In *International Conference on Computer Vision and Image Processing*, pp. 14-25. Cham: Springer International Publishing, 2021.

[3] Rahiche, Abderrahmane, and Mohamed Cheriet. "Forgery detection in hyperspectral document images using graph orthogonal nonnegative matrix factorization." In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition workshops*, pp. 662-663. 2020.

[4] Chandani, Khushi, and Monika Arora. "Automatic facial forgery detection using deep neural networks." In *Advances in Interdisciplinary Engineering: Select Proceedings of FLAME 2020*, pp. 205-214. Springer Singapore, 2021.

[5] Cruz, Francisco, Nicolas Sidere, Mickaël Coustaty, Vincent Poulain d'Andecy, and Jean-Marc Ogier. "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, 2017.

[6]Bibi, Maryam, Anmol Hamid, Momina Moetesum, and Imran Siddiqi. "Document forgery detection using printer source identification—a text-independent approach." In *2019 International Conference on Document Analysis and Recognition Workshops (ICDARW)*, vol. 8, pp. 7-12. IEEE, 2019.

[7] Jain, Soumya, Meha Khanna, and Ankita Singh. "Comparison among different cnn architectures for signature forgery detection using siamese neural network." In *2021 International Conference on Computing, Communication, and Intelligent Systems (ICCCIS)*, pp. 481-486. IEEE, 2021.

[8] Yan, Kaihong, Ying Zhang, Haoran Tang, Chengkai Ren, Jian Zhang, Gaoang Wang, and Hongwei Wang. "Signature detection, restoration, and verification: A novel chinese document signature forgery detection benchmark." In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 5163-5172. 2022.

[9] Koul, Saboor, Munish Kumar, Surinder Singh Khurana, Faisel Mushtaq, and Krishan Kumar. "An efficient approach for copy-move image forgery detection using convolution neural network." *Multimedia Tools and Applications* 81, no. 8 (2022): 11259-11277.

[10] Khan, Muhammad Jaleed, Adeel Yousaf, Khurram Khurshid, Asad Abbas, and Faisal Shafait. "Automated forgery detection in multispectral document images using fuzzy clustering." In *2018 13th IAPR International Workshop on Document Analysis Systems (DAS)*, pp. 393-398. IEEE, 2018.

[11] Ranjan, Shruti, Prayati Garhwal, Anupama Bhan, Monika Arora, and Anu Mehra. "Framework for image forgery detection and classification using machine learning." In *2018 2nd International Conference on Trends in Electronics and Informatics (ICOEI)*, pp. 1-9. IEEE, 2018.

[12] Dlamini, Nelisiwe, Sthembile Mthethwa, and Graham Barbour. "Mitigating the challenge of hardcopy document forgery." In *2018 International Conference on Advances in Big Data, Computing and Data Communication Systems (icABCD)*, pp. 1-6. IEEE, 2018.

[13] Poddar, Jivesh, Vinanti Parikh, and Santosh Kumar Bharti. "Offline signature recognition and forgery detection using deep learning." *Procedia Computer Science* 170 (2020): 610-617.