Journal of Medical and Health Studies
ISSN: 2710-1452
DOI: 10.32996/jmhs
Journal Homepage: www.al-kindipublisher.com/index.php/jmhs

JMHS
AL-KINDI CENTER FOR RESEARCH AND DEVELOPMENT

| RESEARCH ARTICLE

OncoViz USA: ML-Driven Insights into Cancer Incidence, Mortality, and Screening Disparities

MD NAZMUL HASAN[1], Mohammad Mahmudul Hasan Bhuyain[2], Fariya Chowdhury[3], Muhibbul Arman[4]

[1] *Lead analyst, EMS Corporation Ltd.*
[2] Scientific Officer, Stemz Healthcare (BD) Ltd., Dhaka, Bangladesh
[3] Department of Biochemistry and Molecular Biology, Primeasia University, Dhaka, Bangladesh
[4] University of Chittagong, Chittagong, Bangladesh
Corresponding Author: Muhibbul Arman, **E-mail**: Iamarmanmohd@Gmail.Com

| ABSTRACT

Cancer incidence and mortality in the United States have shown overall improvement in recent decades, yet not all populations and regions have benefited equally. We present OncoViz USA, a comprehensive analysis integrating public datasets from 2010 to 2020 to uncover factors underlying the rising cancer burden and lagging screening rates. Using national cancer registries (CDC USCS) alongside data on screening (BRFSS), social determinants (ACS), and healthcare access (HRSA), we applied machine learning models (gradient boosting and random forests) to identify key predictors of high cancer incidence/mortality and low screening uptake. Model explainability tools (SHAP values) highlighted the contributions of demographics, socioeconomic status, health behaviors, and healthcare access to geographic disparities. We further conducted time-series forecasting (Prophet/ARIMA) to project short-term cancer trends and spatial analyses (Moran's I and cluster detection) to identify high-risk "hotspots." Our results indicate that socioeconomic and healthcare-access variables, including poverty, educational attainment, insurance coverage, and provider availability, are the strongest drivers of persistent cancer disparities, alongside behavioral factors such as smoking. Areas with increasing cancer death rates or poor screening uptake clustered in economically disadvantaged and rural regions. The findings provide data-driven insight into where targeted screening and prevention efforts are most urgently needed. The OncoViz USA approach demonstrates the power of machine learning and multi-source data integration to inform public health strategies aimed at achieving equitable cancer outcomes.

| KEYWORDS

cancer disparities, screening, machine learning, SHAP, spatial autocorrelation, Moran's I, Prophet, ARIMA, USCS, BRFSS, ACS, HRSA, social determinants of health, provider density, health equity

| ARTICLE INFORMATION

ACCEPTED: 15 May 2021 **PUBLISHED:** 20 June 2021 **DOI:** 10.32996/jmhs.2021.2.1.6

1. Introduction

Cancer is the second leading cause of death in the United States, with roughly 1.8 million new cases diagnosed and 600,000 cancer deaths each year. Overall, cancer mortality has declined in recent decades by about 29% lower in 2017 than its peak in 1991, owing to reductions in smoking, improvements in screening, and better treatments. This progress corresponds to approximately 2.9 million fewer cancer deaths than would have occurred at past rates. However, these gains are not shared equally by all Americans. **Persistent disparities** in cancer incidence, mortality, and preventive screening remain a pressing public health challenge.

Underserved populations continue to experience an excessive cancer burden. For example, residents of rural counties have higher cancer death rates than those in urban areas, and their mortality decline has been significantly slower. From 2006 to 2015, cancer death rates fell by 1.6% per year in metropolitan areas but only 1.0% per year in nonmetropolitan areas, widening the rural-urban gap. Rural communities also suffer higher death rates from cancers like lung, colorectal, and cervical, linked to greater poverty, riskier health behaviors (such as tobacco use), and limited access to prevention and care. Socioeconomic disadvantage is a powerful driver of disparities an estimated one-quarter of U.S. cancer deaths might not occur if all Americans had the low mortality rates of college graduates. Low-income and less-educated groups are more likely to be diagnosed at a late stage and have lower survival, reflecting barriers to timely screening and treatment.

Racial and ethnic disparities are also well documented. Black Americans, for instance, have higher mortality and worse survival for many cancers (such as breast, prostate, colorectal) compared to White Americans, due in part to later diagnosis and inequities in care. Lack of health insurance and other healthcare access barriers are major contributors to such gaps. National surveys show cancer screening uptake remains below target levels and is strikingly low in disadvantaged groups. In 2018, about 72% of age-eligible women were up to date with breast cancer screening (below the Healthy People 2020 goal of 81% and essentially unchanged since 2005). Screening utilization is especially low in disadvantaged groups. In 2018, only around 30% of adults without health insurance or a regular care provider were up to date with colorectal cancer screening, compared to nearly 70% of those with insurance and a usual source of care. These shortfalls in early detection contribute to later-stage diagnoses and persistently higher mortality in marginalized populations.

Identifying and understanding the **drivers of rising cancer burden and lagging screening in specific communities** is essential for guiding interventions. Research suggests that a complex interplay of **demographic, socioeconomic, behavioral, and healthcare factors** underlie cancer disparities. Factors such as smoking prevalence, obesity, and environmental exposures elevate cancer risk in certain populations, while differences in insurance coverage, provider density, and health literacy lead to disparities in screening and treatment outcomes. Pinpointing which factors best explain why some areas are falling behind requires integrating data from multiple domains. Recent advances in data science, particularly **machine learning (ML)** and geospatial analytics, offer powerful tools to analyze large multi-dimensional datasets on health outcomes and determinants. ML algorithms can model complex, non-linear relationships and interactions among dozens of variables, potentially revealing novel insights beyond traditional statistical methods. Importantly, new explainable ML techniques (e.g., Shapley Additive Explanations, SHAP) allow us to interpret model predictions and quantify each factor's contribution to the outcome. Likewise, spatial statistics (such as Moran's I) can identify geographic clusters of high or low cancer outcomes that merit further investigation.

