AI Implementation in U.S. Hospitals: Regional Disparities and Health Equity Implications

Yeon-Mi Hwang, PhD[1], Madelena Y. Ng, DrPH[1], Malvika Pillai, PhD[1,2], Michelle P. Sahai, BA[1], Tina Hernandez-Boussard, PhD[1,3]

[1]Department of Medicine (Biomedical Informatics), Stanford University, Stanford, CA, USA

[2]Veterans Affairs Palo Alto Health Care System, Palo Alto, California, USA

[3]Department of Biomedical Data Science, Stanford University, Stanford, CA, USA

Correspondence to.

Dr. Yeon-Mi Hwang, PhD

Center for Biomedical Informatics Research

School of Medicine, Stanford University

3180 Porter Drive, Palo Alto, CA, 94304

Email: ymh@stanford.edu

Abstract

AI has the potential to improve healthcare delivery, but uneven geographic adoption and implementation can reinforce existing care gaps and inefficiencies. We analyzed data from 3092 U.S. hospitals using the 2023 American Hospital Association (AHA) Data, community-level socioeconomic indicators, and 2023-2025 CMS hospital quality metrics to assess: where AI is implemented, what factors drive implementation, and how AI is associated with hospital quality. We found that hospital AI implementation is significantly clustered, with over 67% misaligned with critical healthcare needs like provider shortages. Clustering- and hotspot-analysis-mapped regions with high and low AI adoption are used to identify specific areas for targeted equity interventions. Geographically weighted regression reveals that factors driving AI implementation vary by region, suggesting universal strategies may be ineffective. AI holds transformative potential for healthcare, yet adoption is geographically clustered and often misaligned with areas of greatest need. These findings underscore the urgent need for region-specific strategies to ensure AI advances health equity rather than worsening it.

1. Introduction

Hospitals across the United States are rapidly adopting artificial intelligence (AI) and machine learning (ML) technologies to enhance clinical decision-making and streamline operations[1]. However, this technological revolution is not happening equally everywhere.

Existing evidence shows notable disparities between socioeconomic groups and age cohorts in digital health engagement. Prior studies demonstrate that patients from lower socioeconomic backgrounds exhibit markedly lower digital health literacy[2], while older patients are considerably less likely to utilize digital health technologies compared to younger populations[3]. These demographic patterns represent only one dimension of a more complex geographic phenomenon. At the institutional level, high-resource hospitals are more likely to successfully implement and evaluate AI systems than low-resource facilities [4].

Geographic disparities in AI implementation are also becoming increasingly critical. Recent work by Chen and Yan[5] using the American Hospital Association (AHA) Information Technology (IT) Supplement data show that hospitals in areas with high-deprivation neighborhoods were significantly less likely to implement AI and reveal uneven distribution of AI implementation patterns across the US. Geographic disparities are particularly concerning because they compound existing health inequities. For example, rural areas have higher percentages of elderly residents and lower-income populations, and already experience worse health outcomes. When these communities also lack access to beneficial AI-enabled healthcare technologies, existing disparities are further amplified.

Our prior research has further demonstrated that AI implementation varies significantly at the county level and correlates with regional health outcomes, suggesting this technological disparity may translate directly into health inequities.[6] Ensuring equitable access to quality healthcare remains a critical public health objective, making it essential to understand and address concerning geographic patterns in AI implementation.

To address this gap, we conducted a comprehensive geospatial analysis of AI implementation patterns across US hospitals, integrating multiple data sources to examine both the geography of AI adoption and its implications for health equity. We hypothesized that AI implementation follows a disproportionate

geographic distribution with identifiable clusters of high and low AI implementation. Using geospatial analytical techniques with data from AHA, area-level socioeconomic indicators, community characteristics, and Centers for Medicare & Medicaid Services (CMS) quality metrics, we aimed to: (1) characterize the spatial distribution of hospital AI implementation and identify specific geographic clusters, hotspots, and coldspots that can serve as actionable targets for intervention; (2) examine spatial variation in factors driving AI implementation to inform regionally tailored strategies; (3) assess alignment between healthcare need and AI implementation patterns; and (4) evaluate longitudinal associations between AI adoption and hospital quality performance to understand downstream impacts. Rather than simply documenting disparities, this study provides stakeholders with concrete geographic targets, evidence-based guidance for promoting equitable AI adoption, and critical insights into the real-world consequences of implementation decisions.

2. Method

2.1 Data Source

This study utilized data from the 2023 American Hospital Association (AHA) Annual Survey and its Information Technology (IT) Supplement. The AHA Annual Survey collects comprehensive information on institutional characteristics and capabilities from hospitals nationwide, while the IT Supplement specifically focuses on IT infrastructure, applications, and operational capacities. While the AHA Data previously included basic questions on predictive modeling, the 2023 version expanded this coverage to include detailed questions on predictive modeling for the first time, covering use cases, development approaches, and evaluation methodologies. Our analysis included all hospital facilities that responded to the 2023 AHA IT Supplement.

2.2 Variables

2.2.1. Hospital Predictive Model and AI Implementation Measures

Hospitals were classified into three categories based on their predictive modeling capabilities using data from the 2023 AHA IT Supplement: no predictive models (0), non-AI predictive models (1), or AI models (2). Hospitals that responded to the IT Supplement but did not answer this specific survey question were

classified as having no predictive models (imputed as 0). This AI implementation level served as our primary variable of interest, functioning as both a dependent variable when examining associations with geographic and institutional factors, and as an independent variable when assessing relationships with quality metrics.

We developed three exploratory secondary measures to capture key implementation dimensions, as validated composite measures for these specific aspects of AI implementation were not available in the existing literature. These secondary measures assessed predictive model implementation breadth, development approach, and evaluation practices using structured scoring systems as proxy indicators for these implementation dimensions. All secondary measures incorporated the primary AI implementation score (0-2) as a foundational component. The breadth score (range 0-4) added 0.25 points for each reported model use case beyond the base score. The development score (range 0-6.5) assigned additional points based on model source: 2 points for internally developed models, 1 point for EHR vendor or third-party solutions, and 0.5 points for public domain models. The evaluation score (range 0-4) incorporated reported assessments of model accuracy and bias, with points assigned incrementally (0.25 to 1) based on the comprehensiveness of evaluation activities across models (Table S1). Since the survey questions underlying these secondary measures encompass both AI and non-AI predictive models, we use the term "model" rather than "AI model" when referring to these broader implementation measures.

2.2.2. Hospital Characteristics

Hospital characteristics were obtained from the AHA Annual Survey dataset to serve as proxy indicators for institutional resource levels. We included organizational attributes such as teaching hospital status, critical access hospital designation, rural referral center status, ownership type, bed capacity, health system affiliation, frontline hospital designation, Joint Commission accreditation status, presence of subsidiary hospitals, community hospital designation, and delivery system model. Detailed definitions and categorizations of these variables are provided in Table S2. We additionally calculated Core Index and Friction Index as developed by Strawley et al.,[7]. The Core Index measures adoption of foundational interoperability capabilities, with higher scores indicating greater interoperability. The Friction Index

quantifies barriers to interoperability, with higher scores indicating greater barriers. We included these interoperability indices, calculated from AHA IT Supplement responses, as they efficiently summarize multiple dimensions of IT infrastructure and exchange capabilities that may influence a hospital's readiness to implement AI systems.

