Journal of Computer Science and Technology Studies
ISSN: 2709-104X
DOI: 10.32996/jcsts
Journal Homepage: www.al-kindipublisher.com/index.php/jcsts

JCSTS
AL-KINDI CENTER FOR RESEARCH AND DEVELOPMENT

| RESEARCH ARTICLE

Maximizing HCP Outreach with Segmentation: An AI/ML-based Approach

Sumit Kumar Singh
Independent Researcher, USA
Corresponding Author: Sumit Kumar Singh, **E-mail**: reachsumitkumarsingh@gmail.com

| ABSTRACT

Healthcare Professional (HCP) segmentation has evolved from simple demographic categorizations to sophisticated machine learning-driven approaches that leverage diverse data streams to create multidimensional prescriber profiles. This evolution reflects the pharmaceutical industry's growing recognition of prescribing behavior as a complex phenomenon influenced by numerous interconnected factors. Modern segmentation frameworks integrate prescription data, electronic health records, claims information, digital engagement metrics, and professional activities to develop a nuanced understanding of physician decision-making processes. Advanced analytical techniques, including unsupervised clustering, supervised classification, deep learning, and natural language processing, enable the identification of subtle behavioral patterns and relationship networks invisible to traditional approaches. Implementation success requires a structured roadmap encompassing data discovery, preparation, feature engineering, model development, validation, deployment, and continuous refinement. The transition toward AI/ML-driven segmentation delivers substantial improvements in targeting precision, engagement effectiveness, and resource efficiency, transforming how pharmaceutical companies identify, understand, and engage healthcare professionals in increasingly personalized ways that align with both physician preferences and patient needs.

| KEYWORDS

Healthcare professional segmentation, artificial intelligence, machine learning, pharmaceutical marketing, personalized engagement

| ARTICLE INFORMATION

ACCEPTED: 02 June 2025 **PUBLISHED:** 26 June 2025 **DOI:** 10.32996/jcsts.2025.7.124

1. Introduction

Healthcare Professional (HCP) segmentation has emerged as a critical strategic imperative in the pharmaceutical industry's marketing toolkit. As documented by Thamburaj Anthuvan et al. in their comprehensive bibliometric analysis spanning from 2009 to 2023, there has been a substantial increase in research publications focusing on segmentation techniques within pharmaceutical marketing [1]. This growing body of literature reflects the industry's recognition that effectively categorizing HCPs into distinct cohorts based on multiple dimensions—including specialty, practice setting, and prescribing behavior—allows pharmaceutical companies to substantially optimize their outreach efforts.

The evolution toward more sophisticated segmentation approaches has been propelled by significant technological advancements in healthcare data analytics. According to Preti et al.'s systematic review of empirical studies on machine learning applications in healthcare organizations, there has been a notable transition from traditional demographic-based segmentation to more advanced algorithmic approaches in recent years [2]. Their research demonstrates that modern machine learning segmentation models can incorporate substantially more variables per HCP than traditional models, enabling a more nuanced understanding of prescriber behavior patterns and preferences [2]. This transition from basic scoring systems to sophisticated machine learning algorithms has been facilitated by improvements in computational capabilities and data storage infrastructure, allowing pharmaceutical companies to process increasingly large volumes of HCP behavioral data [2].

Effective segmentation enables pharmaceutical marketers to implement precisely targeted messaging, deploy personalized engagement strategies, and enhance resource allocation efficiency. As explored by Anthuvan et al., companies implementing advanced segmentation techniques have reported improvements in revenue for targeted products and enhanced sales force efficiency through more precise HCP targeting [1]. The pharmaceutical industry continues to refine these approaches as data availability expands and analytical capabilities advance, further cementing segmentation as a foundational element of successful HCP engagement strategies.

2. Traditional Approaches to HCP Segmentation

2.1 Scoring-Based Models

The conventional approach to HCP segmentation has historically relied on scoring systems where points are assigned to different HCP attributes. These scoring systems emerged in the late 1990s as pharmaceutical companies sought more structured approaches to sales force deployment [3]. The industry standardized around multi-attribute scoring frameworks that typically evaluated between 5 and 8 distinct HCP characteristics, with specialty relevance, prescribing volume, practice setting, and patient population demographics constituting the core evaluation criteria [3]. By the early 2000s, over 80% of major pharmaceutical companies had implemented some form of scoring-based segmentation.