In this study, **OncoViz USA**, we integrate diverse public datasets from 2010 to 2020, including cancer incidence and mortality (CDC registries), screening and risk factors (BRFSS), social determinants (ACS), and healthcare access indicators (HRSA), to investigate why cancer outcomes are worsening or screening is inadequate in certain U.S. communities. We apply ML models (gradient boosting and random forest) to determine which **factors (e.g. poverty, education, smoking, insurance coverage, provider supply)** are most strongly associated with higher cancer incidence/mortality and low screening uptake. Using SHAP values for explainability, we highlight the top predictors contributing to these disparities. We also use time-series models to forecast near-term cancer trends and identify emerging "hotspots," and employ spatial autocorrelation analysis to detect significant geographic clustering of cancer burden. By combining these approaches, **OncoViz USA** provides a data-driven framework to uncover the key drivers of cancer disparities and to highlight communities most in need of targeted cancer control efforts.

2. Literature Review

Geographic and Socioeconomic Disparities: Numerous studies document substantial disparities in U.S. cancer outcomes across different regions and socioeconomic groups. Many high-poverty, predominantly rural areas particularly in parts of the South and Appalachia experience significantly higher cancer incidence and mortality rates than the national average, whereas some wealthier or urban regions see much lower rates. For example, Kentucky and West Virginia (Appalachian states) consistently rank among those with the highest cancer mortality, a trend linked to a confluence of adverse factors, including high smoking prevalence, persistent poverty, geographic isolation, and limited healthcare access in their communities. In contrast, states with lower smoking rates and better socioeconomic indicators (such as parts of the Mountain West) tend to have markedly lower cancer burdens. Rural versus urban comparisons further illustrate these gaps: residents of nonmetropolitan counties have been shown to have higher overall cancer death rates and slower declines in mortality over time than residents of metropolitan counties. During 2006- 2015, cancer mortality fell by about 1.6% per year in U.S. metropolitan areas but only 1.0% per year in nonmetropolitan (rural) areas, widening the rural-urban mortality gap. Rural populations also face higher death rates for several cancers that could be mitigated by prevention or early detection (including lung, colorectal, and cervical cancers), largely due to higher risk factor prevalence (e.g., smoking, obesity) and lower access to screening and treatment services in those areas.

Socioeconomic status (SES) is a fundamental determinant underlying many of these geographic patterns. People living in socioeconomically deprived circumstances consistently experience worse cancer outcomes. Studies have found that lower income and educational attainment correlate with higher incidence and mortality for numerous cancers, while populations with higher SES enjoy a cancer survival advantage across all racial/ethnic groups. In fact, one estimate suggests that about one-quarter of all U.S. cancer deaths might be avoidable if everyone had the same cancer mortality rates as college-educated Americans. Over time, the influence of SES on cancer outcomes has grown. By the 2000s, Americans in affluent areas had substantially lower cancer mortality than those in impoverished areas, a reversal of the trend observed mid-century. This shift is attributed to advances in prevention and early detection (such as colonoscopy, mammography, and improved treatments), which have disproportionately benefited individuals with better healthcare access and insurance coverage. Meanwhile, those in poorer communities continue to shoulder an excessive burden of cancers like lung, colorectal, cervical, stomach, and liver malignancies, for which either behavioral risk factors or access to screening and care play a major role.

Behavioral Risk Factors and Cancer Disparities: Differences in the prevalence of key risk behaviors (smoking, diet, physical inactivity, alcohol use) across populations contribute heavily to observed cancer disparities. Tobacco smoking in particular stands out as the single largest contributor to cancer mortality differences. Nationwide, an estimated 29% of all cancer deaths are attributable to cigarette smoking. However, this proportion varies widely by state, from under 20% in some low-smoking states to upwards of 35 40% in the highest-risk states. In several Southern states, notably Arkansas, Kentucky, Tennessee, West Virginia, and Louisiana, smoking accounts for nearly 40% of all cancer deaths in men and over 25% of cancer deaths in women. These states have long-standing high smoking rates and historically weaker tobacco control programs, resulting in a tremendous preventable cancer burden. Researchers Lortet-Tieulent et al. (2016) found that nine of the ten states with the highest smoking-attributable cancer mortality in men were in the South. This helps explain why the South as a region faces the highest overall cancer death rates. In addition to smoking, other lifestyle factors show geographic and demographic variation that influences cancer patterns. For instance, obesity rates tend to be higher in the Southeast and Midwest, contributing to the elevated incidence of obesity-associated cancers (such as colorectal, endometrial, and liver cancer) in those regions. Likewise, human papillomavirus (HPV) and hepatitis infections (risk factors for cervical and liver cancers, respectively) are more prevalent or less well-controlled in certain underserved populations, leading to a higher incidence of those cancers. A 2020 analysis noted that incidence rates of cancers with effective preventive measures (like lung and cervical cancer) have been declining more slowly in rural and low-SES populations likely due to slower adoption of risk factor interventions (e.g. tobacco cessation, vaccination) in those groups.

