

# Transforming images into words: optical character recognition solutions for image text extraction

Jyoti Wadmare<sup>1</sup>, Sunita Ravindra Patil<sup>2</sup>, Dakshita Kolte<sup>1</sup>, Kapil Bhatia<sup>1</sup>, Palak Desai<sup>1</sup>,  
Ganesh Wadmare<sup>3</sup>

<sup>1</sup>Department of Computer Engineering, K.J. Somaiya Institute of Technology, Mumbai, India

<sup>2</sup>SVKM's Narsee Monjee Institute of Management Studies, Dhule, India

<sup>3</sup>Department of Artificial Intelligence and Data Science, K.J. Somaiya Institute of Technology, Mumbai, India

## Article Info

### Article history:

Received Mar 5, 2024

Revised Feb 13, 2025

Accepted Mar 15, 2025

### Keywords:

Named entity recognition

Natural language processing

Optical character recognition

Text extraction

Text recognition

## ABSTRACT

Optical character recognition (OCR) tool is a boon and greatest advancement in today's emerging technology which has proven its remarkability in recent years by making it easier for humans to convert the textual information in images or physical documents into text data making it useful for analysis, automation processes and improvised productivity for different purposes. This paper presents the designing, development and implementation of a novel OCR tool aiming at text extraction and recognition tasks. The tool incorporates advanced techniques such as computer vision and natural language processing (NLP) which offer powerful performance for various document types. The performance of the tool is subject to metrics like analysis, accuracy, speed, and document format compatibility. The developed OCR tool provides an accuracy of 98.8% upon execution providing a character error rate of 2.4% and word error rate (WER) of 2.8%. OCR tool finds its applications in document digitization, personal identification, archival of valuable documents, processing of invoices, and other documents. OCR tool holds an immense amount of value for researchers, practitioners and many organizations which seek effective techniques for relevant and accurate text extraction and recognition tasks.

This is an open access article under the [CC BY-SA](https://creativecommons.org/licenses/by-sa/4.0/) license.



## Corresponding Author:

Jyoti Wadmare

Department of Computer Engineering, K. J. Somaiya Institute of Technology

Somaiya Ayurvihar Complex, Eastern Express Highway, Sion (East), Mumbai 400 022, India

Email: jyoti@somaiya.edu

## 1. INTRODUCTION

Computer vision, a subfield of artificial intelligence, uses machine learning and neural networks to train systems how to extract useful information from digital inputs such as images and videos. These algorithms can then suggest activities or identify problems in visual data. Similarly, to enable computers comprehend, interpret and produce meaningful human language, natural language processing (NLP) is used [1]. NLP techniques include text understanding, translation, sentiment analysis, speech recognition, text generation, information retrieval, named entity recognition (NER), and question answering. Applications include virtual assistants, translation services, sentiment analysis tools, and information retrieval systems.

Optical character recognition (OCR) is a combination of computer vision and NLP that converts printed or handwritten text from photos into editable, machine-readable text. OCR extracts text from images or documents by locating text areas, separating text, identifying characters, and outputting the recognized text. It digitizes documents and enables text searches in images. The proposed OCR system was created to collect photographs, extract text using Tesseract OCR using Pytesseract, and clean them. In order for spaCy to train

the model, NER data is manually labeled using beginning, inside, outside (BIO) tagging and organized. The model predicts entities, which are represented in displaCy and highlighted in images. These tasks are simplified by web software, which saves extracted text in an Excel file. The primary algorithms used are Canny edge detection and Douglas-Peucker for edge detection and polyline simplification.

OCR has a wide range of uses, including digitizing student records, medical documents, and invoicing in education, healthcare, and finance, as well as benefiting other industries by improving digitalization and data accessibility [2]. The conversion of handwritten and printed papers into editable, searchable data simplifies a variety of human operations by speeding data entry, record-keeping, and information retrieval. By improving accuracy and efficiency, this change lowers the possibility of errors and facilitates faster access to essential documents. It also improves data organization and administration, making it easier to evaluate and distribute information across multiple platforms and systems [3]. While OCR works well with modern text, it is less accurate with historic documents and outdated digitalization technologies. As a result, noisy digitized texts require post-correction to improve OCR results, which are critical for information retrieval and NLP applications [4].

The structure of this document is as follows: section 1 is the introduction to what is OCR. Section 2 we study about the various tools for OCR already available and their literature survey, section 3 provides an overview of the system which is developed, section 4 discusses the results of the tool made. Section 5 presents the conclusions.