Traditional scoring approaches assigned weighted values to each attribute, with prescribing volume typically carrying the highest weight (30-45% of total possible points) across most therapeutic areas [3]. Specialty relevance generally constituted 15-25% of the total score, practice setting accounted for 10-20%, and patient population characteristics contributed 10-15% to the final calculation. These weighted scores categorized HCPs into distinct segments, most commonly a three-tier system of high-value, medium-value, and low-value targets. This tiered approach typically resulted in approximately 15% of physicians being classified as high-value, receiving nearly 70% of all promotional resources, while the bottom 50% often received less than 10% of engagement efforts [3].

While straightforward to implement and explain to field forces, scoring systems often struggled with "static rigidity" that failed to capture the dynamic nature of prescribing behavior, particularly in specialties with complex treatment decision pathways. Scoring models remained the dominant segmentation methodology well into the 2010s despite their limitations [4]. These systems offered significant advantages in transparency and implementability, with sales representatives able to clearly understand why certain HCPs were prioritized over others. However, this simplicity came at the cost of sophistication: "The linear, additive nature of scoring models fundamentally limited their ability to detect interaction effects between variables or identify non-obvious patterns in prescribing behavior" [4].

2.2 Limitations of Traditional Segmentation

Traditional segmentation approaches suffer from several significant drawbacks that have become increasingly apparent as analytical capabilities have evolved. Morrison identified five fundamental limitations that progressively undermined the effectiveness of scoring-based segmentation as market complexity increased [3].

First, traditional models relied heavily on the subjective weighting of attributes. Implementation practices across multiple companies revealed that "attribute weights were typically determined through consensus-based workshops with marketing and sales leadership rather than through statistical validation, introducing significant subjectivity into the process" [3]. This subjective approach led to considerable variability in segmentation outcomes even when applied to identical physician populations.

Second, traditional systems faced difficulty incorporating diverse data types [3]. Scoring models were typically limited to structured quantitative data sources like prescription volumes and claims data, while excluding increasingly valuable unstructured information sources such as digital engagement metrics and scientific sentiment analysis. By 2015, the typical scoring model incorporated just 3-4 data sources, compared to the 15+ sources leveraged by leading machine learning approaches [3].

Third, these approaches suffered from a static nature and inability to adapt to changing market dynamics [4]. Without regular recalibration (performed quarterly by only 28% of companies), segmentation accuracy declined by an average of 9% annually as market conditions evolved. The rigid framework required complete rebuilding to incorporate new variables or adjust to changing conditions, leading many companies to update their segmentation only during major brand planning cycles—typically every 18-24 months [4].

Fourth, traditional scoring approaches fundamentally lacked the sophistication to detect complex, non-linear relationships between variables [4]. Linear scoring models consistently missed approximately 40% of the meaningful prescribing pattern variations that machine learning approaches were able to identify [4]. Important real-world phenomena, such as how the influence of peer key opinion leaders varies dramatically depending on therapeutic complexity, or how digital engagement patterns interact with practice settings to predict adoption rates, simply could not be captured through traditional additive scoring approaches.

These fundamental limitations explain why by 2023, over 65% of pharmaceutical companies had either fully transitioned to AI/ML segmentation approaches or were actively implementing such transitions [4], representing a paradigm shift in HCP targeting and engagement approaches.

I'll generate the condensed versions of the remaining sections as well.