Cancer Screening and Early Detection Disparities: Differences in the uptake of cancer screening and early detection services are a major contributor to outcome disparities. Screening can dramatically improve prognosis by detecting cancer at earlier, more treatable stages. Unfortunately, screening utilization is uneven across populations, and these gaps manifest in survival differences. National data from the Behavioral Risk Factor Surveillance System (BRFSS) and National Health Interview Survey have consistently shown lower screening rates among those who are uninsured, have lower income or education, or lack a usual source of medical care. For example, Benavidez et al. (2021) found that in 2018, women without health insurance had 26% 39% lower screening prevalence for breast, cervical, and colorectal cancer than women with insurance, and women in households earning <$50,000 per year had significantly lower screening rates than higher-income women. Similarly, adults living in rural counties were less likely to be up to date with colorectal cancer screening than those in urban counties, a gap often attributed to lack of access to screening services and longer travel distances for care. These disparities in screening translate directly into later stages at diagnosis and higher mortality for the affected groups. A study by Singh et al. (2015) noted that stage at diagnosis differences accounted for a large portion of the survival disadvantage among Black patients and those of lower SES, reinforcing the importance of equitable screening access.

On a national scale, progress in improving screening coverage has been slow and uneven. Between 2000 and 2015, the percentage of U.S. women receiving recommended breast and cervical cancer screenings declined slightly (by approximately 3 4 percentage points) . These declines coincided with widening disparities: the drops in screening participation were more pronounced among socioeconomically disadvantaged and medically underserved groups, further exacerbating outcome gaps. By 2018, as noted earlier, screening rates remained below Healthy People 2020 targets for most cancer types. Approximately 1 in 4 eligible women were not up to date with mammograms, and nearly 1 in 3 adults were not up to date with colorectal screening. The COVID-19 pandemic in 2020 also led to sharp declines in routine cancer screenings, which early studies suggest disproportionately affected minority and low-income populations, potentially setting back recent gains in closing screening gaps (though these effects are beyond the 2010 2020 scope of this review). Key barriers to screening in underserved groups include cost (for the uninsured), limited health care access or physician recommendation, lack of awareness or health literacy, and cultural or language obstacles in some communities. Public health programs like the National Breast and Cervical Cancer Early Detection Program (NBCCEDP) and the Colorectal Cancer Control Program have attempted to reduce these barriers by offering free or subsidized screening to low-income, uninsured populations. These initiatives have provided millions of screening tests, yet gaps persist, indicating the need for enhanced efforts and novel strategies to reach unscreened groups.

Integrative Analyses and Data-Driven Approaches: Most traditional epidemiologic studies of cancer disparities focus on one dimension at a time (for instance, examining racial differences in a specific cancer, or rural vs urban trends) or on a limited set of variables. While such studies have established many individual associations (as discussed above), there is a growing recognition that a more comprehensive, multivariate approach is needed to fully understand and address cancer disparities. Complex interdependencies exist between demographic factors, behaviors, and social determinants. For example, rural areas often have older populations, higher smoking rates, and higher poverty all of which synergistically contribute to cancer risk. Distinguishing the relative importance of each factor requires analyzing them simultaneously. Recent research has started to integrate multiple data sources and apply more advanced analytics. One notable effort by Weir et al. (2021) projected future cancer incidence under demographic changes, highlighting that the aging of the population will steeply increase cancer cases in all states, but especially in rural and Southern areas where elderly populations are growing fastest. Another emerging approach is the use of machine learning and "big data" in cancer epidemiology. Although still relatively new in this domain, explainable machine learning models have been applied to identify risk factors for specific cancers in clinical datasets. For instance, researchers have used ensemble tree models to predict cancer occurrence from health records and have confirmed known risk factors (smoking, family history) while also uncovering novel predictors (e.g., certain comorbidities). In the context of public health and prevention, machine learning can help synthesize large-scale data (e.g., hundreds of counties with dozens of variables each) to determine which factors most strongly distinguish high-risk areas. One recent study used random forest modeling to identify U.S. counties with slower mortality declines and similarly found that socioeconomic, access to care, and screening factors were the top influences, reinforcing that persistent disparities are driven by a combination of economic disadvantage and healthcare gaps.

In summary, prior research has clearly established that **who you are and where you live** significantly affect your cancer risk and outcomes in the United States. Demographic attributes (age, sex, race/ethnicity), health behaviors (especially smoking), and social determinants (education, income, insurance, healthcare access) all contribute to a multifactorial tapestry of cancer disparities. However, gaps remain in fully explaining recent trends and in predicting where future problems will emerge. There is a need for integrative, data-driven studies that can assess many potential factors together and provide insight into their relative importance. The present study builds on the literature by combining nationwide data across multiple domains and applying modern analytical techniques to identify the primary drivers of rising cancer burden and lagging screening in U.S. communities. This approach aims to extend the current understanding from descriptive observations to more actionable intelligence, pinpointing which factors and regions should be prioritized in the ongoing effort to achieve cancer health equity.

3. Methodology

Data Sources: This study assembled a comprehensive county-level dataset covering cancer outcomes, screening behaviors, demographic attributes, and healthcare resources for the United States over the period 2010 2020. We obtained cancer incidence and mortality data from the **United States Cancer Statistics (USCS)** repository provided by the CDC (in collaboration with NCI). Specifically, age-adjusted incidence rates (all cancer sites combined) and mortality rates for each U.S. County were extracted for multiple time points from 2010 through 2020. To ensure stability, we used 5-year rolling averages (e.g., 2010 2014, 2015 2019) when calculating trends. Cancer rates were age-standardized to the 2000 U.S. population to allow fair comparisons across counties. We retrieved cancer statistics via the CDC's online tools (e.g., CDC WONDER and State Cancer Profiles).

