Peertechz

Health Informatics and Information Management

OPEN ACCESS JOURNAL

Journal DOI: https://dx.doi.org/10.17352/hiim

Research Article

Proposed EthMed Platform: An AI-Enabled, Blockchain-Governed and Digital-Twin-Ready Health Platform for Ethnic Minority Communities in Remote and Hard-To-Reach Areas

Le Ngoc Hung[1]*, Vo Quy Quang[1], Nguyen Thanh Tuan[1], Le Ngoc Hieu[1]* and Do Ngoc Thuy[2]

[1]VIPTAM Institute of Technology Application, TT3.9, Nam Cuong New Urban Area, Nghia Do, Ha Noi, Vietnam

[2]High Technology Innovation Center, Vietnam Academy of Science and Technology, 18, Hoang Quoc Viet, Nghia Do, Hanoi, Vietnam

Submitted : 11 July, 2026
Accepted : 12 August, 2026
Published : 13 August, 2026

***Corresponding authors:** Le Ngoc Hung, VIPTAM Institute of Technology Application, TT3.9, Nam Cuong New Urban Area, Nghia Do, Ha Noi, Vietnam, E-mail: ngoc10hung@yahoo.com

Le Ngoc Hieu, VIPTAM Institute of Technology Application, TT3.9, Nam Cuong New Urban Area, Nghia Do, Ha Noi, Vietnam, E-mail: dothuy1992831@gmail.com

Keywords: Digital health; Ethnic minority health; Artificial intelligence; Digital health twin; Vietnam; Design science; Health equity; Blockchain governance

https://www.healthdisgroup.com/hiim

Check for updates

Abstract

Objectives: Access to quality healthcare remains limited for ethnic minority populations in remote and mountainous areas because of geographical barriers, shortages of healthcare resources, fragmented medical information, limited connectivity, and linguistic or cultural barriers. The objective of this study was to design and justify EthMed, an integrated digital health platform that combines community screening, interoperable electronic health records (EHRs), telemedicine, artificial intelligence (AI), permissioned blockchain governance and a staged digital-health-twin pathway for hard-to-reach populations.

Methods: A design-science research approach was used to transform requirements from the target setting into a modular platform artefact and an evaluation framework. The methodology comprised requirement elicitation, requirement-to-design traceability, architecture design, technology selection, prototype demonstration, technical and clinical-workflow evaluation, and staged validation. The proposed pilot includes at least 3,000 participants; this number is treated as an initial development and validation cohort rather than as a universal sample size for every disease model. AI development uses prespecified targets, leakage-controlled data partitioning, internal validation, calibration, subgroup fairness assessment, and external/prospective validation before clinical use. The blockchain layer is limited to consent, provenance and audit functions, while clinical data remain off-chain. The digital-twin roadmap begins with risk twins and requires longitudinal state updating, uncertainty estimation and validation before physiology-twin functions are introduced.

Results: The resulting six-layer architecture integrates community access, interoperability, secure storage, AI analytics, telemedicine/referral coordination and governance. Priority pathways include hypertension, diabetes, cardiovascular and stroke risk, chronic respiratory and kidney disease, malnutrition and anaemia, maternal-child health, and selected musculoskeletal and prolonged-fatigue pathways. The design establishes explicit technical, clinical, user, fairness and governance evaluation criteria and a traceable pathway from community requirements to platform functions.

Conclusion: EthMed provides a research and implementation framework for equitable digital health in remote ethnic minority communities. Its principal contribution is not a single algorithm or technology, but a governed ecosystem linking longitudinal community data, AI-assisted risk stratification, telemedicine, secure interoperability and a progressively validated digital-twin architecture. The platform remains a conceptual artefact pending prospective clinical, ethical and implementation validation.

Citation: Hung LN, Quang VQ, Tuan NT, Hieu LN, Thuy DN. Proposed EthMed Platform: An AI-Enabled, Blockchain-Governed and Digital-Twin-Ready Health Platform for Ethnic Minority Communities in Remote and Hard-To-Reach Areas. Health Inform Inf Manag. 2026;2(1):1-8.
Available from: https://dx.doi.org/10.17352/hiim.000002

Abbreviations

AI: Artificial I; HER: Electronic Health Record; FHIR: Fast Healthcare Interoperability Resources; DICOM: Digital Imaging and Communications in Medicine; XAI: Explainable Artificial Intelligence; MFA: Multi-Factor Authentication; TRIPOD+AI: Transparent Reporting of a Multivariable Prediction Model for Individual Prognosis Or Diagnosis-Artificial Intelligence.

Introduction

Digital health has moved from isolated hospital information systems toward integrated ecosystems combining EHRs, telemedicine, AI-assisted decision support, mobile applications, cloud data infrastructure and remote monitoring. Global platforms such as Epic/MyChart, Teladoc Health, Babylon Health, Ada Health, Human API and Patientory illustrate different development paths: some prioritise EHR access, others symptom assessment and telemedicine, while some experiment with blockchain-based personal health-record control [1]. Machine-learning and deep-learning approaches can support diagnosis, triage, prediction, workflow optimisation and personalised care when they are evaluated carefully and embedded within safe clinical governance [2]. The equity problem is especially important in remote ethnic-minority communities. Telemedicine can reduce geographic barriers by connecting community health posts, district facilities, provincial hospitals and specialists, but persistent gaps remain for populations with limited connectivity, low digital literacy, language barriers and incomplete clinical histories. For these communities, telemedicine should therefore operate as a hybrid workflow: local health workers collect and contextualise information, AI supports structured risk assessment, and remote specialists provide advice when predefined thresholds or clinical complexity justify referral.

Many AI-enabled health platforms are developed from datasets collected in urban hospitals, insured populations or technologically mature environments. Their assumptions may not hold in remote communities where laboratory results are scattered, connectivity is intermittent, language affects communication and traditional health practices coexist with biomedical care. Disease risk is shaped not only by biological variables but also by ethnicity, climate, altitude, occupation, housing, nutrition, access to clean water, health-seeking behaviour and distance to high-quality care.

Blockchain has been proposed for health records, consent management, audit trails, provenance tracking and secure data sharing. However, blockchain should not be treated as a substitute for conventional security architecture. EthMed therefore stores clinical information in secure databases and data lakes while using a permissioned ledger for hashes, consent transactions, identity references, timestamps and audit events. The rationale for this selective use is strongest where several independent institutions need a shared, tamper-evident governance record without giving one institution unilateral control of the audit history. Where a single trusted data controller is sufficient, a conventional append-only or WORM audit system remains a simpler alternative [3].

Digital twins in medicine refer to computational representations that are updated with data from the represented individual or system and can support prediction, simulation and personalised decision-making. Because evidence, data integration and validation remain early-stage, EthMed does not equate a longitudinal risk score with a digital twin. Instead, the proposed twin must have an identifiable state representation, repeated data assimilation, an explicit update mechanism, uncertainty estimates, validation against observed trajectories and a defined clinical use case [4]. Multimodal AI can subsequently support integration of clinical text, laboratory results, imaging, physiological signals, wearable data, lifestyle and environmental information [5].

Lao Cai province in northern Vietnam provides a relevant setting because mountainous, border and hard-to-reach communes coexist with a substantial ethnic-minority population and a dual burden of communicable, maternal-child, nutritional and environmental conditions alongside increasing non-communicable diseases. Fragmented data across hospital information systems, laboratory information systems, radiology/PACS systems, electronic medical records and provincial health records limit predictive and preventive pathways.

The objective of this study was to develop a design-science framework for EthMed and specify how it can be implemented and validated without overstating clinical performance. Four questions guided the work: (1) what technical and community requirements should determine the architecture; (2) how can AI, interoperability, telemedicine and governance be combined into a reproducible artefact; (3) what sample-size and validation strategy is appropriate for multiple prediction targets; and (4) what minimum defining characteristics are required before the term digital twin is used. This paper clarified that EthMed is a conceptual and implementation framework, not a validated clinical decision system. First, it provides a requirement-to-design traceability matrix linking community needs to platform components. Second, it separates the pilot dataset size from model-specific sample-size requirements. Third, it defines an AI development and validation framework aligned with TRIPOD+AI principles. Fourth, it justifies permissioned blockchain as a governance option rather than a generic security solution and specifies identity, key, consensus, smart-contract and consent controls. Fifth, it defines a staged digital-twin methodology with state representation, longitudinal updating, uncertainty handling and clinical validation. These additions.

Materials and methods

Design-science research study

This study uses design-science research because its primary output is an artefact, the EthMed platform architecture and associated governance and evaluation model, rather than an estimate of treatment effectiveness. The design process follows a staged cycle adapted to the EthMed context: (1) problem identification and requirement elicitation; (2) objective definition and design requirements; (3) artefact design and technology selection; (4) prototype demonstration

Citation: Hung LN, Quang VQ, Tuan NT, Hieu LN, Thuy DN. Proposed EthMed Platform: An AI-Enabled, Blockchain-Governed and Digital-Twin-Ready Health Platform for Ethnic Minority Communities in Remote and Hard-To-Reach Areas. Health Inform Inf Manag. 2026;2(1):1-8.
Available from: https://dx.doi.org/10.17352/hiim.000002

in representative workflows; (5) evaluation against technical, clinical, user and governance criteria; and (6) iterative refinement and validation. This structure is consistent with established design-science principles [6,7]. Requirements were derived from the target setting and the four linked project documents underlying the original concept: an AI health-platform proposal for ethnic-minority communities in Lao Cai; a blockchain governance proposal for digital health data and clinical research; a digital-health-twin and multimodal-AI proposal; and a personal-platform concept note. Requirements were organised into six domains: access and usability; clinical data interoperability; AI-assisted risk assessment; referral and telemedicine; security and governance; and future digital-twin readiness (Table 1).

The matrix is intended to make the architecture auditable: every major platform requirement has a corresponding design response and a proposed verification mechanism. A component is not considered clinically deployable merely because the software function operates; it must also meet the relevant evaluation criteria.

The 3,000-participant target is justified as an initial platform-development and model-building cohort rather than as a universal sample size for all endpoints. For a descriptive community estimate with an assumed prevalence of 50%, a two-sided 95% confidence interval and an absolute precision of approximately 1.8 percentage points, the simple random-sample requirement is approximately 2,964 participants; a target of 3,000 therefore provides a transparent baseline for descriptive estimation. This calculation does not automatically account for clustering, non-response or subgroup-specific precision, so the final recruitment target should be inflated if cluster sampling or substantial attrition is expected. For prediction modelling, sample size must be evaluated separately for each outcome because the number of outcome events, not the total sample alone, determines information available for model development. As an initial planning assumption, an outcome prevalence of 10% in 3,000 participants yields approximately 300 events. With 20 candidate predictors, this corresponds to about 15 events per candidate predictor and can support an initial regularised model, subject to calibration and model-specific sample-size analysis. If prevalence is 5%, only approximately 150 events would be available, and the same 20-predictor specification would be substantially less stable. Rare outcomes such as stroke or selected cancers therefore require pooled multi-site data, longer follow-up, or a larger cohort rather than being presented as adequately powered by the 3,000-person pilot. Accordingly, the study will perform an outcome-specific sample-size calculation before locking each prediction model, considering anticipated outcome prevalence, candidate predictor degrees of freedom, shrinkage/overfitting control, expected model performance, missingness and validation requirements. The 3,000-person cohort is therefore a minimum pilot target; it is not a claim that every proposed disease model is sufficiently powered. This distinction is important for transparent reporting and avoids overinterpretation of the initial cohort.

Setting and target population

The initial setting is Lao Cai province, Vietnam, with priority attention to highland, border, and hard-to-reach communes. The target population includes ethnic-minority residents and the Kinh population for comparison. Enrollment should be stratified by sex, age, major ethnic groups represented locally, highland versus lower-altitude communes, healthy individuals, and people with priority conditions. The proposed initial dataset contains at least 3,000 individuals.

Data sources

EthMed integrates six categories of data: (1) personal and administrative data, including demographics, insurance identifiers, household context and consent status; (2) clinical history, including symptoms, previous diagnoses, allergies, medications, immunisation and family history; (3) laboratory data, including haematology, biochemistry, immunology, microbiology and point-of-care tests; (4) imaging and signal data, including X-ray, ultrasound, CT, MRI, ECG, respiratory function and other DICOM/PACS-compatible information; (5) wearable and home-monitoring data, including blood pressure, heart rate, heart-rate variability, SpO2, blood glucose, sleep and activity indicators; and (6) contextual data, including occupation, altitude, climate, water source, housing, cooking fuel, nutrition, traditional medicine use and distance to health facilities.

Table 1: Requirement-to-design traceability.

Requirement	Design response	Verification and evaluation
Geographic and connectivity barriers	Assisted-use mobile/web access, offline-tolerant data capture and commune-to-specialist telemedicine workflow	Task completion, connectivity resilience, referral completion
Fragmented clinical information	FHIR/DICOM/API integration plus harmonised analytical layer	FHIR conformance, integration success, data completeness
Heterogeneous local risk factors	Structured demographic, clinical, laboratory, wearable and contextual data model	Missingness, completeness, subgroup coverage
Need for early risk stratification.	XGBoost/LightGBM/CatBoost models with calibrated risk outputs	Discrimination, calibration, clinical utility
Language/cultural barriers	Multilingual education, health-worker assisted use and contextual recording of traditional medicine.	Usability, comprehension, cultural appropriateness
Multi-institutional data governance	Permissioned blockchain for consent/provenance/audit with off-chain clinical storage	Audit integrity, consent traceability, access transparency
Need for longitudinal personalised care.	Risk-twin and later physiology-twin state models	Temporal update accuracy, uncertainty calibration, trajectory validation
Clinical safety	Human-in-the-loop review, confidence scores, data-quality warnings and referral rules	Unsafe-event review, override monitoring, clinician agreement

Citation: Hung LN, Quang VQ, Tuan NT, Hieu LN, Thuy DN. Proposed EthMed Platform: An AI-Enabled, Blockchain-Governed and Digital-Twin-Ready Health Platform for Ethnic Minority Communities in Remote and Hard-To-Reach Areas. Health Inform Inf Manag. 2026;2(1):1-8.
Available from: https://dx.doi.org/10.17352/hiim.000002

Interoperability and storage architecture

The proposed architecture uses HL7 FHIR for structured health-data exchange, DICOM for medical images, APIs/webhooks for integration with hospital and laboratory systems, PostgreSQL for structured EHR and registry data, MinIO or equivalent object storage for medical data-lake files, and Elasticsearch for search and retrieval. Legacy prescriptions, laboratory reports and PDF records can be digitised through OCR before structured extraction and clinical validation. EthMed does not replace hospital information systems; it creates a harmonised analytical and coordination layer.

AI modelling framework

The AI framework is explicitly separated into clinical application, target definition, data handling, model development, validation, performance assessment, fairness assessment and deployment. This structure follows the reporting logic recommended by TRIPOD+AI and related prediction-model guidance [8] (Table 2).

Data handling will prevent patient-level leakage by ensuring that repeated observations from the same individual remain in the same partition and that temporal information used for prediction is available only before the prediction time. Missingness will be described and clinically interpreted; imputation, when appropriate, will be performed within training folds. Candidate predictors will be prespecified, and feature selection will not use the held-out test set. Internal validation will use bootstrap or nested cross-validation as appropriate, while temporal and/or external validation will be performed before clinical deployment. Performance assessment will include discrimination (e.g., AUROC and, for imbalanced outcomes, area under the precision-recall curve), calibration-in-the-large, calibration slope and calibration plots, sensitivity, specificity, positive and negative predictive values at clinically meaningful thresholds, Brier score and decision-curve analysis where appropriate. Fairness assessment will compare discrimination, calibration and error rates across sex, age, major ethnic groups, highland/lower-altitude setting and other relevant subgroups. Model cards, versioning, data provenance, uncertainty estimates and post-deployment drift monitoring will be maintained. Deployment will use a staged gate: research-only model → silent prospective evaluation → clinician-visible decision support → controlled clinical pilot → broader deployment. AI outputs will display confidence/uncertainty information, data-quality warnings and recommended next steps. The final clinical decision remains with qualified health professionals.

Blockchain governance and security model

Blockchain is justified only for governance functions that benefit from a shared, tamper-evident record across organisations. Conventional centralised audit logs remain preferable when one trusted authority controls all participants and rapid correction of audit records is required. Alternative technologies include append-only databases, WORM storage, signed event logs and conventional PKI-backed audit systems. EthMed proposes permissioned blockchain because community health posts, district/provincial facilities, specialist institutions and research partners may require independent verification of consent and access events without placing clinical records on a public ledger. The network is envisaged as a permissioned consortium. Nodes may be operated by authorised health-system or research institutions rather than anonymous participants. A crash-fault-tolerant or Byzantine-fault-tolerant permissioned consensus mechanism appropriate to the selected platform should be configured after institutional requirements are defined; the exact consensus protocol will be fixed during implementation rather than assumed at the conceptual stage. This avoids claiming a particular performance profile before benchmarking. Identity and key management will use institutional identity providers, PKI, role- and attribute-based access control, MFA and managed cryptographic keys. Key rotation, revocation, recovery and separation of duties will be documented. Smart contracts will be limited to explicit governance functions: recording consent status and purpose, authorising or denying access requests, recording revocation events, and writing audit references. Smart contracts will not contain full clinical data or irreversible personal identifiers. Consent management will be purpose-specific and revocable. The off-chain consent service stores the human-readable consent record, while the ledger stores a cryptographic reference, timestamp, authorised purpose, version, and revocation status. A clinician or researcher must therefore satisfy both institutional access controls and the consent policy before data are released. Data minimisation, encryption in transit and at rest, TLS 1.3, AES-256, PKI, data masking, differential privacy for research datasets, incident response and periodic access review remain mandatory; blockchain does not replace these controls.

Table 2: AI modelling design.

Model family	Clinical application	Primary targets and outputs	Development and validation
XGBoost / LightGBM / CatBoost	Community risk stratification and early warning	Hypertension, diabetes, cardiovascular risk, stroke risk, chronic respiratory disease, CKD, anaemia, malnutrition and selected fatigue pathways	Site/time-aware split; missingness analysis; imputation within training folds; hyperparameter tuning within cross-validation; calibration; temporal/external validation
Medical LLMs	Structured history-taking, record summarisation, explanation and education	Structured information extraction, summary quality, question completion, patient-facing explanation	Prompt/version locking; curated test sets; factuality and safety review; clinician evaluation; hallucination/error monitoring; no autonomous diagnosis
Computer vision / MONAI	Selected imaging decision support	Only clinically validated imaging tasks with adequate labelled data	Radiologist-supervised dataset; patient-level split; external test set; sensitivity/specificity/AUC; subgroup analysis; prospective safety assessment

Citation: Hung LN, Quang VQ, Tuan NT, Hieu LN, Thuy DN. Proposed EthMed Platform: An AI-Enabled, Blockchain-Governed and Digital-Twin-Ready Health Platform for Ethnic Minority Communities in Remote and Hard-To-Reach Areas. Health Inform Inf Manag. 2026;2(1):1-8.
Available from: https://dx.doi.org/10.17352/hiim.000002

Digital-twin methodology

EthMed uses the term digital twin only for a computational representation that has: (1) a defined state vector representing the individual or clinical system; (2) a persistent linkage to the real-world subject through governed identity references; (3) repeated data assimilation and state updating; (4) an explicit model that maps current state and new observations to updated state or predicted trajectories; (5) uncertainty estimation; (6) validation against observed longitudinal outcomes; and (7) a defined clinical or operational use case. A conventional one-time risk score that predicts an outcome from baseline variables is therefore not labelled a digital twin. The first stage is a risk twin. Its state may contain age, diagnoses, blood pressure, glucose, BMI, smoking or exposure variables, laboratory indicators and contextual determinants, and it produces continuously updated probabilities for selected conditions. The second stage is a physiology twin, which models trajectories such as blood pressure, glucose, BMI, oxygen saturation or functional status. The third stage may integrate genomic or pharmacogenomic information only after governance, consent, technical validation and clinical utility are established. State updating will occur when new validated observations enter the longitudinal record. Each update will carry a timestamp, data-quality status and provenance. Uncertainty will be quantified through predictive intervals, ensemble variance or calibrated probabilistic outputs depending on the model family. Validation will compare predicted trajectories with subsequent observed measurements using calibration, trajectory error and clinically relevant thresholds. Model drift will trigger revalidation rather than silent updating. Clinical applicability will be restricted to use cases in which a clinician can interpret the state, uncertainty, and recommended action (Table 3).

Evaluation framework

Evaluation is conducted at five levels. Technical evaluation measures uptime, data completeness, FHIR/DICOM conformance, latency, integration success, cybersecurity controls and audit-log integrity. AI evaluation measures discrimination, calibration, sensitivity, specificity, F1 score where appropriate, subgroup fairness, uncertainty quality and clinical utility. Clinical workflow evaluation measures referral appropriateness, time to consultation, chronic-disease follow-up, adherence and health-worker workload. User evaluation measures acceptability, usability, trust, language support and perceived cultural appropriateness. Governance evaluation measures consent comprehension, data-access transparency, privacy incidents, institutional compliance and research readiness. Demonstration will use representative end-to-end scenarios: community registration and consent; collection of symptoms and vital signs; import of laboratory/imaging information; AI risk assessment; low-, medium- and high-risk routing; teleconsultation; consent revocation; and research-access audit. Validation will proceed from technical verification to retrospective model validation, then silent prospective validation and finally controlled clinical implementation. No clinical deployment is authorised solely based on software demonstration.

Results

Overall architecture

The EthMed architecture contains six layers. Layer 1 is community access through mobile applications, web portals and assisted-use interfaces at commune health stations. Layer 2 is data acquisition and interoperability connecting symptoms, EHRs, laboratory results, imaging, wearable devices and contextual variables. Layer 3 is secure data storage using structured databases and medical data-lake storage. Layer 4 is AI analytics for risk prediction, stratification, early warning, summarisation and decision support. Layer 5 is telemedicine and referral coordination linking commune, district, provincial and specialist services. Layer 6 is governance for consent, identity, provenance, audit trails and controlled research access (Table 4).

Priority health pathways

EthMed prioritises health pathways where early detection and longitudinal monitoring can produce substantial benefits. For adults, the initial pathways include hypertension, diabetes, cardiovascular risk, stroke risk, chronic respiratory disease, chronic kidney disease, musculoskeletal disease and prolonged fatigue. For children and maternal-child health, the initial pathways include malnutrition, anaemia, respiratory infection risk, immunisation follow-up and maternal risk screening. For all groups, the platform records social and environmental determinants relevant to mountainous communities, including cold climate exposure, indoor smoke from cooking fuel, water safety, workload, nutrition, distance to care and traditional medicine use (Table 5).

These pathways are prioritised because they combine a meaningful prevention or early-detection opportunity with data that can reasonably be collected or imported in primary-care settings. They should not be interpreted as evidence that EthMed improves outcomes for these diseases; they define the initial implementation scope to be evaluated.

Table 3: Digital-twin design.

Stage	Core state	Update mechanism	Validation gate	Permitted use
Risk twin	Longitudinal risk factors and disease probabilities	New EHR, laboratory, wearable and contextual observations	Calibration, discrimination, temporal validation	Risk monitoring and referral prioritisation
Physiology twin	Dynamic physiological trajectories	Repeated vital signs, laboratory and sensor data	Trajectory error, uncertainty calibration, prospective validation	Monitoring and decision support
Precision twin	Physiology + genomic/ pharmacogenomic context	Validated molecular and clinical data	Clinical utility, safety, governance and prospective validation	Precision-medicine research; later controlled clinical use

Citation: Hung LN, Quang VQ, Tuan NT, Hieu LN, Thuy DN. Proposed EthMed Platform: An AI-Enabled, Blockchain-Governed and Digital-Twin-Ready Health Platform for Ethnic Minority Communities in Remote and Hard-To-Reach Areas. Health Inform Inf Manag. 2026;2(1):1-8.
Available from: https://dx.doi.org/10.17352/hiim.000002

Table 4: Proposed EthMed platform architecture.

Layer	Main functions	Core technologies	Primary verification
Community access	Mobile/web access; assisted use; multilingual education; symptom entry	Mobile app, web portal, chatbot, telemedicine interface	Usability, language support, offline/low-bandwidth tests
Interoperability	Connect EHR, HIS, LIS, PACS, labs, wearables, and legacy records	HL7 FHIR, DICOM, API/webhook, OCR	Conformance and integration testing
Storage	Structured records, medical files, imaging, search and retrieval	PostgreSQL, MinIO, Elasticsearch	Availability, integrity, access control
AI analytics	Risk prediction, triage, early warning, summarisation, education	XGBoost, LightGBM, CatBoost, medical LLMs, MONAI	Model validation, calibration, fairness, safety
Telemedicine	Remote consultation, referral, follow-up and chronic care	Video consultation, care-pathway engine	Referral appropriateness and response time
Governance	Consent, identity, provenance, access logs, research auditability	Permissioned blockchain, IAM, MFA, PKI, encryption, privacy controls	Audit integrity and consent traceability

Table 5: Priority health pathways.

Pathway	Goal	Core inputs	AI and workflow output
Hypertension and cardiovascular risk	High value of repeated community measurement and early referral	BP, age, BMI, smoking/exposure, labs, history	Risk score, repeat measurement, referral threshold
Diabetes and metabolic risk	Longitudinal glucose and metabolic monitoring is feasible at community level.	Glucose/HbA1c when available, BMI, BP, family history	Risk stratification, follow-up and testing recommendation
Stroke risk	Potential benefit from identification of modifiable vascular risk factors	BP, diabetes, prior events, age, cardiovascular variables	Risk flag and specialist referral; not autonomous diagnosis
Chronic respiratory disease	Remote settings may have substantial smoke, occupational and environmental exposure.	Symptoms, SpO2, respiratory tests, cooking fuel, occupation	Risk flag, monitoring and teleconsultation
Chronic kidney disease	Often requires longitudinal laboratory and BP monitoring	Creatinine/eGFR when available, BP, diabetes, history	Monitoring and referral prioritisation
Child malnutrition and anaemia	Community screening can identify children needing early intervention	Anthropometry, diet, symptoms, Hb where available	Risk classification and follow-up
Maternal-child health	Structured screening and follow-up can bridge geographic access gaps	Maternal history, symptoms, vital signs, scheduled visits	Risk flag, reminders and referral
Musculoskeletal/prolonged fatigue	Relevant to workload and functional health in rural communities	Symptoms, occupation, functional indicators, longitudinal history	Structured assessment and referral support

Expected operational workflow

A typical workflow begins with community registration and purpose-specific consent. A health worker or user enters demographic information, symptoms, vital signs, and available historical records. Laboratory and imaging data are imported when available. The data-quality layer checks completeness and plausibility. The AI layer then generates a preliminary risk profile with uncertainty and data-quality warnings. Low-risk users receive education and scheduled follow-up; medium-risk users receive additional testing or commune-level follow-up; high-risk users are referred for teleconsultation or in-person care. Governance services record consent and authorised data-access events. Doctors review the AI-supported summary, verify clinical relevance and make final decisions.

Expected outputs and validation

The expected outputs are: a minimum viable EthMed platform; a community health dataset of at least 3,000 individuals; an interoperability module using FHIR and DICOM-compatible design; disease-specific AI models developed only where outcome-specific sample-size requirements are met; a telemedicine workflow; a governed consent and audit module; and a staged roadmap toward validated digital health twins and multimodal AI. Additional outputs may include implementation guidelines for commune health stations, user manuals, software copyright registration, and peer-reviewed publications. The framework adds explicit validation gates. Gate 1 verifies data and interoperability; Gate 2 verifies model development and retrospective performance; Gate 3 evaluates temporal/external validity and fairness; Gate 4 evaluates silent prospective performance and workflow safety; Gate 5 evaluates controlled implementation and post-deployment monitoring. Failure at a gate requires refinement or restriction of the relevant component rather than progression by default.

Discussion

The EthMed concept addresses a structural gap in digital health: the mismatch between advanced AI platforms and the realities of remote, culturally diverse populations. Telemedicine platforms often emphasise convenience, EHR systems focus on hospital administration, and AI applications narrow to predictive or imaging tasks. EthMed integrates these components around prevention, early detection, remote consultation, secure data sharing, and long-term personalised care. Compared with Epic/MyChart, EthMed places less emphasis on patient-portal functions embedded in mature hospital systems and more emphasis on community-level screening, AI risk stratification and assisted access. Compared with Teladoc Health, EthMed treats teleconsultation as part of a longitudinal record and risk-management pathway rather than primarily as an on-demand consultation service. Compared with symptom-checking platforms, EthMed combines symptoms with laboratory, imaging, wearable, contextual and

Citation: Hung LN, Quang VQ, Tuan NT, Hieu LN, Thuy DN. Proposed EthMed Platform: An AI-Enabled, Blockchain-Governed and Digital-Twin-Ready Health Platform for Ethnic Minority Communities in Remote and Hard-To-Reach Areas. Health Inform Inf Manag. 2026;2(1):1-8.
Available from: https://dx.doi.org/10.17352/hiim.000002

EHR data. Compared with blockchain-centred health-record concepts, EthMed deliberately keeps full clinical data off-chain and limits the ledger to governance and provenance.

The priority health pathways illustrate why an integrated rather than disease-isolated platform is appropriate. Hypertension, diabetes and cardiovascular risk share repeated measurements and referral infrastructure; respiratory risk can incorporate environmental and occupational exposures; child malnutrition and maternal-child pathways require community-level follow-up; kidney disease requires longitudinal laboratory and blood-pressure monitoring. A common interoperability and governance layer can support all of these pathways while allowing disease-specific models to be validated independently. This is preferable to claiming that one AI model can safely cover all conditions. The methodology also changes the interpretation of the 3,000-person target. It is sufficient as a transparent initial platform cohort and can support several common-outcome models under appropriate assumptions, but it is not automatically adequate for rare outcomes or highly granular subgroup analysis. This staged approach is consistent with the principle that prediction-model sample size should be tied to outcome events and model complexity rather than a single arbitrary participant number.

The AI framework strengthens reproducibility by separating development data from validation data, preventing patient-level and temporal leakage, reporting discrimination and calibration, and requiring subgroup fairness assessment. This is particularly important for ethnic-minority health because an apparently strong overall model can conceal systematic miscalibration or unequal error rates in smaller groups. Local language support, assisted use and community engagement are therefore part of model safety rather than merely interface design. The blockchain decision is similarly narrower and more defensible after revision. A conventional centralised audit log may be cheaper and simpler when a single trusted controller exists. A permissioned ledger becomes useful when multiple institutions need a shared record of consent and access events and no single institution should be able to rewrite the historical governance record unilaterally. Even in that case, blockchain contributes integrity and shared provenance, not complete cybersecurity; IAM, encryption, endpoint security, privacy engineering and institutional accountability remain essential. The digital-twin roadmap is intentionally conservative. A longitudinal risk model is not automatically a digital twin. EthMed requires persistent state representation, data assimilation, updating, uncertainty, and validation before the terminology is used. The risk-twin stage can therefore be implemented earlier, while physiology and precision twins remain research stages. This staged architecture reduces the risk of premature claims and creates measurable validation requirements.

As the study results, it recommends the following policy and implementation steps. First, provincial health authorities should establish interoperable data-governance agreements before large-scale data integration. Second, community health workers should be treated as essential participants in data quality, interpretation and referral rather than passive users of technology. Third, AI procurement should require transparent model documentation, subgroup validation, uncertainty reporting and human override. Fourth, digital-health programmes should fund connectivity, devices, training and maintenance alongside software. Fifth, traditional-medicine information should be captured respectfully and subjected to scientific evaluation without allowing it to replace appropriate biomedical diagnosis. Sixth, research governance should distinguish clinical care, service improvement and secondary research, with purpose-specific and revocable consent. These recommendations align with the broader direction of digital-health ecosystem development and the need to assess implementation rather than technology alone [9-11]. The proposed architecture is therefore intended to be scalable beyond Lao Cai, but transfer to other regions should require local requirements assessment, language adaptation, data-quality evaluation and governance review [12-20].

Limitations and future work

This paper presents a conceptual architecture rather than clinical trial outcomes. No claims are made about diagnostic accuracy, reduction in morbidity, cost-effectiveness or patient outcomes. The proposed AI models require locally collected, ethically approved and quality-controlled datasets. The 3,000-person cohort is an initial target, not a universal power calculation for all endpoints. Rare diseases, complex imaging tasks and detailed subgroup analyses may require substantially larger multi-site datasets. The initial AI capabilities are expected to rely primarily on established machine-learning algorithms and foundation models rather than a completely new medical AI model developed from scratch. This reduces development time and can facilitate implementation, but existing models may not represent the epidemiological characteristics, ethnic diversity, environmental conditions, lifestyle factors, traditional-medicine practices and healthcare-access constraints of remote Vietnamese communities. Domain adaptation, local validation and continuous monitoring are therefore essential. Foundation models may also be insufficiently transparent. Future work should incorporate explainable AI, uncertainty quantification, safety monitoring, model cards, version control and human-in-the-loop governance. Prospective validation, independent safety assessment and, where appropriate, randomised or quasi-experimental implementation studies are required before clinical effectiveness claims can be made. For the digital-twin component, future work must establish validated state-space models, data-assimilation methods, uncertainty quantification and longitudinal benchmarks. For blockchain, future work should benchmark the permissioned architecture against conventional append-only audit systems on latency, cost, availability, recovery and governance burden. For the overall platform, implementation research should quantify referral completion, time to specialist consultation, health-worker workload, user trust and subgroup equity.

Conclusion

EthMed is proposed as an AI-enabled, blockchain-governed and digital-twin-ready health platform for ethnic-minority

Citation: Hung LN, Quang VQ, Tuan NT, Hieu LN, Thuy DN. Proposed EthMed Platform: An AI-Enabled, Blockchain-Governed and Digital-Twin-Ready Health Platform for Ethnic Minority Communities in Remote and Hard-To-Reach Areas. Health Inform Inf Manag. 2026;2(1):1-8.
Available from: https://dx.doi.org/10.17352/hiim.000002

communities and other remote populations with limited direct access to high-quality healthcare. The framework integrates community screening, EHR interoperability, AI risk prediction, telemedicine, secure storage, governed consent and a staged digital-twin pathway. Its principal novelty is the alignment of these technologies with an equity-oriented implementation model for mountainous and culturally diverse settings.

This paper clarifies that EthMed is a design-science artefact and research framework, not a clinically validated product. Its proposed 3,000-person cohort is an initial platform and common-outcome modelling target, while disease-specific sample-size calculations are required before each model is locked. AI deployment is gated by retrospective, temporal/external and prospective validation; blockchain is limited to governance functions for which a shared ledger has a defensible advantage; and the digital-twin label is reserved for longitudinal, updating, uncertainty-aware and validated computational representations. The next step is therefore a staged pilot followed by rigorous technical, clinical, ethical, fairness and implementation evaluation before any clinical or policy claims are made.

Conflict of Interest

The authors are members of the research team developing the EthMed platform. No commercial product has been released, and no financial benefit has been received from the technology described in this manuscript.

Acknowledgement

The project was supported by VIPTAM Institute of Technology Application, grant No. 03/VT-KHCN.

Author contributions

Conceptualisation (Vo Quy Quang, Nguyen Thanh Tuan), platform design (Vo Quy Quang, Le Ngoc Hung), manuscript drafting and correction (Le Ngoc Hung, Le Ngoc Hieu, Do Ngoc Thuy).

References

1. Bajwa J, Munir U, Nori A, Williams B. Artificial intelligence in healthcare: transforming the practice of medicine. Future Healthc J. 2021;8(2):e188-e194. Available from: https://doi.org/10.7861/fhj.2021-0095

2. Chen T, Guestrin C. XGBoost: a scalable tree boosting system. In: Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. 2016. p. 785-794. Available from: https://dl.acm.org/doi/10.1145/2939672.2939785

3. Tagde P, Tagde S, Bhattacharya T, Tagde P, Kaur A, et al. Blockchain and artificial intelligence technology in e-health. Environ Sci Pollut Res Int. 2021;28:52810-52831. Available from: https://doi.org/10.1007/s11356-021-16223-0

4. Sun T, He X, Song X, Shu L, Li Z. The digital twin in medicine: a key to the future of healthcare? Front Med (Lausanne). 2022;9:907066. Available from: https://doi.org/10.3389/fmed.2022.907066

5. Simon BD, Ozyoruk KB, Gelikman DG, Harmon SA, Turbey B. The future of multimodal artificial intelligence models for integrating imaging and clinical metadata: a narrative review. Diagn Interv Radiol. 2025;31(4):303-312. Available from: https://doi.org/10.4274/dir.2024.242631

6. Hevner AR, March ST, Park J, Ram S. Design science in information systems research. MIS Q. 2004;28(1):75-105.

7. Peffers K, Tuunanen T, Rothenberger MA, Chatterjee S. A design science research methodology for information systems research. J Manag Inf Syst. 2007;24(3):45-77.

8. Collins GS, Moons KGM, Dhiman P, et al. TRIPOD+AI statement: updated guidance for reporting clinical prediction models that use regression or machine learning methods. BMJ. 2024;385:e078378.

9. Mwogosi A, Mambile C. Digital ecosystems for healthcare communication and collaboration: a scoping review. Digit Health. 2025;11:20552076251377933. Available from: https://doi.org/10.1177/20552076251377933

10. Maurya N, Chauhan AVS, Puri I, Rohil MK, Kochar SK, Mahapatra T. Impact assessment of digital ecosystem in healthcare services: a qualitative case study of hospital data management in Bikaner District in India. Inform Med Unlocked. 2026;61:101735. Available from:

11. World Health Organization. Global strategy on digital health 2020-2025. Geneva: World Health Organization; 2021. Available from: https://www.who.int/publications/i/item/9789240020924

12. Ke G, Meng Q, Finley T, Wang T, Chen W, Ma W, et al. LightGBM: a highly efficient gradient boosting decision tree. In: Advances in Neural Information Processing Systems. 2017. p. 3146-3154.

13. Naithani N, Sinha S, Misra P, Vasudevan B, Sahu R. Precision medicine: concept and tools. Med J Armed Forces India. 2021;77(3):249-257. Available from: https://doi.org/10.1016/j.mjafi.2021.06.021

14. Prokhorenkova L, Gusev G, Vorobev A, Dorogush AV, Gulin A. CatBoost: unbiased boosting with categorical features. In: Advances in Neural Information Processing Systems. 2018;31:6638-6648.

15. Singhal K, Azizi S, Tu T, et al. Large language models encode clinical knowledge. Nature. 2023;620:172-180. Available from: https://doi.org/10.1038/s41586-023-06291-2

16. Thakur S, Lahiry S. Digital clinical trial: a new norm in clinical research. Perspect Clin Res. 2021;12(4):184-188. Available from: https://doi.org/10.4103/picr.picr_278_20

17. Health Level Seven International. FHIR release 5. Ann Arbor (MI): Health Level Seven International; 2023.

18. Digital Imaging and Communications in Medicine Standards Committee. DICOM standard. Rosslyn (VA): National Electrical Manufacturers Association; 2024.

19. Hyperledger Foundation. Hyperledger Fabric documentation. San Francisco (CA): Linux Foundation; 2024.

20. Al-Nafjan A, Aljuhani A, Alshebel A, Alharbi A, Alshehri A. Artificial intelligence in predictive healthcare: a systematic review. J Clin Med. 2025;14(19):6752. Available from: https://doi.org/10.3390/jcm14196752

Citation: Hung LN, Quang VQ, Tuan NT, Hieu LN, Thuy DN. Proposed EthMed Platform: An AI-Enabled, Blockchain-Governed and Digital-Twin-Ready Health Platform for Ethnic Minority Communities in Remote and Hard-To-Reach Areas. Health Inform Inf Manag. 2026;2(1):1-8.
Available from: https://dx.doi.org/10.17352/hiim.000002