iJRASET
International Journal For Research in Applied Science and Engineering Technology

# INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

**Volume:** 11 **Issue:** VI **Month of publication:** June 2023

**DOI:** https://doi.org/10.22214/ijraset.2023.54383

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: 7.538*
*Volume 11 Issue VI Jun 2023- Available at www.ijraset.com*

# Digital Image Text Recognition Using Machine Learning Algorithms

Chaitanya U[1], Emmanuel Alisetti[2], Harsitha Ballam[3] Maneesha Dodda[4]
[1]*Assistant Professor,* [2, 3, 4]*UG student, Department of Information Technology, Mahatma Gandhi Institute of Technology*

***Abstract:*** ***Text recognition is used in various fields like document analysis, picture labeling, and text analysis, etc. The process of recognizing text from image is very crucial.. Employing Maximally Stable Extremal Regions (MSER) and Optical Character Recognition (OCR) algorithms have drastically changed the accuracy and reliability of text detection from photos. Our proposed work suggests a model for text recognition that makes use of both the MSER and OCR algorithms' benefits for the purpose of accuracy improvement and reliability of the digital text extraction procedure. OCR follows cutting-edge techniques to distinguish and identify individual characters, serving as the essential building block for text recognition. However, these algorithms may struggle when dealing with complex backdrops, blurry photos, or text that is structured in an atypical way. We use MSER approach, which excels at recognizing text sections by finding maximally stable regions across various severities and scales, to solve these constraints.***

***The suggested model employs a multi-stage methodology. The MSER algorithm is used for extracting the likely text spots from the input image first. To boost OCR performance, these zones are then fine-tuned using pre-processing techniques including noise reduction and picture enhancement. The OCR system next processes the cleaned-up sections, identifying each region's text using machine learning and pattern recognition methods. The text that is recognized is then further processed to increase the accuracy and refine these findings. Thus, compared to most other models, the text recognition model that is built utilizing the MSER and the CNN (OCR, a component of CNN) algorithm performs better.***

***Keywords: Text recognition; Convolution neural network; Optical character recognition; Maximally Stable Extremal Regions; Text localization;***

## I. INTRODUCTION

Due to the abundance of technologies taking images and saving them is a common activity that is seen all around us in the modern digital world. These images may include text, figures, or other types of visual data. Usually, the snaps of those photos, we save them for later use. But the issue emerges when we try to extract the information from those images and when we try to reuse them. These photographs can be examined manually and withdraw the text from those images, which is beneficial if there are only a limited number of images. However, it becomes challenging when the count of images is too high and we want to extract the digital text from those huge number of pictures, as this requires a lot of manpower. Therefore, a system that can automatically extract text from photos is necessary for this purpose. These images could have been taken from bank checks, house number plates, vehicle licence plates, etc.

Text extraction from photographs is not a simple operation, there are numerous complexities involved, such as text font-related concerns, image quality issues, backdrop colour issues, etc. Therefore, in order to recognize the text from those photos by getting the control of all these difficulties, a computer-based model must be built. The most important thing to accomplish first is to separate the text from the scene image's background.

## II. LITERATURE SURVEY

Zhang, et al. proposed a text detection system in natural scene images based on colour prior guided MSER Digital text extraction from naturally scene images is a very challenging task. So to create textures that resemble strokes, we modify the Stroke Width Transform (SWT) and introduce colour for text candidate extraction [1]. Text verification uses deep learning to distinguish between non-text text candidates and text candidates. To improve accuracy of classification, results of many CNN tasks are combined.

Wenyuan, et al. proposed a deep learning based text recognition model for extracting the text from laboratory reports of patients. Deep learning is utilized for extraction of textual information from medical laboratory reports data [2]. This also contains the historical medical records of patients and also record of patients currently in their care.

S.Y.Arafat, et al. proposed a text recognition model for the Urdu ligature. Being a non-Latin script it might be challenging to identify and locate the Urdu text it in images of natural scenes. So to solve this issue a mechanism was introduced for detection, anticipation of the orientation and recognition of the Urdu ligature in that images [3]. In addition to CNNs like Squeeze Net, Resnet50, and Google Net, the Faster RCNN algorithm was used. On datasets including ligatures that were randomly oriented, a customised RRNN is trained and evaluated for ligature orientation prediction.

