AADHAAR UID MASKING TOOL

[1]Students- D.Saikiran, S.Aravind, S.Vamshikumar, K.Premkumar
[2]Assistant Professor -N. Devender, J.Naveen

[1] BTech Student, Department of CSE, Balaji Institute of Technology and Science, Laknepally, Warangal, India

[2]Department of CSE, Balaji Institute of Technology and Science, Laknepally, Warangal, India

Abstract-The world is now using technologies that are not only effective but also easy to use and understand for people with basic technical knowledge. Data security and privacy issues are the most important ones, especially for personal information. This project is to develop a full fledged system for detecting and anonymizing Aadhaar numbers in images and documents using Python. Aadhaar numbers are secret identifiers that are used a lot in India and their protection is necessary to avoid their misuse. The suggested solution is based on Optical Character Recognition (OCR) technology, which uses Tesseract for identifying Aadhaar numbers in PDF, TIFF or image documents. The identified numbers are then checked using a Verhoeff checksum algorithm. The first eight digits of an Aadhaar number are covered once an Aadhaar number is recognised to preserve the privacy of the client while still allowing for the use of documents. To this end, a Graphical User Interface (GUI) is developed with the help of Tkinter such that users can select files and process them with minimal technical knowledge. This project focuses on proper document management and the usability of the application. With the help of technologies such as OCR, super resolution models, and GUI, it makes it easier for individuals with varying levels of technical aptitude to perform the task of anonymizing Aadhaar documents. Abstract words: Python, Aadhaar masking, document anonymization, OCR, GUI, Tkinter, Verhoeff algorithm, privacy protection.

Index terms – Python, Aadhaar masking, document anonymization, OCR, GUI, Tkinter, Verhoeff algorithm, privacy protection**.**

1.INTRODUCTION

In a world where digital data is at the heart of everything we do, protecting personal information has become more important than ever. Aadhaar numbers, unique identifiers widely used in India, carry sensitive information that, if mishandled, could lead to serious privacy breaches. Safeguarding this data isn't just a technical challenge—it's a responsibility.

This project is all about making data protection accessible to everyone, not just tech experts. By using Optical Character Recognition (OCR) technology powered by Tesseract, the system can scan and detect Aadhaar numbers in documents like PDFs, images, and TIFF files. It then uses the Verhoeff checksum algorithm to confirm the numbers and masks the first eight digits to ensure privacy while keeping the documents functional.

What makes this solution truly user-friendly is its Graphical User Interface (GUI), built with Tkinter. Even those with minimal technical know-how can easily upload their files and anonymize sensitive data in just a few clicks. By blending advanced technologies like OCR, super-resolution models, and a simple, intuitive interface, this project aims to make protecting personal data straightforward, efficient, and accessible to everyone[1-27].

2. PROBLEM STATEMENT

In an era where data security and privacy are increasingly critical, the protection of sensitive

Doi:10.48047/ijiee.2025.15.5.10

personal identifiers, such as Aadhaar numbers, is of paramount importance to prevent their misuse. Despite the growing reliance on technological solutions, individuals with minimal technical expertise often find it challenging to effectively manage and safeguard such data. The absence of a simple yet efficient tool for detecting and anonymizing Aadhaar numbers in documents further complicates the issue. There is a pressing need for a user-friendly system that can address these concerns by ensuring privacy protection without compromising the usability of documents.

3. LITERATURE SURVEY

"AADHAAR CARD MASKING TOOL" (2023) This study introduces a tool that uses convolutional neural networks (CNNs) and the Verhoeff algorithm to detect and mask Aadhaar numbers in digital images. It emphasizes privacy protection by replacing the first eight digits of Aadhaar numbers with characters like "XXXX-XXXX" while retaining the last four digits for identity verification.

"Aadhar Card Masking Tool" (2023) Published in the International Journal of Novel Research and Development, this research highlights a deep learning-based approach for Aadhaar masking. It combines OCR and image processing to ensure accurate identification and masking of sensitive information while maintaining document readability.

