iJRASET
International Journal For Research in Applied Science and Engineering Technology

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 9 **Issue:** VI **Month of publication:** June 2021

DOI: https://doi.org/10.22214/ijraset.2021.36023

www.ijraset.com

Call: 08813907089 | E-mail ID: ijraset@gmail.com

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.429
Volume 9 Issue VI June 2021- Available at www.ijraset.com

Postal Address Identification and Sorting

Mr. Onkar Deshpande[1], Prof.Milind Kulkarni[2]
[1, 2]*Department of Computer Engineering, Vishwakarma Institute of Technology, Pune, India.*

Abstract: ***In this fast-moving world, a normal man can take considerable time to find a postal card in a bunch of postcards with significant issues like unclear handwriting, having trouble recognizing some uncommon or ambiguous names. Also, in postal offices or industries, it negatively impacts the efficiency of the postal system. I am making a system for Indian postal automation based on recognizing pin-code on the postcard. In India, there are multiple languages were speak. Indian postcards are mainly written in three languages the state's official language, English, and Devanagari language. In India, more than 50% of people write Pincode digits in either English or Devanagari language, so I am making such a system that sorts both English and Devanagari language postcards. Moreover, the system is mature enough to recognize handwritten as well as printed digits. As a result, the system gets an accuracy of 92.59% on the English language postcards, 90% accuracy on the Devanagari language postcards e and the digit recognition model gives accuracy 99.23% Devanagari numerals and 99.43% accuracy on English numerals.***

Keywords: CNN, MLP, Neural Network, OpenCV, classification

I. INTRODUCTION

Postal automation is an inspiring study topic for many years, and many researchers also publish their work on postal automation of other than Indian language documents [3, 8, 9, 11-15]. Many system modules for automation of postal service are available in developed countries like the USA, France, UK, Australia, and Canada. However, there is less work has been done on the Indian postal system. I am making a postal system that takes an input as a postcard and gives Pincode details and sorts and store the postcard. One of the primary tasks in postal automation is to locate the pin-code box (PCB) and extract the pin-code from this part. In India, a pin-code is a six-digit number uniquely identifying a postal zone. However, there are many difficulties locating PCB on the postcard because the postcard contains various other essential regions such as postage stamp region, address region, graphics. Besides this, there is a wide difference due to several kinds of writing techniques, writing tools, and the format of the various postal cards. Therefore, the discovery of pin-code from the PCB is also a complex problem.

System module building for Indian automation of postal services is more complex than other countries' problems because of multi-language behavior. For example, see Fig.2 and Fig 3**,** which wrote destination address in the English and Devanagari language. So, the model building for the Indian postal services is a complex problem. In this paper, I propose an approach for Indian postal services where, at first, I build a Convolution Neural Network for digit recognition. After this, based on the positional data of PCB region is positioning and extracting from postcards. Then Pincode from the pin-code box is extracted. Finally, checks predicted Pincode in our Pincode directory, sorts and stores the postcard, and shows the postcard's Pincode details. Examples of English and Devanagari numerals show in Fig. 1 to get an idea of the handwriting variability of both English and Devanagari language integers.

English Numerals	0	1	2	3	4	5	6	7	8	9
Devanagari Numerals	०	१	२	३	४	५	६	७	८	९

Fig 1. English and Devanagari Numerals

II. LITERATURE SURVEY

K. Roy [1] and his team proposes a system for the Indian postal service. In this proposed system, they use RLSA Algorithm, and they decompose the image into blocks, and then the non-text block is identified. Also, they think India is a multi-language and can write the address part by inter-mixing two languages. They created a two-stage MLP classifier to identify Bangla and Arabic numerals. At present, the accuracy of the handwritten digit recognition module is 92.10%.

U. Pal [2] and his team proposes a system module under the three-language formula. English, Hindi, and the state's official language are used to write the postal document's address in these three languages. This paper introduced English, Hindi, and Bangla language 6-digit full pin-code string recognition, and they got 99.01% accuracy from the system when error and rejection rates are 0.83% and 15.27%, respectively.