Perepu Pavan Kumar, et al. proposed a deep learning based model to extract text from natural scene images. Photos of natural scenes pose a challenging difficulty for digital text extraction. CNN's are being used to assess whether the digital text is light on a dark background or vice versa, and some deep learning methodologies are employed to examine the text polarity [4]. With the use of feature deletion, CNNs were trained using the sample images taken from benchmarking dataset..

Huang, et al. proposed an end-to-end trainable system called the Effective Parts Attention Network (EPAN) encodes contextual information using a character effective parts decoder (CEPD) and a text image encoder [5]. It includes a text image encoder and a CEPD, and it is trainable end-to-end. Even though comprehensive medical records are not necessary, electronic health records (EHRs) have emerged as a crucial development in contemporary medicine.

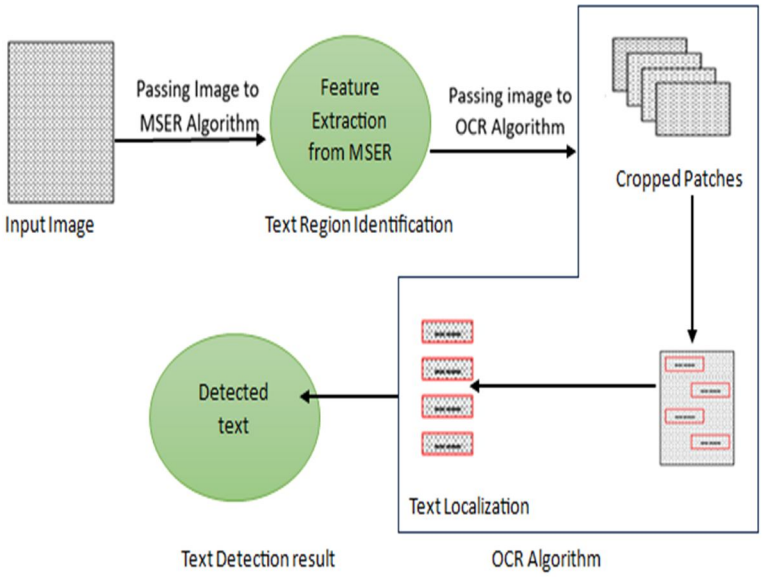



Fig.1. Propose architecture for text recognition model

## III. PROPOSED SYSTEM

Recognition of text from natural scene images is not an easy task due to various difficulties like images may be blurry, background may be dull, etc. So a text recognition system was proposed for the purpose of extracting digital text from these images by facing all these difficulties. The proposed system initially takes an input image and processes it using MSER algorithm where the text spots and identified and return rectangular bounding boxes where the text is present. This mainly helps in enhancement of the picture quality and noise reduction from the image.

MSER is an algorithm that detects stable regions with similar intensity values in images, allowing for the identification of text. It excels at handling complex backgrounds and varying lighting conditions.

Then, Post-processing the image is passed to OCR algorithm for further text identification by using machine learning and pattern matching techniques. CNN a deep learning algorithm utilizes convolutional layers to extract image features, which are then processed by a classification layer to detect text. CNN surpasses traditional methods, particularly in recognizing distorted or rotated text.

Each algorithm has its own strengths and weaknesses, with the choice depending on specific application requirements. MSER offers computational efficiency and excels in complex backgrounds, while CNN demonstrates superior accuracy for distorted or rotated text.

### A. Dataset Description

SVHN(Street View House Numbers) dataset is one of two used to train the proposed model. This file includes several digitised photographs that were taken from Google Street View pictures. The SVHN dataset consists of 531,131 additional training pictures in addition to the 732,000 training photos and 26,032 testing images. The collection's images each has resolution of 32x32 pixels in RGB (colour) format and a label that identifies a set of visible numbers. The labels may include one digit or several (up to five digits), depending on the number depicted in the image. From this dataset, we have only utilised up to 14,560 photos for the training purpose and up to 1000 images to assess the effectiveness of the model. For this dataset, the average accuracy for every image evaluated was 90%.

This SVHN dataset was obtained from Kaggle.com and utilised as well, is another one that is highly suited for OCR models. The total dataset can reach 6.7GB in size. However, only 2376 of the dataset's photos are used to train any model that we need.

### B. Image Pre-Processing

Here, picture visualisation and presentation is the pre-processing method employed. Subplots are initially constructed using the matplotlib library's subplots () method. Subplots allow for the grid-style display of several images. The fig size option specifies the size of the figure as well as the number of columns and rows for the subplots.