## 2. RELATED WORK

The strong foundation to develop OCR tool came through a thorough survey of research papers. It helped in understanding the need for the development of the OCR tools for text recognition from the images. The text obtained from the extraction process is subsequently stored within an Excel file.

Oladayo [5] focuses on using OCR technology to convert and keep many papers in historical archives digitally. Regular scanners scan images from the documents such that they cannot be used on screen or edited with any software used for any other document type. This study unveils an OCR software which is able to convert offline typed and handwritten documents into text forms that can be edited. By utilizing a morphological correlation technique, this system enhances the efficiency of text mapping and recognition [5].

Adjetey and Manu [6] present a novel technique to enhance image retrieval systems (IRSs). Their approach integrates a Tesseract OCR engine, and an enhanced text-matching algorithm, leveraging the levenshtein algorithm. Experimental results demonstrate a 100% success rate in retrieving the appropriate file based on partial query images, showcasing the effectiveness of this integrated approach in improving image retrieval accuracy [6].

Zhu *et al* [7] introduce ShotVis, a novel approach to capture text images from mobile devices and process these text images to store characters as structured data. It allows users to link visual forms to the underlying data and generate visualizations through touch-based interactions. With a simple click of the camera, ShotVis swiftly summarizes text from images into word clouds, scatterplots, and various other visualizations, enabling interactive exploration of text data captured via smartphone cameras [7].

Suddul and Seguin [8] recommend a process of customer registration using deep learning-based OCR technology which can be utilized for automatic text extraction from images of ID cards. The first step involves the text spots identification by a proprietary U-Net image segmentation algorithm and the other step is to recognize characters and formation of words in convolutional recurrent neural network (CRNN) with long short-term memory (LSTM) cells. The experiment was carried out on Mauritius' national identity card and it yields 0.70 intersection over union (IoU) score and 98% pixel accuracy [8].

Satirapiwong and Siriborvornratanakul [9] addressed the challenges of processing Thai invoices for business payments, which traditionally requires extensive manual efforts and template matching. The paper introduces bidirectional long short-term memory-conditional random field (BiLSTM-CRF) deep learning model which uses mixing of words and characters specifically for Thai invoices. This model highlighted accurate F1-score metrics, precision score and recall. The quality of OCR was highlighted using F1-score [9].

Lee [10] defines the use of many different libraries, specifically using interlibrary loan (ILL). To test the process, 20 copies of articles were sent to test the accuracy of Adobe Acrobat Pro DC to create searchable PDFs. The accuracy of automated OCR results was calculated and manual corrections were made after that to avoid problems which would provide a good initiative for ILL to provide patrons with materials that are accessible [10].

Manivannan *et al.* [11] proposed an energy-efficient IoT model for predicting handwritten prescription of doctors. It considers making use of a triboelectric smart recognition system for recognition of medical terms and is considered to be robust. The system results in digital twin development for monitoring systems to track usage where individual prescriptions can be developed and analyzed. The return on investment

(ROI) of the system is evaluated on the basis of different parameters such as OCR, accuracy, sensitivity, specificity, and sensitivity analysis [11].

Poodikkalam and Loganathan [12] focused on cognitive processing for OCR. The authors identified scale-invariant transforming feature (SIFT) descriptors which uses two functions. The other process, namely RootSIFT gives exceptional results without storage requirements or computational complexity. Artificial bee colony (ABC) is used for identification of English language characters. The accuracy of numbers, alphanumeric characters including small and big letters is tested and it is found that ABC algorithm has maximum efficiency of around 97.3077% [12].

Mohd *et al.* [13] focusses on Quranic OCR which is developed using convolutional neural network (CNN) and recurrent neural network (RNN). Based on the printed version of the Holy Quran, a new dataset was developed which recognizes the Quranic image's diacritic text. Two models compared for Arabic text recognition were LSTM and gated recurrent unit (GRU) where a public database was built and it achieved accuracy of 98% with validation data and 95% word recognition rate (WRR) and character recognition rate (CRR) of 99% in the test dataset [13].

Hassan *et al.* [14] highlights that Arabic scene text recognition is a complex part in understanding scene systems. The use of Arabic involving Latin characters is limited in deep learning methods. The dataset evaluates three parameters-use of deep learning techniques, identifying challenges in Arabic text and investigation of bilingual models. The dataset used helps in providing directions for future research [14].