"Aadhaar-UID-Masking-Tool" (2023) This GitHub project focuses on extracting, verifying, and masking Aadhaar numbers from scanned documents and images. It employs PyTesseract for OCR and OpenCV for image processing, achieving a high accuracy rate of 94.6% for both training and validation

4. EXISTING SYSTEM

- **UIDAI's Masked Aadhaar** The UniqueIdentification Authority of India (UIDAI) provides an official option to download a masked Aadhaar. This version replaces the first eight digits of the Aadhaar number with "XXXX-XXXX," leaving only the last four digits visible. It ensures privacy while allowing limited identity verification.
- **Aadhaar Masking APIs** Several companies, such as SurePass and SignDesk, offer Aadhaar masking APIs. These APIs are designed for businesses handling Aadhaar data, enabling them to automatically mask the first eight digits of Aadhaar numbers in various document formats like PDFs and images. They are easy to integrate with existing systems and ensure compliance with privacy regulations.
- **Custom Masking Tools** Platforms like DeepVue provide customizable Aadhaar masking solutions. These tools use algorithms to detect Aadhaar numbers in documents and replace sensitive digits with placeholders. They support multiple formats and are tailored to organizational needs, ensuring secure handling of Aadhaar data.

5.PROPOSED SYSTEM

There are two stages in QR code generation:

a. Uploading Document

The user begins by interacting with a simple and intuitive graphical user interface (GUI) built using Tkinter. Through the interface, the user can browse and select the document they wish to process. This could be in formats such as PDF, image, or TIFF. The tool ensures compatibility with various file types to meet diverse user needs, making the first step straightforward and user-friendly.

Doi:10.48047/ijiee.2025.15.5.10

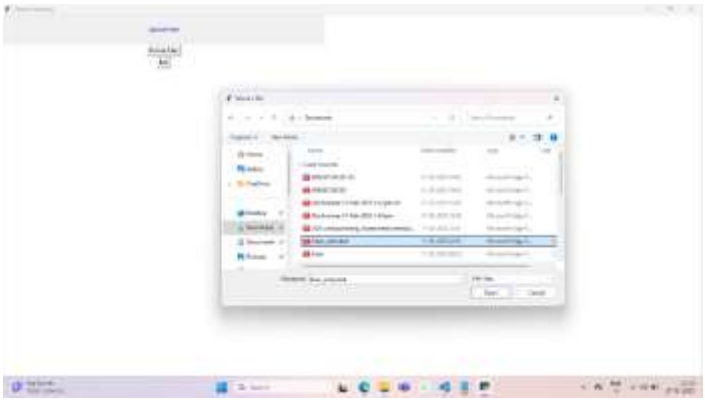


Figure 1

b. Scanning Document

After uploading, the document is processed using Optical Character Recognition (OCR) technology, specifically through Tesseract. To enhance the accuracy of text extraction, the tool applies preprocessing techniques such as noise reduction, converting images to grayscale, and improving resolution with super-resolution methods. OCR scans the document meticulously, identifying all text elements present within it, including potential Aadhaar numbers.

