Signature Verification and Fraud Detecting Using OpenCV and Machine Learning

[1]Tasmiya Anjum H N,[2]Sachin R Gowda,[3]Sai Chethana S P,[4]Sangeetha R,[5]Sathvik R Bharadwaj,

Information Science and Engineering, Malnad College Of Engineeing,

Hassan-573202, India

Email id:thn@mcehassan.ac.in,sachingvr81@gmail.com, saichethanasp@gmail.com, sangeetha7102003@gmail.com, sathvikrb15@gmail.com

---***---

Abstract - In the realm of modern financial transactions, the reliable verification of bank cheques is crucial for maintaining transactional integrity and preventing fraudulent activities. This study introduces a novel approach verification and classification of signatures on bank cheques, leveraging a comprehensive framework that integrates Optical Character Recognition (OCR), line sweep techniques, and Convolutional Neural Networks (CNN). Our proposed methodology aims to improve the accuracy and efficiency of signature verification processes by harnessing the capabilities of advanced technologies. The system begins by utilizing OCR to extract textual information from cheque images, facilitating the identification of key components such as account details and text-based elements. Complementing this, the line sweep technique is employed to analyse spatial relationships within cheque images, focusing on critical features like signatures and account numbers. The cornerstone of our approach lies in the application of Convolutional Neural Networks. By training CNNs on a diverse dataset of signatures, the system learns intricate patterns and characteristics associated with genuine and counterfeit samples. This deep understanding enables the system to make informed decisions regarding the authenticity of signatures, thereby contributing to a more robust and accurate verification process. Through the integration of OCR, line sweep analysis, and CNNs, our proposed system represents a significant advancement in the domain of signature classification and bank cheque verification. The automation and augmentation of verification procedures promise to bolster security, reduce errors, and foster increased confidence in financial transactions. The subsequent sections provide a detailed exposition of our methodology, supported by empirical evidence of its efficacy and potential impact.
.

Key Words: Bank Cheques, CNN, OCR, Line Sweeping etc.

1.INTRODUCTION

In the dynamic landscape of modern financial transactions, ensuring the authenticity and security of bank cheques is of paramount significance. Bank cheques, as a primary instrument for transferring funds and conducting monetary exchanges, necessitate robust verification mechanisms to mitigate errors and prevent fraudulent activities. Traditional manual verification processes are not only time-consuming but also prone to human errors and subjectivity. To address these challenges, advanced technologies are being harnessed to revolutionize the verification process, enhancing accuracy and efficiency.

The use of handwritten signatures for personal identification and authentication is quite common. Verifying signatures is necessary for many documents, including legal transactions and bank checks. However, given the sheer volume of records, it is a highly challenging and time-consuming operation. Consequently, a strong automatic signature verification technology that strives to lower fraud in all connected financial transaction industries is essential.

This study presents an innovative and comprehensive approach to signature classification and verification of bank cheques, leveraging a synergistic combination of Optical Character Recognition (OCR), line sweep techniques, and Convolutional Neural Networks (CNN). By integrating these cutting-edge technologies, we aim to create a robust and automated system that can accurately classify and verify signatures on bank cheques, thereby enhancing the overall security and reliability of financial transactions.

The proposed methodology encompasses multiple layers of analysis. First, OCR technology is employed to extract textual information from cheque images, facilitating the identification of account details and relevant textual components. Complementing this, the line sweep technique is utilized to assess the spatial arrangement of critical features, such as signatures and account numbers, which are vital indicators of a cheque's authenticity.

Furthermore, the power of Convolutional Neural Networks is harnessed to achieve advanced signature classification and verification. CNNs, renowned for their ability to extract intricate patterns from images, are trained on a diverse dataset of signatures to develop a deep understanding of genuine and counterfeit samples. This enables the system to make informed decisions about the legitimacy of signatures, contributing to a more robust and accurate verification process.

Through the fusion of OCR, line sweep analysis, and CNNs, our proposed approach aims to set new standards for the verification of bank cheques and signature authentication. By automating and enhancing the verification process, we aspire to provide financial institutions with a powerful tool to bolster security, minimize errors, and establish a higher level of trust in the realm of financial transactions. In the subsequent sections, we delve into the intricate details of our methodology, followed by empirical results that showcase the effectiveness and potential of this innovative system.