Malhotra and Addis [15] highlight the Ethiopic handwritten text recognition using sequential feature extraction and efficient recognition using an end-to-end strategy. The architecture of the model includes an attention mechanism and a connectionist temporal classification utilizing seven CNNs and two RNNs are used for model training. The accuracy of character error rate (CER) obtained was 17.95% for test set I and 29.95% for test set II [15].

Wang *et al.* [16] attempted to improve Chinese OCR accuracy by creating a hybrid recognition model that was suited to the language's distinctive features. This approach pre-filters image interference and modifies character aspect ratios prior to OCR processing. Experiments revealed that image processing raised Tesseract-OCR's correct identification rate by about 12%, whilst NLP increased accuracy by about 5% [16].

Shahira and Lijiya [17] propose that for the ease of communication, textual data is supported with graphical representations but this is not applicable for blind or visually impaired people. The paper focuses on extracting valuable information or critical data from charts or graphs. Localization and classification are techniques that can be implemented using deep learning. The paper suggests the use of human computer interaction and artificial intelligence techniques to automate extraction of data and provide its description for visually impaired sections of society [17].

Polancic *et al.* [18] investigates the transformation of hand-drawn diagrams to digitally drawn diagrams using OCR. It suggests suitable solutions based on TensorFlow which provide accurate results for different elements or sections of hand-drawn diagrams and their elements. It makes use of different statistical approaches like Bayesian classifier, decision tree classifier, neural network classifier, nearest neighbors classifier, syntactic approach for text recognition [18].

Ueda *et al.* [19] investigates the text-based image captioning method which is used to provide captions to the images in the form of text making use of OCR. It uses a pre-trained contrastive language-image pre-training (CLIP) model to improve and enhance images using linguistic features of OCR. It also introduces two new attention models to strengthen the transformation architecture of representation of images in which the proposed system outperforms the TextCaps dataset [19].

Wu *et al.* [20] proposes a two-level rectification attention network (TRAN) to rectify and identify texts. It consists of two levels-first is two-level rectification network (TORN) which is used to resolve geometrical constraints using pixel level adjustment and give clear text and second is attention-based recognition network (ABRN) which is used to recognize text in rectified images. To handle other variations, a new channel and kernel wise attention unit is developed. The state-of-art performance is achieved as a result of this experimentation conducted [20].

Zhang *et al.* [21] focused-on challenges faced due to text reading of different text images. Sequence-like images are difficult to predict and conventional methods do not align them as character information. The method used to align sequential images is novel adversarial sequence-to-sequence domain adaptation (ASSDA) which mines local regions containing characters and aligns them in an adversarial manner. After performing extensive text recognition, it is proved that ASSDA is efficient to transfer sequence knowledge [21].

Yıldız [22] puts forward a novel technique to employ correction of grammatical errors often found in OCR which involves correcting syntax as well as semantics by considering how often specific combinations of words occur in sentences alongside applying recursion. It computes frequency for every pair of words that occur one after another within any given body of texts before setting up a correctional hub which consists

mostly of high frequency word pairs. The method judge's rectification and accuracy rate which is found to be 98 and 96% respectively [22].

Nguyen *et al.* [23] concentrates on different issues that arise during OCR as a result of inadequately scanned images or limitations in the OCR software that has been designed. The research proposes a hill climbing algorithm-based unsupervised model for OCR error correction. Correcting suggestions are ranked and scored using a weighted objective function, for which optimal weight combinations are determined heuristically [23].

After extensive literature survey, it is found that there is a need for development of OCR tool to identify text from images. This tool is useful for analysis, automation processes and improvised productivity for different purposes. A comparison table of different OCR tools already available based on the parameters such as accuracy, handling complex layouts, speed, ease of use, and cost is shown in Table 1.

From the comparison, it is evident that Tesseract OCR offers an accuracy range of 85-99%, with a processing speed of 2.5 pages per second. It is a free tool and can handle complex layouts up to 90%. It is also relatively easy to use compared to other OCR tools. Tesseract's accuracy is influenced by factors such as image quality, language, and noise. For high-quality printed text, accuracy can exceed 95%, whereas for handwritten text, low-quality or noisy images, it may drop to 70-85% or lower. This survey highlights the need for additional research in OCR technologies to improve them, especially to lower error rates, deal with complex layouts and enhancing speed while maintaining high accuracy.

