Health Technology and Social Research

https://healthsocres.com

# Virtual Support Groups for Perceived Belonging in Cancer Survivor Networks

**Derek Ng***

Department of Food and Health Sciences, Faculty of Science and Technology, Technological and Higher Education Institute of Hong Kong, Hong Kong, Hong Kong SAR, China

Corresponding author: derek.ng@thei.edu.hk

**Abstract**

The transition from active oncology treatment to long term survivorship is often characterized by a profound reduction in clinical contact, leaving many cancer survivors to navigate complex psychosocial challenges in relative isolation. Virtual support networks have emerged as vital platforms for peer connection, yet the dynamics of how these digital interactions translate into a sustained sense of perceived belonging remain poorly understood. This study employs a prospective cohort design tracking 450 cancer survivors over a six month period on a specialized online support platform. By integrating behavioral tracking, natural language processing of user generated text, and temporal network analysis, we construct a predictive modeling framework to estimate individual levels of perceived belonging. Our predictive models, particularly those leveraging gradient boosted decision trees, achieve high accuracy in forecasting belonging trajectories. The findings highlight the critical role of reciprocal interaction patterns, first person plural pronoun usage, and emotional stability in posts as early indicators of strong peer integration. This research provides a quantitative foundation for designing proactive digital oncology interventions that can automatically identify isolated survivors and tailor support mechanisms to enhance their psychological well-being.

Keywords: Virtual Support Networks; Cancer Survivorship; Social Belonging; Predictive Modeling; Perceived Belonging

## 1. Introduction

The transition from active clinical treatment to post treatment survivorship represents a critical juncture in the oncology care continuum. While the eradication or control of malignant disease is a major clinical milestone, the cessation of active therapies such as chemotherapy, radiation, or surgery often coincides with a dramatic decrease in contact with healthcare providers. This sudden reduction in professional medical oversight, colloquially referred to as the clinical cliff, frequently leaves survivors feeling abandoned, anxious, and socially isolated [1]. The physical sequelae of cancer treatment, including chronic fatigue, cognitive changes, and sexual dysfunction, persist long after the last therapeutic intervention, yet survivors are expected to resume normal societal and occupational roles without the continuous support systems present during active treatment [2]. Consequently, the psychosocial burden of survivorship can equal or even exceed the psychological distress experienced during the diagnostic and acute treatment phases.

In response to these unmet psychosocial needs, virtual support networks have proliferated as accessible, low barrier environments where survivors can exchange information, share emotional

burdens, and seek peer validation [3]. These online platforms remove geographical barriers, accommodate varying physical energy levels, and offer a degree of anonymity that can facilitate the disclosure of highly stigmatized or deeply personal concerns. However, the mere availability of virtual spaces does not automatically translate into meaningful peer integration or psychological support. Within any virtual platform, engagement patterns vary widely, ranging from passive reading of posts, often termed lurking, to highly active content creation and community facilitation. Understanding how these diverse interaction modalities foster a subjective sense of perceived belonging is essential for optimizing the design and therapeutic efficacy of digital oncology spaces.

Perceived belonging, defined as the subjective feeling of being valued, respected, and integrated into a peer collective, is a robust predictor of psychological resilience and overall quality of life in chronic illness populations [4]. In the context of virtual cancer communities, a strong sense of belonging has been linked to reduced symptoms of depression, enhanced self efficacy in managing long term side effects, and lower levels of cancer related distress. Conversely, survivors who participate in online groups but fail to establish a sense of belonging may experience heightened feelings of alienation and secondary rejection. Despite the acknowledged importance of this psychological construct, longitudinal research tracing the progression of perceived belonging in online settings remains scarce. Most existing literature relies on cross sectional surveys that capture static snapshots of user sentiments, failing to account for the dynamic, evolutionary nature of peer interactions over time.

To address these empirical gaps, this study presents a comprehensive prospective cohort design utilizing behavioral tracking, natural language processing, and advanced machine learning techniques to predict perceived belonging among cancer survivors. By systematically analyzing the linguistic properties of user contributions, temporal participation metrics, and network structural characteristics over a six month observational period, we develop a predictive framework capable of forecasting belonging outcomes. This predictive capability is highly valuable for clinical coordinators and platform developers, as it enables the early identification of users at risk of social isolation and facilitates the deployment of targeted, personalized peer interventions. Ultimately, this research bridges the gap between digital behavioral data and subjective psychological states, advancing our capability to cultivate supportive online spaces for cancer survivors.

## 2. Literature Review and Theoretical Framework

### 2.1 Psychosocial Challenges in Post-Treatment Survivorship

The completion of primary cancer therapy is often celebrated as a milestone, yet psychological research reveals that this transition frequently triggers a secondary crisis characterized by existential dread and somatic hypervigilance [5]. Survivors must reconstruct their identities in the wake of life altering diagnoses while simultaneously managing the long term toxicities of treatment. The lack of structured rehabilitative frameworks in many traditional oncology care pathways leaves a critical gap in psychosocial support. Peer support has been identified as a uniquely effective mechanism for addressing these challenges, as it provides experiential knowledge and emotional validation that family members and medical professionals cannot easily replicate. Virtual environments have adapted to fill this gap, offering a continuous source of peer connection that accommodates the erratic schedules and physical limitations of recovering survivors.

### 2.2 Theoretical Foundations of Virtual Social Belonging

This study grounds its exploration of online support dynamics in Social Identity Theory and the concept of computer mediated social support. According to Social Identity Theory, individuals derive a significant portion of their self concept and self esteem from the social categories and groups to

which they perceive themselves belonging [6]. For cancer survivors, the shared experience of surviving a life threatening illness serves as a powerful unifying category, fostering a deep, empathetic in group connection. In online domains, this social identification is mediated by textual and symbolic interactions. The asynchronous nature of online discussion boards allows for deliberate, reflective communication, which can enhance self disclosure and emotional expression [7]. Reciprocal self disclosure is a critical driver of intimacy and belonging; when a survivor shares a vulnerability and receives immediate, empathetic validation from peers, the psychological distance between the participants shrinks rapidly, reinforcing a shared group identity [8].

### 2.3 Predictive Modeling in Health Communities

The advent of digital health platforms has enabled researchers to move beyond traditional self report methods to observe naturalistic communication and behavioral patterns directly [9]. Prior research in computational social science has demonstrated that user behavioral traces, such as login frequency, message lengths, and network centrality, can serve as reliable proxies for underlying psychological states. Natural language processing techniques, including sentiment analysis and linguistic style matching, have been successfully applied to analyze online health forum posts to detect signs of clinical depression, anxiety, and cognitive decline. For example, the frequent use of first person singular pronouns has been consistently linked to self focus and depressive symptoms, whereas the shift toward first person plural pronouns often signals a growing sense of collective identity and group cohesion [10]. However, few studies have integrated these diverse data streams into a cohesive, longitudinal predictive framework designed specifically to forecast the development of perceived belonging over time in a medical survivor cohort.

## 3. Methodology and Research Design

### 3.1 Participant Recruitment and Cohort Characteristics

A prospective cohort study was conducted over a six month period, utilizing data from a specialized, moderated virtual support network dedicated to cancer survivors. Participants were recruited during their first week of registration on the platform. Eligible individuals were adult cancer survivors who had completed primary treatments within the past twenty four months and consented to have their de-identified platform interaction data and periodic survey responses analyzed for research purposes. To capture the baseline psychological state and demographic details, participants completed an initial intake survey upon enrollment.

A total of 450 participants met the inclusion criteria and remained active enough on the platform to be included in the final analysis. Active participation was defined as logging into the platform at least once every two weeks during the first month of the study. The demographic distribution of the cohort was highly representative of diverse cancer survivorship profiles, encompassing various cancer sites, stages of initial diagnosis, and age groups.

