2023

Discovering mHealth Users’ Privacy and Security Concerns through Social Media Mining

Mitchell Damion

Omar El-Gayar

Discovering mHealth Users’ Privacy and Security Concerns through Social Media Mining

Damion R. Mitchell
Dakota State University
damion.mitchell@trojans.dsu.edu

Omar F. El-Gayar
Dakota State University
omar.el-gayar@dsu.edu

Abstract

The purpose of this study is to explore the various privacy and security concerns conveyed by social media users in relation to the use of mHealth wearable technologies, using Grounded Theory and Text Mining methodologies. The results of the emerging theory explain that the concerns of users can be categorized as relating to data management, data surveillance, data invasion, technical safety, or legal & policy issues. The results show that over time, mHealth users are still concerned about areas such as security breaches, real-time data invasion, surveillance, and how companies use the data collected from these devices. Further, the results from the emotion and sentiment analyses revealed that users generally exhibited anger and fear, and sentiments that were negatively expressed. Theoretically, the results also support the literature on user acceptance of mHealth wearables as influenced by the distrust of companies and their utilization of personally harvested data.

Keywords: mhealth, privacy, security, wearables

1. Introduction

The World Health Organization (WHO) defines mHealth as “medical and public health practice supported by mobile devices, such as mobile phones, patient monitoring devices, personal digital assistants, and other wireless devices” (World Health Organization, 2011). With the dawn of miniaturized devices, low-power body-area wireless networks, and ubiquitous smartphones, the growing field of mHealth wearable technologies has attracted remarkable commercial activity, consumer interest, and acceptance by major healthcare providers (Kotz et al., 2016). As such, consumers have accepted wearables into their everyday lives, considering them essential to their daily routines, wellness, and health. The global market value of wearables technology in 2015 was over $24 billion (Perez & Zeadally, 2018). By 2026, the market value of wearables is projected to increase by over 250% to $150 billion.

Although mHealth wearable technologies may indeed improve quality of healthcare and quality of life, they also engender security and privacy issues. Wearables, by constantly collecting, transmitting, and storing data, handle information that are cogitated as personal, private, sensitive, or confidential which can lead to severe privacy implications, threats, and risks. Therefore, privacy and security of personal data while using mHealth wearable devices continue to be of great concern. Owing to the high data sensitivity and the mobility of the devices, privacy concerns have proved to be more important in the context of health wearables than other technological devices (Miltgen et al., 2013). Accordingly, understanding the users’ expressed privacy and security concerns with mHealth wearables, can help to get further insights on how to minimize or alleviate these concerns.

As such, social media affords new prospects for analyzing numerous facets of, and patterns in communication. Studies (Correia et al. 2020; Domalewska 2021) have looked at various health related challenges through the lens of social media mining. Consequently, social media mining includes an array of analyses, from simple counting of the likes, retweets, and users’ demographics to more sophisticated measuring of quantifiable information such as sentiment, popularity, or reach (Domalewska, 2021). To the best of our knowledge, no considerable work has been conducted to explore the privacy and security concerns of users, associated with mHealth wearable technologies, using social media mining.

The purpose of this study is to investigate the various privacy and security concerns expressed by social media users in relation to the use of mHealth wearable technologies, using Grounded Theory and Social Media mining techniques. Further, this research seeks to fill the gap by examining mHealth security and privacy related topics, and compare and contrast the findings with extant literature, and propose an emergent theoretical framework that explains these user expressed concerns.

The remainder of the paper is organized as follows: the next section provides a brief literature review, followed by a detailed description of the research design and methodology including data collection and social media analytics. The results section summarizes the findings from an analysis point of view that aims to evaluate the expressed concerns, the general sentiments towards mHealth wearables privacy and security related issues, and how these issues evolved over time. The paper concludes with a discussion of the findings, and a presentation of limitations and future research.

2. Literature review

The emerging field of wearable technologies, has become popular in several application domains, including healthcare, entertainment, and others (V. Motti & Caine, 2015). These devices provide users with the ability to measure and monitor their lifestyles in a methodical manner (Kari et al., 2017). However, extant literature frequently cited privacy and security problems as leading barriers to the insistent acceptance of mhealth wearables (Kang et al., 2013; L. Lee et al., 2015). The notion of privacy is not new, and it has commonly been defined as a person's ability to control the terms by which their personal information is captured and used (Westin, 1968). It was suggested by Bunnig and Cap (2009) that privacy encompasses protecting personal information from being misused by malicious entities and permitting certain authorized entities to access that personal information by making it visible to them. Research has revealed that privacy concerns and sentiments of security threats can impede the usage of e-commerce systems (Eastlick et al. 2006), online health information systems (Bansal et al., 2010) and especially, location-based services (LBS) of mHealth technologies (Zhou, 2012).

Privacy and security issues inhibit the acceptance and distribution of technology in the IT domain (Cho et al., 2009; C. Lee et al., 2011). Owing to the high data sensitivity and the flexibility of the devices, privacy concerns have proved to be more significant in the perspective of health wearables than other technological devices (Miltgen et al., 2013). It was postulated that privacy-related threats can be classified as identity threats, where patients may lose their identity credentials, thus allowing access to their personal health information (PHI); and access threats, where patients have ultimate control on the collection, use, and disclosure of PHI, but if they fail to express their consent broader-than-intended access may be granted (Plachkinova et al., 2015).

