IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 6, Issue 6, November 2026
Impact Factor: 8.2

CURA-ID: An Electronic Health Record System using Random Forest Algorithm

Prof. Joshi G. R.[1], Ugle Priya Rajabhau[2], More Priyanka Popat[3], Madake Gayatri Shivaji[4], Nishi Satish Warale[5]
[1]Project Guide, Department of Computer Engineering
[2-5]Students, Department of Computer Engineering
HSBPVT's GOI Faculty of Engineering, Kashti, India

Abstract : *This paper presents "CURA-ID," an intelligent, role-secured Electronic Health Record (EHR) web platform integrated with an ensemble Random Forest clinical risk prediction engine. Traditional clinics and primary health centers predominantly rely on fragmented paper prescriptions and passive local filing utilities, which delay emergency history retrieval and overlook early indicators of chronic conditions such as cardiovascular disease and type 2 diabetes. CURA-ID bridges this gap by issuing every registered patient a collision-free unique identifier (CURA-ID) paired with a scannable QR card for instantaneous longitudinal timeline lookup. Architected on a decoupled three-tier full-stack framework, the platform comprises a Next.js (App Router) and React.js frontend with role-segregated portals (Administrator, Doctor, and Patient), Next.js Serverless API Routes secured by JWT authentication and Role-Based Access Control (RBAC), and a normalized PostgreSQL database protected by Row-Level Security (RLS). Upon logging clinical vitals and routine blood biomarkers, an embedded ONNX Runtime engine executes a pre-trained Scikit-Learn Random Forest classifier in sub-50 milliseconds to deliver calibrated disease risk probabilities and leading biomarker drivers directly within the consultation workflow.*

Keywords: Electronic Health Record (EHR), Random Forest Algorithm, Unique Patient Identifier (CURA-ID), Clinical Decision Support, Next.js, PostgreSQL Row-Level Security, ONNX Runtime, Chronic Disease Risk Prediction.

I. INTRODUCTION

In modern outpatient clinics and primary healthcare centers, physicians must rapidly evaluate patient demographics, drug allergies, blood pressure trajectories, glycemic trends, and prior prescriptions before making clinical decisions. However, across small and mid-sized healthcare facilities, patient records remain scattered across fragile paper slips, unindexed laboratory printouts, and non-interoperable local terminals, leading to consultation delays and redundant diagnostic testing.

While existing hospital management suites provide basic digital record filing, they operate strictly as passive storage repositories without real-time predictive intelligence. Conversely, academic machine learning classifiers achieve strong accuracy on clinical datasets [5], [6], [14] but remain isolated inside offline research notebooks detached from authentication layers, relational patient databases, and live consultation interfaces.

To address this critical translational gap, we propose "CURA-ID." By uniting a portable QR-indexed Electronic Health Record platform with an embedded Random Forest risk stratification engine, CURA-ID enables clinicians to retrieve complete longitudinal patient histories in milliseconds and automatically evaluate multi-parameter clinical vitals during active visits. This paper details the three-tier software architecture, ensemble mathematical formulation, algorithmic workflow, and comparative evaluation strategy for the platform.

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 6, Issue 6, November 2026
Impact Factor: 8.2

II. LITERATURE SURVEY

The conceptual and algorithmic foundation of CURA-ID is supported by 15 contemporary studies (2024–2026) spanning Electronic Health Record modeling and tree-ensemble clinical risk prediction [1]–[15].

Recent healthcare informatics literature confirms that structured longitudinal EHR variables—demographics, vitals, diagnostic codes, and blood biomarkers—encode reliable early indicators of clinical deterioration [2], [8], [12]. Furthermore, synthetic EHR validation studies by Kavitha Bai et al. [4] and Delos Reyes et al. [7] demonstrate that Random Forest classifiers preserve high predictive stability and transparent biomarker importance across cardiovascular risk cohorts.

Simultaneously, comparative clinical machine learning benchmarks validate the superiority of tree-based ensembles on tabular medical records. Mulakala et al. [14] reported Random Forest achieving 91% accuracy on chronic disease screening, outperforming competing baseline models. Studies by Priyadharshini and Vijayan [5], Raman et al. [6], Afridi et al. [3], Salim and Ibrahim [9], and Tasmurzayev et al. [11] further demonstrate that Random Forest pipelines resist overfitting, capture nonlinear biomarker interactions, and align with established cardiology and diabetes risk factors [1], [10], [13], [15].