Table 1. Comparison of different OCR tools

OCR method	Accuracy (%)	Handling complex layouts (%)	Speed (pages/s)	Ease of use	Cost
Tesseract OCR [24]	85-99	60	2.5	7/10	Free
EasyOCR	90-95	85	1.5	8/10	Free
Amazon Textract [24]	95-98	90	1.8	8/10	\$1.50 per 1000 pages
Adobe Acrobat OCR	90-95	75	2.5	9/10	\$14.99 per month
OCR.Space	85-90	70	2.7	9/10	Free (limited)
Google Document AI [24]	95-98	90	2.0	9/10	Pricing varies (pay-as-you-go)

### 3. METHODOLOGY

The research paper focuses on the development of an OCR tool using computer vision and NLP. It is developed in Python language with different libraries. Python libraries of computer vision used are OpenCV, NumPy and Pytesseract and libraries of NLP used include spaCy, Pandas, regular expression, and string. The extracted text can be collectively downloaded in Microsoft Excel for ease of management of text extracted from the images. The workflow for the development of an OCR tool is depicted in the Figure 1.

- i) Step 1: data preparation
  - Collect images containing certificates or text to be processed.
  - Pytesseract, a Python wrapper for Google's Tesseract OCR engine, extracts text from images.
  - Text extracted from images is preprocessed to remove noise, formatting, and irrelevant data.
- ii) Step 2: labeling NER data
  - NER data is labelled manually using the BIO tagging scheme.
  - B-Beginning: denotes the start of an entity.
  - I-Inside: indicates the continuation of an entity.
  - O-Outside: marks areas not part of any entity.
- iii) Step 3: data preprocessing
  - The labelled NER data is formatted to align with spaCy's training format.
  - The labelled data is converted into a format compatible with spaCy for NER model training.
- iv) Step 4: NER model training
  - Define the architecture and parameters of the NER model using spaCy.
  - The NER model is trained on prepared data with optimizations for performance.
- v) Step 5: NER predictions and data pipeline
  - The trained NER model is loaded to make predictions on new data.
  - spaCy's displaCy module is utilized to render and serve NER predictions visually.
  - Bounding boxes are overlaid on images to highlight recognized entities.
  - Recognized entities are extracted and parsed from the text for further processing or display.
- vi) Step 6: Web App creation: the developed components are integrated to create a user-friendly web application allowing users to upload certificate images, extract text, identify entities, and visualize the results effectively.

vii) Step 7: The text obtained from the extraction process is subsequently stored within an Excel file.  
 Different libraries, their functions and their algorithms used in the development of an OCR tool are: Canny edge detection algorithm and Douglas-Peucker algorithm. Canny edge detection algorithm computes the gradient magnitude and direction for each pixel using techniques such as Sobel operators. It then suppresses non-maximum gradient values to thin out detected edges, retaining only the local maxima along the edges. It follows edges by linking adjacent pixels with gradient magnitudes above a high threshold and potentially weak edges above a low threshold and determine which weak edges to retain based on their connectivity to strong edges. Douglas-Peucker algorithm identifies a polyline defined by a sequence of points in a plane. The output is that the algorithm generates a simplified polyline by retaining critical points that define the shape accurately.

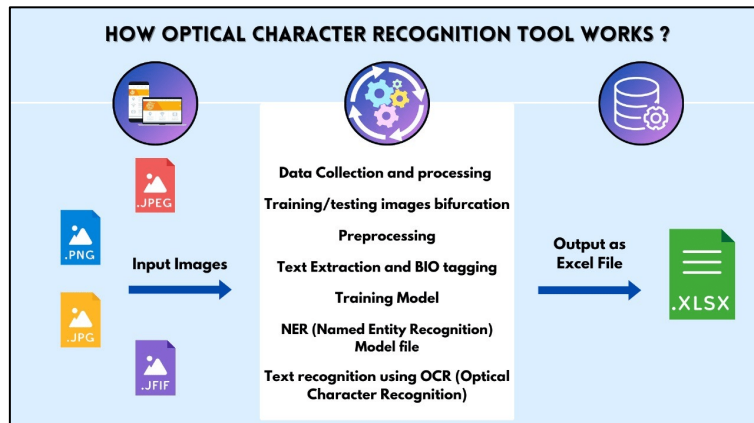


Figure 1. Working of OCR tool

**4. RESULTS AND DISCUSSIONS**

The OCR tool is specifically crafted to extract text from images and store the obtained text in Microsoft Excel. It is trained on the dataset of over 8,000 images which includes 6,500 images used for training, and 1,500 images used for testing. The step-by-step process to use OCR tool is given as follow:

- i) Step 1: upload the image for text extraction process by clicking “Upload Image” button and click “Wrap Certificate and Extract Text”. After that, select the boundaries of the image indicating the part from where the text is to be extracted.
- ii) Step 2: extracted text from loaded image is shown in table format as shown in Figure 2.
- iii) Step 3: click “Download as Excel” to download extracted text in Excel format and click the “Back to Home” button to go back to main home page as shown in Figure 3.

The screenshot shows the interface of the OCR tool. At the top, it says 'KJSIT's Student Achievement Portal and Optical Recognition Tool'. Below this, there are two images of a 'CERTIFICATE OF PARTICIPATION' for Priyanshu Singh. Below the images are buttons for 'Download as Excel' and 'Back to Home'. At the bottom, there is a table showing the extracted text from the certificate.

STU	DES	COI	DAT	NCER	TYP
Priyanshu Singh	This Certificate is Presented To, We Appreciate Your Dedication And Hard Work That Resulted In Successful Completion Of Course A Keton, Fuelformation www.fuelformation.in	santoshhurali	appphchairman,	ashoksfellowchiefmentor,	CertificateOfParticipation

Figure 2. Text extraction from image in table format

19	Yash Chauhan	Of Vv, Participa	The Institution	onmarch25	vikrantsankhe	ParticipationCertificate
20	Yash Chauhan	Of Vv, Participa	The Institution	onmarch25	vikrantsankhe	ParticipationCertificate
21	Deep Dhar	Technology Aac	Engineering An	onmarch25	vikrantsankhe	Certificate
22	Deep Dhar	Wee Prenat, Ha	Engineering An	onmarch25	vikrantsankhe	ParticipationCertificate

Figure 3. Report generation in Excel format

**4.1. Result validation of OCR tool**

The accuracy of the OCR tool is evaluated using two parameters: CER and word error rate (WER) [25]. The CER is used to calculate the total character count that is mismatched or detected incorrect during text extraction to the total number of characters in the original text. It is used to measure the count of characters that are substituted, deleted or incorrect characters inserted during the text extraction process. CER is given by (1).

$$CER = \frac{S(char) + I(char) + D(char)}{N(char)} \tag{1}$$

Where S(char) stands for total character count that are substituted from the original text; I(char) denote the number of incorrect characters that are inserted in the extracted text; D(char) signifies the number of characters not recognized or missing in the extracted text; and N(char) indicates the total character count present in the original text. The typical CER should be in the range of 2-10%. The CER of OCR tool developed is 2.4%.

The WER is used to calculate the total number of words that are mismatched or detected incorrect during text extraction to the total number of words in the original text. It is used to measure the count of words that are substituted, deleted or incorrect words inserted during the text extraction process. WER is given by (2).

$$WER = \frac{S(word) + I(word) + D(word)}{N(word)} \tag{2}$$

Where S(word) stands for the number of words that are substituted from the original text; I(word) denote the number of incorrect words that are inserted in the extracted text; D(word) signifies the number of words not recognized or missing in the extracted text; and N(word) indicates total words count present in the original text. The target WER should be less than 5%. The WER of OCR tool developed is 2.8%. The accuracy of the overall OCR tool developed is 98.8%.

**4.2 Comparison with existing system**

Table 2 compares various OCR approaches and their accuracies on different types of text. It combines standard OCR algorithms, custom-built models, and transformed machine learning techniques. The performance measurements include typewritten and handwritten text, as well as specific datasets. Table 2 demonstrates the dataset used and accuracy of each OCR technique. Traditional approaches, such as Tesseract OCR, achieve good accuracy for typewritten text, although sophisticated models with U-Net and CRNN structures perform competitively. These insights help choose the best OCR technique based on particular application needs and text characteristics.