2.2.3. Geographic and Community Characteristics

Rural-urban classification was obtained from the AHA Annual Survey. Hospital geographic coordinates were derived from street addresses using the Positionstack API[8]. We linked hospitals with their service area, county, state, census division, and census region using responses from the AHA Annual Survey. Area deprivation index (ADI)[9,10] and social vulnerability indices (SVI)[11] were first aggregated to five-digit zip codes and then to hospital service areas (HSAs) using a ZIP code-HSA crosswalk from Dartmouth Atlas[12]. We used median values for this aggregation because deprivation indices typically include numerous small areas with similar socioeconomic profiles within each service region, making the median more robust to outliers while representing the typical deprivation level.

Healthcare Professional Shortage Area (HPSA) and Medically Underserved Area (MUA) scores were obtained from the Health Resources and Services Administration[13]. These designations represent geospatial boundaries with scores reflecting the degree of healthcare shortage and underservice, respectively. We spatially joined these areas to HSAs wherever they overlapped. For hospital service areas intersecting with multiple HPSAs or MUAs, often with substantially different scores, we calculated mean values to capture the combined influence of all shortage designations. Hospitals without overlapping HPSA or MUA designations received a score of 0. We additionally collected county-level digital infrastructure metrics from 2019-2024 U.S. Census Bureau data[14]: internet access percentage (households with any type of internet subscription), broadband percentage (households with high-speed internet subscription), and computing device percentage (households with at least one computing device such as desktop, laptop, tablet, or smartphone). These digital infrastructure variables were linked to hospitals based on their county location from the AHA Annual Survey (Table S2).

2.2.4. Hospital Quality Metrics

We obtained 47 continuous hospital quality of care metrics from the Centers for Medicare & Medicaid Services (CMS). These standardized metrics assess clinical performance, care quality, and patient safety across U.S. hospitals. For our cross-sectional and longitudinal analyses, we collected quarterly data from 2022, 2023, 2024, and the first two quarters of 2025, yielding 14 measurement timepoints.

2.3 Analysis

2.3.1 Descriptive statistics

We generated descriptive statistics comparing hospital and community characteristics across regions and AI implementation categories. Continuous variables are presented as means with standard deviations, while categorical variables are shown as frequencies with percentages. Statistical comparisons across groups were conducted using ANOVA for continuous variables and chi-square tests for categorical variables. For the state-level analysis, states with less than 20 hospitals were dropped to avoid unreliable statistical estimates due to small sample sizes. This included Alaska, DC, Delaware, Hawaii, Maine, Rhode Island, Vermont, and Wyoming.

2.3.2 Healthcare Need-AI Adoption Mismatch Assessment

We evaluated whether AI adoption aligns with areas of greatest healthcare need by comparing AI implementation levels against established need indicators. Given that AI implementation scores represented distinct and meaningful categories (no predictive models, non-AI models, AI models) with reasonably balanced distribution across hospitals, we retained these original classifications to preserve interpretability. Healthcare need measures were divided into rank-based tertiles to enable systematic comparison across need levels.

We generated cross-tabulations between AI implementation and need tertiles and visualized these patterns using heatmaps. Our primary metric was the mismatch rate, calculated as the percentage of hospitals where need and implementation levels do not correspond. High mismatch rates indicate systematic misalignment and suggest disparities in AI resource allocation across communities with different levels of need. We selected HPSA, MUA, ADI, and SVI as need indicators because areas with

provider shortages (HPSA), insufficient healthcare resources (MUA), neighborhood disadvantage (ADI), and social vulnerability (SVI) may benefit most from AI implementation to help address healthcare capacity and resource constraints.

2.3.3 Spatial Clustering and Hotspot Analysis

2.3.3.1 Geographic Visualization and Distribution

We visualized the geographic distribution of AI implementation through heatmaps across the United States and within individual census divisions. It should be noted that areas with higher hospital density may naturally show greater concentrations of AI implementation. Census divisions are categorized into nine regional groups (New England, Middle Atlantic, East North Central, West North Central, South Atlantic, East South Central, West South Central, Mountain, and Pacific).

2.3.3.2 Spatial Clustering Analysis

To assess spatial clustering of hospitals, we conducted a nearest neighbor analysis using hospital coordinates. We calculated the observed mean nearest neighbor distance using a KDTree algorithm. For evaluating clustering of hospital AI implementation specifically, we calculated Global Moran's I using projected hospital coordinates and AI scores. This analysis used k=5 nearest neighbors to define local spatial relationships, allowing detection of localized spatial clustering while maintaining statistically significant Moran's I and z-score values (Figure S1).

We applied DBSCAN clustering to identify geographically proximate hospitals with similar AI implementation patterns. These clusters serve as practical units for stakeholders to target interventions and develop region-specific approaches to AI implementation. To address the uneven geographic distribution of hospitals, we computed local density estimates for each hospital using its k=5 nearest neighbors, then upweighted AI implementation scores in higher-density regions to make dense hospital clusters more identifiable and reinforce DBSCAN's density-based clustering principle. After standardizing these density-adjusted AI scores along with hospital coordinates, we performed DBSCAN clustering with eps=0.25. We selected these parameters based on the inflection point where the number of clusters

stabilized and noise remained acceptably low (Figure S2). For the state-level analysis, states with less than 20 hospitals were dropped to avoid unreliable statistical estimates due to small sample sizes.

2.3.3.3 Hotspot Analysis

We complemented the clustering analysis with a Getis-Ord Gi* hotspot analysis to identify statistically significant spatial concentration of high and low AI implementation. Hotspots can guide policymakers toward successful regional examples, while coldspots highlight areas needing target support and resources. Using projected hospital coordinates and AI implementation scores, we computed Gi* z-scores based on each hospital's five nearest neighbors, consistent with our DBSCAN parameter. Hotspot (high implementation) and coldspot (low implementation) classifications were assigned based on z-score significance thresholds (90%, 95%, 99%). For visualization purposes, we aggregated these classifications to the HSA level. For descriptive statistics, we further aggregated counts to state, census division, and national levels. For the state-level analysis, states with less than 20 hospitals were dropped to avoid unreliable statistical estimates due to small sample sizes.

2.3.4 Predictive Modeling and Feature Importance Analysis

We employed a random forest classification model to identify the most influential factors predicting hospital AI implementation levels. The model incorporated both hospital-level characteristics and geospatial features as predictors. We calculated SHapley Additive exPlanations (SHAP) values to determine feature importance and understand the directional relationships between predictors and AI implementation. SHAP values quantify each feature's contribution to individual predictions while accounting for feature interactions.

