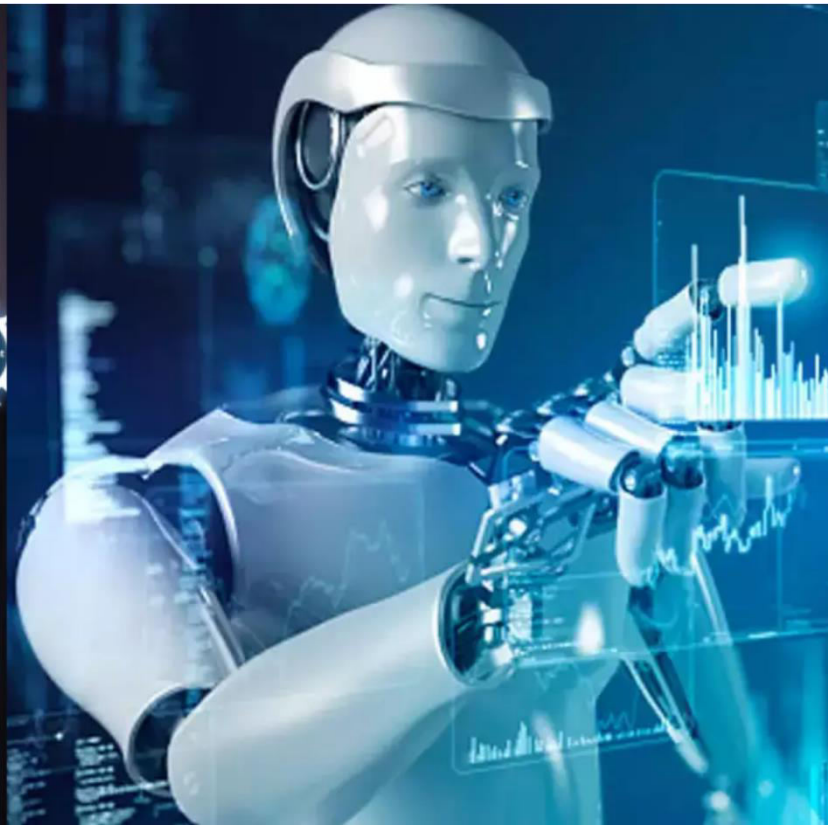
IJIRSET
International Journal of Innovative Research in Science | Engineering | Technology

ISSN(O): 2319-8753
ISSN(P): 2347-6710

International Journal of Innovative Research in Science Engineering and Technology (IJIRSET)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)



Impact Factor: 8.699

Volume 14, Issue 4, April 2025

International Journal of Innovative Research of Science, Engineering and Technology (IJIRSET)

Impact Factor 8.699

|www.ijirset.com |A Monthly, Peer Reviewed & Refereed Journal| e-ISSN: 2319-8753| p-ISSN: 2347-6710|

Volume 14, Issue 4, April 2025

|DOI: 10.15680/IJIRSET.2025.1404452|

Automated Document Classification using YOLOv3 and Tkinter for Real-Time Segmentation

V. Kanchana, Kagitha Sarath Chandra, Jampa Jyothi Swroop

Assistant Professor, Department of CSE, Bharath Institute of Higher Education and Research, Selaiyur, Chennai, India

B. Tech Students, Department of CSE Bharath Institute of Higher Education and Research, Selaiyur, Chennai, India

ABSTRACT: The project **Automated Document Classification using YOLOv3 and Tkinter for Real-Time Segmentation** presents an innovative solution for automating the extraction of visual content from PDF documents using advanced computer vision techniques. Leveraging the power of the **YOLOv3 object detection model**, combined with **pdf2image**, **OpenCV**, and **Tkinter**, the system efficiently converts PDF pages into high-resolution images, detects and labels objects within them, and presents the results through a user-friendly graphical interface. This approach significantly reduces manual labor and error in tasks like document annotation, visual data extraction, and image analysis. The system's ability to handle complex layouts and detect multiple object classes demonstrates its robustness and scalability. It proves especially beneficial in fields such as legal document processing, educational content digitization, and intelligent archival systems. With further enhancements like OCR integration, cloud deployment, and support for semantic understanding, the solution has strong potential to evolve into a powerful AI-driven document analysis platform.