Rabeeh Ayaz Abbasi [3] shows various information extraction techniques that can also be used to address standardization and mainly focus on a statistical model, Hidden Markov Model, and two rule-based methods, RAPIER GRID, that extract information from free text. This paper also discusses some personal experiences to address standardization.

Nabin Sharma [4] discusses the challenges and complexity of developing a postal automation system for a multi-language and multi-script country like India. This paper presents the Deep CNN module for detecting pin-code in structured, unstructured postal envelopes. Region-based Convolutional Neural Networks are used for detecting the many significant regions, namely Pin-code blocks, destination address block, seal, and stamp in a postal document. Zeiler and Fergus, Visual Geometry Group, examined these three network architectures for analysis and identifying their potential.

K. Roy and S. Vajda [5] create a system module for the Indian postal service based on the recognition of Pincode and the city name on the postal document. At first, this recommended system identifies the non-text region and the Address Block from the postal document. After this, from the address region words, and lines are segmented. It is a challenging problem to identify the pin-code piece is where written in the postal document. So, they have used two-stage artificial Neural Network-based classifiers to recognize pin-code digits written in English and Bangla. The Non-Symmetric Half Plane-Hidden Markov Model-based technique is used to recognize city names.

Nosheen Abid [6] and his team presents the trainable neural network-based architecture for postal address system that tackles any issues and can be applied to any Named Entity Recognition problem. This model is trained on a synthetically created dataset and tested on real-world addresses. The system module has also been tested on the NER dataset and gave 90.44% accuracy.

Umapada Pal [7] create a model under the three-language formula: English, Hindi, and the state's official language. The statistical analysis found that 12.37%, 76.32%, and 10.21% of postal letters are written in Bangla language, English language, and Devanagari language script, respectively. They tested this system on 16132 Indian trilingual city names and got 92.25% overall recognition accuracy.

Xiaojie Xia [8] is recommended an efficient handwritten Japanese address recognition system that combines a general ICR engine and special processing functions. The rough recognition result is obtained from an available ICR engine based on an over-segmentation strategy. In the experiments on 670 handwritten Japanese address images, the proposed method successfully divided the entire address into three levels and achieved string level correct rates of 81.6% and 72.5% at a high level.

Edward. J. Kuebert [9] presents a system model for Handwritten address interpretation for United States Postal Service. The information of this technology provided algorithms for control structure recognizers and databases.

Sri Rama Prasanna [10] and his team propose an Automated Postal System that reduces manual mail sorting time along with reducing human errors. They created an Automatic Mail Processor unit (AMP), scans postal documents, and describes the essential fields of the destination address such as the Pin Code, City name, Locality name, and Street name. The system is fulfilled by representing the characters in the form of chain codes and using their Fourier Descriptors for alphanumeric matching with the aid of a Neural Network.

Chin Keong Lee [11] describes an algorithmic approach for optical character recognition system integrated with a Singapore handwritten address interpretation system. This recommended system creates multiple hypotheses of postcodes, verifies the hypotheses using a postal dictionary, and uses the address features to choose among the hypotheses. The performance of this system on a set of 450 fictitious but realistic handwritten mail pieces show an improvement in letter sorting performance from 81.6% correct, 6.0% reject 12.4% error using OCR alone on the postcode to 70.9% accurate; 28.5% reject; 0.7% error describing a vital change in the error rate.

M.F.A. Ifhaam [12] proposed a system for Sri Lankan post office automation. The majority of Sri Lankans use the Sinhala language for their day-to-day activities. In this module, they used a Genetic Algorithm to generate more optimized results faster with higher accuracy. The algorithm shows an accuracy of over 92% for addresses that are recognized with three misrecognized characters. This algorithm can be used in practice scenarios as the AI Recognition has more than 79 % accuracy.

Mustain Billah [13] has been proposed Bangla handwritten digit recognition based on a post office automation system for Bangladeshi Post offices. This proposed system automatically sorts mails according to the postcode. Thus the system can save time and money, reducing the necessity of manual sorting. Furthermore, for recognizing purposes, Multi-Layer Neural Network is applied. As a result, this proposed system gains a higher accuracy of 99.71%.

