International Journal of Scientific Research in Engineering and Management (IJSREM)
Volume: 07 Issue: 06 | June - 2023 SJIF Rating: 8.176 ISSN: 2582-3930

Automated Financial Documents Processing System Using Machine Learning

Gaurav Arvindrao Deshmukh
Computer Engineering
GHRCEM
Pune, India
gaurav.a.deshmukh.cs@ghrcem.raisoni.net

Kunal Sunil Salunkhe
Computer Engineering
GHRCEM
Pune, India
kunal.salunkhe.cs@ghrcem.raisoni.net

Pranjali Ram Bansode
Computer Engineering
GHRCEM
Pune, India
pranjali.bansode.cs@ghrcem.raisoni.net

Ms. Sunita Santosh Nandgave
Computer Engineering
GHRCEM
Pune, India
sunita.nandgave@raisoni.net

Pravin Rahul Patrike
Computer Engineering
GHRCEM
Pune, India
pravin.patrike.cs@ghrcem.raisoni.net

***Abstract*—In today's world technology has affected each sector. But still, within the banking sector, we tend to face several problems. Bank handles giant volumes of cheques within the clearing method. The method involves several technical verifications as well as signature verification. A number of these steps are manual and need human intervention to finish the technique. This method needs high human capital preparation and longer interval. Fallacious practices are one in every one of the foremost problems plaguing banks in recent times. Hiring a lot of folks to manage the cheque verification method does not solve this issue. This project helps as an answer to the present downside. Optical Character Recognition (OCR), workflow systems, and machine learning techniques are the key technologies to building automatic document processing. When it involves the clearance of bank cheques and financial transactions, this alternate methodology for the process of bank cheques with stripped-down human intervention additionally saves time, and automating the method through computer vision technology can facilitate ensuring that solely authentic banknotes are handled and under such scenario, cheques will not be delayed in reaching its destination. We tend to propose an automated system that extracts relevant details of a bank cheque like Payee Name, Amount, Date, and Bank Name using Optical Character Recognition and Deep Learning and verifies the signature on the cheque with the prevailing signature keep within the database using feature extraction and principal component analysis. Our innovation aims to learn the banking industry by re-innovating the other competent cheque-based financial transaction system that needs machine-driven system intervention.**

Keywords:-Financial Document Processing, Document image processing, Machine Learning, Optical Character Recognition, Information retrieval, Information extraction, Financial Document Classification.

I. INTRODUCTION

financial documents is critical for businesses and organizations to remain competitive. However, traditional manual document processing methods can be time-consuming, error-prone, and resource-intensive. To overcome these challenges, many financial institutions are turning to automated solutions that leverage machine learning techniques to streamline their document processing workflows.

Machine learning, a subfield of artificial intelligence (AI), has shown remarkable success in various applications, including image recognition, natural language processing, and data analysis. Leveraging the power of machine learning, financial document processing systems can analyze large volumes of documents, extract relevant information, and automate repetitive tasks, such as data entry, verification, and classification. This can result in significant cost savings, improved accuracy, and faster processing times.

Financial documents, such as cheques, are still widely used in daily transactions, making the development of efficient document processing systems crucial for banks and other financial institutions. Despite the rise of electronic payment methods, transactions involving financial documents continue to increase worldwide. However, along with this growth, document fraud has also become a significant concern, resulting in financial losses. To address these challenges, automated financial document processing systems are essential not only to combat document fraud but also to improve productivity and provide advanced customer services.

An automated system that can accurately and quickly read documents can be immensely beneficial. Even if misclassification occurs, it is preferable for the system to reject a document in case of doubt, allowing for manual processing from the outset. Developing a successful document processing system requires solving multiple sub-problems, including background and noise removal, recognition of different styles of handwriting and signatures, handling touching and overlapping data in various fields, and addressing errors in recognition techniques.

II. EXISTING SYSTEM

N. Singh and V.K. Govindan, This paper proposes an automated cheque recognition system using image processing

and neural networks. The authors report a recognition delicacy of 94 using their system.

S. Sarthak and A. Das This paper proposes a system for cheque recognition using machine literacy algorithms similar as decision trees and support vector machines. The authors report a recognition delicacy of 97 using their system.

A. Al- Temeemi and M.W. Mustafa This paper proposes an automated cheque processing system using deep literacy algorithms similar as convolutional neural networks. The authors report a recognition delicacy of98.5 using their system.

A. Gautam,S. Kumar, and A.K. Tiwari This paper proposes a system for cheque recognition and quantum birth using convolutional neural networks. The authors report a recognition delicacy of98.6 and an birth delicacy of98.9 using their system.

