ESMO GOOD SCIENCE BETTER MEDICINE BEST PRACTICE

ANNALS OF ONCOLOGY DRIVING INNOVATION IN ONCOLOGY

ORIGINAL ARTICLE

Machine learning reveals country-specific drivers of global cancer outcomes

M. S. Patel[1,2*], C. S. Pramesh[3], N. N. Sanford[4], E. J. G. Feliciano[5,6], P. L. Nguyen[7], P. Iyengar[2], T. P. Kingham[8], J. Willmann[9,10], B. A. Mahal[11,12], N. Y. Lee[2], M. J. K. Magsanoc-Alikpala[13,14], M. Mutebi[15], J. F. Wu[16], J. P. G. Robredo[5,17] & E. C. Dee[2*]

[1]Department of Molecular Biosciences, The University of Texas at Austin, Austin; [2]Department of Radiation Oncology, Memorial Sloan Kettering Cancer Center, New York, USA; [3]National Cancer Grid, Tata Memorial Centre, Homi Bhabha National Institute, Mumbai, India; [4]Department of Radiation Oncology, The University of Texas Southwestern Medical Center, Dallas, USA; [5]School of Medicine and Public Health, Ateneo de Manila University, Pasig City, Philippines; [6]Department of Medicine, NYC Health + Hospitals/Elmhurst, Icahn School of Medicine at Mount Sinai, Queens; [7]Department of Radiation Oncology, Dana-Farber Cancer Institute/Harvard Cancer Center, Boston; [8]Global Cancer Disparities Initiatives, Memorial Sloan Kettering Cancer Center, New York, USA; [9]Department of Radiation Oncology, University Hospital Zurich, University of Zurich, Zurich, Switzerland; [10]Department of Medical Physics, Memorial Sloan Kettering Cancer Center, New York; [11]Sylvester Comprehensive Cancer Center, University of Miami Miller School of Medicine, Miami; [12]Department of Medicine, University of Miami Miller School of Medicine, Miami, USA; [13]ICanServe Foundation, Manila; [14]Cancer Coalition Philippines, Makati, Philippines; [15]Department of Surgery, Aga Khan University, Nairobi, Kenya; [16]Division of Hematology and Oncology, Department of Medicine, Medical College of Wisconsin, Milwaukee; [17]Ariadne Labs, Harvard T.H. Chan School of Public Health, Boston, USA

Available online XXX

Background: Global inequities in access to cancer diagnostics and treatment contribute to wide variation in cancer mortality-to-incidence ratios (MIRs), a proxy for survival. We aimed to develop an interpretable machine learning framework to quantify country-specific health system contributors to MIR and inform policy prioritization.

Materials and methods: We assembled national MIRs from GLOBOCAN 2022 for 185 countries and health system indicators from multilateral sources, including gross domestic product (GDP) per capita, universal health coverage (UHC) index, radiotherapy centers per population, health spending (%GDP), out-of-pocket expenditure, work force densities (physicians; nurses/midwives; surgical work force), pathology availability, Human Development Index, and gender inequality index. A CatBoost gradient-boosting model was trained with repeated leave-one-country-out cross-validation (10 repeats; 1850 predictions). Nested hyperparameter optimization and strict leakage control were used. Model interpretability employed SHapley Additive exPlanations (SHAP; TreeExplainer) to generate global and country-level feature attributions. SHAP values, model-derived metrics quantifying each factor's contribution to cancer outcomes, were generated. Performance metrics included R^2, root mean squared error (RMSE), mean absolute error, and Pearson correlation; uncertainty was estimated by bootstrap resampling.

Results: The model showed strong out-of-sample performance [$R^2 = 0.852$, 95% confidence interval (CI) 0.801-0.891; RMSE 0.057, 95% CI 0.050-0.064]; correlation between predicted and observed MIRs was $r = 0.923$ ($P = 8.30 \times 10^{-78}$). Global SHAP contributions ranked GDP per capita (22.5%), radiotherapy centers per population (15.4%), and UHC index (12.9%) as the leading determinants. Country-specific SHAP profiles revealed substantial heterogeneity in dominant drivers across settings, enabling tailored policy levers (e.g. infrastructure, coverage expansion, or financial protection). An accompanying web interface provides country-level SHAP summaries for decision support.

Conclusions: An explainable machine learning approach accurately predicts national MIRs and decomposes predictions into country-specific health system attributions. While ecological and noncausal by design, the SHAP profiles translate population-level associations into actionable hypotheses for prioritizing investments—highlighting, across many contexts, radiotherapy capacity and UHC expansion as recurrent levers, and underscoring that higher total health spending alone may be insufficient without strategic allocation. Prospective, country-specific evaluations are warranted to test whether targeting model-identified drivers improve cancer outcomes.

Key words: cancer outcomes, machine learning, SHAP analysis, health systems, global oncology, universal health coverage

*Correspondence to: Mr. Milit S. Patel, Department of Molecular Biosciences, The University of Texas at Austin, Austin, TX, 78712, USA. Tel: +1 (512) 471-4641. E-mail: militpatel@utexas.edu (M. S. Patel).

Dr. Edward Christopher Dee, Department of Radiation Oncology, Memorial Sloan Kettering Cancer Center, New York, NY, 10065, USA. Tel: +1 (212) 639-2000 E-mail: deee1@mskcc.org (E. C. Dee).

INTRODUCTION

Cancer is a leading cause of morbidity and mortality worldwide. Global disparities in access to cancer diagnostics and care, along with associated cancer outcomes, are immense.[1] Recent evidence from the Global Burden of Disease Study 2023 underscores that, as of 2023, there were an estimated 18.5 million new cancer cases and 10.4 million cancer deaths around the world, with >65% of these deaths occurring in low- and middle-income countries.[2] Despite a decline in age-standardized mortality rates in many high-income regions, progress elsewhere has been limited, and global numbers of cases and deaths are projected to rise sharply through 2050, especially in resource-constrained areas.[2] Thus, identifying actionable cancer systems factors specific to each country and context is needed.

Prior work on pan-cancer ecological findings established universal health coverage (UHC) and gross domestic product (GDP) per capita as robust, independent predictors of cancer outcomes in linear models.[3] However, while classical regression provides vital baseline associations, it is constrained by assumptions of linearity and does not fully account for the nuanced, context-specific interplay between health system factors and cancer outcomes.

To capture the multifaceted nature of national cancer systems, we applied machine learning approaches capable of modeling nonlinear and context-dependent relationships among health system indicators. Specifically, gradient-boosting methods such as CatBoost enable the characterization of complex interactions between economic, infrastructural, and policy variables that shape cancer outcomes. To enhance the interpretability of these models, we incorporated SHapley Additive exPlanations (SHAP), which quantify the relative contribution of each input to country-specific cancer mortality-to-incidence ratios. This framework thus integrates predictive precision with transparency, enabling actionable insights into the health system determinants of global cancer outcomes.[4]

We sought to move beyond the traditionally 'black box' machine learning paradigm into one capable of actionable, country-tailored insight. Our overarching goal is to translate these interpretable, country-specific predictions into a practical prioritization of health system levers—guiding policymakers toward the highest yield, context-appropriate investments to reduce cancer mortality and close equity gaps.

MATERIALS AND METHODS

Data sources and variables

We used pan-cancer mortality-to-incidence ratio (MIR) as our primary outcome variable, based on GLOBOCAN 2022.[5] The MIR represents the proportion of cancer cases that result in death, serving as a proxy for cancer care effectiveness.[3] Health system metrics were collected from the World Health Organization, the World Bank, United Nations agencies, and the Directory of Radiotherapy Centres (DIRAC). The following health system variables were included: health spending as percentage of GDP, physicians per 1000 population, nurses and midwives per 1000 population, surgical work force per 1000 population, UHC index (a composite measure of health service coverage and financial protection), pathology services availability, human development index, radiotherapy centers per 1000 population, GDP per capita, gender inequality index, and out-of-pocket expenditure percentage.

Machine learning pipeline and cross-validation

Our machine learning pipeline implemented several methodological advances over traditional regression approaches. We used repeated leave-one-country-out cross-validation (10 repeats, 1850 total predictions) to ensure robust performance estimation and prevent overfitting. Cross-validation is a technique that systematically excludes portions of data to test model performance on unseen cases, while overfitting occurs when a model memorizes training data rather than learning generalizable patterns. This approach provides unbiased estimates of model performance by systematically excluding each country from training and using it as a test case across multiple iterations.

Within each cross-validation fold, we implemented a rigorous preprocessing pipeline to prevent data leakage—the inadvertent use of future or test information during model training. Missing values were handled using CatBoost's native missing value processing, which treats missing values as a separate category during tree construction rather than requiring explicit imputation. This approach leverages CatBoost's ability to learn optimal splits for missing data patterns, often outperforming traditional imputation methods by preserving the informative nature of missingness patterns in health system data.

After handling missing values, we applied variance inflation factor-based (VIF) feature selection (threshold = 10) to address multicollinearity—the presence of highly correlated predictor variables that can destabilize model interpretability. VIF measures how much the variance of a regression coefficient increases due to collinearity, with values >10 indicating problematic correlation levels. Features with VIF exceeding the threshold were iteratively removed, prioritizing the removal of the highest VIF feature until all remaining features had acceptable collinearity levels.

To ensure transparency in feature selection, we tracked feature retention across all cross-validation iterations. VIF-based filtering demonstrated perfect stability, with all features retained in 100% of cross-validation folds (1850/1850 iterations). This complete retention indicates that the initial feature set—comprising GDP per capita, UHC index, radiotherapy centers per 1000 population, health spending as percentage of GDP, pathology services availability, physicians per 1000 population, nurses and midwives per 1000 population, surgical work force per 1000 population, out-of-pocket expenditure percentage, and gender inequality

index—exhibited acceptable multicollinearity levels (VIF <10) across all national contexts. The consistent feature stability across diverse country compositions in each cross-validation fold supports the robustness of these health system indicators as independent predictors of cancer MIR.

Model selection and hyperparameter optimization

We selected CatBoost as our primary modeling framework due to its superior handling of missing values, categorical features, and resistance to overfitting. CatBoost is a gradient-boosting algorithm that builds models by iteratively combining weak predictors (typically decision trees) to create a strong ensemble predictor. For each outer cross-validation fold, we carried out nested hyperparameter optimization using Optuna with 50 trials per fold. Hyperparameters are configuration settings that control model behavior (such as learning speed and model complexity) rather than parameters learned from data. The specific hyperparameter ranges, optimization scheme, and full software references are provided in Supplementary Methods, available at https://doi.org/10.1016/j.annonc.2025.11.014. The complete hyperparameter search space is detailed in Supplementary Table S1, available at https://doi.org/10.1016/j.annonc.2025.11.014. The optimization process used Bayesian optimization with Tree-structured Parzen Estimator (TPE) sampling—a probabilistic approach that learns from previous trials to efficiently explore the hyperparameter space.

Five-fold inner cross-validation with countries as groups ensured that hyperparameter selection was based on genuine out-of-sample performance. This ensured that all data from the same country appears in only one fold, preventing information leakage between training and validation sets. Early stopping (50 rounds) prevented overfitting during model training by halting the training process when validation performance on a held-out fold stopped improving. All random seeds were fixed to ensure reproducibility across the 10 repeated cross-validation cycles.

SHAP analysis for model interpretability

SHAP analysis was selected over alternative explainability methods due to its theoretical foundation and unique properties.[6] SHAP is a method for explaining individual predictions by quantifying each feature's contribution to the prediction. Unlike permutation importance or linear coefficients, SHAP value, which are model-derived metrics quantifying each factor's contribution to cancer outcomes, satisfy mathematical axioms ensuring that feature contributions sum exactly to the difference between each prediction and the model's baseline—the average prediction the model would make without any input features. This enables precise quantification of how much each health system factor contributes to a country's predicted cancer mortality, providing the foundation for evidence-based policy recommendations.

To generate interpretable predictions, we made use of the SHAP framework, which brings unprecedented granularity, revealing substantial heterogeneity. For some countries, radiotherapy infrastructure may eclipse economic indicators in explanatory weight, while for others, work force density or out-of-pocket costs emerge as dominant barriers. This level of detail enables health policymakers to prioritize interventions not just generically, but in a data-driven, context-specific manner, targeting the precise bottlenecks limiting their national cancer control efforts.

SHAP values were computed for each country prediction using TreeExplainer,[7] which provides exact calculations for tree-based models like CatBoost. TreeExplainer leverages the tree structure to compute SHAP values efficiently without approximation. We aggregated SHAP values across repeated predictions to generate stable feature importance rankings and country-specific explanations. Force plots visualized individual country predictions, showing how each health system factor pushed the prediction above or below the global baseline.

Statistical analysis and performance evaluation

Model performance was evaluated using multiple metrics calculated at the country level by averaging predictions across the 10 repeats. Primary metrics included R^2 (the proportion of variance in outcomes explained by the model), root mean squared error (RMSE, the square root of average squared prediction errors), and mean absolute error (the average absolute difference between predicted and actual values). Bootstrap resampling (2000 iterations) generated 95% confidence intervals (CIs) for all performance metrics. Bootstrap resampling involves repeatedly sampling with replacement from the original data to estimate the uncertainty around our performance statistics.

Statistical significance was assessed using Pearson correlation between actual and predicted values across countries. Feature selection stability was quantified by calculating the percentage of cross-validation folds in which each variable was selected after VIF filtering.

Methodological innovations and clinical translation

By embedding SHAP analysis within CatBoost, our study directly addresses major limitations of prior ecological research: (i) it provides a defensible, axiomatic basis for feature attribution grounded in game theory, mathematical principles ensuring fair allocation of contributions among predictors; (ii) it supports scenario modeling to estimate expected gains from policy shifts (e.g. scaling up radiotherapy access); and (iii) it paves the way for integrating additional dimensions such as within-country inequities across rural—urban divides and insurance coverage gaps, disparities that may rival or exceed between-country differences in magnitude and require targeted subnational interventions, as well as policy implementation tracking and cost-effectiveness analyses. Translational value refers to the practical application of research findings to real-world health care policy and practice. Taken together, this approach provides an interpretable set of associations to

inform health system strengthening and generate hypotheses for future causal research.

All analyses were conducted in Python using scikit-learn,[8] CatBoost,[9] Optuna,[10] statsmodels,[11] and SHAP[4] libraries with parallelization using joblib[12] for computational efficiency and tqdm[13] for utility. All data manipulation was handled by pandas[14] and numpy[15] libraries. All plots were created using the matplotlib,[16] seaborn,[17] and scipy[18] libraries. The computational environment and model configuration parameters are summarized in Supplementary Table S2, available at https://doi.org/10.1016/j.annonc.2025.11.014.

Ethics approval

This study did not require ethical approval given the use of publicly available data that does not constitute patient health information or human subject research.

RESULTS

The machine learning model demonstrated robust predictive performance with $R^2 = 0.852$ (95% CI 0.801-0.891) and RMSE = 0.057 (95% CI 0.050-0.064) (Figure 1). The correlation between predicted and actual cancer mortality ratios was highly significant ($r = 0.923$, $P = 8.30 \times 10^{-78}$).

Global feature performances are presented in Figure 2. However, the key advancement lies in SHAP's ability to provide country-specific policy guidance. For each nation, SHAP values reveal exactly which health system factors most contribute to higher or lower cancer mortality rates. These are visualized in Supplementary Figure S1, available at https://doi.org/10.1016/j.annonc.2025.11.014 (heatmap figure). We have made country-specific estimates available in a usable web interface.[19]

SHAP values, derived from repeated model validation, reveal highly specific patterns in how health system factors affect cancer outcomes for each country. A negative SHAP value corresponds to a lower MIR, whereas a positive SHAP value corresponds to a higher MIR. In Turkey, the analysis shows that the current number of radiotherapy centers per 1000 people is a major determinant of cancer mortality relative to incidence, with a mean SHAP value of −0.0337.[20,21] This suggests that policies designed to strategically increase radiotherapy access may be associated with significant improvements in outcomes. On the other hand, Turkey's positive SHAP value for health spending as a percentage of GDP (+0.0258) indicates that higher spending has not translated into better MIR ratios, highlighting the need to reassess how current resources are deployed rather than simply increasing budgets.

In Brazil, the UHC index holds the greatest impact among evaluated factors, with a SHAP value of −0.0230.[22,23] This highlights universal health coverage as a priority area for policy consideration in Brazil, given its strong association with MIR in our ecological model. However, intervention studies would be required to establish causality for outcome improvements. Pathology services, with a modest positive SHAP value (+0.0093), may not substantially limit Brazil's outcomes, potentially directing policy attention toward broader coverage solutions.

Poland's strengths are rooted in radiotherapy center density, as evidenced by a SHAP value of −0.0246.[24] This suggests that recent efforts at the national level to strengthen health insurance coverage and to support service access may have led to more pronounced gains than general health spending, which shows a small, more negligible positive effect (+0.0029). For Japanese policymakers, the data portray an altogether different picture: all assessed health system attributes correlate only with improvement in cancer outcomes.[25,26] The analysis reveals that radiotherapy center density (−0.0579) and nurse and midwife density (−0.0539) are the strongest correlates of Japan's low MIR.[27] This suggests an association between targeted infrastructure investments in cancer treatment facilities and health care work force development and Japan's cancer outcomes. The negative SHAP values across all predictors suggest these factors are working in Japan's favor by decreasing MIR. Therefore, Japan's strategy should prioritize preserving and enhancing these effective drivers.

In Malaysia, the analysis highlights GDP per capita as a key lever, with a negative SHAP value of −0.0256.[28,29] This supports growth-focused policies for long-term cancer control. However, higher health spending as a share of GDP carries a positive effect (+0.0158), suggesting a need for better strategic use of available resources, possibly directing funds toward the most impactful technologies or services rather than across-the-board budget increases.

In Ghana, SHAP analysis reveals that the UHC index has a strong positive SHAP value (+0.0205) in relation to cancer MIR, indicating that current gaps in insurance coverage or service accessibility are major contributors to poorer cancer outcomes.[30] Conversely, the availability of nurses and midwives per 1000 people has a negative SHAP value (−0.0092), suggesting that increases in nursing and midwifery work force are associated with small but meaningful improvements in cancer outcomes. Together, these findings suggest that Ghana should prioritize health policies that advance UHC as their primary lever for improving cancer survival, while also recognizing the continued importance of investing in essential frontline health workers to support long-term gains.

Thailand's SHAP profile also emphasizes the value of broader UHC, with a strong negative impact (−0.0351).[31] The slightly positive association between health spending as a percentage of GDP and MIR (+0.073) mirrors trends seen in other countries; it indicates the importance not just of how much is spent, but how efficiently resources are directed.

In China, SHAP analysis highlighted GDP per capita (−0.0245) and UHC index (−0.0221) as the dominant negative drivers of mortality-to-incidence ratio, indicating that continued economic growth and further expansion of universal health coverage synergistically contribute to reductions in cancer mortality. By contrast, out-of-pocket health expenditure exhibited a substantial positive SHAP value (0.0225), suggesting that high direct costs for patients remain a critical barrier to optimal cancer outcomes, even amidst national improvements in health financing and access. These findings

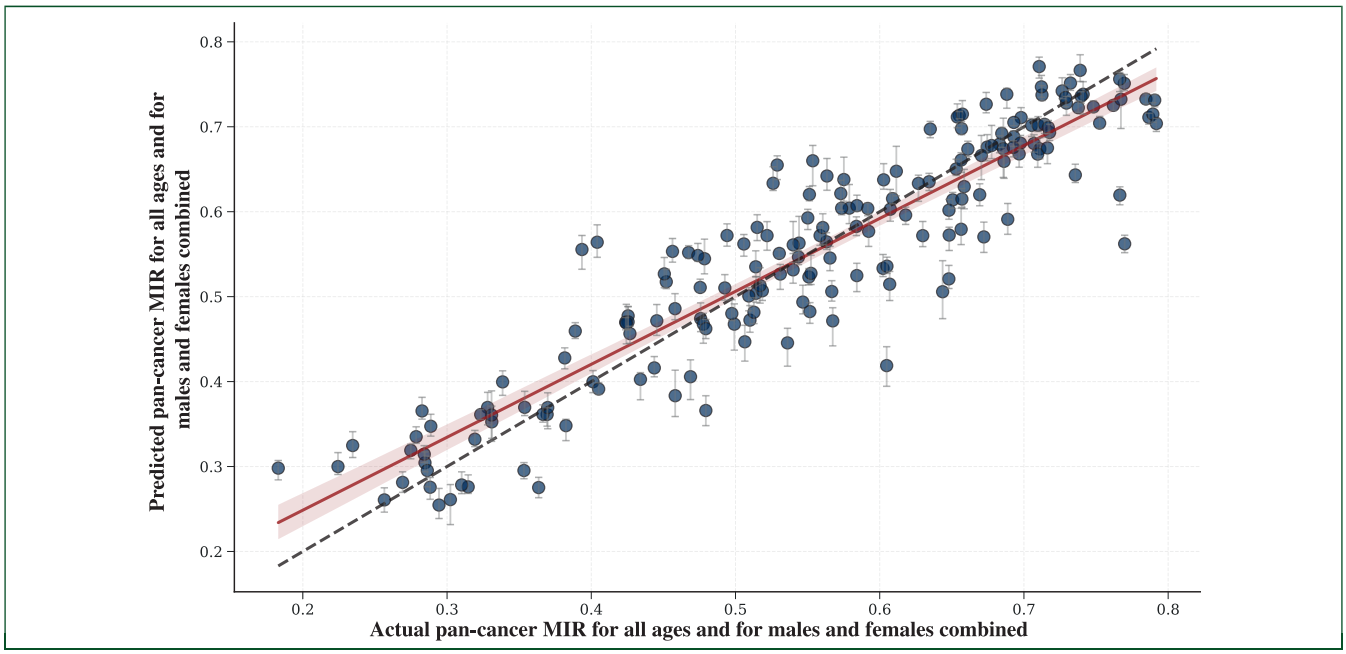

Figure 1. Model performance for predicting country-level average mortality-to-incidence ratio (MIR) for all-age, mixed-sex pan-cancer cases. The scatter plot compares actual versus predicted MIRs, with each point representing a country and error bars indicating 95% confidence intervals. The solid line shows the regression fit, the dashed line denotes perfect prediction (y = x), and the model achieves $R^2 = 0.852$ [95% confidence interval (CI) 0.801-0.891], root mean squared error (RMSE) = 0.057 (95% CI 0.050-0.064), $P < 0.001$.

underscore that while China's rapid health system development is yielding important gains in cancer control, disparities in financial protection and coverage persist, warranting intensified policy focus on reducing out-of-pocket expenditures and further strengthening UHC implementation to maximize health system impact.[32]

We also include a dendrogram depicting the results of agglomerative hierarchical clustering on 185 countries, providing insight into countries that exhibit similar patterns in their SHAP values (Supplementary Figure S2, available at https://doi.org/10.1016/j.annonc.2025.11.014). Together, these detailed country profiles arm policymakers with the evidence to design precise, high-impact interventions, whether that means investing in infrastructure, expanding coverage, or reorienting spending to maximize reductions in cancer mortality.

Across diverse national contexts, two patterns emerge as consistent policy levers: radiotherapy infrastructure and UHC expansion demonstrate robust negative associations with cancer mortality ratios in most countries analyzed. In contrast, health spending as a percentage of GDP frequently shows positive SHAP values, as observed in Turkey (+0.0258), Malaysia (+0.0158), and Thailand (+0.073), suggesting that higher health expenditure alone does not guarantee improved outcomes. These findings underscore that strategic allocation of resources toward high-impact interventions, rather than simply increasing total health budgets, represents the more actionable pathway for policymakers seeking to reduce cancer mortality.[3] Country-specific predictions and SHAP attributions for countries with the highest mortality-to-incidence ratios are detailed in Supplementary Table S3A, available at https://doi.org/10.1016/j.annonc.2025.11.014. Conversely, countries with the lowest predicted mortality-to-incidence ratios and their corresponding SHAP drivers are presented in Supplementary Table S3B, available at https://doi.org/10.1016/j.annonc.2025.11.014.

Figure 3 demonstrates the heterogeneity in feature impacts across countries, with the same health system factor showing positive contributions to cancer mortality in some nations while reducing it in others.

DISCUSSION

This approach represents a paradigm shift from describing associations to providing actionable intelligence. While previous linear analysis established that UHC and GDP per capita matter globally, SHAP analysis reveals the specific magnitude of their importance for each country and identifies additional factors that may be more impactful in specific national contexts. For instance, while radiotherapy capacity ranks second globally by the SHAP analysis, it may be the primary driver of poor outcomes in specific countries with adequate number of facilities but insufficient cancer infrastructure.

The model's strong performance across diverse global contexts (185 countries spanning all income levels) demonstrates robustness and generalizability. The consistent feature selection across cross-validation runs confirms model stability and reliability of the SHAP attributions.

While this analysis aggregates pan-cancer outcomes to provide broad health system insights, the relative importance of these drivers likely varies substantially by cancer type. For cancers highly responsive to radiotherapy, such as breast, cervical, and certain head and neck malignancies,

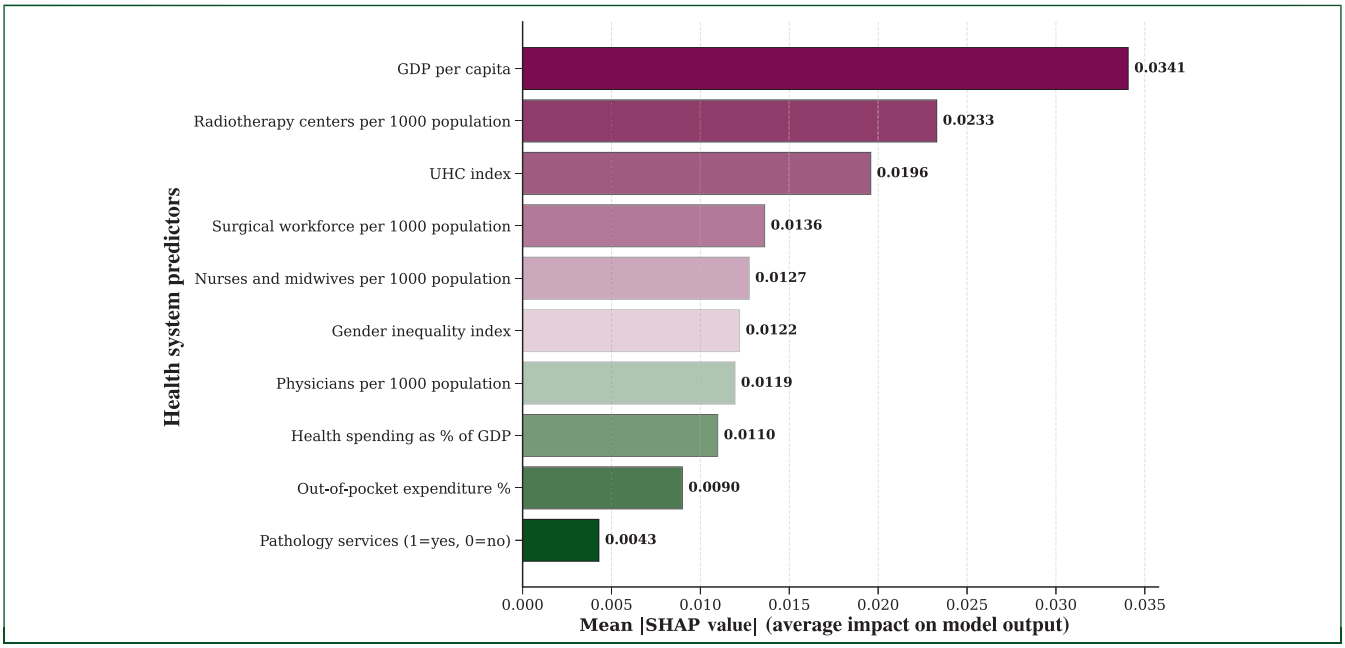

Figure 2. Global feature importance based on mean absolute SHAP values, revealing gross domestic product (GDP) per capita (22.5%), radiotherapy centers per population (15.4%), and universal health coverage (UHC) index (12.9%) as the most impactful factors worldwide. Bars indicate each predictor's average impact on model output for pan-cancer mortality-to-incidence ratio prediction, with health spending, work force, and access indices contributing less.

radiotherapy infrastructure may exert a disproportionately large impact on outcomes.[33] Conversely, for cancers where systemic therapy predominates (such as hematologic malignancies) or where early detection is critical (such as colorectal cancer), factors like pathology services availability and screening program implementation may emerge as more influential drivers of the mortality-to-incidence ratio.[34] Similarly, the impact of UHC expansion may differ between cancers with established screening programs (breast, cervical, colorectal, and lung) versus those without guideline-recommended screening. For instance, cervical cancer outcomes in low-resource settings are particularly sensitive to both screening access and treatment availability, while blood cancer outcomes may be less associated with changes in radiotherapy infrastructure.[35] Future work should employ cancer-specific models to identify which health system levers provide the greatest return on investment for particular malignancies, enabling even more targeted policy recommendations.

Implications for practice and policy

Our findings provide several actionable insights for clinicians and policymakers working to strengthen national cancer control. Firstly, radiotherapy infrastructure and UHC indices emerge consistently as high-yield investments, demonstrating robust associations with improved mortality-to-incidence ratios across a range of country contexts. Secondly, the data highlight that increased health spending, while necessary, is insufficient to drive outcomes unless coupled with strategic and efficient resource allocation; simply increasing budgets without targeting impactful interventions may deliver only marginal gains. Thirdly, the interpretability of country-specific SHAP profiles generated by our machine learning models offers new potential for directly informing national cancer control strategies. Such tools enable policymakers to identify and prioritize interventions tailored to their unique health system bottlenecks and strengths, moving beyond generic recommendations toward precision public health approaches. Ultimately, adoption of explainable artificial intelligence (AI) and country-level modeling can help focus investments in resource-constrained settings, ensuring that expansions in infrastructure, work force, and coverage yield tangible improvements in cancer survival.

Future research directions

While this study leverages country-level data to reveal actionable policy levers for cancer system strengthening, important opportunities remain for further investigation. Future work should focus on subnational analyses to dissect regional disparities, illuminate urban—rural divides, and guide more granular policymaking. Additionally, site-specific modeling is needed to determine which health system factors most strongly influence outcomes for individual cancer types, enabling highly targeted intervention design. Such efforts will require the continued development and harmonization of high-quality, regionally representative cancer registry and health system data to enable robust and context-specific modeling. Finally, the integration of cost-effectiveness analyses will be essential to ensure that recommended interventions not only improve outcomes but also maximize health system efficiency and sustainability. As such, the combination of explainable AI, high-resolution data, and economic

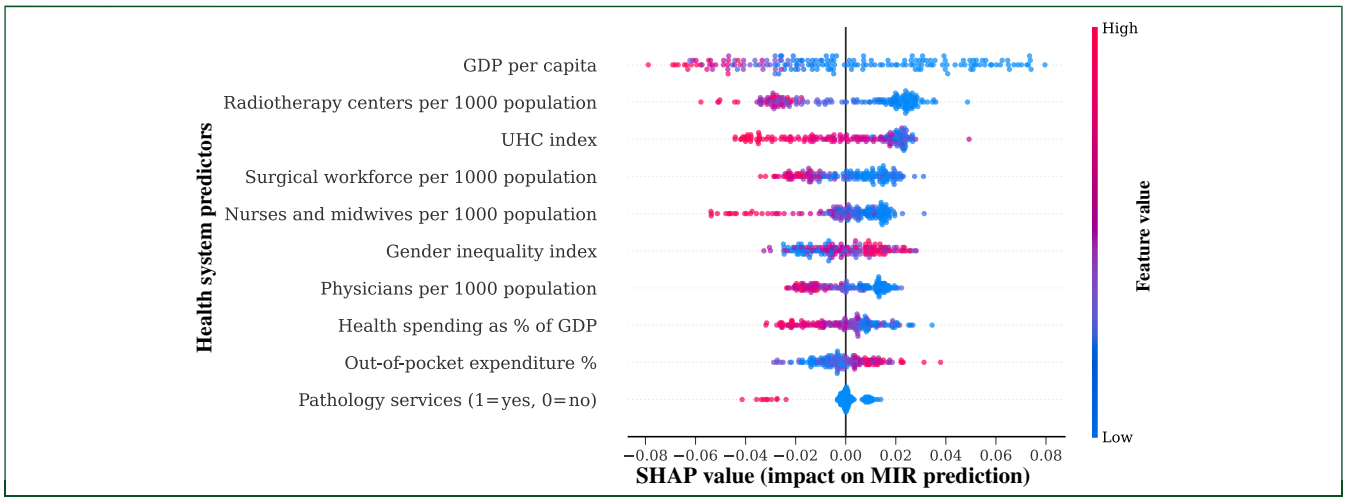

Figure 3. SHAP summary (beeswarm) plot showing the relative impact of health system predictors on country-level cancer mortality-to-incidence ratio (MIR) predictions. Each point is a country, with color encoding the feature value (blue: low, pink: high). Points to the right have positive SHAP values and push MIR higher; points to the left have negative SHAP values and lower MIR. Predictors are ordered by overall influence, with gross domestic product (GDP) per capita, universal health coverage (UHC) index, and radiotherapy centers per population exerting the largest effects. The distribution within each row depicts the magnitude, direction, and variability of each feature's contribution, revealing how health system characteristics shape national cancer burden estimates across 185 countries.

evaluation represents an essential next step to advance precision cancer policy and deepen the global evidence base.

Limitations

Important limitations must be considered when interpreting these findings. Firstly, the ecological study design relies on national-level aggregated data, which precludes inferences about individual-level relationships and is subject to ecological fallacy. MIR is affected by registry completeness, incidence ascertainment, and diagnostic infrastructure, and may not exclusively reflect cancer care effectiveness. Policy interpretation should consider these factors, especially in lower resource settings. SHAP values, while mathematically rigorous, represent model-based feature attributions rather than causal effects. These attributions quantify each variable's contribution to predictions within our model framework but cannot establish whether intervening in these factors would produce the predicted mortality reductions. Country-specific interventional studies or quasi-experimental designs are needed to validate causal relationships before implementing large-scale policy changes based solely on these associations. Additionally, unmeasured confounders at the national level, temporal dynamics, and implementation contexts may substantially modify the actual impact of health system interventions. Therefore, our findings should be interpreted as generating hypotheses and identifying priority areas for further investigation rather than definitive evidence for specific policy interventions.

The ecological design introduces several specific sources of potential bias that warrant careful consideration. Unmeasured confounders at the national level—including cultural attitudes toward health care-seeking behavior, governance quality indicators such as corruption indices, and political stability—may significantly influence the model's SHAP attributions and could bias interpretations of health system impacts. As such, our findings should inform the prioritization of pilot interventions aimed at validating high-impact drivers identified by the model. Rigorous evaluation of such interventions will be critical for confirming their effectiveness and ensuring that resource allocation decisions are evidence-based. For instance, countries with similar UHC indices but divergent governance efficiency or institutional capacity could demonstrate markedly different cancer outcomes, which the model may attribute entirely to measured variables rather than these unmeasured contextual factors. Additionally, subnational disparities in health care access, infrastructure distribution, and cancer registry coverage within countries are not captured in our national-level analysis, potentially masking substantial within-country heterogeneity that could be as large or larger than between-country differences. Data quality represents another critical limitation, particularly regarding GLOBOCAN estimates in low- and middle-income countries where cancer registry systems may be incomplete, leading to potential under-reporting of both incidence and mortality. Registry completeness varies substantially across countries, with some nations relying on modeled estimates rather than population-based data, which may introduce systematic biases in MIR calculations. Future studies should incorporate sensitivity analyses that stratify results by data quality metrics, exclude countries with low registry completeness scores, or weight observations by data quality indicators to assess the robustness of findings to these measurement limitations.

Future work should employ causal inference methods, such as difference-in-differences analyses of policy implementations, instrumental variable approaches, or propensity score methods applied to longitudinal data, to move beyond these ecological associations toward causal understanding of health system interventions on cancer mortality.

Conclusions

This machine learning approach with explainable AI transforms global health system research from descriptive to prescriptive, enabling evidence-based, country-specific cancer system strengthening. As the global cancer burden continues to grow, particularly in low- and middle-income countries, such personalized policy guidance becomes increasingly critical for optimizing limited resources and maximizing population health impact.

FUNDING

This work was supported in part through the Cancer Center Support Grant from the National Cancer Institute [grant number P30 CA008748] to MSP, PI, TPK, JW, NYL and ECD; the Swiss National Science Foundation/Schweizerischer Nationalfonds zur Förderung der wissenschaftlichen Forschung [grant number P500PM_203194] to JW; the National, Heart, Lung, and Blood Institute [grant number R38HL167238] to JFW; and in part through the Prostate Cancer Foundation Young Investigator Award (no grant number) to ECD.

DISCLOSURE

PI holds primary appointment at Memorial Sloan Kettering Cancer Center and secondary appointment at University of Texas Southwestern Medical Center. PI serves on advisory boards for Pfizer, AstraZeneca, BioConvergent, Johnson and Johnson, Novocure, and NGM Biosciences. PI's institution receives funding for clinical trials from Incyte, AstraZeneca, NGM Biosciences, and Summit Pharma. PI declares no conflicts of interest relative to this paper. TPK receives honoraria from Olympus Medical Systems. TPK serves in a consulting or advisory role for Physicians' Education Resource. NYL serves as consultant on advisory boards for Merck, Merck Serono, Nanobiotix, Galera, LEO SAB, and Johnson & Johnson. NYL serves on the steering committee for AstraZeneca's eVOLVE trial. NYL holds stock options in LEO SAB. The rest of the authors have no conflicts to declare.

REFERENCES

1. Pramesh CS, Badwe RA, Bhoo-Pathy N, et al. Priorities for cancer research in low- and middle-income countries: a global perspective. *Nat Med*. 2022;28(4):649-657.
2. GBD 2023 Cancer Collaborators. The global, regional, and national burden of cancer, 1990-2023, with forecasts to 2050: a systematic analysis for the Global Burden of Disease Study 2023. *Lancet*. 2025;406(10512):1565-1586.
3. Dee EC, Wu JF, Feliciano EJG, et al. National cancer system characteristics and global pan-cancer outcomes. *JAMA Oncol*. 2025;11(6):650-654.
4. Lundberg SM, Lee S-I. A unified approach to interpreting model predictions. In: Guyon I, Luxburg UV, Bengio S, et al., editors. *Advances in Neural Information Processing Systems*. Long Beach USA: Curran Associates, Inc. Available at https://proceedings.neurips.cc/paper_files/paper/2017/file/8a20a8621978632d76c43dfd28b67767-Paper.pdf; 2017.
5. Bray F, Laversanne M, Sung H, et al. Global cancer statistics 2022: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA Cancer J Clin*. 2024;74(3):229-263.
6. Toumaj S, Heidari A, Jafari Navimipour N. Leveraging explainable artificial intelligence for transparent and trustworthy cancer detection systems. *Artif Intell Med*. 2025;169:103243.
7. Lundberg SM, Erion G, Chen H, et al. From local explanations to global understanding with explainable AI for trees. *Nat Mach Intell*. 2020;2(1):56-67.
8. Pedregosa F, Varoquaux G, Gramfort A, et al. Scikit-learn: machine learning in Python. *J Mach Learn Res*. 2011;12(85):2825-2830.
9. Prokhorenkova L, Gusev G, Vorobev A, Dorogush AV, Gulin A. CatBoost: unbiased boosting with categorical features. *arXiv*. 2017. https://doi.org/10.48550/arXiv.1706.09516.
10. Akiba T, Sano S, Yanase T, Ohta T, Koyama M. Optuna: a next-generation hyperparameter optimization framework. In: *Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*. Anchorage, USA: ACM; 2019. p. 2623-2631.
11. Seabold S, Perktold J. Statsmodels: econometric and statistical modeling with Python. In: *9th Python in Science Conference*. Austin; 2010.
12. *The joblib developers. joblib*. Zenodo. Available at https://zenodo.org/doi/10.5281/zenodo.14915601. Accessed August 10, 2025.
13. da Costa-Luis C, Larroque SK, Altendorf K, et al. *tqdm: a fast, extensible progress bar for Python and CLI*. Zenodo. Available at https://zenodo.org/doi/10.5281/zenodo.595120. Accessed August 11, 2025.
14. *The pandas development team*. pandas-dev/pandas: Pandas. Zenodo. Available at https://zenodo.org/doi/10.5281/zenodo.3509134. Accessed August 11, 2025.
15. Harris CR, Millman KJ, Van Der Walt SJ, et al. Array programming with NumPy. *Nature*. 2020;585(7825):357-362.
16. Hunter JD. Matplotlib: a 2D graphics environment. *Comput Sci Eng*. 2007;9(3):90-95.
17. Waskom M. seaborn: statistical data visualization. *J Open Source Softw*. 2021;6(60):3021.
18. Virtanen P, Gommers R, Oliphant TE, et al. SciPy 1.0: fundamental algorithms for scientific computing in Python. *Nat Methods*. 2020;17(3):261-272.
19. Available at https://cancersystemsai.org/. Accessed December 1, 2025.
20. Goksel F, Koc O, Ozgul N, et al. Radiation oncology facilities in Turkey: current status and future perspectives. *Asian Pac J Cancer Prev*. 2011;12(9):2157-2162.
21. Ozmen V, Dagoglu N, Dede I, et al. Turkish Ministry of Health, 2nd Turkish Medical General Assembly Clinical Oncology Study Group Report. *J Breast Health*. 2016;12(1):9-17.
22. de Fonseca BP, Albuquerque PC, de Saldanha RF, Zicker F. Geographic accessibility to cancer treatment in Brazil: a network analysis. *Lancet Reg Health Am*. 2022;7:100153.
23. De Camargo Cancela M, Bezerra De Souza DL, et al. Can the sustainable development goals for cancer be met in Brazil? A population-based study. *Front Oncol*. 2023;12:1060608.
24. Didkowska J, Wojciechowska U, Michalek IM, Caetano dos Santos FL. Cancer incidence and mortality in Poland in 2019. *Sci Rep*. 2022;12(1):10875.
25. Iwamoto M, Nakamura F, Higashi T. Monitoring and evaluating the quality of cancer care in Japan using administrative claims data. *Cancer Sci*. 2016;107(1):68-75.
26. Tamaki T, Morioka N, Machida A, Kashiwagi M. Trends in the geographic inequality of advanced practice nursing workforce in cancer care in Japan from 1996 to 2022: a panel data analysis. *Hum Resour Health*. 2024;22(1):33.
27. Aoyama T, Koide Y, Shimizu H, et al. Trends and distribution of external radiation therapy facilities in Japan based on Survey of Medical Institutions from the Ministry of Health, Labour and Welfare. *J Radiat Res*. 2024;65(3):328-336.
28. Aminuddin F, Mohd Nor Sham Kunusagaran MSJ, Raman S, et al. The economic toll of cancer: catastrophic expenditure and impoverishment among lower-income households in Malaysia. *BMC Public Health*. 2025;25(1):2216.
29. Fnu N, Kuan W-C, Kong Y-C, et al. Cancer-related costs, the resulting financial impact and coping strategies among cancer survivors living in

a setting with a pluralistic health system: a qualitative study. *Ecancermedicalscience*. 2022;16:1449.

30. Tuck CZ, Akparibo R, Gray LA, Aryeetey RNO, Cooper R. What influences cancer treatment service access in Ghana? A critical interpretive synthesis. *BMJ Open*. 2022;12(10):e065153.
31. Insamran W, Sangrajrang S. National cancer control program of Thailand. *Asian Pac J Cancer Prev*. 2020;21(3):577-582.
32. Wu W, Zhang R, Jin Y, et al. Cancer trends and risk factors in China over the past 30 years (1990-2019). *J Cancer*. 2023;14(10):1935-1945.
33. Kyaw JYA, Rendall A, Gillespie EF, et al. Systematic review and meta-analysis of the association between radiation therapy treatment volume and patient outcomes. *Int J Radiat Oncol Biol Phys*. 2023;117(5):1063-1086.
34. Global Burden of Disease Cancer Collaboration, Fitzmaurice C, Akinyemiju TF, Al Lami FH, et al. Global, regional, and national cancer incidence, mortality, years of life lost, years lived with disability, and disability-adjusted life-years for 29 cancer groups, 1990 to 2016: a systematic analysis for the global burden of disease study. *JAMA Oncol*. 2018;4(11):1553.
35. Frick C, Rumgay H, Vignat J, et al. Quantitative estimates of preventable and treatable deaths from 36 cancers worldwide: a population-based study. *Lancet Glob Health*. 2023;11(11):e1700-e1712.