Table 2. Summary of different OCR techniques and their accuracies

Technique/model used	Dataset	Accuracy
Slant correction layer and character segmentation and recognition [3]	ICDAR2013-1,081 images	96.42%-ICDAR2013 dataset.
U-Net image segmentation [8]	Self made-8,000 images	96.52%-self-made screen rendered dataset
RootSIFT with ABC optimized neural network algorithm [12]	55,000 images	98%
Statistical classification approaches [18]	500 training images	97.31%
CNN with RNN [24]	ICDAR2013-1,081 images	Typewritten text: 97% Handwritten text: 80 to 90%
Tesseract OCR-our method	10,419 images	96.21%
	8,000 images	Typewritten text: 98.8% Handwritten text: 90.6%

**5. CONCLUSION AND FUTURE SCOPE**

The OCR tool plays a crucial role in extracting text from images, improving efficiency in various industries by converting scans, images, and handwriting into editable formats. Recent advancements in machine learning and computer vision offer enhanced accuracy, particularly in challenging scenarios like noisy images

or complex fonts. Broadening language support will further enhance the capabilities of OCR devices to handle a variety of scripts. By integrating OCR with emerging technologies like IoT and augmented reality (AR), and training it to detect handwritten and cursive text, its versatility is increased. Accessibility features such as voice control and screen reader compatibility cater to a wide range of users, including those with disabilities. Implementing robust security measures, including encryption and compliance with regulations, is essential for protecting sensitive data. In conclusion, the future of OCR tools looks promising, driven by innovation and user requirements. Emphasizing accuracy, language diversity, integration, customization, accessibility, and security will ensure that OCR tools continue to evolve as indispensable solutions for text processing in the digital landscape.

## ACKNOWLEDGMENTS

The authors fully acknowledge the institute, K. J. Somaiya Institute of Technology for its support on this research.

## FUNDING INFORMATION

Authors state no funding involved.

## AUTHOR CONTRIBUTIONS STATEMENT

This journal uses the Contributor Roles Taxonomy (CRediT) to recognize individual author contributions, reduce authorship disputes, and facilitate collaboration.

Name of Author	C	M	So	Va	Fo	I	R	D	O	E	Vi	Su	P	Fu
Jyoti Wadmare	✓	✓		✓	✓	✓			✓	✓		✓		✓
Sunita Ravindra Patil	✓				✓	✓			✓			✓		✓
Dakshita Kolte	✓	✓	✓	✓	✓	✓	✓	✓	✓		✓			
Kapil Bhatia	✓	✓	✓	✓	✓	✓	✓	✓		✓	✓			
Palak Desai	✓	✓	✓	✓	✓	✓	✓	✓		✓	✓			
Ganesh Wadmare	✓				✓	✓				✓		✓		✓

C : Conceptualization

M : Methodology

So : Software

Va : Validation

Fo : Formal analysis

I : Investigation

R : Resources

D : Data Curation

O : Writing - Original Draft

E : Writing - Review & Editing

Vi : Visualization

Su : Supervision

P : Project administration

Fu : Funding acquisition

## CONFLICT OF INTEREST STATEMENT

Authors state no conflict of interest.

## DATA AVAILABILITY

The data that support the findings of this study are available from the corresponding author, [JW], upon reasonable request.





## REFERENCES

- [1] P. Jain, D. K. Taneja, and D. H. Taneja, "Which OCR toolset is good and why? a comparative study," *Kuwait Journal of Science*, vol. 48, no. 2, Apr. 2021, doi: 10.48129/kjs.v48i2.9589.
- [2] W. Sun, L. Liu, W. Zhang, and J. C. Comfort, "Intelligent OCR processing," *Journal of the American Society for Information Science*, vol. 43, no. 6, pp. 422–431, 1992, doi: 10.1002/(SICI)1097-4571(199207)43:63.O.CO;2-Z.
- [3] T. H. Nguyen, A. Jatowt, M. Coustaty, and A. Doucet, "Survey of post-OCR processing approaches," *ACM Computing Surveys*, vol. 54, no. 6, pp. 1–37, Jul. 2022, doi: 10.1145/3453476.
- [4] T. Hegghammer, "OCR with Tesseract, Amazon Textract, and Google document AI: a benchmarking experiment," *Journal of Computational Social Science*, vol. 5, no. 1, pp. 861–882, May 2022, doi: 10.1007/s42001-021-00149-1.
- [5] O. O. Oladayo, "Optical character recognition of off-line typed and handwritten English text using morphological and template matching techniques," *IAES International Journal of Artificial Intelligence*, vol. 3, no. 3, pp. 121–128, Sep. 2014, doi: 10.11591/ijai.v3.i3.pp121-128.
- [6] C. Adjetej and K. S. A. -Manu, "Content-based image retrieval using Tesseract OCR engine and levenshtein algorithm," *International Journal of Advanced Computer Science and Applications*, vol. 12, no. 7, 2021, doi: 10.14569/IJACSA.2021.0120776.