Yih-Ming Su [14] introduced an automatic postal document-sorting system based on creating a real-time OCR system. This system handles Chinese mail to identify the city and country names in Taiwan on handwritten as well as machine-printed standard Chinese style postal document. Preliminary results proved that the system could sort 5400 mail pieces per hour with a correct rate of 5.6% and an error rate of 0.92%.

R.J.N. Kalberg [15] shows a mail sorting system for addresses on Dutch mail. The system uses grammar to determine the syntax of an address. The choice of the postal code is tested using a test set of 7876 images of addresses from live characters. In 54% of the images, selected the postal code successfully, and 42% of images were rejected; 20% of images were correct because the address did not contain a postal code. In 4% of the cases, the algorithm misclassed a word as a postal code.

A. Comparison Of The Method With Accuracy

Here table1 shows some comparison of models that they have invented with their accuracy.

Algorithms	Accuracy
The run-length smoothing algorithm	92%
Multi-Layer Neural Network	99.71%
Genetic Algorithm	79 %
Neural Network	90.4

Table 1 Compares algorithms with accuracy

III. METHODOLOGY

A. Data Collection

Postcard digitalization for the current work has been done from picked images from mobile cameras for digitalization. The photos are in color format and captured by the 64mega-pixel camera, and stored in jpg Format.

Fig 2 English postcard

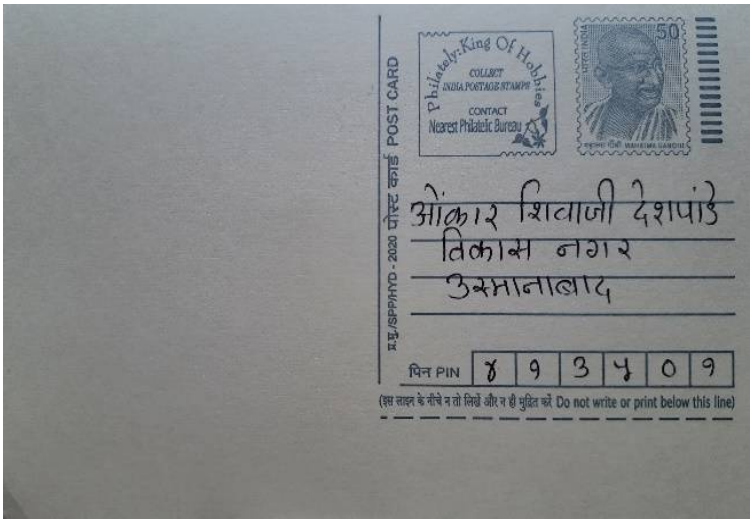


Fig 3 Devanagari postcard

B. Pin-Code Box Detection And Extraction

In Indian postal cards, there are pre-printed boxes to write pin-code. We call these boxes pin-code boxes. People commonly write pin-code inside those boxes. To extract the pin code box, we apply a Gaussian blur filter to reduce noises. After this Adaptivethreshold filter to get threshold next step is to detect rectangular regions of the Pincode. The following fig shows the extraction of the pin block of fig 2.

Fig. 4 extraction of the pin code block

C. Pin Code Digit Detection

After the extraction of the PCB block next step is to remove horizontal and vertical lines in the Pincode box.

Fig 5 Removal of horizontal and vertical lines

After removing horizontal and vertical lines, the next step is to use counters to get digit x, y position, and we use the bounding rectangular function to get digit boundaries.

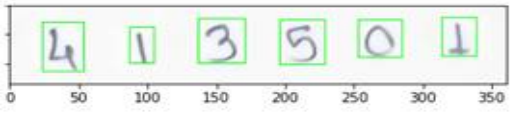


Fig 6 digit detection

D. Numeral Recognition

After extracting digits from the Pincode box of a postcard, continue with their recognition. For recognition, do not compute any feature from the postcard. Convolution Neural Network for recognition uses a fixed size of input; we normalized the digit first to a 28x28 pixel size. Above bounding boxes resizes to 28*28 sizes, and after normalizing this, we pass to our model and predict which digit it is. Fig. 7 is the output of fig 6

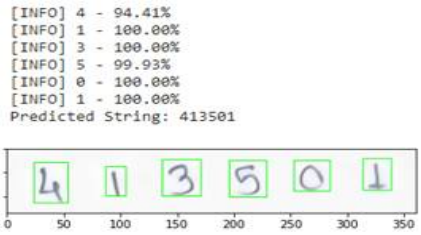


Fig. 7 prediction

E. Predicting Pincode and Storing

After predicting the Pincode, the Pincode is checked in the Pincode directory and shown the postal details as well as sort and store postcard according to their Pincode.

