
Automated Property and Risk Assessment System for Maharashtra

Pranav Divekar, Komal Shendage, Siddhi Kadam, Vedant Dongare & Prof. Prashant Sadaphule

Department of Computer Engineering, AISSMS Institute of Information Technology, Pune, India, pranavd4441@gmail.com, shendagekomal11@gmail.com, siddhikadam1803@gmail.com, vedantdongare10@gmail.com, prashant.sadaphule@aissmsioit.org

Abstract.

This paper presents an Automated Property and Risk Assessment System for Maharashtra that integrates government property records, OCR, computer vision, and machine learning for automated verification and risk assessment. Experiments achieved 96.8% extraction accuracy and 94.8% risk-classification accuracy, demonstrating improved efficiency, consistency, and traceability in property verification.

Keywords. Automated property verification, Maharashtra land records, IGR Maharashtra, OCR, YOLO, OpenCV, risk assessment, property intelligence, multilingual summarization.

1. INTRODUCTION

Property transactions in India require reliable verification of ownership, registration, and land records, but information is often fragmented across government portals and legal documents. This creates challenges for manual due diligence and increases the risk of inconsistencies and disputes [2, 5, 8, 11]. This study proposes a modular Automated Property and Risk Assessment System comprising data retrieval, document integration, multilingual summarization, and risk assessment. The system aims to convert heterogeneous property records into structured information and provide interpretable risk indicators, with the current implementation focusing mainly on automated data retrieval.

2. RELATED WORK

Recent research has applied deep learning, OCR, transformers, web extraction, and machine learning to document understanding, Legal AI, and anomaly detection [1, 3, 4, 6, 7, 9–11]. However, existing studies mainly address individual tasks. The proposed system integrates government-portal retrieval, OCR, document processing, multilingual summarization, and hybrid risk assessment into a unified property-verification workflow. It also supports 7/12 and mutation-record integration, OpenCV and YOLO-based extraction, and modular architecture for scalable property-risk analysis.

3. PROPOSED SYSTEM

The system accepts inputs such as district, taluka, village, property identifier, and time range. These inputs pass through interconnected modules that progressively transform raw government data into structured records and risk indicators

3.1. System Architecture

The architecture includes four modules: Data Retrieval, Document Integration, Multilingual Summarization, and Risk Assessment. The retrieval module is implemented and tested, while the others are designed for future development.

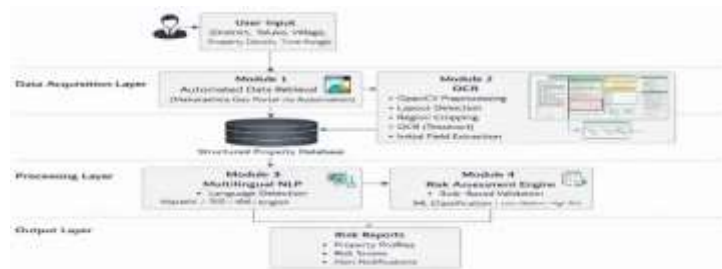


Figure 3.1.1: Workflow and data-flow architecture of the proposed automated property verification system

3.2. Data Retrieval

Selenium, OCR, and retries automate property-data retrieval and extract transaction details into a database.

3.3 Document Integration

7/12 records, mutation certificates, and transactions are combined using OCR, entity extraction, and fuzzy matching.

3.4 Multilingual Summarization

Transformer models generate summaries in English, Hindi, and Marathi.

3.5 Risk Assessment

Rule-based and machine-learning methods identify inconsistencies and classify properties as low, moderate, or high risk..

3.6 Technology Stack and Data Flow

The system uses Python, Selenium, Tesseract, OpenCV, YOLO, REST APIs, relational databases, and NLP models.

4. EXPERIMENTAL RESULTS

4.1 Experimental Setup

Experiments evaluated the retrieval module using 250 IGR Maharashtra property transactions across districts, document types, and registration years. The system used Python 3.11, Selenium, SQLite, REST APIs, and a CPU-based system with 16 GB RAM.

Performance was measured using extraction accuracy, completeness, OCR accuracy, risk-classification performance, and processing time.

4.2 Data Retrieval Results

Under dynamic web conditions, the retrieval module achieved the results in Table 2. The measurements indicate robust extraction and parsing performance across varying portal conditions.

Table 1: Performance metrics of the data retrieval module.

Metric	Value (%)
Extraction Accuracy	96.8
Record Completeness	95.4
Successful Query Rate	97.2
Data Parsing Accuracy	96.1

4.3 Region Detection and OCR

Computer vision segmented 7/12 records for OCR. Among 120 records, Tesseract provided fast and good multilingual recognition, while PaddleOCR had better detection accuracy.

Table 2: Comparison of OCR models.

OCR Model	Detection	Marathi	English	Speed
Tesseract OCR	Moderate	High	High	Fast
PaddleOCR	High	Moderate	High	Moderate
EasyOCR	Moderate	Low	Moderate	Slow

4.4 Risk Classification

The risk module was tested on 150 labeled property profiles. Table 4 shows strong performance, particularly for high-risk cases, attributed to rule-based validation and pattern recognition.

Table 3: Risk classification performance.

Risk Level	Precision (%)	Recall (%)	F1 (%)
Low	94.2	95.1	94.6
Moderate	91.3	89.8	90.5
High	96.8	97.4	97.1

Overall Accuracy: 94.8



Figure 4.4.1: Region detection and segmentation results on a 7/12 document.

4.5 *Efficiency and Scalability*

Average processing time per property was 6.4 s for data retrieval, 4.1 s for document integration, 3.8 s for summarization, and 2.6 s for risk assessment, giving a total of 16.9 s. The results indicate a substantial reduction in processing time compared with manual verification and provide a traceable basis for scaling the workflow.

5. DISCUSSION

The results show that automated retrieval and OCR can effectively structure property records, although poor-quality and handwritten documents remain challenging. Combining rule-based and analytical methods improves high-risk property identification.

6. CONCLUSIONS

The proposed framework achieved 96.8% extraction accuracy and 94.8% risk-classification accuracy, demonstrating effective automated property verification. Future work will improve OCR, risk assessment, database scalability, and the user interface.

7. REFERENCES

- [1] K. Boutalbi, F. Loukil, H. Verjus, D. Teliason and K. Salamanan, "Machine learning for text anomaly detection: A systematic review," in Proc. IEEE 47th Annual Computers, Software, and Applications Conference (COMPSAC), pp. 1319–1324, 2023.
- [2] Department of Land Resources, Government of India, "Digital India Land Records Modernization Programme (DILRMP)," 2023.
- [3] Gupta, S. Furniturewala, V. Kumari and Y. Sharma, "Rhetorical role labeling of legal documents using transform-ers and graph neural networks," arXiv:2305.04100, 2023.
- [4] Majebraries, "Selenium web scraping: How to scrape dynamic sites step-by-step," ScraperAPI Blog, 2025.
- [5] Ministry of Law and Justice, Government of India, "National Judicial Data Grid Statistics," 2024.
- [6] P. Neelamegam and S. J. Nirmala, "Recent trends in Legal AI: A comprehensive review," in Proc. ICAISS, pp. 152–159, 2025.
- [7] S. K. Nigam, T. Dubey, G. Sharma, N. Shallum, K. Ghosh and A. Bhattacharya, "LegalSeg: Unlocking the structure of Indian legal judgments through rhetorical role classification," in Findings of NAACL, pp. 1129–1144, 2025.
- [8] NITI Aayog, "Land Records and Titles in India: Issues and Reforms," Government of India, 2021.
- [9] D. Premasiri, T. Ranasinghe, R. Mitkov, M. El-Haj and I. Frommholz, "Survey on legal information extraction: Current status and open challenges," Knowledge and Information Systems, vol. 67, no. 12, pp. 11287–11358, 2025.
- [10] D. Sporici, E. Cusnic and C.-A. Bolangiu, "Improving the accuracy of Tesseract 4.0 OCR engine using convolution-based preprocessing," Symmetry, vol. 12, no. 5, p. 715, 2020.
- [11] N. Subramani, A. Matton, M. Greaves and A. Lam, "A survey of deep learning approaches for OCR and document understanding," arXiv:2011.13534, 2020.

Biographies



Pranav Divekar B.Tech. in Computer Engineering (2026), AISSMS Institute of Information Technology, Pune, with interests in AI, data processing, automation, and software development.



Vedant Dongare B.Tech. in Computer Engineering (2026), AISSMS Institute of Information Technology, Pune, with interests in AI, software development, data processing, and intelligent systems.



Siddhi Kadam B.Tech. in Computer Engineering (2026), AISSMS Institute of Information Technology, Pune, with interests in intelligent computing, data analysis, software engineering, and automation.



Komal Shendage received the B.Tech degree in Computer Engineering from the AISSMS Institute of Information Technology, Pune, India, in 2026. Her academic interests include computer engineering, intelligent systems, data processing, software development, and emerging technologies.



Prashant Shivling Sadaphule Assistant Professor of Computer Engineering at AISSMS Institute of Information Technology, Pune, with an M.E. in Computer Engineering and 17 years of academic and professional experience.