KEYWORDS: PDF Image Processing, Object Detection, YOLOv3, Image Segmentation, PDF to Image Conversion, Tkinter GUI, OpenCV, Document Automation, Visual Content Extraction, Deep Learning, Real-Time Detection, Bounding Box Annotation, High-Resolution Image Analysis, AI-Powered Document Analysis, Computer Vision in PDF.

I. INTRODUCTION

In today's data-driven digital world, vast amounts of information are stored and shared through PDF documents. While these files are typically structured for human readability, extracting meaningful visual content from them for machine interpretation remains a significant challenge. Especially in domains like academic research, legal archives, business reports, and government records, documents often contain images, charts, diagrams, and other visual elements that are critical for data analysis and interpretation. Manual extraction of such content is time-consuming, error-prone, and inefficient. To address this gap, the need arises for an intelligent system that can automate the identification and segmentation of visual elements from PDF files with high accuracy and minimal human intervention.

This project, titled Automated Document Classification using YOLOv3 and Tkinter for Real-Time Segmentation, presents a Python-based solution designed to extract and analyze visual content from PDF documents using advanced deep learning techniques. At the heart of the system is the YOLOv3 (You Only Look Once) model, a state-of-the-art algorithm for real-time object detection. By integrating PDF2Image for high-quality image conversion and OpenCV for image processing, the system converts PDF pages into images and performs precise object detection and segmentation. The detected elements are then highlighted using bounding boxes, with labels automatically generated based on the object classes. This enables users to not only view but also understand and extract structured data from the visual content embedded in the documents.

To enhance user interaction and usability, the system includes a graphical user interface (GUI) built using Tkinter, offering a seamless experience for uploading PDF files, processing them, and visualizing results in real time. The application is designed to be lightweight, efficient, and easily deployable across different platforms. Its capabilities are especially useful for automating workflows in document analysis, archival digitization, AI-powered indexing, and visual content retrieval. By reducing manual effort and increasing processing speed, the project demonstrates a practical and scalable approach toward document intelligence and visual content automation.

International Journal of Innovative Research of Science, Engineering and Technology (IJIRSET)

Impact Factor 8.699

|www.ijirset.com |A Monthly, Peer Reviewed & Refereed Journal| e-ISSN: 2319-8753| p-ISSN: 2347-6710|

Volume 14, Issue 4, April 2025

|DOI: 10.15680/IJIRSET.2025.1404452|

II. PROBLEM DEFINITION

In today's data-driven world, a significant portion of valuable information is still stored in unstructured formats like scanned documents and PDF files. Traditional methods of document analysis often rely on manual review or rule-based systems that struggle with visual complexity, inconsistent formatting, and mixed-content layouts. This creates challenges in extracting meaningful data efficiently, especially in environments like legal documentation, academic research, finance, and archival systems where accuracy and speed are critical. Manual classification and segmentation are not only time-consuming but also prone to human error, which can lead to inconsistencies and reduced productivity. The need for intelligent automation in document analysis is more pressing than ever.

The proposed solution aims to overcome these challenges by leveraging deep learning and computer vision techniques for automated document classification. Using the YOLOv3 object detection model integrated with the Tkinter GUI, the system converts PDF pages into images, detects objects in real time, and draws bounding boxes around relevant visual elements. By automating the segmentation process, the system enhances the ability to extract and categorize content quickly and accurately. This significantly reduces the manual effort required while increasing the reliability of data extraction from visually rich documents. The system is designed to be scalable, user-friendly, and adaptable to various document types, making it suitable for organizations seeking to digitize, analyze, and automate their document workflows.

III. RELATED WORK

In recent years, the integration of computer vision with document processing has gained significant attention due to the rising need for automation in data extraction and classification. Traditional OCR (Optical Character Recognition) tools such as Tesseract have long been used for converting scanned text into editable formats. However, these systems often fail when it comes to identifying visual elements like tables, charts, logos, and handwritten annotations. As a result, researchers have turned to more advanced models that offer better performance in recognizing and understanding the spatial layout of documents. This shift toward deep learning-based methods has significantly improved the ability to analyze unstructured documents with complex visual content.