2.3.5 Geographically Weighted Regression Analysis

We implemented a geographically weighted regression (GWR) to examine how the influence of community-level structural and technological factors on hospital AI implementation level varies across geographic space. Hospital coordinates were used to compute spatial weights based on Gaussian kernels. We modeled the association between hospital AI base scores and predictor features. All continuous variables were standardized prior to analysis to allow for coefficient comparability. The

resulting spatially varying coefficients were visualized through maps to assess regional heterogeneity in association. We used a bandwidth of 2 decimal degrees, calculated as approximately 222 km radius at mid-US latitudes.[15] This bandwidth encompasses multiple counties or most smaller states, providing a good balance to capture meaningful regional patterns while ensuring sufficient local observations for stable coefficient estimation.

2.3.6 Longitudinal Analysis of Hospital Quality Trends by AI Implementation Status

We examined associations between AI implementation levels and hospital quality metrics using a longitudinal approach. Building on our prior county-level cross-sectional study that established preliminary patterns of associations between AI implementation and health outcomes,[6] this analysis addresses previous methodological constraints by employing hospital-level data across 14 quarterly time points from 2022 to 2025 with adjustment for regional factors.

While our longitudinal approach helps address temporal misalignment issues, important limitations remain. Our data captured AI implementation status as of 2023 but not precise implementation timing, and our AI implementation measure reflects broad categories without detailed information about specific applications or clinical targets.

Despite these constraints, this exploratory analysis serves several important purposes: advancing beyond our previous analysis to examine quality trajectory patterns across hospitals with different AI implementation levels, identifying potential areas where AI may be associated with quality differences over time, and providing enhanced evidence for policymakers about technology adoption and performance relationships.

We analyzed quality metrics with less than 50% missingness in the second quarter 2025 report to balance data completeness with sample size. We strategically selected confounding variables to avoid high-dimensional modeling issues: the top three predictors of AI implementation from section 2.3.4 to control for systematic differences between hospitals, and the top five LASSO-identified outcome-specific predictors for each quality metric. We excluded mediating features such as interoperability capabilities from covariate selection.

We examined quality metric trajectories across 14 quarterly time points from 2022 to 2025 using linear mixed-effects models adjusted for the selected covariates. These models calculated differential slopes by AI implementation level, with hospitals grouped into three categories: AI-0 (no predictive models), AI-1 (non-AI predictive models), and AI-2 (AI predictive models), using AI-0 as the reference group.

3. Results

3.1 Hospital AI Implementation Patterns

Among hospitals responding to the AHA IT Supplement (N = 3092), 941 (30.4%) reported using no predictive models, 867 (28.1%) employed non-AI predictive models, and 1090 (35.2%) utilized AI-based predictive models. The remaining 6.3% (194 hospitals) did not report their AI implementation status and were classified as non-users in our analysis. Table 1 and Figure 1 presents AI implementation levels stratified by U.S. census division. Predictive models were most commonly utilized to predict patient health trajectories (58.6%), with the majority developed by hospitals' EHR vendors (51.6). The majority of hospitals reported no evaluation for model accuracy (50.8%) and bias (53.5%).

Geographic variation was notable across census divisions and states (Figure 1). At the census division level, the South Atlantic division consistently showed the highest scores across all AI implementation measures: base implementation (1.2), breadth (2.0), development (3.1), and evaluation (2.1). State-level analysis revealed additional geographic heterogeneity, with South Carolina achieving the highest overall AI implementation (1.7) and breadth scores (2.5), Colorado leading in development (4.0), and Utah demonstrating the strongest evaluation scores (3.0). In contrast, Montana was lowest across all scores (0.6, 0.9, 1.6, 0.9)**;** detailed scores by census division and state are provided in Table 1 and Table S4.

Descriptive statistics for hospital and community characteristics are stratified by census division (Table S3), by state (Table S4), and by hospital AI implementation level (Table S5). All characteristics differed significantly across geographic regions. However, when stratified by AI implementation level, hospital Medicaid inpatient proportion, overall SVI, and SVI theme 2 (household composition/disability) showed no statistically significant differences in distribution.

Heatmap analysis revealed consistent concentration of AI implementation in metropolitan regions across all census divisions (Figure S3). However, these concentration patterns partly reflect the underlying clustering of hospitals themselves, as confirmed by our nearest neighbor analysis showing significant hospital clustering ($P<0.0001$).

3.2 Healthcare Need-AI Adoption Mismatch Assessment

Figure 2 reveals notable misalignment between healthcare needs and AI implementation across multiple dimensions. Mismatch rates ranged from 67.3% for the SVI to 72.4% for both MUA Elder and MUA Infant scores. The heatmap analysis shows hospitals concentrated in misaligned categories: high need/low AI implementation (11.1-14.7%) and low need/high AI implementation (10.9-13.9%), while ideal alignment categories consistently showed lower percentages (8-10%).

3.3 Spatial Clustering and Hotspot Analysis

3.3.1. Detection of Spatial Clustering Patterns

Moran's I analysis revealed significant spatial autocorrelation for all AI model implementation metrics ($P<0.0001$; Table S6). Model development scores showed the strongest clustering (Moran's I=0.19, z-score=17.67), while breadth scores showed the weakest clustering (Moran's I=0.16, z-score=14.56). Regional analysis confirmed clustering in most census divisions, with variation in degree of clustering across regions and states (Table S6).

3.3.2. Identification of Density-Based Clusters

The adaptive DBSCAN clustering algorithm identified 434 distinct clusters of hospital AI implementation at the base level across 2397 hospitals nationwide (77.52%). Clusters averaged 5.5 hospitals per cluster, with regional variation in cluster size. The Pacific region exhibited the largest average cluster size (14.6 hospitals), while the East South Central region had the smallest clusters (2.9 hospitals). Complete DBSCAN clustering results for all AI model implementation metrics are presented in Table S7.

3.3.3. Hotspot and Coldspot Analysis

Hospital-level analysis identified statistically significant 159 hotspots and 242 coldspots of AI implementation. At the hospital service area level (n=1,842), we detected 85 hotspots and 175 coldspots for AI base score (Figure 3). The South Atlantic census division exhibited the highest mean z-score across all AI implementation measures. At the state level, South Carolina led for base score, breadth score, and evaluation score, while Utah showed the highest scores for development measures. Complete hotspot analysis results are presented in Figure S4-S7 and Table S8-S12.

3.4 Predictive Modeling and Factors Associated with AI Implementation

To identify key factors associated with hospital AI implementation, we developed a random forest classification model incorporating hospital-level, geographic, and community characteristics. The model achieved moderate predictive performance with an area under the curve (AUC) of 0.67.

Figure S8 presents the 10 most influential predictors of AI implementation based on SHAP values. The analysis revealed that both institutional and regional factors serve as important correlates, with 7 hospital-level characteristics and 3 regional factors among the top predictors.