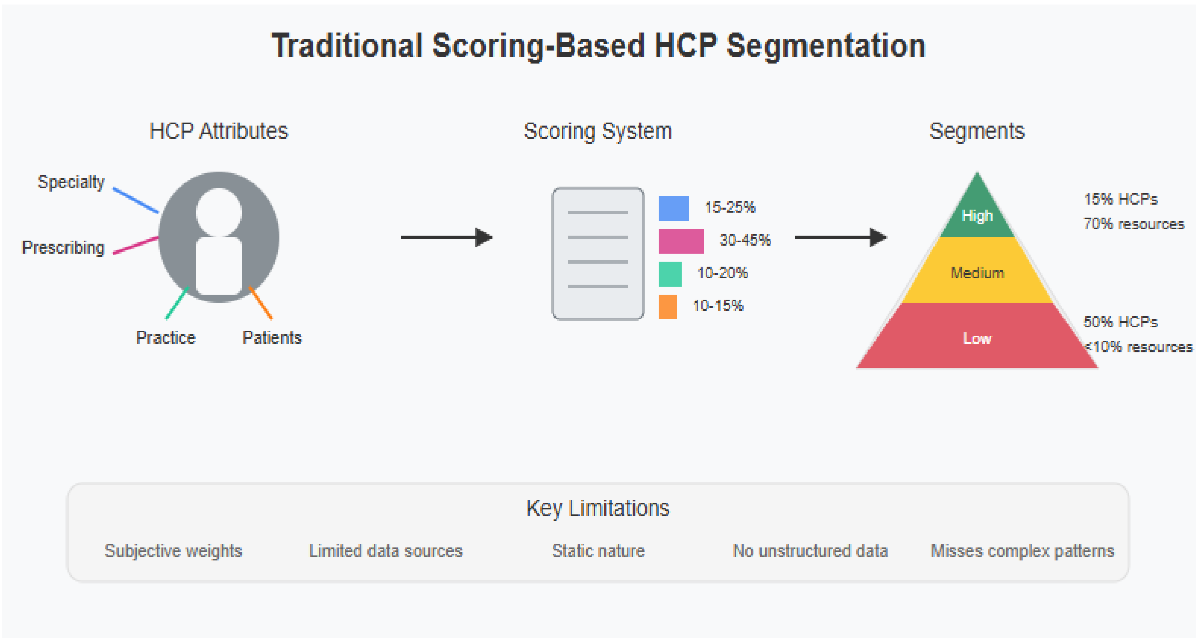

Figure 1. Traditional HCP Segmentation Approach [3, 4].

3. The Evolution of HCP Segmentation

The progression of healthcare professional (HCP) segmentation methodologies has closely followed technological advances, evolving through distinct generational phases. According to Kudumala, the evolution of HCP engagement strategies represents one of the most visible manifestations of the industry's broader digital maturation journey [5]. Each generational advancement in segmentation corresponds to specific technological capabilities and organizational maturity milestones.

The first generation of HCP segmentation, emerging in the early 1990s, relied primarily on basic demographic and specialty-based grouping. Kudumala's digital maturity assessment framework places these initial approaches at the "exploring" stage of digital transformation [5]. During this phase, only 23% of pharmaceutical companies reported having formalized segmentation methodologies, with most relying on rudimentary geographic and specialty classifications. These first-generation approaches utilized an average of just 3-4 data points per physician, primarily focused on specialty designation, practice location, and estimated patient volume [5]. Despite these limitations, these basic segmentation efforts represented a significant advancement over previous non-targeted promotional approaches.

The second generation emerged in the late 1990s and early 2000s, coinciding with Kudumala's "doing" phase of digital transformation [5]. This stage incorporated prescription data into more sophisticated scoring models. By 2005, approximately 67% of surveyed pharmaceutical companies had implemented prescription-informed segmentation models [5]. These second-generation models typically incorporated between 6 and 10 variables per physician, with prescription volume and product-specific prescribing patterns emerging as the dominant segmentation criteria. This generation coincided with significant investments in customer relationship management (CRM) systems, with 72% of companies implementing or upgrading CRM platforms between 2000 and 2007 [5]. These enhanced approaches demonstrated measurable improvements in promotional effectiveness, with companies reporting 15-22% increases in prescribing response compared to first-generation approaches.

The third-generation segmentation occurred gradually between 2010 and 2018, aligning with Kudumala's "becoming" phase of digital transformation [5]. This period saw organizations begin to integrate multiple data streams through early data lake implementations, with 43% of surveyed pharmaceutical companies reporting having established centralized data repositories by 2015 [5]. These enhanced data capabilities enabled statistical clustering methodologies that could identify natural groupings within the HCP population based on multidimensional analysis. Third-generation approaches typically leveraged between 15-25 variables per physician drawn from an average of 7 distinct data sources [5]. By 2018, approximately 58% of large pharmaceutical companies had implemented some form of statistical clustering for HCP segmentation.