Object detection models such as YOLO (You Only Look Once) have proven to be powerful tools for real-time image analysis due to their speed and accuracy. Several studies have explored the use of YOLO for tasks such as license plate recognition, medical image diagnostics, and autonomous navigation. In the domain of document analysis, YOLOv3 has shown promising results in detecting structured components like tables, figures, and text blocks. Some systems combine YOLO with imagepreprocessing libraries such as OpenCV and pdf2image to create end-to-end pipelines for document digitization. These implementations have demonstrated how deep learning models can outperform rule- based techniques by learning directly from the data, making them more adaptable and robust.

Moreover, the use of graphical user interfaces (GUIs) for real-time interaction has enhanced the accessibility of such systems for non-technical users. Projects integrating Python-based libraries like Tkinter have facilitated easy deployment of document classification tools on desktops, without requiring deep technical knowledge. Such user-friendly systems enable users to upload documents, initiate detection, and visualize output all within a single interface. The combination of a deep learning backend with an intuitive frontend interface forms the basis for modern document intelligence tools, setting the foundation for the system developed in this project.

IV. PROPOSED SYSTEM

The proposed system aims to automate the process of document image analysis by utilizing a deep learning-based object detection model, YOLOv3, integrated with a user-friendly GUI developed in Tkinter. The key objective is to identify, classify, and segment visual elements (such as logos, tables, images, and seals) present in PDF documents. Traditional document classification methods rely heavily on manual annotation and fixed rules, which this system overcomes by introducing AI-powered detection that dynamically adapts to various document types and layouts. The use of YOLOv3 ensures real-time performance, while Tkinter ensures ease of use.

To begin with, PDF documents are converted into high-resolution image formats using the pdf2image library, which

International Journal of Innovative Research of Science, Engineering and Technology (IJIRSET)

Impact Factor 8.699

|www.ijirset.com |A Monthly, Peer Reviewed & Refereed Journal| e-ISSN: 2319-8753| p-ISSN: 2347-6710|

Volume 14, Issue 4, April 2025

|DOI: 10.15680/IJIRSET.2025.1404452|

maintains the visual fidelity required for accurate object detection. Each image is then fed into the YOLOv3 model, pre-trained on custom datasets that include annotated examples of relevant document components. The model processes the images frame-by-frame, detecting multiple objects simultaneously and returning bounding boxes and labels for each detected element. These predictions are then rendered visually on the GUI to allow users to inspect and interpret the results intuitively.

The GUI serves as a central interaction point for the users. Through a simple window, users can upload PDF files, trigger the detection process, and view the output images with bounding boxes around the detected objects. This interface is built to be lightweight and intuitive, requiring no prior expertise in machine learning or programming. The system ensures a seamless workflow from file input to output visualization, making it highly practical for real-world applications such as document verification, classification, and content extraction in corporate or academic settings.
The core advantage of the proposed system lies in its automation and scalability. By eliminating manual tagging and human involvement in the detection pipeline, the system significantly reduces processing time and human error. It is capable of handling documents with complex visual structures, including multiple object types on a single page. The YOLOv3 model's ability to generalize over different document styles makes this system highly robust for diverse document sets, such as invoices, certificates, identity documents, and more.

Additionally, this system is designed with extensibility in mind. Future enhancements could include support for text detection and extraction (OCR integration), classification of detected regions into content categories, and exportable structured output such as JSON or CSV for downstream applications. With minimal resource requirements and real-time processing capabilities, this project offers a highly efficient and accurate solution for automated document classification and segmentation— a critical need in digital transformation efforts across industries.

