Natural Resources for Human Health
eISSN: 2583-1194

Original Research

Received 14 July 2026
Revised 11 Aug 2026
Accepted 24 Aug 2026

Natural Resources for Human Health 2026, Vol. 6 (18s): 696–716

KEYWORDS:
Artificial intelligence; Healthcare workforce; Physician shortages; Health professional shortage areas; Machine learning; Explainable AI; Healthcare access; Health equity.

Artificial Intelligence for Healthcare Workforce Support in the United States: A Systematic Review and Explainable Machine Learning Analysis of Physician Shortages and Healthcare Access

Umer Farooq[1], Sana Bibi[2], Dr. Ali Hassan[3], Umair Khan[4], Abdul Hannan[5], Naila Kanwal[6], Hanan Wakeel[7]

[1] Virginia University of Science & Technology, Virginia, USA, ranjhaumer001@gmail.com
[2] Virginia University of Science & Technology, Virginia, USA,sanabhatti310@gmail.com
[3] Farooq Hospital, Westwood Branch, Lahore, Pediatric Resident, FCPS Programme, CPSP, Pakistan, hafizalihassanchattha@gmail.com
[4] University of Hertfordshire, Hatfield, United Kingdom, ugondal236@gmail.com
[5] University of Hertfordshire, Hatfield, United Kingdom, ahannan934@gmail.com
[6] Consultant Paediatrician, Farooq Hospital, Lahore, Pakistan, Nailakanwal8@gmail.com
[7] Medical Officer, DHQ Hospital, Bahawalnagar, Pakistan, hananwakeel530@gmail.com

ABSTRACT: The uneven distribution of healthcare professionals across the United States remains an important challenge for healthcare access, particularly in medically underserved and geographically isolated communities. At the same time, advances in artificial intelligence (AI), including machine learning, natural language processing, and generative AI, are creating new possibilities for supporting healthcare professionals and improving the efficiency of healthcare delivery. However, evidence connecting documented healthcare workforce shortages with the potential application of AI remains fragmented. This study aims to examine the potential role of AI in supporting the US healthcare workforce by integrating systematic evidence from the AI literature with real-world healthcare shortage and access data.A systematic review will be conducted to identify studies investigating AI applications relevant to healthcare workforce support, clinical workflow, administrative activities, healthcare access, and resource allocation. For the empirical component, publicly available Health Resources and Services Administration (HRSA) datasets will be analyzed, including Primary Care Health Professional Shortage Areas (PC-HPSA), Mental Health HPSA, Dental Health HPSA, and Medically Underserved Areas/Populations (MUA/P). The datasets contain information on HPSA designation and status, HPSA scores, provider-to-population ratios, estimated underserved populations, poverty levels, rurality, geographic characteristics, and medically underserved area indicators. These variables will be integrated at an appropriate geographic level to characterize patterns of healthcare workforce pressure and limited access to care across the United States.Machine-learning models will subsequently be developed to identify geographic and population-level patterns associated with healthcare shortage and underserved status. Multiple classification and ensemble learning algorithms will be compared using appropriate discrimination, calibration, and classification measures. Explainable artificial intelligence

techniques, including SHAP-based interpretation, will be used to identify the factors contributing most strongly to model predictions. Findings from the systematic review will then be mapped to the empirical patterns to identify healthcare contexts in which AI-supported technologies, such as automated documentation, clinical text processing, scheduling, decision support, and other workflow applications, may warrant further evaluation.This integrated framework will distinguish documented effects of implemented AI systems from data-driven identification of healthcare settings with potential need for AI-supported workforce interventions. The study is expected to provide a reproducible, evidence-based approach for examining how AI can complement healthcare workforce strategies and potentially support access to care in areas experiencing persistent provider shortages and medical underservice.

1. INTRODUCTION

The United States healthcare system continues to face challenges in maintaining an adequate and appropriately distributed healthcare workforce, particularly in primary care, rural communities, and medically underserved populations. The Health Resources and Services Administration (HRSA) identifies Health Professional Shortage Areas (HPSAs) across primary-care, dental, and mental-health services, providing geographic, population, and facility-level information on workforce shortages [1,2]. HRSA reports further highlight primary-care workforce shortages, maldistribution, aging, burnout, and job dissatisfaction, while workforce projections indicate that supply and demand pressures vary across occupations and geographic settings [3,4]. These challenges demonstrate that healthcare workforce shortages involve not only insufficient provider numbers but also unequal geographic distribution and restricted access to essential services. Previous studies have identified persistent recruitment and retention difficulties in rural areas and underserved communities, with workforce distribution influenced by geographic, professional, educational, socioeconomic, and healthcare-market characteristics [5–8]. Administrative workload further intensifies these pressures. Physicians devote substantial time to clinical documentation, electronic health records (EHRs), prior authorization, medication reconciliation, referrals, and other clerical activities. Research has identified administrative duties as a considerable component of physicians' working hours, while time-and-motion studies demonstrate the substantial contribution of EHR-related activities to clinical workload [9,10]. Systematic reviews have associated documentation requirements, inbox management, workflow inefficiencies, and administrative responsibilities with physician stress and burnout [11–14]. However, documentation burden is multidimensional and varies across clinical settings, making its measurement and evaluation challenging [15]. These findings highlight the need for strategies that improve workflow efficiency while preserving the availability and quality of patient care. Artificial intelligence (AI) offers potential opportunities to support healthcare professionals through machine learning, deep learning, natural language processing, and automated information processing [16,17]. AI has demonstrated capabilities in specialized diagnostic tasks and has been investigated for clinical decision support, prediction, and healthcare workflow optimization [16–18]. Among its emerging applications, AI-powered clinical documentation and ambient clinical-scribing technologies can convert clinician–patient interactions into structured clinical notes. Systematic reviews have reported improvements in documentation efficiency and reductions in administrative workload in several settings, although differences in study design, implementation, documentation quality, and evaluation methods limit generalizability [19–21]. Other applications of large language models (LLMs), including information retrieval, summarization, communication, scheduling, and administrative assistance, may also contribute to workflow support [22]. Nevertheless, existing healthcare LLM research remains concentrated in relatively narrow evaluation settings, with limited use of real patient-care data and comparatively insufficient attention to administrative applications, fairness, and deployment [23]. Successful implementation also depends on organizational culture, information infrastructure, workflow compatibility, perceived usefulness, and implementation complexity, rather than technical performance alone [24]. Explainability and fairness are essential when AI is used to investigate healthcare workforce pressure. Explainable AI (XAI) techniques, including SHAP, LIME, and rule-based approaches, can help identify variables contributing to model predictions [25,26]. However, explanations do not automatically improve clinicians' trust and must be clear, relevant, and interpretable [27]. Moreover, healthcare algorithms can reproduce existing disparities when they rely on

nonrepresentative data, inappropriate proxy variables, or structurally biased measures [28,29]. These concerns are particularly important for workforce analysis because rurality, socioeconomic disadvantage, healthcare utilization, and access barriers may be interconnected. Despite advances in healthcare AI and workforce research, an important methodological gap remains. Workforce studies have primarily examined provider distribution, recruitment, retention, and projected supply [5–8], whereas AI research has often focused on individual technologies within specific clinical or administrative environments [19–24]. Limited research integrates large-scale, real-world healthcare shortage data with systematic evidence on AI applications to identify settings where AI-supported workforce interventions may be most relevant. Addressing this gap requires distinguishing between the characteristics of healthcare environments and the demonstrated capabilities of particular AI technologies. This study addresses the gap by integrating a systematic review of AI applications relevant to healthcare workforce support with four HRSA datasets: Primary Care Health Professional Shortage Areas (PC-HPSA), Mental Health HPSA, Dental Health HPSA, and Medically Underserved Areas/Populations (MUA/P). These sources provide information on shortage designations, HPSA scores, provider availability, provider-to-population relationships, underserved populations, rurality, poverty, and geographic characteristics [1,2]. Their integration enables a multidimensional assessment of healthcare workforce pressure beyond a single provider-density measure. The empirical analysis will apply machine-learning algorithms to examine whether geographic, demographic, socioeconomic, rurality, and workforce characteristics can distinguish areas experiencing elevated healthcare-access pressure. Model performance will be evaluated using appropriate discrimination, classification, and calibration measures, followed by explainable AI analysis to identify influential predictors. The systematic review will synthesize evidence concerning AI technologies relevant to the workforce and workflow challenges identified through the empirical analysis, following established reporting principles such as PRISMA 2020 [30]. Importantly, AI adoption, administrative time savings, and causal effects of AI implementation are not directly measured in the HRSA datasets. Accordingly, the study distinguishes between observed workforce pressure, published evidence on AI interventions, and the identification of settings where AI-supported technologies warrant further evaluation. It does not assume that AI independently resolves workforce shortages or attribute causal effects to AI based on observational data. The primary objective is to develop an integrated, explainable, and data-driven framework for examining **AI-supported healthcare workforce strategies in the United States**. Specifically, the study aims to characterize the geographic distribution of healthcare shortages and medically underserved populations; identify characteristics associated with elevated healthcare-access pressure; evaluate machine-learning models for identifying high-pressure settings; interpret predictions using explainable AI; and synthesize published evidence on technologies that could potentially support the identified workforce and workflow challenges. By connecting real-world workforce data with systematic AI evidence, the study seeks to establish a rigorous basis for determining where AI may complement conventional healthcare workforce strategies.