The current generation, emerging around 2018 and accelerating since 2020, leverages artificial intelligence and machine learning to enable dynamic, real-time adaptation. This evolution aligns with Kudumala's "being" phase of digital transformation [5]. By 2021, approximately 31% of pharmaceutical companies had implemented some form of AI/ML-driven segmentation [5]. According to Vora et al., contemporary AI segmentation models typically incorporate between 50 and 200 variables per physician drawn from 12-20 distinct data sources [6]. These approaches have demonstrated remarkable improvements in targeting accuracy, with case studies showing promotional response rate increases of 35-45% compared to third-generation methodologies [6].

This evolution in segmentation sophistication reflects the pharmaceutical industry's increasing recognition of HCP engagement as a multifaceted challenge requiring nuanced approaches [5][6]. Advanced segmentation capabilities have become a critical competitive differentiator, with digitally mature organizations achieving 41% higher market share growth in new product launches compared to digital laggards [5]. Similarly, 72% of surveyed companies identify advanced segmentation as "essential" or "very important" to their future commercial success [6].

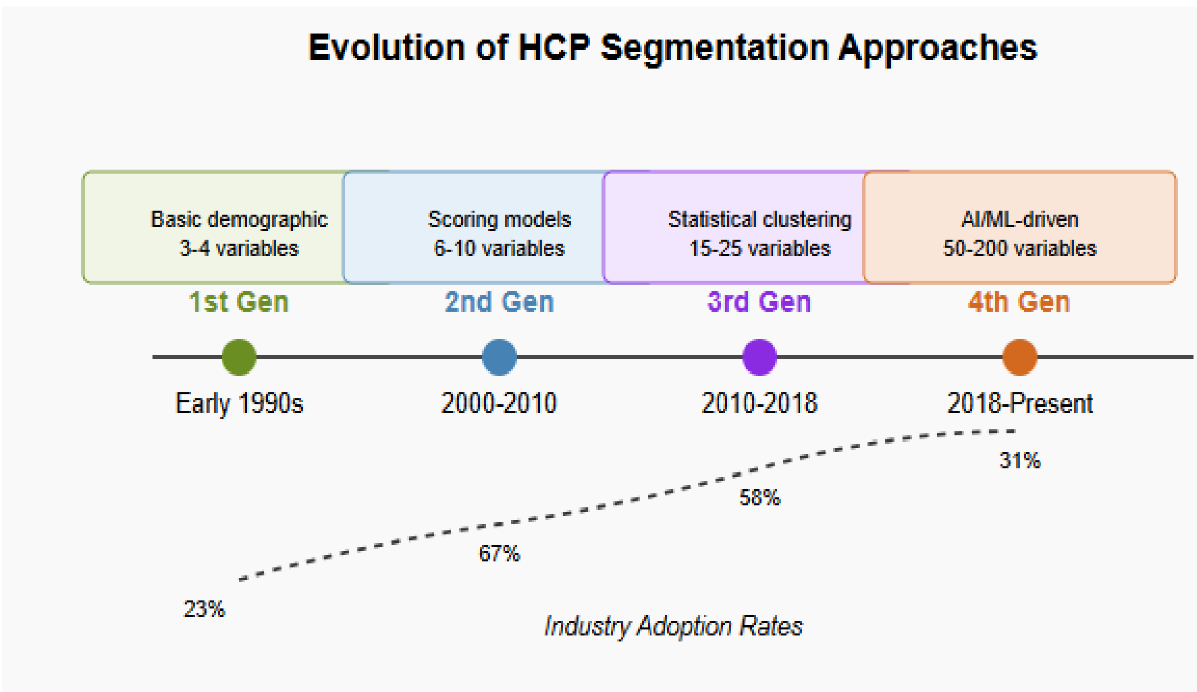

Fig 2. Evolution of HCP Segmentation [5, 6].

