The Journal of Computational Science and Engineering (TJCSE)
ISSN 2583-9055 (Media Online)
Vol 3, No 11, November 2025
PP 1–12

OCR Enabled Health Analysis using ABHA Integration

Yuvraj Singh Chauhan[1], Prajwal Chinchmalatpure[2], Suyash Chinchmalatpure[3], Prashant Yelmar[4], Sai Jagdale[5], Rajendra G. Pawar[6]

[1,2]Northeastern University, Boston, United States, [3]University of Toronto, Rotman, Toronto, Canada, [4,5,6]MIT School of Computing, MIT ADT University, Pune, India.

Corresponding Author *: rgpawar13@gmail.com, prajwalvvc@gmail.com

Abstract

Our research focuses on the advancement of document scanning technology utilizing OCR techniques to improve health analysis [1,2,3,4,5]. We aim to address the challenges encountered in maintaining health records by exploring innovative approaches in data extraction. Main investigation integrates cutting-edge technologies such as Flutter for front-end development and Type Script for backend services. By harnessing OCR technology, specifically powered by Paddle OCR [6,7], we seek to efficiently extract pertinent information from various documents, potentially enhancing diagnostic procedures and data management in healthcare settings. Additionally, our exploration extends to integration with existing healthcare systems, such as the Indian government's Unified Health Interface (UHI) system through ABHA, to facilitate seamless data exchange and elevate the efficacy of health analysis tools [4,10]. With a commitment to user- friendly interfaces, robust security protocols, and future for machine learning integration, our research endeavors to advance health analysis capabilities and provide actionable insights for both patients and healthcare professionals.

Keywords:
OCR technology, Health analysis, Document scanning, Diabetes management, Healthcare technology, Unified Health Interface (UHI) system

1. Introduction

In an era where technological advancements continually reshape the landscape of healthcare, our research initiative embarks on a journey at the crossroads of document scanning technology and health analysis. With a central focus on integrating OCR, we aim to unveil the transformative potential of this technology in enhancing healthcare outcomes. At its core, our research seeks to

revolutionize health analysis by harnessing OCR technology to simplify the extraction of critical health information from a myriad of medical documents. From deciphering lab reports to parsing prescriptions, OCR holds the promise of streamlining data entry processes and mitigating the inherent risks of errors associated with manual transcription.

By automating the extraction of pertinent health data, our endeavor endeavors to empower both patients and healthcare professionals with accurate and readily accessible information. This accessibility not only facilitates informed decision-making but also fosters a collaborative healthcare environment where data-driven insights drive superior patient care. Our exploration transcends the realms of document scanning, extending into the seamless integration of OCR technology within healthcare systems. The Unified Health Interface (UHI) system serves as a focal point for this integration, providing a cohesive framework for secure data exchange and interoperability across various healthcare platforms.

Through this integration, real-time health information becomes readily available, facilitating enhanced communication and coordination among healthcare stakeholders. Beyond its practical applications, our research delves into the broader implications of OCR technology in transforming healthcare workflows and improving patient outcomes. By enabling efficient data management and analysis, OCR-enabled solutions have the potential to revolutionize how healthcare providers diagnose, treat, and monitor patients. This shift towards more personalized and effective care heralds a new era in healthcare delivery.

In summary, our research initiative represents a concerted effort to harness the capabilities of OCR technology to drive innovation in healthcare. Through our interdisciplinary approach, we seek to uncover new avenues for leveraging OCR technology to streamline document scanning processes, enhance health analysis, and ultimately elevate the delivery of healthcare services to unprecedented heights.

2. Literature Survey

Early studies on churn prediction primarily relied on statistical methods and traditional data mining Table1 shows review summarizes research articles focusing on OCR technology and its potential applications in healthcare. Each article contributes valuable insights into OCR systems, their features, performance evaluations, challenges, advancements, and prospects, particularly in the context of digitizing medical documents, automating data extraction, and facilitating health analysis processes.