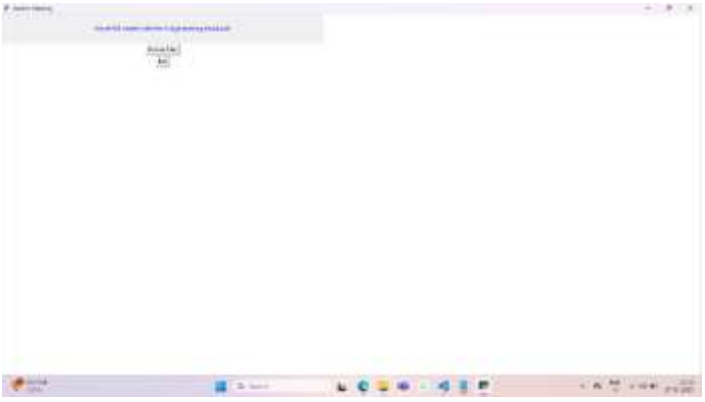
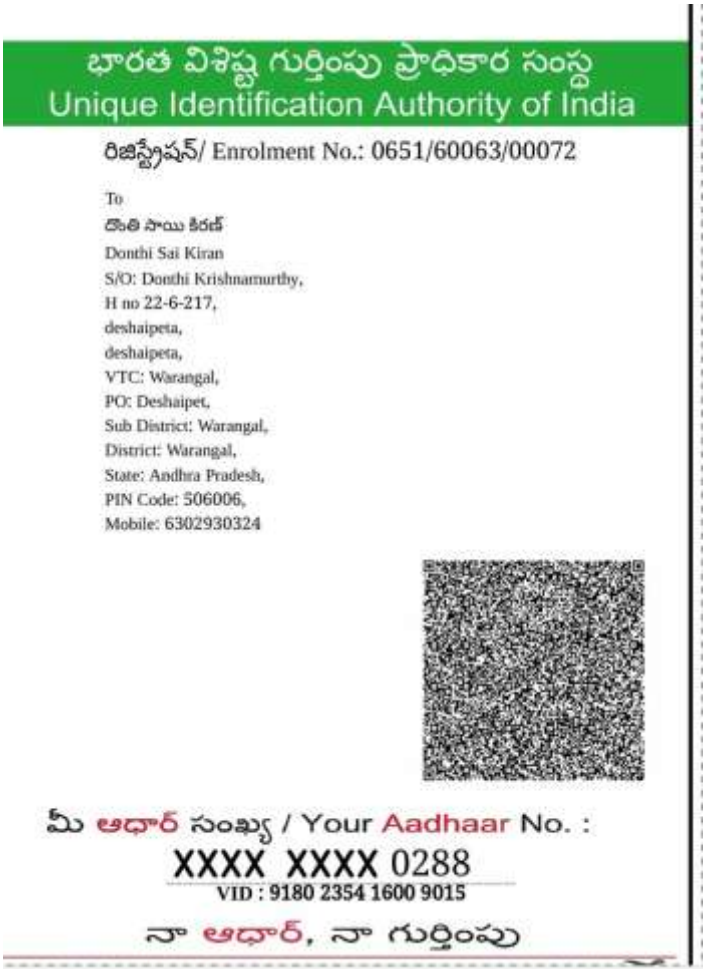


Figure 2

c. Masking Aadhaar card

Once the text is extracted, the tool employs a pattern-matching algorithm, such as regular expressions, to detect Aadhaar numbers. Detected numbers are then validated using the Verhoeff checksum algorithm to ensure they are authentic Aadhaar numbers. For any verified Aadhaar numbers, the masking logic is applied—replacing the first eight digits with "XXXX-XXXX," while retaining the last four digits to allow partial identification without compromising privacy.



4.IMPLEMENTATION

The implementation of this project begins with creating a user-friendly GUI using Tkinter, allowing users to upload documents in formats like PDF or images. The system employs Tesseract OCR to extract text from the documents, enhanced by preprocessing techniques such as noise reduction and resolution improvement for accuracy. Aadhaar

Doi:10.48047/ijiee.2025.15.5.10

numbers are then identified using regular expressions and validated via the Verhoeff checksum algorithm. Once validated, the masking mechanism replaces the first eight digits of Aadhaar numbers with "XXXX-XXXX," retaining only the last four digits to ensure privacy while maintaining document usability. Finally, the processed document is saved or exported, making the tool efficient, accessible, and secure for all users.

6.CONCLUSION

In conclusion, this project successfully addresses the critical need for safeguarding sensitive Aadhaar information by providing an efficient and user-friendly solution for document anonymization. By leveraging advanced technologies like OCR with Tesseract, the Verhoeff checksum algorithm for Aadhaar validation, and a GUI built with Tkinter, the tool ensures accessibility for users with varying technical expertise. The systematic approach to masking Aadhaar numbers, by replacing the first eight digits while retaining the last four, effectively balances privacy protection with document functionality. This innovative tool not only reinforces data security and privacy but also highlights the potential of technology in simplifying complex tasks for everyday users. It stands as a significant contribution towards secure and responsible management of personal data in the digital era..