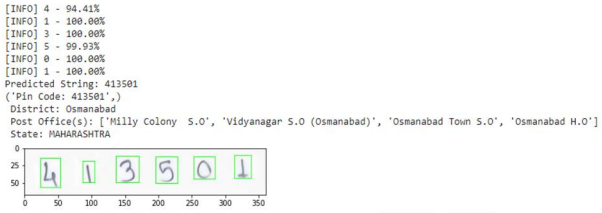


Fig 8 getting Pincode details

IV. FLOW CHART

Fig. 9 is the pictorial presentation of system flow. First, the system takes input as an image then removal of noises. The next step is getting the Pincode block after this removal of horizontal and vertical lines. The next step is getting digit positions in the Pincode box; after this, we predict the digits. The next step is to check an expected string in our Pincode directory. If the string is matched, show the

Pincode details like district, states, and post office circles and the last step is storing postcards Pincode-wise with respective their directories.

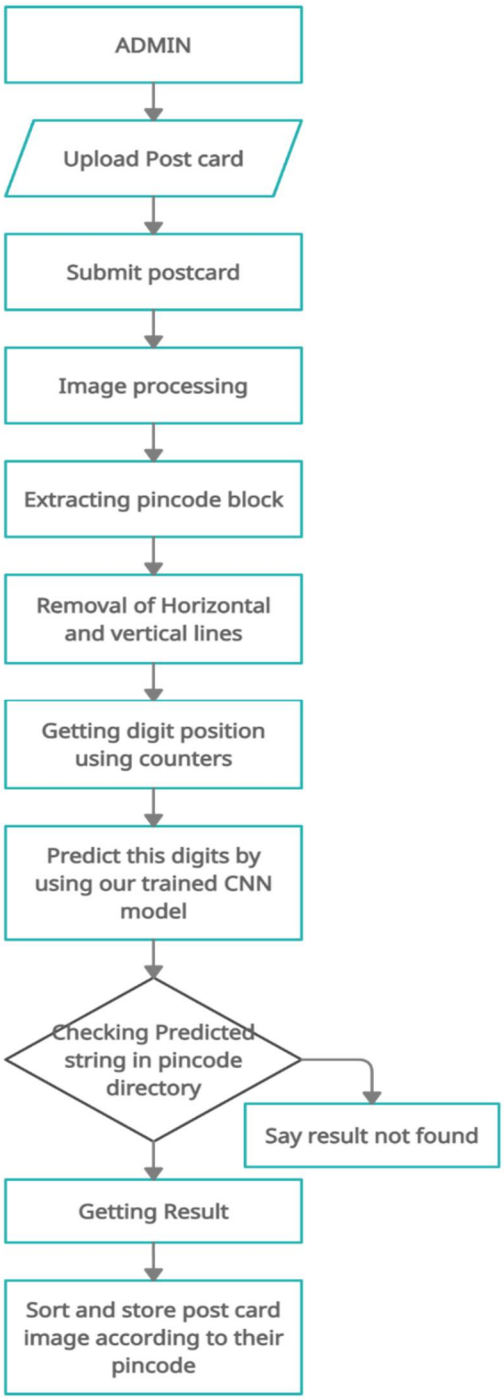


Fig 9. Flowchart of system

V. RESULT

I tested this system on 47 images in that 27 are English, and 20 are Devanagari postcards.
Figure10 shows the correct result of 411037 Pincode and gives relevant details like district, post office, state, etc.