*Table 1: Cohort Demographic Characteristics and Initial Baseline Metrics*

| Parameter | Classification | Value or Distribution |
|---|---|---|
| Stage of Diagnosis | Stage I to II | 0.42 |
| Stage of Diagnosis | Stage III to IV | 0.58 |
| Primary Cancer Site | Breast | 0.38 |
| Primary Cancer Site | Colorectal | 0.22 |
| Primary Cancer Site | Hematological | 0.18 |
| Primary Cancer Site | Gynecological or Urological | 0.22 |
| Mean Age | Years | 48.50 |
| Baseline Perceived Belonging | Scale 1 to 5 | 2.10 |

Throughout the six month tracking period, participants completed brief, validated monthly self report assessments measuring their level of perceived belonging to the online community. The primary

outcome variable, Perceived Belonging, was measured using a five item scale adapted from the Sense of Belonging Instrument, yielding a continuous score ranging from 1 to 5, where higher scores indicate a stronger sense of connection and integration within the virtual community.

### 3.2 Feature Engineering and Behavioral Metrics

To construct a robust predictive model, we extracted a comprehensive set of features from the platform databases, categorizing them into three primary domains: linguistic characteristics, behavioral engagement metrics, and social network structure. These features were aggregated over sliding temporal windows of 30, 60, and 90 days to capture the evolving nature of user interactions.

Linguistic features were extracted from the body of text written by each participant. Utilizing natural language processing pipelines, we computed sentiment valence scores, emotional intensity indices, and specific lexical category frequencies. Lexical categories focused heavily on pronoun usage, specifically analyzing the ratio of first person singular pronouns to first person plural pronouns. We also calculated text readability indices and lexical diversity to capture communication style matching between users.

Behavioral engagement metrics focused on the quantitative patterns of platform use. These metrics included the total number of original posts created, replies written to other users, likes or reactions received, and the mean response latency when replying to threads. We also calculated the ratio of active contributions to passive reading actions to distinguish between highly active contributors, moderate participants, and passive lurkers.

Social network analysis was performed by representing the virtual support network as a directed graph, where nodes represent users and edges represent direct interactions, such as replies or mentions. For each participant, we calculated local network metrics including degree centrality, betweenness centrality, and the local clustering coefficient. These metrics provided a structural representation of the participant position within the social fabric of the platform, indicating whether they were central connectors, members of tight knit subgroups, or peripheral participants.

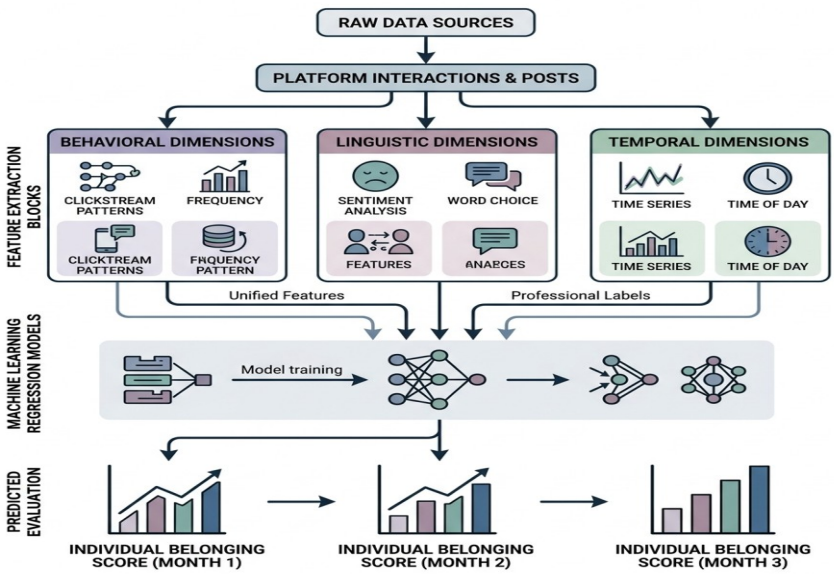


*Figure 1: Architectural Workflow of the Predictive Modeling Pipeline for Perceived Belonging*

The integrated workflow, illustrated above, outlines the sequential transformation of raw digital interactions into refined predictive targets. Behavioral metrics and written texts are continuous

inputs, which are processed through linguistic engines and social network analysis libraries before being combined into unified cohort feature arrays.

### 3.3 Predictive Modeling Framework

The predictive task was formulated as a supervised regression problem, aiming to predict a participant self reported perceived belonging score at month six based on the behavioral, linguistic, and structural network features accumulated during the preceding months. We compared four machine learning algorithms to evaluate their predictive efficacy: Elastic Net Regression, Support Vector Regression with a radial basis function kernel, Random Forest Regression, and Gradient Boosted Decision Trees.

Each model was trained and evaluated using a nested ten fold cross validation scheme to prevent data leakage and ensure generalizability. Within the inner loop of the cross validation, hyperparameter tuning was conducted via grid search. The performance of each model was evaluated using three key metrics: Root Mean Squared Error, Mean Absolute Error, and the Coefficient of Determination, also referred to as the R-squared value. Feature importance analysis was subsequently conducted on the top performing models to identify the most critical behavioral and linguistic markers of social integration.

## 4. Results and Analysis

### 4.1 Model Predictive Accuracy and Performance Metrics

The predictive capabilities of the four machine learning models varied significantly across the evaluated performance metrics. The tree based ensemble methods, specifically Gradient Boosted Decision Trees and Random Forest Regression, systematically outperformed the linear and kernel based models. This performance advantage suggests that the relationships between online interaction characteristics and the psychological state of perceived belonging are highly non-linear and characterized by complex feature interactions that linear frameworks fail to capture.

*Table 2: Performance Comparison of Machine Learning Models in Predicting Perceived Belonging*

| Predictive Model | Root Mean Squared Error | Mean Absolute Error | Coefficient of Determination |
|---|---|---|---|
| Elastic Net Regression | 0.68 | 0.54 | 0.41 |
| Support Vector Regression | 0.59 | 0.46 | 0.55 |
| Random Forest Regression | 0.48 | 0.38 | 0.70 |
| Gradient Boosted Decision Trees | 0.41 | 0.32 | 0.78 |

The Gradient Boosted Decision Trees model achieved the highest predictive accuracy, yielding an R-squared value of 0.78. This indicates that approximately 78 percent of the variance in the perceived belonging scores of cancer survivors at the six month mark can be explained by the behavioral, linguistic, and network features collected during their initial participation phases. The low Mean Absolute Error of 0.32 on a 5 point scale demonstrates the practical utility of this model for real world monitoring systems, as it can reliably identify users whose predicted belonging scores fall below critical thresholds.

### 4.2 Analysis of Primary Behavioral Predictors

To understand the underlying drivers of the top performing model, we analyzed the relative feature importance within the Gradient Boosted Decision Trees framework. The analysis revealed that structural network features and specific linguistic markers were the most critical predictors of long term perceived belonging.

The most influential feature was Reciprocal Reply Count, which measures the frequency with which a participant original post received a reply, followed by a subsequent response from the original poster. This bidirectional interaction pattern indicates active, sustained dialogue rather than isolated, uni-directional statements. Simple posting volume, in contrast, had a much lower predictive power, emphasizing that quality and reciprocity of engagement are far more critical for fostering a sense of community than sheer activity levels.

Linguistically, the Shift in Pronoun Ratio emerged as a highly significant predictor. Participants who exhibited a temporal transition from first person singular pronouns to first person plural pronouns over the first 60 days of participation were predicted to have significantly higher belonging scores at month six [11]. This linguistic shift reflects a psychological transition from self focused distress to collective identification with the survivor network. Furthermore, the average sentiment valence of received replies was strongly predictive of belonging, confirming that positive, supportive feedback from peers directly reinforces a user feelings of being valued and integrated [12].

Conversely, indicators of social isolation and cognitive distress were also successfully captured. A high frequency of cognitive mechanism words, which often indicate an ongoing struggle to make sense of trauma, coupled with a high local network clustering coefficient but low global centrality, predicted a stagnation or decline in perceived belonging [13]. This pattern suggests that some users may become trapped in small, highly insular subgroups that focus extensively on negative emotional venting, which can ultimately hinder broader social integration and long term psychological recovery.

## 5. Discussion and Implications

### 5.1 Implications for Digital Oncology Support Interventions

The empirical findings from this cohort study have significant practical implications for the design and administration of digital oncology support systems. By demonstrating that behavioral tracking and natural language processing can accurately predict perceived belonging, this research offers a pathway toward proactive, algorithmic moderator support. Currently, online community moderators must manually review forum threads to identify isolated or distressed users, a process that is highly resource intensive and prone to oversight. Integrating predictive models into platform administration dashboards can automate this process, flagging users who display behavioral or linguistic patterns associated with low future belonging [14].

Once identified, targeted interventions can be deployed to assist these at risk users. For example, platform administrators can pair a newly registered survivor who exhibits high isolation indicators with a trained peer mentor or veteran user whose communication style matches theirs [15]. Additionally, system algorithms can deliberately highlight posts from isolated users on main community feeds, increasing the probability that they receive the prompt, supportive feedback shown to be critical for fostering group identification [16]. Such targeted strategies can help mitigate the clinical cliff effect by ensuring that survivors receive continuous, personalized peer support during their transition to long term recovery [17].

Furthermore, understanding the importance of reciprocity highlights the need to design interface features that encourage active dialogue over passive consumption. Virtual platforms could incorporate nudges that prompt users to respond to unanswered posts or facilitate collaborative group activities that require mutual engagement [18]. By shifting the focus of digital health platforms from simple user engagement metrics, such as page views or daily active users, to structural and emotional indicators of belonging, designers can create online environments that are more aligned with therapeutic clinical outcomes [19].

### 5.2 Limitations and Future Directions

Despite the promising predictive accuracy achieved in this study, several limitations must be acknowledged. First, the data were collected from a single virtual support network, which may limit the generalizability of the findings to other platforms with different demographics, design features, or moderator guidelines [20]. Additionally, while we controlled for baseline psychosocial metrics, external life events, such as disease recurrence or changes in offline social support systems, were not continuously monitored and could have influenced participants self reported belonging scores [21].

Future research should focus on validating these predictive models across diverse digital oncology platforms and incorporating multi modal data sources, such as physiological stress markers or mobile application usage patterns [22]. Longitudinal studies tracking cohorts over multi year periods would also be beneficial to examine how the relationship between online interaction and perceived belonging changes as survivors transition into very long term survivorship [23]. By continuing to refine these predictive frameworks, researchers can provide clinical teams and digital platform architects with the tools necessary to foster resilient, supportive communities that enhance the well-being of cancer survivors globally.

## References

1. Duchscher, J. A process of becoming: The stages of new nursing graduate professional role transition. J. Contin. Educ. Nurs. 2008, 39, 441–450.

2. Duchscher, J. Transition shock: The initial stage of role adaption for newly graduated Registered Nurses. J. Adv. Nurs. 2009, 65, 1103–1113.

3. Scortegagna, H.D.M.; Cachioni, M.; Cecagno-Zanini, S.C.; Alonso, V.; Melo, R.C.D.; Neri, A.L. Interventions used by health professionals in older adults with low levels of health literacy: A scoping review. Geriatr. Gerontol. Aging 2023, 17, e0230004.

4. Mehralian, G.; Yusefi, A.R.; Davarani, E.R.; Ahmadidarrehsima, S.; Nikmanesh, P. Examining the relationship between health literacy and quality of life: Evidence from older people admitted to the hospital. BMC Geriatr. 2023, 23, 147.

5. Ryff, C.D. Happiness is everything, or is it? Explorations on the meaning of psychological well-being. J. Pers. Soc. Psychol. 1989, 57, 1069–1081.

6. Banco Santander. (2026). Available online: www.impulsa-empresa.es (accessed on 24 May 2026).

7. Kor, M.; Yitmen, I.; Alizadehsalehi, S. An Investigation for Integration of Deep Learning and Digital Twins towards Construction 4.0. Smart Sustain. Built Environ. 2023, 12, 461–487.

8. International Council of Nurses [ICN]. Digital Health Transformation and Nursing Practice. Position Statement. Available online: https://www.icn.ch/sites/default/files/2023-08/ICN%20Position%20Statement%20Digital%20Health%20FINAL%2030.06_EN.pdf (accessed on 1 November 2025).

9. Hu, X.; Olgun, G.; Assaad, R.H. An Intelligent BIM-Enabled Digital Twin Framework for Real-Time Structural Health Monitoring Using Wireless IoT Sensing, Digital Signal Processing, and Structural Analysis. Expert Syst. Appl. 2024, 252, 124204.

10. Yesavage, J.A.; Brink, T.L.; Rose, T.L.; Lum, O.; Huang, V.; Adey, M.; Leirer, V.O. Development and validation of a geriatric depression screening scale: A preliminary report. J. Psychiatr. Res. 1982, 17, 37–49.

11. Chandler, G.E. Succeeding in the first year of practice: Heed the wisdom of novice nurses. J. Nurses Staff. Dev. 2012, 28, 103–107.

12. World Health Organisation (WHO). Ageing and Health. 2022. Available online: https://www.who.int/news-room/fact-sheets/detail/ageing-and-health (accessed on 28 October 2025).

13. Feng, H.; Chen, Q.; de Soto, B.G. Application of Digital Twin Technologies in Construction: An Overview of Opportunities and Challenges. In Proceedings of the International Symposium on Automation and Robotics in Construction, Dubai, United Arab Emirates, 2–4 November 2021.

14. Canadian Association of Schools of Nursing [CASN]. Nursing Informatics Entry-to Practice Competencies for

Registered Nurses. 2025. Available online: https://www.casn.ca/wp-content/uploads/2014/12/Informatics-Entry-to-Practice-Competencies_2025_EN-V3.pdf (accessed on 1 November 2025).

15. Sardareh, M.; Matlabi, H.; Shafiee-Kandjani, A.R.; Bahreini, R.; Mohammaddokht, S.; Azami-Aghdash, S. Interventions for improving health literacy among older people: A systematic review. BMC Geriatr. 2024, 24, 911.

16. Wang, Y.; Tang, Z.; Qian, G.; Xu, W.; Huang, X.; Fang, H. A Prototype of a Lightweight Structural Health Monitoring System Based on Edge Computing. Sensors 2025, 25, 5612.

17. Wang, T.; Huo, J.; Wu, X. A Real-Time Detection Approach for Bridge Crack. Algorithms 2026, 19, 247.

18. Gao, Y.; Xiong, G.; Hu, Z.; Chai, C.; Li, H. Bridge Digital Twin for Practical Bridge Operation and Maintenance by Integrating GIS and BIM. Buildings 2024, 14, 3731.

19. Yang, Z.; Tang, C.; Zhang, T.; Zhang, Z.; Doan, D.T. Digital Twins in Construction: Architecture, Applications, Trends and Challenges. Buildings 2024, 14, 2616.

20. Creswell, J.W.; Poth, C.N. Qualitative Inquiry and Research Design, 4th ed.; SAGE: Los Angeles, CA, USA, 2018.

21. Mousavi, Y.; Gharineiat, Z.; Karimi, A.A.; McDougall, K.; Rossi, A.; Gonizzi Barsanti, S. Digital Twin Technology in Built Environment: A Review of Applications, Capabilities and Challenges. Smart Cities 2024, 7, 2594–2615.

22. Pordata. População: Dados Sobre População Residente, Envelhecimento, Densidade Populacional, Nascimentos, Mortes, Casamentos, Divórcios. 2024. Available online: https://www.pordata.pt/pt/estatisticas/populacao?f[0]=theme%3A287 (accessed on 17 November 2025).

23. National Statistics Institute (Instituto Nacional de Estatística de Portugal—INE). 2023. Available online: https://www.ine.pt/xportal/xmain?xpid=INE&xpgid=ine_destaques&DESTAQUESdest_boui=707460498&DESTAQUESmodo=2&xlang=pt (accessed on 17 November 2025).