Security refers to the protections, methods, and tools used to safeguard against the inappropriate access or release of information. As such it is one of the key factors in guarding the users from any type of uncertainties and risks. mHealth security covers the triads of confidentiality (guaranteeing that the collected data is available only to the authorized entities), integrity (confirming the correctness and trueness of the data being transmitted), and availability (survivability notwithstanding security attacks). mHealth wearable technologies must secure the users' trust and provide surety that they will be safe, so that there will be no hindrance to the adoption of these devices (AlHogail, 2018).

With the proliferation of the use of social media sites, emerging research aims to utilize this media as a rich data source to gain insight into 'real-life' use experiences. Studies such as (Al-Ramahi et al., 2016; V. Motti & Caine, 2015) demonstrate the use of text-mining of online user reviews, which affirms the potential for analyzing posts made on social media platforms as a means to better understand the privacy and security concerns of mHealth wearables as expressed by users.

3. Research design and methodology

The research method used in this study was adapted from Al-Ramahi et al. (2016) and saw the combination of grounded theory and text-mining, which are considered to epistemologically compatible. Therefore, the automation of the extraction and evaluation of social media posts was performed through text mining within the grounded theory framework (Charmaz, 2006). Firstly, data collection was performed based on a specific time and keywords of interest. Further, the compiled tweets were pre-processed and open-coded using text mining. Secondly, the Latent Dirichlet Allocation (LDA) algorithm was implemented for topic modeling, which allowed for the automatic extraction of concepts from the large corpus of text data. In our research we evaluated different LDA models, by varying the number of topics (k) up to a maximum of 50, and evaluated them against the held-out test items. Therefore, the perplexity of a held-out test set was computed to evaluate the generated models. The dataset was divided in which 80% (n=20420) of the data was used to train the models, while the remaining 20% (n=5105) was used for the held-out test set.

Manual coding through ATLAS.ti on a representative sample (n=1,000) was done to confirm the findings of the topic modeling. Thirdly, axial coding and selective coding were conducted to unearth appropriate higher-level categories. Finally, Brandwatch (BW), a social media mining platform was used to examine the data for aspects such as sentiment and trend analysis.

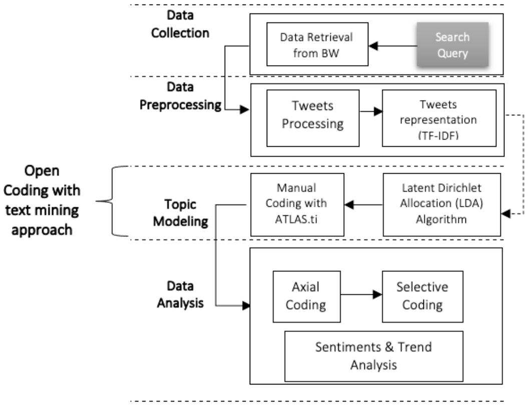

Data Collection
Data Retrieval from BW
Search Query
Data Preprocessing
Tweets Processing
Tweets representation (TF-IDF)
Open Coding with text mining approach
Topic Modeling
Manual Coding with ATLAS.ti
Latent Dirichlet Allocation (LDA) Algorithm
Data Analysis
Axial Coding
Selective Coding
Sentiments & Trend Analysis

Figure 1. Research Methodology

3.1. Data collection and preprocessing

The Twitter microblogging was our target social media platform for data collection. We utilized Brandwatch which provides access to the "Twitter firehose" which makes available every public tweet ever posted on Twitter in any language and from any geographic location that meets the search criteria. Several keywords such as intrusion, theft, surveillance, expose, tracking, or data control were identified by scrutinizing the literature (V. G. Motti & Caine, 2014; Solove, 2006) as well as using online synonym generators such as Merriam-Webster and synonyms.com. The collected tweets were pre-processed by removing stop words, retweets, addresses, and certain words that are not context appropriate. We performed lemmatization and represented each document using the well-known Term Frequency Inverse Document Frequency (TF-IDF) weighting scheme (Haddi et al., 2013). The extracted English tweets were for the period June 1, 2010, to Dec 31, 2021. The following query was used:

```
((mhealth OR wearable* OR smartwatch* OR "mobile
health" OR "m-health" OR Fitbit OR "fitness tracker*")

AND (intru* OR protect* OR (access AND control) OR (
data AND control) OR expose OR transparen* OR consent
OR confidential* OR leak* OR anonym* OR malicious OR
unauthoriz* OR harvest OR hack OR theft OR
privacy* OR breach OR invad* OR captur* OR invasion OR
(location AND tracking) OR breach OR disclose OR manip
ulate OR (third AND party) OR (3rd AND party) OR sensi
tive* OR access OR (unauthorized AND access) OR (priva
cy AND concern) OR vulnerabl* OR violate* OR (privacy
AND policy) OR (data AND collection) OR surveillance
OR (data AND capture) OR (personal AND data))
AND-(RT OR http OR https OR author:(fitbit* OR apple*)
```

3.2. Open coding using text mining

In this phase of the analysis the labeling of the phenomena discovered in the posts was done to create concepts (Charmaz, 2006). According to Myers (2009) the open coding phase form the basic foundation for Grounded Theory construction. Normally, manual content analysis is performed at this stage, however, we used text-mining, which according to Feldman and Sanger (2006) is the process of getting valid information from collection of documents through the discovery and identification of interesting patterns. The text-mining process like that of grounded theory necessitates objectivity which will allow for categories to emerge from the data (Yu et al., 2011).