The five most influential features were Core Index (measuring interoperability capabilities), Friction Index (reflecting barriers to information exchange), delivery system model, bed capacity, and SVI Index theme 3 (Minority Status and Language). SHAP analysis demonstrated that Core Index, bed size, and SVI Index - theme 3 exhibited positive associations with AI implementation, where higher values corresponded to increased likelihood of adoption. Conversely, Friction Index and certain delivery system models showed inverse relationships with AI implementation.

3.5 Spatial Heterogeneity in Factors Influencing AI Implementation

Geographically weighted regression analysis revealed significant spatial heterogeneity in the relationships between hospital-level and community-level factors and AI implementation across the US (Figure S9 and S10). Local regression coefficients varied by geographic region, indicating that the influence of predictor variables on AI adoption was not spatially uniform. The Core Index demonstrated the most spatial stability, maintaining positive coefficients throughout all regions. In contrast, other hospital-level factors exhibited pronounced spatial heterogeneity, with coefficients varying from negative to positive values.

Community-level characteristics similarly demonstrated spatial variation in their associations with AI implementation.

3.6 Longitudinal Analysis of Hospital Quality Trends by AI Implementation Status

Our longitudinal analysis examined associations between AI implementation and changes in hospital quality metrics across 14 quarterly time points from 2022 to 2025. Linear mixed-effects models compared trend trajectories between hospitals without predictive models (AI-0, reference) and those with AI-based predictive models (AI-2). Of 20 quality metrics analyzed, 12 showed statistically significant trajectory differences after FDR correction (Figure S11, Table S10). AI-adopting hospitals demonstrated favorable trends in several domains: reduced heart failure mortality (slope difference: -0.20 [-0.23,-0.17], P<0.0001), pneumonia mortality (slope difference: -0.32 [-0.36, -0.28], P<0.0001), fewer excess acute care days after heart failure discharge (slope difference: -0.90 [-1.3, -0.54], P<0.0001), lower Medicare spending per patient (slope difference: -0.0022 [-0.0033, -0.00096], P=0.0009), and higher influenza vaccination coverage (slope difference: 0.66 [0.44, 0.87], P<0.0001).

However, AI-adopting hospitals showed less favorable trends across multiple areas: prolonged ED median times (slope difference: 1.16 [0.78, 1.50], P<0.0001), higher rates of patients leaving ED without being seen (slope difference: 0.047 [0.014, 0.079], P=0.009), increased unplanned hospital visits (slope difference: 0.048 [0.019, 0.076], P=0.015), increased unsafe opioid prescribing (slope difference: 1.16 [0.95, 1.40], P<0.0001), worsened sepsis care performance (SEP-1: slope difference: 0.91 [0.67, 1.10], P<0.0001; SEV-SEP-3HR: slope difference: 0.42 [0.25, 0.59], P<0.0001), and steeper decline in COVID-19 vaccination coverage (slope difference: -2.61 [-3.72, -1.48], P<0.0001).

4. Discussion

This nationwide study of 3092 U.S. hospitals reveals the uneven landscape of healthcare AI implementation and characterizes geographic disparities that threaten health equity. AI implementation exhibits pronounced regional variation, with distinct spatial clustering patterns across different geographic areas. Analyses demonstrated substantial misalignment between healthcare needs and AI adoption, with mismatch rates exceeding 67% across multiple indicators, indicating these technologies often fail to reach

the communities with greatest need. Beyond mapping current patterns, results show that the factors driving AI implementation vary substantially across geographic space, challenging one-size-fits-all policy approaches and suggesting the need for regionally tailored strategies. The identification of specific hotspots, coldspots, and clusters provides actionable targets for stakeholders to address implementation gaps. Longitudinal analysis revealed mixed associations between AI implementation and hospital quality metrics across clinical domains. Together, these findings provide essential evidence for policymakers, healthcare leaders, and technology developers working to ensure that AI expansion reduces rather than amplifies existing health inequities.

Our findings extend beyond previous spatial analyses of healthcare AI adoption by revealing complex geographic patterns across multiple implementation domains that cannot be captured through conventional statistical approaches. Chen and Yan[5] identified spatial autocorrelation in AI adoption and showed that hospitals in high-deprivation areas were less likely to implement these technologies. We extend this foundational work in several important ways. First, our analysis examines multiple dimensions of AI implementation, including evaluation, development, and breadth measures, rather than a single adoption metric. Second, we move beyond identifying general clustering patterns to precisely map statistically significant hotspots and coldspots across the US. Most importantly, our geographically weighted regression analysis demonstrates that the factors driving AI implementation vary substantially across geographic space, except for interoperability measures. This spatial heterogeneity suggests the need for regionally tailored strategies that account for local contexts and barriers rather than uniform implementation approaches.

The analysis reveals significant and systematic misalignment between healthcare needs and AI implementation across multiple dimensions. Mismatch rates range from 67.3% for the SVI to 72.4% for both MUA Elder and MUA Infant scores, with the highest misalignment occurring in areas with the greatest medical underservice for elder and infant care, particularly vulnerable populations in healthcare. The heatmap reveals consistent misalignment across all eight measures. The highest concentrations of hospitals appear in two problematic patterns: high need with low AI implementation, and low need with high AI implementation. Achieving more equitable AI distribution would require reducing these misaligned

concentrations and improving concordance between healthcare needs and implementation levels. These findings suggest current deployment patterns fail to align with population health priorities, representing a significant missed opportunity to leverage AI for reducing healthcare inequities. Achieving equitable distribution will require targeted interventions to correct these imbalances and establish need-based implementation frameworks.

Our feature importance analysis showed that interoperability measures (Core Index and Friction Index) emerged as the most influential predictors of AI implementation. This finding confirms that foundational data exchange capabilities serve as prerequisites for advanced analytics implementation[16–18]. Hospitals with higher Core Index scores showed higher AI implementation rates regardless of their location. This spatial consistency suggests that targeted investments in interoperability infrastructure could serve as the most effective strategy for reducing geographic disparities in AI adoption, providing a clear and actionable policy pathway.

Our longitudinal analysis across 14 quarterly time points from 2022 to 2025 suggested complex associations between AI implementation and hospital performance, consistent with findings from our previous cross-sectional study assessing county-level AI implementation and health outcomes.[6] This analysis advances our prior work by examining hospital-level quality trajectories over time rather than single-time point county-level associations. While significant methodological limitations detailed in the methods section and below require cautious interpretation, this approach provides important preliminary evidence about potential relationships between AI adoption and hospital performance trajectories. After adjusting for outcome-specific hospital characteristics and regional factors, AI-adopting hospitals showed more favorable trajectories in certain clinical areas, including reduced mortality rates for heart failure and pneumonia patients and decreased excess days in acute care after hospitalization. However, they demonstrated less favorable trends in emergency department efficiency and deteriorating sepsis care performance. Rather than indicating consistent improvements, these mixed associations highlight both the potential promise and complexity of relationships between AI implementation and hospital performance. These trajectory patterns provide important baseline evidence but underscore the critical need for standardized reporting of AI implementation timing and detailed implementation characteristics

to enable future research to understand the true impact of AI implementation on healthcare quality and patient outcomes.