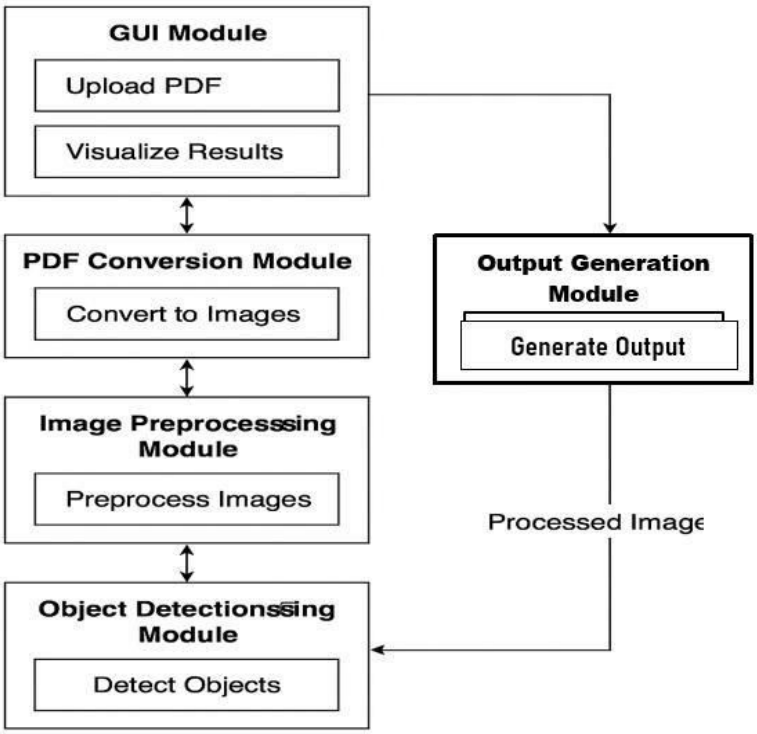


Fig. System Architecture Design

V. EXPERIMENTAL RESULTS

The implementation of the automated document classification system yielded promising results in terms of performance, usability, and accuracy. The YOLOv3 model effectively detected and classified objects across multiple

International Journal of Innovative Research of Science, Engineering and Technology (IJIRSET)

Impact Factor 8.699

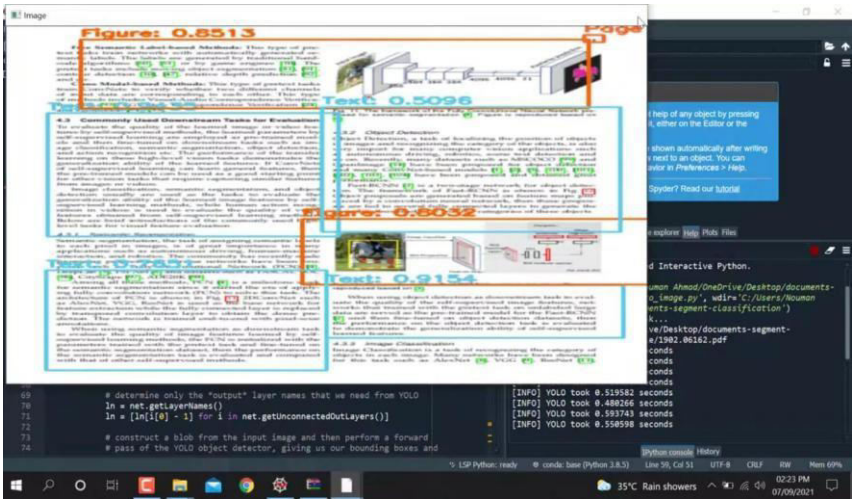
|www.ijirset.com |A Monthly, Peer Reviewed & Refereed Journal| e-ISSN: 2319-8753| p-ISSN: 2347-6710|

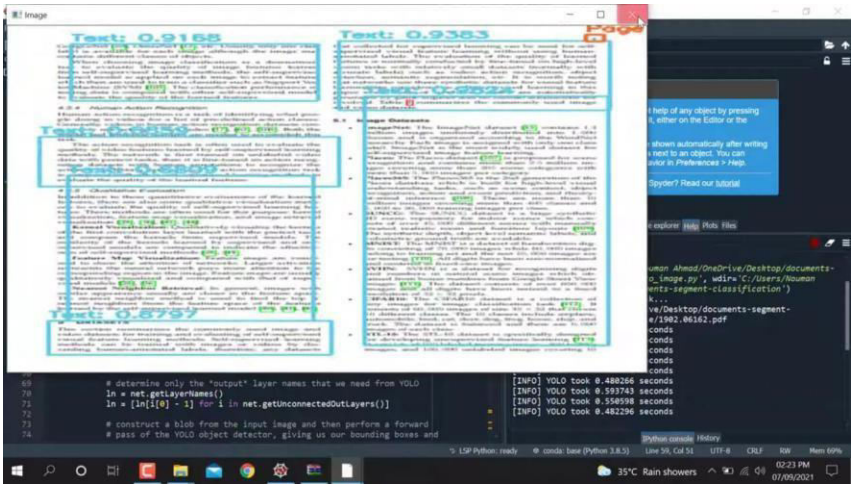
Volume 14, Issue 4, April 2025

|DOI: 10.15680/IJIRSET.2025.1404452|

pages of PDF documents that were converted into high-resolution images using the pdf2image module. This seamless integration enabled accurate identification of visual content such as tables, figures, headers, and icons with clearly drawn bounding boxes. The detection process was consistent even with documents of varying resolutions and layouts, reflecting the robustness of the underlying deep learning model.