The axes of the subplots are then flattened using the flatten () function. By flattening the axes, a single loop may be used to iterate across them.

Then, the programme begins to loop over each subplot. Throughout each cycle, a picture is loaded and shown using the matplotlib im-show () function. The im-show () function accepts the picture that was acquired after the image file was read using the im-read () function as an input. Using the axis('off') function, labels will be removed from the figure when the axis ticks.

Here, query () method is useful to determine the count of annotations associated with the picture, and the picture id is extracted by using the image file name. This information is used for setting the title of each subplot using the set title () function. By providing the picture id and the number of annotations, the title provides more information and context about the image.

### C. Architecture

In our proposed model as shown in Fig.(1) , the digital text is taken from those photos that are collected randomly and might be taken from address signs, licence plates on cars, house numbers, posters, etc. The system will receive these photographs, which were taken at random, as input so it may extract text from them. Text will be extracted from the input images that have been cropped. MSER algorithm is utilized for the process of digital text localization. This is a two-step process. Where in text localization the area with more text concentration in the entire image is discovered. We need to identify the patches where the text is most concentrated in order to use those patches in the subsequent process of text extraction.

Later in Fig (1), Post the text localization process the text localization patches are passed to the CNN algorithm as input, So that CNN algorithm extracts the text present in those images. OCR is used to detect the text from the cropped patches by implementing the CNN. CNN classifier, which has convolutional layers that are primarily utilised for extracting the features using a neural network that is fully connected and can implement the conventional architecture for OCR.

### D. MSER Algorithm

MSER approach locates areas in a image that are stable and has consistent intensity or colour characteristics at different scales. These regions are known as Extremal Regions (ER). This proposed MSER algorithm works by finding the threshold values of the image at varying intensities and isolating the connected regions that have nearby threshold vaues. By analysing changes in these connected components across many ways, the technique identifies regions that are stable and consistent.

The MSER approach has been used to successfully solve a number of computer vision tasks, including text detection, image segmentation, and object recognition. Because of its usefulness and power, it offers a trustworthy and efficient method of detecting regions of text.

### E. Working principle of MSER Algorithm

MSER firstly takes input images and calculates the intensity or colour difference in the picture. The local variations in intensity or colour are mapped out in this difference image by deducting the value of the colour or current pixel's intensity from the values of the pixels around it.

The program then analyses the different image starting at the lowest threshold and moving up to a few more threshold values. For each threshold level, the programme identifies linked areas that satisfy particular stability requirements. These areas are referred to as Extremal Regions (ERs) as they show areas of the picture with constant brightness or hue.

To determine a region's stability, the MSER algorithm searches for areas that remain linked over a range of thresholds. Each threshold level's difference from the one preceding it in terms of region size is evaluated. Regions that alter or stay steady beyond thresholds with little to no change are said to be maximally stable.

The union-find data structure allows the algorithm to maintain track of connections between pixels and regions, ensuring that related regions are appropriately discovered.

As the algorithm advances through the threshold levels, it builds a hierarchy of ideally stable zones. This hierarchy provides information on the size and stability of each zone and shows how regions vary as the threshold changes.

The collection of corresponding stability levels of the maximally stable areas, and other relevant data are included in the MSER algorithm's output. These regions can be further processed or used to several tasks, such as object detection, picture segmentation, and feature extraction.

### *F. OCR Algorithm*

OCR technologies convert printed or handwritten text into digital, editable representations. OCR has crucial importance in various applications, such as documents digitalization, automated data entry and text extraction. The needs of the application determine the OCR algorithm to use. OCR algorithm complexity might vary. Many modern OCR systems combine many strategies to generate high text recognition task accuracy and resilience, including character recognition algorithms based on deep learning.

### *G. Working principle of OCR Algorithm*

1) The OCR algorithm uses a layered architecture to extract digital text from images. These layers make the model to learn and process the input images to recognize the text. These layers are very crucial in extracting the digital text from the input images.
2) The OCR algorithm is implemented using the convolutional layers for extracting the relevant features from the input images where these layers identify the visual features and patterns.
3) Initially the pooling layer helps in reducing the spatial dimensions like width and height of the identified features or characters.
4) Recurrent layer is further used in the next step after reducing the spatial dimensions which is responsible for recognizing the characters in proper order and also to handle the texts having variable lengths.
5) The output layer is the final layer of the OCR model which performs the text recognition process. Where the identified characters will be decoded and matched with the corresponding characters and final output is returned.