- [7] B. Zhu, H. Zhang, W. Chen, F. Xia, and R. Maciejewski, "ShotVis: smartphone-based visualization of OCR information from images," *ACM Transactions on Multimedia Computing, Communications, and Applications*, vol. 12, no. 1s, pp. 1–17, Oct. 2015, doi: 10.1145/2808210.
- [8] G. Suddul and J. F. L. Seguin, "A custom-built deep learning approach for text extraction from identity card images," *International Journal of Informatics and Communication Technology (IJ-ICT)*, vol. 13, no. 1, pp. 34–41, Apr. 2024, doi: 10.11591/ijict.v13i1.pp34-41.
- [9] K. Satirapiwong and T. Siriborvornratanakul, "Information extraction for different layouts of invoice images," *The Imaging Science Journal*, vol. 69, no. 5–8, pp. 417–429, Nov. 2021, doi: 10.1080/13682199.2022.2157367.
- [10] M. C. Lee, "Improving accessibility in interlibrary Loan using OCR," *Journal of Interlibrary Loan, Document Delivery & Electronic Reserve*, vol. 29, no. 1–2, pp. 75–87, Mar. 2020, doi: 10.1080/1072303X.2020.1859426.
- [11] P. Manivannan *et al.*, "Doctor unpredicted prescription handwriting prediction using triboelectric smart recognition," *Production Planning & Control*, pp. 1–17, Apr. 2023, doi: 10.1080/09537287.2023.2202173.
- [12] S. B. Poodikkalam and P. Loganathan, "Optical character recognition based on local invariant features," *The Imaging Science Journal*, vol. 68, no. 4, pp. 214–224, May 2020, doi: 10.1080/13682199.2020.1827814.
- [13] M. Mohd, F. Qamar, I. Al-Sheikh, and R. Salah, "Quranic optical text recognition using deep learning models," *IEEE Access*, vol. 9, pp. 38318–38330, 2021, doi: 10.1109/ACCESS.2021.3064019.
- [14] H. Hassan, A. El-Mahdy, and M. E. Hussein, "Arabic scene text recognition in the deep learning era: analysis on a novel dataset," *IEEE Access*, vol. 9, pp. 107046–107058, 2021, doi: 10.1109/ACCESS.2021.3100717.
- [15] R. Malhotra and M. T. Addis, "End-to-end historical handwritten Ethiopic text recognition using deep learning," *IEEE Access*, vol. 11, pp. 99535–99545, 2023, doi: 10.1109/ACCESS.2023.3314334.
- [16] B. Wang, Y. W. Ma, and H. T. Hu, "Hybrid model for Chinese character recognition based on Tesseract-OCR," *International Journal of Internet Protocol Technology*, vol. 13, no. 2, 2020, doi: 10.1504/IJIPT.2020.106316.
- [17] K. C. Shahira and A. Lijiya, "Towards assisting the visually impaired: a review on techniques for decoding the visual data from chart images," *IEEE Access*, vol. 9, pp. 52926–52943, 2021, doi: 10.1109/ACCESS.2021.3069205.
- [18] G. Polancic, S. Jagecic, and K. Kous, "An empirical investigation of the effectiveness of optical recognition of hand-drawn business process elements by applying machine learning," *IEEE Access*, vol. 8, pp. 206118–206131, 2020, doi: 10.1109/ACCESS.2020.3034603.
- [19] A. Ueda, W. Yang, and K. Sugiura, "Switching text-based image encoders for captioning images with text," *IEEE Access*, vol. 11, pp. 55706–55715, 2023, doi: 10.1109/ACCESS.2023.3282444.
- [20] L. Wu, Y. Xu, J. Hou, C. L. P. Chen, and C.-L. Liu, "A two-level rectification attention network for scene text recognition," *IEEE Transactions on Multimedia*, vol. 25, pp. 2404–2414, 2023, doi: 10.1109/TMM.2022.3146779.
- [21] Y. Zhang, S. Nie, S. Liang, and W. Liu, "Robust text image recognition via adversarial sequence-to-sequence domain adaptation," *IEEE Transactions on Image Processing*, vol. 30, pp. 3922–3933, 2021, doi: 10.1109/TIP.2021.3066903.
- [22] S. Yildiz, "Turkish scene text recognition: introducing extensive real and synthetic datasets and a novel recognition model," *Engineering Science and Technology, an International Journal*, vol. 60, no. 1, Dec. 2024, doi: 10.1016/j.jestch.2024.101881.
- [23] Q.-D. Nguyen, N.-M. Phan, P. Krömer, and D.-A. Le, "An efficient unsupervised approach for OCR error correction of Vietnamese OCR text," *IEEE Access*, vol. 11, pp. 58406–58421, 2023, doi: 10.1109/ACCESS.2023.3283340.
- [24] J. Memon, M. Sami, R. A. Khan, and M. Uddin, "Handwritten optical character recognition (OCR): a comprehensive systematic literature review (SLR)," *IEEE Access*, vol. 8, pp. 142642–142668, 2020, doi: 10.1109/ACCESS.2020.3012542.
- [25] S. Karthikeyan, A. G. S. de Herrera, F. Doctor, and A. Mirza, "An OCR post-correction approach using deep learning for processing medical reports," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 32, no. 5, pp. 2574–2581, May 2022, doi: 10.1109/TCSVT.2021.3087641.