2. **Related Work**

Researchers and practitioners alike have focused a lot of emphasis on the topic of signature categorization and verification, particularly in relation to bank checks. Different strategies have been investigated to deal with the difficulties in maintaining the integrity and legitimacy of signatures on financial documents. The following related works highlight some of the significant contributions in this area:

Our creative tool precisely verifies key components of bank cheques—such as the Branch identifier, cheque serial, lawful and polite sums, bank account identifier, and signature styles—using image processing and deep learning techniques. Using the IDRBT check dataset in conjunction with convolutional neural networks, we achieved a 99.14% accuracy for identifying handwritten numeric characters. Using MATLAB's OCR, we achieved 97.7% accuracy for machine-printed script. Our application of Support Vector Machine (SVM) as a classifier and Scale Invariant Feature Transform (SIFT) for signature feature extraction also produced a 98.10% accuracy for signature verification [1]. Though, even with these high accuracies, the tool might still struggle with very unreadable handwriting, smeared or damaged checks, and unusual signature patterns. The tool's efficacy in real-world situations could also be affected by the quality of the input images and the complexity of the cheque components, which would affect its accuracy. Changes in cheque design or format might also affect the tool's performance, requiring continuous adaptation and enhancement [1].

Aiming to increase accuracy in cheque verification procedures, another approach used a recurrent attention optical character recognition network (RA_OCRN) to extract legal information from cheque images. Showing better outcomes than existing techniques when evaluated on the Kaggle cheque detection dataset, the study employed deep learning for data pretreatment and augmentation. The results showed that RA_OCRN, especially when combined with data augmentation, delivered the highest recognition accuracy [2]. However, this study focused solely on the recognition of legal information and did not account for challenges like damaged or altered cheques, non-standard fonts, or the practical implementation costs of deploying RA_OCRN models in real banking environments [2].

One solution combined dynamic and static characteristics for verifying handwritten signatures to strengthen authenticity and protect consumer property in the banking sector. Unlike traditional methods relying on human judgment, which can result in misunderstandings or financial loss, digital signature verification systems aim to verify transactions, enhance customer service, and promote staff-customer trust [3]. Still, such systems may be prone to impersonation or electronic signature forgery and may suffer from technological complexity due to their dependence on machine learning, image processing, and image acquisition [3].

Signature verification remains a core biometric identification method in financial transactions, supported by various models such as Hopfield Neural Networks, SVM, Hidden Markov Models, and Artificial Neural Networks (ANN). ANN-based signature validation is a notable trend due to its ability to handle non-linear data effectively [4]. That said, signature verification has its pitfalls: it can be susceptible to forgery, lacks standardization, is computationally complex, and heavily depends on well-curated training data [4].

A proposed three-layer signature verification system employs writer-independent, offline verification using graph metrical and FAST features as input. It combines artificial neural networks, Gaussian mixture models, and image matching, achieving 99% overall accuracy [5]. However, this system may face challenges in real-time processing due to image quality variation and complexity in extracting features from overlapping or noisy signatures [5].

Developed using Python, another CNN-based offline signature verification system addressed performance issues for high-volume document validation, achieving 99.70% accuracy [6]. While this improved static verification, the study also highlighted the growing relevance of online biometric systems like fingerprints and iris scans. These approaches, however, come with infrastructure costs, potential physical accessibility issues, and data privacy concerns [6].

Innovatively, a device-independent online handwritten signature verification system (ASSV) was developed using acoustic signals instead of sensor-instrumented pens. Using a chord-based model for estimating phase changes, and discrete cosine transforms for feature extraction, the system achieved an AUC of 98.7% and an Equal Error Rate (EER) of 5.5% [7]. Despite its strengths, the model's real-world performance could be undermined by environmental noise or individuals' unique acoustic signing patterns [7].

In a broader context, signature verification is critical in the face of increasingly sophisticated online threats. Though handwritten signatures are difficult to mimic, even skilled forgeries can defeat automated systems. Enhancing security through user involvement and education may bolster resilience but introduces its own challenges in user compliance and system integration [8].

Automation of cheque processing is another major trend. Systems now integrate preprocessing, information extraction, recognition, and verification to reduce the effort and error rates of manual checks. They enable fast detection of bounced cheques and expedite valid ones [9]. However, such systems may struggle with poorly written cheques and require significant upfront investment and integration effort [9].

Using OCR and deep learning, other proposed systems automate data extraction from bank cheques, including payee name, date, and signature validation. A modified CNN trained on the IAM dataset demonstrated improved efficiency [10]. Yet, these systems are still sensitive to handwriting variations and may raise security concerns due to how signatures and personal data are stored [10].