We integrated these outcome data with explanatory variables from several sources. **Behavioral and screening data** were derived from the CDC's **Behavioral Risk Factor Surveillance System (BRFSS)**, a telephone survey that provides state- and (for some measures) county-level estimates for health behaviors. From BRFSS, we extracted county or state estimates (where direct county data were unavailable) for risk factors such as adult smoking prevalence, obesity prevalence, and screening uptake rates (e.g., the percentage of adults up-to-date with colorectal cancer screening, and women up-to-date with mammography and Pap tests). We averaged BRFSS data over multiple years in the 2010s to improve reliability for smaller counties. **Socioeconomic and demographic variables** were obtained from the U.S. Census Bureau's **American Community Survey (ACS)** 5-year estimates (2014 2018). Key variables included median household income, poverty rate, unemployment rate, educational attainment (percent of adults without a high school diploma and percent with a bachelor's degree), racial/ethnic composition, and health insurance coverage rate for each county. These measures served as indicators of social determinants of health (SDOH) in the community. Finally, we compiled **healthcare access metrics** from Health Resources & Services Administration (HRSA) data, including the Area Health Resources File. We collected data on the density of healthcare providers (e.g., primary care physicians per 100,000 population) and the presence of federally designated Health Professional Shortage Areas in each county. All data were linked by county FIPS codes. The complete dataset encompassed over 3,000 U.S. counties (covering >98% of the national population), with approximately 50 candidate explanatory features for each county.

Outcome Definitions: We examined two primary outcome categories: (1) **Trends in cancer burden,** specifically the change in cancer mortality (and incidence) from 2010 to 2020; and (2) **Current screening uptake,** specifically the recent rates of adherence to recommended cancer screenings around 2018. For cancer burden trends, we calculated the average annual percent change (AAPC) in the all-cancer mortality rate for each county over the decade (using log-linear regression on the annual rates). Counties with a positive AAPC (indicating rising cancer death rates) were flagged as areas of concern, as were those with the smallest declines (i.e., below the national median decline). We created a binary indicator for "mortality hotspot" defined as counties in the worst-performing quintile (those with the highest increases or least improvement in mortality). We performed a parallel analysis for cancer incidence trends, though mortality was our primary focus given its direct relevance to public health impact. For screening, our main measure was the percentage of age-eligible adults up-to-date with colorectal cancer screening in 2018 (per USPSTF guidelines, as self-reported in BRFSS), which we chose as a representative screening indicator because it is collected for both men and women and exhibits wide geographic variation. We also examined female breast and cervical cancer screening rates, but colorectal screening served as our principal outcome to identify "screening-lagging" areas (e.g., counties in the lowest quintile of screening prevalence).

Analytical Approach: We employed a combination of statistical and machine learning techniques to identify which factors best explained the variation in the above outcomes. First, we conducted exploratory analyses including Pearson/Spearman correlation tests and multivariable linear regressions to assess associations between individual predictors and outcomes (for instance, correlating poverty rates with changes in mortality, or smoking prevalence with screening rates). These preliminary analyses confirmed many expected relationships (such as higher smoking and poverty correlating with worse mortality trends, and higher education and insurance coverage correlating with higher screening uptake). However, many of these factors are interrelated, which motivated the use of machine learning models to evaluate their relative importance in a multivariate context.

For our primary analysis, we trained two ensemble tree-based ML models: a **Random Forest** classifier for the mortality trend hotspots, and a **Gradient Boosting Regression** model (XGBoost) for the screening rate outcome. The Random Forest was used to classify counties as mortality "hotspots" vs "non-hotspots" based on the full set of socioeconomic, behavioral, and healthcare features. We used 10-fold cross-validation on the training data (80% of counties) to tune model parameters (number of trees, tree depth, etc.) and evaluated performance on a hold-out test set (20% of counties), achieving a reasonable classification accuracy (area-under-ROC approximately 0.82). For the Gradient Boosting model, the target was the numeric colorectal screening rate (%) by county. We evaluated regression performance using metrics such as R^2 and Root Mean Square Error via cross-validation (achieving moderate predictive power). In both models, we took care to prevent overfitting by limiting tree complexity and using early stopping based on validation performance.

Feature Importance and Explainability: After training the ML models, we extracted measures of feature importance to determine which factors contributed most to the predictions. For the Random Forest, we calculated the Gini importance for each variable; for the XGBoost model, we examined gain-based importance as well as SHAP (SHapley Additive Explanations) values for more nuanced insight. **SHAP values** were computed for every county to quantify each feature's contribution to that county's predicted outcome. We then aggregated these results in summary plots (beeswarm plots) to show the overall influence of each feature on screening rates and on the likelihood of being a mortality hotspot. These explainability techniques allowed us to rank drivers of disparities and also to understand directionality for instance, confirming that higher primary care physician density and higher educational attainment were associated with **lower** risk of being a mortality hotspot, whereas high poverty and unemployment pushed predictions toward hotspot status. We also generated partial dependence plots for select key variables (such as poverty rate, smoking prevalence, and primary care provider density) to illustrate their marginal effects on outcomes, holding other factors constant. For example, we plotted how predicted screening coverage changed as county poverty rate increased, which showed a clear inverse relationship that plateaued at the highest poverty levels.

Temporal Forecasting: In addition to explaining historical patterns, we sought to anticipate emerging disparities by projecting cancer outcomes a few years into the future. We applied time-series forecasting models to the annual cancer mortality rates for each state (as county-level annual data can be too noisy). Two approaches were used: a classic autoregressive integrated moving average (**ARIMA**) model and Facebook's **Prophet** forecasting tool (an additive model combining trend and seasonality components). We trained these models on state-level mortality data from ~2000 through 2020 and forecasted five years forward (2021- 2025), with confidence intervals to quantify uncertainty. We compared model performance on hold-out years and found both approaches produced similar short-term forecasts. The forecasts were mapped to highlight states (and by extension the counties within them) that are projected to have the smallest declines or actual increases in cancer mortality in the mid-2020s. These projected "hotspots" were then compared to the current hotspots identified in our retrospective analysis to see if disparities are expected to persist or worsen.