## IV. RESULTS

The process of digital text extraction from images using the MSER and OCR algorithm can be done initially the input image is processed by the MSER algorithm post-processing it returns the bounding boxes showing the text present in the image , this is called as text localization. After the localization of text areas then the image is further processed using the OCR algorithm where the text is extracted from the localized regions.


Fig.2 Image consists of alphabets

Fig.2 shows an image consists of only alphabets, a signboard randomly taken on a street is passed to our proposed system, and the entire text present in that image is recognized by indicating each word using red coloured annotations showing the word it is representing.


Fig.3 Image consisting of only numbers

Fig.3 shows the image consisting of numbers, where this image is taken from the SVHN data set the number present in that image is identified along with the annotations showing the number.


Fig.4 Image consisting of both numbers and alphabets

In Fig.4 the above image shown the input image consisting of both the numbers and alphabets and the recognized text is indicated using the annotations.

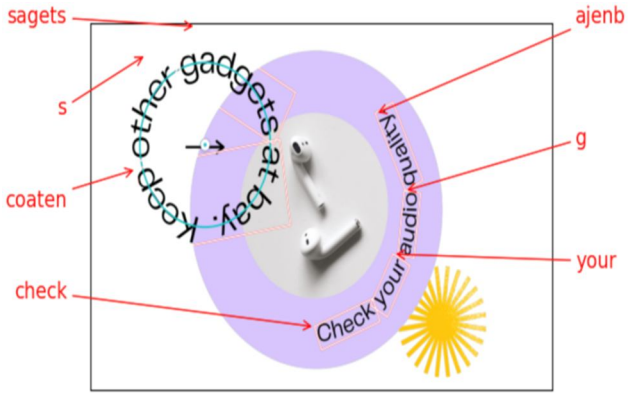



Fig 5. Image consisting of oriented text

Fig.5 showing an image consists of oriented text. Only a few words that are oriented are being detected correctly indicated using the annotations. The accuracy of this image with oriented text is reduced.

The model's performance is measured by calculating the accuracy of the text recognition system where the Equation(1) represents the formula for calculating the accuracy. The number of words that are matched correctly are taken as Matched no.of words indicated in the numerator. The total count of words present in the image is represented using Total no.of words shown in denominator.

$$\text{Accuracy} = \frac{\text{Matched no.of words}}{\text{Total no.of words}} * 100 \quad (1)$$

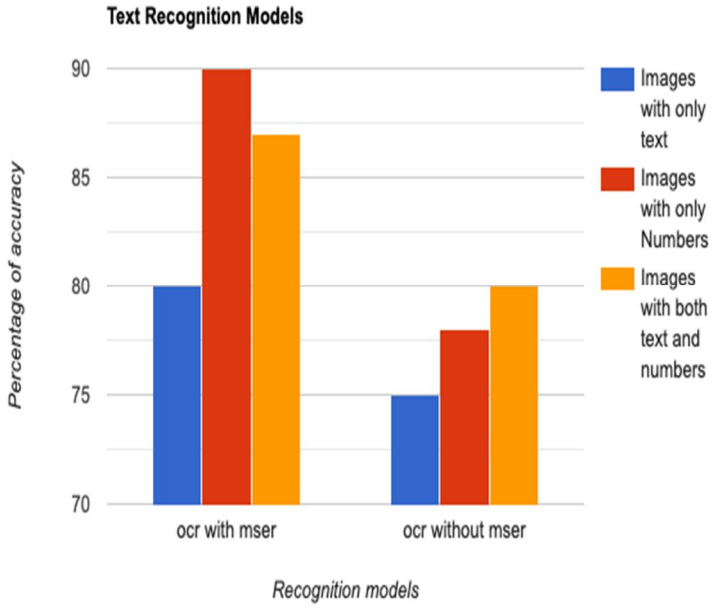



Fig.6 The bar graph of text recognition models

TABLE-I
ACCURACY COMPARISION OF OUR PROPOSED MODEL USING MSER AND WITHOUT USING MSER

| Machine Learning Models | Image consisting of only alphabets | Image consisting of only numbers | Image consisting both numbers and alphabets |
|---|---|---|---|
| OCR along with MSER algorithm | 80 | 90 | 87 |
| OCR without MSER algorithm | 75 | 78 | 80 |

In Fig. 6 we have taken two cases,where the OCRis used with the combination of MSER algorithm and OCR is used without the combination of MSER algorithm, in these situations we consider threee different senarios like by considering different types of images like image consisting of only alphabets, image consisting of only numbers and image consisting both numbers and alphabets. The percentage of accuracy assessed using the accuracy Equation (1) with several models in three distinct circumstances is shown by the numbers. The results are more accurate when we used the MSER algorithm and text present in pictures is more easily recognised, But the accuracy is reduced when the text is detected without using MSER , they are less precise than ones that use the MSER method. The parameter employed here to assess the yield of findings is accuracy.

