Human Factors in Accessibility, Assistive Technology and Digital Environments, Vol. 202, 2026, 173–178
https://doi.org/10.54941/ahfe1007308

Integrated Framework to Identify Attrition Mechanisms in Digital Healthcare-Based Physical Activity

Chanryong Kim

Department of Leisure & Sport Science, Dongeui University, Busan, 47340, South Korea

ABSTRACT

Digital healthcare services are frequently evaluated in terms of functionality, usability, and technical performance, yet their sustainable expansion cannot be fully explained by technical quality alone. This study proposes an integrated framework for identifying attrition mechanisms in digital healthcare by incorporating both service-level features and social-scientific dimensions of user interpretation. To explore this perspective, 23,665 YouTube comments and replies were collected from 16 videos related to Nike Training Club, Strava, and MyFitnessPal and analyzed using BERTopic. The results revealed clear cross-service differences. Strava generated comparatively coherent discourse centered on activity tracking, device integration, and data use, whereas Nike Training Club and MyFitnessPal showed broader thematic dispersion shaped by viral short-form reactions, general wellness talk, and comparisons with competing services. Rather than treating such inconsistency as a simple methodological weakness, this study interprets it as evidence that digital healthcare attrition is socially mediated through trust formation, contextual relevance, comparative evaluation, and platform culture. The findings suggest that sustainable expansion of digital healthcare requires more than technological advancement; it also depends on understanding how services are publicly interpreted, socially positioned, and integrated into everyday routines.

Keywords: Digital healthcare, Attrition mechanism, Sustainability, Sociotechnical perspective, YouTube comments, BERTopic

INTRODUCTION

Digital healthcare services have expanded rapidly across domains such as exercise support, self-tracking, nutrition management, and preventive health promotion, yet long-term retention remains unstable even when initial adoption is high (Eysenbach, 2005). In this context, attrition should not be treated merely as a secondary operational problem, because discontinuation itself is a central outcome that reveals how users evaluate and experience digital interventions over time (Eysenbach, 2005). Previous studies have shown that sustained use of digital health tools depends not only on technical functionality, but also on trust, personalization, accessibility, relevance to daily life, and perceived legitimacy (Adjekum et al., 2018; Vo et al., 2019). This suggests that sustainability in digital healthcare is fundamentally a sociotechnical issue rather than a purely technical one (Shaw and Donia, 2021).

Received May 22, 2026; Revised May 28, 2026; Accepted June 15, 2026; Available online July 20, 2026

Most prior studies on digital health engagement have relied on surveys, interviews, clinical settings, or implementation-focused analyses. Although these approaches are valuable, they are limited in capturing spontaneous public discourse surrounding real-world service use (Connolly et al., 2020). Social media data are increasingly important in health research because they provide unsolicited, real-time, and publicly visible accounts of user perceptions, frustrations, comparisons, and experiences that are often absent from formal healthcare records or controlled studies (Chen, 2021). Among these sources, YouTube comments are particularly meaningful because they are embedded in review culture, quitting narratives, influencer mediation, peer recommendation, and public comparison among services (Chen, 2021).

This study begins from the premise that digital healthcare services can expand sustainably to broader publics only when technological advancement is accompanied by a rigorous social-scientific understanding of how users construct meaning, trust, and commitment around those services (Adjekum et al., 2018; Shaw and Donia, 2021). Accordingly, this paper does not interpret mixed or unstable discourse simply as methodological failure. Instead, it views such discourse as evidence that attrition mechanisms emerge through interactions among service features, user expectations, platform culture, and comparative market narratives (Vo et al., 2019; Connolly et al., 2020). On this basis, the study aims to propose an integrated framework for identifying attrition mechanisms in digital healthcare through the analysis of YouTube comments related to Nike Training Club, Strava, and MyFitnessPal.