## BIOGRAPHIES OF AUTHORS






**Dr. Jyoti Wadmare**     is Assistant Professor in Department of Computer Engineering at KJSIT. She has teaching experience of 17 years with an AI background. Her major domain of interest is the conjunction of AI and computer vision. Testimonials of work includes many conferences' presentations and articles published that quite clearly states advancement in this area by her and has filed a patent and acquired four copyrights. She can be contacted at email: jyoti@somaiya.edu.






**Dr. Sunita Ravindra Patil**     is the Director, NMIMS Deemed to be University, Shirpur, Dhule, Maharashtra. She holds a Ph.D. in Computer Engineering, specializing in data mining, big data, and data science, with around 20 years of teaching and administrative experience. A member of the board of studies in computer engineering at UoM, she has published extensively in esteemed journals and conferences and has visited various international institutions for knowledge exchange. Her focus is on implementing outcome-based academic reforms to benefit society. She can be contacted at email: spatil@somaiya.edu.








**Dakshita Kolte**    is a B.Tech. student in computer engineering at KJSIT. She has developed AI and ML solutions that integrate artificial intelligence with web-based technologies. She has a strong track record of participating in prestigious competitions such as Mastek Project Deep Blue, Aavishkar, and Creative Ideas and Innovations in Action. She has also been honored with the prestigious “Somaiya star girl” Award by Somaiya Management. Additionally, she holds four copyrights for her work. She can be contacted at email: [d.kolte@somaiya.edu](mailto:d.kolte@somaiya.edu).






**Kapil Bhatia**    is a B.Tech. student in computer engineering at KJSIT, specializing in artificial intelligence and machine learning. He excels in developing solutions that integrate web-based technologies, the internet of things, and artificial intelligence. His participation in renowned competitions such as Aavishkar, Mastek Project Deep Blue, and Creative Ideas and Innovations in Action showcases his exceptional expertise. He holds four copyrights for his work and has received widespread appreciation for his innovative project developments. He can be contacted at email: [kapil.bhatia@somaiya.edu](mailto:kapil.bhatia@somaiya.edu).



**Palak Desai**    is a computer engineering student at KJSIT passionate about creativity and technology in the domains UI/UX design, front-end web development, and data analytics. She enjoys creating intuitive, beautiful user interfaces and analyzing data to drive insights. She has participated in competitions like Aavishkar and Creative Ideas and Innovations in Action (CIIA) for her contribution in two institute-level projects. She is dedicated to continuous learning and making impactful, user-centered solutions. She has been granted three copyrights for her work. She can be contacted at email: [palak.pd@somaiya.edu](mailto:palak.pd@somaiya.edu).



**Ganesh Wadmare**    is an Assistant Professor in the Department of Artificial Intelligence and Data Science of KJSIT and a Ph.D. scholar in Savitribai Phule Pune University, with academic experience for over 19 years. He has extensive exposure and experience in the field of artificial intelligence and renewable source of energy. He has published his research papers in both national and international conferences. He can be contacted at email: [gwadmare@somaiya.edu](mailto:gwadmare@somaiya.edu).