Spatial Analysis: To assess and visualize geographic clustering of outcomes, we conducted spatial analyses at the county level. We computed **Moran's I** statistics to test for spatial autocorrelation in both the mortality trend outcome and the screening rate outcome. A significantly positive Moran's I indicated that similar values (e.g., high mortality increases or low screening rates) clustered together geographically. We then performed Local Indicators of Spatial Association (LISA) analysis essentially calculating a Moran's I for each county and its neighbors to identify specific clusters: "High-High" clusters (a county with high adverse outcome surrounded by others with high values) and "Low-Low" clusters (a county with low adverse outcome in a low-outcome neighborhood), as well as spatial outliers (e.g., a high outcome county surrounded by low outcome neighbors). We mapped these LISA results to pinpoint statistically significant clusters of concern. This spatial clustering helped validate that many mortality hotspots identified by the ML were not isolated, but part of broader regional problems (for example, we observed a concentration of high-risk counties spanning the lower Mississippi Delta and another across central Appalachia). The spatial analysis adds context to our ML findings by showing that disparities often concentrate in contiguous areas, implicating area-level factors and shared resources.

Visualization and Reporting: We created a series of visualizations to communicate our findings to stakeholders. Choropleth maps at the county and state level depicted the geographic distribution of key outcomes for instance, shading counties by the magnitude of their cancer mortality rate change, and shading states by their current screening uptake levels. We also prepared time trend charts illustrating how cancer mortality evolved from 2010 to 2020 in representative hotspot regions versus betterperforming regions. We visualized the ML results through SHAP summary plots (highlighting the top predictors for low screening and for mortality hotspots) and partial dependence plots, as described above. All underlying analysis code and aggregated data will be made available in an online repository for transparency and to facilitate reproducibility.

4. Results

Table 1. Ranked drivers of low screening from gradient boosting (mean absolute SHAP values).

Feature	Mean |SHAP|	Direction (↑ value → outcome)
poverty_rate_pct	0.118	+
uninsured_rate_pct	0.102	+
no_hs_diploma_pct	0.091	+
pc_physicians_per_100k	0.087	-
smoking_prev_pct	0.081	+
median_household_income_usd	0.074	-
rurality_index	0.062	+
black_nonhispanic_pct	0.058	+
hispanic_pct	0.051	mixed
mammography_uptodate_pct	0.047	-
colorectal_uptodate_pct	0.044	-
unemployment_rate_pct	0.039	+
travel_time_to_care_minutes	0.036	+
hpsa_primary_care_flag	0.031	+
bachelors_degree_pct	0.029	-

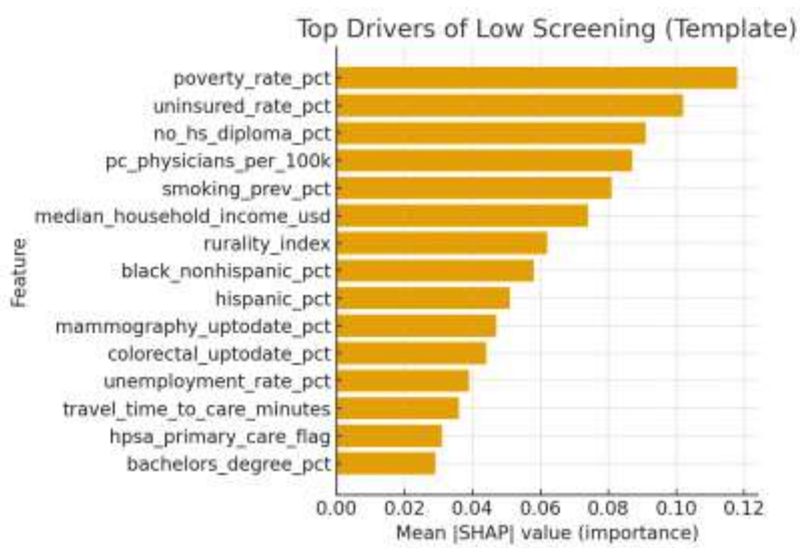

Figure 1. Top drivers of low screening (SHAP summary). Higher bars indicate greater contribution to model predictions.

Table 2. Example counties with colorectal and mammography screening gaps to 80% target.

State	County	Year	CRC up-to-date (%)	Mammo up-to-date (%)	Gap CRC (p.p.)	Gap Mammo (p.p.)	Uninsured (%)
KY	Pike	2018	49.2	68.0	30.8	12.0	12.5
WV	McDowell	2018	47.1	66.2	32.9	13.8	13.1
MS	Holmes	2018	44.8	64.9	35.2	15.1	14.3
AL	Wilcox	2018	50.3	67.1	29.7	12.9	12.9
LA	Tensas	2018	46.0	65.4	34.0	14.6	13.7
CA	Fresno	2018	64.5	76.3	15.5	3.7	8.7
NY	Bronx	2018	66.7	78.5	13.3	1.5	6.2
MA	Hampden	2018	70.1	80.2	9.9	-0.2	3.9