Blei (2012) described topic modeling as statistical algorithms used to uncover hidden thematic structure or topics from large unstructured corpus of documents. In this study, the LDA-based topic Modeling (Blei, 2003) was used. According to Chiru et al. (2014), LDA has the greatest performance among various topic modeling algorithms when dealing with large-scale documents and deciphering identified latent topics. Further, the model produces automatic summaries of topics in terms of a discrete probability distribution over words for each topic, additionally it deduces per-document discrete distributions over topics.

Topics are typically manually labeled to ensure high labeling quality, particularly when such classification requires domain knowledge (Chang et al., 2009). To guarantee that the labeling was not biased, two independent researchers reviewed and labeled the 10 topics generated by the LDA model. The inter-rater agreement (kappa) between the two authors was substantial (kappa = .80) which, according to Landis & Koch (1977), indicates almost perfect agreement between two authors.

3.3. Axial & selective coding

We performed axial coding which involved the development of main categories and their sub-categories (Charmaz, 2006). In grounded theory, selective coding refers to the incorporation of the categories that have been discovered during the axial coding to form the theoretical framework (Pandit, 1996). Selective-coding is the process of integrating categories to build a theory and to refine the theory (Glaser & Strauss, 1967). Its task is to relate categories found in axial-coding to a core category which represents the main theme of research (Pandit, 1996).

3.4. Sentiments and trend analysis

Liu (2010) opined that word-based data can be generally grouped into facts and opinions; where facts are actual expressions, while opinions are biased expressions that communicate people's sentiments, appraisals, or feelings. Automatically attributing positive, negative, or neutral sentiments to segments of text that articulate opinions is classified as sentiment analysis (Jeong et al., 2019). Furthermore, emotion analysis provides an additional layer of related analysis by the use of "Ekman 6" (Anger, Fear, Disgust, Joy, Surprise, and Sadness) basic human emotions (Ekman, 1993).

A supervised algorithm named BrightView, an updated version of the ReadMe algorithm developed by (Hopkins & King, 2010) was used. The algorithm utilizes aggregate analysis which is principally suited when the amount of tweets that fit into specific categories over time is to be depicted. A training set of documents was manually coded into a set of predefined groups. Compared to traditional classification methods that concentrate on maximizing the percent of documents correctly classified into a given set of categories, the ReadMe algorithm underscores the broad categorization about the whole sets of documents (Hopkins & King, 2010).

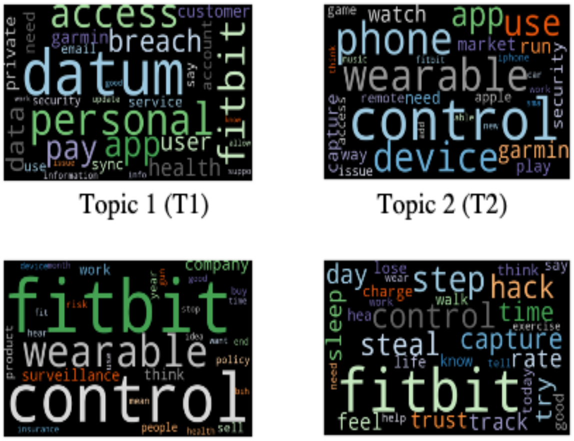
We assigned approximately 20 tweets into each category (derived from the topic modeling stage), after which the Brightview algorithm was run on both past and future posts. The tweets were examined based on the assigned categories, and further training was conducted where necessary.

4. Results

A total of 25,525 unique English tweets which were returned for the period June 1, 2010, to December 31, 2021. Of this amount 37.1% (n=9444) were considered as noisy or irrelevant data, for example "*Forgot to take my Apple Watch off tho and I really look like a spy kid*"

4.1. Open Coding Results

Table 1 illustrates the result of a 10-topic LDA model produced during the open coding phase, where each topic was represented by the top-15 weighted words in its vocabulary distribution. A descriptive word or phrase was then ascribed to each topic to signify the main privacy and security concerns related to mhealth wearable technologies. For example the concept of Surveillance was evidenced in the following tweet: *"I have opted out of sharing my activity data, but perhaps there's a shared tracking cookie that could be leaking my location"*.



Topic 1 (T1) Topic 2 (T2)

Topic 10 (T10) Topic 4 (T4)

Figure 2. Sample Word Clouds

Figure 2 shows four sample generated word clouds, which highlight the most prominent words, as depicted by the words with the highest frequency.

4.2. Axial Coding Results

After completing the open coding step in which ten (10) privacy and security concerns were obtained, these were further abstracted into eight (8) concepts, table 3. For example, *"Company Usage of Data"* and *"Access to Data"* were generalized and placed as *"Accessibility of Data"*; *"Control over wearables" and "Control over Patient Apps" to "Dynamic User Control"*.