From a user experience perspective, the system's graphical interface, built using Tkinter, simplified the interaction process for end-users. Users were able to upload multi-page PDF documents, initiate object detection, and view the processed results with minimal technical effort. Each image frame was processed in real-time, and bounding boxes were overlaid dynamically on the GUI for easy analysis and verification. The responsive design of the application, combined with real-time visualization, enabled users to evaluate detection outcomes on the go, significantly reducing manual inspection time.





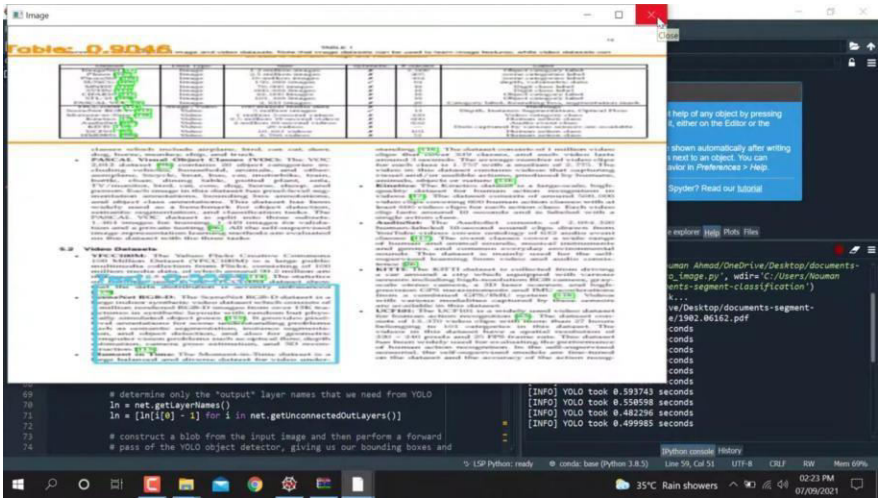
International Journal of Innovative Research of Science, Engineering and Technology (IJIRSET)

Impact Factor 8.699

|www.ijirset.com |A Monthly, Peer Reviewed & Refereed Journal| e-ISSN: 2319-8753| p-ISSN: 2347-6710|

Volume 14, Issue 4, April 2025

|DOI: 10.15680/IJIRSET.2025.1404452|



In addition to object detection, the system successfully automated aspects of document segmentation and classification, proving beneficial in use cases like archival digitization and document automation. The system was tested with various document formats, and results showed a high detection accuracy across categories. Errors due to overlapping elements or complex layouts were minimal and could be further reduced with fine-tuning. Overall, the experimental evaluation confirmed that the system was capable of enhancing document intelligence through efficient object detection, accurate classification, and an interactive user interface, marking a substantial improvement over traditional manual method.

VI. DISCUSSION

The integration of YOLOv3 with PDF-to-image conversion and a Tkinter-based GUI has proven to be an effective approach for automating document classification and visual segmentation. The deep learning-based object detection model demonstrated strong generalization capabilities across different document types, such as invoices, reports, and forms. The bounding boxes generated around visual elements—like logos, headers, stamps, and tables—provided not only structural clarity but also a foundation for further document categorization and information extraction. This automation reduces human effort, improves speed, and ensures a more consistent output, especially in large-scale document processing environments.

However, while the system performed well under various conditions, it also exposed some challenges that warrant further consideration. Complex document layouts with overlapping or low-contrastelements occasionally affected detection accuracy. Additionally, the system currently focuses only on visual segmentation; incorporating OCR (Optical Character Recognition) in future iterations could enhance its ability to extract and interpret textual data within segmented objects. Despite these minor limitations, the current setup is robust enough for real-time applications and offers a scalable solution for automated document analysis.