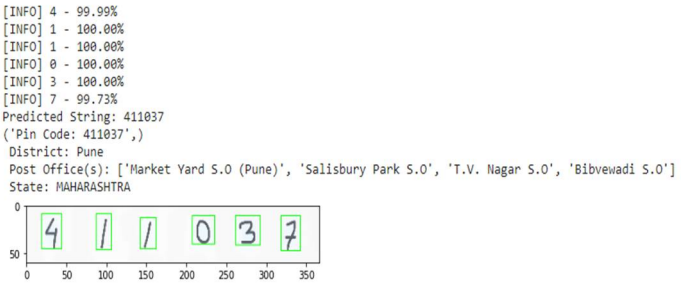


Fig 10 Pincode details

In figure 11, the 230532 Pincode is not any post office code or not present in our directory, so the result is not found.

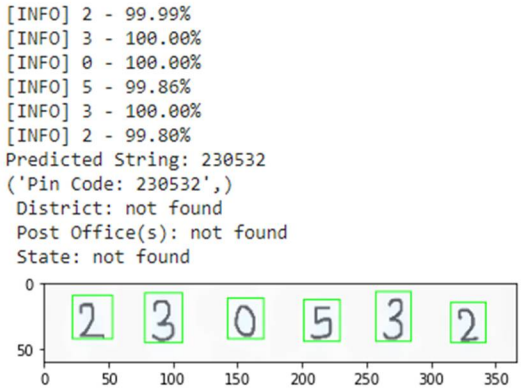


Fig 11. if Pincode not present in the directory

Figure12 is devanagari language postcard image and shows the correct result of 413501 Pincode and gives relevant details like district, post office, state, etc

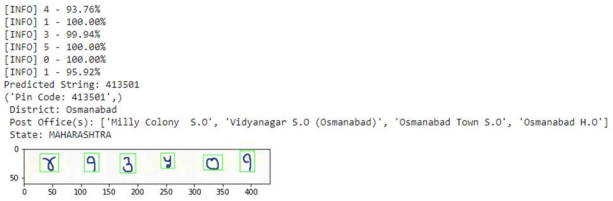


Fig 12. Devanagari Pincode prediction

Figure13 is printed postcard image and shows the correct result for 413501 Pincode and gives relevant details like district, post office, state, etc

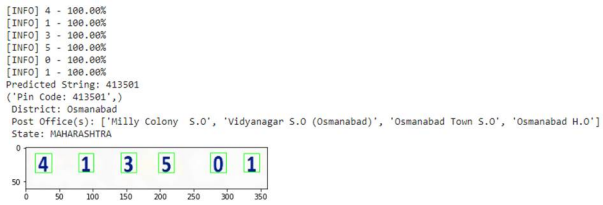


Fig 13. Printed postcard Pincode prediction

Our model gets 92.59% accuracy in the English language and 90% accuracy in the Devanagari language, as shown in Table 2

	Total cards	Correct predicted	Accuracy
English	27	25	92.59%
Devanagari	20	18	90%

Table 2 Accuracy Table

VI. FEATURE SCOPE

The project currently has a simple GUI based on Tkinter, but it can be made much more user-friendly and easily navigable by using many other modules. After increasing the Efficiency and Integrating the project to a RaspberryPi will be more practical and handy to use. The idea of this project can be implemented in Post Offices to categorize the postcards according to their destination automatically.

The faster the processor, the quicker one will get the result so that a faster processing speed can be used in practical cases. Currently, the project is only giving District, State, and Post Office from CSV files but other things like Post office's phone number, division, circle, taluk, etc., can also be printed by modifying the code accordingly.

VII. CONCLUSION

In this study, I am detecting a pin-code box on postcards modeled as a standard object detection problem. Analysis of Pincode box detection technique is explored in this paper to understand its importance for Indian Postal Automation. With the complexity concerned with sorting Indian postal cards in real-time, the use of CNN looks very promising. Using a single pipeline to detect pin-code boxes on a postal document is also examined for the first time, to the best of our knowledge. The results collected from the test cases are very encouraging. The system is also able to detect handwritten Pin-codes as well as printed Pin-code and Pincode box successfully. This is the first study that considers CNNs to develop a real-time Indian Postal automation system to benefit our information. The outcome of the present study is the basis for our future research. Correction of the CNN architecture and multi-language OCR for the Pin-codes will also be considered. I am preparing a larger dataset for further experiments and will surely improve the performance of the CNN networks.