A comprehensive review of 56 studies from 2014 to 2019 showed that deep learning models consistently perform well in offline signature verification (OfSV). It highlighted five key pillars of successful OfSV systems: dataset selection, preprocessing, feature extraction, classification models, and evaluation metrics [11]. Still, the review may be outdated and omits newer advances or unpublished works [11].

Handwritten signature verification, positioned as a two-class classification problem, continues to evolve through robust machine learning techniques. A survey of over 20 papers compared multiple datasets and feature extraction approaches, pointing to the need for more resilient, interpretable, and computationally efficient models [12].

The Cheque Truncation System (CTS) has introduced standardization and security features, yet remains vulnerable to sophisticated cloning and forgery methods. A particular study demonstrated the ease with which a machine-generated cheque could mimic a genuine one, exposing limitations in CTS's current detection abilities [13].

One signature verification tool combined SURF and SIFT algorithms for edge detection and feature extraction, showing strong performance in speed and accuracy [14]. However, these tools remain vulnerable to adversarial attacks, variations in signature style, and the challenges of scaling to large databases [14].

The digital banking revolution necessitates tools for automated cheque data extraction. A system using CNNs and the Tesseract OCR framework enhanced extraction accuracy for key cheque regions like payee and date fields [15]. However, the diversity of cheque layouts and the need for real-time operation remain significant challenges [15].

Another holistic solution, the Signature Forgery Detection and Verification system, integrates image processing, OCR, and ML for high-fidelity signature validation, aiding sectors like banking and administration [16]. Yet, limitations remain, especially in detecting advanced forgeries and in the computational demands of large-scale deployment [16].

As a unique personal identifier, signatures continue to face issues of imitation, degradation, and variability. These challenges are compounded in digital formats, where legal and technical standards for e-signatures vary across regions [17].

Finally, as biometric verification gains traction, signature-based authentication presents itself as a secure and user-friendly method. It avoids pitfalls of passwords and tokens, but still faces issues like spoofing, high deployment costs, and user resistance due to privacy or cultural concerns [18].

To streamline cheque validation, another system leverages OCR and Easy OCR alongside OpenCV, using a custom bank cheque dataset for performance evaluation. While effective, the system's performance still hinges on input quality, cheque formats, and integration complexity [19].

3. Proposed Methodology

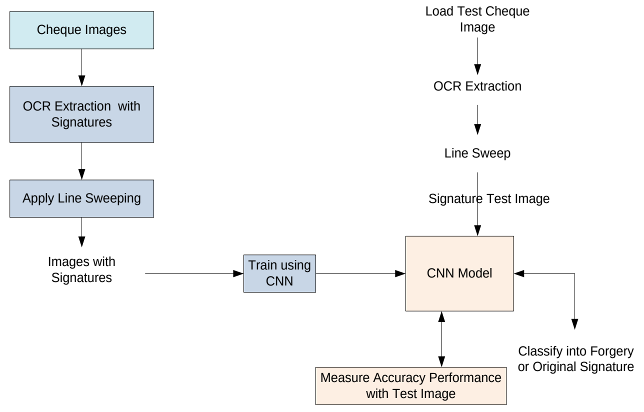

Fig 1: The Proposed System Architecture

3.1 Data Augmentation:

The subsequent code presents the implementation of the data augmentation process employed to generate images and artificially expand the dataset's size. In this context, the Image Data Generator is utilized to create novel images. The rotation range is configured to 360 degrees since the images, being spiral in nature, can be rotated to various angles without altering the inherent meaning of the image. It is advisable to explore other image transformations within the Image Data Generator class; however, caution is advised while applying augmentations as certain transformations might potentially reduce the CNN model's overall accuracy. Post augmentation, the distribution of data is adjusted.

The images are also resized to a uniform dimension of (128, 128, 1), and subsequent to resizing, image normalization is executed before incorporating the dataset into the model.

3.2 CNN Model Architecture:

The implementation adopts a CNN model architecture characterized by the ensuing attributes:

- The model encompasses four Convolutional Layers housing 128, 64, 32, and 32 filters, respectively.

- Diverse filter sizes are employed across the convolutional layers.

- Each convolutional layer is succeeded by a MaxPool2D layer.

- A pair of Fully Connected layers ensue after the convolutional block.

Specification of the Model using Keras:

3.3 Model Training:

The model undergoes training with a learning rate of 3.15e-5 using the Adam optimizer. The number of epochs is set at 70, and the batch size is established as 128.