Table 3. Axial Coding Results

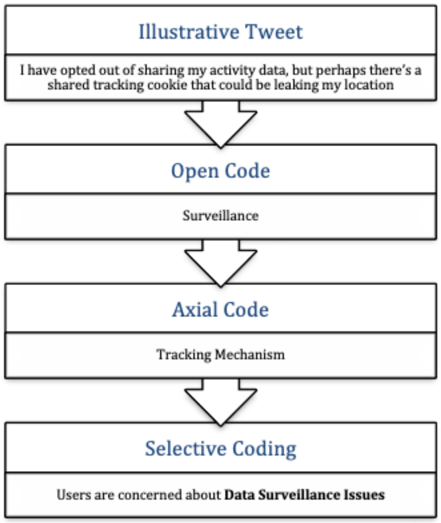
Open Code	Axial Code
Misuse of Data	Data Harvesting
Control over wearables	Dynamic User Control
Control over Patient app	
Real Time Data	Privacy Invasion
Capture of Personal Data	Information Gathering Consent
Access to Data	Data Accessibility
Company Use of Data	
Data Protection	Data & Privacy Protection
Security Breaches	Security Vulnerabilities
Surveillance	Tracking Mechanisms

Table 1. Open coding results with descriptive word or phrase for user concerns

Misuse of Data (T1)	**Control over wearables (T2)**	**Control over Patient Apps (T3)**	**Company Use of Data (T4)**	**Real Time Data (T5)**	**Capture of Personal Data (T6)**	**Data Access (T7)**	**Data Protection (T8)**	**Security breach (T9)**	**Surveillance (T10)**
datum	control	control	fitbit	time	datum	access	protection	wearable	fitbit
personal	wearable	patient	datum	private	capture	new	Datum	security	control
health	gesture	app	personal	make	people	change	Privacy	breach	wearable
information	device	help	hack	datum	need	body	Data	privacy	surveillance
use	technology	mobile	health	real	personal	level	Issue	tech	day
device	come	way	steal	enable	good	medical	consumer	wear	apple
collect	let	phone	company	tool	think	year	Say	use	business
secure	use	allow	information	invasion	activity	increase	healthcare	bring	diabete
record	thing	use	buy	store	want	work	Security	human	disease
harvest	project	health	sell	use	track	info	Concern	connect	product
know	glass	love	google	thank	acquire	company	Share	device	use
share	home	improve	wellness	protect	include	tech	challenge	camera	monitor
wearable	remote	device	acquisition	hand	really	phone	Need	create	market
patient	smawatch	monitoring	trust	data	point	monitor	Service	data	long
protect	want	care	say	people	research	personal	High	capture	trust

4.3. Selective Coding Results

The selective coding section resulted in a final emergent theory on the privacy and security concerns of healthcare wearable technologies as expressed by social media users. At this stage the eight (8) concepts from the axial coding phase were abstracted to form five (5) higher level categories to include issues in: *Data Management, Technical Safety, Data Invasion, Legal & Policy, and Data Surveillance.* Figure 3 shows an illustration of how the theoretical construct emerged from an example tweet through the open, axial, and selective coding phases. The final emergent theoretical diagram is shown in Figure 6.



I have opted out of sharing my activity data, but perhaps there's a shared tracking cookie that could be leaking my location

Users are concerned about **Data Surveillance Issues**

Figure 3. Grounded Theory Phases Illustration

4.5. Emotion and Sentiment Analyses

The expressed emotions of mHealth users relating to their privacy and security are captured in figure 4. The chart shows that 53% of the posts were depicting anger, with another 14% demonstrating fear. Further, 6% of the posts conveyed sadness, with 4%, expressing share disgust. For example, one tweet highlighted the concern of how their personal data was being handled by companies, *"Fitbit is just another acquisition that will give Google access to hugely valuable sensitive data about us."* Some were fearful about what it meant to have all their location and fitness data being harvested through data surveillance. Interestingly, 22% of the posts were showing expressed joy by the users, which indicates that some of the users were pleased with the benefits gained from these devices.

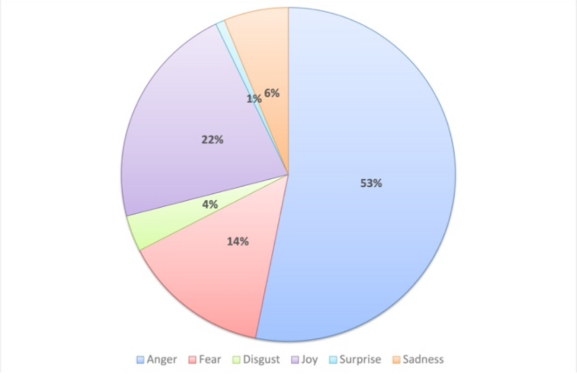


Figure 4. Emotion Analysis

For example: "*I definitely suggest an Apple Watch. I love that it captures all of my calories lost for the day.*" Based on the sentiments analysis performed through the Brandwatch tool, 71% of the posts depicted a negative sentiment, whereas 29% were positive, which coincides with the emotion analysis, where some users are satisfied with the benefits to be derived from these devices, even with expressed privacy and security concerns. This behaviour is in line with the Privacy Calculus theory, where individuals always rationally weigh the potential benefits and potential risks of data disclosure decisions (Culnan & Armstrong, 1999).

4.4. Dominant mHealth Privacy & Security Themes

An assessment of the dominant mHealth privacy and security themes discussed over the period under examination, shows that 22% of the posts were related to surveillance issues being expressed by the users, while 20% were concerned about how much control they truly have over these wearable devices. Further, the concern about the invasion of real time data accounted for 11% of the posts, while security breach and company use of the data were represented by 10% each. Other issues included the misuse of data, 9%, lack of data protection, 8%, personal data capture, 6%, and data access, 4%. The results also show that users seemingly were not utilizing nor having major issues with the control over patients apps, as less than 1% of the posts were related to that concern. The findings also show that as the popularity of mHealth devices grew between 2013 and 2017, more users were expressing concerns about lack of data protection and also data surveillance challenges.