In the above mentioned scenarios, the accuracy % is measured in both the cases and compared and shown on the y-axis. The OCR model is implemented as a combination of MSER technique and also implemented without using MSER technique represented in this models on x-axis.

The results are shown in Table 1 as a percentage of accuracy. These figures are calculated by averaging the accuracy of 50 photos from each category. The accuracy of the first 50 text-containing pictures is determined for each image using both MSER and non-MSER methods, and the average of these accuracies is then determined. The same is true for photos that simply include numbers and for images that have both text and numbers.As a result, we may infer from table (1) that an OCR-based text recognition model by itself does not demonstrate adequate accuracy. But when OCR and MSER are coupled, it produces excellent outcomes with high accuracy.

## V. CONCLUSION

The proposed model introduces a method for extracting digital text from images or scanned documents. It combines two algorithms, namely OCR and MSER, to achieve accurate results. The first step in the process involves using the MSER algorithm to analyze the input image containing text. MSER identifies the text regions and generates bounding boxes that precisely enclose these areas. These bounding boxes are rectangular in shape and serve as an accurate representation of the text's location within the image. In this image containing the original text along with the bounding boxes obtained from the MSER algorithm, is then fed into the OCR algorithm. The OCR algorithm undergoes several phases, including text localization and text detection, to process the image and extract the text present in the identified areas. Finally, the text that is recognized is presented as the output of the model. By combining the MSER and OCR algorithms, the proposed model achieves a higher level of accuracy compared to using the OCR algorithm alone. It is particularly effective when the input image consists of alphabets, numbers, or a combination of both. However, the accuracy of the model decreases when the characters in the image are oriented or skewed.

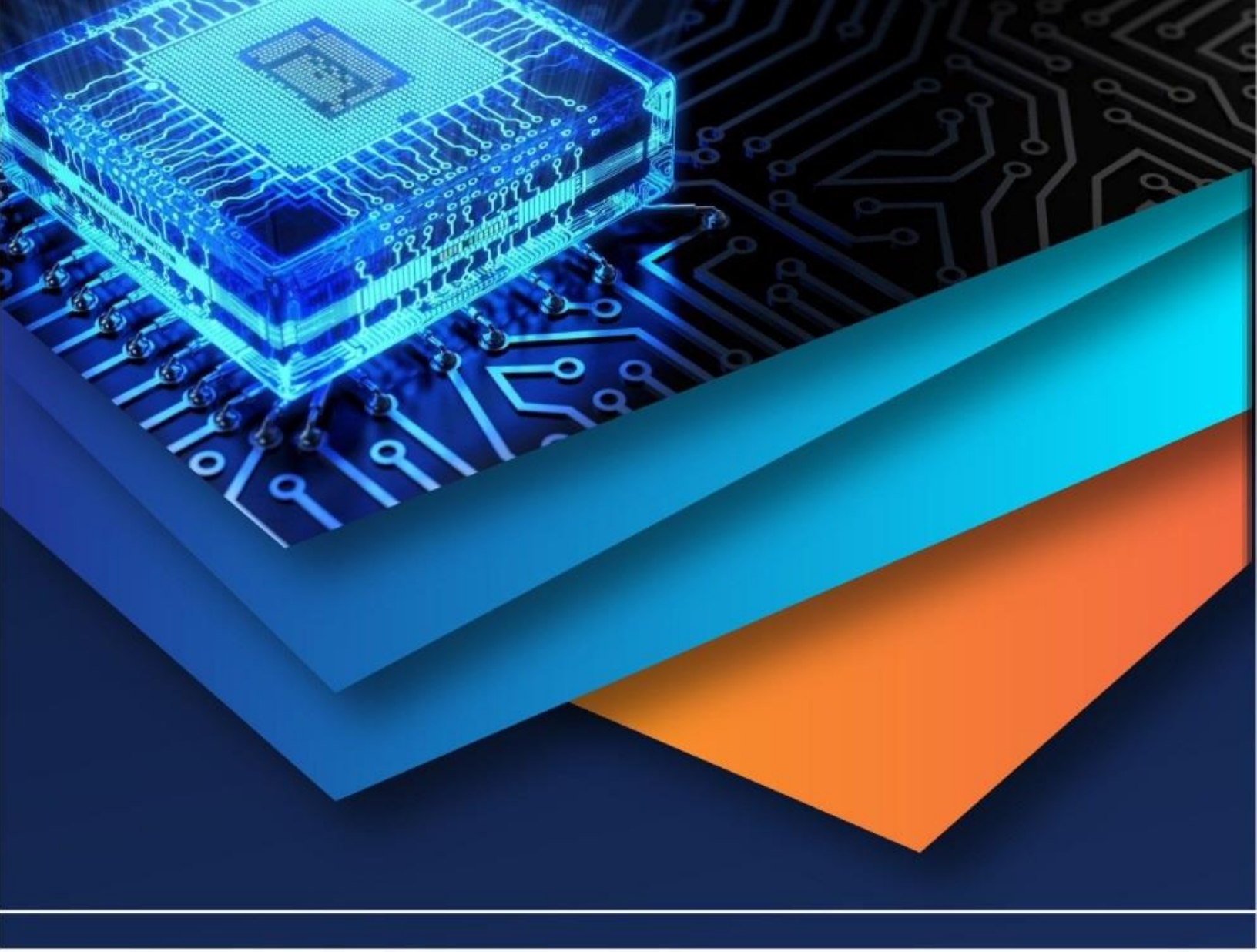
The proposed model empowers the transformation of scanned images into digital text. It finds applications in various domains such as document digitization, number plate detection, scanning of passport and id scanning, invoice scanning, receipt, and data entry automation, as well as mining of text and text analysis. Its versatility makes it a valuable tool for streamlining processes and extracting meaningful information from visual data.