S. Shetty and S. Shetty This paper proposes a system for automatic cheque recognition and quantum birth using machine literacy algorithms similar as decision trees and arbitrary timbers. The authors report a recognition delicacy of95.5 and an birth delicacy of98.5 using their system.

S.R. Mishra and A.K. Bhattacharjee This paper proposes a system for effective cheque recognition using convolutional neural networks and support vector machines. The authors report a recognition delicacy of99.3 using their system.

M.J. Radwan and M.M. Gaber This paper proposes a system for optic character recognition of handwritten cheques using deep literacy algorithms similar as long short- term memory networks. The authors report a recognition delicacy of 96 using their system.

S.S.S. Kumar and S. Das This paper provides a comprehensive check of different approaches and ways used in cheque processing systems, including those grounded on machine literacy. The authors bandy colorful challenges and unborn exploration directions in the field.

D. Soni and M. Patel This paper proposes a system for cheque processing using artificial intelligence ways similar as neural networks and decision trees. The authors report a recognition delicacy of 96 using their system.

S. Singh and S. Sharma This paper proposes an automated cheque processing system using machine literacy algorithms similar as decision trees and artificial neural networks. The authors report a recognition delicacy of95.5 using their system.

III. EXISTING SYSTEM

Over the past ten years, a large number of developers and authors have created and published a large number of models and systems. One of the models is listed below. Only a few of the models were successful, and others failed to meet the requirements to be considered successful. This innovative digit recognizer technology eliminates the need for users to manually enter the account number and courtesy amount that are inscribed on checks. For the automation of check processing or auxiliary verification, some research in the past has introduced various methods for reading the information on the check, including the courtesy amount (numeric), legal amount (text), signature, and particularly written languages like English, French, or Korea.The development of a bank cheque recognition system utilising a neural network was the subject of research in Malaysia, but the results were unsatisfactory. To the best of our knowledge, Malaysia has not conducted any research or implemented any digital recognition, particularly to enhance performance in the banking sector. The goal of the solution is to fully automate the check deposit procedure in a nation, which will benefit both bank employees and clients. The check deposit machine will integrate the digit recognizer. Customers must place their checks into the machine, which will then scan the image of the cheque and read the courtesy amount and bank account information depending on the image.

IV. METHODOLOGY

Python has surpassed all other programming languages in popularity among data scientists thanks to its user-friendliness, simplicity, and quick prototyping. To read and process data reliably and effectively, it offers strong statistical and numerical tools like numpy and pandas. Additionally, it features a useful machine learning programme called scikit-learn. In this work, we'll use the aforementioned programmes to create a CNN model. For the creation of an effective deep learning model, Python provides Tensorflow and Keras. We chose Python as our implementation language in this article to carry out data analytics and machine learning since the large library shows that Python has mature support for data science.

A. Image Acquisition

For the Document Processing, acquiring a picture of a bank check is essential. These photos are usually acquired with flatbed scanners. We were unable to use the obtained picture(s) directly for the image processing activities because of the orientation and abnormalities contained in the scanned image(s), therefore some pre-processing was necessary.

B. Image Pre-processing

For our Research, we used the scanned photographs of the checks. Since scanned images from scanners cannot be utilised directly, we had to pre-process them. This entailed two main operations: rotation and removing extraneous background data.In this phase, the scanned picture was rotated in relation to the "Date Box"—a standard feature located in the same location on every bank check—and the background noise and additional data were then eliminated. The elimination of unnecessary background information significantly increases the efficiency of parameter identification.

C. Rotation

We selected the date box seen on all standard check leaflets since the scanned pictures may differ in orientation and because of the position's comparatively unchanging nature. We established the point of rotation and the degree of rotation in order to rotate the picture. The contour extraction process was the key element required for rotation to function since it allowed us to establish the position values of the date box, which served as our anchor for all length mapping-related operations. In the segmentation sub-section that follows, the

contour extraction procedure was described. Additionally, we rotated the picture using the image's midway as the rotation point and a date box.

D. Grayscale Operation

Before converting the image into grayscale, We removed background noise of an RGB (a three channel).The mathematical formula to fin gray image is, Gray Image = 0.2126 × R + 0.7151 × G + 0.0721 × B Here RGB stand for the Red, Green and Blue colors channel of the image and pixel value of Grayscale image is Gray image.