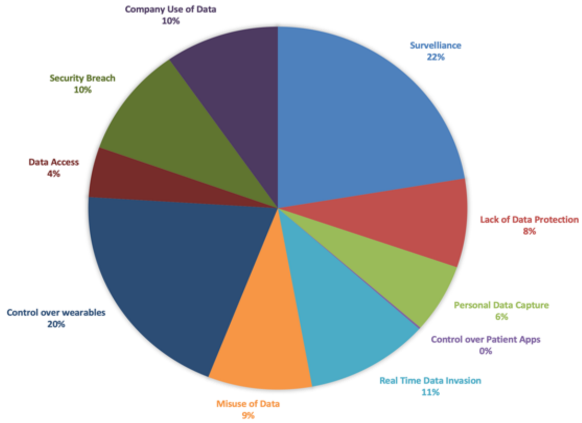


Figure 5. Dominant mHealth Privacy and Security Issues

There were many discussions around 2015, which could be due to the increased download of fitness applications (Krebs & Duncan, 2015). Furthermore, another detected peak in 2020 in the surveillance domain, could be attributed to the acquisition of Fitbit by Google at the end of 2019.

5. Discussion

The emergent theoretical model shown in Figure 6, highlights five (5) abstracted categories, namely, Data Management, Technical Safety, Data Surveillance, Data Invasion, and Legal & Policy issues. Data Management issues encapsulates data harvesting, data accessibility, and information gathering consent. A study conducted by de Arriba-Pérez et al. (2016) showed that users of healthcare wearables are worried that data that is harvested via the sensors available in these devices may be misused by different individuals. Further, data accessibility continues to be of great concern for users which is supported by Arora et al. (2014) and Kotz et al. (2009). In addition, Abdolkhani et al. (2020) shared from their research, that users lament the lack of transparency on who owns and has access to the data, also, the lack of information gathering consent for continuous data collection and use. This concern is intensified since sensors in wearable devices allow the collection of a wide array of user data ubiquitously and unobtrusively on a continuum basis, and in most cases, without the explicit consent of the user.

The Technical Safety concern refers to the process of data collection, and is due to the lack of control over devices and data permissions, where users cannot choose to shut down a sensor individually or cancel data collection, making it difficult to authorize the viewing and use of data (Jiang & Shi, 2021). Therefore, users are concerned that they do not have dynamic control over wearables and patient apps which all have the ability to sense, collect, and store data which are often personal, confidential or sensitive; that is the user interaction with a wearable. On the other hand, users should have influence that will readily allow them to apply fine-grained control about what is collected and shared (V. Motti & Caine, 2015).

The findings also demonstrated that surveillance through different tracking mechanisms results in Data Invasion issues where information is collected most times without the knowledge of the users (Datta et al., 2018). Surveillance can be seen as "any collection and processing of personal data, whether identifiable or not, for the purposes of influencing or managing those whose data have been garnered" (Lyon, 2001, p. 2). This is confirmed by Young (2018) where the top five wearable vendors were analysed to understand how they amass digital data on their users through surveillance assemblage, from which many concerns were discovered.

Our findings revealed that real-time data is affected by privacy invasion and security breach for healthcare wearables are caused by different security vulnerabilities, which all present data invasion concerns for users. This was confirmed by a study conducted by Ching & Singh (2016) outlining security and privacy vulnerabilities on wearable devices. It was shown that there exists some security weakness that makes wearable devices vulnerable to attack. One of the critical attacks on wearable technology is authentication issues.

Users are always concerned about their data and privacy protection, but it was apparent that there are legal and policy issues. Legal & Policy issues refer to a lack of policies and regulations on data security and privacy protection for wearable devices, especially healthcare wearables devices, once the manufacturers sell user data privately (Jiang & Shi, 2021). This concern was amplified by Lazzarotti (2015), in which it was suggested that HIPAA does not apply directly to wearable devices, but may be applied to wearables and their collection of health-related data only when related to a group health plan.

Myers (2009) recommended two vital conditions that must be met during the evaluation of grounded theory research: 1) rigor and validity; 2) generalization. In this study, the rigor and validity of the data analysis was realized through the use of a text-mining approach where concepts were extrapolated from a large corpus of tweets. This was also supported by the systematic approach in conducting the different Grounded Theory phases. Additionally, several tweets were identified which supported the privacy and security concerns deduced. Importantly, compared to manual coding with limited occurrences of the data, a higher degree of consistency and reliability can be realized through the mining of knowledge from a sizable volume of data (Yu et al. , 2011). In terms of generalizability, we developed an emergent theoretical framework by extracting knowledge from the large corpus of text data. The framework demonstrated five (5) overarching privacy and security concerns: data management, technical safety, data invasion, data surveillance, and legal and policy issues.