The Journal of Computational Science and Engineering (TJCSE)
ISSN 2583-9055 (Media Online)
Vol 3, No 11, November 2025
PP 1-12

Table 1. Literature Review

Title	Focus
OCRopus - A Free Document Analysis and OCR System [1]	Introduction of OCRopus, an open-source system for document analysis and OCR, highlighting its features and performance evaluations across datasets.
Intelligent OCR System for Automated Document Classification and Data Extraction [2]	Proposes an intelligent OCR system integrating machine learning with OCR for automated document classification and data extraction, demonstrating promising results.
A Detailed Analysis of Optical Character Recognition Technology [3]	Provides a comprehensive analysis of OCR techniques, discussing preprocessing, segmentation, feature extraction, and recognition algorithms, along with challenges, advancements, and future directions.
Optical Character Recognition: Recent Advances and Future Prospects [4]	Reviews recent advances in OCR, including traditional and deep learning-based approaches, examining improvements in accuracy, efficiency, applications, challenges, offering insights into leveraging OCR for enhanced health analysis.

3. Survey Methodology

Table 2. Health Analysis Survey Methodology

Stage	Description
Questionnaire Development	Tailored questionnaires created for medical professionals and users. - Medical professional questionnaires focused on familiarity, feature preferences, and perceived benefits of OCR tools. User questionnaires focused on document management practices, challenges, desired functionalities, and adoption willingness.
Initial Feedback Collection	Questionnaires distributed within personal networks (family, friends, and colleagues). This approach ensured familiarity and willingness to contribute insights.
Continuous Feedback	Encouraged ongoing dialogue with participants to address questions and gather additional insights. - This iterative process led to a deeper understanding of participant's needs.
Feedback Integration	Feedback from both groups incorporated into the research framework. This ensured alignment with target user needs and expectations.

We developed a questionnaire tailored for both medical professionals and users to gather insights on the utilization of OCR technology in health analysis shown in above table 2

4. Experimental Setup and Implementation

The identification of the problem that prompted our research on scanning documents using OCR and benefiting health analysis stemmed from a thorough examination of the healthcare landscape in India, with a specific focus on the challenges related to health record fragmentation and disease management. Several key factors contributed to the recognition of this pressing issue:

High Prevalence of Chronic Diseases: India has been experiencing a significant rise in the prevalence of chronic diseases, including diabetes, cardiovascular diseases, and others. According to recent data, India is grappling with a substantial burden of chronic conditions, underscoring the need for effective disease management solutions.

Fragmented Health Records: A notable challenge observed in India's healthcare system was the fragmentation of health records among various healthcare providers and facilities. Patients often encountered difficulties in accessing their complete health information, which was scattered across different clinics, hospitals, and diagnostic centers. This fragmentation posed challenges to comprehensive disease management and hindered seamless information exchange between patients and healthcare providers.

Inefficient Information Exchange: Communication and information exchange between patients and healthcare professionals were found to be inefficient and often cumbersome. Patients struggled to share their health records with different healthcare providers, leading to disjointed care and missed opportunities for optimal treatment outcomes. Healthcare professionals face challenges in accessing comprehensive patient histories, impacting their ability to make well-informed clinical decisions.

Limited Technological Solutions: Existing digital health solutions available in the Indian market often fell short in addressing the complexities of health record management and disease monitoring. Many lacked robust features, user-friendly interfaces, or failed to cater to the specific needs of Indian patients and healthcare providers. This highlighted the need for innovative technological solutions that could streamline health record management and enhance disease management practices. Recognizing these challenges and understanding the importance of a cohesive approach to health record management and disease monitoring, our research focused on exploring the potential of OCR technology to address these issues effectively. Extensive market analysis, user surveys, and consultations with healthcare experts further

The Journal of Computational Science and Engineering (TJCSE)
ISSN 2583-9055 (Media Online)
Vol 3, No 11, November 2025
PP 1–12