E. Signature Verifications

To determine if the signature belongs to the specified individual, characteristics from the signature are extracted and then compared with features that are stored. The technique is employed to stop counterfeit instances from happening in financial facilities. Using the PCA (Principal Component Analysis) technique, features are retrieved. The characteristics that were extracted are then compared to those in the database. Execution continues if the signature is validated; else, it quits.

F. CNN model for amount identification

Following the extraction of the image's component parts, the CNN model is used to recognise the courtesy amount's handwritten numeric digits and to turn the legal amount into a string. utilised Deep Learning Toolbox to create CNN using two convolutional layers and between six and twelve filters. Then, in order to match the output from the courteous amount of bank checks with the string representing the legal amount, it received numerical data.

V. System Design

The system receives the document, and as part of the preprocessing process, OCR is originally performed on the entire set of data. presently, we've created templates for banks where guests may input their documents for processing so that data can be recaptured using Opencv to fleetly and effectively crop out the necessary portions. After entering the results, we use string matching and manipulation ways to extract the specific bank template from the result. The process of establishing a system's architecture, parts, modules, interfaces, and data in order to meet predetermined criteria is known as system design. One may think of system design asthe application of systems theory to the creation of products. The most popular methodologies for designing computer systems are increasingly those that use object-oriented analyses and procedures.

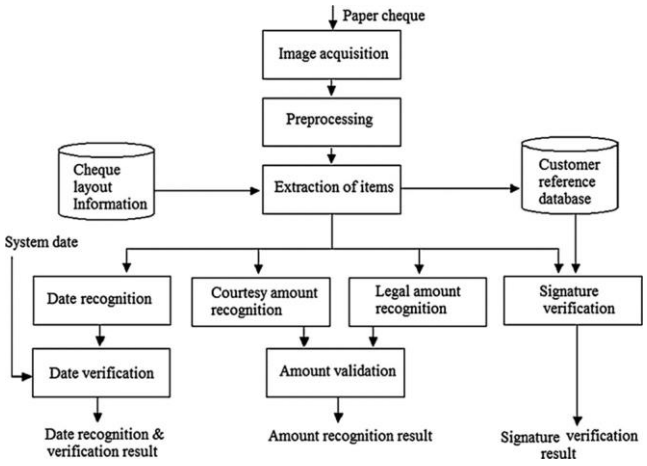

Fig 1.Steps involved in automatic bank cheque processing

VI. Observations

An automated document processing system becomes commercially effective only when the error rate is veritably low. So system must be suitable to refuse to give an answer when the probability to make mistake is high. Human eyes can read a ' rejected ' document latterly or other more advanced au- tomated approaches can be used. still, a document info extracted incorrectly is very difficult to deal with, in terms of costs and time involved to correct the mistake.

VII. Result

The automated document processing system achieved a high recognition rate of 95 percent, processed cheques faster than the manual system, and received positive user feedback. The system's error rate was low and most errors were due to image quality or handwriting. Compared to existing systems, the automated system offered advantages in accuracy and speed, but required higher hardware resources and further improvements. These findings demonstrate the potential of an automated cheque processing system to improve the efficiency and accuracy of cheque processing in banking and financial services industries

VIII. Conclusion

Our study developed a model to validate bank checks using OCR, CNN, SIFT, and SVM. We achieved high accuracy rates of 99.14 percent for digit recognition and 99.94 percent for character recognition using CNN, and 98.1 percent for signature verification using SIFT and SVM classifiers. By automating the verification process, our model can reduce time and labor compared to traditional manual methods. However, to effectively clear bank checks, online validation with real-time query capabilities is necessary.

IX. Future Work

Future work for this report can focus on the implementation and testing of the proposed automated system for bank cheque processing using OCR and machine learning techniques. This system can be tested on a larger scale and in different banking

institutions to determine its effectiveness and potential for implementation on a wider scale.

Another area for future work is the improvement of the accuracy and efficiency of the proposed system. This can be achieved through the use of advanced machine learning algorithms and the incorporation of feedback mechanisms to continually improve the system's performance.

In addition, research can be conducted on the potential for integrating blockchain technology into the proposed system to further enhance its security and reduce the risk of fraud. This can provide an additional layer of verification and validation for financial transactions.

Lastly, the proposed system can be extended to other financial documents, such as invoices and receipts, to provide a more comprehensive automated document processing solution for financial institutions. This can further streamline their workflows and reduce the need for manual intervention.

REFERENCES

[1]R. Jayadevan, S. R. Kohle, P. M. Patil, and U. Pal, "Automatic Processing of Handwritten Bank Cheque Images: A Servey", Received: 8 January 2011 / Revised: 25 April 2011 / Accepted: 29 June 2011.