3.4 Model Performance:

To assess the model's performance, various metrics are employed, including Loss and Accuracy Plots, Classification Report, and Confusion Matrix. These evaluations provide insights into the model's effectiveness and accuracy.

4. Algorithms of Proposed Work

Creating a mathematical model for the "Signature Classification and Verification of Bank Cheques using OCR, Line Sweep, and CNN" involves representing the various components of the methodology mathematically.

Let's break down the key aspects of the research and formulate a mathematical model for each:

1. OCR Component:

Let's denote the OCR process as a function that takes an input image of a bank cheque and produces extracted textual information:

OCR (image) → Textual Information

Here, the OCR function analyzes the input image and outputs relevant textual details such as account numbers, payee names, and other relevant information.

2. Line Sweep Component:

The Line Sweep technique involves systematically scanning and extracting features from signatures. Let's define the Line Sweep process as follows:

Line Sweep(signature) → Extracted Features

The Line Sweep function takes a signature image as input and performs a line-by-line scan, extracting features such as stroke directions, lengths, and angles. The output is a set of features that represent the signature's characteristics.

3. CNN Model Architecture:

The Convolutional Neural Network (CNN) model can be represented using mathematical notation for each layer:

- Convolution Layer: Conv (input, filter) → Feature Map

- MaxPooling Layer: MaxPool(input) → Pooled Feature Map

- Fully Connected Layer: FC(input, weights) → Output

The CNN model consists of multiple layers, each performing convolution, pooling, and fully connected operations. The output of the final layer represents the model's classification decision.

A Convolutional Neural Network (CNN) is a type of deep learning model commonly used for image classification, object detection, and other computer vision tasks. Here, I'll provide a mathematical model for a basic CNN architecture, including key operations and components involved. Please note that this is a simplified representation for illustration purposes.

Let's consider a simple CNN architecture with convolutional layers, pooling layers, and fully connected layers. We'll use mathematical notation to represent each step:

1. Convolutional Layer:

Given an input image X of size $H_{in} * W_{in} * C_{in}$, where H_{in} height W_{in} is the width, and C_{in} is the number of input channels (e.g., 3 for RGB images), and a set of K filters $W_{(1)} = \{W_1, W_2, W_3, \ldots, W_k\}$ of size F*F*C_{in} , where F is the filter size:

Apply convolution operation with each filter W_k to get feature maps $Z_{k(1)}$:

$$Z_{k(1)} = f(W_K * X + b_k \quad (1)$$

where b_k is the bias term associated with filter W_K f is the activation function (e.g., ReLU), and * represents the convolution operation.

2. Pooling Layer (MaxPooling):

Given feature maps$Z_{k(1)}$ of size $H_{in} * W_{in}$ after the convolutional layer, perform max pooling with a pooling window of size P*P:

Apply max pooling operation to get pooled feature maps $P_{k(1)}$:

$$P_{k(1)}(i,j) = max_{m,n} * Z_{k(1)}(i * P + m, j * P + n) \quad (2)$$

where $P_{k(1)}(i,j)$ is the value of the pooled feature map at position(i,j) .

3. Flatten Layer:

Flatten the pooled feature maps $P_{k(1)}$to create a 1D vector of size $H_{out} * W_{out} * K$ to prepare for fully connected layers:

F = flatten($P_{1(1)}, P_{1(1)}, \ldots, P_{k(1)}$) (3)

4. Fully Connected Layer:

Given the flattened vector F and weights W^2 and b^2 biases , perform matrix multiplication and apply activation function to get the output of the fully connected layer $A_{(2)}$:

$$A_{(2)} = f(W_{(2)} * F + b_{(2)}) \quad (4)$$

5. Output Layer:

Given the output of the fully connected layer $A^2$and weights W^3 and biases b^3,apply matrix multiplication and activation function (e.g., softmax) to obtain the final class probabilities Y:

Y = softmax($W_{(3)} * A_{(3)} + b_{(3)}$) (5)

This mathematical model captures the basic operations and components of a CNN. The model's parameters($W_{(1)}, b_{(1)}, W_{(2)}, b_{(2)}, W_{(3)}, b_{(3)}$)are learned through training using optimization algorithms to minimize a loss function that measures the difference between predicted and actual class labels. This model can be further extended with more layers, skip connections, and other architectural elements for more complex tasks.