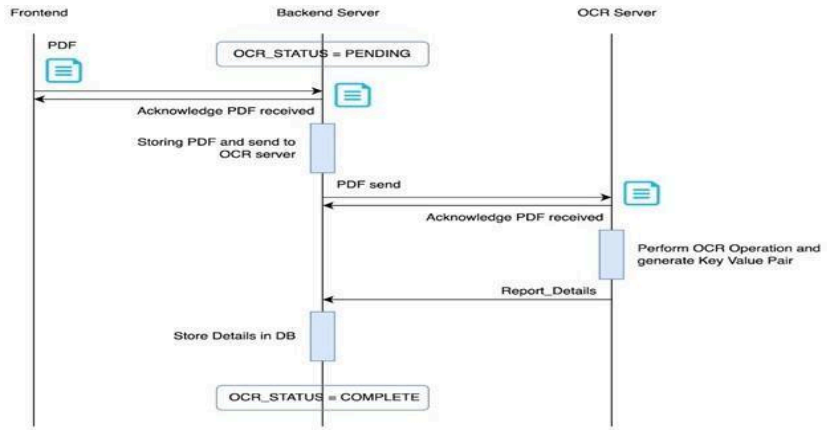
validated the need for a comprehensive solution that could streamline document scanning using OCR and facilitate health analysis in the Indian context.

By identifying and understanding the unique challenges faced by individuals' managing chronic diseases in India, our research aims to develop innovative solutions that leverage OCR technology to revolutionize health record management and disease monitoring practices. The integration of OCR technology into healthcare workflows holds the promise of enhancing efficiency, improving communication between patients and healthcare providers, and ultimately, advancing healthcare outcomes in India.

4. Implementation

The methodology employed for the implementation of the OCR-enabled health analysis system adheres to an iterative development approach, ensuring flexibility and adaptability throughout the project lifecycle. The iterative process involves continuous feedback loops and collaboration among stakeholders, and end users. Regular reviews and retrospectives are conducted to assess progress and identify areas for improvement. Additionally, user-centric design principles guide the development process, prioritizing the needs and preferences of end users to deliver a solution that effectively addresses their requirements and expectations

Algorithms: In the implementation of the OCR-enabled health analysis system, various algorithms play crucial roles in enabling key functionalities. Some of the algorithms utilized include OCR: The system leverages advanced OCR algorithms, such as Paddle OCR, Tesseract OCR, Easy OCR or other OCR services provided by Google, Amazon, Azure and many more to accurately ex- tract pertinent information from scanned medical documents. These algorithms employ machine learning techniques to process images and recognize text data efficiently.



The Journal of Computational Science and Engineering (TJCSE)
ISSN 2583-9055 (Media Online)
Vol 3, No 11, November 2025
PP 1–12

Fig. 1. Application workflow

Data Visualization: The system applies algorithms for data aggregation, trend analysis, and visualization to provide insightful representations of health metrics and trends. Visualization techniques like charts, graphs, and histograms are employed to enhance data interpretation and analysis.

User Authentication and Security: Cryptographic algorithms, including hashing and encryption techniques, are utilized to ensure secure user authentication and data protection. These algorithms safeguard sensitive user information and facilitate secure communication between the system components. It is important to acknowledge that the selection and implementation of algorithms may vary depending on specific project requirements and technical considerations. The algorithms mentioned offer a glimpse into the diverse range of algorithms involved in enabling document scanning using OCR and deriving valuable health insights for enhanced healthcare analysis.