REFERENCES

1. Unique Identification Authority of India (UIDAI), "What is Masked Aadhaar?" Available at https://uidai.gov.in, accessed on 12-03-2025.
2. K. Patel, S. Joshi, "Image Preprocessing Techniques for Improved OCR Accuracy," International Journal of Advanced Research in Computer Science, 2020.
3. Aadhar Card Masking Tool, International Journal of Novel Research and Development, 2023
4. Ramdas Vankdothu,Dr.Mohd Abdul Hameed, Husnah Fatima" A Brain Tumor Identification and Classification Using Deep Learning based on CNN-LSTM Method" Computers and Electrical Engineering , 101 (2022) 107960
5. Ramdas Vankdothu,.Mohd Abdul Hameed "Adaptive features selection and EDNN based brain image recognition on the internet of medical things", Computers and Electrical Engineering , 103 (2022) 108338.
6. Ramdas Vankdothu,.Mohd Abdul Hameed,Ayesha Ameen,Raheem,Unnisa " Brain image identification and classification on Internet of Medical Things in healthcare system using support value based deep neural network" Computers and Electrical Engineering,102(2022) 108196.
7. Ramdas Vankdothu,.Mohd Abdul Hameed" Brain tumor segmentation of MR images using SVM and fuzzy classifier in machine learning" Measurement: Sensors Journal,Volume 24, 2022, 100440 .
8. Ramdas Vankdothu,.Mohd Abdul Hameed" Brain tumor MRI images identification and classification based on the recurrent convolutional neural network" Measurement: Sensors Journal,Volume 24, 2022, 100412 .
9. Bhukya Madhu, M.Venu Gopala Chari, Ramdas Vankdothu,.Arun Kumar Silivery,Veerender Aerranagula " Intrusion

Doi:10.48047/ijiee.2025.15.5.10

detection models for IOT networks via deep learning approaches ” Measurement: Sensors Journal,Volume 25, 2022, 100641

10. Mohd Thousif Ahemad ,Mohd Abdul Hameed, Ramdas Vankdothu” COVID-19 detection and classification for machine learning methods using human genomic data” Measurement: Sensors Journal,Volume 24, 2022, 100537
11. S. Rakesh [a], NagaratnaP. Hegde [b], M. VenuGopalachari [c], D. Jayaram [c], Bhukya Madhu [d], MohdAbdul Hameed [a], Ra mdas Vankdothu [e], L.K. Suresh Kumar “Moving object detection using modified GMM based background subtraction” Measurement: Sensors ,Journal,Volume 30, 2023, 100898
12. Ramdas Vankdothu,Dr.Mohd Abdul Hameed, Husnah Fatima “Efficient Detection of Brain Tumor Using Unsupervised Modified Deep Belief Network in Big Data” Journal of Adv Research in Dynamical & Control Systems, Vol. 12, 2020.
13. Ramdas Vankdothu,Dr.Mohd Abdul Hameed, Husnah Fatima “Internet of Medical Things of Brain Image Recognition Algorithm and High Performance Computing by Convolutional Neural Network” International Journal of Advanced Science and Technology, Vol. 29, No. 6, (2020), pp. 2875 – 2881
14. Ramdas Vankdothu,Dr.Mohd Abdul Hameed, Husnah Fatima “Convolutional Neural Network-Based Brain Image Recognition Algorithm And High-Performance Computing”, Journal Of Critical Reviews,Vol 7, Issue 08, 2020(Scopus Indexed)
15. Ramdas Vankdothu, Dr.Mohd Abdul Hameed “A Security Applicable with Deep Learning Algorithm for Big Data Analysis”,Test Engineering & Management Journal,January-February 2020
16. Ramdas Vankdothu, G. Shyama Chandra Prasad “ A Study on Privacy Applicable Deep Learning Schemes for Big Data” Complexity International Journal, Volume 23, Issue 2, July-August 2019
17. Ramdas Vankdothu, Dr.Mohd Abdul Hameed, Husnah Fatima “ Brain Image Recognition using Internet of Medical Things based Support Value based Adaptive Deep Neural Network” The International journal of analytical and experimental modal analysis, Volume XII, Issue IV, April/2020
18. Ramdas Vankdothu,Dr.Mohd Abdul Hameed, Husnah Fatima” Adaptive Features Selection and EDNN based Brain Image Recognition In Internet Of Medical Things “ Journal of Engineering Sciences, Vol 11,Issue 4 , April/ 2020(UGC Care Journal)
19. Ramdas Vankdothu, Dr.Mohd Abdul Hameed “ Implementation of a Privacy based Deep Learning Algorithm for Big Data Analytics”, Complexity International Journal , Volume 24, Issue 01, Jan 2020
20. Ramdas Vankdothu, G. Shyama Chandra Prasad” A Survey On Big Data Analytics: Challenges, Open Research Issues and Tools” International Journal For Innovative Engineering and Management Research,Vol 08 Issue08, Aug 2019.
21. Vankdothu, R., Hameed, M.A. “An Effective Congestion and Interference Secure Routing Protocol for

