Machine Learning Models for Predicting Health Insurance Coverage Gaps

Tresna Dey

Abstract

In 2026, the "Price of Health" in Bangladesh remains the most volatile currency, defining not only access to medical care but also the very survival of millions. Despite ongoing policy efforts toward Universal Health Coverage (UHC), the statistical landscape is stark: nearly 74% of all health expenditures are still paid out-of-pocket (OOP), and less than 1% of the population is formally insured.

For the approximately 70 million citizens facing financial hardship due to medical costs, the absence of insurance is more than a lack of a policy—it is a structural barrier that translates directly into poverty, lost opportunity, and avoidable suffering. In this context, Predictive Analytics and Machine Learning (ML) are emerging as critical tools to identify and ultimately bridge these health coverage gaps, enabling policymakers to move from reactive interventions to proactive, data-driven strategies.

The anatomy of the coverage gap reveals a complex interplay of socio-economic factors, employment status, and geographic disparities. Data from 2025–2026 highlights the depth and scale of the challenge. Approximately 41.7% of the population faced catastrophic health expenditures during this period, defined as spending exceeding 10% of household income on health care. For a daily wage earner, a single hospital admission can trigger "acute poverty," forcing families to sell livestock, mortgaging land, or compromising on essential nutrition. The formal-informal divide is particularly stark: while 85.3% of professionals in the formal sector express a willingness to pay for community-based health insurance, informal workers—the backbone of the national economy—remain largely unreached. Out-of-pocket expenses continue to surge despite incremental increases in the national health budget, rising to an average of BDT 1,564 per month per household, with medicines alone accounting for nearly 60% of this burden.

Machine Learning offers a transformative approach to addressing these gaps by identifying the "likely uninsured" before they experience catastrophic financial shocks. By leveraging household, demographic,

and socio-economic data, ML models can detect patterns and predictors of coverage exclusion that traditional methods may overlook. Ensemble tree-based models, particularly Random Forest and XGBoost, have demonstrated superior performance in this context, effectively capturing non-linear relationships between income, employment type, and insurance access. These models achieve mean accuracies exceeding 91% and AUC (Area Under the Curve) scores of 0.96, significantly outperforming linear regression models that struggle with high-dimensional and correlated social data.

A key innovation lies in understanding feature importance through techniques like SHAP (SHapley Additive exPlanations), which reveal that household wealth quintile, maternal education, and employment sector are among the strongest predictors of health insurance exclusion. For example, households led by informally employed heads or with low maternal literacy are disproportionately likely to remain uninsured. These insights transform health insurance from a purely administrative challenge into a measurable, data-driven equity issue. By quantifying who is left behind and why, policymakers can design interventions tailored to the specific vulnerabilities of each community.

Ensuring fairness in ML models is critical. Historical insurance data reflects systemic inequities, meaning that naive predictive models may inadvertently reinforce existing gaps, overlooking marginalized groups who have historically been excluded from formal coverage. To counter this, researchers are integrating fairness-aware layers into the algorithms, applying constraints that ensure equitable identification of “at-risk” populations across gender, region, and socio-economic strata. For instance, a woman in rural Sylhet should have the same probability of being flagged for targeted insurance outreach as a man in urban Dhaka, regardless of historical representation in insurance datasets. These fairness adjustments not only improve the accuracy of intervention targeting but also uphold social justice principles, ensuring that AI contributes to reducing, rather than perpetuating, inequality.

The practical application of these predictive models extends across multiple policy domains. By mapping the “likely uninsured,” government programs such as the Shasthyo Suroksha Karmachari (SSK) scheme can pre-emptively allocate insurance subsidies to households most at risk of catastrophic expenditure. Predictive modeling also helps differentiate between structural barriers—such as affordability—and informational gaps, such as a lack of awareness about available schemes, allowing community-based health insurance programs to tailor their outreach and design. By integrating these insights directly into the Bangladesh Universal Health Coverage Roadmap 2026–2035, policymakers gain evidence-based guidance to allocate resources strategically, maximizing what can be termed the “Equity ROI” of each intervention.

Machine Learning-driven health insurance strategies also offer the opportunity to leverage the observed willingness to pay among formal-sector professionals. Data indicates that approximately 5.6% of income could be mobilized to support pooled-fund insurance schemes, creating a scalable financial base for community-based models. Coupled with a National Health Information Exchange, ML algorithms could

track insurance enrollment in real time, flag coverage gaps, and monitor claims data for compliance, fraud, and efficiency. Such digital integration ensures that insurance expansion is not only targeted but also transparent and accountable, maintaining the public trust essential for a sustainable national insurance system.

The long-term implications of successfully addressing coverage gaps are profound. For individuals, insurance mitigates the financial shock of illness, preventing households from falling into poverty due to preventable medical costs. At a societal level, pre-emptive coverage reduces overreliance on informal borrowing, limits catastrophic expenditures, and promotes better utilization of preventive and primary health services. Economically, comprehensive insurance coverage can stabilize household consumption, preserve productivity, and support inclusive growth, while also reducing the hidden costs of untreated disease.

Health insurance in Bangladesh is not simply a fiscal instrument; it is a mechanism for social protection and resilience. In an environment where medical costs can determine the trajectory of a family's life, ML-based risk prediction turns previously invisible vulnerabilities into actionable intelligence. By anticipating where financial protection is lacking, the system can intervene before households endure the economic and health consequences of being uninsured. This proactive approach represents a shift from reactive crisis management toward anticipatory, equity-focused policy design, where interventions are both data-informed and human-centered.

Looking toward the broader horizon, the integration of predictive analytics into the national health strategy offers multiple synergies. Real-time dashboards can guide emergency response, identify underserved regions, and coordinate the expansion of insurance services with other social protection programs. Furthermore, these insights can inform public education campaigns to raise awareness about insurance benefits, encourage enrollment, and foster a culture of preventive health-seeking behavior. By coupling AI predictions with targeted subsidies, digital enrollment platforms, and policy oversight, Bangladesh can create a health insurance ecosystem that is both inclusive and sustainable.

In conclusion, health insurance in Bangladesh remains the "missing shield" in the fight against poverty and illness. The widespread reliance on out-of-pocket payments exposes millions to catastrophic risk, perpetuating cycles of vulnerability and inequality. Machine Learning provides the tools to detect, map, and ultimately address these gaps with unprecedented precision. By applying ensemble models, fairness-aware algorithms, and real-time data integration, we are moving toward a future in which health coverage is not a privilege but a right. In the era of "Smart Bangladesh," the predictive frameworks we build today are more than code—they are the financial safety net that safeguards health, dignity, and opportunity for every citizen.