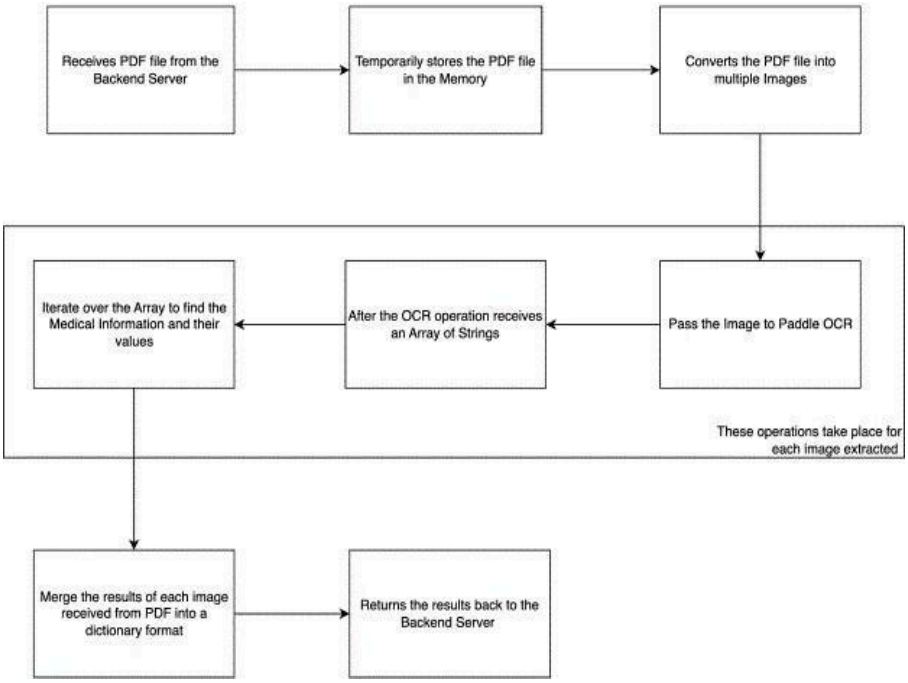


Fig. 2. OCR implementation Logic

ABHA: In the context of scanning documents using OCR for health analysis, the ABHA sandbox provided by governmental authorities serves as a pivotal platform facilitating integration with various government APIs. ABHA acts as a bridge between the OCR-enabled system and the government's Unified Health Interface (UHI), enabling seamless interaction and

The Journal of Computational Science and Engineering (TJCSE)
ISSN 2583-9055 (Media Online)
Vol 3, No 11, November 2025
PP 1–12

data exchange. Through ABHA, the system gains access to essential government APIs pertaining to healthcare and patient data. These APIs grant the system the capability to retrieve vital information such as patient records and diagnostic reports stored within the UHI system

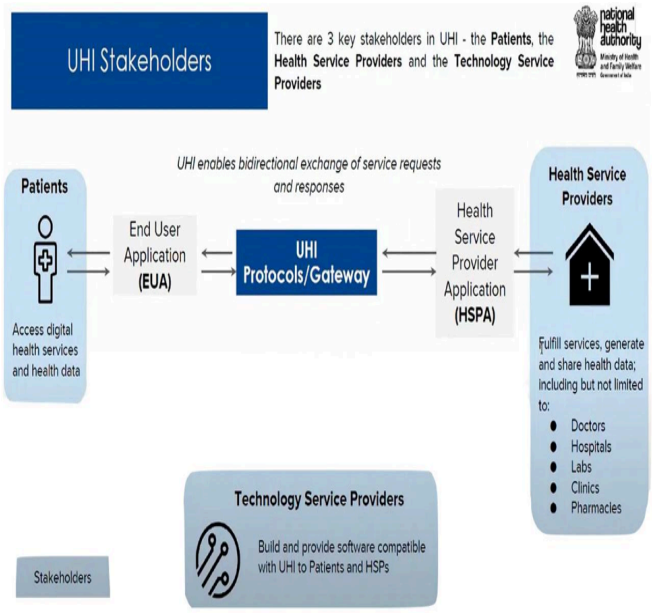


Fig. 3. Unified Health Interface

By integrating with these APIs, the OCR-enabled system empowers users to securely access and share their health information with healthcare providers, thereby facilitating accurate diagnosis and treatment. Leveraging the UHI system's extensive healthcare data enhances the system's functionality, enriching health analysis capabilities.

The Journal of Computational Science and Engineering (TJCSE)
ISSN 2583-9055 (Media Online)
Vol 3, No 11, November 2025
PP 1–12

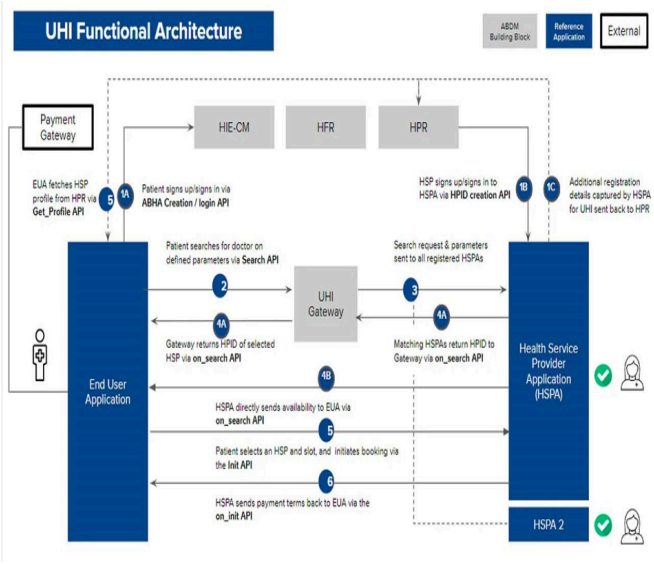


Fig. 4. UHI Functional Architecture (Source: National Health Authority (NHA))

