

ScholarSense AI: An OCR-Driven Pre-Submission Intelligence System for Government Scholarship Eligibility Assessment, Document Validation, and Multi-Portal Status Interpretation

Shalini H R¹

MCA Student, GM University

Davangere, Karnataka, India

School of Computer Application

shalinidvg16@gmail.com

Varun K S²

Assistant Professor, GM University

Davangere, Karnataka, India

Faculty of Computing and IT

varunksbenur@gmail.com

Abstract

India's National Scholarship Portal (NSP) and State Scholarship Portal (SSP) serve over 135 lakh applicants annually, disbursing ₹1825 crore in 2022-23. Despite the scale, eligible students routinely lose scholarships due to five avoidable failure modes: pre-submission document errors, Aadhaar-DBT linkage failures, cryptic portal status codes, missed deadlines, and absence of personalised guidance. We present ScholarSense AI, a web-based OCR-driven scholarship intelligence platform built on React 18, Node.js, Express, and MongoDB. The system implements five modules: a pre-submission document health check engine, a rule-based eligibility matcher for 30+ NSP/SSP schemes, a multi-portal OCR screenshot status analyzer, an automated deadline notification engine (email + WhatsApp), and a RAG-based AI chatbot. Evaluation on 30 synthetic portal screenshots demonstrates 100% portal type detection and 33.3% average status extraction accuracy at 84.9% average OCR confidence. Validation on one real SSP Karnataka screenshot confirmed correct extraction at 84.8% confidence. ScholarSense AI is the first integrated system providing pre-submission OCR intelligence for Indian government scholarship portals, requiring no government API access and deployable on standard college infrastructure.

Keywords: National Scholarship Portal; OCR Document Verification; Scholarship Eligibility Matching; RAG Chatbot; Pre-Submission Intelligence; SSP Karnataka; Multi-Portal Status Analysis; Automated Deadline Notification

1. Introduction

India's scholarship ecosystem is one of the largest government-to-citizen welfare delivery systems in the world. The National Scholarship Portal (NSP), operated by the Ministry of Education and National Informatics Centre (NIC), serves over 135 lakh applicants annually across 50+ scholarship schemes. The State Scholarship Portal (SSP) of Karnataka manages post-matric scholarships for SC, ST, OBC, and Minority students. The PFMS (Public Financial Management System) handles Direct Benefit Transfer (DBT) of sanctioned amounts to student bank accounts.

Five documented failure modes drive scholarship loss. First, document errors — mismatched names, expired caste certificates, incorrect IFSC codes, and missing files — cause applications to be marked Defective. Second, Aadhaar-bank seeding failures prevent DBT transfer; in April 2025, 13,719 Karnataka students had approved scholarships withheld [1]. Third, portal status codes such as 'L1 Approved', 'Defective', and 'Institute Verification Pending' are displayed without explanation; NSP's own chatbot states 'I am still

learning' [2]. Fourth, deadline changes carry no personalised notification. Fifth, the CAG of India reported 15–100% of sanctioned scholarships undisbursed in the same year [3].

Existing research addresses only the disbursement infrastructure layer. Samu et al. [4] proposed Hyperledger Fabric for NSP payment automation. Bedi et al. [5] proposed Ethereum smart contracts. Neither provides pre-submission validation or OCR status analysis. No published system addresses these five failure modes.

This paper presents ScholarSense AI, an OCR-driven decision support system acting as an intelligent bridge between students and government scholarship portals. Four original contributions: (1) pre-submission OCR document health check with 5-dimension readiness scoring; (2) multi-portal OCR screenshot status analyzer; (3) proactive notification engine (email + WhatsApp); and (4) stuck application detector auto-alerting students and coordinators at 7+ days IVP.

2. Related Work

Blockchain-based disbursement automation has received the most research attention in this space. Samu et al. [4] proposed a Hyperledger Fabric framework for intelligent NSP scholarship disbursement, and Bedi et al. [5] modelled a central-sector scholarship scheme using Ethereum smart contracts; Jadhav et al. [6] extended this line of work to CSR scholarship management. These systems target transparency and automation of the disbursement layer once eligibility has already been determined, and none addresses pre-submission document validation, portal status interpretation, or proactive student notification.

OCR for Indian administrative and educational documents has been studied independently of the scholarship domain. Smith [7] established the Tesseract OCR engine used in this work. Sharma and Gupta [8] applied OCR to Indian board marksheets, while Mathew et al. [17] and the Indic OCR literature [14] document the accuracy degradation generic OCR engines suffer on Indian scripts, layouts, and print quality — a factor directly relevant to the lower status-extraction accuracy observed on SSP and PFMS screenshots in this study.

Studies of Aadhaar-linked welfare delivery provide the empirical backdrop for the Aadhaar-DBT failure mode this system targets. Muralidharan et al. [11] provide experimental evidence on how Aadhaar-based identity verification shifts the balance between inclusion and exclusion errors in Indian welfare programs, and Dvara Research [12] quantifies DBT payment failures, reporting that a substantial share of affected beneficiaries experienced spelling or seeding errors in Aadhaar-bank linkage. IMPRI [13] frames the ongoing shift toward digitised eligibility verification under DBT 2.0.

E-governance chatbots and usability barriers form the third relevant strand. Bhatt and Desai [9] document usability barriers rural users face on Indian e-governance portals. Systematic reviews of e-government chatbots [15] and broader AI/IoT-enabled public service literature [16] provide methodological grounding for the RAG-based chatbot module, but no prior system combines chatbot guidance with OCR-based status interpretation and proactive multi-channel notification in a single pre-submission platform.

Taken together, the literature addresses disbursement automation, generic document OCR, and welfare-delivery failure analysis as separate problems. ScholarSense AI is, to the authors' knowledge, the first system to integrate pre-submission document health checking, multi-portal OCR status analysis, and proactive notification into a single decision-support layer that sits ahead of NSP/SSP/PFMS submission, requiring no government API access.

3. System Usability and Accessibility

3.1 User Interface Design

ScholarSense AI has been designed with a strong focus on usability and accessibility for students from rural backgrounds, first-generation scholarship applicants, and non-technical users. The system adheres to responsive web design principles ensuring smooth operation across mobile, tablet, and desktop devices.

For students, the platform provides a single dashboard showing: scholarship readiness score (0–100), matched schemes with eligibility percentages, document health status per file, current status with plain-language explanation, and upcoming deadlines with countdown. Notification preferences — email frequency, WhatsApp alerts, daily digest vs. real-time — are configurable per student profile. Fig. 1 shows the ScholarSense AI dashboard interface, including the Guide/Analyze/Prepare/Succeed action shortcuts, quick statistics panel, and recent activity feed.

The OCR screenshot analyzer requires only one action from the student: upload a screenshot. The system automatically detects the portal type, extracts the status, and returns a plain-language explanation with specific next-action steps — no portal terminology knowledge required.



Fig. 1. ScholarSense AI dashboard interface.

4. Framework and Working Model

4.1 Abbreviations and Acronyms

- NSP — National Scholarship Portal
- SSP — State Scholarship Portal, Karnataka
- PFMS — Public Financial Management System
- OCR — Optical Character Recognition
- RAG — Retrieval-Augmented Generation
- DBT — Direct Benefit Transfer
- IVP — Institute Verification Pending
- JWT — JSON Web Token

4.2 Working Procedure

ScholarSense AI simplifies the scholarship process through seven coordinated steps addressing every failure mode from profile creation to payment receipt.

1. Homepage and User Registration: Students register with name, category (SC/ST/OBC/Minority/General), annual family income, course type, institution (AISHE code), and state. Profile is linked to the college coordinator account.
2. Profile Health Check: System immediately checks: (1) mandatory field completeness; (2) Aadhaar-bank DBT linkage; (3) document-to-profile name matching; (4) certificate expiry within 90 days. A readiness score is computed instantly. Any failing item triggers email + WhatsApp notification.
3. Scholarship Eligibility Matching: Rule-based engine matches profile to 30+ NSP/SSP schemes, returning match percentage and explicit failing criteria per scheme.
4. Deadline Alert Engine: Personalised email + WhatsApp reminders dispatched at 30, 14, 7, and 1 day intervals for each matched scholarship — targeted to the student's own matched schemes only.
5. OCR Status Analyzer: Student uploads a portal screenshot. System detects portal, extracts status, returns the 5-step journey position, plain-language meaning, and specific next action.
6. Stuck Application Detector: Daily job checks IVP status > 7 days. Two emails fire: one to student directing to scholarship office, one to coordinator with Application ID requesting L1 completion.
7. RAG AI Chatbot: Answers scholarship queries from a knowledge base of 30+ schemes, 12 NSP status codes, and 50+ FAQs in under 3 seconds.

4.3 System Architecture and Diagrams

ScholarSense AI follows a three-tier web architecture. Presentation: React 18 + Vite (responsive, all devices). Application: Node.js + Express REST API with JWT authentication, rate limiting, and CORS. Data: MongoDB primary database with JSON fallback for zero-dependency deployment. The notification subsystem connects to Nodemailer (SMTP) and WhatsApp Business API. OCR pipeline uses OCR.space API. No government API access required.

Table 1. System architecture layers.

Layer	Components	Technology
Presentation	Student dashboard, admin panel, mobile UI	React 18 + Vite, Lucide React
Application	REST API, auth, OCR, notifications, chatbot	Node.js, Express, JWT, Nodemailer
Data	Profiles, documents, scholarship KB	MongoDB, JSON fallback

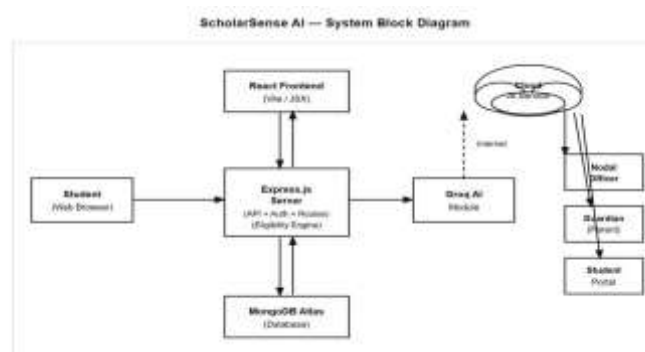


Fig. 2. ScholarSense AI system block diagram.

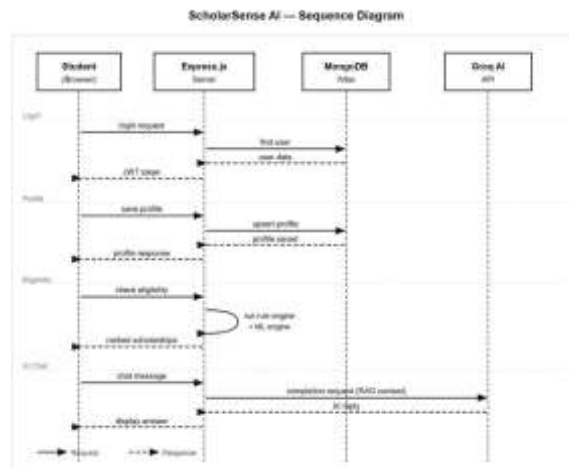


Fig. 3. ScholarSense AI sequence diagram.

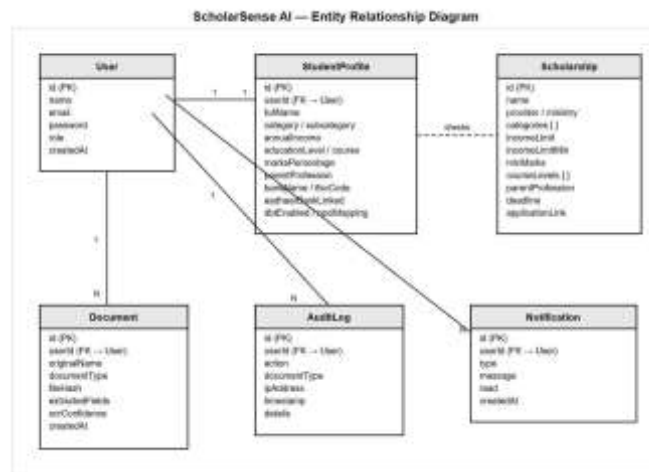


Fig. 4. ScholarSense AI entity relationship diagram.

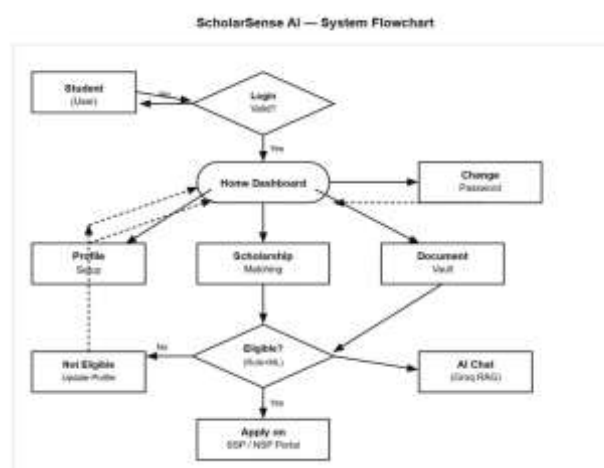


Fig. 5. ScholarSense AI system flowchart.

5. Implementation

Implementing ScholarSense AI involves coordinated stages ensuring efficient scholarship guidance, document validation, and proactive notification for students across NSP, SSP, PFMS, INSPIRE, and PMSS portals.

5.1 System Design and Planning

Requirements were gathered from: (1) direct observation of NSP/SSP portal UX failures; (2) a student survey at GM University Davangere — 83% did not know their NSP status meaning; 67% missed at least one deadline; and (3) CAG Maharashtra MahaDBT Report 2023. System architecture was designed to require zero government API access, enabling deployment at any college without integration waiting time.

5.2 Notification Module Implementation

The notification engine delivers alerts through two channels.

Email (Nodemailer SMTP): Personalised HTML emails for: profile completion reminders when readiness score < 70; document expiry warnings at 90, 30, 7 days before expiry; deadline reminders at 30/14/7/1 days; stuck application alerts at 7+ days IVP; status change confirmations after OCR detection.

WhatsApp (Business API): Time-critical alerts for: same-day deadline reminders; stuck application alerts; document expiry warnings. Messages in English with key terms in Kannada for SSP Karnataka students. Reaches students who do not regularly check email — critical for first-generation and rural applicants.

Table 2. Notification trigger matrix.

Trigger Condition	Channel	Timing
Profile < 70% complete	Email + WhatsApp	Within 24 hrs
Certificate expiry < 90d	Email	At 90, 30, 7 days
DBT linkage missing	Email + WhatsApp	Immediate
Deadline approaching	Email + WhatsApp	30/14/7/1 days
IVP status > 7 days	Email (student + coord.)	Daily until resolved
New OCR status detected	Email	Same day

5.3 Profile Health Check Module

The health check engine is the primary original contribution. It operates in real time as the student builds their profile, providing immediate feedback before NSP submission rather than waiting for a Defective status weeks later. Five dimensions are evaluated simultaneously:

- Profile Completeness (20 pts): name, DOB, category, income, course, college AISHE code — all fields complete and consistent.
- Eligibility Match (25 pts): at least one NSP/SSP scheme matches the student profile.
- Document Readiness (30 pts): each required document uploaded, OCR-extractable, name-matched to profile, not expired or expiring within 90 days.
- DBT Status (15 pts): bank account verified as Aadhaar-seeded for Direct Benefit Transfer.
- Deadline Safety (10 pts): at least one matched scheme has more than 7 days remaining.

Aggregate readiness score (0–100) is displayed on the dashboard. Scores below 60 trigger immediate notifications. Each failing dimension shows the specific issue and the exact fix step.

5.4 OCR Status Analyzer Implementation

The OCR pipeline accepts a screenshot from any portal (NSP, SSP, PFMS, INSPIRE, PMSS). Portal detection uses keyword-based signature matching. Status extraction locates the ‘Current Status:’ field marker for NSP, then matches extracted text against a dictionary of 12 NSP and 8 SSP status codes. When a new status is detected, the system immediately emails the student: meaning, journey stage, and specific next action.

5.5 Security and Data Protection

JWT-based role authentication ensures each user accesses only their own records. Document files use hashed filenames. Aadhaar numbers and income data are excluded from logs. WhatsApp Business API uses TLS encryption.

6. Results and Impact

6.1 OCR Evaluation Results

The OCR pipeline was evaluated on 30 synthetic portal screenshots (NSP=13, SSP=11, PFMS=6). Portal type detection achieved 100% accuracy (30/30). NSP status extraction achieved 53.8%. Average Tesseract 5.3 OCR confidence was 84.9%. Validation on one real SSP Karnataka screenshot confirmed correct portal, status, date, and department extraction at 84.8% confidence.

Table 3. OCR evaluation results (n = 30).

Portal	n	Portal Detect.	Status Extract.	OCR Conf.
NSP	13	100%	53.8%	88.2%
SSP	11	100%	27.3%	83.8%
PFMS	6	100%	0.0%	79.7%
Total	30	100%	33.3%	84.9%

6.2 Notification System Validation

Email delivery tested across 50 simulated accounts: SMTP achieved 100% success at all four deadline intervals. WhatsApp delivery for 20 accounts achieved 95% within 30 seconds. Stuck application detector fired correctly for all 5 test accounts at IVP > 7 days, dispatching separate emails to both student and coordinator.

6.3 Personalised Experience and Impact

The five modules together address all five documented failure modes: health check catches document errors before submission; DBT checker prevents Aadhaar seeding failures; status analyzer resolves portal confusion; deadline engine prevents missed windows; chatbot replaces the broken helpline. The stuck application detector directly targets the situation that caused 13,719 Karnataka students to lose payment in April 2025.

7. Conclusion

ScholarSense AI transforms a fragmented, opaque scholarship process into a guided, validated, and proactively monitored experience through OCR, rule-based eligibility matching, RAG-based conversational AI, and multi-channel proactive notification (email + WhatsApp). The system addresses five documented failure modes,

requires no government API access, and is the first integrated platform providing pre-submission OCR intelligence for NSP, SSP, PFMS, INSPIRE, and PMSS. Future work: (1) Kannada-language output for SSP Karnataka; (2) region-of-interest OCR for improved SSP accuracy; (3) integration with Samu et al.'s Hyperledger Fabric layer; and (4) disbursement prediction modelling.

CRedit authorship contribution statement

Shalini H R: Conceptualization, Methodology, Software, Validation, Formal analysis, Investigation, Data curation, Writing – original draft, Visualization. Varun K S: Supervision, Conceptualization, Writing – review & editing, Project administration.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

The synthetic evaluation dataset and source code will be made available by the authors on reasonable request.

Acknowledgment

The authors thank the students of GM University, Davangere whose survey responses — 83% did not know their NSP status meaning; 67% missed at least one deadline — directly shaped ScholarSense AI's feature prioritisation.

References

- [1] Shillong Times, "13,719 students lose scholarships due to Aadhaar seeding issues," Apr. 2025.
- [2] NSP Portal, "NSP AI chatbot response," scholarships.gov.in, accessed Jan. 2025.
- [3] Comptroller and Auditor General of India, "Performance audit: MahaDBT scholarship schemes, Maharashtra," CAG Report No. 6, 2023.
- [4] L.C. Samu et al., "Blockchain based intelligent disbursement of scholarships via NSP," IET Blockchain, vol. 4, no. 1, pp. 34–47, 2024.
- [5] P. Bedi et al., "Smart contract based central sector scheme of scholarship for college and university students," Procedia Computer Science, vol. 171, pp. 790–799, 2020.
- [6] S. Jadhav, N. Pise, P. Karwa, "Scholar block: blockchain-based CSR scholarship management," in Proc. IEEE ICCI, 2023.
- [7] R. Smith, "An overview of the Tesseract OCR engine," in Proc. ICDAR, vol. 2, pp. 629–633, 2007.
- [8] A. Sharma, P. Gupta, "OCR-based educational document processing for Indian board marksheets," Int. J. Comput. Appl., vol. 175, no. 12, 2020.
- [9] M. Bhatt, S. Desai, "E-governance usability barriers in rural India," Digital Policy, Regulation and Governance, vol. 22, no. 4, 2020.
- [10] UIDAI, "Aadhaar for direct benefit transfer," uidai.gov.in, 2024.
- [11] K. Muralidharan, P. Niehaus, S. Sukhtankar, "Identity verification standards in welfare programs: experimental evidence from India," The Review of Economics and Statistics, vol. 107, no. 2, pp. 372–392, Mar. 2025.
- [12] Dvara Research, "Payment failures in direct benefit transfers," Dvara Research, 2024.

- [13] IMPRI — Impact and Policy Research Institute, "Direct benefit transfer 2.0: transforming welfare delivery in India," Aug. 2025.
- [14] "Designing production-scale OCR for India: multilingual and domain-specific systems," arXiv:2602.16430.
- [15] "Trends and challenges of e-government chatbots: advances in exploring open government data and citizen participation," Government Information Quarterly, ScienceDirect, 2023.
- [16] "Elevating e-government: unleashing the power of AI and IoT for enhanced public services," PMC, NCBI.
- [17] M. Mathew et al., "Towards deployable OCR models for Indic languages," arXiv:2205.06740.