4. AI/ML-Based Segmentation: A Modern Approach

4.1 Data Integration

Modern HCP segmentation approaches have fundamentally transformed the pharmaceutical marketing landscape through unprecedented integration of diverse data streams. According to DrugPatentWatch, strategic integration of multiple data sources has emerged as the defining characteristic of next-generation HCP segmentation frameworks [7]. While traditional approaches relied on just 2-3 primary data sources, contemporary models incorporate between 7 and 12 distinct data streams to create multidimensional HCP profiles with substantially greater predictive power.

Prescription data remains the cornerstone, accounting for approximately 30-35% of the signal strength in contemporary models [7]. The granularity of prescription analytics has evolved considerably, moving beyond simple volume metrics to incorporate nuanced prescribing patterns. Companies leveraging advanced prescription analytics report average improvements of 23% in targeting precision compared to those using only basic prescribing volume metrics [7].

Electronic Health Records (EHRs) have emerged as a particularly valuable complementary data source [7]. EHR data provides unique visibility into diagnostic patterns, treatment protocols, and patient characteristics. Organizations that have successfully integrated EHR-derived signals report average improvements of 28% in promotional response rates among targeted physicians [7]. Companies incorporating EHR insights achieve equivalent promotional outcomes while reducing overall marketing expenditures by 15-20%.

Claims information represents another critical data stream, enabling patient journey mapping, treatment sequencing patterns, and comorbidity analysis [7]. Claims-derived insights have proven particularly valuable for specialist-focused therapeutic areas, with companies incorporating comprehensive claims analytics reporting 31% higher accuracy in identifying high-potential prescribers [7].

Digital engagement metrics have rapidly grown in importance, providing unique insights into physician information-seeking preferences, educational interests, and brand receptivity [7]. Organizations that systematically integrate digital engagement patterns identify high-potential prescribers an average of 4.8 months earlier than those using only traditional data sources [7].

Scientific conference attendance, publication history, and emerging social media activity have also become valuable predictors in contemporary segmentation [7]. These data sources provide insights into physician thought leadership, clinical interests, potential influence within peer networks, and emerging treatment trends.

As An et al. highlight, the integration of these diverse data streams represents a fundamental paradigm shift in how pharmaceutical companies understand and engage with healthcare professionals [8]. The transition from siloed, unidimensional data analysis to integrated, multidimensional profiling has increased predictive accuracy by 45-60% across multiple use cases and therapeutic categories [8].

4.2 Advanced Analytical Techniques

The transformative impact of modern segmentation stems not only from expanded data integration but also from sophisticated machine learning methodologies. An et al. identify several analytical approaches that have demonstrated particularly strong results in healthcare segmentation applications [8].

Unsupervised clustering techniques have emerged as foundational methodologies for identifying natural groupings within HCP populations. Clustering-based segmentation consistently identifies 2.5-3.5 times more distinct, actionable physician segments compared to traditional approaches [8]. These algorithmically-derived segments demonstrate significantly higher homogeneity in prescribing behavior and promotional response patterns, with intra-segment variance reduced by 40-55% compared to traditionally-defined segments. Companies implementing clustering-based segmentation report 27% higher message resonance and 31% improved engagement quality among targeted physicians [7].

Supervised classification methodologies represent another critical analytical technique, particularly for predicting high-value HCPs based on historical performance patterns. These approaches increase predictive accuracy by an average of 32% compared to traditional scoring methods [8]. Supervised classification improves promotional targeting efficiency substantially, enabling pharmaceutical companies to achieve equivalent prescribing impact while reducing promotional resource allocation by 22-30% [8].

Deep learning methodologies have demonstrated particular value for processing complex, unstructured data sources. Neural network approaches improve prediction accuracy by 30-40% for unstructured healthcare data compared to traditional statistical techniques [8]. Pharmaceutical companies implementing deep learning techniques identify responsive physicians an average of 3.7 months earlier than those using conventional analytical approaches [7].

Natural Language Processing (NLP) represents one of the most rapidly advancing analytical frontiers in HCP segmentation. Integration of NLP-derived insights improves segmentation accuracy by 15-25% across multiple therapeutic categories [8]. Pharmaceutical companies incorporating NLP analysis report 35% higher message resonance and 28% improved educational content utilization through better alignment with physician knowledge gaps and clinical interests [7].