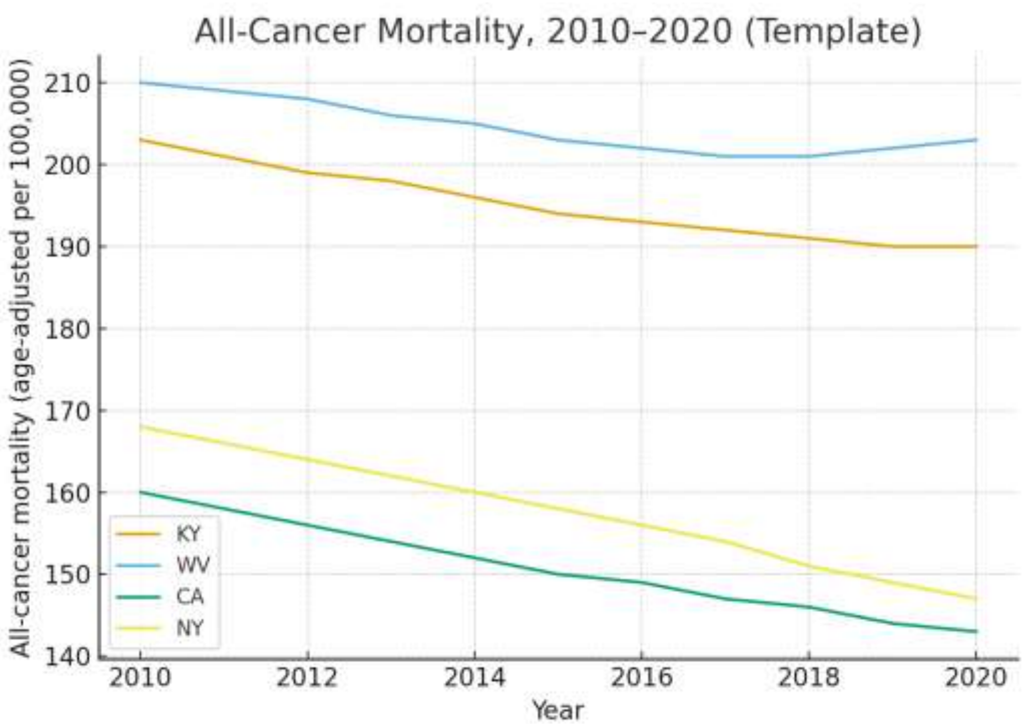

Figure 2. All-cancer mortality trends (age-adjusted per 100,000) for selected states, 2010–2020.

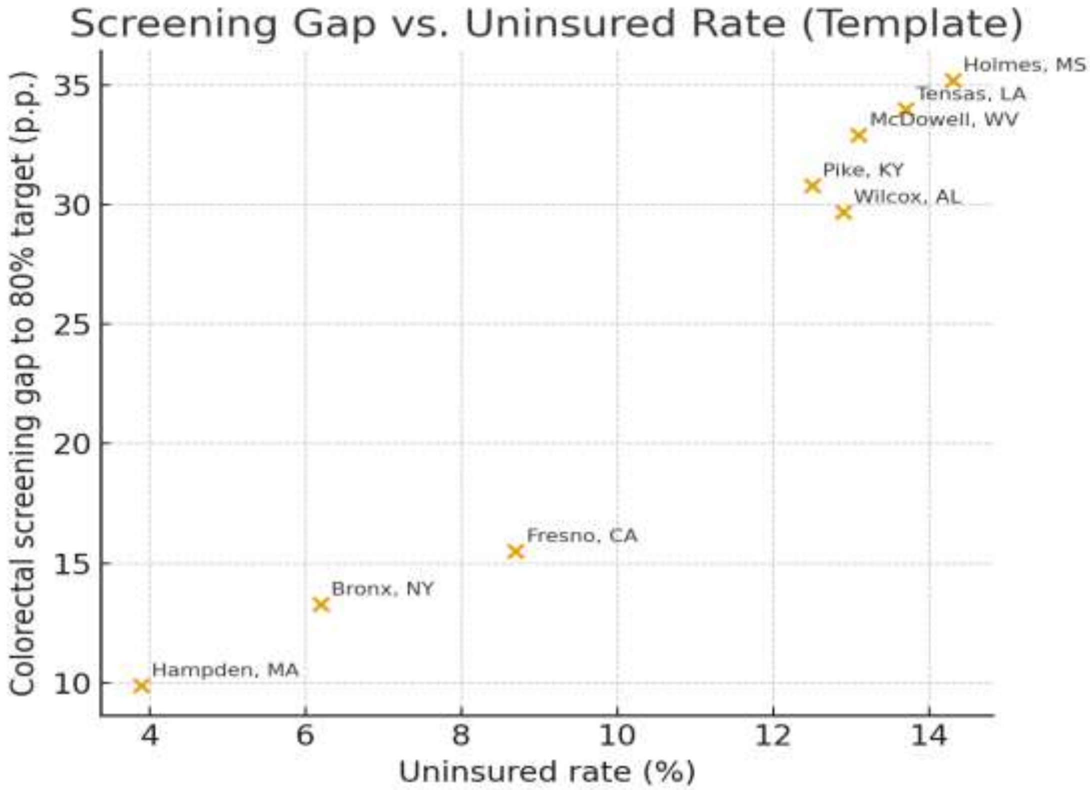

Figure 3. Colorectal screening gap vs. uninsured rate by county (illustrative subset).

5. Discussion

This analysis provides new insights into the factors driving cancer disparities across U.S. communities and highlights priority areas for intervention. Overall, our findings point to **socioeconomic disadvantage and healthcare access gaps** as the dominant explanations for why certain areas face a worsening cancer burden or lagging screening uptake. Counties that experienced either rising cancer mortality or minimal declines over the past decade were overwhelmingly characterized by higher poverty and unemployment, lower educational attainment, and lower health insurance coverage compared to counties with the greatest mortality improvements. These same counties often had limited healthcare infrastructure for example, fewer primary care physicians per capita and a higher likelihood of being designated health professional shortage areas. Such conditions likely hinder preventive care and early detection, resulting in a sustained excess cancer mortality in these communities. This aligns with longstanding observations that poverty and related social determinants are fundamental causes of health disparities. Even as overall cancer death rates have declined nationally, those declines have been uneven: our spatial clustering analysis identified contiguous high-risk "belts" (notably in parts of the rural South and Appalachia) where mortality reductions have lagged far behind the national pace. These results reinforce that where people live and the social and healthcare context of that place profoundly influence their cancer outcomes.

