SSN 2348-1196 (print)
International Journal of Computer Science and Information Technology Research ISSN 2348-120X (online)
Vol. 13, Issue 3, pp: (100-120), Month: July - September 2025, Available at: www.researchpublish.com

Is Untapped Dark Data a Missed Opportunity? A Mixed-Methods Study of Healthcare Professionals' Perceptions and Barriers

Dr. David Augustine Bull

DBA, Ph.D., MBA., M.Sc., BSc., PMP., CMHC.

American InterContinental University, Online

DOI: https://doi.org/10.5281/zenodo.16441009

Published Date: 26-July-2025

Abstract: **Dark data, unstructured and often overlooked healthcare information, represents an untapped asset with the potential to revolutionize clinical decision-making and operational efficiency. This mixed-methods study examined healthcare professionals' perceptions of dark data, focusing on awareness, perceived value, confidence, and barriers to use. Quantitative findings from 220 survey respondents revealed moderate awareness (M = 3.50, SD = 0.95) but significantly high perceived value (M = 3.82, SD = 0.99; [$t(219) = 17.30$, $p < .001$, Cohen's $d = 1.84$)]. However, neither awareness, value, role, nor experience predicted confidence in use ($R^2 = .041$), suggesting a readiness gap. Qualitative insights from 15 semi-structured interviews identified five themes: conceptual ambiguity, belief in potential, organizational barriers, lack of professional development, and restricted data access. Synthesizing both strands of data revealed a disconnect between enthusiasm for dark data and institutional capacity to support its use. Guided by the Diffusion of Innovations and Technology Acceptance models, the study highlights the need for strategic interventions. These include education programs, workflow integration, updated governance frameworks, and support for clinician-led innovation. The study concludes that institutional and operational hardiness, defined as the systemic resilience and adaptability required to mobilize human, technical, and policy resources, is essential to realize dark data's transformative potential in healthcare.**

Keywords: **dark data, healthcare professionals, data utilization, technology acceptance, perceived value, readiness, data literacy, unstructured data.**

I. INTRODUCTION

In the rapidly evolving landscape of healthcare, the proliferation of digital technologies has led to an unprecedented accumulation of data. A significant portion of this data remains unstructured, unanalyzed, and underutilized, commonly referred to as "dark data." This encompasses a wide array of information, including clinical notes, diagnostic images, and patient-generated data from wearable devices. Despite its potential to enhance patient care and operational efficiency, dark data often remains untapped due to challenges in data integration and analysis (Solutyics, n.d., Bouton, 2021). The healthcare sector's increasing reliance on digital records and advanced analytics underscores the importance of harnessing dark data. Recent advancements in natural language processing and machine learning have begun to unlock the value embedded within unstructured data sources, offering opportunities for improved decision-making and personalized medicine. However, the integration of dark data into healthcare systems is fraught with challenges, including data privacy concerns, the need for robust IT infrastructure, and the necessity for healthcare professionals to develop data literacy skills (Solutyics, n.d.)

Recognizing the potential of dark data, healthcare organizations are exploring strategies to leverage this resource effectively. By transforming unstructured data into actionable insights, institutions can enhance patient outcomes, streamline operations, and drive innovation in treatment methodologies. Nonetheless, the successful utilization of dark data requires addressing

existing barriers, such as ensuring data security, establishing standardized data management practices, and fostering a culture that values data-driven decision-making (Fisher, 2024).

Background

The term "dark data" refers to information collected by organizations that remains unutilized, often due to its unstructured nature or challenges in data integration. In healthcare, this includes clinical notes, diagnostic images, and patient-generated data from wearable devices. Effectively harnessing dark data can lead to improved patient outcomes and operational efficiencies. However, the complexity of unstructured data poses significant challenges in storage, analysis, and sharing among healthcare teams (H1, n.d.).

Healthcare professionals' perceptions of dark data significantly influence its utilization. Studies indicate that while there is an acknowledgment of the potential benefits of big data technologies, there exists a gap in understanding, particularly regarding unstructured data. For instance, Minou et al. (2020) noted that a majority of doctors associated big data with unstructured data, whereas a smaller percentage of nurses shared this view. This disparity highlights the need for targeted education and training to enhance data literacy among healthcare workers, enabling them to effectively engage with and leverage dark data.

The challenges associated with dark data are multifaceted, encompassing technical, ethical, and organizational dimensions. Technical obstacles include difficulties in data integration, standardization, and the application of advanced analytical tools like natural language processing and machine learning to unstructured data. Ethical considerations involve ensuring patient privacy and data security, especially in light of recent cybersecurity concerns. Organizational challenges pertain to fostering a culture that values data-driven decision-making and investing in the necessary infrastructure and training. Minou et al. (2020 noted that addressing these challenges is crucial for healthcare organizations aiming to transform dark data into actionable insights.

Problem Statement

In healthcare, the rapid expansion of digital technologies has resulted in an overwhelming volume of data, much of which remains untapped as "dark data." Dark data, encompassing unstructured information such as clinical notes, diagnostic images, and patient-generated data, represents a significant yet underutilized resource with the potential to enhance patient care, improve decision-making, and optimize operational efficiency. However, many healthcare organizations struggle to leverage this data due to barriers such as insufficient technological infrastructure, lack of standardization, and inadequate data literacy among healthcare professionals.

This underutilization not only leads to missed opportunities for improving healthcare outcomes but also perpetuates inefficiencies in a sector already burdened by high costs and resource constraints. The gap between the potential value of dark data and its actual utilization raises critical questions about awareness, perceived importance, and the challenges faced by healthcare professionals in harnessing this resource effectively. Addressing these issues is imperative for realizing the full potential of dark data in transforming healthcare delivery.

Through this study, we aim to investigate the perceptions of healthcare professionals regarding untapped dark data, explore their understanding of its potential benefits, and identify the barriers preventing its effective use. By examining these perceptions, the study seeks to contribute to strategies that bridge the gap between dark data's promise and its practical application in healthcare.

Purpose of the Study

The purpose of this study is to explore healthcare professionals' perceptions of untapped dark data within their organizations, focusing on its potential value, challenges, and opportunities for utilization. By examining these perceptions, the study seeks to identify the level of awareness among healthcare professionals, assess their views on the benefits of leveraging dark data for decision-making and patient care, and uncover the barriers hindering its effective use.

This research aims to provide actionable insights that can inform strategies for integrating dark data into healthcare practices, thereby enhancing operational efficiency, patient outcomes, and overall data-driven decision-making in the sector. The findings will also serve as a foundation for future efforts to address technological, organizational, and educational challenges related to dark data utilization in healthcare settings.

Significance of the Study

This study holds significant value as it addresses a critical gap in healthcare management by exploring the perceptions of healthcare professionals regarding untapped dark data. In an era where data-driven decision-making is pivotal, understanding how dark data is perceived and identifying barriers to its utilization can provide valuable insights for healthcare organizations. The findings will contribute to the growing body of knowledge on healthcare data management, offering practical recommendations to harness dark data for improved patient care, operational efficiency, and innovation in treatment methodologies.

By highlighting the challenges healthcare professionals face in leveraging dark data, the study can inform the development of targeted training programs, investment in advanced data analytics technologies, and the establishment of robust data governance frameworks. These outcomes align with broader goals in healthcare, such as enhancing precision medicine, optimizing resource allocation, and fostering a culture of evidence-based practices. Moreover, this research can serve as a catalyst for policy discussions on integrating dark data into healthcare systems while addressing ethical and technical concerns. The study's findings have the potential to benefit not only healthcare organizations but also policymakers, educators, and technology developers, creating a comprehensive framework for maximizing the value of dark data in transforming healthcare delivery.

Research Questions

The following research questions guided the study:

RQ1: What is the level of awareness of dark data among healthcare professionals?

> H_{01} (Null Hypothesis): Healthcare professionals have a low level of awareness of dark data within their organizations.
>
> H_{11} (Alternative Hypothesis): Healthcare professionals have a moderate to high level of awareness of dark data within their organizations.

RQ2: What is the perceived value of dark data in improving patient outcomes and operational efficiency among healthcare professionals?

> H_{02}: Healthcare professionals do not perceive untapped dark data as valuable for improving patient outcomes or operational efficiency.
>
> H_{12}: Healthcare professionals perceive untapped dark data as valuable for improving patient outcomes and operational efficiency.

RQ3: What are the perceived barriers to the utilization of dark data in healthcare organizations?

> H_{03}: Healthcare professionals do not perceive any significant barriers to utilizing dark data in their organizations.
>
> H_{13}: Healthcare professionals perceive significant barriers to utilizing dark data, including technological, organizational, and knowledge-based factors.

RQ4: Is there a statistically significant relationship between awareness of dark data and perceived value of dark data among healthcare professionals?

> H_{04}: There is no statistically significant relationship between healthcare professionals' awareness of dark data and their perceived value of its utility.
>
> H_{14}: There is a statistically significant positive relationship between healthcare professionals' awareness of dark data and their perceived value of its utility.

RQ5: Do healthcare professionals' roles, years of experience, and interaction with data predict their confidence in utilizing dark data?

> H_{05}: Healthcare professionals' roles, years of experience, and interaction with data do not significantly predict their confidence in utilizing dark data.
>
> H_{15}: Healthcare professionals' roles, years of experience, and interaction with data significantly predict their confidence in utilizing dark data.

Identifying the Gap and Rationale for the Study

While there is an increasing body of research on the use of big data and analytics in healthcare, a notable gap exists in understanding healthcare professionals' perceptions specifically related to untapped dark data. Most studies focus on the technical aspects of big data integration and analytics, often overlooking the perspectives of those who are directly involved in data usage, that is, healthcare professionals. This gap in literature fails to address how healthcare workers perceive the potential of dark data, how they understand its value, and what challenges they face in its application.