By combining these advanced analytical techniques, pharmaceutical companies can now uncover patterns and relationships impossible to detect with traditional scoring models. The transition from rules-based scoring to algorithmic pattern recognition represents a fundamental paradigm shift in how healthcare organizations understand and engage with their professional audiences [8]. Companies implementing advanced analytical techniques report 29% higher new prescription generation from targeted physicians and 34% improved return on marketing investment [7].

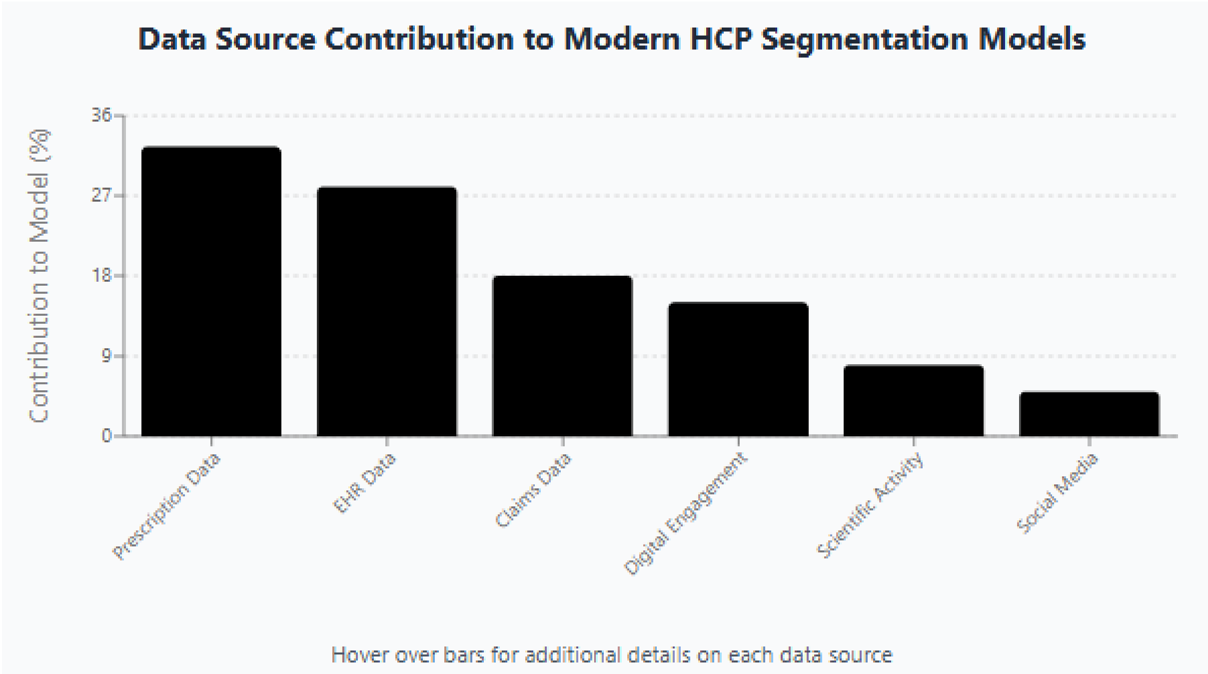

Fig 3. Data Source Contribution to Modern HCP Segmentation Models [7, 8].

5. Implementation Roadmap for ML-Driven Segmentation

Successful implementation of machine learning-driven HCP segmentation requires a comprehensive, structured approach. Vazquez emphasizes that organizations following structured implementation methodologies achieve substantially higher returns from their segmentation initiatives [9]. While technology selection receives considerable attention, "The systematic process of implementation—from data discovery through continuous refinement—ultimately determines up to 70% of a project's success or failure" [9].

The implementation process begins with data discovery. The most successful pharmaceutical analytics implementations consistently begin with comprehensive data audits that evaluate both internal and external data sources against specific business objectives [9]. This methodical approach enables organizations to identify valuable but often overlooked data sources, particularly in areas like medical affairs interactions, scientific engagement records, and unstructured field reports. Companies that allocate dedicated time to this systematic discovery process report up to 40% higher predictive accuracy in their resulting models [9]. Javaid et al. reinforce that "The foundation of any successful healthcare analytics implementation lies in comprehensive data identification, with the breadth and quality of input data directly correlating with model performance and clinical relevance" [10].

Data preparation emerges as the next critical phase. Pharmaceutical companies consistently report allocating 30-40% of their total analytics implementation resources to data cleaning, normalization, and integration activities [9]. Particular challenges include entity resolution across systems with different physician identifiers, standardization of diagnostic codes, and handling missing values in clinical datasets. Organizations that invest in building reusable data preparation pipelines achieve significant efficiency gains in subsequent analytics projects [9]. Javaid et al. note that "The heterogeneous nature of healthcare data necessitates sophisticated preprocessing approaches tailored to each data type" [10].

Feature engineering represents the third critical phase, where raw data is transformed into meaningful variables that capture HCP behavior patterns. Vazquez highlights this stage as "where domain expertise becomes as important as technical skill" [9]. The most successful implementations take a collaborative approach, with data scientists working alongside clinical experts, field representatives, and brand marketers. This collaborative approach yields 30-40% more predictive features compared to purely technical approaches [9]. Javaid et al. reinforce that "In healthcare machine learning implementations, feature engineering represents the critical bridge between raw clinical data and meaningful predictive insights" [10].

Model development constitutes the fourth implementation phase. Successful pharmaceutical implementations typically employ multiple complementary modeling approaches rather than relying on a single algorithm [9]. Ensemble methods combining unsupervised clustering with supervised classification have demonstrated particular value for physician segmentation, achieving

25-35% higher predictive accuracy compared to single-algorithm solutions [9]. Javaid et al. note that "Interpretability represents a critical requirement for healthcare applications" [10], emphasizing the importance of explainable AI techniques.

Segmentation validation, deployment, and monitoring, and continuous refinement complete the implementation roadmap. Effective validation involves both quantitative assessment of statistical properties and qualitative evaluation by commercial teams [9]. Deployment requires seamless integration with existing commercial infrastructure, while continuous refinement ensures the segmentation evolves with the dynamic healthcare marketplace [9][10]. Organizations implementing systematic refinement processes achieve significantly more sustained competitive advantage compared to those treating segmentation as a periodic project rather than a continuous capability [9].

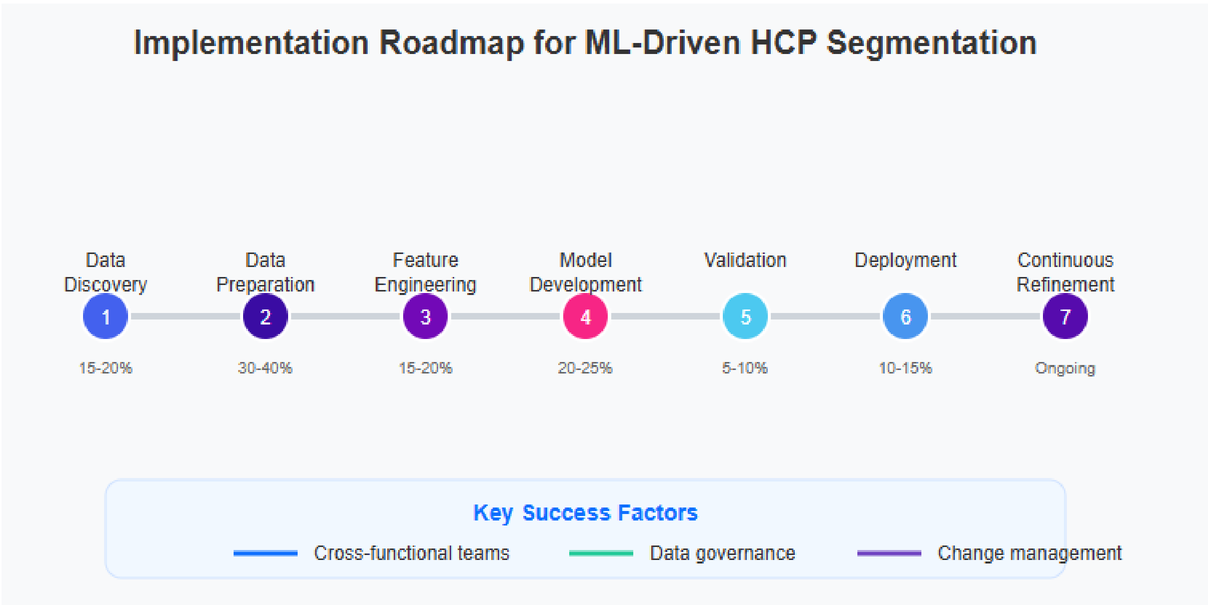

Fig 4. Implementation Roadmap for ML-Driven HCP Segmentation [9, 10].

Conclusion

Healthcare Professional segmentation has transformed from rudimentary classification systems into sophisticated, dynamic frameworks that capture the multidimensional nature of prescriber behavior and preferences. The progression through generations of segmentation approaches mirrors broader technological advancements, with each evolutionary stage addressing limitations of previous methods while introducing new capabilities. Contemporary AI/ML-driven segmentation leverages unprecedented data integration and advanced analytical techniques to uncover complex relationships between variables, identify subtle behavioral patterns, and predict future prescribing potential with remarkable precision. Successful implementation demands thoughtful attention to multiple dimensions, including data quality, algorithm selection, organizational adoption, and continuous improvement processes. The value delivered extends far beyond improved targeting efficiency, enabling truly personalized engagement strategies tailored to individual physician characteristics, preferences, and information needs. As healthcare markets continue evolving with new treatments, changing guidelines, and shifting practice patterns, dynamic segmentation capabilities provide essential competitive advantages through deeper customer understanding and more meaningful professional relationships. The future promises even greater personalization as segmentation technologies continue advancing, ultimately transforming promotional interactions into valued professional exchanges that enhance both physician satisfaction and patient outcomes.

Funding: This research received no external funding.
Conflicts of Interest: The authors declare no conflict of interest.
Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers.

References

[1] Aditya Kudumala et al., "Biopharma digital transformation: Gain an edge with leapfrog digital innovation," Deloitte Insights, 2021. [Online]. Available: https://www2.deloitte.com/us/en/insights/industry/life-sciences/biopharma-digital-transformation.html

[2] Anton Morrison, "The Evolving World of Pharma Marketing," PharmacyTimes, 2020. [Online]. Available: https://www.pharmacytimes.com/view/the-evolving-world-of-pharma-marketing

[3] DrugPatentWatch, "New Business Models for Pharmaceutical Marketing: Transforming Data into Market Domination," 2024. [Online]. Available: hhttps://www.drugpatentwatch.com/blog/new-business-models-for-pharmaceutical-marketing-transforming-data-into-market-domination/

[4] Lalitkumar K Vora et al., "Artificial Intelligence in Pharmaceutical Technology and Drug Delivery Design," National Library of Medicine, 2023. [Online]. Available: https://pmc.ncbi.nlm.nih.gov/articles/PMC10385763/

[5] Luigi M Preti et al., "Implementation of Machine Learning Applications in Health Care Organizations: Systematic Review of Empirical Studies," National Library of Medicine, 2024. [Online]. Available: https://pmc.ncbi.nlm.nih.gov/articles/PMC11629039/

[6] Mercury Williams, "AI-Driven Analytics: From Competitive Advantage to Essential Business Tool," PharmExec.com, 2025. [Online]. Available: https://www.pharmexec.com/view/ai-driven-analytics-from-competitive-advantage-to-essential-business-tool

[7] Mohd Javaid et al., "Significance of machine learning in healthcare: Features, pillars and applications," ScienceDirect, 2022. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S2666603022000069

[8] Qi An et al., "A Comprehensive Review on Machine Learning in Healthcare Industry: Classification, Restrictions, Opportunities and Challenges," MDPI, 2023. [Online]. Available: https://www.mdpi.com/1424-8220/23/9/4178 Jacob Vazquez, "6 Reasons Why Advanced Analytics Is Essential In Pharma," p360, 2019. [Online]. Available: https://www.p360.com/birdzai/6-reasons-why-advanced-analytics-is-essential-in-pharma/

[9]

Thamburaj Anthuvan et al., "Trends in pharmaceutical marketing and branding research: a bibliometric analysis (2009–2023)," Emerald Insight, 2024. [Online]. Available: https://www.emerald.com/insight/content/doi/10.1108/irjms-03-2024-0030/full/html