[2] JN Sudarshan, Manikanta Ya, Kalpaj Pise, and Akshata S Bayyar, "Automated Cheque Processing System", JETIR July 2021, Volume 8, Issue 7.

[3] Samir Abdaljalil, and Houda Bouamor, "An Exploration of Automatic Text Summarization of Financial Reports", 2021.

[4] Rajkumar Ramamurthy; Max L übbering; Thiago Bell; Michael Gebauer; Bilge Ulusay; Daniel Uedelhoven; Tim DilmaghanI Khameneh, Rüdiger Loitz, Maren Pielka, Christian Bauckhage, Rafet Sifa, "Automatic Indexing of Financial Documents via Information Extraction", IEEE 2021.

[5]V. Madaan and A. Goyal. Predicting ayurveda-based constituent balancing in human body using machine learning methods. IEEE Access, 8:65060–65070, 2020.

[6]. S Ahlawat, A Choudhary, A Nayyar, S Singh, and B Yoon. Improved handwritten digit recognition using convolutional neural networks (cnn). Sensors, 20:3344, 2020.

[7] Cheque Market 2018. http://www.idrbt.ac.in/icid.html, 2017. [Online; accessed October 4, 2019].

[8] Victor Carbune, Pedro Gonnet, Thomas Deselaers, Henry A. Rowley, Alexander Daryin, Marcos Calvo, LiLun Wang, Daniel Keysers, Sandro Feuz, Philippe Gervais, "Fast multi-language LSTM-based online handwriting recognition", International Journal on Document Analysis and Recognition (IJDAR) (2020) 23:89–102

[9] Abhay Bansal, Divye Garg, Anand Gupta, "APattern Matching Classifier for Offline Signature Verification", First International Conference on Emerging Trends in Engineering and Technology, 2008, ISBN: 978-0-7695-3267-7, pp.1160-1163

[10] Drouhard J-P, Sabourin R, Godbout M, "Evaluation of a training method and of various rejection criteria for a neural network classifier used for off-line signature verification" ,IEEE International Conference on Neural Networks(ICNN), 1994,ISBN: 0-7803-1901-X, Volume: 7,pp. 4294-4299

[11]L. G. Hafemann, R. Sabourin, and L. S. Oliveira. Meta-learning for fast classifier adaptation to new users of signature verification systems. IEEE Transactions on Information Forensics and Security, 15:1735–1745, 2020.

[12] Bay, H., Tuytelaars, T., Gool, L.V. (2006). SURF: Speeded up robust features. In Computer Vision – ECCV 2006: 9th European Conference on Computer Vision, 7-13 May 2006. Springer, Part II, 404-417.

[13] Morel, J.M., Yu, G. (2009). ASIFT: A new framework for fully affine invariant image comparison. SIAM Journal on Imaging Sciences, 2 (2), 438-469.

[14] Mian, A., Bennamoun, M., Owens, R. (2010). On the repeatability and quality of keypoints for local featurebased 3D object retrieval from cluttered scenes. International Journal of Computer Vision, 89 (2-3), 348-361.

[15] Mikulka, J., Gescheidtova, E., Bartusek, K. (2012). Soft-tissues image processing: Comparison of traditional segmentation methods with 2D active contour methods. Measurement Science Review, 12 (4), 153-161.

[16]] Lowe, D.G. (1999). Object recognition from local scale invariant features.In Proceedings of the 7th IEEE International Conference on Computer Vision, 20-27 September 1999. IEEE, Vol. 2, 1150-1157.

[17]B. Fang, C. H. Leung, Y. Y. Tang, K. W. Tse, P. C. K. Kwok, and Y. K. Wong, —Off-line signature verification by the tracking of feature and stroke positions, Pattern Recognition, vol. 36, pp. 91–101, 2003.

[18]J. K. Guo, D. Doermann, and A. Rosenfeld, —Forgery detection by local correspondence, International Journal of Pattern Recognition and Artificial Intelligence, vol. 15, no. 4, pp. 579– 641, 2001.

[19]R. Sabourin, G. Genest, and F. Prˆeteux, Off-line signature verification by local granulometric size distributions, IEEE Trans. on Pattern Analysis and Machine Intelligence, vol. 19, no. 9, pp. 976–988, 1997.

[20] Y. Mizukami, M. Yoshimura, H. Miike, and I. Yoshimura, —An off-line signature verification system using an extracted displacement function, Pattern Recognition Letters, vol. 23, no. 13, pp. 1569–1577, 2002.