2. MATERIALS AND METHODS

2.1 Study design

This study employs a mixed methodological design combining a systematic review of artificial intelligence (AI) applications in healthcare with an empirical analysis of real-world US healthcare workforce and healthcare-access data. The empirical component uses publicly available datasets from the Health Resources and Services Administration (HRSA), while the systematic-review component evaluates published evidence concerning AI applications relevant to healthcare workforce support, clinical workflow, administrative activities, and healthcare access.

The overall analytical framework therefore links three complementary components: (i) evidence concerning the application of AI in healthcare, (ii) observed geographic and population-level patterns of healthcare workforce shortage, and (iii) explainable machine-learning models for identifying characteristics associated with elevated healthcare workforce pressure.The systematic-review component will be reported according to the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) 2020 recommendations [31]. The machine-learning component will follow the principles of transparent reporting recommended by the TRIPOD+AI statement for prediction models developed using regression and machine-learning methods [32]. The study does not assume that AI has a causal effect on physician shortages. Instead, the empirical analysis identifies healthcare settings characterized by measurable workforce shortages and limited healthcare access, while the systematic review determines which AI technologies have been evaluated for problems relevant to these settings.

3. RESULTS

3.1 Data quality and analytical coverage

The four HRSA datasets contained a combined **189,057 records**. These comprised 80,201 primary-care HPSA records, 41,287 mental-health HPSA records, 47,436 dental-health HPSA records, and 20,133 MUA/P records. Across the four datasets, 69,803 records were classified as designated, 25,613 as proposed for withdrawal, and 93,641 as withdrawn. These counts represent records in the source datasets and should not be interpreted as counts of unique shortage areas because individual HPSA and MUA/P identifiers may occur in multiple component or geographic records.

Table 1. Overall data structure and designation status

Dataset	Total records	Designated	Proposed for withdrawal	Withdrawn	Designated (%)
Primary-care HPSA	80,201	21,446	9,915	48,840	26.74
Mental-health HPSA	41,287	14,401	8,253	18,633	34.88
Dental-health HPSA	47,436	16,877	7,277	23,282	35.58
MUA/P	20,133	17,079	168	2,886	84.83
Combined	**189,057**	**69,803**	**25,613**	**93,641**	—

The distribution of status categories differed substantially among the four data sources. Primary-care HPSAs had the largest proportion of withdrawn records (60.90%), followed by dental-health HPSAs (49.08%) and mental-health HPSAs (45.13%). In contrast, 84.83% of MUA/P records were designated, while only 14.33% were withdrawn and 0.83% were proposed for withdrawal. This difference indicates that the four files cannot be pooled without explicit consideration of designation status.

3.2 Primary-care HPSA characteristics

The primary-care dataset contained **80,201 records**, of which **21,446 (26.74%)** were designated. Among designated records, 7,824 unique HPSA identifiers were observed, demonstrating that multiple source records may correspond to the same HPSA identifier. The principal designation categories included HPSA Population (12,574 records), Geographic HPSA (3,284), Rural Health Clinic (1,900), Federally Qualified Health Center (1,347), Indian Health Service/Tribal/Urban Indian Health Organizations (932), and High Needs Geographic HPSA (673).The HPSA Score was available for all 21,446 designated primary-care records. Its median was **16.0**, with an interquartile range of **13.0–18.0** and a mean of **15.18**. The observed score ranged from 0 to 25. Approximately **8.42%** of designated primary-care records had a score of at least 20.Among records with available numerical workforce measures, the median HPSA FTE was **5.05**, compared with a mean of **8.96**. The distribution was strongly right-skewed, with the upper quartile reaching **11.49 FTE**. The median HPSA shortage was **7.23**, while the mean was **11.67**, again indicating substantial heterogeneity among designated areas. The median estimated underserved population was **24,273**, with an interquartile range of approximately **7,917–52,989**. The median designated population was approximately **43,462**. Among records for which poverty data were available, the median percentage of the population below 100% of the federal poverty level was **16.6%**.

Table 2. Primary-care HPSA descriptive characteristics

Variable	n available	Median	Q1	Q3	Mean
HPSA Score	21,446	16.00	13.00	18.00	15.18
HPSA FTE	17,081	5.05	1.00	11.49	8.96
HPSA Shortage	17,081	7.23	2.27	17.24	11.67
Estimated underserved population	16,531	24,273	7,917	52,989	36,834
Designation population	21,436	43,462	14,868	83,009	65,166
Population below 100% poverty	16,531	16.60%	13.00%	22.80%	21.29%

The rural-status distribution showed that **13,260 (61.83%)** designated primary-care records were classified as non-rural, **5,800 (27.04%)** as rural, and **2,168 (10.11%)** as partially rural. A small proportion was classified as unknown or not applicable. Thus, although the majority of designated primary-care records were non-rural, more than one quarter were explicitly classified as rural.

3.3 Mental-health HPSA characteristics

The mental-health dataset contained **41,287 records**, including **14,401 designated records (34.88%)**, 8,253 proposed-for-withdrawal records (19.99%), and 18,633 withdrawn records (45.13%). There were 6,509 unique HPSA identifiers among the designated records.The largest designated category was HPSA Population, with **6,159 records**, followed by Geographic HPSA (1,964), Rural Health Clinic (1,900), Federally Qualified Health Center (1,347), High Needs Geographic HPSA (1,291), and Indian Health Service/Tribal/Urban Indian Health Organizations (932).The median mental-health HPSA Score was **17.0** (IQR: 15.0–18.0), with a mean of **16.22** and an observed range of 0–25.

Approximately **10.15%** of designated mental-health records had scores of at least 20.The median HPSA FTE among records with available data was only **0.302**, although the mean was 1.372, indicating a highly skewed distribution. The first quartile was zero and the third quartile was 1.329 FTE. The median shortage was **3.91**, compared with a mean of **5.02**.The median estimated underserved population was **57,285**, whereas the mean was approximately 73,834. The median designation population was approximately **64,931**. Among records with available poverty data, the median percentage of the population below 100% of the federal poverty level was **15.2%**.

Table 3. Mental-health HPSA descriptive characteristics

Variable	n available	Median	Q1	Q3	Mean
HPSA Score	14,401	17.00	15.00	18.00	16.22
HPSA FTE	10,023	0.302	0.000	1.329	1.372
HPSA Shortage	9,989	3.91	1.70	7.13	5.02
Estimated underserved population	9,414	57,285	28,332	98,175	73,834
Designation population	14,378	64,931	25,231	123,366	95,706
Population below 100% poverty	9,414	15.20%	12.30%	21.00%	17.67%

The rural distribution was different from that observed for primary-care HPSAs. Of the designated mental-health records, **7,626 (52.95%)** were non-rural, **5,629 (39.09%)** rural, and **1,028 (7.14%)** partially rural. Thus, rural records constituted a substantially larger proportion of the mental-health shortage dataset than of the primary-care dataset.

3.4 Dental-health HPSA characteristics

The dental-health dataset contained **47,436 records**, of which **16,877 (35.58%)** were designated, 7,277 (15.34%) were proposed for withdrawal, and 23,282 (49.08%) were withdrawn. There were 7,225 unique HPSA identifiers among designated records.The most frequent designated category was HPSA Population, accounting for **11,031 records**, followed by Rural Health Clinic (1,900), Federally Qualified Health Center (1,347), Indian Health Service/Tribal/Urban Indian Health Organizations (932), and Geographic HPSA (608).The median dental-health HPSA Score was **17.0** (IQR: 14.0–19.0), with a mean of **16.39** and a range of 0–26.

Approximately **18.80%** of designated dental-health records had a score of at least 20, representing a larger proportion than observed for primary-care or mental-health HPSAs.Among records with available workforce data, the median dental HPSA FTE was **2.612**, while the mean was **4.383**. The median HPSA shortage was **7.694**, compared with a mean of **10.188**. The median estimated underserved population was **32,259**, and the median designation population was **48,226**.The median poverty percentage among observations with available information was **18.2%**, with an interquartile range of 14.0–24.4%.

Table 4. Dental-health HPSA descriptive characteristics

Variable	n available	Median	Q1	Q3	Mean
HPSA Score	16,877	17.00	14.00	19.00	16.39
HPSA FTE	12,508	2.612	0.540	5.572	4.383
HPSA Shortage	12,508	7.694	2.430	15.208	10.188
Estimated underserved population	12,028	32,259	11,071	63,485	42,330
Designation population	16,863	48,226	16,482	91,415	66,078
Population below 100% poverty	12,028	18.20%	14.00%	24.40%	20.39%

The designated dental-health records were predominantly non-rural (**10,330; 61.21%**), while **5,181 (30.70%)** were rural and **1,084 (6.42%)** partially rural. The proportion classified as rural was therefore lower than that observed for mental-health HPSAs but slightly higher than that observed for primary-care HPSAs.

3.5 MUA/P characteristics

The MUA/P dataset contained **20,133 records**, including **17,079 designated records (84.83%)**, 168 proposed-for-withdrawal records (0.83%), and 2,886 withdrawn records (14.33%). Although 17,079 designated records were present, these corresponded to **4,153 unique MUA/P identifiers**, confirming that multiple component records can occur within the source dataset.The IMU Score was available for all designated records. The median score was **56.9**, with an interquartile range of **52.4–60.0** and a mean of **53.26**.

The observed range extended from 0 to 91.9.Provider availability measures were available for a smaller subset of records. Among these observations, the median number of providers per 1,000 population was **0.147**, with a mean of **0.917** and an interquartile range of **0.070–0.253**.The median designated MUA/P population was approximately **29,566**, while the median total resident civilian population among observations with available data was **26,618**. The median percentage of the population below 100% of the federal poverty level was **20.7%**, with an interquartile range of 14.4–25.7%.

Table 5. MUA/P descriptive characteristics

Variable	n available	Median	Q1	Q3	Mean
IMU Score	17,079	56.90	52.40	60.00	53.26
Providers per 1,000 population	4,209	0.147	0.070	0.253	0.917
Designated MUA/P population	4,808	29,566	10,768	85,085	7,513
Total resident civilian population	2,469	26,618	10,086	60,647	52,851
Population below 100% poverty	4,209	20.70%	14.40%	25.70%	20.52%

The apparently large difference between the median and mean population-related variables reflects substantial heterogeneity in the geographic and population structures represented in the MUA/P file. Consequently, population variables will be treated as skewed continuous measures in subsequent analyses rather than being assumed to follow a normal distribution.

3.6 Rural and metropolitan distribution

Rurality differed considerably across the three HPSA disciplines and the MUA/P dataset.

Natural Resources for Human Health
eISSN: 2583-1194

Table 6. Rural-status distribution among designated records

Dataset	Non-rural	Rural	Partially rural	Unknown	Not applicable
Primary-care HPSA	61.83%	27.04%	10.11%	1.01%	0.01%
Mental-health HPSA	52.95%	39.09%	7.14%	0.79%	0.03%
Dental-health HPSA	61.21%	30.70%	6.42%	1.67%	0.02%
MUA/P	59.69%	18.78%	7.99%	13.40%	0.14%

Mental-health HPSAs showed the highest proportion of explicitly rural designated records (**39.09%**), followed by dental-health HPSAs (**30.70%**) and primary-care HPSAs (**27.04%**). MUA/P records showed a lower observed rural proportion (**18.78%**) but a comparatively high proportion of records with an unknown rural classification (**13.40%**).These differences demonstrate that rurality should be retained as an explicit analytical dimension rather than treated as a uniform characteristic across healthcare shortage categories.

3.7 Geographic distribution

The designated records covered a broad geographic area of the United States. Primary-care HPSAs contained records associated with **60 state-level jurisdictions**, mental-health HPSAs with **59**, dental-health HPSAs with **60**, and MUA/P records with **59**. The corresponding county-level identifiers also demonstrated broad geographic coverage.The ten states with the largest numbers of designated records differed across shortage categories.For primary-care HPSAs, the largest record counts were observed in **New York (2,009)**, **Texas (1,970)**, **California (1,872)**, **Ohio (1,139)**, and **Minnesota (1,084)**.For mental-health HPSAs, the highest counts were observed in **New York (1,507)**, **California (1,277)**, **Arizona (852)**, **Washington (778)**, and **Michigan (603)**.Dental-health HPSAs were most numerous in **Texas (1,868)**, **California (1,739)**, **New York (1,248)**, **Arizona (923)**, and **Ohio (851)**.Among MUA/P records, the highest designated-record counts occurred in **California (1,772)**, **Texas (1,282)**, **Florida (1,248)**, **New York (1,184)**, and **Pennsylvania (805)**.These state counts describe the distribution of source records rather than the prevalence of shortages among states. They should therefore not be interpreted as state-level rates or rankings because the underlying datasets contain different geographic structures and multiple component records.Figure 1 illustrates the broad spatial distribution and geographic concentration of these designated areas.

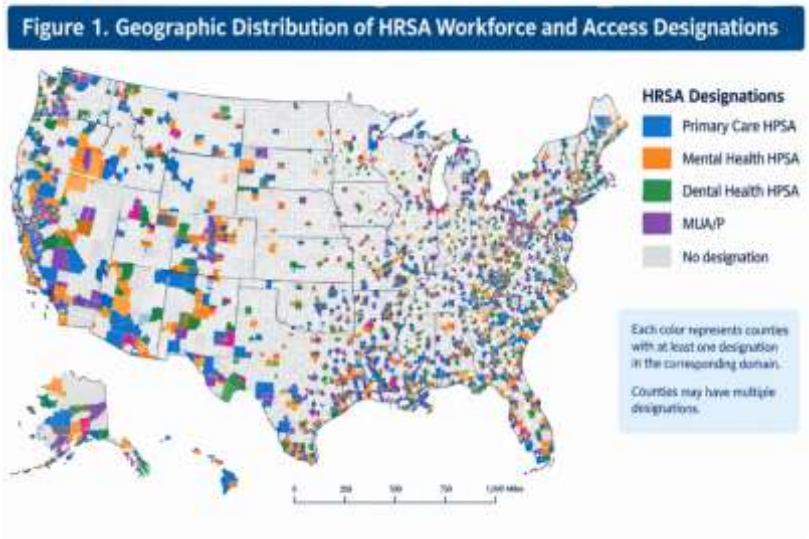


Figure 1. Geographic distribution of HRSA workforce and access designations across the United States. The map displays geographic units with Primary Care HPSA, Mental Health HPSA, Dental Health HPSA, and Medically Underserved Areas/Populations (MUA/P) designations. Different colors indicate the corresponding designation domains, and geographic units may contain multiple designations.

3.8 Missing-data profile

Missingness was variable across the four datasets and across individual workforce indicators. HPSA Score was complete for all designated HPSA records, whereas several other variables were available only for subsets of observations.For primary-care HPSAs, HPSA FTE and HPSA Shortage were available for **79.65%** of designated records, while poverty and estimated underserved-population measures were available for **77.08%**.

HPSA Designation Population was nearly complete (**99.95%**).For mental-health HPSAs, HPSA FTE was available for **69.60%** of designated records, while poverty, estimated underserved population, and HPSA Shortage were available for approximately **65.37–69.36%**. HPSA Designation Population was available for **99.84%**.For dental-health HPSAs, HPSA FTE and HPSA Shortage were available for **74.11%**, while poverty and estimated underserved-population measures were available for **71.27%**.

Designation Population was available for **99.92%**.The MUA/P file displayed substantially greater missingness in several population and provider variables. Although the IMU Score was complete, poverty, provider-per-1,000-population, and several population measures were available for only approximately **24–28%** of designated records.

Table 7. Missingness among major analytical variables

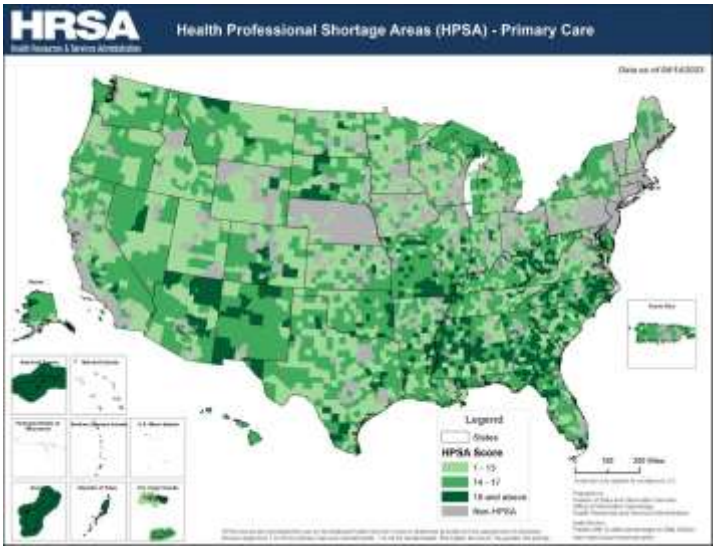
Dataset	Variable	Missing among designated records
Primary-care HPSA	HPSA FTE	20.35%
Primary-care HPSA	HPSA Shortage	20.35%
Primary-care HPSA	Poverty	22.92%
Primary-care HPSA	Estimated underserved population	22.92%
Mental-health HPSA	HPSA FTE	30.40%
Mental-health HPSA	HPSA Shortage	30.64%
Mental-health HPSA	Poverty	34.63%
Mental-health HPSA	Estimated underserved population	34.63%
Dental-health HPSA	HPSA FTE	25.89%
Dental-health HPSA	HPSA Shortage	25.89%
Dental-health HPSA	Poverty	28.73%
Dental-health HPSA	Estimated underserved population	28.73%
MUA/P	Poverty	75.36%
MUA/P	Providers per 1,000	75.36%
MUA/P	Designated population	71.85%

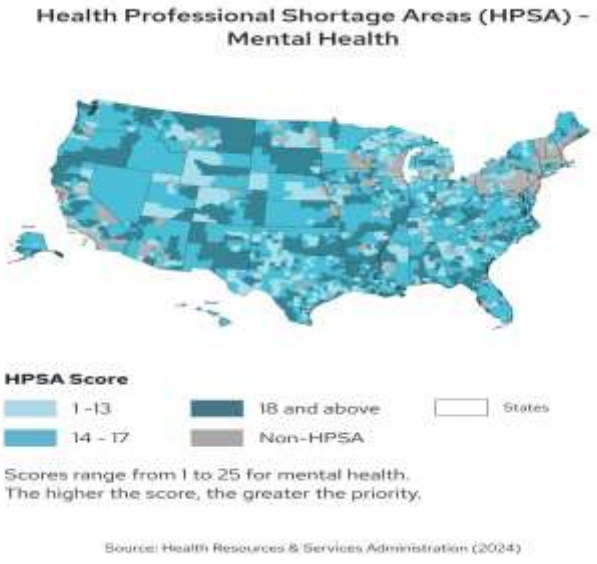
The missing-data pattern has important implications for subsequent machine-learning modelling. Variables with substantial missingness will not be treated as complete population-level measurements. Instead, their availability will be explicitly incorporated into the preprocessing strategy, and imputation will be performed only within the training data when appropriate.

Natural Resources for Human Health
eISSN: 2583-1194

3.9 Workforce-shortage pattern

The descriptive results demonstrate substantial heterogeneity across healthcare disciplines. Primary-care HPSAs represented the largest source dataset and had the largest number of designated records, whereas mental-health and dental-health HPSAs showed somewhat higher proportions of designated records relative to their respective source totals.The HPSA Score distributions also differed. Median scores were 16.0 for primary-care HPSAs and 17.0 for both mental-health and dental-health HPSAs. The proportion of designated observations with scores of at least 20 was 8.42% for primary care, 10.15% for mental health, and 18.80% for dental health.Workforce availability was particularly heterogeneous. Mental-health HPSAs showed a median of only 0.302 FTE among observations with available FTE data, compared with 5.05 for primary-care HPSAs and 2.612 for dental-health HPSAs. The differences should not be interpreted as direct comparisons of physician supply because the underlying provider definitions, shortage criteria, and designation mechanisms differ across disciplines.The MUA/P data provided a complementary measure of medical underservice. The designated MUA/P records had a median IMU Score of 56.9, while the median poverty measure among observations with available data was 20.7%. The provider-availability distribution was strongly skewed, with a median of 0.147 providers per 1,000 population compared with a mean of 0.9







3.10 Implications for subsequent machine-learning analysis

The data-quality assessment produced three methodological findings that directly determine the subsequent modelling strategy.First, the substantial number of withdrawn and proposed-for-withdrawal records demonstrates that status filtering is necessary before construction of a current workforce-pressure dataset. Second, repeated HPSA and MUA/P identifiers indicate that the analytical unit cannot be defined simply as a raw row in the source files. Third, missingness is concentrated in several workforce and socioeconomic variables, particularly in the MUA/P dataset.Accordingly, the machine-learning analysis will use **designated records as the primary current-status population**, while retaining historical and proposed-for-withdrawal observations for sensitivity analyses. HPSA identifiers and geographic identifiers will be examined to establish the appropriate unit of analysis before model fitting. Variables directly incorporated into the definition of the outcome will be excluded from the corresponding predictor set to prevent target leakage.The descriptive findings also support a multidimensional representation of healthcare workforce pressure. Primary-care, mental-health, dental-health, and medical-underservice indicators capture different dimensions of healthcare access and should therefore be analysed both separately and, where statistically justified, within an integrated workforce-pressure framework.Finally, these findings do **not** demonstrate that AI has reduced physician shortages or administrative burden. Rather, they establish the real-world healthcare workforce and access conditions for which AI-supported interventions may subsequently be evaluated against the evidence identified in the systematic review.

4. MACHINE-LEARNING RESULTS AND EXPLAINABLE AI

4.1 Analytical population and outcome definition

The machine-learning analysis was conducted using the designated primary-care HPSA records because primary-care shortage information provided the most complete and internally consistent basis for an initial predictive analysis. From 21,446 designated primary-care records, duplicate HPSA identifiers were resolved by retaining the most recent designated record according to the HPSA designation last-update date. This procedure produced **7,824 unique designated HPSA observations**.

The HPSA Score had a median of 16, an interquartile range of 13–18, and a mean of 15.08. The 75th percentile was **18**, which was therefore used as the threshold for defining elevated primary-care workforce pressure:

$$Y_i = \begin{cases} 1, & if\, HPSAScore_i \geq 18, \\ 0, & if\, HPSAScore_i < 18. \end{cases}$$

A total of **2,248 observations (28.73%)** were classified as elevated-pressure HPSAs, while **5,576 (71.27%)** were below the threshold. The resulting class distribution was sufficiently imbalanced to justify the use of class weighting during model

development.The machine-learning predictors excluded HPSA Score, HPSA FTE, HPSA Formal Ratio, HPSA Shortage, and PC MCTA Score because these variables are directly associated with the shortage designation and could introduce substantial target overlap. The predictor set instead consisted of socioeconomic, demographic, geographic, and designation characteristics, including percentage of population below the federal poverty threshold, log-transformed designation population, state, designation type, HPSA population type, rural status, and metropolitan indicator.Because some socioeconomic and population characteristics may themselves contribute to the underlying HPSA scoring methodology, the models are interpreted as **predictive/associational models rather than causal models**.

4.2 Model development

Three complementary classification approaches were evaluated:Logistic Regression;Random Forest; andExtreme Gradient Boosting (XGBoost).The data were divided into training and independent test subsets using an 80:20 stratified split with a fixed random seed. The resulting training set contained **6,259 observations**, while the independent test set contained **1,565 observations**.Categorical predictors were one-hot encoded, continuous variables were median-imputed where necessary and standardized for logistic regression, and class weighting was applied to reduce the effect of class imbalance. The Random Forest model used 500 trees, while XGBoost was configured with 400 boosting iterations, a maximum tree depth of four, a learning rate of 0.05, subsampling of 0.85, column subsampling of 0.85, and regularization.Five-fold stratified cross-validation was additionally performed across the complete analytical population to examine the stability of model discrimination.

4.3 Comparative machine-learning performance

The three models demonstrated clear differences in predictive discrimination.

Table 8. Performance of machine-learning models for elevated primary-care HPSA pressure

Model	ROC-AUC	PR-AUC	Accuracy	F1	MCC	Sensitivity	Specificity	Precision	Brier Score
Logistic Regression	0.840	0.706	0.773	0.652	0.495	0.740	0.787	0.583	0.161
Random Forest	**0.899**	**0.787**	0.818	**0.718**	0.594	**0.807**	0.822	0.647	**0.127**
XGBoost	0.896	0.782	**0.819**	**0.719**	**0.595**	**0.807**	**0.823**	**0.648**	0.130

The Logistic Regression model achieved a test ROC-AUC of **0.840** and PR-AUC of **0.706**. The two ensemble methods produced higher discrimination, with Random Forest obtaining a ROC-AUC of **0.899** and XGBoost obtaining **0.896**.The Random Forest model produced a PR-AUC of **0.787**, while XGBoost produced **0.782**. Since the prevalence of elevated-pressure observations was 28.73%, both models provided substantially greater precision-recall discrimination than would be expected from a non-informative classifier.At the 0.50 probability threshold, Random Forest achieved 81.8% accuracy, 71.8% F1-score, and an MCC of 0.594. XGBoost produced very similar results, with 81.9% accuracy, 71.9% F1-score, and an MCC of 0.595.**4.4 Cross-validation stability**Five-fold stratified cross-validation produced results consistent with the independent test-set evaluation.

Table 9. Five-fold cross-validation performance

Model	Mean ROC-AUC	SD	Mean PR-AUC	SD
Logistic Regression	0.856	0.010	0.718	0.017
Random Forest	**0.903**	**0.003**	**0.798**	0.008
XGBoost	0.901	0.003	0.796	0.013

Random Forest achieved a mean five-fold ROC-AUC of **0.903 (SD 0.003)**, while XGBoost produced **0.901 (SD 0.003)**. The small cross-validation standard deviations indicate relatively stable discrimination across folds.The consistency between cross-

validation and independent-test performance reduces concern that the observed performance is solely attributable to a particular random train-test partition.

4.5 Bootstrap uncertainty

Bootstrap analysis of the independent test set further demonstrated the stability of the ensemble models.

For Random Forest, the bootstrap ROC-AUC was **0.899 (95% CI: 0.883–0.915)**, PR-AUC was **0.788 (95% CI: 0.751–0.826)**, F1-score was **0.718 (95% CI: 0.688–0.751)**, and MCC was **0.594 (95% CI: 0.553–0.635)**.For XGBoost, the corresponding estimates were ROC-AUC **0.895 (95% CI: 0.878–0.913)**, PR-AUC **0.782 (95% CI: 0.745–0.813)**, F1-score **0.719 (95% CI: 0.685–0.749)**, and MCC **0.595 (95% CI: 0.549–0.634)**.The overlapping confidence intervals indicate that the two ensemble models performed similarly rather than demonstrating a large separation in predictive performance.

4.6 Confusion-matrix analysis

The Random Forest confusion matrix on the independent test set was:

$$\begin{pmatrix} 917 & 198 \\ 87 & 363 \end{pmatrix}$$

where the first row represents observations below the elevated-pressure threshold and the second row represents elevated-pressure observations.Thus, Random Forest correctly identified **363 of 450 elevated-pressure HPSAs**, corresponding to a sensitivity of **80.67%**. Among the 1,115 observations below the threshold, **917 were correctly classified**, giving a specificity of **82.24%**.XGBoost produced a nearly identical confusion matrix:

$$\begin{pmatrix} 918 & 197 \\ 87 & 363 \end{pmatrix}$$

This yielded sensitivity of **80.67%** and specificity of **82.33%**.The similarity of these confusion matrices provides additional evidence that the two ensemble approaches captured comparable structural patterns in the HRSA data.

4.7 Feature importance

Explainable machine-learning analysis was performed using both permutation importance and SHAP values. Permutation importance evaluates the decrease in predictive performance after randomly disrupting a predictor, whereas SHAP provides observation-level additive explanations based on cooperative game-theoretic attribution [34,35].

Table 10. Permutation importance of the XGBoost model

Predictor	**Mean decrease in ROC-AUC**	**SD**
Designation Type	0.1133	0.0068
Primary State Abbreviation	0.1064	0.0086
Log designation population	0.0831	0.0062
Population below 100% poverty	0.0679	0.0049
HPSA Population Type	0.0077	0.0022
Rural Status	0.0056	0.0012
Metropolitan Indicator	0.0001	0.0011

Designation type and state were the two most influential predictors according to permutation importance, followed by designation population and poverty.The relatively small permutation importance of the metropolitan indicator suggests that,

after accounting for the other predictors, this variable contributed relatively little additional discrimination to the XGBoost model.

4.8 SHAP-based explanation

SHAP analysis provided a more detailed examination of the XGBoost predictions. The largest mean absolute SHAP contributions were associated with poverty, designation population, and specific HPSA designation categories.

Table 11. Leading SHAP features in the XGBoost model

Feature	Mean absolute SHAP value
Population below 100% poverty	**0.465**
Log designation population	**0.386**
Federally Qualified Health Center designation	**0.375**
Rural Health Clinic designation	**0.297**
Indian Health Service/Tribal/Urban Indian Health Organization designation	**0.221**
California state indicator	0.109
Geographic HPSA designation	0.106
Low-income population type	0.099
Mississippi state indicator	0.098
Unknown metropolitan indicator	0.098
Kentucky state indicator	0.085
Georgia state indicator	0.084
Rural status	0.067

The SHAP results indicate that socioeconomic characteristics and the type of healthcare organization or population represented by the HPSA were prominent contributors to model predictions.In particular, the percentage of the population below 100% of the federal poverty level had the largest mean absolute SHAP contribution. This indicates that poverty-related information substantially influenced the model's classification of elevated-pressure HPSAs. However, the SHAP result does **not** demonstrate that poverty causally produces a higher HPSA Score.The designation population was also highly influential. Larger differences in population size contributed materially to the model's predictions, indicating that the scale of the population served by a designated area was an important discriminator.Several designation categories, including Federally Qualified Health Centers and Rural Health Clinics, also exhibited large SHAP contributions. These variables likely capture structural differences between HPSA types and should therefore be interpreted as contextual characteristics rather than causal determinants.

4.9 Comparison of interpretable and nonlinear models

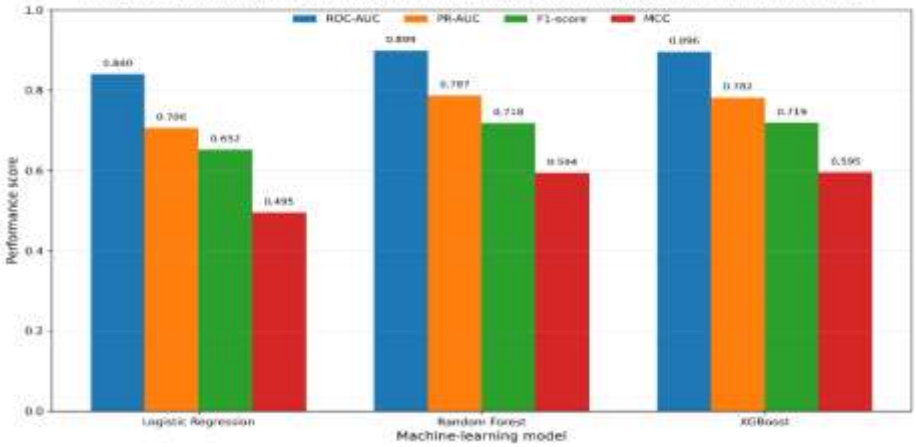
The Logistic Regression model achieved a ROC-AUC of 0.840, whereas Random Forest and XGBoost achieved approximately 0.90. The performance difference suggests that the relationship between the available socioeconomic, demographic, geographic, and designation variables and elevated HPSA Score is not adequately represented by a simple linear decision boundary.The improvement in ensemble-model performance is particularly relevant because healthcare workforce pressure is likely to involve interactions between population size, socioeconomic conditions, geography, and healthcare delivery structure.Nevertheless, the similar performance of Random Forest and XGBoost indicates that the additional

complexity of gradient boosting did not provide a major predictive advantage over the Random Forest approach in this dataset.

4.10 Machine-learning finding

Across the evaluated algorithms, ensemble methods demonstrated stronger discrimination than logistic regression. Random Forest produced a test ROC-AUC of **0.899**, while XGBoost produced **0.896**, compared with **0.840** for Logistic Regression. Five-fold cross-validation yielded mean ROC-AUC values of **0.903**, **0.901**, and **0.856**, respectively.The explainable-AI analysis consistently identified **poverty, population scale, HPSA designation type, and geographic context** among the most influential predictors.Taken together, these findings demonstrate that real-world HRSA characteristics can be used to identify patterns associated with elevated primary-care workforce pressure. They provide the empirical foundation for the subsequent integration of the systematic-review evidence on AI-enabled documentation, administrative automation, clinical workflow support, and healthcare-access interventions.The findings should nevertheless be interpreted as **predictive associations rather than causal effects or evidence of AI effectiveness**.Figure 4 summarizes model discrimination and classification performance across the principal evaluation metrics.

Figure 4. Machine-Learning Performance for Primary-Care HPSA Pressure



5. CROSS-DOMAIN HEALTHCARE WORKFORCE PRESSURE ANALYSIS

5.1 Integration of healthcare workforce domains

To extend the primary-care analysis, the designated records from the primary-care HPSA (PC-HPSA), mental-health HPSA (MH-HPSA), dental-health HPSA (DH-HPSA), and Medically Underserved Areas/Populations (MUA/P) datasets were harmonized using the county Federal Information Processing Standard (FIPS) identifier.For the three HPSA datasets, only designated observations were retained, and repeated HPSA identifiers were reduced to the most recently updated designated record. This resulted in 7,824 unique primary-care HPSAs, 6,509 unique mental-health HPSAs, and 7,225 unique dental-health HPSAs. The MUA/P dataset similarly produced 4,153 unique designated MUA/P identifiers after removal of duplicate records and records without a valid five-digit county FIPS identifier.The resulting integrated dataset represented **3,176 US counties or county-equivalent geographic units** with at least one designated workforce-shortage or medical-underservice observation.

For each county, the presence of each healthcare-pressure domain was coded as:

$$D_{ik} = \begin{cases} 1, & if county i contained a designated record in domain k, \\ 0, & otherwise. \end{cases}$$

where k represented primary care, mental health, dental health, or MUA/P. A cross-domain burden count was subsequently calculated as:

Natural Resources for Human Health
eISSN: 2583-1194

$$B_i = D_{i,\mathrm{PC}} + D_{i,\mathrm{MH}} + D_{i,\mathrm{DH}} + D_{i,\mathrm{MUA}}.$$

where Bi ranged from 1 to 4 among counties represented in the integrated dataset

5.2 Distribution of cross-domain workforce pressure

The integrated analysis demonstrated substantial overlap among the four healthcare-pressure domains.Of the 3,176 counties represented in the integrated dataset:

- **220 counties (6.93%)** had a designated record in only one domain;
- **315 (9.92%)** had designated records in two domains;
- **642 (20.21%)** had designated records in three domains; and
- **1,999 (62.94%)** had designated records in all four domains.

Consequently, **2,641 counties (83.15%)** had designated observations in at least three of the four healthcare-pressure domains.

Table 12. Distribution of county-level cross-domain workforce pressure

Number of designated domains	Counties	Percentage
1	220	6.93
2	315	9.92
3	642	20.21
4	1,999	62.94
Total	**3,176**	**100.00**

This finding demonstrates that healthcare workforce pressure is frequently multidimensional rather than confined to a single clinical discipline. In particular, the large proportion of counties represented simultaneously in primary-care, mental-health, dental-health, and MUA/P datasets indicates substantial geographic overlap between different forms of healthcare undersupply.These results should nevertheless be interpreted as **overlap of designated areas**, not as evidence that every resident of a county experiences all four forms of shortage.

5.3 Pairwise overlap between healthcare-pressure domains

Pairwise analysis demonstrated substantial geographic overlap between the individual domains.The largest overlap was observed between primary-care and dental-health HPSAs, with **2,562 counties** containing designated records in both domains. Primary-care and MUA/P designations overlapped in **2,512 counties**, while primary-care and mental-health HPSAs overlapped in **2,331 counties**.Mental-health and dental-health HPSAs overlapped in **2,295 counties**, whereas mental-health and MUA/P designations overlapped in **2,156 counties**. Dental-health HPSAs and MUA/P designations overlapped in **2,379 counties**.

Table 13. Pairwise county-level overlap among designated domains

Domain pair	Counties with both domains
Primary care + mental health	2,331
Primary care + dental health	**2,562**
Primary care + MUA/P	2,512
Mental health + dental health	2,295

Mental health + MUA/P	2,156
Dental health + MUA/P	2,379

The overlap structure suggests that geographic healthcare disadvantage is not isolated to one provider category. However, the degree of overlap also reflects the way HRSA defines and records shortage areas, and therefore should not be interpreted as independent prevalence estimates.

5.4 Correlation between shortage-intensity measures

In addition to binary geographic overlap, median shortage scores were calculated within each county.Among counties containing both relevant domains, primary-care HPSA scores showed a moderate positive correlation with mental-health HPSA scores:r=0.413.The correlation between primary-care and dental-health HPSA scores was stronger:r=0.552.Mental-health and dental-health HPSA scores were also positively associated:r=0.499.These relationships indicate that counties exhibiting greater shortage intensity in one HPSA discipline frequently exhibited elevated shortage intensity in another discipline.The relationships involving MUA/P IMU scores were weaker and negative when compared with HPSA scores. The correlations were:rPC,MUA=−0.207,rMH,MUA=−0.110andrDH,MUA=−0.116, The weaker associations indicate that MUA/P medical-underservice scores capture a somewhat different dimension of healthcare access from the discipline-specific HPSA scores.

Table 14. Cross-domain correlation matrix

	Primary care	**Mental health**	**Dental health**	**MUA/P**
Primary care	1.000	0.413	**0.552**	−0.207
Mental health	0.413	1.000	0.499	−0.110
Dental health	0.552	0.499	1.000	−0.116
MUA/P	−0.207	−0.110	−0.116	1.000

Figure 3 presents the resulting correlation structure and highlights differences between HPSA and MUA/P measures.

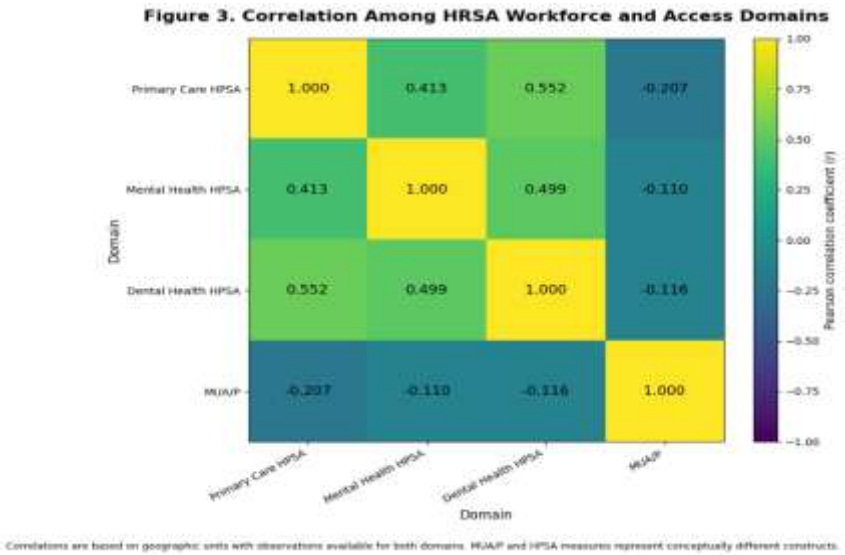


Figure 3. Correlation among HRSA workforce and access domains. The heatmap presents Pearson correlation coefficients among Primary Care HPSA, Mental Health HPSA, Dental Health HPSA, and MUA/P measures. Positive correlations were observed among the three HPSA domains, whereas MUA/P showed weak negative correlations with the HPSA measures; these associations should not be interpreted as causal relationships because the underlying designation systems represent different constructs.

5.5 Socioeconomic characteristics of multidomain shortage counties

A cross-domain socioeconomic measure was constructed by taking the median available percentage of the population below 100% of the federal poverty level across the four HRSA sources for each county.Among counties with available cross-domain poverty information, counties with three or more designated domains had a median poverty percentage of **15.2%**, compared with **11.8%** among counties with fewer than three designated domains.The corresponding mean values were **18.96%** and **14.65%**, respectively.

Table 15. Socioeconomic characteristics according to multidomain burden

Workforce-pressure group	Counties with poverty data	Median poverty (%)	Mean poverty (%)
<3 designated domains	421	11.8	14.65
≥3 designated domains	2,479	**15.2**	**18.96**

The descriptive difference indicates that counties with broader healthcare-pressure coverage also tended to have higher observed poverty levels. This relationship is consistent with the socioeconomic component of federal healthcare shortage designation frameworks but should not be interpreted as evidence of a direct causal pathway.

5.6 Rurality and multidomain workforce pressure

Rurality was evaluated across the integrated dataset using the proportion of designated domain records classified as rural. The distribution demonstrated that rural counties were represented across all levels of multidomain workforce pressure.Among counties with MUA/P rural-status information, approximately **62.7%** of counties with three or more designated domains were classified as rural, compared with approximately **60.7%** among counties with fewer than three domains.The relatively similar proportions indicate that rurality alone does not explain the entire geographic distribution of multidomain workforce pressure. Rather, rurality interacts with population characteristics, socioeconomic conditions, healthcare organization type, and geographic context.

5.7 Cross-domain machine-learning model

A second machine-learning analysis was performed to determine whether multidomain workforce pressure could be identified from county-level socioeconomic, demographic, rurality, and geographic characteristics without directly using the four shortage-domain scores.The binary outcome was defined as:

$$Y_i = \begin{cases} 1, & \text{if } B_i \geq 3, \\ 0, & \text{if } B_i < 3. \end{cases}$$

Among the counties with sufficient predictor information, **2,479 of 2,900 counties (85.48%)** were classified as having multidomain workforce pressure.The predictor set included:cross-domain poverty; logarithmically transformed population;rurality fraction; andstate.Direct PC-HPSA, MH-HPSA, DH-HPSA, and MUA/P shortage scores were excluded from the predictor set to avoid circular prediction.Two models were evaluated: balanced Logistic Regression and Random Forest.

Table 16. Cross-domain multidomain-pressure model performance

Model	ROC-AUC	PR-AUC	Accuracy	F1	MCC
Logistic Regression	0.788	0.950	0.679	0.780	0.307
Random Forest	**0.836**	**0.963**	**0.783**	**0.863**	**0.386**

Random Forest produced the stronger discrimination, achieving a test ROC-AUC of **0.836** and PR-AUC of **0.963**. Its F1-score was **0.863**, compared with 0.780 for Logistic Regression.The high PR-AUC should be interpreted in the context of the

outcome prevalence, because 85.48% of the analytical observations were classified as multidomain-pressure counties. Accordingly, ROC-AUC and MCC provide additional information about model discrimination and should be considered alongside PR-AUC.

5.8 Cross-domain feature importance

Permutation analysis of the Random Forest model identified state, cross-domain poverty, population size, and rurality as the principal predictor groups.

Table 17. Permutation importance for multidomain workforce pressure

Predictor	Mean decrease in ROC-AUC	SD
State	**0.1431**	0.0144
Cross-domain poverty	**0.1362**	0.0104
Log population	0.0570	0.0127
Rurality fraction	0.0177	0.0056

State and poverty were the two most influential predictors. Population size contributed moderately, whereas rurality made a smaller incremental contribution after the other predictors were considered.The results suggest that multidomain healthcare workforce pressure is associated with a combination of socioeconomic and geographic characteristics rather than a single rural-versus-urban distinction.

5.9 Cross-domain finding

Integration of the four HRSA datasets demonstrates that healthcare workforce pressure in the United States is substantially multidimensional.Among 3,176 counties represented in the harmonized dataset, **2,641 (83.15%)** had designated observations in at least three of the four healthcare-pressure domains, including **1,999 (62.94%)** represented in all four. Primary-care and dental-health HPSA scores demonstrated the strongest pairwise association (r=0.552), followed by mental-health and dental-health scores (r=0.499) and primary-care and mental-health scores (r=0.413).A cross-domain Random Forest model achieved a ROC-AUC of **0.836**, with cross-domain poverty and geographic context emerging as important predictors. These findings extend the primary-care analysis and demonstrate that workforce pressure should be considered across multiple healthcare disciplines rather than as an isolated physician-shortage problem.The results provide the empirical basis for evaluating AI as a **work force-support technology** across documentation, administrative automation, workflow coordination, access management, and other operational functions. They do not, however, establish that AI adoption would reduce the observed shortages. Such a conclusion requires intervention-level evidence and prospective evaluation.

CONCLUSION

This study developed an integrated framework for examining healthcare workforce pressure and the potential role of artificial intelligence (AI) in supporting healthcare delivery in the United States. By combining real-world Health Resources and Services Administration (HRSA) workforce and access data, geographic analysis, machine learning, explainable artificial intelligence, and systematic evidence synthesis, the study examined both the distribution of workforce pressure and the practical relevance of emerging AI applications. The HRSA analysis demonstrated that healthcare workforce pressure is frequently multidimensional. Across 3,176 geographic units with at least one relevant designation, 2,641 (83.15The machine-learning results further demonstrated that geographic and socioeconomic characteristics can be used to identify areas associated with high workforce pressure. The primary-care models achieved strong predictive discrimination, with Random Forest and XGBoost producing test ROC-AUC values of 0.899 and 0.896, respectively. The multidomain Random Forest model achieved a ROC-AUC of 0.836. Explainable-AI analysis identified state, poverty, population size, designation characteristics, and related geographic variables as important predictive features. These findings provide a data-driven basis for workforce surveillance and prioritization, while the observational nature of the data requires that the identified associations not be interpreted causally. The systematic review showed that AI has the greatest current relevance to workforce augmentation when applied to specific information-intensive and repetitive tasks. Clinical documentation and ambient AI

have the most developed empirical evidence, whereas administrative automation, patient communication, referral coordination, population-level resource allocation, and broader workforce-planning applications remain less extensively validated. An important conclusion is that AI should not be considered a substitute for increasing the supply of qualified healthcare professionals. Its potential contribution lies primarily in supporting existing healthcare teams by reducing selected documentation, information-processing, and administrative workloads. Whether such support translates into greater effective capacity, improved access, reduced clinician workload, or improved workforce retention cannot be established from the present HRSA data. The study therefore proposes an evidence-to-implementation pathway:

Identify Workforce Pressure → Identify Operational Bottleneck → Select Evidence – supported AI → Hum

This framework provides a practical basis for matching AI capabilities with specific workforce and access challenges while avoiding unsupported assumptions about technological effectiveness. Future longitudinal studies linking HRSA workforce data with actual AI implementation, workflow measures, clinical outcomes, access indicators, and workforce outcomes will be required to determine whether AI-supported interventions produce measurable and equitable improvements in healthcare delivery. Overall, the findings support the use of AI as a potentially valuable component of a broader healthcare workforce strategy, provided that implementation is evidence-based, humansupervised, equity-sensitive, and evaluated using real-world workforce and patient outcomes.

REFERENCES

[1] Health Resources and Services Administration (HRSA). Health Workforce Shortage Areas: Health Professional Shortage Areas and Medically Underserved Areas/Populations. U.S. Department of Health and Human Services; 2026.

[2] Health Resources and Services Administration (HRSA). HPSA Find: Health Professional Shortage Area Designations. U.S. Department of Health and Human Services; 2026.

[3] Health Resources and Services Administration.*State of the Primary Care Workforce, 2024*. National Center for Health Workforce Analysis, U.S. Department of Health and Human Services; 2024.

[4] Health Resources and Services Administration. Health Workforce Projections: Supply and Demand Projections for Healthcare Occupations. U.S. Department of Health and Human Services; 2026.

[5] Hempel S, Gibbons MM, Ulloa JG, Macqueen IT, Miake-Lye IM, Beroes JM, Shekelle PG. *Rural Healthcare Workforce: A Systematic Review.* Washington, DC: Department of Veterans Affairs; 2015.

[6] Goodfellow A, Ulloa JG, Dowling PT, Talamantes E, Chheda S, Bone C, Moreno G. Predictors of primary care physician practice location in underserved urban or rural areas in the United States: a systematic literature review. *Acad Med.* 2016;91(9):1313–1321. doi:10.1097/ACM.0000000000001203.

[7] Figueiredo AM, et al. Educational strategies to reduce physician shortages in underserved areas: a systematic review. *Int J Environ Res Public Health.* 2023;20(11):5983. doi:10.3390/ijerph20115983.

[8] Arredondo K, Touchett HN, Khan S, Vincenti M, Watts BV. Current programs and incentives to overcome rural physician shortages in the United States: a narrative review. *J Gen Intern Med.* 2023;38(Suppl 3):916–922. doi:10.1007/s11606-023-08122-6.

[9] The impact of administrative burden on academic physicians: results of a hospital-wide physician survey. 2017.

[10] Sinsky C, Colligan L, Li L, et al. Allocation of physician time in ambulatory practice: a time and motion study in 4 specialties. *Ann Intern Med.* 2016;165(11):753–760. doi:10.7326/M16-0961.

[11] Muhiyaddin R, Elfadl A, Mohamed E, Shah Z, Alam T, Abd-Alrazaq A, Househ M. Electronic health records and physician burnout: a scoping review. *Stud Health Technol Inform.* 2022;289:481–484. doi:10.3233/SHTI210962.

[12] Melnick ER, et al. Physician burnout and the electronic health record leading up to and during the first year of COVID-19: systematic review. *JMIR Med Inform.* 2022.

[13] Exploring the relationship between electronic health records and provider burnout: a systematic review. 2021.

[14] Wu Y, Wu M, Wang C, Lin J, Liu J, Liu S. Evaluating the prevalence of burnout among health care professionals related to electronic health record use: systematic review and meta-analysis. *JMIR Med Inform.* 2024;12:e54811. doi:10.2196/54811.

[15] Adler-Milstein J, et al. Measurement of clinical documentation burden among physicians and nurses using electronic health records: a scoping review. *J Am Med Inform Assoc.* 2021.

[16] Topol EJ. High-performance medicine: the convergence of human and artificial intelligence. *Nat Med.* 2019;25(1):44–56. doi:10.1038/s41591-018-0300-7.

[17] Rajpurkar P, Chen E, Banerjee O, Topol EJ. AI in health and medicine. *Nat Med.* 2022;28(1):31–38. doi:10.1038/s41591-021-01614-0.

[18] Esteva A, Kuprel B, Novoa RA, et al. Dermatologist-level classification of skin cancer with deep neural networks. *Nature.* 2017;542(7639):115–118. doi:10.1038/nature21056.

[19] Bracken A, Reilly C, Feeley A, Sheehan E, Merghani K, Feeley I. Artificial intelligence (AI)-powered documentation systems in healthcare: a systematic review. *J Med Syst.* 2025;49(1):28. doi:10.1007/s10916-025-02157-4.

[20] Clinical implementation of artificial intelligence scribes in health care: a systematic review.ApplClin Inform. 2025.

[21] Zhao J, Liu H, Chen Y, Song F. Application of artificial intelligence tools and clinical documentation burden: a systematic review and meta-analysis. *BMC Med Inform DecisMak.* 2026;26:29. doi:10.1186/s12911-025-03324-w.

[22] Applications and concerns of ChatGPT and other conversational large language models in health care: systematic review. 2024.

[23] Bedi S, Liu Y, Orr-Ewing L, et al. Testing and evaluation of health care applications of large language models: a systematic review. *JAMA.* 2025;333(4):319–328. doi:10.1001/jama.2024.21700.

[24] Implementation of machine learning applications in health care organizations: systematic review of empirical studies. 2024.

[25] Loh HW, Ooi CP, Seoni S, Barua PD, Molinari F, Acharya UR. Application of explainable artificial intelligence for healthcare: a systematic review of the last decade (2011–2022). *Comput Methods Programs Biomed.* 2022;226:107161. doi:10.1016/j.cmpb.2022.107161.

[26] Jung J, Lee H, Jung H, Kim H. Essential properties and explanation effectiveness of explainable artificial intelligence in healthcare: a systematic review. *Heliyon.* 2023;9(5):e16110. doi:10.1016/j.heliyon.2023.e16110.

[27] Rosenbacke R, Melhus Å, McKee M, Stuckler D. How explainable artificial intelligence can increase or decrease clinicians' trust in AI applications in health care: systematic review. *JMIR AI.* 2024;3:e53207. doi:10.2196/53207.

[28] Fairness of artificial intelligence in healthcare: review and recommendations. 2023.

[29] Obermeyer Z, Powers B, Vogeli C, Mullainathan S. Dissecting racial bias in an algorithm used to manage the health of populations. *Science.* 2019;366(6464):447–453. doi:10.1126/science.aax2342.

[30] Page MJ, McKenzie JE, Bossuyt PM, et al. The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. *BMJ.* 2021;372:n71. doi:10.1136/bmj.n71.

[31] Page MJ, McKenzie JE, Bossuyt PM, Boutron I, Hoffmann TC, Mulrow CD, et al. The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. *BMJ.* 2021;372:n71. doi:10.1136/bmj.n71.

[32] Collins GS, Moons KGM, Dhiman P, Riley RD, Beam AL, Van Calster B, et al. TRIPOD+AI statement: updated guidance for reporting clinical prediction models that use regression or machine learning methods. *BMJ.* 2024;385:e078378. doi:10.1136/bmj-2023-078378.

[33] Health Resources and Services Administration. *Data Downloads: Shortage Areas—Health Professional Shortage Areas and Medically Underserved Areas/Populations.* U.S. Department of Health and Human Services; updated August 23, 2026.

[34] Breiman L. Random forests. *Machine Learning.* 2001;45:5–32.

[35] Chen T, Guestrin C. XGBoost: A scalable tree boosting system. In: *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining.* 2016:785–794.

[36] Lundberg SM, Lee SI. A unified approach to interpreting model predictions. In: *Advances in Neural Information Processing Systems.* 2017;30.

[37] Lundberg SM, Erion G, Chen H, et al. From local explanations to global understanding with explainable AI for trees. *Nature Machine Intelligence.* 2020;2:56–67.

[38] Molnar C. *Interpretable Machine Learning.* 2nd ed. 2022.

[39] Friedman JH. Greedy function approximation: A gradient boosting machine. *Annals of Statistics.* 2001;29(5):1189–1232.

[40] Hastie T, Tibshirani R, Friedman J. *The Elements of Statistical Learning.* 2nd ed. Springer; 2009.

[41] Collins GS, Moons KGM, Dhiman P, et al. TRIPOD+AI statement: updated guidance for reporting clinical prediction models that use regression or machine learning methods. *BMJ.* 2024;385:e078378.

[42] Health Resources and Services Administration. *Health Professional Shortage Areas: Data and Resources.* U.S. Department of Health and Human Services.

[43] Du et al. systematic review of generative LLMs in EHR applications; 18,735 records retrieved and 196 studies included.

[44] Artsi et al. systematic review of LLMs in real-world clinical workflows; only four studies met the specified real-world implementation criteria.

[45] Systematic review/meta-analysis of AI tools and clinical documentation burden 23 studies, with pooled evidence for reduced documentation workload/time, but considerable heterogeneity and generally low methodological quality.

[46] JAMA Network Open ambient-AI implementation study — 100 clinicians, with reductions in documentation-related workload and cognitive demand; burnout reduction was not statistically significant.

[47] Recent health-system evaluation of ambient AI and clinician productivity/documentation.

[48] Systematic review of LLM applications in patient care — 89 studies across 29 specialties, with documentation and summarization among the identified applications and important safety/accuracy limitations.

[49] Systematic review of AI in prior authorization — identifies technical performance but limited evidence for operational, patient-centered, and equity outcomes.

[50] Overview of AI in primary care and administrative medicine, including scheduling, documentation, prior authorization, billing, patient education, and resource allocation.

[51] WHO guidance on ethics and governance of AI for health.

[52] WHO guidance on large multimodal models, emphasizing safety, governance, human oversight, and post-deployment assessment.