4. Overall Signature Classification:

Combining the OCR, Line Sweep, and CNN components, the signature classification process can be represented as a composite function:

SignatureClassification(image) = CNN(LineSweep(OCR(image))) (6)

Here, the OCR component extracts textual information, the Line Sweep component extracts signature features, and the CNN model processes the combined features to classify the signature as genuine or fraudulent.

5. Training and Optimization:

The mathematical optimization process involves minimizing a loss function using an optimization algorithm (e.g., stochastic gradient descent). Let's denote the optimization process as follows:

Optimize(parameters) = Minimize(LossFunction(parameters)) (7)

The parameters represent the weights and biases of the CNN model, and the LossFunction calculates the difference between predicted and actual classifications.

6. Performance Metrics:

The evaluation of the model's performance involves calculating various metrics such as accuracy, precision, recall, and F1-score. These metrics can be defined as follows:

Accuracy = (True Positives + True Negatives) / Total Samples (8)

Precision = True Positives / (True Positives + False Positives) (9)

Recall = True Positives / (True Positives + False Negatives) (10)

F1-Score = 2 * (Precision * Recall) / (Precision + Recall) (11)

These metrics quantify the model's ability to correctly classify genuine and fraudulent signatures.

By combining these mathematical representations, we create a comprehensive model that captures the signature classification and verification process using OCR, Line Sweep, and CNN techniques. The model encompasses image analysis, feature extraction, deep learning, optimization, and performance assessment, providing a structured framework for the research methodology.

5. Results

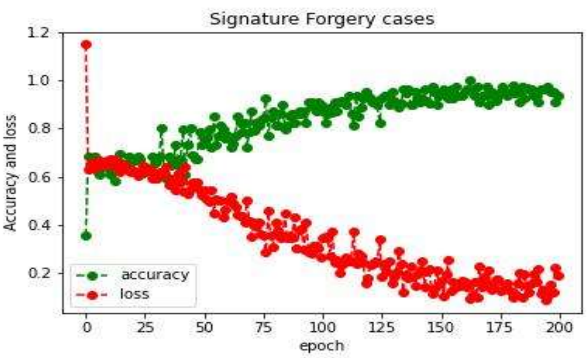

Fig 2: Accuracy and Loss Graph for Signature Forgery Cases with dataset

Fig 3: Detected as genuine with the browsed Cheque Image

Fig 4: Detected as forgery with the browsed Cheque Image

3. CONCLUSIONS

In this, we present a comprehensive approach to address the critical task of signature classification and verification in the context of bank cheques. The research aims to enhance the efficiency and accuracy of signature validation processes, ensuring the security and reliability of financial transactions. The combination of Optical Character Recognition (OCR), Line Sweep techniques, CNN constitutes a multi-faceted methodology that contributes to achieving these goals. Through the implementation of OCR, the study effectively extracts textual information from bank cheques, enabling the identification of key details such as account numbers and payee names. This integration enhances the overall signature verification process by incorporating additional contextual information, thereby refining the precision/accuracy of results.

Furthermore, the innovative application of Line Sweep techniques offers a unique and efficient approach to capturing signature attributes. By systematically scanning and extracting features from signatures, this methodology contributes to a more comprehensive representation of signatures, facilitating better differentiation between genuine and fraudulent ones.

are automatically asked, whether they would like to order the pdf, and are given instructions as to how to do so.

ACKNOWLEDGEMENT

We have made efforts in this project. However, it would not have been
possible without the kind support and help of many individuals and organizations.
We would like to extend our sincere thanks to all of them.

REFERENCES

[1] Prateek Agrawal, Deepak Chaudhary, Vishu Madaan, Anatoliy Zabrovskiy, Radu Prodan, Dragi Kimovski and Christian Timmerer. "Automated bank cheque verification using image processing and deep learning methods". (2021) Multimedia Tools and Applications. 80. 1-32. 10.1007/s11042-020-09818-1.

[2] Hitesh Chaitanyaswami, and Ashwin Dobariya. "Deep Learning Recurrent Attention Optical Character Recognition Network with Data Augmentation for Cheque Data Extraction." *2023 International Conference on Computer, Electronics & Electrical Engineering & their Applications (IC2E3)* https://doi.org/10.1109/IC2E357697.2023.10262478

[3] Dr.Thangavel V, "Use of Digital Signature Verification System (DSVS) in Various Industries: Security to Protect against Counterfeiting: Research". (April 24, 2023). SSRN: https://ssrn.com/abstract=4461014 or http://dx.doi.org/10.2139/ssrn.4461014