Overall, the integration of ABHA with government APIs fosters a seamless connection between users, their health records, and healthcare providers. This integration plays a crucial role in improving healthcare management and decision-making processes within the context of OCR-enabled health analysis systems. OCR Technology: In the realm of scanning documents for health analysis, the implementation of Optical Character Recognition (OCR) technology is a vital process with several crucial steps aimed at extracting pertinent information from uploaded documents. Here's an overview of the OCR implementation process. Document Processing: Upon document upload, the system processes the document, extracting individual pages and converting them into image formats suitable for OCR analysis. This step ensures that each page can be independently processed by the OCR algorithm.

OCR Algorithm: The system employs advanced OCR algorithms, such as Paddle OCR, which utilize deep learning techniques including convolutional neural networks (CNN) and recurrent neural networks (RNN) to recognize and extract text from images. Pre-processing: Prior to OCR analysis, the system may apply pre-processing techniques to enhance image quality and improve OCR accuracy. These techniques may include resizing, noise reduction, contrast adjustment and binarization. Text Extraction: The OCR algorithm analyzes preprocessed images, identifying text regions and performing text recognition to convert image-based text into machine-readable data.

Information Extraction: Extracted text is further processed to identify and extract specific information relevant to health analysis. This may include extracting values such as blood sugar levels, HbA1c values, and other health metrics crucial for analysis and management.

The Journal of Computational Science and Engineering (TJCSE)
ISSN 2583-9055 (Media Online)
Vol 3, No 11, November 2025
PP 1–12

Data Utilization: Extracted data is then utilized within the system for various purposes, including generating insights, visualization, and sharing with healthcare providers. The seamless integration of OCR-extracted data enhances the system's capabilities for health analysis and management.

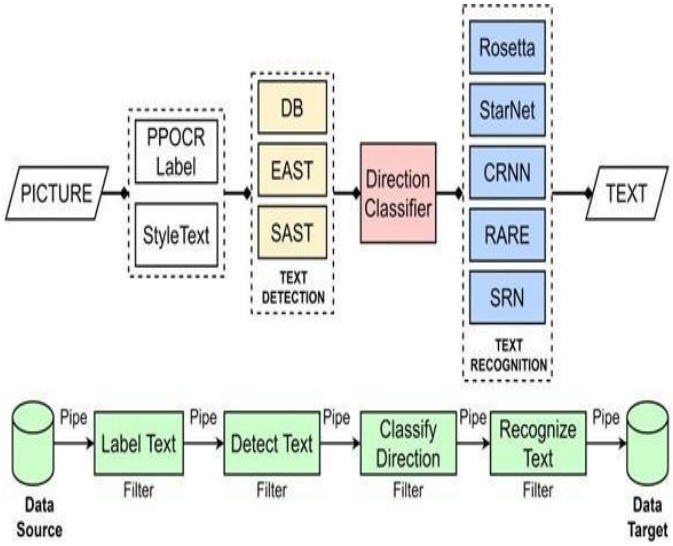


Fig. 5. Workflow of Paddle OCR

Internal Architecture of Paddle OCR By leveraging advanced OCR algorithms and preprocessing techniques, the system effectively automates data extraction from scanned documents, contributing to streamlined health analysis processes and improved healthcare outcomes [11,13].

User Engagement and Satisfaction: The research incorporates feedback from users involved in testing the OCR-enabled health analysis system. Gathering user feedback through surveys, interviews, and iterative feedback loops aids in identifying strengths, weaknesses, and areas for enhancement in system functionality and user experience.