Several limitations warrant consideration when interpreting our findings. First, our analysis is limited to hospitals that responded to the 2023 AHA IT Supplement, which may introduce selection bias as responding hospitals may differ systematically from non-responding hospitals in their IT infrastructure, resources, or organizational characteristics. Second, our cross-sectional measurement of AI implementation status from the 2023 AHA IT Supplement captures a single point in time and does not provide information on when implementation occurred or how it evolved. While our longitudinal quality analysis partially addresses this limitation, the precise timing of implementation remains unknown. Third, our quality metrics analysis was limited to variables with less than 50% missingness, which may introduce selection bias and limit the generalizability of quality-related findings, though this threshold still preserved sufficient statistical power to detect statistically significant associations. Fourth, despite our extensive data integration efforts, unmeasured confounders likely influence both implementation decisions and quality outcomes, which may also explain the moderate predictive performance of our random forest model (AUC = 0.67); however, the primary purpose of this model was to identify feature importance rather than achieve high predictive accuracy. Fifth, random forest and SHAP analysis assumes feature independence, which may not hold for correlated hospital, geographic, and community variables, potentially affecting the reliability of feature importance rankings. Sixth, our quality metric analysis did not account for specific AI use cases within hospitals because the survey questions about model use cases were not exclusive to AI models. Additionally, the use cases were too broad to link to specific quality metrics. Finally, while our geospatial methods identified significant patterns, they cannot establish causal relationships between geographic factors and AI implementation, or between AI implementation and quality outcomes. Future research should address these limitations through longitudinal tracking of AI implementation, standardized assessment of implementation maturity, and examination of specific AI use cases in relation to targeted outcomes.

5. Conclusion

Significant spatial clustering characterizes current hospital AI implementation patterns, with over 67% of deployments demonstrating misalignment with critical healthcare needs, particularly provider shortages. This creates a digital health divide that disproportionately limits vulnerable populations' access to beneficial emerging technologies. Geographically weighted regression showed substantial regional variation, indicating that universal implementation strategies will likely be ineffective. While longitudinal analysis showed improved mortality outcomes, decreased performance in emergency care and sepsis management highlights the complex and varied impact of AI on care delivery. These findings have immediate policy implications. Successful AI equity initiatives must be designed at the regional level, addressing locally specific barriers. Furthermore, AI implementation should be coupled with robust monitoring systems to detect and mitigate potential negative impacts. Without targeted efforts to deploy AI in underserved areas using region-specific strategies, current trends may worsen rather than improve health disparities.

6. Disclosures

Author Contributions: Drs. Yeon-Mi Hwang and Madelena Y. Ng had full access to all of the data in the study and take responsibility for the integrity of the data and the accuracy of the data analysis.

Concept and design: Hwang, Ng, Pillai, Hernandez-Boussard.

Acquisition, analysis, or interpretation of data: Hwang, Ng.

Drafting of the manuscript: Hwang.

Critical review of the manuscript for important intellectual content: All authors.

Statistical analysis: Hwang.

Obtained funding: Ng, Hernandez-Boussard.

Administrative, technical, or material support: Hernandez-Boussard.

Supervision: Hernandez-Boussard.

Conflict of Interest Disclosures: None reported.

Human Ethics and Consent to Participate declarations: Not applicable.

Funding/Support: Research reported in this publication was supported by The SCAN Foundation under Award Number G24-28. The content is solely the responsibility of the authors and does not necessarily represent the official views of the funder.

Role of the Funder/Sponsor: The funder had no role in the design and conduct of the study; collection, management, analysis, and interpretation of the data; preparation, review, or approval of the manuscript; and decision to submit the manuscript for publication.

Data Sharing Statement: Except for AHA data, all data are publicly available through the sources cited in the references. AHA data requires a subscription and can be obtained through AHA. Information on data access is available in the cited references.

7. References

1. Bajwa J, Munir U, Nori A, Williams B. Artificial intelligence in healthcare: transforming the practice of medicine. Future Healthc J. 2021 Jul;8(2):e188–94.

2. Estrela M, Semedo G, Roque F, Ferreira PL, Herdeiro MT. Sociodemographic determinants of digital health literacy: A systematic review and meta-analysis. Int J Med Inform. 2023 Sep 1;177(105124):105124.

3. Badr J, Motulsky A, Denis JL. Digital health technologies and inequalities: A scoping review of potential impacts and policy recommendations. Health Policy. 2024 Aug 1;146(105122):105122.

4. Nong P, Adler-Milstein J, Apathy NC, Holmgren AJ, Everson J. Current Use And Evaluation Of Artificial Intelligence And Predictive Models In US Hospitals: Article examines uses and evaluation of artificial intelligence and predictive models in US hospitals. Health Aff (Millwood). 2025 Jan 1;44(1):90–8.

5. Chen J, Yan AS. Hospital artificial intelligence/machine learning adoption by neighborhood deprivation. Med Care. 2025 Mar 1;63(3):227–33.

6. Hwang YM, Ng MY, Sahai MP, Pillai M, Hernandez-Boussard T. AI Implementation in U.S. Healthcare and Its Association With Elder Mortality and Quality of Care [Internet]. medRxiv. 2025 [cited 2025 Jun 26]. p. 2025.06.24.25330241. Available from: https://www.medrxiv.org/content/10.1101/2025.06.24.25330241v1.abstract

7. Strawley CE, Adler-Milstein J, Holmgren AJ, Everson J. New indices to track interoperability among US hospitals. J Am Med Inform Assoc. 2025 Feb 1;32(2):318–27.

8. positionstack: Real-Time & Accurate Geocoding API [Internet]. [cited 2025 May 30]. Available from: https://positionstack.com/

9. University of Wisconsin School of Medicine and Public Health. Area Deprivation Index [Internet]. 2020 [cited 2025 Feb 15]. Available from: https://www.neighborhoodatlas.medicine.wisc.edu/

10. Kind AJH, Buckingham WR. Making neighborhood-disadvantage metrics accessible - the neighborhood atlas. N Engl J Med. 2018 Jun 28;378(26):2456–8.