Doi:10.48047/ijiee.2025.15.5.10

Internet of Things Applications in Wireless Sensor Network " Wireless Personal Communication Journal 140, 143–161 (2025)

22. Vankdothu, R., Bhukya, H. & Bhukya, R.R. "Hybrid TDR-MI Based Wireless Sensor Network for Underground Water Pipeline Leakage Detection and Localization Using Pressure Residuals and Classifiers Wireless Personal Communications 139, 803–823 (2024).
23. Vankdothu, R., Cheng, X. "Energy Efficient TDMA and Secure Based MAC Protocol for WSN Using AQL Coding and ASGWI Clustering". Wireless Personal Communications 136, 2125–2143 (2024)
24. Vankdothu, R., Hameed, M.A., Fatima, H. *et al.* Multicast Scaling in Heterogeneous Wireless Sensor Networks for Security and Time Efficiency. Wireless Personal Communications (2025).
25. Vankdothu, R., Hameed, M.A., Fatima, H. *et al.* Multicast Scaling in Heterogeneous Wireless Sensor Networks for Security and Time Efficiency. Wireless Personal Communications (2025)
26. Ramdas Vankdothu, Mohd Abdul Hameed" Brain MRI Images for Tumor Detection using Storage Optimization Technique",Mobile Radio Communications and 5G Networks,Lecture Notes in Networks and Systems,425-437, Springer .
27. Bandi Krishna , Ramdas Vankdothu , Varun Revuri and B. Prashanth" A brain tumor identification using convolution neural network in the deep learning" MATEC Web of Conferences 392, 01131 (2024) ,https://doi.org/10.1051/matecconf/202439201131 ICMED 2024

8.BIBILIOGRAPHY

I'm Donthi Saikiran. I am currently in my 8th semester of Computer Science in the BachelorsDegreeatBalajiInstituteofTechnologyandScience.Myresearchinterestis done based on "AADHAAR UID MASKING TOOL".

Doi:10.48047/ijiee.2025.15.5.10

I'm Suraashi Aravind. I am currently in my 8th semester of Computer Science in the BachelorsDegreeatBalajiInstituteofTechnologyandScience.Myresearchinterestis done based on "AADHAAR UID MASKING TOOL”.

I'mShivaratriVamshiKumar.Iamcurrentlyinmy8thsemesterofComputerSciencein theBachelorsDegreeatBalajiInstituteofTechnologyandScience.Myresearchinterest is done based on "AADHAAR UID MASKING TOOL”.

I'm Kota Premkumar. I am currently in my 8th semester of Computer Science in the BachelorsDegreeatBalajiInstituteofTechnologyandScience.Myresearchinterestis done based on "AADHAAR UID MASKING TOOL”.

Doi:10.48047/ijiee.2025.15.5.10