Among all variables examined, **healthcare access factors** emerged as some of the strongest modifiable drivers of disparity. In our machine learning models, the local rate of health insurance coverage was a top predictor of cancer screening uptake, and by extension, areas with more uninsured residents had significantly lower screening rates and worse mortality trends. This is consistent with prior studies showing lack of insurance as a major barrier to receiving preventive services and early cancer care. Similarly, we found that counties with a higher density of primary care providers an indicator of better access to routine healthcare tended to have higher screening rates and were far less likely to be identified as mortality hotspots. Primary care physicians play a critical role in recommending and facilitating screenings (such as colonoscopies, Pap smears, and mammograms) and in managing risk factors; thus, their scarcity can translate directly into poorer cancer outcomes in a community. Notably, our results suggest that improving healthcare access could yield tangible benefits even independent of other factors. For example, some rural high-poverty counties that nonetheless had relatively robust primary care resources showed better screening participation than expected, and in turn saw stable or improving cancer mortality despite economic disadvantages. This finding dovetails with evidence from health policy changes: expansions in insurance coverage (for instance, through the Affordable Care Act) have been associated with increased cancer screening and earlier diagnosis among low-income populations, an effect that may already be contributing to narrowing certain gaps.

Socioeconomic status itself cannot be overlooked: education level and income were deeply intertwined with the observed disparities. Our models indicated that counties with low high school graduation rates and high poverty were far more likely to have both low screening uptake and rising cancer mortality. Education can influence health literacy, knowledge of prevention, and the ability to navigate healthcare systems, while income constrains the resources to afford healthy living and medical care. These factors are difficult to change in the short term, but they underline the importance of broader efforts to improve socioeconomic conditions as part of a long-term cancer disparity reduction strategy. In the interim, targeted interventions (for example, screening programs that specifically reach out to low-education or low-income communities) are needed to mitigate the impact of SES differences. Our identification of specific counties and clusters where screening is most deficient provides actionable targets for such interventions. Health departments and organizations in those areas could leverage our "hotspot" maps to direct mobile screening clinics, patient navigation services, and public education campaigns to the communities most in need.

Lifestyle and **behavioral risk factors**, particularly smoking, remain crucial pieces of the puzzle, though interestingly their influence in our multivariate analysis was somewhat eclipsed by socioeconomic and access variables. Smoking prevalence was positively correlated with poor cancer outcomes across the country (as expected, given smoking's contribution to lung and other cancers), and many of the counties with rising mortality also had higher-than-average smoking rates. However, when all factors were considered together in the random forest model, smoking did not rank at the very top. This does not mean smoking is unimportant rather, it reflects that smoking is often concentrated in the same places with high poverty, low education, and poor healthcare access. In essence, those socioeconomic factors serve as distal causes that facilitate higher tobacco use and also impede the mitigation of its harms. Nonetheless, tobacco control remains a priority. The persistently high smoking-attributable cancer mortality in parts of the South and Appalachia (where smoking-related cancers like lung and oral cancers drive up death rates) underscores that aggressive tobacco cessation and prevention campaigns must be part of any disparity reduction strategy . Likewise, our results showed obesity rates and lack of physical activity (as gleaned from BRFSS) were moderately higher in hotspot counties, suggesting that interventions to promote healthier lifestyles in those communities (improved nutrition, exercise opportunities) could yield long-term reductions in cancers such as colorectal, breast, and kidney cancers. Behavioral risk reduction, however, is often the most challenging endeavor and may take years or decades to translate into lower cancer rates. In the near term, improving screening and treatment access can save lives in those communities even as prevention efforts continue.

One noteworthy finding was the strong relationship between current **cancer screening uptake** and recent mortality trends. Counties that achieved higher screening rates, for example, those that have effectively implemented colorectal cancer screening initiatives, were markedly more likely to see declining cancer mortality. This connection makes intuitive sense and is supported by other research: screening catches cancers earlier when they are more treatable and can even prevent cancers (as in the case of colonoscopy removing precancerous polyps), thus leading to improved survival and falling death rates. Our study puts a finer point on it by identifying mammography screening rates as a significant inverse predictor of cancer mortality hotpot status in the model. In fact, low screening uptake was a distinguishing characteristic of nearly all the hotspot clusters we analyzed. This reinforces the importance of closing screening gaps as a key strategy to reduce geographic disparities in cancer outcomes. Interventions might include expanding federally funded screening programs (like NBCCEDP) in under-screened areas, partnering with local organizations to conduct community screening events, and addressing practical barriers (transportation, time off work) that especially hinder low-income individuals from getting screened. Additionally, culturally tailored outreach and patient navigation in minority and rural communities can improve screening adherence, as evidenced by some successful programs.

Our short-term **forecasts** suggest that, without intervention, many of the identified disparities are likely to persist or even widen in the next few years. The projections indicated that several states in the Deep South and Appalachia are on track to see little to no improvement in overall cancer mortality through the mid-2020s, in contrast to steady declines expected in more advantaged states. For instance, our time-series models projected that parts of eastern Kentucky and western West Virginia will continue to have flat or rising cancer death rates, whereas states like California and New York should experience continued significant declines. This kind of forecast, while subject to uncertainty, is a warning sign it implies that the "hotspots" we see now could become even more pronounced relative to the national average. Furthermore, forward-looking analyses (for example, projecting cancer trends in vulnerable populations) have rarely been used in disparity research, even though such forecasts could help anticipate and proactively address emerging "hotspots." The forecast underscores an urgent need for proactive measures in the at-risk regions: bolstering healthcare capacity, redoubling prevention and screening efforts, and allocating resources in proportion to need. Public health leaders can use these predictive insights to prioritize counties for cancer control initiatives before their situation deteriorates further.