The proposed model currently accepts image inputs for text extraction, and it has the potential to expand its capabilities to recognize text with special characters, different font styles, text orientations, and languages not only English, But also others such as Tamil, Kannada, and Sanskrit. However, future extensions can involve incorporating video inputs, enabling the detection and extraction of text from the video form of input. Furthermore, the model could be utilized for recognizing incomplete letters or letters obscured by shadows. For real-time applications, the OpenCV module in Python can be employed and additionally, intelligent error correction can be implemented using other algorithms for spell-check.

## REFERENCES

[1] Zhang, Xiangnan, Xinbo Gao, and Chunna Tian. "Text detection in natural scene images based on colour prior guided MSER." Neurocomputing 307 (2018): 61-71.
[2] Xue, Wenyuan, Qingyong Li, and Qiyuan Xue. "Text detection and recognition for images of medical laboratory reports with a deep learning approach." IEEE Access 8 (2019): 407-416.
[3] Arafat, Syed Yasser, and Muhammad Javed Iqbal. "Urdu-text detection and recognition in natural scene images using deep learning." IEEE Access 8 (2020): 96787-96803.
[4] Perepu, Pavan Kumar. "Deep learning for detection of text polarity in natural scene images." Neurocomputing 431 (2021): 1-6.
[5] Huang, Yunlong, Zenghui Sun, Lianwen Jin, and Canjie Luo. "EPAN: Effective parts attention network for scene text recognition." neurocomputing 376 (2020): 202-213..
[6] U. Bhattacharya, S. K. Parui and S. Mondal, "Devanagari and Bangla Text Extraction from Natural Scene Images," 2009 10th International Conference on Document Analysis and Recognition, Barcelona, Spain, 2009, pp. 171-175, doi: 10.1109/ICDAR.2009.178
[7] J. Memon, M. Sami, R. A. Khan and M. Uddin, "Handwritten Optical Character Recognition (OCR): A Comprehensive Systematic Literature Review (SLR)," in IEEE Access, vol. 8, pp. 142642-142668, 2020, doi: 10.1109/ACCESS.2020.3012542.
[8] K. S. Satwashil and V. R. Pawar, "Integrated natural scene text localization and recognition," 2017 International conference of Electronics, Communication and Aerospace Technology (ICECA), Coimbatore, India, 2017, pp. 371-374, doi: 10.1109/ICECA.2017.8203708



10.22214/IJRASET

45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429

# INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY