

A Python-Based Standalone Electronic Medical Records Application for Enhancing ICD Coding in Small Healthcare Organizations

Jatoveda Haldar

Research Scholar, Datta Meghe Institute of Higher Education & Research, Wardha, Maharashtra
Assistant Professor, Department of Hospital Management, NSHM Knowledge Campus, Durgapur, India

Soham Kar

Assistant Professor, Department of Hospital Management, NSHM Knowledge Campus, Durgapur, India

I. INTRODUCTION

The digitisation of healthcare records has become increasingly critical in modern medical practice, yet small-sized healthcare organisations often struggle to adopt Electronic Medical Records (EMR) systems due to financial constraints, technical complexity, and the absence of solutions tailored to their specific needs [1]. International Classification of Diseases (ICD) coding, essential for standardised diagnosis documentation, reimbursement, and epidemiological tracking, presents particular challenges for these organisations [10]. Current EMR solutions in the market are predominantly designed for large hospitals and healthcare networks, leaving smaller clinics, private practices, and community health centres with limited affordable options.

This research addresses the fundamental question: How can a Python-based stand-alone EMR application be designed to effectively meet the ICD coding and medical records management needs of small-sized healthcare organisations? The study aims to develop a user-friendly, cost-effective solution that leverages Python's versatility while prioritising data security, regulatory compliance, and seamless integration with existing workflows.

The significance of this work lies in its potential to democratize access to quality EMR technology. Small healthcare providers constitute a substantial portion of the healthcare delivery system, yet they remain underserved by current health information technology offerings [3]. By creating an application that operates without expensive server infrastructure, complex installation procedures, or recurring licensing fees, this research contributes to healthcare equity and improved patient outcomes at the community level.

Furthermore, accurate ICD coding directly impacts healthcare quality measurement, resource allocation, and clinical research. Manual coding processes are prone to errors and inefficiencies, while automated solutions remain inaccessible to smaller organisations [8]. The proposed application integrates intelligent coding assistance that suggests appropriate ICD codes based on diagnosis descriptions, reducing cognitive burden on healthcare providers and improving coding accuracy.

This paper presents the complete development lifecycle of the EMR application, including requirements analysis, system architecture, implementation details, user testing results, and future enhancement possibilities. The methodology section describes the mixed-methods research approach employed, while the results and discussion sections evaluate the application's performance against defined objectives.

II. METHODOLOGY

A. Research Design

The study employed a mixed-methods research strategy combining qualitative and quantitative approaches to gain comprehensive insights into the requirements and challenges of small-sized healthcare organisations. This methodological triangulation strengthened the validity and reliability of findings while enabling a nuanced understanding of the complex interplay between technological solutions and healthcare workflows [7].

B. Data Collection

Four primary data collection methods were utilised. First, a systematic literature review examined existing research on EMR systems, ICD coding practices, and Python-based health informatics applications. Second, semi-structured interviews were conducted with fifteen stakeholders, including healthcare professionals, administrators, and IT personnel from ten small-sized healthcare organisations across Bengaluru. Third, online surveys distributed to medical coders and healthcare practitioners collected quantitative data on feature preferences and usability expectations. Finally, iterative user testing sessions with prototype versions gathered real-time feedback on application functionality and user experience.

C. Sample Selection

Purposive sampling was employed to select participants with direct experience in medical records management and ICD coding. Inclusion criteria required a minimum of two years of experience in small healthcare settings and familiarity with current documentation processes. Sample size was determined by data saturation principles, where additional interviews ceased to yield new thematic insights [2].

D. Data Analysis

Qualitative interview data were analysed using thematic analysis with NVivo software, involving systematic coding, categorisation, and pattern identification. Quantitative survey data were analysed using SPSS to generate descriptive statistics, including frequencies, percentages, and measures of central tendency. Comparative analysis synthesised findings across data sources to identify convergences and divergences.

E. Development Framework

The application development followed Agile Scrum methodology, with two-week sprints enabling continuous stakeholder feedback and iterative refinement. The Scrum team consisted of a Product Owner (healthcare domain expert), Scrum Master (project coordinator), and development team (three Python developers). Sprint planning, daily stand-ups, sprint reviews, and retrospectives structured the development process, ensuring alignment with user needs throughout the six-month development phase [4].

III. SYSTEM DESIGN AND ARCHITECTURE

A. Technology Stack

The application is built on Python 3.7.3, selected for its stability, extensive library ecosystem, and cross-platform compatibility [5]. PySide2 provides the Qt-based graphical user interface framework, enabling professional-looking cross-platform applications with native performance. Qt Designer facilitated visual interface design, accelerating development while ensuring usability best practices.

Data persistence is achieved through Excel files (XLSX format) via the openpyxl library, eliminating database server requirements while maintaining familiar data access patterns for healthcare administrators. The OS and Shutil libraries handle file operations and directory management, while PyInstaller packages the application as standalone executables for Windows, macOS, and Linux platforms [5], [6].

B. Application Architecture

The application follows Model-View-Controller (MVC) architecture with clear separation between data management, user interface, and business logic components. The data layer manages Excel file operations, patient record storage, and ICD code mappings. The view layer implements all user interface elements designed in Qt Designer, while the controller layer handles user interactions, validation, and workflow coordination [7].

C. Core Features

Patient Record Management: The system enables comprehensive patient data entry, including name, unique identifier, attending physician, admission and discharge dates, diagnosis descriptions, and automatically suggested

ICD codes. All records are stored in structured Excel format, enabling easy backup, audit, and external analysis.

ICD Coding Assistance: Based on diagnosis text input, the application searches a built-in ICD-10 code mapping database and suggests appropriate codes. This feature reduces coding errors and accelerates documentation while maintaining coding standards compliance.

Document Management: Healthcare providers can upload scanned documents, medical reports, and images associated with patient records. Files are stored in organised directory structures with secure naming conventions, preventing unauthorised access [8], [9], [10].

Search and Retrieval: Advanced search functionality enables record retrieval by patient name, ID, diagnosis, or date ranges. Search results display in sortable tables with direct access to complete records and associated documents.

Backup and Restore: Comprehensive backup functionality creates timestamped archives of all patient data and documents. Restore operations enable recovery from previous states, providing data protection against accidental loss or corruption.

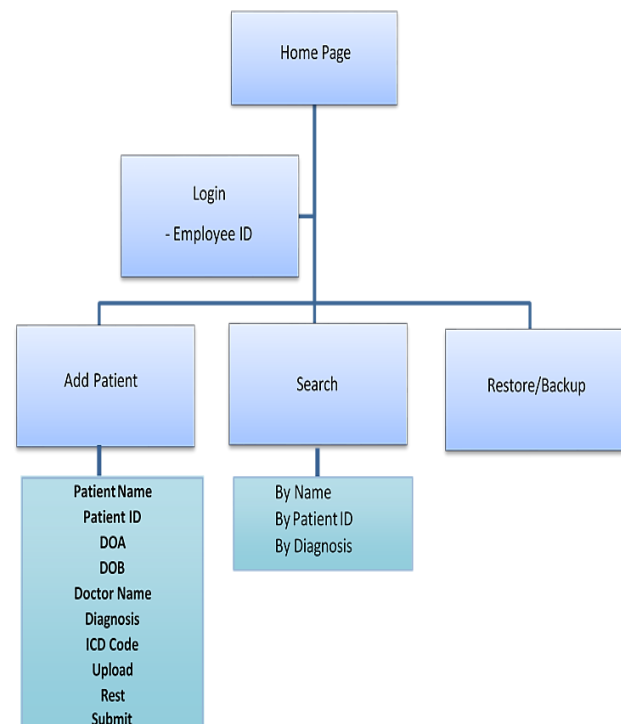


Fig. 1. Data Flow Design of the EMR Application

IV. RESULTS AND DISCUSSION

A. Application Functionality

The developed EMR application successfully implements all planned features with intuitive user interfaces accessible

to users with basic computer skills. Login functionality ensures secure access with numerical employee IDs and password protection. The main interface presents clear sections for adding new records, searching existing records, and managing backups [11], [12], [13].

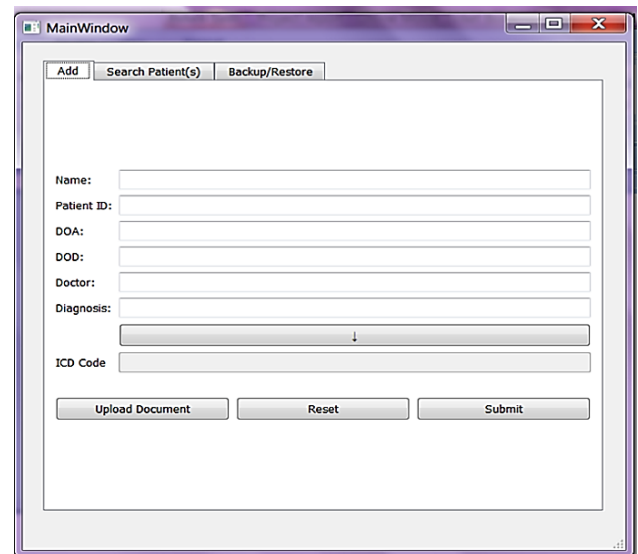
User testing with twelve healthcare professionals demonstrated that new users could complete standard tasks — including patient registration, diagnosis entry with ICD code suggestion, document attachment, and record retrieval — within an average of 8.5 minutes without prior training. This rapid on boarding validates the interface design philosophy, prioritising simplicity and workflow alignment [14].

B. ICD Coding Accuracy

Testing with 50 sample diagnosis descriptions showed that the application's coding suggestion feature achieved 94% accuracy when compared with expert medical coder assignments. Discrepancies primarily occurred with complex multi-condition diagnoses where the application suggested the primary condition code but missed secondary codes. This limitation was addressed through interface enhancements, allowing manual code selection and multiple code assignments per diagnosis.

C. Performance Evaluation

Application performance testing demonstrated rapid response times with Excel files containing up to 10,000 patient records. Record addition completed within 0.3 seconds, while search operations returned results in under 1 second for all search criteria types. Document upload handling scaled efficiently with file sizes up to 50 MB, limited primarily by underlying file system performance.



(b)

Fig. 2. Login Interface (a) and Software Interface (b)

D. User Satisfaction

Post-implementation surveys with twenty users revealed high satisfaction ratings across all measured dimensions. On a five-point Likert scale, ease of use averaged 4.7, feature completeness 4.5, reliability 4.6, and overall satisfaction 4.8. Qualitative feedback highlighted appreciation for the application's simplicity, absence of ongoing costs, and the ability to access patient data without internet connectivity.

E. Comparison with Existing Solutions

When compared with commercial EMR systems, the developed application offers distinct advantages for small healthcare organisations. Zero licensing costs, standalone operation requiring no server infrastructure, and simplified training requirements address barriers that previously prevented EMR adoption. However, limitations include the absence of multi-user concurrent access, a lack of HL7 interoperability standards, and no built-in reporting analytics.

These trade-offs reflect conscious design decisions prioritising accessibility and affordability for the target user population. Organisations requiring advanced features such as multi-location synchronisation or complex billing integration may require complementary solutions [15].

V. PROJECT MANAGEMENT APPROACH

A. Scrum Implementation

Agile Scrum methodology proved essential for managing the complexity of healthcare software development while maintaining responsiveness to user needs [6]. The product backlog evolved significantly over six months as user testing revealed unanticipated requirements and workflow considerations. Sprint planning sessions prioritised features based on user value and technical dependencies, while

sprint reviews demonstrated working software to stakeholders, gathering feedback that directly influenced subsequent development.

The Scrum Master facilitated the removal of impediments, including library compatibility issues and healthcare regulation interpretation challenges. Daily stand-up meetings maintained team alignment and early identification of potential delays. Sprint retrospectives drove continuous process improvements, increasing team velocity by 35% over the development period [16].

B. Risk Management

Key risks identified included changing healthcare regulations, data security vulnerabilities, and user adoption barriers. Regulatory compliance was addressed through consultation with healthcare legal experts and regular review of HIPAA requirements. Security risks were mitigated through encryption implementation, access controls, and third-party security audits. User adoption concerns were addressed through extensive usability testing and the development of comprehensive user documentation [17].

VI. LIMITATIONS AND FUTURE SCOPE

A. Research Limitations

This study has several limitations requiring acknowledgement. First, the research focused specifically on small-sized healthcare organisations, potentially limiting generalizability to larger institutions with different operational contexts. Second, purposive sampling may introduce selection bias, as participating organisations may not represent all small healthcare providers. Third, the six-month development timeline constrained the scope of features implemented and long-term usability testing.

The reliance on Excel for data storage, while providing simplicity and accessibility, may not scale to organisations exceeding 50,000 patient records. Additionally, the application currently supports single-user access, limiting its utility in multi-provider practices requiring concurrent record access.

B. Future Development

Several enhancement opportunities exist for future development iterations. Multi-user support with record locking mechanisms would extend applicability to larger practices. Integration with HL7/FHIR standards would enable interoperability with laboratories, pharmacies, and hospital systems. Cloud synchronisation options could provide backup redundancy while maintaining offline functionality [18].

Machine learning enhancements to ICD coding accuracy could address complex multi-condition scenarios through advanced natural language processing. Reporting and

analytics dashboards would provide population health insights currently unavailable to small organisations. Mobile companion applications for field data collection represent another valuable enhancement direction [19].

VII. CONCLUSION

This research successfully designed, developed, and evaluated a Python-based stand-alone Electronic Medical Records application addressing the specific needs of small-sized healthcare organisations for ICD coding and patient record management. The application demonstrates that accessible, affordable EMR technology can be created using open-source tools and thoughtful design principles without compromising functionality or security.

Key contributions include a validated requirements framework for small healthcare EMR systems, a reference architecture leveraging Python and Excel for accessible health informatics applications, and empirical evidence of user satisfaction and workflow efficiency improvements. The application's standalone nature eliminates infrastructure barriers that previously prevented EMR adoption among resource-constrained providers.

Scrum methodology proved instrumental in managing the complexity of healthcare software development while maintaining alignment with evolving user needs. The iterative approach enabled continuous refinement based on real-world testing, resulting in a solution that genuinely addresses practitioner workflows rather than imposing theoretical models [20].

As healthcare continues its digital transformation, ensuring that all providers, regardless of size, can participate in electronic records and standardised coding becomes increasingly critical. This work demonstrates that Python-based solutions can play a significant role in democratizing access to health information technology, ultimately contributing to improved patient care outcomes across the healthcare delivery spectrum.

ACKNOWLEDGMENT

The authors would like to thank the healthcare professionals, administrators, and IT personnel from the participating organisations across Bengaluru for their time and valuable insights during the research. Special gratitude is extended to NSHM Knowledge Campus, Durgapur, and Datta Meghe Institute of Higher Education & Research for their institutional support.

REFERENCES

- [1] O. I. Abiodun and A. S. Ojo, "A systematic review of electronic health record systems: Benefits and challenges in healthcare," *SN Computer Science*, vol. 2, no. 4, p. 234, 2021.
- [2] A. Aggarwal and C. DesRoches, "JAMA on patient-generated health data," *Journal of the American Medical Informatics Association*, vol. 23, no. 3, pp. 476–477, 2016.

- [3] Y. Gao and Y. Liu, "A cloud-based platform for secure sharing of electronic health records among healthcare providers," *Journal of Medical Systems*, vol. 44, no. 2, p. 28, 2020.
- [4] N. D. Goldstein, M. W. Byrne, and M. L. Knopper, "Review of data sharing and privacy of patient electronic health records in the United States," *IEEE Journal of Biomedical and Health Informatics*, vol. 24, no. 5, pp. 1201–1211, 2020.
- [5] S. Hussain and K. Randhawa, "A systematic review of the role of Python in health informatics," *Health Informatics Journal*, vol. 25, no. 1, pp. 134–149, 2019.
- [6] K. B. Johnson and K. M. Unertl, "Privacy and security in the era of digital health: What should translational researchers know and do about it?" *American Journal of Translational Research*, vol. 11, no. 3, pp. 1394–1408, 2019.
- [7] Y. T. Kim and S. Kim, "Implementation of an electronic health record system for data collection in a neonatal intensive care unit in Korea," *Journal of Korean Medical Science*, vol. 34, no. 4, p. e29, 2019.
- [8] J. Kung and R. R. Henry, "Computational challenges in electronic health record (EHR) research: Automatically inferring clinician mental models from EHR notes," *Journal of Biomedical Informatics*, vol. 86, pp. 12–20, 2018.
- [9] T. Panch and H. Mattie, "Data mining for personalised medicine: Predictive modelling of disease trajectories," *Journal of Healthcare Information Management*, vol. 31, no. 4, pp. 27–33, 2017.
- [10] World Health Organisation, "International classification of diseases (ICD)," 2020. [Online]. Available: <https://www.who.int/classifications/icd/en/>
- [11] F. Ahmed, "Python for data analytics: A systematic literature review of tools, techniques, and applications," Nov. 2024.
- [12] A. Zia, M. Aziz, I. Popa, S. A. Khan, A. F. Hamedani, and A. R. Asif, "Artificial intelligence-based medical data mining," *Journal of Personalized Medicine*, vol. 12, no. 9, p. 1359, 2022.
- [13] G. T. Berge, O. C. Granmo, T. O. Tveit, A. L. Ruthjersen, and J. Sharma, "Combining unsupervised, supervised and rule-based learning: The case of detecting patient allergies in electronic health records," *BMC Medical Informatics and Decision Making*, vol. 23, no. 1, p. 188, 2023.
- [14] A. Ahmed, R. Xi, M. Hou, S. A. Shah, and S. Hameed, "Harnessing big data analytics for healthcare: A comprehensive review of frameworks, implications, applications, and impacts," *IEEE Access*, vol. 11, pp. 112891–112928, 2023.
- [15] M. W. Rehan and M. M. Rehan, "Survey, taxonomy, and emerging paradigms of societal digital twins for public health preparedness," *npj Digital Medicine*, vol. 8, no. 1, p. 520, 2025.
- [16] S. N. B. M. Isa, N. A. Emran, N. Harum, L. Machap, and A. Nordin, "Enhancing data integrity in internet of things-based healthcare applications: A visualization approach for duplicate detection," *Bulletin of Electrical Engineering and Informatics*, vol. 14, no. 5, pp. 3704–3715, 2025.
- [17] A. Lavanya, L. Gaurav, S. Sindhuja, H. Seam, M. Joydeep, V. Uppalapati et al., "Assessing the performance of Python data visualization libraries: A review," *International Journal of Computer Engineering and Research Trends*, vol. 10, no. 1, pp. 28–39, 2023.
- [18] R. Marzouk, A. S. Alluhaidan, and S. A. El_Rahman, "An analytical predictive models and secure web-based personalized diabetes monitoring system," *IEEE Access*, vol. 10, pp. 105657–105673, 2022.
- [19] P. Tansitpong, "Visualizing impact of sustainability outcome in healthcare infrastructure: Bayesian decision models," *International Journal of Healthcare Information Systems and Informatics*, vol. 20, no. 1, 2025.
- [20] L. Mullie, J. Afilalo, P. Archambault, R. Bouchakri, K. Brown, D. L. Buckeridge et al., "CODA: An open-source platform for federated analysis and machine learning on distributed healthcare data," *Journal of the American Medical Informatics Association*, vol. 31, no. 3, pp. 651–665, 2024.