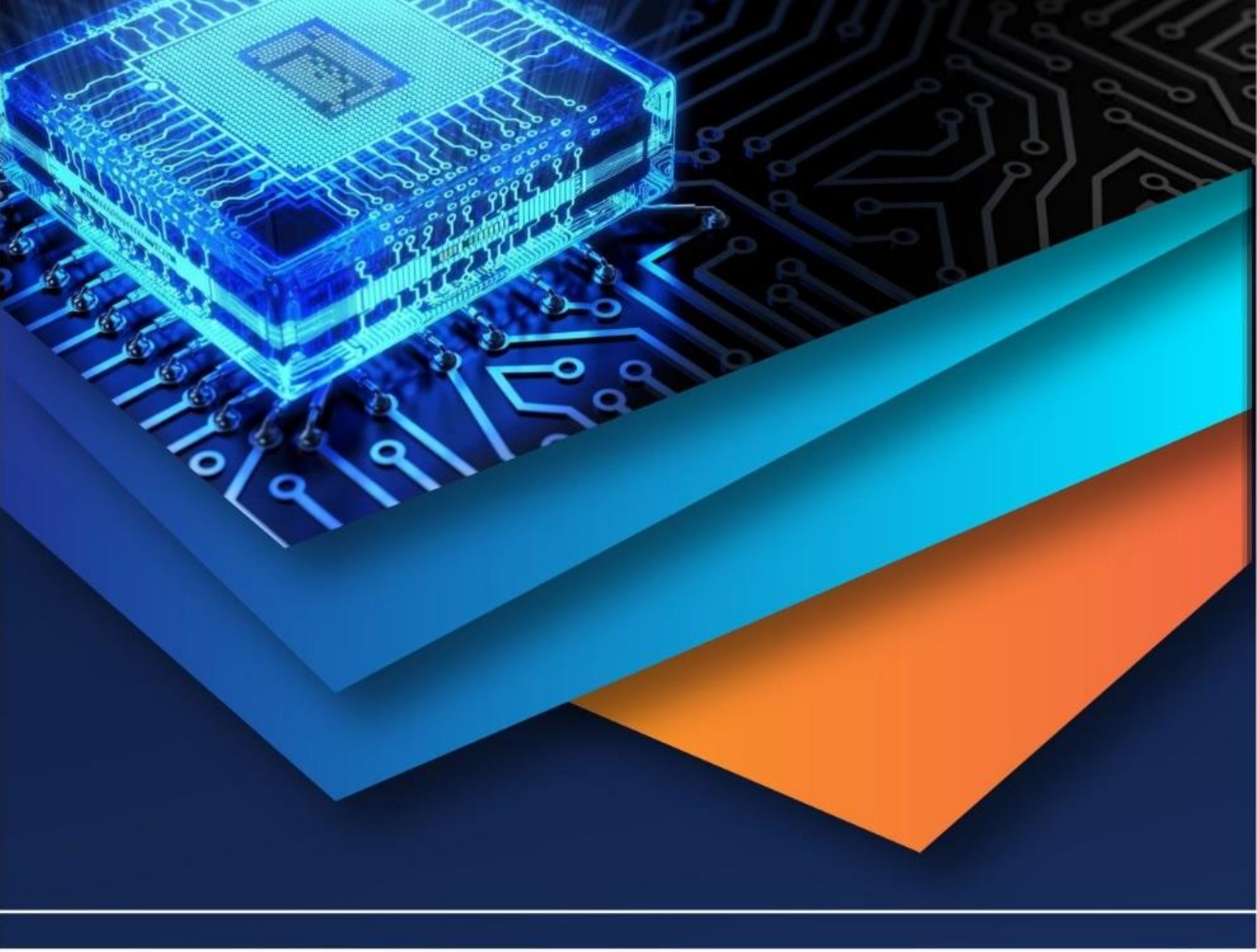
REFERENCES

[1] K. Roy, S. Vajda, U. Pal (December 2004), "A System towards Indian Postal Automation," Ninth International Workshop on Frontiers in Handwriting Recognition, 10.1109/IWFHR.2004.12 URL: https://ieeexplore.ieee.org/document/1363974
[2] U. Pal1, R. K. Roy1, K. Roy2 and F. Kimura3 (October 2009), "Indian Multi-Script Full Pin-code String Recognition for Postal Automation," 2009 10th International Conference on Document Analysis and Recognition, URL: https://ieeexplore.ieee.org/document/5277632
[3] Rabeeh Ayaz Abbasi(March 2007), "Information Extraction Techniques for Postal Address Standardization," URL: https://ieeexplore.ieee.org/document/4133470
[4] Nabin Sharma, Abira Senguptay, Rabi Sharmay, Umapada Paly and Michael Blumenstein (July 2018), "Pincode detection using Deep CNN for Postal Automation," 2017 International Conference on Image and Vision Computing New Zealand (IVCNZ), URL: https://ieeexplore.ieee.org/document/8402501
[5] K. Roy, S. Vajda*, U. Pal, B. B. Chaudhuri and A. Belaid(January 2006), "A System for Indian Postal Automation," Eighth International Conference on Document Analysis and Recognition (ICDAR'05), URL: https://ieeexplore.ieee.org/document/1575706
[6] Nosheen Abid_, Adnan ul Hasany, and Faisal Shafait_y, "DeepParse: A Trainable Postal Address Parser," URL: https://ieeexplore.ieee.org/document/8615844
[7] Umapada Pal, Ramit Kumar Roy, Fumitaka Kimura (February 2013), "Multi-lingual City Name Recognition for Indian Postal Automation," 2012 International Conference on Frontiers in Handwriting Recognition, URL: https://ieeexplore.ieee.org/document/6424387
[8] Xiaojie Xia, Xiaoyi Yu, Wei Liu, Chunyan Zhang, Jun Sun, Satoshi Naoi(August 2002), "An efficient off-line handwritten Japanese address recognition system," 2019 International Conference on Document Analysis and Recognition (ICDAR) URL: https://ieeexplore.ieee.org/document/8977985