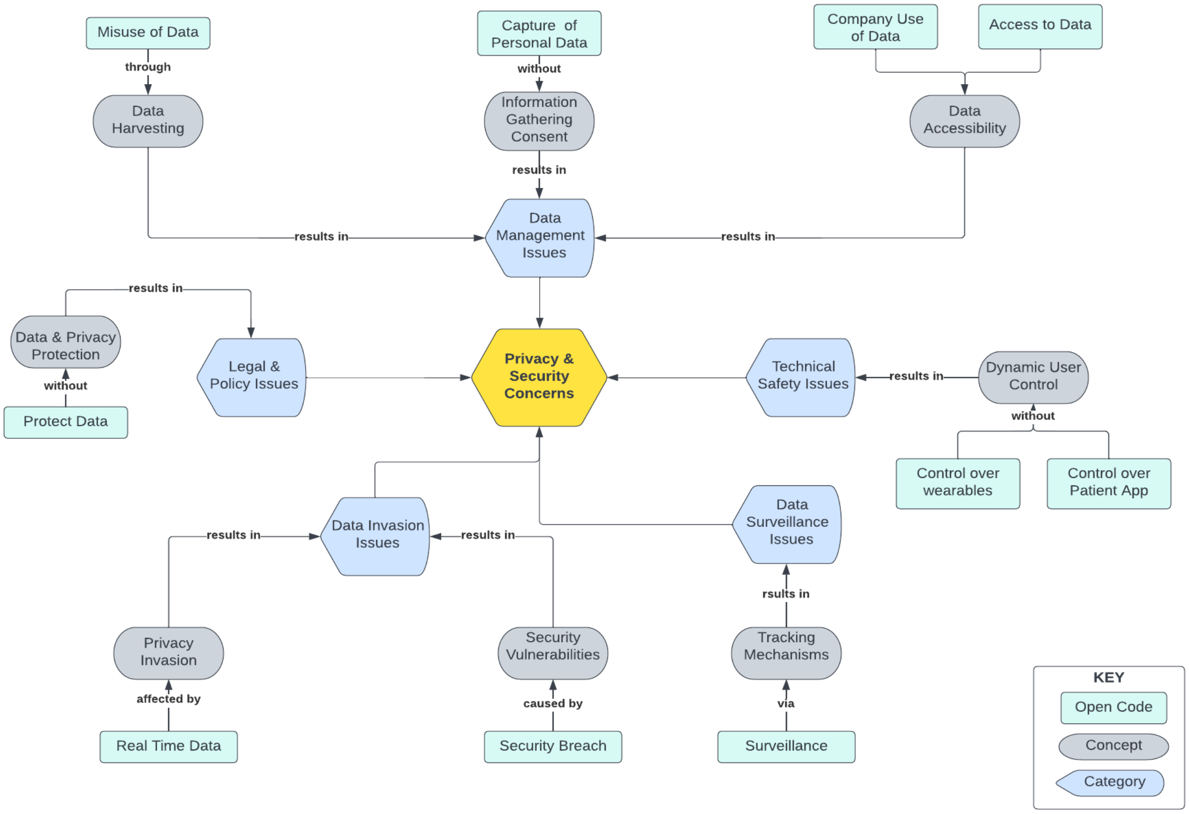


Figure 6. Emerging Theoretical Model

6. Conclusion

The purpose of this study is to explore the various privacy and security concerns conveyed by social media users in relation to the use of mHealth wearable technologies, using Grounded Theory and Text Mining methodologies. To confirm the concepts from the open-coding phase a representative sample of tweets were analyzed using ATLAS.ti. The results of the emerging theory explains that the concerns of users can be categorized as relating to data management, data surveillance, data invasion, technical safety, or legal & policy issues. The sentiment and emotion analysis results demonstrated that the collected tweets were largely associated with anger and fear, and an overall negative experience.

Methodologically, the capability of text mining within the grounded theory context was utilized. We used the LDA algorithm for topic modeling, to automatically extract concepts from large amounts of text data, instead of manually analyzing and coding the tweets, which is time-consuming and subjective.

Theoretically, the findings provide evidence through the emergent theoretical framework, that users of mHealth wearables are concerned about data management, technical safety, data invasion, data surveillance, and legal and policy issues. Further, users' distrust of companies and how their data is utilized is likely to influence the acceptance of these technologies, therefore the results also contribute to the literature of users' acceptance of health consumer technology.

Practically, it can help policy makers with developing comprehensive guidelines to govern data collection, dissemination, and processing on these devices. Furthermore, better indicators of the acceptance and use of mHealth devices can be established through available data on the web which provides opportunities for tracking and analyzing actual users' opinions about a phenomenon (Motiwalla et al., 2019).

A limitation of the study is the potential noise that accompanies social media posts and the impact of pulling data from only one social media platform, which could impact the findings. Therefore, future research will further explore the generalizability of these findings. In addition, understanding the concerns from users on other popular social media platforms like Reddit and Facebook may be beneficial.

Future research may also investigate other factors relating to privacy and security concerns in healthcare wearables usage and adoption such as the role of age, gender, and culture. Further studies will examine the relationships that exist between expressed sentiments and each privacy and security concerns. Further studies can investigate the generalizability of the developed emergent theory.

7. References

Abdolkhani, R., Gray, K., Borda, A., & DeSouza, R. (2020). Quality Assurance of Health Wearables Data: Participatory Workshop on Barriers, Solutions, and Expectations. *JMIR MHealth and UHealth*, *8*(1). https://doi.org/10.2196/15329

Al-Ramahi, M., El-Gayar, O., & Liu, J. (2016). Discovering Design Principles for Persuasive Systems: A Grounded Theory and Text Mining Approach. *2016 49th Hawaii International Conference on System Sciences (HICSS)*, 3074–3083. https://doi.org/10.1109/HICSS.2016.387

Arora, S., Yttri, J., & Nilsen, W. (2014). Privacy and Security in Mobile Health (mHealth) Research. *Alcohol Research : Current Reviews*, *36*(1), 143–151.

Bansal, G., Zahedi, F. "Mariam," & Gefen, D. (2010). The impact of personal dispositions on information sensitivity, privacy concern and trust in disclosing health information online. *Decision Support Systems*, *49*(2), 138–150. https://doi.org/10.1016/j.dss.2010.01.010

Blei, D. M. (2003). *Latent Dirichlet Allocation*. 30.

Blei, D. M. (2012). Probabilistic topic models. *Communications of the ACM*, *55*(4), 77–84. https://doi.org/10.1145/2133806.2133826

Bunnig, C., & Cap, C. H. (2009). Ad Hoc Privacy Management in Ubiquitous Computing Environments. *2009 Second International Conference on Advances in Human-Oriented and Personalized Mechanisms, Technologies, and Services*, 85–90. https://doi.org/10.1109/CENTRIC.2009.20

Chang, J., Boyd-Graber, J., Gerrish, S., Wang, C., & Blei, D. M. (2009). *Reading Tea Leaves: How Humans Interpret Topic Models*. 10.

Charmaz, K. (2006). *Constructing grounded theory*. Sage Publications.

Ching, K. W., & Singh, M. M. (2016). Wearable Technology Devices Security and Privacy Vulnerability Analysis. *International Journal of Network Security & Its Applications*, *8*(3), 19–30. https://doi.org/10.5121/ijnsa.2016.8302

Chiru, C., Rebedea, T., & Ciotec, S. (2014). Comparison between LSA-LDA-lexical chains. In *WEBIST 2014—Proceedings of the 10th International Conference on Web Information Systems and Technologies* (Vol. 2).

Cho, H., Rivera, M., & Lim, S. S. (2009). A multinational study on online privacy: Global concerns and local responses. *New Media & Society - NEW MEDIA SOC*, *11*, 395–416. https://doi.org/10.1177/1461444808101618

Correia, R., Wood, I., Bollen, J., & Rocha, L. (2020). *Mining social media data for biomedical signals and health-related behavior*.

Culnan, M. J., & Armstrong, P. K. (1999). Information Privacy Concerns, Procedural Fairness, and Impersonal Trust: An Empirical Investigation. *Organization Science*, *10*(1), 104–115. https://doi.org/10.1287/orsc.10.1.104

Datta, P., Namin, A. S., & Chatterjee, M. (2018). A Survey of Privacy Concerns in Wearable Devices. *2018 IEEE International Conference on Big Data (Big Data)*, 4549–4553. https://doi.org/10.1109/BigData.2018.8622110

de Arriba-Pérez, F., Caeiro-Rodríguez, M., & Santos-Gago, J. M. (2016). Collection and Processing of Data from Wrist Wearable Devices in Heterogeneous and Multiple-User Scenarios. *Sensors (Basel, Switzerland)*, *16*(9), 1538. https://doi.org/10.3390/s16091538

Domalewska, D. (2021). An analysis of COVID-19 economic measures and attitudes: Evidence from social media mining. *Journal of Big Data*, *8*. https://doi.org/10.1186/s40537-021-00431-z

Eastlick, M. A., Lotz, S. L., & Warrington, P. (2006). Understanding online B-to-C relationships: An integrated model of privacy concerns, trust, and commitment. *Journal of Business Research*, *59*(8), 877–886. https://doi.org/10.1016/j.jbusres.2006.02.006

Ekman, P. (1993). Facial expression and emotion. *American Psychologist*, *48*(4), 384–392. https://doi.org/10.1037/0003-066X.48.4.384

Feldman, R., & Sanger, J. (2006). *The Text Mining Handbook: Advanced Approaches in Analyzing Unstructured Data*. Cambridge University Press. https://doi.org/10.1017/CBO9780511546914

Glaser, B. G., & Strauss, A. L. (1967). *The discovery of grounded theory: Strategies for qualitative research*.

Haddi, E., Liu, X., & Shi, Y. (2013). The Role of Text Pre-processing in Sentiment Analysis. *Procedia Computer Science*, *17*, 26–32. https://doi.org/10.1016/j.procs.2013.05.005

Hopkins, D. J., & King, G. (2010). A Method of Automated Nonparametric Content Analysis for Social Science. *American Journal of Political Science*, *54*(1), 229–247. https://doi.org/10.1111/j.1540-5907.2009.00428.x

Jeong, B., Yoon, J., & Lee, J.-M. (2019). Social media mining for product planning: A product opportunity mining approach based on topic modeling and sentiment analysis. *International Journal of Information Management*, *48*, 280–290. https://doi.org/10.1016/j.ijinfomgt.2017.09.009

Jiang, D., & Shi, G. (2021). Research on Data Security and Privacy Protection of Wearable Equipment in Healthcare. *Journal of Healthcare Engineering*, *2021*. https://doi.org/10.1155/2021/6656204

Kang, K., Pang, Z., & Wang, C. (2013). Security and privacy mechanism for health internet of things. *The Journal of China Universities of Posts and Telecommunications*, *20*, 64–68. https://doi.org/10.1016/S1005-8885(13)60219-8

Kari, T., Kettunen, E., Moilanen, P., & Frank, L. (2017). Wellness Technology Use in Everyday Life: A Diary Study. *Digital Transformation – From Connecting Things to Transforming Our Lives*, 279–293. https://doi.org/10.18690/978-961-286-043-1.20