Effectiveness of OCR-Based Health Analysis: The research evaluates the effectiveness of OCR technology in capturing relevant health information from scanned documents. This assessment encompasses the accuracy of data extraction, including blood sugar levels, medication adherence, and dietary patterns. It aims to gauge how well the OCR-enabled system supports users in managing their health-related information effectively[8,9].

Data Interpretation and Visualization: Analysis of the extracted data enables the identification of trends, correlations, and actionable insights. Statistical methods are employed to discern patterns in health metrics, such as blood sugar fluctuations and medication adherence rates. Visual representations, such as charts and graphs, aid in comprehending trends and informing decision-making processes[12].

The Journal of Computational Science and Engineering (TJCSE)
ISSN 2583-9055 (Media Online)
Vol 3, No 11, November 2025
PP 1–12

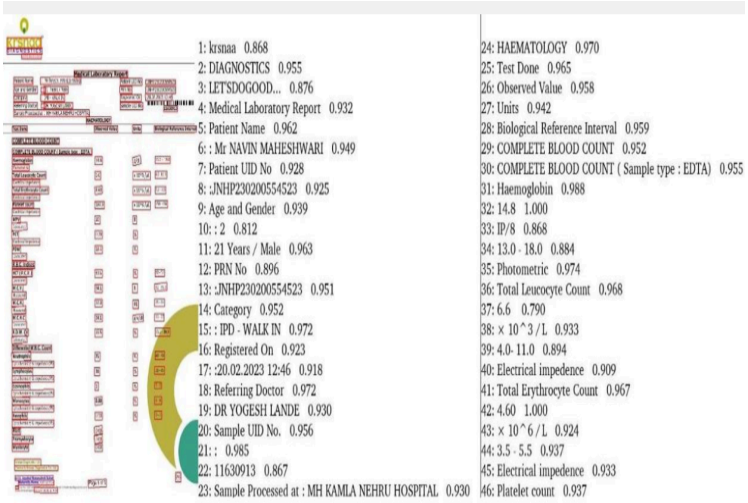


Fig. 6. OCR Result

Impact on Health Management: The research assesses the impact of employing OCR technology for health analysis on overall health management. This entails evaluating improvements in health outcomes, such as better blood sugar control and adherence to treatment regimens. Additionally, user-reported benefits, such as heightened awareness and improved communication with healthcare providers, are considered to understand the broader impact of OCR-enabled health analysis.

Integration with Healthcare Systems: The analysis evaluates the system's efficacy in integrating with existing healthcare systems and facilitating seamless data exchange with healthcare providers. It examines the ease of sharing health data and reports, support enhanced diagnosis and treatment decision-making processes.

The results and analysis provide valuable insights into the performance, user satisfaction, and implications of employing OCR technology for health analysis through document scanning. These findings inform the refinement of system features, usability enhancements, and strategic decisions for future advancements in OCR-enabled health analysis systems.

5. CONCLUSION

The utilization of OCR technology for health analysis through document scanning marks a significant advancement in healthcare management. The project's endeavors demonstrate a concerted effort to streamline data entry processes, enhance user experience, and integrate healthcare systems for improved health outcomes. By extracting pertinent information from scanned documents, the technology simplifies data acquisition and augments health analysis capabilities.

The project's adoption of state-of-the-art technologies underscores its commitment to innovation and effectiveness. Leveraging cross-platform compatibility, advanced OCR algorithms, and secure data integration, it establishes a robust framework for comprehensive health analysis and management. Moreover, the project's alignment with government initiatives, such as the UHI through ABHA, underscores its potential to facilitate seamless information exchange and collaborative healthcare efforts. Through user-centric design, meticulous feedback gathering, and consultation with healthcare experts, the project ensures its relevance and effectiveness in addressing user needs. Furthermore, stringent security measures safeguard user privacy and confidentiality, in- stilling trust and confidence in the system.

Looking ahead, the project holds promise for broader applications beyond its cur- rent scope. Its potential spans across chronic disease management, telemedicine, health analytics, and collaborative healthcare endeavors. By fostering innovation and collaboration, the project contributes to a future where healthcare is personalized, integrated, and accessible, ultimately leading to improved health outcomes and well-being for individuals across diverse healthcare domains.