[4] Vijaya Dhopte1 , Prof. Shaila P. Kharde, "Signature Verification of Bank Cheque" IJIRSET Volume 10, Issue 10, October 2021 DOI:10.15680/IJIRSET.2021.1010093

[5] Bramara Neelima K and S Arulselvi "Signature Extraction and Recognition from Bank Cheque Image" IJITEE 2019 DOI: 10.35940/ijitee.B7766.129219

[6] Eman Alajrami, Belal A. M. Ashqar, Bassem S. Abu-Nasser, Ahmed J. Khalil, Musleh M. Musleh, Alaa M. Barhoom, Samy S. Abu-Naser "Handwritten Signature Verification using Deep Learning" International Journal of Academic Multidisciplinary Research (IJAMR) ISSN: 2643-9670 Vol. 3 Issue 12, December – 2019.

[7] Feng Ding, Dong Wang, Qian Zhang, and Run Zhao. 2019. "ASSV: Handwritten Signature Verification Using Acoustic Signals". Proc. ACM Interact. (September 2019), https://doi.org/10.1145/3351238

[8] Yash Borse, Anjali Patil, Sumeet Shah, Anand Gharu "Signature Verification Using Deep Learning" IJCRT 2023 Volume 11, Issue 4 April 2023 ISSN: 2320-2882

[9] Phalguni Nikam, Prof. Pramod Patil, Meetali Patidar, Aishwarya Nanoskar, Pranav Parmar and Jayesh More. "Checkue Bounce Detection System Using Image Processing". IRJET-volume 7 issue:01 Jan 2020 DOI https://doi.org/10.1007/s10032-011-0170-8.

[10] Mukesh Jha, Madhur Kabra, Sahil Jobanputra and Rupali Sawant "Automation of Cheque Transaction using Deep Learning and Optical Character Recognition." *International Conference on Smart Systems and Inventive Technology (ICSSIT)* (2019): DOI https://doi.org/10.1109/ICSSIT46314.2019.8987925

[11] M. Muzaffar Hameed, Rodina Ahmad, Miss Laiha Mat Kiah and Ghulam Murtaza (2021). "Machine learning-based offline signature verification systems: A systematic Review", Research Gate 2021 DOI: http://dx.doi.org/10.1016/j.image.2021.116139

[12] Zainab Hashim, Hanaa M.Ahmed and Ahmed Hussein Alkhayyat, "A Comparative Study among Handwritten Signature Verification Mrthods Using Machine Learning Techniques", Hindwai Vol.2022, DOI: https://doi.org/10.1155/2022/8170424

[13]Vishnu Dev Yadav, Khushboo Phore and Loganathan Lingan. "Bank Cheque Fraud and Cheque Truncation System to Illuminate the Fraudulent Transaction: A Case Study" J Forensic Res 14 (2023), DOI: 10.37421/2157-7145.2023.14.562

[14] Walid Hussein, Mostafa A. Salama and Osman Ibrahim "Image Processing Based Signature Verification Technique to Reduce Fraud in Financial Institutions"MATEC Web Conf. Volume 76, 2016 https://doi.org/10.1051/matecconf/20167605004

[15] Prajwal V Suranagi, and Dr T R Arunkumar "Automated Cheque Image Processing AI Driven Recognition For Efficient banking" irjmets Volume:05/Issue:09/September-2023 DOI : https://www.doi.org/10.56726/IRJMETS44875

[16] Navya V K, Abhilasha Sarkar, Aditi Viswanath, Akshita Koul, and Amipra Srivastava "Signature Verification And Forgery Detection System" IJCRT Volume 11, Issue 6 June 2023 ISSN: 2320-2882

[17] M. B. Sudhan, Abhilasha Sarkar, Aditi Viswanath, Akshita Koul, and Amipra Srivastava "A Survey on Signature Verification and Forgery Detection System" International Journal of Research Publication and Reviews, Vol 4, no 4, pp 1735-1738, April 2023 https://doi.org/10.55248/gengpi.2023.4.4.35359

[18] Shraddha Jagtap, Sejal Kalyankar, Tejal Jadhav and Ashwini Jarali "Signature Verification And Detection System"

[19] P. Kunekar, K. Vayadande, O. Kulkarni, K. Ingale, R. Kadam and S. Inamdar, "OCR based Cheque Validation using Image Processing," 5th Biennial International Conference on Nascent Technologies in Engineering (ICNTE), 2023, pp. 1-5,doi : 10.1109/ICNTE56631.2023.10146687.