11. Flanagan BE, Gregory EW, Hallisey EJ, Heitgerd JL, Lewis B. A Social Vulnerability Index for Disaster Management. J Homel Secur Emerg Manag [Internet]. 2011 Jan 5;8(1). Available from: http://dx.doi.org/10.2202/1547-7355.1792

12. Supplemental Data [Internet]. Dartmouth Atlas DATA. Dartmouth Atlas Project; 2021 [cited 2025 Jun 4]. Available from: https://data.dartmouthatlas.org/supplemental/

13. What Is Shortage Designation? [Internet]. [cited 2025 May 30]. Available from: https://bhw.hrsa.gov/workforce-shortage-areas/shortage-designation

14. U.S. Census Bureau. Census Bureau Data [Internet]. United States Census Bureau. [cited 2025 Mar 30]. Available from: https://data.census.gov/

15. How much distance does a degree, minute, and second cover on your maps? [Internet]. USGS. [cited 2025 May 23]. Available from: https://www.usgs.gov/faqs/how-much-distance-does-a-degree-minute-and-second-cover-your-maps

16. Mandl KD, Gottlieb D, Mandel JC. Integration of AI in healthcare requires an interoperable digital data ecosystem. Nat Med. 2024 Mar 30;30(3):631–4.

17. Anzalone AJ, Geary CR, Dai R, Watanabe-Galloway S, McClay JC, Campbell JR. Lower electronic health record adoption and interoperability in rural versus urban physician participants: a cross-sectional analysis from the CMS quality payment program. BMC Health Serv Res. 2025 Jan 23;25(1):128.

18. Rehburg F, Graefe A, Hübner M, Thun S. How Interoperability can enable Artificial Intelligence in clinical applications. Stud Health Technol Inform. 2024 Aug 22;316:596–600.

Census Region			***Northeast***		***South***			***Midwest***		***West***		
Census Division	**Missing**	**Overall**	**NE**	**MA**	**SA**	**ESC**	**WSC**	**ENC**	**WNC**	**MTN**	**PAC**	**P-Value**
States			CT, ME, MA, NH, RI, VT	NJ, NY, PA	DE, DC, FL, GA, MD, NC, SC, VA, WV	AL, KY, MS, TN	AR, LA, OK, TX	IL, IN, MI, OH, WI	IA, KS, MN, MO, NE, ND, SD	AZ, CO, ID, MT, NV, NM, UT, WY	AK, CA, HI, OR, WA	
Number of Hospitals		3092	130	310	497	192	459	429	494	276	305	
AI Implementation Level												
AI implementation score, mean (SD)	194	1.1 (0.8)	1.1 (0.8)	1.1 (0.9)	1.2 (0.7)	1.0 (0.8)	0.9 (0.9)	1.0 (0.8)	1.0 (0.9)	1.1 (0.8)	0.9 (0.8)	<0.0001*
AI implementation level, n (%)												<0.0001*
AI/ML model		1090 (35.3)	44 (33.8)	133 (42.9)	198 (39.8)	60 (31.2)	157 (34.2)	135 (31.5)	183 (37.0)	105 (38.0)	75 (24.6)	
Non-AI/ML predictive model		867 (28.0)	43 (33.1)	63 (20.3)	191 (38.4)	57 (29.7)	68 (14.8)	155 (36.1)	102 (20.6)	72 (26.1)	116 (38.0)	
Do not know/Neither		941 (30.4)	33 (25.4)	101 (32.6)	89 (17.9)	64 (33.3)	194 (42.3)	119 (27.7)	167 (33.8)	84 (30.4)	90 (29.5)	
Secondary Model Implementation Measures												
Model breadth implementation, mean (SD)	202	1.7 (1.3)	1.6 (1.1)	1.8 (1.4)	2.0 (1.1)	1.4 (1.2)	1.4 (1.4)	1.8 (1.3)	1.6 (1.4)	1.7 (1.3)	1.6 (1.2)	<0.0001*
Model development score, mean (SD)	194	2.7 (2.1)	2.5 (1.8)	2.7 (2.2)	3.1 (1.9)	2.1 (1.8)	2.1 (2.2)	2.8 (2.1)	2.4 (2.2)	2.9 (2.3)	3.0 (2.3)	<0.0001*
Model evaluation score, mean (SD)	194	1.8 (1.5)	2.0 (1.4)	1.9 (1.6)	2.2 (1.4)	1.8 (1.6)	1.7 (1.7)	1.9 (1.4)	1.6 (1.4)	2.0 (1.6)	1.6 (1.3)	<0.0001*
Predictive Model Use Case												
Predicting health trajectories or risks for inpatients, n (%)												
Yes		1812 (58.6)	83 (63.8)	179 (57.7)	355 (71.4)	107 (55.7)	214 (46.6)	279 (65.0)	268 (54.3)	144 (52.2)	183 (60.0)	<0.0001*
Identify high risk outpatients to inform follow-up care, n (%)												
Yes		1608 (52.0)	51 (39.2)	165 (53.2)	298 (60.0)	96 (50.0)	156 (34.0)	269 (62.7)	253 (51.2)	144 (52.2)	176 (57.7)	<0.0001*
Monitor health, n (%)												

Yes		648 (21.0)	26 (20.0)	95 (30.6)	88 (17.7)	37 (19.3)	74 (16.1)	134 (31.2)	47 (9.5)	64 (23.2)	83 (27.2)	<0.0001*
Recommend treatments, n (%)												
Yes		830 (26.8)	22 (16.9)	141 (45.5)	144 (29.0)	38 (19.8)	80 (17.4)	156 (36.4)	104 (21.1)	80 (29.0)	65 (21.3)	<0.0001*
Simplify or automate billing procedures, n (%)												
Yes		673 (21.8)	17 (13.1)	77 (24.8)	116 (23.3)	20 (10.4)	66 (14.4)	118 (27.5)	99 (20.0)	37 (13.4)	123 (40.3)	<0.0001*
Facilitate scheduling, n (%)												
Yes		903 (29.2)	21 (16.2)	115 (37.1)	193 (38.8)	29 (15.1)	108 (23.5)	138 (32.2)	137 (27.7)	71 (25.7)	91 (29.8)	<0.0001*
Other operational process optimization, n (%)												
Yes		465 (15.0)	19 (14.6)	44 (14.2)	109 (21.9)	NA	54 (11.8)	100 (23.3)	47 (9.5)	58 (21.0)	31 (10.2)	<0.0001*
Other clinical use case, n (%)												
Yes		580 (18.8)	21 (16.2)	53 (17.1)	124 (24.9)	22 (11.5)	116 (25.3)	101 (23.5)	50 (10.1)	67 (24.3)	26 (8.5)	<0.0001*
Model Developer												
EHR developer, n (%)												
Yes		1597 (51.6)	76 (58.5)	154 (49.7)	294 (59.2)	92 (47.9)	173 (37.7)	275 (64.1)	256 (51.8)	118 (42.8)	159 (52.1)	<0.0001*
Third-party developer, n (%)												
Yes		1051 (34.0)	36 (27.7)	111 (35.8)	228 (45.9)	34 (17.7)	119 (25.9)	172 (40.1)	88 (17.8)	117 (42.4)	146 (47.9)	<0.0001*
Self-development, n (%)												
Yes		978 (31.6)	26 (20.0)	107 (34.5)	182 (36.6)	36 (18.8)	103 (22.4)	137 (31.9)	142 (28.7)	113 (40.9)	132 (43.3)	<0.0001*
Public domain, n (%)												
Yes		84 (2.7)	NA	NA	19 (3.8)	NA	NA	30 (7.0)	15 (3.0)	NA	NA	<0.0001*
Model Evaluation												
Model accuracy, n (%)												
All models		876 (28.3)	37 (28.5)	99 (31.9)	207 (41.6)	62 (32.3)	126 (27.5)	119 (27.7)	75 (15.2)	104 (37.7)	47 (15.4)	<0.0001*
Most models		332 (10.7)	26 (20.0)	26 (8.4)	45 (9.1)	12 (6.2)	29 (6.3)	70 (16.3)	66 (13.4)	16 (5.8)	42 (13.8)	

Some models		209 (6.8)	NA	25 (8.1)	29 (5.8)	NA	21 (4.6)	23 (5.4)	NA	30 (10.9)	NA	
Few models		105 (3.4)	NA	14 (4.5)	26 (5.2)	NA	20 (4.4)	15 (3.5)	NA	10 (3.6)	NA	
Do not know/None/No response		1570 (50.8)	63 (48.5)	146 (47.1)	190 (38.2)	101 (52.6)	263 (57.3)	202 (47.1)	309 (62.6)	116 (42.0)	180 (59.0)	
Model bias, n (%)												
All models		652 (21.1)	29 (22.3)	72 (23.2)	152 (30.6)	58 (30.2)	115 (25.1)	76 (17.7)	42 (8.5)	77 (27.9)	31 (10.2)	<0.0001*
Most models		249 (8.1)	19 (14.6)	11 (3.5)	37 (7.4)	NA	36 (7.8)	62 (14.5)	58 (11.7)	16 (5.8)	NA	
Some models		319 (10.3)	NA	44 (14.2)	41 (8.2)	NA	27 (5.9)	43 (10.0)	52 (10.5)	NA	NA	
Few models		219 (7.1)	NA	24 (7.7)	52 (10.5)	15 (7.8)	15 (3.3)	28 (6.5)	27 (5.5)	NA	NA	
Do not know/None/No response		1653 (53.5)	68 (52.3)	159 (51.3)	215 (43.3)	105 (54.7)	266 (58.0)	220 (51.3)	315 (63.8)	130 (47.1)	175 (57.4)	

Table 1. Descriptive statistics of hospitals on AI model implementation characteristics across census division.

Census division abbreviations: NE=New England, MA=Mid Atlantic, SA=South Atlantic, ENC=East North Central, WNC=West North Central, ESC=East South Central, WSC=West South Central, MTN=Mountain, PAC=Pacific. State abbreviations: AL=Alabama, AK=Alaska, AZ=Arizona, AR=Arkansas, CA=California, CO=Colorado, CT=Connecticut, DE=Delaware, FL=Florida, GA=Georgia, HI=Hawaii, ID=Idaho, IL=Illinois, IN=Indiana, IA=Iowa, KS=Kansas, KY=Kentucky, LA=Louisiana, ME=Maine, MD=Maryland, MA=Massachusetts, MI=Michigan, MN=Minnesota, MS=Mississippi, MO=Missouri, MT=Montana, NE=Nebraska, NV=Nevada, NH=New Hampshire, NJ=New Jersey, NM=New Mexico, NY=New York, NC=North Carolina, ND=North Dakota, OH=Ohio, OK=Oklahoma, OR=Oregon, PA=Pennsylvania, RI=Rhode Island, SC=South Carolina, SD=South Dakota, TN=Tennessee, TX=Texas, UT=Utah, VT=Vermont, VA=Virginia, WA=Washington, WV=West Virginia, WI=Wisconsin, WY=Wyoming.

Model scores are defined in Table S1 and variables in Table S2. Cells with counts below 10 were suppressed to prevent re-identification through cross-tabulation; neighboring cells were also suppressed as needed. P-values were calculated using chi-square tests for categorical variables and ANOVA for continuous variables.

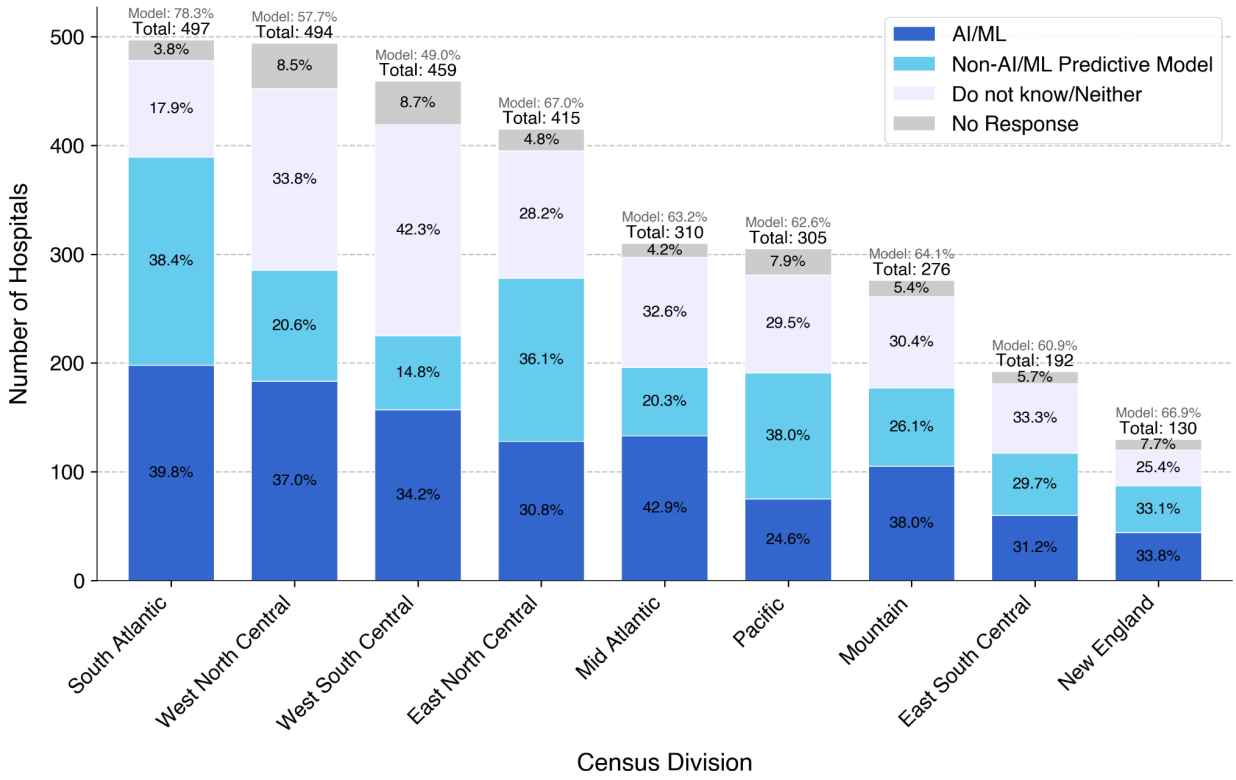

Figure 1. Hospital distribution by AI/ML implementation level across U.S. census divisions.

Stacked bars show the proportion of hospitals with no predictive models (dark blue), non-AI predictive models (light blue), and AI-based predictive models (lavender). The gray bar at the top represents hospitals with unknown model use. Total number of hospitals and overall model adoption rate are shown above each bar. AI/ML implementation varies substantially across regions, with the South Atlantic and East South Central divisions having the highest and lowest model adoption rates, respectively.

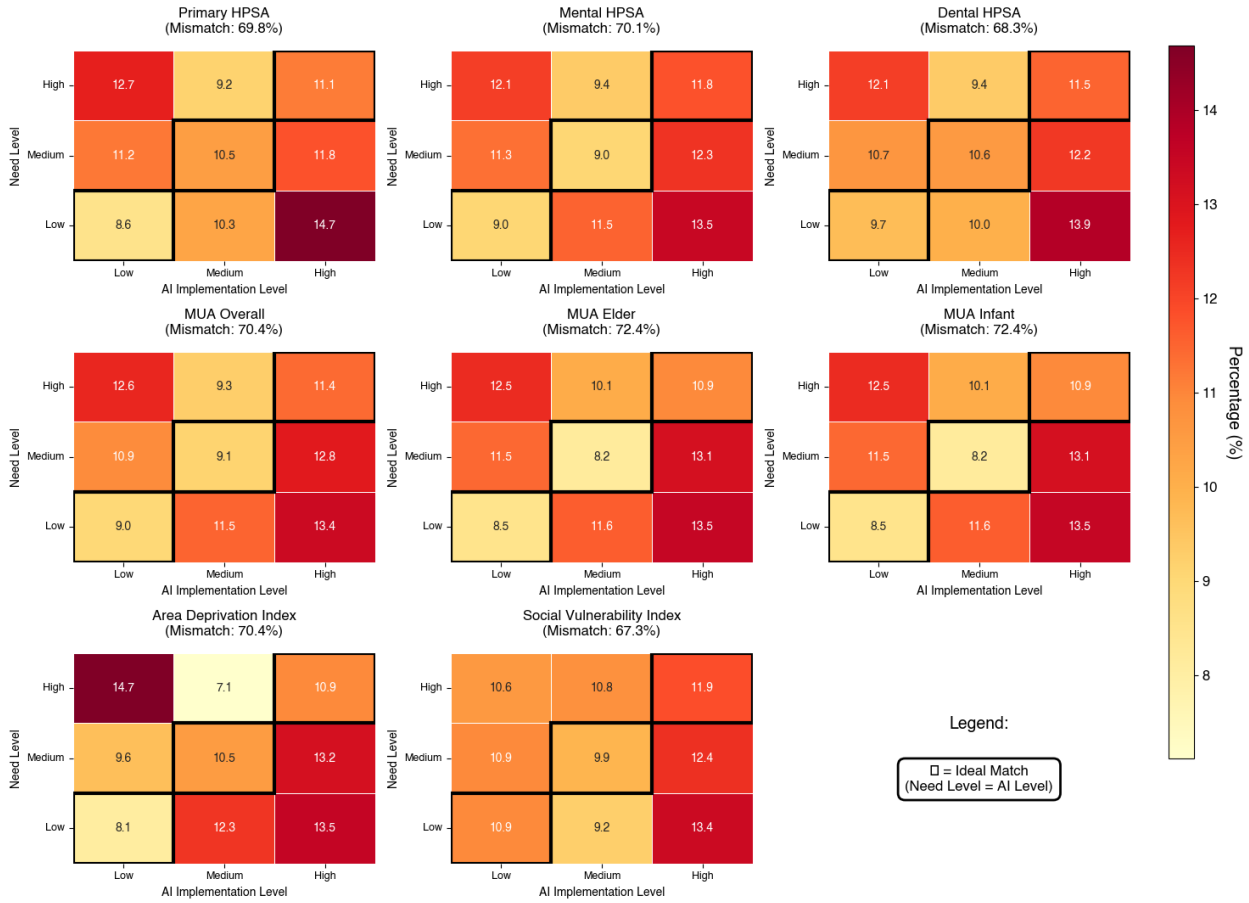

Figure 2. Alignment between healthcare need and AI implementation.

AI Artificial intelligence; HPSA Healthcare professional shortage area; MUA Medical underservice area
Each heatmap compares AI implementation levels (x-axis) to healthcare need levels (y-axis) for different indicators, including HPSA scores, MUA scores, ADI, and SVI. Darker red indicates a higher percentage of hospitals within each category. Cells outlined in bold represent ideal alignment, where the need level matches the AI implementation level. Mismatch rates, defined as the proportion of hospitals where need and AI implementation are not aligned, ranged from 67.3% to 72.4%. While primary, mental, and dental HPSA scores are separate metrics, MUA Elder and MUA Infant are components of the overall MUA score.

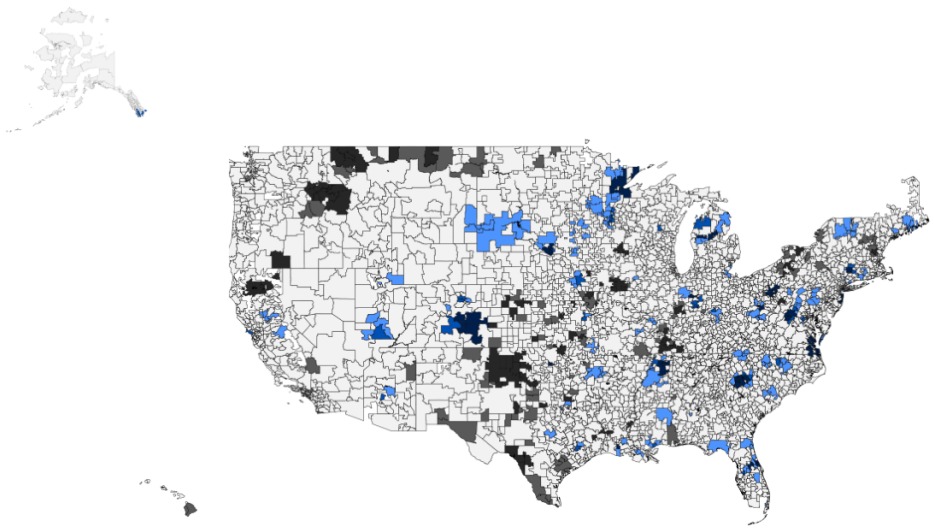

Figure 3. Geographic distribution of AI implementation hotspots and coldspots across U.S. hospital service areas.

Black areas represent coldspots (low AI implementation), and blue areas represent hotspots (high AI implementation) identified using Getis-Ord Gi* statistics.