However, a critical engineering gap remains: while offline Random Forest accuracy is well established and longitudinal EHR modeling is proven, existing prototypes lack unique QR-based patient indexing, role-secured relational database persistence, and sub-50ms web inference integrated into clinical portals. CURA-ID directly targets this intersection.

III. PROPOSED ARCHITECTURE

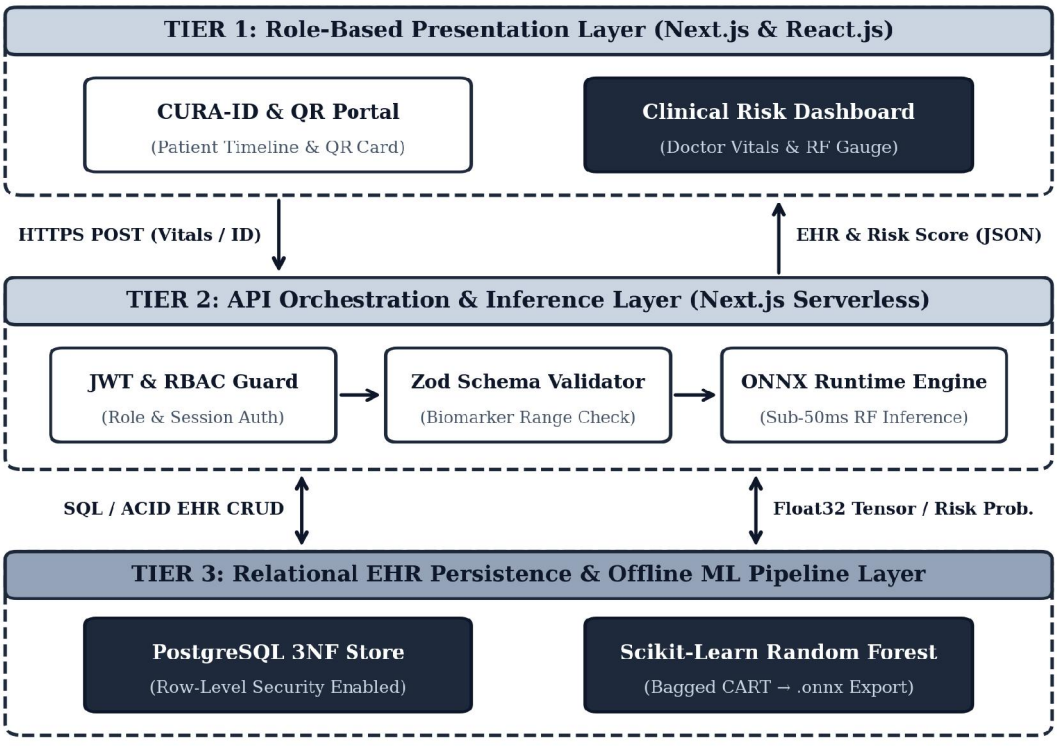


Fig. 1 Decoupled 3-Tier Web and Machine Learning Architecture of CURA-ID

To ensure strict patient confidentiality, low-latency inference, and modular scalability, CURA-ID abandons monolithic desktop designs in favor of a decoupled three-tier web and machine learning architecture.

Copyright to IJARSCT
www.ijarsct.co.in
DOI: 10.48175/IJARSCT-41119

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429 Volume 6, Issue 6, November 2026 Impact Factor: 8.2

A. Role-Based Presentation Layer The client-side interface is engineered using Next.js (App Router), React.js, and Tailwind CSS, providing role-segregated portals for Administrators, Doctors, and Patients. Each registered patient receives a collision-free CURA-ID and printable QR card, enabling clinicians to retrieve complete longitudinal visit timelines, pinned allergy alerts, lab PDFs, and color-coded risk gauges instantaneously.

B. Serverless API and ONNX Inference Layer The middleware utilizes Next.js Serverless API Routes acting as a unified security and inference gateway. Every request undergoes JSON Web Token (JWT) verification, Role-Based Access Control (RBAC) enforcement, and Zod physiological range validation. Instead of calling an external Python microservice prone to cold-start delays, the API executes the exported Random Forest model natively via ONNX Runtime in under 50 milliseconds.

C. Relational Persistence and Offline ML Pipeline Layer Patient demographics, clinical encounters, prescriptions, risk scores, and diagnostic attachments are persisted in a Third Normal Form (3NF) PostgreSQL database protected by native Row-Level Security (RLS) policies. In parallel, an offline Python and Scikit-Learn pipeline handles dataset preprocessing, stratified cross-validation, hyperparameter tuning, and ONNX serialization via skl2onnx.

IV. MATHEMATICAL MODEL AND ALGORITHM

The core operation of CURA-ID is formalized as a 6-tuple $S = \{I, O, F, C, S_{\text{succ}}, S_{\text{fail}}\}$ uniting deterministic relational EHR state transitions with supervised ensemble classification.

A. Ensemble Random Forest Mapping Let $I = \{(\text{id}, \mathbf{x})\}$ represent a consultation input, where id is the unique patient CURA-ID and $\mathbf{x} = [x_1, x_2, \ldots, x_d]^T$ denotes the d-dimensional clinical biomarker vector (age, systolic blood pressure, fasting glucose, serum cholesterol, BMI, and clinical indicators). Over N labeled training samples, the Random Forest constructs B decorrelated decision trees $\{h_1, h_2, \ldots, h_B\}$ on bootstrapped subsets using $m = \lfloor \sqrt{d} \rfloor$ candidate features per node by minimizing Gini Impurity: $G(D) = 1 - \sum_{k=1}^{K} p_k^2$. During point-of-care inference, the ensemble aggregates tree votes via majority rule: $\hat{y} = \text{mode}\{h_b(\mathbf{x})\}_{b=1}^{B}$.

B. Clinical Risk Probability and Screening Sensitivity Beyond binary classification, CURA-ID computes a calibrated disease risk probability $P(y = 1 \mid \mathbf{x})$ across all B decision trees to stratify patients into Low (< 35%), Moderate (35%–69%), and High (≥ 70%) clinical risk tiers:

$$P(y = 1 \mid \mathbf{x}) = \frac{1}{B} \sum_{b=1}^{B} \mathbb{I}(h_b(\mathbf{x}) = 1)$$

$$\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}}, \quad \text{F1} = \frac{2 \times \text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}}$$

Prioritizing Recall alongside ROC-AUC minimizes false-negative omissions during preventive outpatient screening while keeping online inference complexity strictly polynomial at O(B · D_max), where D_max denotes the maximum tree depth.

C. System Algorithm

The clinical encounter pipeline enforces authentication, biomarker range validation, ensemble inference, and ACID database persistence within a single atomic workflow.

Algorithm 1: Secure EHR Logging and Random Forest Risk Prediction

1. **Input:** Authenticated Doctor JWT T_jwt, Patient CURA-ID, and clinical vitals vector x.
2. **Authorization:** Verify JWT signature and RBAC role (DOCTOR/ADMIN); otherwise return HTTP 401/403.
3. **Validation:** Validate physiological bounds via Zod schema; if any vital is out of range, return HTTP 422.
4. **Preprocessing:** Encode categorical indicators and assemble normalized float32 feature tensor v.
5. **Ensemble Inference:** Execute cached ONNX Random Forest session (B = 100 trees, D_max = 12).
6. **Risk Stratification:** Compute calibrated probability P_risk and map to Low (< 0.35), Moderate (0.35–0.69), or High (≥ 0.70) risk tier.
7. **Persistence**: Commit encounter record and risk audit log atomically inside a PostgreSQL transaction.
8. **Output:** Return HTTP 201 Created with the calibrated risk score and top contributing biomarker drivers.

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 6, Issue 6, November 2026
Impact Factor: 8.2

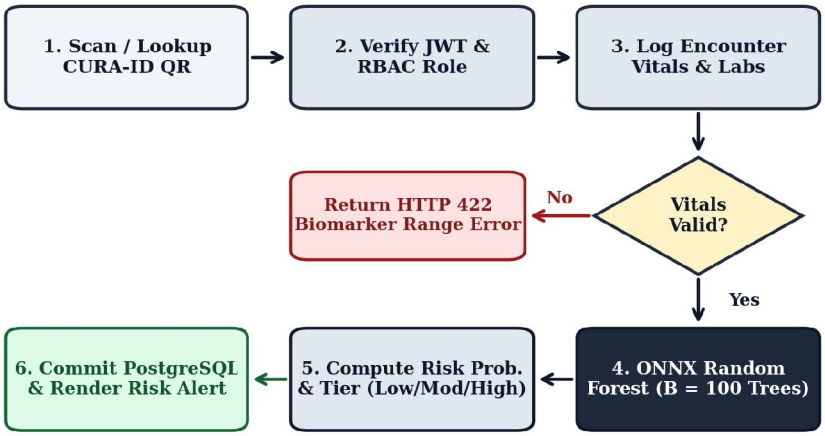


Fig. 2 UML Activity Diagram illustrating point-of-care EHR validation and ONNX Random Forest inference.

V. SYSTEM REQUIREMENTS AND COMPARATIVE ANALYSIS

CURA-ID is built upon a lightweight, purely software-based stack (Next.js v14+, React.js v18, Tailwind CSS, Node.js v20 LTS, PostgreSQL with RLS, Python Scikit-Learn, and ONNX Runtime) operable on any standard clinic desktop, laptop, or tablet without specialized medical sensors. Table I contrasts CURA-ID against existing clinical record and machine learning approaches.

TABLE I
COMPARATIVE ANALYSIS: EXISTING HEALTHCARE SYSTEMS VS. CURA-ID

Evaluation Parameter	Paper-Based Charts	Legacy Clinic ERP	Standalone ML Scripts	CURA-ID (Proposed)
Patient Retrieval	Manual folder search	Local terminal ID	None (static CSV files)	Instant QR / CURA-ID Lookup
Predictive Analytics	None	Passive filing only	Offline batch execution	Real-Time ONNX Random Forest
Access Security	Physical cabinet lock	Basic login password	None	JWT + RBAC + PostgreSQL RLS
Platform Accessibility	On-site clinic only	Intranet desktop app	Developer IDE only	Responsive Multi-Role Web App
Longitudinal View	Fragmented slips	Siloed visit tables	Static snapshot	Unified Health Timeline + Alerts

VI. ADVANTAGES AND LIMITATIONS

1. **Instant QR & CURA-ID Lookup:** Eliminates scattered paper prescriptions by indexing every patient's longitudinal history under a collision-free QR identifier.
2. **Proactive Clinical Decision Support**: Embeds calibrated Random Forest risk probabilities directly into routine outpatient vitals logging.
3. **Defense-in-Depth Data Privacy**: Combines JWT session verification, RBAC role guards, and native PostgreSQL Row-Level Security to prevent unauthorized record exposure.
4. **Sub-50ms Lightweight Inference:** Executes ONNX Runtime predictions natively inside Node.js without external Python cold-start latency or GPU hardware.

Copyright to IJARSCT
www.ijarsct.co.in
DOI: 10.48175/IJARSCT-41119

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429 Volume 6, Issue 6, November 2026 Impact Factor: 8.2

Limitations and Future Scope: Currently, clinical vitals rely on structured manual entry or CSV batch uploads, and class imbalance in clinical datasets requires careful threshold calibration during training to minimize false negatives.

Future Expansion Vectors:

1. **HL7 FHIR & ABHA Interoperability**: Extending API endpoints to exchange standardized health summaries with national digital health registries.
2. **Wearable Medical IoT Telemetry**: Integrating automated ingestion pipelines for continuous blood pressure and blood glucose monitoring devices.
3. **OCR Digitization & SHAP Explainability:** Adding optical character recognition for handwritten prescriptions alongside per-patient SHAP feature attribution charts.

VII. CONCLUSION

The CURA-ID platform presents a scalable, secure, and intelligent solution to patient record fragmentation and passive clinical data storage. By uniting collision-free QR and UUID patient indexing, three-tier PostgreSQL Row-Level Security, and sub-50ms ONNX Random Forest inference inside a responsive Next.js and React.js web application, the system transforms routine EHR documentation into proactive preventive healthcare support. The current development phase has established the complete architectural design, ensemble mathematical formulation, UML behavioral workflows, and normalized relational database schemas, providing a rigorous foundation for empirical classifier benchmarking and production deployment.

VIII. ACKNOWLEDGMENT

We would like to express our sincere gratitude to our internal guide, Prof. Joshi G. R., for providing invaluable technical direction and continuous mentorship. We also extend our thanks to Dr. Sayyad J. I., Head of the Computer Engineering Department, and Dr. Dhagate M. D., Principal of HSBPVT'S GOI Faculty of Engineering, Kashti, for providing the necessary laboratory infrastructure and institutional encouragement to pursue this research.

REFERENCES

1. P. Rezaei et al., "Enhancing Type 2 Diabetes Diagnosis with Evolutionary Algorithms and Machine Learning," arXiv preprint arXiv:2610.34123, 2026. URL: https://arxiv.org/pdf/2610.34123
2. R. Goncalves et al., "EHRAdapt: Adapting Pretrained Language Models to Electronic Health Records with Semantic Priors for Rare Clinical Events," arXiv preprint arXiv:2609.34007, 2026. URL: https://arxiv.org/pdf/2609.34007
3. A. Afridi, M. Usman, D. J. Rubin, and M. R. R. Rana, "A Robust and Explainable Framework for Predicting Hospital Readmission in Patients with Diabetes," Diabetology, vol. 7, no. 10, Art. 197, 2026. URL: https://www.mdpi.com/2673-4540/7/10/197/pdf
4. S. Kavitha Bai et al., "Realistic Synthetic Electronic Health Record Data Generation for Cardiovascular Risk Prediction," Journal of Vascular Diseases, vol. 5, no. 4, Art. 29, 2026. URL: https://www.mdpi.com/2813-2475/5/4/29/pdf
5. U. Priyadharshini and R. Vijayan, "Structured-to-text ClinicalBERT Embeddings with Random Forest for Heart Disease Prediction: A Proof-of-Concept Study on the UCI Statlog Dataset," Frontiers in Artificial Intelligence, vol. 9, Art. 1844707, 2026. URL: https://www.frontiersin.org/journals/artificial-intelligence/articles/10.3389/frai.2026.1844707/pdf
6. S. Raman, D. Thakkar, J. Calixte, and R. Kumar, "Machine Learning for Coronary Heart Disease Prediction: Comparative Analysis of Framingham and Cleveland Subset of the UCI Dataset with SHAP-Based Interpretability," Epidemiologia, vol. 7, no. 3, Art. 75, 2026. URL: https://www.mdpi.com/2673-3986/7/3/75/pdf
7. R. Delos Reyes, D. Capurro, and N. Geard, "Generating Synthetic Electronic Health Record Data Using

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 6, Issue 6, November 2026
Impact Factor: 8.2

Agent-Based Models to Evaluate Machine Learning Robustness Under Mass Casualty Incidents," arXiv preprint arXiv:2605.09951, 2026. URL: https://arxiv.org/pdf/2605.09951

8. J. Zhou, E. Wang, and X. Wang, "Semi-supervised Method for Risk Prediction with Doubly Censored EHR Data," arXiv preprint arXiv:2605.08046, 2026. URL: https://arxiv.org/pdf/2605.08046
9. S. S. Salim and A. A. Ibrahim, "A Machine Learning Approach for Predicting 30-Day Hospital Readmission in Patients with Diabetes," Healthcare, vol. 14, no. 9, Art. 1185, 2026. URL: https://www.mdpi.com/2227-9032/14/9/1185/pdf
10. E. Chowdhury et al., "Risk Prediction of Cardiovascular Disease for Diabetic Patients with Machine Learning and Deep Learning Techniques," arXiv preprint arXiv:2511.04971, 2025. URL: https://arxiv.org/pdf/2511.04971
11. N. Tasmurzayev, B. Imanbek, A. Boltaboyeva, and G. Dikhanbayeva, "Explainable AI for Coronary Artery Disease Stratification Using Routine Clinical Data," Algorithms, vol. 18, no. 11, Art. 693, 2025. URL: https://www.mdpi.com/1999-4893/18/11/693/pdf
12. P. D. Sood, S. Liu, H. Lehmann, and H. Kharrazi, "Assessing the Impact of Computable Type 2 Diabetes Phenotypes on Predicting Healthcare Utilization Using Electronic Health Records and Administrative Claims," Healthcare, vol. 13, no. 18, Art. 2292, 2025. URL: https://www.mdpi.com/2227-9032/13/18/2292/pdf
13. O. B. Ayoade, S. Shahrestani, and C. Ruan, "Machine Learning and Deep Learning Approaches for Predicting Diabetes Progression: A Comparative Analysis," Electronics, vol. 14, no. 13, Art. 2583, 2025. URL: https://www.mdpi.com/2079-9292/14/13/2583/pdf
14. S. V. Mulakala, G. Neeharika, P. V. Kumar, and A. B. Kiran, "Chronic Diseases Prediction Using ML," arXiv preprint arXiv:2502.10481, 2025. URL: https://arxiv.org/pdf/2502.10481
15. R. Jose, F. Syed, A. Thomas, and M. Toma, "Cardiovascular Health Management in Diabetic Patients with Machine-Learning-Driven Predictions and Interventions," Applied Sciences, vol. 14, no. 5, Art. 2132, 2024. URL: https://www.mdpi.com/2076-3417/14/5/2132/pdf