Kotz, D., Avancha, S., & Baxi, A. (2009). A privacy framework for mobile health and home-care systems. *Proceedings of the First ACM Workshop on Security and Privacy in Medical and Home-Care Systems - SPIMACS '09*, 1. https://doi.org/10.1145/1655084.1655086

Kotz, D., Gunter, C. A., Kumar, S., & Weiner, J. P. (2016). Privacy and Security in Mobile Health: A Research Agenda. *Computer*, *49*(6), 22–30. https://doi.org/10.1109/MC.2016.185

Krebs, P., & Duncan, D. T. (2015). Health App Use Among US Mobile Phone Owners: A National Survey. *JMIR MHealth and UHealth*, *3*(4), e4924. https://doi.org/10.2196/mhealth.4924

Landis, J. R., & Koch, G. G. (1977). The measurement of observer agreement for categorical data. *Biometrics*, *33*(1), 159–174.

Lazzarotti, J. (2015). *Wearables, Wellness and Privacy*. The National Law Review. https://www.natlawreview.com/article/wearables-wellness-and-privacy

Lee, C., Eze, U., & Ndubisi, N. (2011). Analyzing key determinants of online repurchase intentions. *Asia Pacific Journal of Marketing and Logistics*, *23*, 200–221. https://doi.org/10.1108/13555851111120498

Lee, L., Egelman, S., Lee, J. H., & Wagner, D. (2015). Risk Perceptions for Wearable Devices. *ArXiv:1504.05694 [Cs]*. http://arxiv.org/abs/1504.05694

Liu, B. (2010). *Sentiment Analysis and Subjectivity*. 38.

Lyon, D. (2001). *Surveillance society: Monitoring Everyday Life* (1st edition). Open University Press.

Malhotra, N., Kim, S., & Agarwal, J. (2004). Internet Users' Information Privacy Concerns (IUIPC): The Construct, the Scale, and a Causal Model. *Information Systems Research*, *15*, 336–355. https://doi.org/10.1287/isre.1040.0032

Miltgen, C., Popovič, A., & Oliveira, T. (2013). Determinants of end-user acceptance of biometrics: Integrating the "Big 3" of technology acceptance with privacy context. *Decision Support Systems*, *56*, 103–114. https://doi.org/10.1016/j.dss.2013.05.010

Motiwalla, L., Deokar, A. V., Sarnikar, S., & Dimoka, A. (2019). Leveraging Data Analytics for Behavioral Research. *Information Systems Frontiers*, *21*(4), 735–742. https://doi.org/10.1007/s10796-019-09928-8

Motti, V., & Caine, K. (2015). *Users' Privacy Concerns About Wearables: Impact of form factor, sensors and type of data collected* (Vol. 8976). https://doi.org/10.1007/978-3-662-48051-9_16

Motti, V. G., & Caine, K. (2014). Human Factors Considerations in the Design of Wearable Devices. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, *58*(1), 1820–1824. https://doi.org/10.1177/1541931214581381

Myers, M. D. (2009). *Qualitative research in business & management* (pp. xii, 284). Sage Publications Ltd.

Pandit, N. (1996). The Creation of Theory: A Recent Application of the Grounded Theory Method. *The Qualitative Report*, *2*(4), 1–15. https://doi.org/10.46743/2160-3715/1996.2054

Perez, A. J., & Zeadally, S. (2018). Privacy Issues and Solutions for Consumer Wearables. *IT Professional*, *20*(4), 46–56. https://doi.org/10.1109/MITP.2017.265105905

Plachkinova, M., Andrés, S., & Chatterjee, S. (2015). A Taxonomy of mHealth Apps – Security and Privacy Concerns. *2015 48th Hawaii International Conference on System Sciences*, 3187–3196. https://doi.org/10.1109/HICSS.2015.385

Sarker, A., O'Connor, K., Ginn, R., Scotch, M., Smith, K., Malone, D., & Gonzalez, G. (2016). Social Media Mining for Toxicovigilance: Automatic Monitoring of Prescription Medication Abuse from Twitter. *Drug Safety*, *39*(3), 231–240. https://doi.org/10.1007/s40264-015-0379-4

Solove, D. J. (2006). A Taxonomy of Privacy. *University of Pennsylvania Law Review*, *154*(3), 477. https://doi.org/10.2307/40041279

Westin, A. (1968). Privacy And Freedom. *Washington and Lee Law Review*, *25*(1), 166.

World Health Organization. (2011). *mHealth: New horizons for health through mobile technologies: second global survey on eHealth*. World Health Organization. https://www.cabdirect.org/globalhealth/abstract/20113217175

Young, S. (2018). Agency and the Digital Alter Ego: Surveillance Data and Wearable Technologies. *International Journal of Sociotechnology and Knowledge Development*, *10*(3), 41–53. https://doi.org/10.4018/IJSKD.2018070103

Yu, C., Jannasch-Pennell, A., & DiGangi, S. (2011). Compatibility between Text Mining and Qualitative Research in the Perspectives of Grounded Theory, Content Analysis, and Reliability. *The Qualitative Report*, *16*(3), 730–744. https://doi.org/10.46743/2160-3715/2011.1085

Zhou, T. (2012). Examining Location-Based Services Usage From The Perspectives Of Unified Theory Of Acceptance And Use Of Technology And Privacy Risk. *Journal of Electronic Commerce Research*, *13*(2), 10.