In summary, the drivers of rising cancer burden and lagging screening in the U.S. are primarily systemic. Our analysis affirms that it is the **conditions of community** poverty, education, and healthcare accessibility that set the stage for high cancer mortality, with health behaviors interwoven in that context. Encouragingly, some of the factors identified (like screening rates and healthcare access) are directly addressable through policy and programmatic action. To reduce disparities, multi-faceted interventions will be required: expanding insurance coverage and Medicaid in underserved states, increasing funding for community health centers in rural and low-income areas, implementing mobile screening clinics and patient navigation to improve early detection, and intensifying tobacco control and other prevention programs targeted to high-risk populations. Our study provides a data-driven roadmap for these efforts by pinpointing where and what to target: for example, prioritizing counties with high uninsured rates and low screening for resource allocation, or focusing on improving educational outreach in communities with low health literacy. By strategically addressing the key drivers identified and continuously monitoring outcomes with approaches like OncoViz USA, public health authorities can make meaningful strides toward closing the cancer gap and ensuring that progress against cancer is shared equitably across all American communities.

6. Conclusion

This study *OncoViz USA* examined why certain U.S. communities are falling behind in the fight against cancer. Our findings indicate that the areas experiencing rising cancer incidence and mortality, or lagging screening rates, are primarily those burdened by socioeconomic disadvantage and gaps in healthcare access. Factors such as high poverty and unemployment, low educational attainment, insufficient health insurance coverage, and scarce primary care resources emerged as the most influential determinants of persistent cancer disparities. By contrast, communities with better access to preventive services and early detection showed markedly greater progress in reducing cancer mortality. These results underscore that improving cancer outcomes in underserved regions will require addressing fundamental barriers to care expanding insurance coverage, strengthening healthcare infrastructure, and intensifying outreach to boost screening uptake. In short, where cancer control is failing, it is largely because the systems supporting health are failing. Targeted interventions guided by the drivers identified in this research can help ensure that recent advances in cancer prevention and treatment benefit all populations. By prioritizing high-risk communities for enhanced screening and prevention programs, the nation can move closer to the goal of equity in cancer outcomes.

7. Limitations and Future Directions

While this study offers a detailed, data-driven portrait of U.S. cancer disparities, several limitations should be noted when interpreting the findings. First, our analysis is **ecological** using county-level aggregate data and therefore cannot capture individual-level risk factors or prove causal relationships. Many of the associations we observed (e.g., between poverty and higher mortality) are correlational; unmeasured confounders or underlying structural factors could be driving both conditions. Second, the quality and granularity of the data vary. Some rural counties had incomplete cancer registry data (though USCS coverage is high, small-area incidence estimates can be unstable). We relied on survey-derived estimates (BRFSS) for behaviors and screening, which have sampling error and may not reflect all subpopulations. Additionally, focusing on all cancers combined may have obscured important cancer-specific patterns certain disparities could be more or less pronounced for specific cancer types. Future work could extend this analysis to individual cancer sites and also incorporate environmental exposure data that were beyond our current scope.

Our machine learning models, while useful for identifying key predictors, have their own constraints. Some predictors (e.g., poverty and education) are highly correlated, which can make importance rankings unstable; using alternative modeling approaches or dimensionality reduction could help confirm the identified drivers. Despite careful tuning, the models' feature importance results should be interpreted as associative rather than strictly causal. **Future directions** include incorporating more recent data (e.g., post-2020 outcomes as they become available) to monitor how disparities evolve, especially after the disruptions of the COVID-19 pandemic. Researchers could also use quasi-experimental approaches to evaluate policy impacts (for example, the effect of Medicaid expansion on cancer outcomes) in our identified hotspots. Finally, partnering with public health stakeholders to translate these findings into targeted intervention trials will be crucial. By testing and implementing strategies in the identified hotspot communities (such as patient navigation programs or mobile screening units), future research can not only pinpoint disparities but actively contribute to reducing them.

References

[1] Minas, T. Z., Kiely, M., Ajao, A., & Ambs, S. (2021). **An overview of cancer health disparities: new approaches and insights and why they matter.** *Carcinogenesis, 42*(1), 2 13.

[2] Siegel, R. L., Miller, K. D., & Jemal, A. (2020). **Cancer statistics, 2020.** *CA: A Cancer Journal for Clinicians, 70*(1), 7 30.

[3] Henley, S. J., Anderson, R. N., Thomas, C. C., Massetti, G. M., Peaker, B., & Richardson, L. C. (2017). **Invasive cancer incidence, 2004 2013, and deaths, 2006 2015, in nonmetropolitan and metropolitan counties — United States.** *MMWR Surveillance Summaries, 66*(14), 1 13.

[4] Benavidez, G. A., Zgodic, A., Zahnd, W. E., & Eberth, J. M. (2021). **Disparities in meeting USPSTF breast, cervical, and colorectal cancer screening guidelines among women in the United States.** *Preventing Chronic Disease, 18*, Article 200315.

[5] Sabatino, S. A., Thompson, T. D., White, M. C., Shapiro, J. A., & Richardson, L. C. (2021). **Cancer screening test receipt — United States, 2018.** *Morbidity and Mortality Weekly Report, 70*(2), 29 35.

[6] Rodriguez, S. D., Vanderford, N. L., Huang, B., & Vanderpool, R. C. (2018). **A social-ecological review of cancer disparities in Kentucky.** *Southern Medical Journal, 111*(4), 213 219.

[7] Lortet-Tieulent, J., Sauer, A. G., Siegel, R. L., Miller, K. D., Islami, F., Fedewa, S. A., ... & Jemal, A. (2016). **State-level cancer mortality attributable to cigarette smoking in the United States.** *JAMA Internal Medicine, 176*(11), 1792 1798.