METHODS

Three widely recognized digital health and fitness services were selected as empirical cases: Nike Training Club, Strava, and MyFitnessPal. These services were chosen because they represent different but overlapping dimensions of digital health practice, including guided exercise, activity tracking, and nutrition management. In addition, they have substantial public visibility on social media platforms, making them suitable for discourse-based analysis.

To identify relevant public discourse, YouTube videos were searched using combinations of service names and evaluative phrases such as “review,” “why I quit,” and related comparison-oriented terms. Candidate videos were screened using visibility indicators including view count and comment volume in order to gather discourse with sufficient scale for exploratory topic modeling. A total of 16 videos were selected, including five associated with MyFitnessPal, five with Nike Training Club, five with Strava, and one cross-labeled video associated with both Nike Training Club and Strava. This strategy was effective for securing a large corpus, but it also introduced a structural limitation because highly visible videos are not always the most service-relevant videos. This limitation is analytically important because platform popularity can shape the discourse environment before topic modeling begins (Chen, 2021).

Comments were collected using the YouTube Data API through the commentThreads.list and comments.list methods, which enable retrieval of top-level comments and replies from public YouTube videos (Google

Developers, 2026a, b). The final dataset contained 23,665 text items, including 9,544 top-level comments and 14,121 replies. At the service level, 16,977 items were associated with Nike Training Club, 3,861 with Strava, 2,638 with MyFitnessPal, and 189 with the cross-labeled sample. The corpus therefore provided a large but uneven body of public discourse across the three services. Such imbalance is common in platform-based digital trace data and should be interpreted as part of the empirical environment rather than as a purely technical inconvenience (Chen, 2021).

Table 1: Corpus composition by service.

Service	Number of Videos	Number of Comments/Replies	Main Service Orientation
Nike Training Club	5	16,977	Guided exercise and training content
Strava	5	3,861	GPS-based tracking and social fitness logging
MyFitnessPal	5	2,638	Nutrition and calorie management
Nike Training Club | Strava	1	189	Cross-service comparison

The collected text was preprocessed by removing URLs, special characters, and other non-substantive noise. Duplicate-like fragments and extremely short strings were reduced where necessary in order to improve topic interpretability. This step was particularly important because short-text social media data are often characterized by sparse word co-occurrence, slang, heterogeneous expression, and non-meaningful tokens, all of which complicate topic modeling (Albalawi et al., 2020). After preprocessing, BERTopic was applied to identify semantically coherent clusters of discourse. BERTopic was selected because it combines transformer-based document embeddings, clustering, and class-based TF-IDF representations to generate interpretable topic structures in short-text settings (Grootendorst, 2022). In addition to the computational outputs, a service-level analytical dashboard was used to visually compare topic distributions across the three services and to assist the interpretation of thematic concentration and dispersion in the corpus.

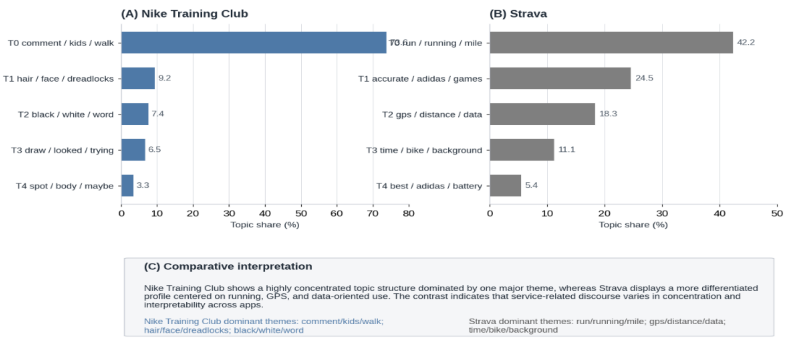

Figure 1: Simplified service-level topic profiles in YouTube comments.

The figure compares dominant topic distributions in Nike Training Club and Strava comments and highlights differences in thematic concentration and interpretive spread across services.

RESULTS

Figure 1 provides a simplified comparison of service-level topic profiles derived from the YouTube comment corpus. The visual summary shows that the two services generated markedly different discourse structures. Nike Training Club exhibited a highly concentrated topic distribution, with one dominant topic accounting for a substantial share of the visible discourse, whereas Strava displayed a more differentiated pattern distributed across several function-related topics. This contrast indicates that user discourse was not evenly organized across services and that topic concentration itself may serve as an important descriptive indicator of how publicly visible digital health services are interpreted in platform environments.

More specifically, the Nike Training Club profile was characterized by strong concentration around a small number of recurrent expressions, suggesting that the discourse was shaped less by detailed discussion of service functionality than by broader and sometimes peripheral reactions circulating around the video content. In contrast, the Strava profile showed a clearer distribution across themes associated with running, GPS, distance, and data-related use. This implies that Strava was discussed through more recognizable service-centered experiential frames, whereas Nike Training Club was more vulnerable to discursive spillover from video format, audience behavior, and platform-specific attention dynamics.

From an attrition perspective, this distinction is analytically meaningful. A more differentiated and function-oriented discourse structure may indicate that users engage with a service through relatively stable expectations related to practical use, while a highly concentrated or weakly service-specific structure may reflect unstable interpretive contexts in which service identity is diluted by non-functional or socially diffuse discussion. Accordingly, Figure 1 suggests that discontinuation cannot be understood solely in terms of technical shortcomings. Rather, the public discursive environment in which a service is encountered may also influence how users evaluate relevance, usefulness, and continuity of use (Adjekum et al., 2018; Vo et al., 2019).

DISCUSSION

The findings support the central claim of this study that digital healthcare attrition should be interpreted through an integrated sociotechnical framework rather than through a narrow technology-centered model. At first glance, the uneven topic structures shown in Figure 1 might be treated as methodological noise or as a limitation of comment-based analysis. However, a more productive interpretation is that such unevenness reflects the social conditions under which digital health services are publicly interpreted, compared, and normalized. In this sense, discourse concentration and

thematic dispersion are not merely descriptive artifacts; they are part of the empirical evidence needed to understand why some services are more stably integrated into everyday user narratives than others (Shaw and Donia, 2021).

This interpretation extends prior research in several ways. Previous studies have emphasized that trust, usability, accessibility, personalization, and contextual relevance shape adoption and continued engagement in digital health technologies. The present findings are consistent with that literature, but they further suggest that these factors are mediated through public discourse and platform visibility. Users do not encounter digital healthcare services only as technical systems; they also encounter them as socially circulated objects of comparison, recommendation, skepticism, and reputational judgment. Figure 1 is therefore valuable because it makes visible the extent to which service discourse may remain function-centered or, alternatively, become absorbed into more diffuse interpretive fields (Adjekum et al., 2018; Vo et al., 2019; Chen, 2021).

The figure also strengthens the methodological contribution of the study. Large-scale social media data are often criticized for inconsistency, brevity, and weak topical control, especially when short-text topic modeling is applied. Yet the present analysis suggests that these characteristics should not automatically be treated as defects to be removed. In studies of sustainability and attrition, discursive instability may itself be theoretically meaningful because it reveals how service identity becomes fragmented across media contexts. The contrast between the highly concentrated Nike Training Club pattern and the more differentiated Strava pattern demonstrates that platform-based discourse can illuminate differences in interpretability, not just differences in volume. This supports the use of public comment data as a complementary source for digital healthcare research when the objective is to understand socially mediated mechanisms of engagement and disengagement (Albalawi et al., 2020; Grootendorst, 2022).

At the same time, the findings should be interpreted with appropriate caution. The dataset was shaped by popularity-based video selection, and the resulting discourse was influenced not only by service use itself but also by creators, audiences, video format, and recommendation logics. Therefore, the present results should not be read as definitive evaluations of the intrinsic quality of the services. Rather, they should be understood as evidence of how digital healthcare services are situated within wider sociotechnical and communicative environments. Future research should combine stricter relevance screening, discourse-type classification, and qualitative validation in order to distinguish more clearly between function-specific dissatisfaction and broader socially mediated attrition processes (Connolly et al., 2020; Ndiaye et al., 2022).