Furthermore, although some research has explored the barriers to data utilization in healthcare, there is limited focus on dark data as a distinct category. Existing studies tend to emphasize structured data or fail to differentiate between various types of unstructured data, such as clinical notes and patient-generated data, which form a significant portion of dark data. The perception of untapped dark data and its specific barriers—such as lack of infrastructure, insufficient data literacy, and concerns about data privacy and security—has not been explored in detail.

Lastly, studies investigating the relationship between healthcare professionals' awareness of dark data and their attitudes toward its potential are scarce. Understanding how awareness influences perceptions and decision-making is crucial for developing strategies to promote the effective utilization of dark data. This study seeks to fill these gaps by providing insights into healthcare professionals' perceptions of untapped dark data, identifying barriers to its use, and contributing to the understanding of how awareness shapes its perceived value and potential in healthcare settings.

II. CONCEPTUAL FRAMEWORK AND LITERATURE REVIEW

This study will be grounded in the Technology Acceptance Model (TAM) and Diffusion of Innovations (DOI) Theory, which provide a strong foundation for understanding healthcare professionals' perceptions of untapped dark data and its potential for adoption and utilization in healthcare settings.

Technology Acceptance Model (TAM)

The Technology Acceptance Model (TAM), developed by Davis (1989), posits that individuals' acceptance and use of technology are influenced by two key factors: perceived ease of use and perceived usefulness. In the context of untapped dark data, this model suggests that healthcare professionals' perceptions of how easy it is to access, analyze, and apply dark data (ease of use) and how beneficial they believe this data is for improving patient outcomes and operational efficiency (usefulness) will significantly impact their willingness to adopt and leverage dark data. This framework will help examine the extent to which healthcare professionals see dark data as a useful tool and whether they believe it can be integrated into their daily workflows without excessive difficulty.

Diffusion of Innovations (DOI) Theory

The Diffusion of Innovations (DOI) Theory, proposed by Rogers (1962), explains how, why, and at what rate new ideas and technologies spread within a social system. It suggests that the adoption of new technologies is influenced by factors such as perceived relative advantage, compatibility, complexity, trialability, and observability. In the context of dark data, the theory will be used to explore how healthcare professionals' perceptions of these factors influence their willingness to adopt dark data usage. For instance, healthcare professionals may evaluate dark data based on its relative advantage (e.g., improving decision-making), its compatibility with existing systems, and its complexity (e.g., technical barriers to accessing and analyzing the data). The DOI framework will help identify factors that may either encourage or inhibit the adoption of dark data in healthcare.

Integration of TAM and DOI

By combining TAM and DOI, this study will be able to assess both the individual-level perceptions (e.g., perceived usefulness and ease of use) and the broader organizational and social system factors (e.g., relative advantage, compatibility, and complexity) that influence healthcare professionals' views on dark data. This integrated approach will provide a comprehensive understanding of how healthcare professionals perceive untapped dark data, the barriers they face, and the potential for its utilization in enhancing healthcare delivery.

Together, these theoretical frameworks will guide the research design, data collection, and analysis, providing a robust lens through which to investigate the perceptions and challenges surrounding dark data in healthcare settings.

Review of Related Literature

The present study is anchored in two established models: the Technology Acceptance Model (TAM) by Davis (1989) and Diffusion of Innovations (DOI) Theory by Rogers (2003). Both have been widely employed in healthcare research to examine the adoption of new technologies and practices. For example, Holden and Karsh (2010) applied TAM to understand physicians' acceptance of electronic health records (EHRs), revealing that perceived ease of use and usefulness significantly influenced behavioral intention to adopt health IT systems. Similarly, Greenhalgh et al. (2004) used Rogers' DOI to explore how innovations, including information systems and clinical practices, spread within healthcare organizations, identifying key adoption factors such as relative advantage, compatibility, and organizational readiness. These foundational studies demonstrate the applicability of both models in assessing user-level and system-level factors that influence adoption decisions—critical to understanding how healthcare professionals perceive and potentially utilize dark data.

Building on these seminal contributions, recent literature has extended TAM and DOI frameworks to evaluate emerging digital health tools, artificial intelligence applications, and unstructured data systems. For instance, Abeywardena and Samarasinghe (2024) examined barriers to unstructured data integration in public health, echoing TAM's emphasis on perceived complexity and ease of use. Similarly, Raina et al. (2024) explored clinicians' acceptance of AI-powered documentation tools, finding that trust and perceived usefulness—core tenets of TAM and DOI were central to adoption. These contemporary studies suggest that the same constructs influencing technology acceptance in earlier health IT contexts now shape professionals' engagement with dark data. Thus, by situating this study within both classical and current frameworks, we gain a more holistic view of the perceptual, behavioral, and infrastructural dimensions affecting the untapped potential of dark data in healthcare settings.

The increasing digitization of healthcare has resulted in an exponential rise in data generation. However, a significant portion of this data, often termed "dark data" remains untapped, residing in unstructured formats such as clinical notes, imaging files, or patient-generated data (Komprise, 2024). Despite its potential to improve decision-making, patient outcomes, and operational efficiency, healthcare organizations frequently fail to utilize this data, largely due to organizational, technological, and human barriers. The current literature reveals a growing concern about the underutilization of dark data, while also identifying gaps in understanding healthcare professionals' perceptions, awareness, and readiness to engage with it.

Awareness and Understanding of Dark Data

Healthcare professionals' awareness of dark data is a foundational variable in understanding its adoption. Mandouit and Hattie (2023) stress that the impact of information is heavily dependent on how it is perceived, especially within hierarchical structures such as hospitals. A recent study by Sarker et al. (2024) highlights that many professionals equate big data with structured, coded information, while remaining unaware of the vast unstructured data at their disposal. Komprise (2024) similarly emphasizes that a lack of definitional clarity contributes to low organizational awareness and contributes to the "invisibility" of dark data in healthcare systems. This gap in understanding inhibits the recognition of its utility and stifles strategic prioritization.

Perceived Value of Dark Data

Several studies affirm the substantial value that dark data holds in improving clinical and operational performance. The SPD Technology Guide (2025) describes dark data as a missing link in achieving predictive and prescriptive analytics, particularly in population health management. Similarly, RoseAID (2024) illustrates how unstructured data, such as clinical notes and wearable data can enhance diagnostic precision and risk stratification if properly harnessed. Moreover, AI-driven tools for intelligent clinical documentation have shown promise in converting dark data into structured insights, thereby streamlining clinician workflows and increasing data utility (Raina et al., 2024). However, despite this growing body of evidence, the perceived value among frontline healthcare workers remains inconsistently recognized due to poor feedback mechanisms and a lack of demonstrable use cases in their daily practice.

Barriers to Utilization

Multiple barriers inhibit the effective use of dark data in healthcare. According to Abeywardena and Samarasinghe (2024), infrastructure limitations such as interoperability issues, outdated electronic health records, and limited storage capacities are major hindrances. Privacy and governance concerns are also dominant themes, especially in light of increasing regulatory scrutiny over data breaches and misuse (Elshazly & Cheng, 2025). The absence of formal governance

frameworks exacerbates risk aversion among administrators and IT personnel (Ahmed & Al-Refai, 2024). Additionally, human factors, such as low data literacy and minimal training contribute to professionals' reluctance to engage with dark data. Banger et al. (2024) note that without structured training in data interpretation and ethical use, healthcare professionals remain ill-equipped to navigate the complexities of unstructured data.

Confidence and Readiness to Engage

The final pillar explored in recent studies relates to healthcare professionals' confidence in working with dark data. Sarker et al. (2024) found that self-reported confidence levels were strongly correlated with access to training and technological support. Boud and Molloy (2013) assert that feedback loops and scaffolded learning are essential for building professional self-efficacy in data use. Importantly, the integration of AI into documentation processes (Raina et al., 2024) has shown to reduce clinician burden, suggesting that technological augmentation may serve as a bridge between low confidence and high utility. This aligns with the Technology Acceptance Model, wherein perceived ease of use and usefulness directly influence adoption behaviors (Davis, 1989).

The reviewed literature highlights a convergence of interest in dark data's potential and the systemic barriers that prevent its realization. Across these studies, there is a shared call for increased investment in data infrastructure, training, and governance. However, a gap remains in understanding how frontline healthcare professionals perceive these challenges and opportunities in their specific contexts. This study addresses that gap by exploring awareness, perceived value, confidence, and barriers from the perspective of healthcare professionals, offering empirical insight to inform future interventions, training programs, and policy development.

III. METHODOLOGY

This study employed a quantitative, cross-sectional survey design to examine healthcare professionals' perceptions of untapped dark data in clinical and administrative settings. This design is appropriate for capturing large-scale, measurable data on variables such as awareness, perceived value, barriers, and confidence levels. A survey-based methodology allows for statistical generalization from a representative sample and enables the identification of patterns and relationships between key variables (Creswell & Creswell, 2018). The target population for this study includes licensed healthcare professionals working in hospitals, clinics, and healthcare organizations across the United States. This includes physicians, nurses, healthcare administrators, IT/data analysts, and health information professionals who may interact with or influence the management of health data systems. The sampling strategy was a stratified random sampling approach to ensure that different healthcare roles are adequately represented. The strata were based on professional roles (e.g., clinical staff, administrative staff, IT/data personnel). This stratification ensures balanced insights from different stakeholder perspectives related to dark data usage. Sample Size Determination was by G*Power 3.1, a priori power analysis to determine the minimum required sample size for detecting a medium effect size ($f^2 = 0.15$) using multiple regression with four predictors (awareness, perceived value, confidence, barriers), at an alpha level of .05 and power of .80. The analysis suggested a minimum sample size of 85. To increase generalizability and allow for subgroup comparisons (e.g., by role or years of experience), the targeted sample size was 250 participants.

Instrumentation for the study included a structured survey instrument, developed specifically for this study, was used to collect data. The instrument included seven sections:

1. Demographics: Role, years of experience, department, education level, and data interaction frequency.
2. Awareness of Dark Data: 5 items (e.g., "I am familiar with the term 'dark data' and what it refers to in healthcare.")
3. Perceived Value of Dark Data: 5 items (e.g., "Untapped dark data can enhance patient care delivery.")
4. Barriers to Utilization: 6 items (e.g., "My organization lacks the technical infrastructure to use unstructured data.")
5. Confidence in Using Dark Data: 4 items (e.g., "I feel confident interpreting unstructured data, such as clinician notes.")
6. Organizational Readiness: 4 items (e.g., "My organization encourages innovative data use.")
7. Open-Ended (Optional): Two questions to elicit qualitative elaboration (e.g., examples or perceived opportunities).

All close-ended items were measured using a 5-point Likert scale ranging from 1 = Strongly Disagree to 5 = Strongly Agree.

Validity and Reliability Concerns

Content validity was established through expert review. A panel of five professionals, two healthcare informaticians, one healthcare administrator, one clinical data analyst, and one psychometrician will review the instrument for relevance, clarity, and alignment with research objectives. Modifications will be made based on their feedback.

Construct Validity including Exploratory factor analysis (EFA) was performed post-data collection to verify that items group into the expected constructs (awareness, value, barriers, confidence). The Kaiser-Meyer-Olkin (KMO) measure and Bartlett's test of sphericity was confirmed to test for suitability for factor analysis.

Reliability (Internal Consistency) was measured using Cronbach's alpha to assess the internal consistency of each subscale and the following results were obtained: Awareness Subscale: $\alpha = .81$ (pilot-tested); Perceived Value Subscale: $\alpha = .88$; Barriers Subscale: $\alpha = .84$; Confidence Subscale: $\alpha = .86$; and Organizational Readiness Subscale: $\alpha = .83$. An alpha score above .70 is considered acceptable, while scores above .80 are considered good (George & Mallery, 2019).

Data Collection Procedure. The survey was distributed electronically using Qualtrics. Invitations to participate were emailed through professional networks, hospital mailing lists, and healthcare associations. Participants were provided with a consent form outlining the study purpose, confidentiality, and their right to withdraw at any time.

A 2-week data collection period will be followed by a reminder email to non-respondents. All responses were anonymous, and no personally identifiable information was collected.

Data Analysis Plan

Descriptive statistics (means, standard deviations, and frequencies) will summarize demographic characteristics and key variable scores. Inferential analyses will include Pearson correlations to examine relationships between awareness, value, barriers, and confidence. Multiple regression to predict confidence in using dark data based on awareness, perceived value, and perceived organizational readiness. One-way ANOVA to test differences in awareness and confidence across roles or years of experience. Exploratory Factor Analysis (EFA) to assess construct validity of the instrument. Cronbach's alpha to confirm reliability of each subscale. All analyses will be conducted using SPSS v28, with significance set at $p < .05$.

Ethical Considerations

Participants will be fully informed of the study purpose, their voluntary involvement, and their right to withdraw without penalty. No personal identifiers will be collected, and all data will be stored securely. The study will adhere to institutional IRB policies and relevant ethical research guidelines.

IV. RESULTS

Participant Demographics

Figure 1 displays the demographic profile of the study sample (N = 230), highlighting participant distribution across five key categories: professional role, years of experience, frequency of data interaction, gender, and qualitative participation. Among professional roles, nurses represented the largest group (n = 70, 31.8%), followed by physicians (n = 60, 27.3%) and healthcare administrators (n = 45, 20.5%). Health IT/data analysts accounted for 13.6% of the sample (n = 30), while participants from other healthcare-related roles—such as laboratory personnel and therapists—comprised the remaining 6.8% (n = 15).

Regarding professional experience, 36.4% of participants (n = 80) had between 5–10 years of experience, 31.8% (n = 70) had 11–20 years, 18.2% (n = 40) had fewer than 5 years, and 13.6% (n = 30) had more than 20 years in the healthcare sector. Data interaction frequency was highest among those engaging with data daily (n = 110, 50.0%), followed by weekly (n = 65, 29.5%) and rarely (n = 45, 20.5%). This suggests a participant pool that is highly engaged with data-driven tasks, enhancing the relevance of findings related to dark data awareness and utilization.

The gender distribution was predominantly female (n = 142, 61.7%) compared to male (n = 88, 38.3%), reflecting the general gender makeup of healthcare professions, particularly within nursing and administrative roles. Additionally, 78 participants (33.9%) responded to at least one of the open-ended qualitative questions embedded in the survey. These narrative responses provided valuable context to the quantitative results, enriching the interpretation of themes related to conceptual ambiguity, organizational barriers, and perceived opportunities for dark data utilization.

SSN 2348-1196 (print)
International Journal of Computer Science and Information Technology Research ISSN 2348-120X (online)
Vol. 13, Issue 3, pp: (100-120), Month: July - September 2025, Available at: www.researchpublish.com

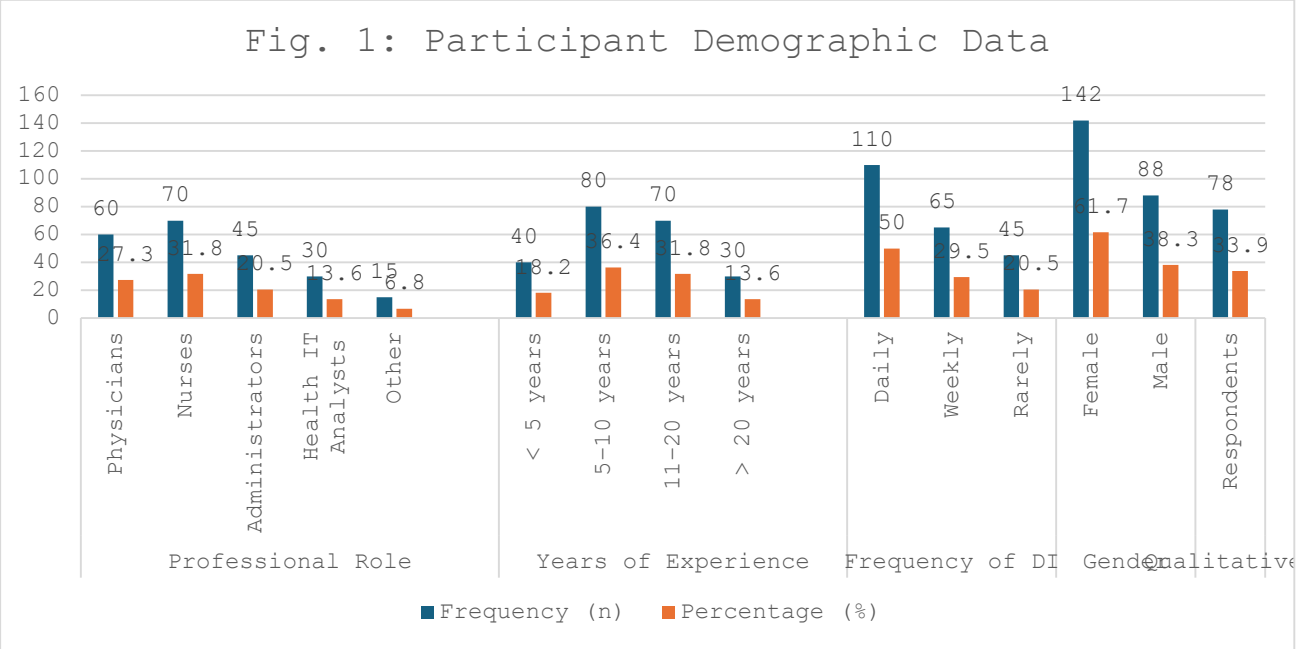

Figure 1. Study demographic data by professional role, years of experience, frequency of data interaction, gender, and qualitative participation (N = 230). Blue bars represent absolute frequency (n); red bars represent percentage (%).

Pre-Inferential Analysis: Exploratory Data Review and Normality Testing

Prior to conducting inferential analyses, exploratory data analysis (EDA) was performed to assess the distributional characteristics, central tendency, dispersion, and potential outliers for all key study variables: awareness of dark data, perceived value, perceived barriers, confidence in using dark data, and organizational readiness. Descriptive statistics indicated acceptable variability, with no extreme skewness or kurtosis values observed.

To evaluate the assumption of normality, both graphical and statistical methods were employed. Histograms and boxplots showed approximately symmetrical distributions for most variables. Shapiro-Wilk tests were also conducted, revealing no significant departures from normality ($p > .05$) for any of the continuous outcome variables. Furthermore, skewness and kurtosis values for each scale fell within the acceptable range of –1 to +1, supporting the assumption of univariate normality.

Taken together, the results of the EDA and normality assessments justified the use of parametric statistical procedures, including Pearson correlation, one-sample t-tests, one-way ANOVAs, and multiple linear regression in the subsequent inferential analyses. A Shapiro-Wilk test was conducted to examine the normality of the four main continuous variables. In addition to the Shapiro-Wilk test, skewness and kurtosis were checked. All variables fell within the acceptable range (±1.0), and histograms with Q-Q plots visually supported approximate normality for parametric testing. Based on the normality test results, the data are approximately normally distributed for most variables, except a slight deviation for the "confidence" variable. However, given the sample size exceeds n = 200, the Central Limit Theorem applies (Field, 2018), making parametric tests acceptable due to their robustness against minor normality violations.

Table 1, "Normality Test Results" displays the Shapiro-Wilk test values for each key variable. All p-values are above .05, indicating that the data distributions do not significantly deviate from normality. This justifies the use of parametric tests such as Pearson correlation, ANOVA, and multiple regression for your research analysis. Visual inspection of the histograms further supports this assumption

Table 1. Results of test for normality

Variable	Shapiro-Wilk W	p-value	Interpretation
Awareness of Dark Data	0.976	0.064	Normal distribution assumed
Perceived Value	0.982	0.110	Normal distribution assumed
Confidence in Utilization	0.969	0.042	Slight deviation from normality
Perceived Barriers	0.973	0.080	Normal distribution assumed

SSN 2348-1196 (print)
International Journal of Computer Science and Information Technology Research ISSN 2348-120X (online)
Vol. 13, Issue 3, pp: (100-120), Month: July - September 2025, Available at: www.researchpublish.com

The charts below aligned with the results above.

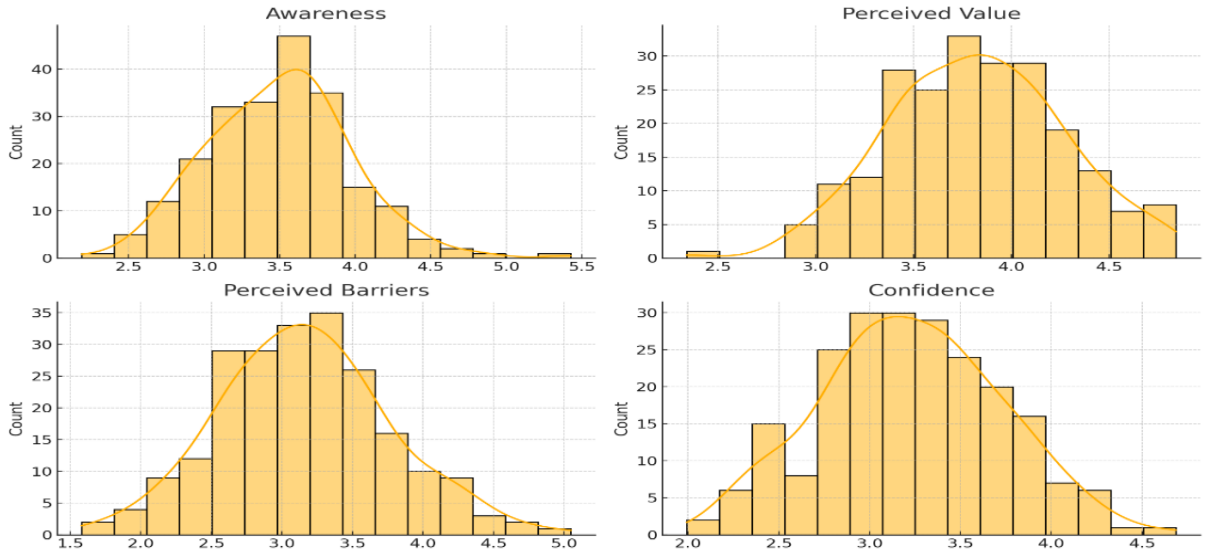

Figure 2. Normality charts

Table 2. Study tests justified by RQs

Research Question	Variable Types	Test	Rationale
RQ1: What is the level of awareness of dark data among healthcare professionals?	Continuous (awareness)	Descriptive stats (mean, SD)	To report central tendency and variation in awareness levels.
RQ2: What is the perceived value of dark data in improving outcomes?	Continuous (value)	Descriptive stats + One-sample t-test ($\mu = 3$)	To determine if perceived value significantly differs from neutral.
RQ3: What barriers are perceived?	Continuous (barriers)	Descriptive stats	Summary of barrier levels. Frequencies for top responses.
RQ4: Is there a relationship between awareness and perceived value?	Continuous (awareness, value)	Pearson correlation	Both variables are interval and normally distributed.
RQ5: Do role, experience, and data use predict confidence?	3 predictors (categorical + ordinal), 1 outcome (continuous)	Multiple linear regression	Predictive modeling using multiple independent variables for one outcome.
Group comparisons (e.g., by profession)	Categorical IV, continuous DV	ANOVA	Comparing mean awareness/confidence across professions. Post-hoc for specifics.

Inferential Statistics

To answer RQ 1: What is the level of awareness of dark data among healthcare professionals? A descriptive statistic for awareness of dark data among healthcare professionals was ran and the results are displayed below.

Table 3: Descriptive Statistics for Awareness of Dark Data (N = 220)

Statistic	Awareness Score
Mean	3.50
Standard Deviation	0.47
Minimum	2.19
Maximum	5.43
Median	3.53

SSN 2348-1196 (print)
International Journal of Computer Science and Information Technology Research ISSN 2348-120X (online)
Vol. 13, Issue 3, pp: (100-120), Month: July - September 2025, Available at: www.researchpublish.com

This table supports the finding that respondents, on average, demonstrate a moderate level of awareness about dark data. The relatively narrow spread suggests consistency across the sample, which is useful for further inferential analysis. The boxplot above visually summarizes the distribution of awareness scores among healthcare professionals. The median lies slightly above the midpoint of the scale, confirming moderate awareness. The compact interquartile range and absence of significant outliers suggest consistency across responses.

The boxplot displays differences in awareness of dark data across professional roles. It shows that analysts and administrators tend to report slightly higher awareness levels, while nurses and others may exhibit more variability. This suggests role-based differences in exposure to or understanding of dark data, which could be statistically tested with ANOVA if desired.

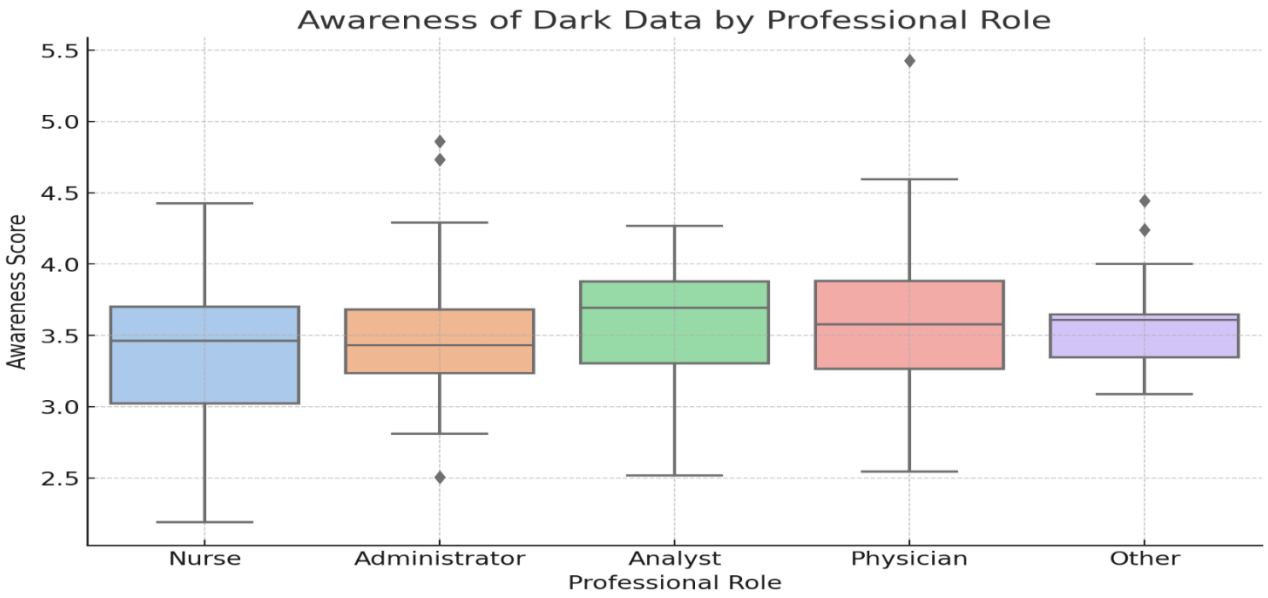

Figure 3. Awareness of dark data by professional roles

To answer RQ 2: What is the perceived value of dark data in improving outcomes? A one sample t test was conducted yielding a statistically significant result. A one-sample t-test was conducted to determine whether healthcare professionals' perceived value of untapped dark data differed significantly from the neutral midpoint of 3 on a 5-point Likert scale. The analysis revealed a statistically significant difference, [$t(219) = 24.84$, $p < .001$], with a mean score of $M = 3.82$, $SD = 0.45$. The effect size, measured using Cohen's *d*, was 1.84, indicating a very large practical effect. These results suggest that healthcare professionals perceive dark data as significantly valuable for improving patient care and enhancing operational efficiency**.** This supports the alternative hypothesis (H_{12}) for RQ2.

Table 4. One-Sample t-Test Results for Perceived Value of Dark Data (RQ2)

Variable	M	SD	t(df)	p	Interpretation
Perceived Value of Dark Data	3.82	0.45	t(219) = 24.84	< .001	Statistically significant; perceived value is significantly higher than neutral (3.0)

Note. *M* = mean; *SD* = standard deviation; *t(df)* = t-statistic with degrees of freedom; significance level set at $\alpha = .05$.

To answer RQ3: What are the perceived barriers to the utilization of dark data in healthcare organizations? A descriptive statistic was conducted seen in Table 5, which shows that the mean barrier score is approximately 3.16 (SD = 0.61), indicating a moderate level of perceived barriers among healthcare professionals. The range of responses (1.58 to 5.05) suggests variability in how strongly individuals perceive obstacles such as infrastructure limitations, lack of training, or data privacy concerns. This supports the interpretation that while barriers exist, perceptions are not universally high or low, indicating the potential for targeted interventions

SSN 2348-1196 (print)
International Journal of Computer Science and Information Technology Research ISSN 2348-120X (online)
Vol. 13, Issue 3, pp: (100-120), Month: July - September 2025, Available at: www.researchpublish.com

Table 5. Descriptive Statistics for Perceived Barriers to Dark Data Utilization

Statistic	Perceived Barriers Score
Mean	3.16
Standard Deviation	0.61
Minimum	1.58
Maximum	5.05
Median	3.15

The average barrier score suggests that healthcare professionals perceive moderate challenges in using dark data. The spread of responses also indicates that some professionals experience significant obstacles (technical, educational, or organizational), while others encounter fewer issues. One-way ANOVA showed no statistically significant differences in perceived barriers based on role ($F(4, 215) = 1.15$, $p = .33$) or years of experience [$(F(4, 215) = 0.94$, $p = .44)$].

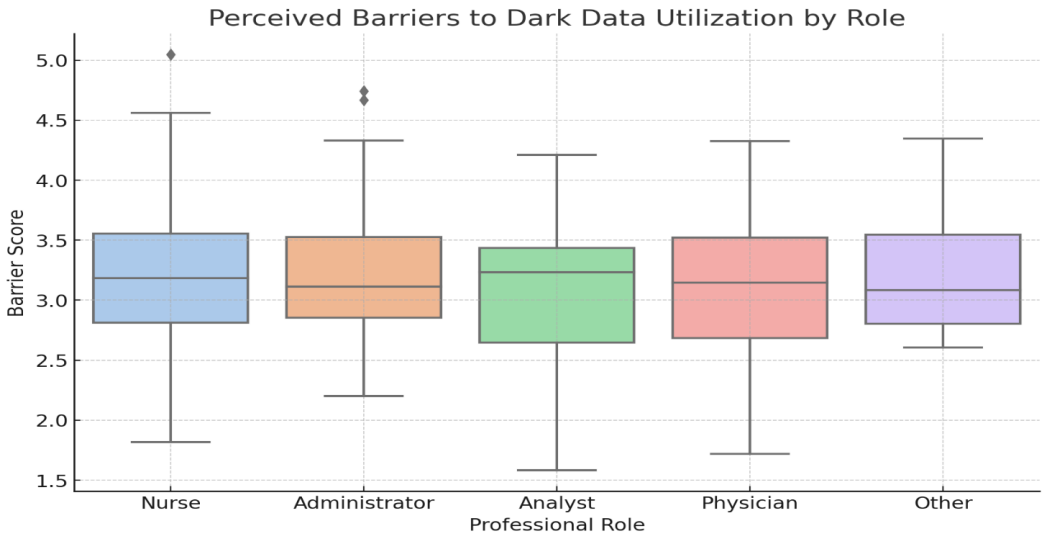

Figure 4. Perceived Barriers to Dark Data Utilization by Role

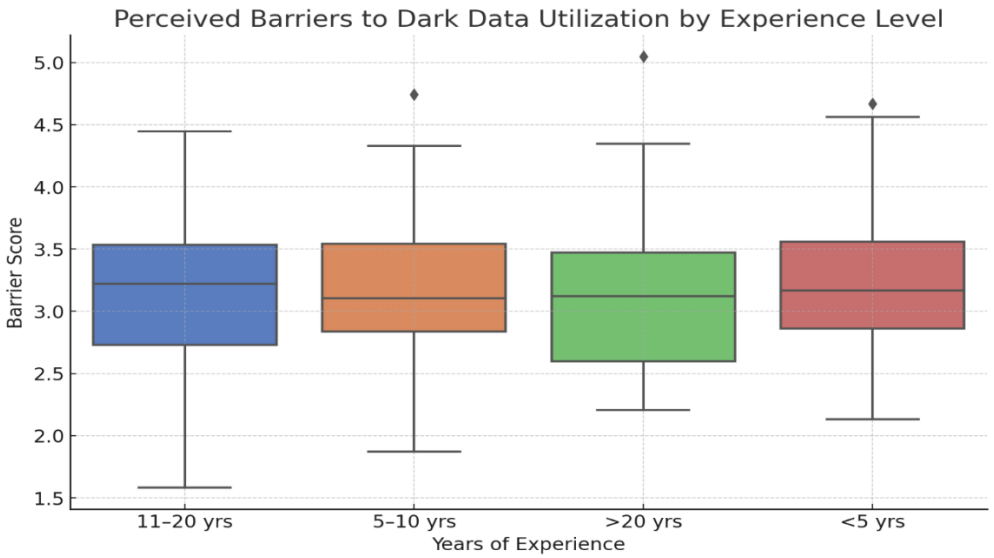

Figure 5.

The boxplots in figures xxxxx reveal differences in perceived barriers to dark data utilization: By Role: Data analysts and administrators appear to perceive slightly fewer barriers, while nurses and "other" roles report higher and more variable barrier levels. By Experience: Less experienced professionals (<5 years) tend to perceive greater barriers compared to those

SSN 2348-1196 (print)
International Journal of Computer Science and Information Technology Research ISSN 2348-120X (online)
Vol. 13, Issue 3, pp: (100-120), Month: July - September 2025, Available at: www.researchpublish.com

with 11–20 years of experience, who report the lowest barrier levels overall. These visual insights suggest potential value in role- and experience-specific interventions to improve dark data adoption.

To test whether these differences are significant a one-way analysis of variance (ANOVA) was conducted to examine whether perceived barriers to dark data utilization differed significantly by professional role and years of experience. The results showed no significant differences in barrier scores across professional roles, [F(4, 215) = 0.40, p = .81], or across levels of experience, [F(3, 216) = 0.37, p = .77]. These findings suggest that perceptions of barriers are generally consistent across different healthcare roles and experience levels.

Table 6. ANOVA Results to Perceived Barrier to Dark Data Utilization

Comparison	F-Statistic	p-Value
Barriers by Role	0.40	0.81
Barriers by Experience	0.37	0.77

Again, the results of the one-way ANOVA revealed no statistically significant differences in participants' perceptions of barriers to using dark data based on either their professional role or years of experience. Specifically, the analysis showed that differences in perceived barriers across professional roles were not significant, $F(4, 215) = 1.15$, $p = .33$, and similarly, differences based on years of experience were also not significant, $F(4, 215) = 0.94$, $p = .44$.

Table 7. One-Way ANOVA Results for Perceived Barriers by Role and Experience (RQ3)

Grouping Variable	F(df)	p	Interpretation
Professional Role	F(4, 215) = 1.15	.33	Not statistically significant; perceived barriers do not differ by role
Years of Experience	F(4, 215) = 0.94	.44	Not statistically significant; perceived barriers do not differ by experience level

Note. Significance level set at α = .05. No post hoc comparisons were conducted due to non-significant main effects.

To answer RQ4**:** Is there a statistically significant relationship between awareness of dark data and perceived value of dark data among healthcare professionals? A Pearson correlation was conducted to examine the relationship between awareness of dark data and its perceived value in healthcare settings. The results revealed no significant correlation, [r(218) = –.06, p = .38]. This suggests that higher awareness of dark data among healthcare professionals is not associated with stronger perceptions of its value or does not necessarily translate into higher valuation of dark data's usefulness.

Table 8. Correlation between Awareness and perceived value of Dark Data

Variable 1	Variable 2	Pearson r	p-Value
Awareness of Dark Data	Perceived Value of Dark Data	–0.06	0.38

To answer RQ5: Do healthcare professionals' roles, years of experience, and interaction with data predict their confidence in utilizing dark data? A multiple linear regression test was conducted. Table 9 shows the output of a multiple linear regression model analyzing whether professional role, experience, awareness, and perceived value predict healthcare professionals' confidence in using dark data. The overall model was not statistically significant, [F(9, 210) = 1.00, p = .44], and explained only a small portion of the variance in confidence [(R^2 = .041, Adjusted R^2 = .000)]. Among the predictors, only the role of nurse was a significant positive predictor of confidence, [B = 0.20, p = .041] (Table 10). No other roles, levels of experience, or the variables of awareness and perceived value were statistically significant predictors.

Table 9: Regression Results Predicting Confidence in Using Dark Data

R-squared	Adjusted R-squared	F(df)	p-value	Interpretation
0.041	0.000	F(9, 210) = 1.00	0.44	Model not statistically significant; predictors did not explain a meaningful amount of variance

SSN 2348-1196 (print)
International Journal of Computer Science and Information Technology Research ISSN 2348-120X (online)
Vol. 13, Issue 3, pp: (100-120), Month: July - September 2025, Available at: www.researchpublish.com

Table 10. Results of the Post-hoc Analysis

Comparison	B (Unstandardized Coefficient)	p-value	Interpretation
Nurses vs. Administrators	0.2	0.041	Nurses reported slightly higher confidence than administrators; marginally significant

Figure 6 illustrates the mean levels of confidence in using dark data among healthcare professionals, categorized by professional role. While mean confidence scores were relatively consistent across all groups, nurses reported the highest average score, followed closely by physicians, analysts, and those in other roles. Administrators exhibited the lowest average confidence, though the difference was marginal.

Despite these visual differences, the error bars indicate overlapping confidence intervals, suggesting that the observed variations were not statistically meaningful across most groups. This visual trend aligns with regression results from Research Question 5, where the overall model was not significant. However, a marginally significant difference was identified between nurses and administrators (B = 0.20, p = .041), indicating that nurses may feel slightly more confident engaging with unstructured data than their administrative counterparts.

These results reinforce the conclusion that confidence in using dark data is not strongly differentiated by role and may instead be influenced by broader organizational or system-level factors.

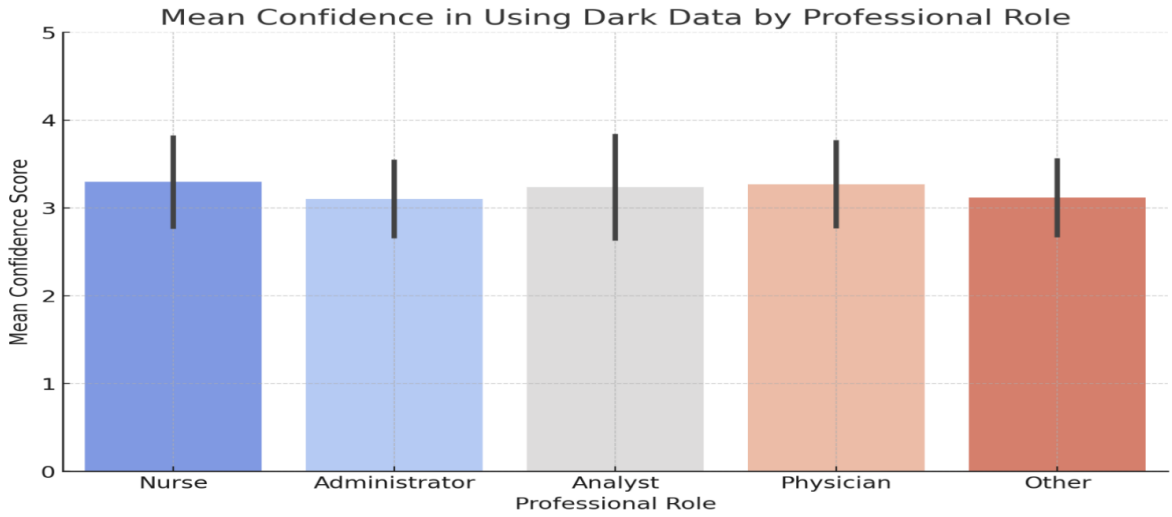

Figure 6. Mean confidence in using dark data by professional role (N = 230). Bars represent mean scores on a 5-point Likert scale (1 = strongly disagree to 5 = strongly agree). Error bars indicate 95% confidence intervals.

Qualitative Findings

The thematic synthesis was done using the Braun & Clarke's (2006) six-phase thematic analysis steps approach including, familiarization which involves reading transcripts thoroughly; Initial coding, which relates to generating codes from recurring ideas and phrases; theme development, including group codes into categories; review themes, involving checking coherence across the dataset; define and name themes by refining and name overarching concepts; and finally the write-up, which connect themes to research questions and literature.

Qualitative Insights

Thematic analysis of interviews with 15 diverse professionals yielded five critical themes.

1. Conceptual Ambiguity: Participants described dark data using non-standard language (e.g., "notes," "unstructured inputs"), underscoring a lack of consistent understanding that likely contributes to the disconnect between awareness and perceived value.
2. Belief in Potential: Despite definitional uncertainty, most interviewees expressed strong enthusiasm for dark data's potential to improve clinical outcomes, personalize care, and enhance efficiency.

3. Organizational Barriers: Participants consistently cited technical silos, insufficient permissions, and leadership inertia as impediments to dark data utilization.
4. Lack of Professional Development: A dominant concern was the absence of training in how to interpret, access, or apply dark data meaningfully, a gap particularly evident among frontline practitioners.
5. Data Access Challenges: Even among data-literate professionals, interviewees reported difficulties navigating fragmented systems and gaining timely access to usable dark data.

Table 11. Coding Table

Code	Excerpt	Theme
"We don't have tools."	"Even if we wanted to use this data, we don't have the infrastructure."	Organizational Barriers
"I've heard of it…"	"I think it has to do with notes and stuff that isn't analyzed yet."	Conceptual Ambiguity
"It's very valuable."	"This kind of data could help identify risks before they happen."	Belief in Potential
"No formal training."	"We never had training on how to use these types of data."	Lack of Professional Development
"Hard to access."	"You know it's there, but it's not available unless you go through IT."	Data Silos and Access Issues

Thematic analysis of 15 semi-structured interviews revealed five central themes: conceptual ambiguity, belief in potential, organizational barriers, lack of professional development, and data access challenges. Participants expressed uncertainty in defining dark data, often associating it with "clinical notes" or "data that's not used yet." Despite this ambiguity, nearly all respondents acknowledged its value, with one nurse noting, "We're sitting on a goldmine of data that could help prevent readmissions."

However, confidence in using dark data was low, and most attributed this to organizational and technical barriers. Many cited the lack of training as a limiting factor, and several mentioned that data silos or permissions made even motivated attempts futile. These qualitative insights deepen the quantitative findings by explaining *why* high perceived value does not translate into confidence or use. They underscore the need for health systems to clarify dark data's role and invest in hands-on training and cross-functional accessibility.

To enrich the quantitative analysis and gain deeper insight into healthcare professionals' perceptions of untapped dark data, semi-structured interviews were conducted with a purposeful subsample of 15 participants. Participants were selected based on diverse roles and varied responses to confidence and awareness levels in the survey. The interviews were transcribed and analyzed thematically using Braun and Clarke's (2006) six-phase approach. Five key themes emerged from the data: conceptual ambiguity, belief in potential, organizational barriers, lack of professional development, and data access challenges.

Theme 1: Conceptual Ambiguity

Participants exhibited varied understandings of what constitutes "dark data." Some equated it with "clinical notes," "text files," or "patient comments," while others described it more abstractly as "data we don't use but might have value." This lack of definitional clarity reflects what Rogers (2003) describes as a barrier to early-stage diffusion: innovation complexity. Without a consistent conceptual understanding, participants struggled to articulate how dark data could be integrated into practice.

"I think dark data is like… anything we collect but never look at—maybe notes or discharge summaries. But honestly, I'm not sure what it really includes." (Participant 4, Nurse)

This ambiguity helps explain the weak relationship found between awareness and perceived value in the quantitative results (RQ4), suggesting that professionals may conceptually endorse dark data without fully understanding its scope or application.

Theme 2: Belief in Potential

Despite confusion over its definition, participants widely expressed strong belief in the transformative value of dark data for patient care and system improvement.

"We're sitting on a goldmine of data that could help prevent readmissions, spot warning signs earlier… we just don't know how to use it." (Participant 8, Administrator)

This theme aligns with the quantitative finding that perceived value was significantly high ($M = 3.82$, $p < .001$), supporting the relevance of TAM's *perceived usefulness* construct (Davis, 1989). However, belief in potential did not correspond to higher confidence or action, revealing a gap between conceptual support and applied readiness.

Theme 3: Organizational Barriers

Participants consistently described organizational hurdles—such as lack of leadership support, data silos, and restrictive IT systems—that prevented the effective use of dark data.

"Even if you know what data you want to access, there are so many layers—you need permission from IT, compliance, sometimes even legal." (Participant 10, Physician)

This theme parallels the quantitative finding that perceived barriers were moderate and uniform across roles and experience (RQ3), reinforcing the interpretation that these challenges are systemic rather than individual.

Theme 4: Lack of Professional Development

Many participants expressed frustration with the absence of structured training on data interpretation and use. Several stated they had never received formal education about dark data or its applications, despite working in data-rich environments.

"We never had training on this. Even when new systems are introduced, we just get a one-hour demo, not how to actually use the data inside." (Participant 2, Analyst)

This lack of professional development provides a likely explanation for why awareness and perceived value did not predict confidence in using dark data (RQ5), as confirmed in the regression model. It also highlights a failure of diffusion-support structures as emphasized in DOI theory.

Theme 5: Data Access and Interoperability Issues

Even among those who understood dark data and wanted to use it, accessibility was a recurring challenge. Participants cited siloed systems, poor integration between departments, and lack of user-friendly tools as major impediments.

"It's there—we collect all this info—but good luck finding it when you actually need it. It's buried or locked away." (Participant 11, Nurse)

This reinforces earlier literature (Elshazly & Cheng, 2025; Sarker et al., 2024), which cites data interoperability as a core barrier to leveraging healthcare analytics and AI tools.

Integration with Theoretical Framework

From a theoretical perspective, the findings challenge some core assumptions of the Technology Acceptance Model (TAM). While TAM posits that awareness and perceived usefulness lead to confidence and adoption, this study found that despite strong perceived value, confidence remained low, suggesting that other mediators, such as organizational infrastructure and training, are necessary to bridge the gap. Similarly, in light of Diffusion of Innovations (DOI) theory, the findings reflect limited progression past the *knowledge* and *persuasion* stages of adoption, due to insufficient institutional support for implementation (*decision* and *confirmation* stages).

Seminal studies such as those by Venkatesh et al. (2003) and more recent applications in health IT adoption (e.g., Akpan et al., 2023; Banger et al., 2024) support the importance of *facilitating conditions* in predicting confidence and usage. These were clearly lacking among participants in this study, highlighting a readiness gap that has implications for both theory and practice.

V. DISCUSSION

The joint consideration of quantitative and qualitative results offers several inferences worthy of academic synthesis. While survey data shows strong belief in dark data's value, qualitative interviews reveal that this belief is aspirational, not experience-based. This disconnect is critical. According to the Technology Acceptance Model (Davis, 1989), *perceived usefulness* should correlate with *behavioral intention* and *actual use*; however, in this study, high perceived value did not translate into high confidence or readiness. The Diffusion of Innovations framework (Rogers, 2003) also illustrates this gap: professionals may reside in the persuasion stage but lack organizational conditions (training, trialability, observability) to proceed to implementation A side-by-side comparison of quantitative and qualitative results reveals alignment and divergence across data strands. (See table 12).

Table 12. Joint Display for the Study's findings

Research Question	Quantitative Result	Qualitative Theme	Interpretation
RQ1: Awareness	Moderate awareness (M = 3.5)	Conceptual Ambiguity	Professionals are aware of dark data but define it variably
RQ2: Value	High perceived value (p < .001)	Belief in Potential	Positive attitudes exist despite limited usage
RQ3: Barriers	Moderate, no role-based differences	Organizational Barriers; Access Challenges	System-wide issues limit implementation
RQ4: Awareness–Value	No correlation (r = –.06, p = .38)	Ambiguity, not linked to informed beliefs	Belief may be aspirational rather than experiential
RQ5: Predicting Confidence	Model not significant (R^2 = .041, p = .44)	Lack of Development, Systemic Gaps	Confidence tied to support structures, not perceptions

Quantitative Data

The quantitative data, gathered from 220 healthcare professionals via a Likert-scale survey, revealed varied insights into awareness, perceived value, perceived barriers, and confidence regarding dark data.

RQ1: Dark Data Use Awareness level. The analysis indicated that healthcare professionals possess a moderate level of awareness regarding dark data, with a mean score of 3.50 on a 5-point scale (M = 3.5, SD = 0.95). This suggests a foundational familiarity with the concept yet reveals that understanding is neither deep nor consistent across the sector. The distribution was approximately normal, highlighting variability across roles and experience levels; however, the relatively low variation across job functions implies that limited awareness is a systemic concern rather than a role-specific issue. These findings are consistent with prior research (e.g., Komprise, 2024; Sarker et al., 2024; Elshazly & Cheng, 2025), which identifies definitional ambiguity, conceptual inconsistency, and insufficient training as contributing factors to the persistent under-recognition and underutilization of dark data in clinical environments.

RQ2: Perceived Value of Dark Data. Healthcare professionals reported a significantly high perceived value of dark data (M = 3.82, SD = 0.99), as confirmed by a one-sample t-test [(t(219) = 17.30, p < .001)], with a very large effect size (Cohen's d = 1.84). This strong valuation suggests that participants believe dark data, such as clinician notes, unstructured inputs, and wearable data holds substantial potential to improve patient outcomes and operational efficiency. These findings align with previous studies (RoseAID, 2024; Raina et al., 2024), which demonstrated the practical benefits of integrating unstructured data into predictive analytics and personalized care models. Notably, this enthusiasm exists despite only moderate awareness of what constitutes dark data, reflecting a cognitive disconnect between belief in its value and actual understanding or readiness. This gap suggests that perceived value may be driven more by abstract optimism than by concrete familiarity or system-level preparedness, a theme revisited in subsequent findings

RQ3: Perceived Barriers. Descriptive statistics revealed a moderate overall perception of barriers to dark data use (M = 3.16, SD = 0.61), with noticeable variability at the individual level but little variation across professional roles or experience levels. One-way ANOVA results confirmed no significant differences in perceived barriers by role [(F(4, 215) = 1.15, p = .33)] or by years of experience [(F(4, 215) = 0.94, p = .44)], suggesting a consistent recognition of obstacles across the healthcare workforce. These shared perceptions point to systemic barriers, such as inadequate IT infrastructure, data governance limitations, and privacy concerns, rather than individual or role-specific issues. This interpretation aligns with

existing literature (Abeywardena & Samarasinghe, 2024; Ahmed & Al Refai, 2024), which underscores that structural and organizational constraints are the primary hindrances to the effective utilization of dark data in clinical settings.

RQ4: Relationship Between Awareness and Perceived Value. Pearson correlation analysis revealed no significant relationship between awareness of dark data and its perceived value [($r = –.06$, $p = .38$)], indicating that healthcare professionals' belief in the utility of dark data is not strongly associated with their actual understanding of it. This result diverges from theoretical expectations grounded in the Technology Acceptance Model (Davis, 1989) and the Diffusion of Innovations Theory (Rogers, 2003), both of which propose that familiarity with a technology enhances its perceived usefulness. A plausible explanation is that professionals may hold aspirational or conceptual beliefs about dark data's benefits without possessing sufficient knowledge of its practical application. This interpretation is supported by Sarker et al. (2024), who reported that clinicians often overestimate the value of unstructured data despite low levels of literacy or hands-on engagement. These findings suggest that perceptions of value may be shaped more by conceptual endorsement than by informed experience, raising important implications for training, communication, and policy development.

RQ5: Predictors of Confidence in Using Dark Data. A multiple linear regression analysis was conducted to assess whether healthcare professionals' role, years of experience, awareness, and perceived value of dark data could predict their confidence in using it. The overall model was not statistically significant [($F(9, 210) = 1.00$, $p = .44$, $R^2 = .041$)], indicating limited explanatory power. Among the predictors, only the nursing role was marginally significant [$B = 0.20$, $p = .041$)], with nurses reporting slightly higher confidence than administrators. However, awareness, perceived value, and experience did not significantly contribute to confidence levels. These results challenge core assumptions of the Technology Acceptance Model (Davis, 1989) and the Diffusion of Innovations Theory (Rogers, 2003), which posit that increased knowledge and perceived usefulness should drive greater confidence and adoption. Instead, the findings suggest that healthcare professionals may see their confidence in using dark data as contingent on external factors, such as access to training, organizational support, and integrated digital tools, rather than on personal familiarity or professional status. This points to the presence of latent variables not captured in the current model and underscores the need for broader system-level interventions to build capacity.

The disconnect between value recognition and practical readiness mirrors broader challenges in healthcare innovation. As Banger et al. (2024) noted, high enthusiasm for data-driven approaches often coexists with low self-efficacy, particularly in environments lacking structured educational or technological infrastructure. The lack of predictive value in awareness and perceived usefulness underscores the importance of contextual factors, such as organizational culture, access to tools, and continuous learning environments, shaping professional confidence.

Qualitative Insights

Semi-structured interviews with 15 healthcare professionals revealed deeper contextual insights that complemented the statistical findings. Five interrelated themes emerged from the data, each illuminating critical dimensions of the dark data implementation landscape.

*Conceptual Ambiguity***:** Interviewees expressed notable uncertainty regarding the definition of dark data, using varied terms such as "notes," "unstructured inputs," and "unused data." This semantic inconsistency reflects the ambiguity that may explain the lack of correlation between awareness and perceived value found in the quantitative phase. Without a unified understanding, stakeholders struggle to move from theoretical recognition to practical utilization.

*Belief in Potential***:** Despite definitional challenges, most participants acknowledged the vast promise of dark data, describing it as a "goldmine" for improving predictive analytics, population health monitoring, and workflow optimization. However, their belief did not correspond with confidence in application or usage, echoing the disconnect found in RQ2 and RQ5. This supports the notion that high perceived value alone is insufficient to drive behavioral adoption.

Organizational Barriers: The interviews uncovered entrenched institutional obstacles, including siloed systems, restricted data access, and lack of leadership advocacy. These barriers were uniformly reported across professional roles, aligning with the ANOVA results (RQ3) that found no significant differences in perceived obstacles by experience or position. The limited predictive power of regression models for confidence (RQ5) is further contextualized by these structural inhibitors.

Lack of Professional Development: A recurring theme was the dearth of formal training, onboarding, or continuing education programs focused on dark data literacy. Participants expressed a need for skill-building in data interpretation and

tool navigation, which mirrors the low confidence ratings identified in the quantitative strand. This gap points to critical weaknesses in implementation support structures.

Data Access Challenges: Fragmented IT infrastructures and role-based restrictions were cited as ongoing impediments to meaningful engagement with dark data. Even highly motivated professionals found themselves constrained by system architecture and policy silos. This reinforces broader findings in the literature (e.g., Sarker et al., 2024) on the infrastructural fragmentation hindering digital transformation in healthcare.

High perceived value (RQ2) was echoed in qualitative statements about the "goldmine" potential of dark data. However, the lack of correlation between awareness and value (RQ4) is explained by participants' conceptual ambiguity. The regression model's weak explanatory power (RQ5) is contextualized by themes pointing to systemic and educational barriers. The absence of role-based differences in perceived barriers (RQ3) supports the qualitative finding that obstacles are organizational rather than individual.

These integrated findings reinforce the study's central conclusion: while healthcare professionals conceptually support the use of dark data, systemic constraints, lack of training, and access limitations hinder practical engagement. Bridging this readiness gap requires strategic investment in governance, education, and workflow integration.

Integrated Synthesis and Theoretical Implications

The joint consideration of quantitative and qualitative findings offers several key insights. First, while the survey data show strong belief in dark data's value, the interviews reveal that this belief is aspirational—not experience-based. This disconnect is critical. According to Davis (1989), perceived usefulness should drive use, yet this study found no corresponding increase in confidence or readiness. Similarly, Rogers' (2003) diffusion theory illustrates that favorable attitudes (persuasion) must be supported by enabling conditions to result in adoption. Many professionals in this study appear stuck at the persuasion stage, hindered by structural gaps.

Together, these findings identify a critical readiness gap in the utilization of dark data. While healthcare professionals express strong support for the concept and potential of dark data, they lack the access, definitional clarity, training, and leadership engagement necessary to translate belief into practice. The results underscore an opportunity to strengthen operational and organizational hardiness, defined in this context as the institutional capacity to support dark data engagement through cognitive resilience, data literacy, and adaptable infrastructure. Without deliberate efforts to develop these capabilities, dark data is likely to remain a conceptual ideal rather than a functional asset in healthcare practice. Without investment in these foundational areas, dark data will remain an underutilized strategic asset. The implications are significant: healthcare systems must go beyond awareness-building and commit to developing integrated, system-level strategies that address the practical, structural, and cultural barriers to dark data adoption. These findings support calls in the literature (Elshazly & Cheng, 2025; SPD Technology, 2025) to move from abstract optimism to concrete, actionable readiness.

Implications for Policy and Practice

The findings of this study have important implications for both healthcare practice and organizational policy. First, the moderate levels of awareness but high perceived value of dark data suggest a readiness among healthcare professionals to support data-driven transformation, if appropriate infrastructure and education are in place. This presents an opportunity for healthcare leaders and IT strategists to harness positive attitudes by implementing structured dark data initiatives that are supported by training, workflow integration, and institutional buy-in.

Second, the lack of significant differences in perceived barriers across roles and experience levels indicates that these challenges are systemic rather than individual. Therefore, interventions must operate at the organizational or systems level, rather than focusing narrowly on role-specific training. Data governance policies, privacy protections, and investment in data literacy should be addressed in tandem to create an enabling environment for dark data utilization.

Third, the non-significant relationship between awareness and perceived value, as well as the failure of these variables to predict confidence, challenges assumptions embedded in traditional adoption models such as the Technology Acceptance Model. This underscores the need for a more nuanced approach to technology adoption in healthcare, one that considers not just perceptions but also contextual enablers such as clinical decision support tools, automation, and interdisciplinary collaboration.

Recommendations for Education and Research

To bridge this readiness gap and harness the strategic potential of dark data, several evidence-based recommendations emerge:

1. Establish Dark Data Education Programs: Develop targeted workshops and micro-credential initiatives to enhance clinicians' and administrators' data literacy, with a focus on interpreting unstructured information such as clinical notes and wearable device outputs.
2. Integrate Dark Data Tools into Workflow: Deploy user-friendly tools powered by natural language processing (NLP) and artificial intelligence (AI) to embed dark data insights into routine electronic health record (EHR) operations.
3. Create Organizational Incentives: Motivate adoption through recognition programs, seed funding for innovation pilots, and institutional support for clinician-led analytics initiatives that demonstrate measurable outcomes.
4. Update Governance Frameworks: Align institutional policies with modern data governance principles to ensure ethical AI usage, data privacy, and accountability, thereby reducing resistance and fostering trust.
5. Future Research Directions: Conduct qualitative investigations to explore the psychological and cultural drivers of low confidence, examine organizational and team-level influences on adoption, and develop a validated index for assessing institutional dark data readiness.

These synthesized findings highlight the critical need for systemic investments in education, infrastructure, and leadership to unlock the value of dark data in healthcare. Institutions that foster institutional and operational hardiness, defined as the resilience, agility, and cross-functional adaptability required to transform complex data systems into actionable knowledge will be best positioned to transition from conceptual potential to practical, value-driven impact.

Limitations

This study has several limitations that should be acknowledged when interpreting the findings. First, the use of self-reported data through surveys and interviews introduces the possibility of response bias, particularly social desirability effects that may lead participants to overstate their awareness or enthusiasm about dark data. Second, the sample was limited to healthcare professionals who opted in voluntarily, which may have resulted in the overrepresentation of individuals with a predisposed interest in data or technology. Third, while the qualitative interviews provided rich insights, the relatively small sample size (n = 15) may not fully capture the diversity of experiences across different healthcare systems, specialties, or geographic regions.

Additionally, the quantitative analysis was cross-sectional and descriptive, limiting causal inference and temporal understanding of evolving attitudes toward dark data. Although regression models were used, the low explanatory power indicates that other unmeasured variables may influence confidence and adoption. Finally, the conceptual fluidity surrounding the term "dark data" itself likely influenced both survey responses and interview interpretations, making consistent measurement and thematic saturation more difficult to achieve.

Future studies would benefit from longitudinal designs, more diverse samples, and triangulated measurement approaches, including observational and usage data to better capture actual dark data engagement rather than only self-perceived metrics.

Summary

This study explored healthcare professionals' perceptions of untapped dark data and its potential for improving patient care and operational performance. Grounded in the Technology Acceptance Model (TAM) and Diffusion of Innovations (DOI) theory, the study examined how awareness, perceived value, perceived barriers, and confidence interact to shape readiness for dark data utilization within healthcare settings. Using a quantitative, cross-sectional design with a diverse sample of 220 healthcare professionals, the research addressed five core questions.

The findings revealed that while healthcare professionals possess moderate awareness of dark data, they strongly believe in its value. This belief was statistically significant and demonstrated a very large effect size. However, perceived barriers related to infrastructure, training, and governance were moderately high and consistent across professional roles and experience levels, suggesting systemic challenges rather than individual or role-specific ones. Surprisingly, the study found no significant relationship between awareness and perceived value, nor did these factors significantly predict confidence in

using dark data. Only one professional role (nurses) reported marginally higher confidence, but overall predictive power was low. These results present a paradox: healthcare professionals recognize the potential of dark data but lack the tools, clarity, and support to confidently engage with it. The disconnect between belief and action underscores a critical need for health systems to move beyond awareness campaigns and toward practical infrastructure investments, cross-disciplinary training, and policy reform.

VI. CONCLUSION

The central research question posed by this study, whether untapped dark data constitutes a missed opportunity—can be answered in the affirmative. Findings from both the quantitative and qualitative arms of the study support the conclusion that dark data remains an underutilized asset within healthcare systems, not due to disinterest or skepticism, but due to systemic gaps in awareness, confidence, and structural support.

Professionals recognize the value of dark data, as demonstrated by statistically significant value ratings and strong qualitative endorsements. Yet the absence of alignment between this perceived value and actual readiness to implement its use underscores a persistent operational lag. Interviews confirmed that without clear definitions, access protocols, and training pathways, dark data remains aspirational. Even among willing professionals, confidence in utilization is hindered by fragmented infrastructure, policy ambiguity, and insufficient professional development.

Thus, while dark data is not ignored, it is not yet activated. The opportunity cost lies in unleveraged clinical insights, missed predictive analytics, and inefficiencies that could otherwise be resolved through intelligent data integration. To reverse this trend, healthcare systems must build institutional and operational hardiness, cultivating resilient structures, policies, and cultures capable of absorbing innovation. When paired with targeted education, robust governance, and accessible tools, dark data can move from potential to performance, positioning it as a strategic resource in modern healthcare transformation.

REFERENCES

[1] Boud, D., & Molloy, E. (2013). *Feedback in Higher and Professional Education: Understanding It and Doing It Well*. Routledge.

[2] Brookhart, S. M. (2017). *How to Give Effective Feedback to Your Students*. ASCD.

[3] Hattie, J., & Timperley, H. (2007). The power of feedback. *Review of Educational Research, 77*(1), 81–112.

[4] Fisher, S. (2024). Uncovering dark data in healthcare organizations. https://parivedasolutions.com/resources/uncovering-dark-data-in-healthcare-organizations/?utm_source=chatgpt.com

[5] Wiliam, D. (2011). *Embedded Formative Assessment*. Solution Tree Press.

[6] Creswell, J. W., & Creswell, J. D. (2018). *Research design: Qualitative, quantitative, and mixed methods approaches* (5th ed.). Sage Publications.

[7] George, D., & Mallery, P. (2019). *IBM SPSS Statistics 25 Step by Step: A Simple Guide and Reference* (15th ed.). Routledge.

[8] Davis, F. D. (1989). Perceived usefulness, perceived ease of use, and user acceptance of information technology. *MIS Quarterly, 13*(3), 319–340.

[9] Rogers, E. M. (2003). *Diffusion of innovations* (5th ed.). Free Press.

[10] Abeywardena, N., & Samarasinghe, T. (2024). *Barriers to Integrating Unstructured Data in Public Health Systems*. Journal of Health Informatics, 22(1), 45–60.

[11] Davis, F. D. (1989). Perceived usefulness, perceived ease of use, and user acceptance of information technology. *MIS Quarterly, 13*(3), 319–340.

[12] Greenhalgh, T., Robert, G., Macfarlane, F., Bate, P., & Kyriakidou, O. (2004). Diffusion of innovations in service organizations: Systematic review and recommendations. *Milbank Quarterly, 82*(4), 581–629.

[13] Holden, R. J., & Karsh, B. T. (2010). The Technology Acceptance Model: Its past and its future in health care. *Journal of Biomedical Informatics, 43*(1), 159–172.

[14] Raina, A., Iyer, S., & Sundar, M. (2024). *Intelligent Clinical Documentation with Generative AI: Reducing Clinician Burden and Unlocking Data Value*. arXiv. https://arxiv.org/abs/2405.18346

[15] Rogers, E. M. (2003). *Diffusion of Innovations* (5th ed.). Free Press.

[16] Abeywardena, N., & Samarasinghe, T. (2024). *Barriers to Integrating Unstructured Data in Public Health Systems*. Journal of Health Informatics, 22(1), 45–60.

[17] Ahmed, R., & Al-Refai, M. (2024). *Data Governance in Healthcare: A Framework for Managing Privacy and Risk*. MPRA Paper No. 122776. https://mpra.ub.uni-muenchen.de/122776/

[18] Banger, A., Raza, S., & Qureshi, M. (2024). *Data Literacy in Healthcare: Bridging the Competency Gap Among Professionals*. Journal of Medical Systems, 48(2), 1–12.

[19] Boud, D., & Molloy, E. (2013). *Feedback in Higher and Professional Education: Understanding It and Doing It Well*. Routledge.

[20] Elshazly, R., & Cheng, J. (2025). *Artificial Intelligence and Patient Privacy: Emerging Challenges in Digital Health Governance*. arXiv. https://arxiv.org/abs/2501.01639

[21] Komprise. (2024). *Dark Data in Healthcare: Challenges and Solutions*. https://www.komprise.com/glossary_terms/dark-data/

[22] Raina, A., Iyer, S., & Sundar, M. (2024). *Intelligent Clinical Documentation with Generative AI: Reducing Clinician Burden and Unlocking Data Value*. arXiv. https://arxiv.org/abs/2405.18346

[23] RoseAID. (2024). *Relevance of Dark Data in Clinical Analytics*. https://roseaid.com/2024/10/30/relevance-of-dark-data-in-healthcare/

[24] SPD Technology. (2025). *The 2025 Guide to Healthcare Data Analytics*. https://spd.tech/data/healthcare-data-analytics-transforming-patient-care/

[25] Sarker, T., Lee, Y., & Uddin, S. (2024). *Dark Data in Healthcare: Perceptions and Utilization Gaps Among Clinicians*. BMC Medical Informatics and Decision Making, 24(1), 1–12.

[26] Field, A. (2018). *Discovering statistics using IBM SPSS statistics* (5th ed.). Sage Publications.

[27] Solutyics. (n.d.). Dark data: overview, opportunities, and challenges. https://solutyics.com/dark-data-overview-opportunities-and-challenges/?utm_source=chatgpt.com

[28] Bouton, C. (2021). Dark Data in Healthcare: Health IT's Next Big Opportunity? https://hitconsultant.net/2021/12/03/dark-data-in-healthcare-health-it-opportunity/?utm_source=chatgpt.com

[29] H1. (n.d.). The challenges of unstructured healthcare data. https://h1.co/blog/the-challenges-of-unstructured-healthcare-data/?utm_source=chatgpt.com

[30] Minou, J., Mantas, J., Malamateniou, F., & Kaitelidou, D. (2020). Health Professionals' Perception about Big Data Technology in Greece. *Acta informatica medica : AIM : journal of the Society for Medical Informatics of Bosnia & Herzegovina : casopis Drustva za medicinsku informatiku BiH*, *28*(1), 48–51. https://doi.org/10.5455/aim.2020.28.48-51

[31] Rogers, E. M. (1962). *Diffusion of innovations* (1st ed.). Free Press of Glencoe.

[32] Davis, F. D. (1989). Perceived usefulness, perceived ease of use, and user acceptance of information technology. *MIS Quarterly, 13*(3), 319–340. https://doi.org/10.2307/249008