References

[1] Verma, Rakesh, and Virender Ranga. "Intelligent OCR System for Automated Document Classification and Data Extraction." IEEE Access 8 (2020): 17064-17072.

[2] R. Pawar et al., "Crop Advancement with Machine Learning," 2023 7th International Conference on Computing, Communication, Control and Automation (ICCUBEA), Pune, India, 2023, pp. 1-6, Doi: 0.1109/ICCUBEA58933.2023.10392151.

[3] Pawar, R. G., Chinchmalatpure, P., Chinchmalatpure, S., Pujeri, R. V., Wategaonkar, D., Jadhav, N., Kapare, S., & Naik, D. N. (2025). Social threat identification on X via sentiment analysis: An SVM and LSTM-based approach. International Journal of Applied Mathematics, 38(3s), 565–581. DOI: https://doi.org/10.12732/ijam.v38i3s.167

[4] Limkar, S. et al. (2021). Classification and Prediction of Rice Crop Diseases Using CNN and PNN. In: Satapathy, S., Zhang, YD., Bhateja, V., Majhi, R. (eds) Intelligent Data Engineering and Analytics. Advances in Intelligent Systems and Computing, vol 1177. Springer, Singapore. https://doi.org/10.1007/978-981-15-5679-1_4

[5] Kumar, Ajit, et al. "A Detailed Analysis of Optical Character Recognition Technology." ResearchGate, 2016.

[6] Mishra, Deepak, and Rajesh Kumar. "Optical Character Recognition: Recent Advances and Future Prospects." Frontiers in Genetics 9 (2018): 515.

[7] Smith, Ray, et al. "OCR: A Review of Traditional Techniques and Recent Advancements." International Journal of Document Analysis and Recognition (IJDAR) 22.2 (2019): 121-154.

[8] Kuldeep Vayadande, Amol A. Bhosle, Rajendra G. Pawar, Deepali J. Joshi, Preeti A. Bailke, Om Lohade, Innovative approaches for skin disease identification in machine learning: A comprehensive study, Oral Oncology Reports, Volume 10, 2024, 100365, ISSN 2772-9060, https://doi.org/10.1016/j.oor.2024.100365

[9] S. Tyagi, K. Panchal, P. Chitta, S. Todi, S. Priya and R. Pawar, “Emotionomics: Pioneering Depression Detection Through Facial Expression Analytics," 2023 7th International Conference On Computing, Communication, Control And Automation (ICCUBEA), Pune, India, 2023, pp. 1-4, doi: 1109/IC-CUBEA58933.2023.10392157.

[10] Limkar, S., Baser, S., Jhamnani, Y., Shinde, P., Jithu, R., Chinchmalatpure, P. (2020). Chatbot: Integration of Applications Within Slack Channel. In: Bhateja, V., Satapathy, S., Zhang, YD., Aradhya, V. (eds) Intelligent Computing and Communication. ICICC 2019. Advances in Intelligent Systems and Computing, vol 1034. Springer, Singapore. https://doi.org/10.1007/978-981-15-1084-7_53

[11] Dixit, B. ., G. Pawar, R. ., Gayakwad, M. ., Joshi, R. ., Mahajan, A. ., & Chinchmalatpure, 6Suyash V. . (2023). Challenges and a Novel Approach for Image Captioning Using Neural Network and Searching Techniques. International Journal of Intelligent Systems and Applications in Engineering, 11(3), 712–720.

[12] P. Paygude, M. Gayakwad, D. Wategaonkar, R. Pawar, R. Pujeri, R. Joshi Dried fish dataset for Indian seafood: a machine learning application Data Br., 55 (2024), Article 110563, 10.1016/j.dib.2024.110563 ISSN 2352-3409

[13] Limkar, S., Hunashimarad, S., Chinchmalatpure, P., Baj, A., Patil, R. (2021). Potential of Robust Face Recognition from Real-Time CCTV Video Stream for Biometric Attendance Using Convolutional Neural Network. In: Satapathy, S., Zhang, YD., Bhateja, V., Majhi, R. (eds) Intelligent Data Engineering and Analytics. Advances in Intelligent Systems and Computing, vol 1177. Springer, Singapore. https://doi.org/10.1007/978-981-15-5679-1_2