CONCLUSION

This study proposed an integrated framework for identifying attrition mechanisms in digital healthcare by combining computational topic analysis with a social-scientific interpretation of public discourse. The findings show that the sustainability of digital healthcare services cannot be adequately

understood through technical performance alone. Instead, sustained use and disengagement are shaped through broader sociotechnical processes involving trust, contextual relevance, comparative evaluation, public narrative, and platform culture (Eysenbach, 2005; Adjekum et al., 2018; Shaw and Donia, 2021).

A key contribution of this study is that it reframes discursive inconsistency as an analytically productive finding rather than as a simple methodological weakness. The comparatively coherent discourse observed for Strava and the more diffuse patterns found for Nike Training Club and MyFitnessPal indicate that different digital healthcare services are socially interpreted in different ways. This suggests that public discourse data can reveal not only what users say about a service, but also how that service is positioned, compared, and normalized in everyday digital environments.

Future research should strengthen sample relevance screening, distinguish functional from socially mediated discourse categories before topic modeling, and combine computational methods with qualitative validation. Such an approach would provide a more robust basis for explaining digital healthcare attrition and for designing sustainability strategies that integrate both technical innovation and social understanding (Connolly et al., 2020; Ndiaye et al., 2022). Ultimately, if digital healthcare services are to expand sustainably to wider publics, social-scientific inquiry must be treated not as a supplementary perspective, but as a core analytical foundation alongside technology development itself.

REFERENCES

Adjekum, A., Blasimme, A. and Vayena, E. (2018), Determinants of Trust in Digital Health: A Scoping Review, Journal of Medical Internet Research, 20(12), e11254.

Albalawi, R., Yeap, T.H. and Benyoucef, M. (2020), Topic Modeling for Short Text: A Review, Frontiers in Artificial Intelligence, 3, 42. https://doi.org/10.3389/frai.2020.00042

Chen, J. (2021), Social Media Use for Health Purposes: Systematic Review, Journal of Medical Internet Research, 23(5), e17917. https://doi.org/10.2196/17917

Connolly, S.L., Miller, C.J. and Bauer, M.S. (2020), Factors Influencing Mental Health App Uptake: A Narrative Review of the Literature, BMC Psychiatry, 20, 443.

Eysenbach, G. (2005), The Law of Attrition (Or: Why Do eHealth Users Discontinue Usage?), Journal of Medical Internet Research, 7(1), e11.

Google Developers (2026a), YouTube Data API: commentThreads.list. Available at: https://developers.google.com/youtube/v3/docs/commentThreads/list

Google Developers (2026b), YouTube Data API: comments.list. Available at: https://developers.google.com/youtube/v3/docs/comments/list

Grootendorst, M. (2022), BERTopic: Neural Topic Modeling with a Class-Based TF-IDF Procedure, arXiv, 2203.05794. https://doi.org/10.48550/arXiv.2203.05794

Ndiaye, N.P., Sow, K., Bah, A., Niane, B. and Pulford, H. (2022), Barriers and Facilitators for the Sustainability of Digital Health Interventions in Low and Middle-Income Countries: A Systematic Review, Frontiers in Digital Health, 4, 1014375. https://doi.org/10.3389/fdgth.2022.1014375

Shaw, J. and Donia, J. (2021), A Sociotechnical Ethics of Digital Health, Frontiers in Digital Health, 3, 722659. https://doi.org/10.3389/fdgth.2021.722659

Vo, V., Auroy, L. and Sarradon-Eck, A. (2019), Patients' Perspectives on mHealth Apps: A Systematic Review, JMIR mHealth and uHealth, 7(7), e13817. https://doi.org/10.2196/13817