[9] Sargur. N. Srihari, Edward. J. Kuebert(August 2002), "Integration of Hand-Written Address Interpretation Technology into the United States Postal Service Remote Computer Reader System," Proceedings of the Fourth International Conference on Document Analysis and Recognition, URL: https://ieeexplore.ieee.org/document/620640
[10] P. Sri Rama Prasanna, S.Balaji, Thejavor Haralu Khezhie, C.Vasanthanayaki, SAnnadurai (March 2004), "Destination Address Interpretation For Automating the Sorting Process Of Indian Postal System," TENCON 2003. Conference on Convergent Technologies for Asia-Pacific Region, URL:https://ieeexplore.ieee.org/document/1273300
[11] Chin Keong Lee and Graham Leedham(August 2002), "Empirical Design of a Holistic Verifier for Automatic Sorting of Handwritten Singapore Postal Addresses," Proceedings of the Fifth International Conference on Document Analysis and Recognition. ICDAR '99 (Cat. No.PR00318), URL:https://ieeexplore.ieee.org/document/791892
[12] M.F.A. Ifhaam, S. Jayalal (September 2019), "Sinhala handwritten Postal Address Recognition for Postal sorting," 2019 International Research Conference on Smart Computing and Systems Engineering (SCSE), URL: https://ieeexplore.ieee.org/document/8842746

[13] Mustain Billah, Md. Kamruzzaman Ruman, Nazmus Sadat, Md. Mahfuzul Islam(April 2019), "Bangladeshi Post Office Automation System Using Neural Network," 2019 International Conference on Electrical, Computer and Communication Engineering (ECCE), URL: https://ieeexplore.ieee.org/document/8679350

[14] Yih-Ming Su and Jhing-Fa Wang (August 2002), "Recognition Of Handwritten Chinese Postal Address Using Neural Networks," URL: https://ieeexplore.ieee.org/document/703885

[15] R.J.N. Kalberg, G.H. Quint, H. Scholten (August 2002), "Automatic interpretation of Dutch addresses," 11th IAPR International Conference on Pattern Recognition. Vol. II. Conference B: Pattern Recognition Methodology and Systems, URL: https://ieeexplore.ieee.org/document/201794



10.22214/IJRASET

45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY