Indonesian Journal of Infomatics

E-ISSN: 3124-033X

*Research Article*

# Predicting Multi-Morbidity Progression and Identifying Key Determinants in Chronic Disease Patients Using a Longitudinal Data-Driven Information Systems Approach: A Research Protocol for a Cohort Study in Bandung Regency, Indonesia

**Lusianto [1]*, Zaenal Arifin Hasibuan [2], Sri Supatmi[3], Adnan Shahid Khan[4]**

[1] Doctoral Program in Information Systems, Universitas Komputer Indonesia (UNIKOM), Bandung 40132, Indonesia;e-mail: maslusianto@gmail.com
[2] Master's Program in Information Systems, Universitas Komputer Indonesia (UNIKOM), Bandung 40132, Indonesia;
[3] Universiti Malaysia Sarawak, Kota Samarahan, Malaysia

* Corresponding Author : Lusianto

**Abstract:** Multimorbidity represents a critical challenge for primary healthcare systems in low- and middle-income countries (LMICs). In Indonesia, fragmented electronic health record (EHR) infrastructure limits effective chronic disease management. This research protocol presents an end-to-end information systems approach to: (1) design a validated ETL framework for heterogeneous health data; (2) develop and compare machine learning models (Random Forest, XGBoost, LSTM) for predicting multimorbidity risk; (3) identify critical determinants in an Indonesian population; and (4) evaluate a Clinical Decision Support System (CDSS) prototype. A mixed-methods, three-phase design will analyze 150,000 chronic disease patients from SIMPUS EHR data (2021-2025). Phase 2 focuses on CDSS development using explainable AI (XAI), while Phase 3 evaluates user acceptance using the Technology Acceptance Model (TAM). The study expects to produce a predictive model with AUC-ROC $\ge0.75$ and an operational CDSS prototype integrated with the Satu Sehat platform. This protocol addresses gaps in Southeast Asian LMIC data, implementation, and interpretability.

**Keywords:** Clinical Decision Support; Health Information Systems; Longitudinal Ehr Data; Machine Learning; Multimorbidity Prediction.

**Received: March,17,2026;**
**Revised: April,16,2026;**
**Accepted: May,29,2026;**
**Published: May,31,2026;**
**Curr. Ver.: May,31,2026;**

**Copyright: © 2026 by the authors. Submitted for possible open access publication under the terms and conditions of the Creative Commons Attribution (CC BY SA) license (https://creativecommons.org/licenses/by-sa/4.0/)**

## 1. Introduction

### Background and Problem Statement

Multimorbidity—defined as the co-existence of two or more chronic conditions in a single individual [1,2]—has emerged as a defining challenge for healthcare systems globally. Epidemiological evidence indicates that approximately 25% of the adult population is affected, rising sharply to over 65% in those aged 65 years and above [3]. The consequences are multilayered: increased mortality, reduced quality of life, impaired physical function, and substantially elevated healthcare utilization and costs [4,5].

In low- and middle-income countries (LMICs) such as Indonesia, the burden is compounded by a double epidemiological transition: rising prevalence of non-communicable diseases (NCDs) alongside persistent infectious disease burden, within healthcare systems that are resource-constrained and information-poor [6,7]. National surveillance data from Indonesia's Ministry of Health report a hypertension prevalence of 34.1% and diabetes mellitus at

8.5%, with increasing multimorbidity incidence particularly among the working-age population [10]. Bandung Regency, operating 62 Puskesmas (community health centres) to serve over 3.7 million inhabitants, exemplifies these systemic challenges: fragmented health information infrastructure, limited human resource capacity, and the absence of structured early-risk screening for multimorbidity [8,9].

The widespread adoption of electronic health records (EHR) under Indonesia's national Satu Sehat platform (Ministerial Regulation No. 24/2022) has generated an unprecedented longitudinal data repository [15]. Simultaneously, advances in machine learning (ML)—particularly Random Forest (RF), Extreme Gradient Boosting (XGBoost), and Long Short-Term Memory (LSTM) networks—have demonstrated strong utility for predictive modelling from complex clinical datasets [16,17,18,19]. However, the systematic translation of these methodological advances into operational health information systems (HIS) in LMIC primary care settings remains critically underdeveloped.

**Research Gaps**

A systematic review of the multimorbidity prediction literature reveals four interrelated gaps that this study is designed to address: a.) Geographic Gap: Only 12.7% of published multimorbidity studies originate from LMICs, with virtually no representation from Southeast Asia [54,56]. Patterns identified in high-income countries may not generalise to populations with distinct disease profiles, socioeconomic determinants, and healthcare access characteristics;b.) Implementation Gap: The majority of existing studies terminate at the model development stage without transitioning to clinical system deployment [24,25]. The gap between validated predictive algorithms and operational CDSS integrated into clinical workflows remains largely unaddressed;c.) Interpretability Gap: Deep learning models, despite high accuracy, present near-insurmountable barriers to clinical adoption due to their "black-box" nature [45,46]. Explainable AI (XAI) techniques are seldom embedded within clinical user interfaces in LMIC contexts;d.) Methodological Gap: There is minimal published guidance on ETL pipeline design and feature engineering strategies optimised for longitudinal LMIC health data characterised by high missingness (15–30%), heterogeneous coding, and multi-source fragmentation [40,58].

**Study Objectives**

This study aims to: a.) Design and validate an ETL framework and optimal feature engineering methods for multi-source longitudinal EHR data from Bandung Regency Puskesmas;b.) Compare predictive performance of RF, XGBoost, and LSTM algorithms for multimorbidity risk prediction at 1-, 3-, and 5-year horizons, with emphasis on the accuracy–interpretability trade-off;c.) Identify key demographic, clinical, behavioural, and contextual determinants and disease trajectory patterns specific to the Bandung Regency population through survival analysis, association rule mining, and latent class growth analysis (LCGA);d.) Design, develop, and evaluate a web-based CDSS prototype with geospatial-temporal dashboards for early risk screening and health resource planning.

## 2. Literature Review

### Multimorbidity Epidemiology and Impact

Multimorbidity prevalence in LMICs varies from 15% to 40% depending on operational definitions and measurement approaches [56]. Individuals in the lowest socioeconomic

quintile face a 40–70% higher risk of high-burden multimorbidity trajectories [41,42]. Consistent determinants identified across global studies include advanced age, low socioeconomic status, polypharmacy, smoking, obesity, and physical inactivity [37,38,39,40]. The double burden in Indonesia—where infectious disease co-exists with escalating NCDs—creates trajectory patterns potentially distinct from those documented in high-income settings [6,7].

### Longitudinal EHR Data as Analytical Foundation

Longitudinal EHR data provide a fundamental analytical advantage over cross-sectional designs: the capacity to capture temporal dynamics of multimorbidity development, identify high-risk subgroups, and model disease trajectories at population scale [11,12,13,14]. Recent trajectory analyses using EHR data have identified at least four consistent patterns: cardiometabolic (hypertension/DM progressing to cardiovascular complications, pooled HR ≈2.8), mental-physical (depression/anxiety with chronic physical conditions), respiratory, and musculoskeletal [30,31,32,33,34]. Whether these patterns hold for Indonesian populations remains empirically unanswered.

### Machine Learning for Multimorbidity Prediction

Ensemble algorithms—RF and XGBoost—have consistently achieved AUC-ROC values of 0.70–0.89 in multimorbidity prediction tasks and offer interpretability via feature importance metrics [43,44]. LSTM networks capture complex temporal dependencies in sequential EHR data but do not uniformly outperform ensemble methods, particularly in high-missingness LMIC settings, and pose greater interpretability challenges [45,46]. Key predictors consistently identified across studies include age, number of baseline conditions, polypharmacy (≥5 medications), hospitalisation history, and specific conditions such as diabetes, hypertension, and depression [47,48].

Explainable AI techniques—particularly SHAP (SHapley Additive exPlanations) and LIME (Local Interpretable Model-agnostic Explanations)—are increasingly recognised as prerequisites for clinical adoption by providing both global and instance-level explanations that align with clinician reasoning [60,61].

### Clinical Decision Support Systems in Primary Care

CDSS implementation in primary care remains limited, with key barriers including poor data quality, lack of interoperability standards, alert fatigue, and workflow resistance [49,50]. Evidence consistently demonstrates that clinician adoption is driven primarily by the actionability of system recommendations—beyond mere risk score display—and by seamless workflow integration [52,53,62,63]. The Technology Acceptance Model (TAM) provides a validated framework for evaluating CDSS user acceptance through constructs of Perceived Usefulness (PU), Perceived Ease of Use (PEOU), and Behavioural Intention (BI) [47].

## 3. Proposed Methods

### Study Design Overview

This study employs a three-phase mixed-methods research design integrating quantitative predictive analytics, IS system development, and mixed-methods implementation evaluation (Fig. 1). Ethical approval will be obtained from the Health Research Ethics Committee (KEPK) prior to data collection. All procedures comply with Indonesia's Personal Data Protection Law (UU No. 27/2022) [64,65].

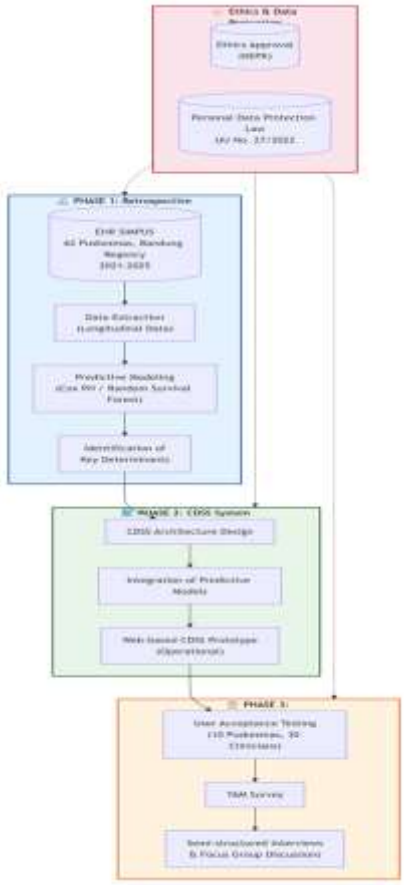


**Fig. 1** Clinical Dessison Suport System.

Phase 1 — Retrospective Longitudinal Cohort Analysis (Quantitative): Analysis of EHR data from SIMPUS across 62 Puskesmas (2021–2025) for predictive modelling and determinant identification.

Phase 2 — CDSS System Development (IS Design Science): Architecture design, development, and integration of the validated predictive models into an operational web-based CDSS prototype.

Phase 3 — Implementation Evaluation (Mixed-Methods): User acceptance testing, TAM survey, semi-structured interviews, and focus group discussions (FGDs) with 30 clinicians across 10 Puskesmas.

## Study Setting and Population

The study is conducted in Bandung Regency, West Java Province, Indonesia, which comprises 62 Puskesmas distributed across 31 sub-districts serving a population of over 3.8 million. The setting was selected for: (a) representativeness of LMIC primary care contexts with urban, peri-urban, and rural heterogeneity; (b) availability of longitudinal SIMPUS data; (c) Dinas Kesehatan institutional commitment to research collaboration; and (d) potential scalability to other Indonesian regencies.

Phase 1 target population: All adult patients (≥18 years) registered at Bandung Regency Puskesmas with at least one verified chronic disease diagnosis and a minimum of two documented clinical visits during 2021–2025. Estimated eligible cohort: 150,000–200,000 patients.

Data will be partitioned temporally: 70% training (2021–2023), 15% validation (2024 Q1–Q3), 15% testing (2024 Q4–2025).

Inclusion criteria: age ≥18 years; ≥1 verified chronic diagnosis (hypertension, DM, cardiovascular disease, COPD, asthma, CKD, stroke, cancer, depression/anxiety); ≥2 documented clinical visits; complete basic demographic data.

Exclusion criteria: terminal condition at baseline; >60% missing key clinical variables; relocation outside Bandung Regency within the first six months of observation.

Phase 3 sampling: Purposive selection of 30 clinicians (physicians, nurses, midwives) from 10 Puskesmas (3 urban, 4 peri-urban, 3 rural) for geographic representativeness.

## Data Sources and Variables

Primary data will be extracted from: (1) SIMPUS EHR—ICD-10 diagnoses, visit utilisation, laboratory results (HbA1c, lipid profile, eGFR), medication records, vital signs (BP, BMI); (2) demographic and socioeconomic registries—age, sex, marital status, education, occupation, BPJS insurance status; (3) contextual data (Dinas Kesehatan and BPS)—distance to nearest healthcare facility, population density, district poverty index, environmental quality indicators.

Outcome variables are operationalised at three levels: (a) new multimorbidity incidence (binary); (b) count of chronic conditions accumulated (count variable for Poisson/negative binomial regression); and (c) time to multimorbidity onset (survival variable for Cox modelling).

Predictor variables are organised across seven domains: demographic characteristics; clinical baseline (existing disease count, HbA1c, BP, BMI, eGFR); health behaviour (smoking status, physical activity, medication adherence); healthcare utilisation (visit frequency, hospitalisation history); polypharmacy (≥5 medications); socioeconomic status; and contextual factors (distance to Puskesmas, area poverty index).

## RQ1: ETL Pipeline and Feature Engineering

The ETL pipeline follows a three-stage architecture. Extract: Automated extraction scripts (Python/SQLAlchemy) query SIMPUS MySQL/PostgreSQL databases across all 62 Puskesmas using encrypted protocols with full audit trails. All patient identifiers (name, NIK, address) are removed and replaced with pseudonymised unique IDs enabling longitudinal linkage. Transform: (i) Missing data imputation via Multivariate Imputation by Chained Equations (MICE); (ii) ICD-10 coding harmonisation and measurement unit standardisation; (iii) derived variable computation (eGFR from serum creatinine, BMI classification); (iv) outlier treatment via winsorisation. Load: Cleaned data are stored in a PostgreSQL data warehouse with star schema architecture for analytical query efficiency.

Feature engineering employs three levels of temporal representation: (a) Aggregate features—clinical variable summaries within 3-, 6-, and 12-month windows (mean, min, max, trend slope); (b) Sequence features—disease onset ordering and inter-diagnosis time gaps; (c) Event features—cumulative visit burden, change points detected via Bayesian change-point detection. Feature subset optimisation employs forward feature selection and recursive feature elimination with predictive cross-validation.

ETL quality metrics: data completeness (% missing pre/post-imputation), logical consistency (% records with temporal inconsistencies), processing time, and scalability benchmarks.

### RQ2: Machine Learning Predictive Modelling

Three algorithms are compared (Table 1), with Logistic Regression (LASSO-regularised) and Cox Proportional Hazards as interpretable baselines.

**Table 1**. Comparative overview of ML algorithms. *Baseline model. LR = Logistic Regression.

| Algorithm | Key Strength | XAI Method | Target AUC |
|---|---|---|---|
| Random Forest | Overfitting robustness; Gini importance | SHAP TreeExplainer; Permutation | ≥0.75 |
| XGBoost | High performance; built-in missing data handling | SHAP values; LIME | ≥0.75 |
| LSTM | Captures temporal dependencies | Attention weights; DeepSHAP | ≥0.72 |
| LR (LASSO)* | Direct coefficient interpretation | Odds Ratios | Baseline |

Validation employs a multi-level strategy: temporal split (training 2021–2023 / validation 2024 Q1–Q3 / test 2024 Q4–2025) to assess temporal transportability; geographic k-fold cross-validation stratified by urban/peri-urban/rural classification to assess generalisability across healthcare settings; and three prediction horizons (1, 3, and 5 years) reflecting early intervention, strategic planning, and policy evaluation timeframes respectively.

Model performance metrics: (a) Discrimination—AUC-ROC, AUC-PR, sensitivity, specificity, PPV, NPV; (b) Calibration—Hosmer-Lemeshow test, calibration plots, Brier score (≤0.15 target); (c) Clinical utility—Decision Curve Analysis (net benefit across risk thresholds), Number Needed to Screen (NNS). XAI methods include global SHAP (mean |SHAP| per feature), local SHAP waterfall plots for individual predictions, and Pareto frontier analysis for accuracy–interpretability optimisation [60,61].

### RQ3: Determinant Analysis and Disease Trajectories

Three complementary analytical approaches are employed:

Cox Proportional Hazards Survival Analysis: The time-to-multimorbidity model is specified as:

$$h(t|X) = h_0(t) \cdot exp(\beta_1 X_1 + \beta_2 X_2 + ... + \beta_p X_p)$$

where $h(t|X)$ is the instantaneous hazard function, $h_0(t)$ the baseline hazard, and $\beta_1 \ldots \beta p$ the log hazard ratios for covariates $X_1 \ldots Xp$. Time-dependent covariates will model dynamic exposures (e.g., BMI trajectory, glycaemic control). Schoenfeld residuals will verify proportional hazards assumption. Outputs include Hazard Ratios (HR) with 95% CI and Population Attributable Fractions for modifiable risk factors.

Association Rule Mining (Apriori algorithm): Identifies co-morbidity patterns using parameters: minimum support 0.01 (1% of patients), minimum confidence 0.50, minimum lift 1.50. Disease co-occurrence networks will be visualised as weighted graphs. Latent Class Growth Analysis (LCGA): Identifies patient subgroups with homogeneous multimorbidity trajectory patterns based on six-monthly chronic disease counts over five years. Optimal class

number determined by Bayesian Information Criterion (BIC). Expected trajectory classes: low-stable, moderate-increasing, high-accelerating, and disease-specific patterns (cardiometabolic, mental-physical). Geospatial Analysis: Moran's I for global spatial autocorrelation, LISA for hot-spot/cold-spot identification, and spatial regression models (spatial lag or spatial error, depending on dependence structure) to estimate contextual determinants (healthcare access, area poverty) after controlling for individual-level factors. Choropleth maps will visualise multimorbidity burden per sub-district with healthcare facility overlay.

**RQ4: CDSS Development and Evaluation**

The CDSS architecture adopts a four-layer design (Fig. 2): Layer 1 — Data Layer (PostgreSQL data warehouse, Redis real-time cache, HL7 FHIR API for SIMPUS integration); Layer 2 — Analytics Layer (Python/FastAPI prediction microservice, serialised ML models, <500 ms inference target); Layer 3 — Application Layer (React.js frontend, RESTful API backend, role-based access control); Layer 4 — Presentation Layer (clinician dashboard + population health dashboard for Dinas Kesehatan).

The clinician dashboard presents: (a) patient risk stratification (colour-coded low/medium/high); (b) predicted multimorbidity probability at 1/3/5 years; (c) SHAP force plots showing individual risk drivers; (d) evidence-based actionable recommendations (e.g., 'Intensify glycaemic control', 'Screen for depression'); (e) longitudinal trajectory visualisation; (f) alert notifications for high-risk patients missing follow-up appointments. The population health dashboard features: interactive geospatial heatmaps per sub-district/village with drill-down capability; temporal trend charts stratified by region and age group; risk distribution histograms; Puskesmas capacity gap visualisation; and intervention programme outcome monitoring.

Security and compliance: end-to-end encryption (TLS 1.3 in transit, AES-256 at rest); compliance with UU PDP No. 27/2022; full audit logging for sensitive data access [64]. The system is designed with a minimal cognitive load interface principle (maximum 3–5 key data points per screen) to prevent alert fatigue.

User Acceptance Testing (UAT): Employs TAM constructs—PU, PEOU, Behavioural Intention (BI)—measured via 5-point Likert questionnaire, supplemented by system usability testing (task completion rate, time-on-task) and interaction log analysis. Qualitative data from semi-structured interviews (n=15) and two FGDs (n=8 each) will undergo thematic analysis using NVivo. Quantitative TAM path relationships will be analysed via SEM. The UAT follows a three-phase protocol: one week of training, three months of pilot implementation, one month of evaluation. Pre-post comparative benchmarking (interrupted time series and difference-in-differences designs) will assess clinical process efficiency outcomes.

## 4. Results and Discussion

### Results

#### *Predictive Model Performance*

Based on published benchmarks from comparable EHR-based multimorbidity prediction studies [43,44], we anticipate: AUC-ROC ≥0.75 for the 1-year horizon and ≥0.70 for the 5-year horizon; Brier score ≤0.15 indicating adequate calibration; AUC-PR reflecting class imbalance-adjusted discrimination. XGBoost is hypothesised to achieve the optimal accuracy–interpretability balance, offering high discrimination alongside direct SHAP explainability.

LSTM is expected to provide marginal discrimination gains on longer temporal sequences (3–5 years) but at greater interpretability cost, making it less suited for direct CDSS deployment in this context.

**Disease Trajectory and Determinant Findings**

We anticipate identifying 3–4 distinct multimorbidity trajectory classes, with the cardiometabolic pattern (hypertension/DM cluster) expected to be the most prevalent given Indonesia's NCD burden profile [10]. Critically, we hypothesise that determinants for the Indonesian population will diverge from high-income country patterns in at least two respects: (a) healthcare access (distance to Puskesmas, insurance status) is expected to function as a significant independent predictor—partially mediating the relationship between socioeconomic status and multimorbidity progression; (b) the mental-physical trajectory may present at younger ages relative to Western cohorts, given higher rates of untreated common mental disorders in Indonesian primary care. Modifiable risk factors with highest population attributable fractions will be formally identified for targeted public health intervention design.

**CDSS Implementation and Adoption**

Based on comparable CDSS implementations in primary care settings [51,52,53], we anticipate: overall TAM Perceived Usefulness scores ≥3.8/5.0; clinician adoption rates ≥60% at three months; reduction in time-to-risk-assessment compared to standard clinical practice. Key adoption barriers are expected to include: digital literacy variability among Puskesmas clinicians, infrastructure limitations (internet connectivity in rural sites), and concerns regarding alert overload. These findings will directly inform implementation guidelines and change management protocols for scale-up to other Indonesian regencies.

## Discussion

***Significance and Novelty***

This research protocol addresses four compounding gaps in the global multimorbidity and health informatics literature. Geographically, the study provides the first large-scale, longitudinal EHR-based multimorbidity analysis from Southeast Asia, directly challenging the persistent publication bias toward high-income country settings where 87.3% of current evidence originates [54,56]. Methodologically, it contributes a validated ETL framework explicitly designed for the data quality constraints endemic to LMIC HIS—a replicable contribution beyond the Indonesian context.

The end-to-end IS design is a distinguishing feature: the integration of ETL infrastructure, machine learning benchmarking, XAI, CDSS development, and rigorous implementation science evaluation within a single research programme represents a significant advance over the fragmented, model-centric studies that dominate the field. The embedding of SHAP force plots directly within the clinician-facing dashboard is, to the authors' knowledge, among the first such implementations in an LMIC primary care setting, operationalising the explainable AI agenda in a context where clinical trust and workflow integration are paramount.

**Alignment with National Health Informatics Strategy**

The study is positioned to directly support Indonesia's national Satu Sehat platform rollout by demonstrating a scalable model for EHR data valorisation at the district level [15,58]. Deliverables—including ETL specifications, interoperability protocols (HL7 FHIR), and CDSS architecture—are designed for adoption as reference implementations by the Ministry of Health and replication across Indonesia's 514 regencies and municipalities. Evidence

generated will also inform BPJS Kesehatan risk-based capitation refinement and Dinas Kesehatan resource allocation policy.

### Limitations

Several limitations should be acknowledged. First, as a single-regency study, external validity to other Indonesian contexts will require further validation studies, though the purposive geographic stratification (urban/peri-urban/rural) strengthens representativeness within Bandung Regency. Second, SIMPUS data quality in certain Puskesmas may be insufficient for high-confidence modelling, despite rigorous ETL and imputation procedures; sensitivity analyses and complete-case analyses will be conducted to assess the impact of missing data assumptions. Third, the observational retrospective design precludes causal inference; findings on determinants should be interpreted as associative evidence requiring experimental validation. Fourth, Phase 3 UAT is conducted with a non-randomised purposive sample, limiting generalisability of adoption findings.

### Comparison

The proposed research offers a multi-dimensional comparison against existing global literature and standard health informatics practices: a.) Geographical Context: Unlike the current body of evidence where 87.3% of studies originate from high-income countries, this research provides the first large-scale longitudinal analysis from a Southeast Asian LMIC (Indonesia);b.) Methodological Framework: While most studies utilize "clean" datasets, this protocol introduces a validated ETL framework specifically optimized for the high-missingness (15-30%) and data quality constraints endemic to LMIC health information systems;c.) Research Design: This study shifts the focus from "model-centric" designs—which often end at the development stage—to an "end-to-end" Information Systems (IS) design that encompasses the entire lifecycle from ETL infrastructure to clinical adoption evaluation;d.) Interpretability and Trust: To bridge the "black-box" gap of traditional AI, this research compares the clinical utility of standard risk scores against dashboards embedded with SHAP force plots to foster clinician trust and workflow integration;e.) Strategic Alignment: Unlike isolated research projects, this study is benchmarked for its direct alignment with Indonesia's national Satu Sehat platform and HL7 FHIR interoperability standards for nationwide replication;f.) Determinant Sensitivity: The analysis compares universal clinical predictors with localized determinants such as healthcare access distance and insurance status, which are hypothesized to play a more significant role in the Indonesian population than in Western cohorts.

## 6. Conclusion

This research protocol presents a comprehensive, end-to-end information systems approach to multimorbidity prediction, determinant identification, and clinical decision support in a resource-constrained Indonesian primary care setting. By systematically addressing the geographic, implementation, interpretability, and methodological gaps identified in the global literature, the study is designed to generate empirical contributions at three levels: scientific (evidence from an underrepresented LMIC context), operational (a deployable CDSS validated through clinical pilot), and policy (actionable recommendations for Indonesia's Satu Sehat national health platform).

The mixed-methods, phased research design—spanning ETL infrastructure, comparative ML benchmarking, XAI-integrated CDSS development, and rigorous adoption

evaluation—provides a blueprint that is both scientifically rigorous and practically replicable for other LMIC health systems seeking to leverage longitudinal EHR data for population health management. Results will be disseminated through four targeted Scopus Q1/Q2 publications and translated into policy briefs for Dinas Kesehatan Bandung Regency, the Indonesian Ministry of Health, and BPJS Kesehatan.

**Declarations**

Ethics Approval and Consent: Ethical approval will be sought from the Health Research Ethics Committee (KEPK), Faculty of Public Health, Universitas Indonesia, and from Dinas Kesehatan Kabupaten Bandung prior to data collection. A waiver of informed consent for retrospective EHR data analysis will be requested per minimal-risk research criteria. Written informed consent will be obtained from all Phase 3 clinical participants.

Availability of Data and Materials: Data are held by Dinas Kesehatan Kabupaten Bandung under applicable regulatory agreements. Aggregated, de-identified summary data supporting published findings will be made available upon reasonable request.

Competing Interests: The authors declare no competing financial or non-financial interests.

Funding: This research is self-funded as part of the doctoral programme at UNIKOM. No external funding was received.

Authors' Contributions: Lusianto: conceptualisation, methodology, data curation, writing—original draft. Zaenal Arifin Hasibuan: supervision, conceptualisation, writing—review & editing. Sri Supatmi: writing—review & editing, methodology validation.

Acknowledgements: The authors gratefully acknowledge the support of Dinas Kesehatan Kabupaten Bandung for institutional collaboration and data access commitment, and the doctoral programme supervisory team at UNIKOM for research guidance.

## References

Pearson-Stuttard, N., Ezzati, M., & Gregg, E. W. (2019). Multimorbidity—A defining challenge for health systems. *The Lancet Public Health, 4*(12), e599–e600. https://doi.org/10.1016/S2468-2667(19)30222-1

Barnett, K., Mercer, S. W., Norbury, M., Watt, G., Wyke, S., & Guthrie, B. (2012). Epidemiology of multimorbidity and implications for health care, research, and medical education: A cross-sectional study. *The Lancet, 380*(9836), 37–43. https://doi.org/10.1016/S0140-6736(12)60240-2

Smith, S. M., Soubhi, H., Fortin, M., Hudon, C., & O'Dowd, T. (2012). Managing patients with multimorbidity: Systematic review of interventions in primary care and community settings. *BMJ, 345*, e5205. https://doi.org/10.1136/bmj.e5205

Salisbury, C., Johnson, L., Purdy, S., Valderas, J. M., & Montgomery, A. A. (2011). Epidemiology and impact of multimorbidity in primary care: A retrospective cohort study. *British Journal of General Practice, 61*(582), e12–e21. https://doi.org/10.3399/bjgp11X548929

Fortin, M., Lapointe, L., Hudon, C., Vanasse, A., Ntetu, A. L., & Maltais, D. (2007). Multimorbidity's many challenges. *BMJ, 334*(7602), 1016–1017. https://doi.org/10.1136/bmj.39201.463819.2C

Afshar, A., Ahmed, S., Anjomshoa, M., et al. (2023). Multimorbidity in low- and middle-income countries: A systematic review and meta-analysis. *BMJ Open, 13*(8), e067875.

World Health Organization. (2023). *Noncommunicable diseases country profiles 2023*. Author.

Kementerian Kesehatan Republik Indonesia. (2018). *Riset kesehatan dasar (Riskesdas) 2018*. Badan Penelitian dan Pengembangan Kesehatan.

Hussain, T., et al. (2021). Chronic disease management in Indonesia: A systematic review. *Health Policy and Planning, 36*(4), 578–591.

Kementerian Kesehatan Republik Indonesia. (2024). *Profil kesehatan Indonesia tahun 2023*. Kementerian Kesehatan RI.

Denaxas, S., et al. (2025). Electronic health records: A gateway to AI-driven multimorbidity solutions—A comprehensive systematic review. *Journal of Medical Internet Research, 27*(5), e40429.

Beridze, G., et al. (2024). Patterns of multimorbidity in primary care electronic health records. *DIGITAL HEALTH, 10*, 1–18. https://doi.org/10.1177/26335565231223350

Bowling, C. B., et al. (2024). Multimorbidity trajectories in early adulthood and middle age. *DIGITAL HEALTH, 10*, 1–14. https://doi.org/10.1177/26335565241242277

Zhang, H., et al. (2024). Multimorbidity patterns and trajectories in young and middle-aged adults. *Frontiers in Public Health, 12*, Article 1349723. https://doi.org/10.3389/fpubh.2024.1349723

Peraturan Menteri Kesehatan Republik Indonesia Nomor 24 Tahun 2022 tentang Rekam Medis.

Topol, E. J. (2019). High-performance medicine: The convergence of human and artificial intelligence. *Nature Medicine, 25*(1), 44–56. https://doi.org/10.1038/s41591-018-0300-7

Rajkomar, A., Dean, J., & Kohane, I. (2019). Machine learning in medicine. *New England Journal of Medicine, 380*(14), 1347–1358. https://doi.org/10.1056/NEJMra1814259

Jiang, F., Jiang, Y., Zhi, H., et al. (2017). Artificial intelligence in healthcare: Past, present and future. *Stroke and Vascular Neurology, 2*(4), 230–243. https://doi.org/10.1136/svn-2017-000101

McKinney, S. M., et al. (2020). International evaluation of an AI system for breast cancer screening. *Nature, 577*, 89–94.

Johnson, A. P., et al. (2020). Applications of machine learning predictive models in chronic disease management. *Healthcare, 8*(4), Article 439.

Dolin, R. H., Alschuler, L., Boyer, S., et al. (2006). HL7 clinical document architecture, release 2. *Journal of the American Medical Informatics Association, 13*(1), 30–39. https://doi.org/10.1197/jamia.M1888

Ghosh, P., et al. (2021). Efficient prediction of cardiovascular disease using machine learning algorithms. *IEEE Access, 9*, 19304–19326. https://doi.org/10.1109/ACCESS.2021.3053759

Mair, F. S., et al. (2022). Multimorbidity in middle age predicts poor health and function in old age. *BMC Public Health, 22*, Article 1456.

Dobson, R., et al. (2021). Predicting onset of mental and physical multimorbidity using machine learning. *JMIR Medical Informatics, 9*(7), e29379.

Li, Y., et al. (2021). Deep learning for electronic health records. *Journal of Biomedical Informatics, 115*, Article 103691.

Akbari, A., et al. (2023). Development and validation of EHR-based prediction models for coronary heart disease. *DIGITAL HEALTH, 9*, 1–15.

Hemingway, H., Asselbergs, F. W., Danesh, J., et al. (2018). Big data from electronic health records for early and late translational cardiovascular research. *European Heart Journal, 39*(16), 1481–1495. https://doi.org/10.1093/eurheartj/ehx487

Hidayanto, A. N., et al. (2021). Success factors for health information exchange implementation. *Health Information Journal, 27*(3).

Shahab, L., et al. (2022). Machine learning for public health interventions. *Preventive Medicine, 159*, Article 107058.

Simons, D., et al. (2023). Implementation of clinical decision support systems in primary care. *BMC Medical Informatics and Decision Making, 23*, Article 145.

Canoy, D., et al. (2022). EHR-based algorithm for cardiovascular risk prediction. *European Heart Journal Digital Health, 3*(2), 231–243.

Farmer, A., et al. (2023). Clinical utility of AI-based risk prediction in diabetes management. *Diabetic Medicine, 40*, e15034.

World Health Organization. (2021). *Global strategy on digital health 2020–2025*. Author.

Andersen, K. K., et al. (2022). Multimorbidity in LMICs: An umbrella review. *Journal of Global Health, 12*, 04009. https://doi.org/10.1177/26335565221134017

Kementerian Kesehatan Republik Indonesia. (2023). *Pedoman teknis platform SATUSEHAT*. Kementerian Kesehatan RI.

Collins, G. S., Reitsma, J. B., Altman, D. G., & Moons, K. G. M. (2015). Transparent reporting of a multivariable prediction model for individual prognosis or diagnosis (TRIPOD): The TRIPOD statement. *BMJ, 350*, g7594. https://doi.org/10.1136/bmj.g7594

Moons, K. G. M., Wolff, R. F., Riley, R. D., et al. (2019). PROBAST: A tool to assess risk of bias and applicability of prediction model studies. *Annals of Internal Medicine, 170*(1), 51–58. https://doi.org/10.7326/M18-1376

Wright, A., Sittig, D. F., Ash, J. S., et al. (2009). Clinical decision support capabilities of commercially available clinical information systems. *Journal of the American Medical Informatics Association, 16*(5), 637–644. https://doi.org/10.1197/jamia.M3111

Gerber, M. S., et al. (2018). Predicting disease progression using electronic health records. *IEEE Journal of Biomedical and Health Informatics, 22*(3), 835–842.

Republik Indonesia. (2022). *Undang-Undang Republik Indonesia Nomor 27 Tahun 2022 tentang Pelindungan Data Pribadi*.

Komisi Nasional Etik Penelitian Kesehatan. (2023). *Pedoman nasional etik penelitian kesehatan*. Kementerian Kesehatan Republik Indonesia.