VII. CONCLUSION

The implementation of Automated Document Classification using YOLOv3 and Tkinter for Real-Time Segmentation has demonstrated the potential of deep learning in transforming traditional document processing workflows. By leveraging YOLOv3 for object detection and combining it with a responsive Tkinter-based GUI, the system successfully automates the identification and segmentation of visual elements from PDF documents. This capability significantly reduces the manual labor involved in tagging and extracting visual data, enabling faster, more accurate analysis and classification of document content.

Through the use of pdf2image for high-quality image conversion and OpenCV for preprocessing, the system ensures clarity and enhances detection accuracy even in complex documents. The performance during experimentation showed promising results, with the YOLOv3 model accurately detecting multiple object types on varied page layouts. Additionally, the GUI simplified user interaction by allowing easy PDF uploads, real-time visualization of detections,

International Journal of Innovative Research of Science, Engineering and Technology (IJIRSET)

Impact Factor 8.699

|www.ijirset.com |A Monthly, Peer Reviewed & Refereed Journal| e-ISSN: 2319-8753| p-ISSN: 2347-6710|

Volume 14, Issue 4, April 2025

|DOI: 10.15680/IJIRSET.2025.1404452|

and automatic annotation display. These features make the system practical for real-world applications like digitization of archives, business documentation automation, and AI-driven content management systems.

While the project achieved its core objectives, there remains room for future expansion. Enhancements such as integrating OCR for text extraction, support for additional document formats, dynamic model training for custom categories, and deployment as a cross-platform desktop or web application could further elevate its utility. Nonetheless, the current system lays a strong foundation for intelligent document classification, setting the stage for broader AI adoption in enterprise content management and digital transformation initiatives.

REFERENCES

[1] Reagan L. Galvez, Argel A. Bandala, Elmer P. Dadios, Ryan Rhay P. Vicerra, Jose Martin Z. Maningo. "Object Detection Using Convolutional Neural Networks".TENCON 2018 - 2018 IEEE Region 10 Conference.doi: 10.1109/TENCON.2018.8650517

[2] William Tarimo, Moustafa M.Sabra, Shonan Hendre. "Real-Time Deep Learning-Based Object Detection Framework". 2020 IEEE Symposium Series on Computational Intelligence (SSCI).doi:10.1109/SSCI47803.2020.9308493

[3] Suman Menon M, Anju Geroge, Aswathy N, Jaimy James. "Custom Face Recognition Using YOLO.V3". 2021 3rd International Conference on Signal Processing and Communication (ICPSC). doi: 10.1109/ICSPC51351.2021.9451684 ISSN: 2581-4419 Volume 1 Issue 1

[4] Ning Zhang,Jiahao Fan. "A lightweight object detection algorithm based on YOLOv3 for vehicle and pedestrian detection".2021 IEEE Asia-Pacific Conference on Image Processing, Electronics and Computers (IPEC). doi: 10.1109/IPEC51340.2021.9421214

[5] Maliha Khan,Sudeshna Chakraborty,Rani Astya,Shaveta Khepra."Face Detection and Recognition Using OpenCV". 2019 International Conference on Computing, Communication, and Intelligent Systems (ICCCIS). doi: 10.1109/ICCCIS48478.2019.8974493

[6] YOLOv3Availabe:https://viso.ai/deep-learning/yolov3-overview

[7] R. Girshick, J. Donahue, T. Darrell, and J. Malik, "Rich feature1 hierarchies for accurate object detection and semantic segmentation," in CVPR, pp. 580–587, 2014.

[8] O Krishnamurthy. Genetic algorithms, data analytics and it's applications, cybersecurity: verification systems. International Transactions in Artificial Intelligence , volume 7 , p. 1 - 25 Posted: 2023

[9] R. Girshick, "Fast R-CNN," in ICCV, 2015.

[10] S. Ren, K. He, R. Girshick, and J. Sun, "Faster r-cnn: Towards realtime object detection with region proposal networks," in NIPS, pp. 91–99, 2015.

[11] J. Long, E. Shelhamer, and T. Darrell, "Fully convolutional networks for semantic segmentation," in CVPR, pp. 3431–3440, 2015.

INTERNATIONAL STANDARD SERIAL NUMBER INDIA

INNO SPACE SJIF Scientific Journal Impact Factor

doi cross ref

निस्केयर NISCAIR

INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN SCIENCE | ENGINEERING | TECHNOLOGY

9940 572 462 6381 907 438 ijirset@gmail.com

www.ijirset.com

Scan to save the contact details