"I Don't Think I Can Share My Health Information": Understanding Users' Risk Perceptions about Personal Health Records Shared on Social Networking Services

ABSTRACT

Social networking services (SNS) provide an active online environment to exchange personal health information. However, it would be dangerous to share users' personal health record (PHR) on SNS without considering security and privacy risks.

To understand users' risk perceptions about sharing their PHR on SNS, we first conducted a qualitative user study by interviewing 16 participants. Next, we conducted a large-scale online user study with 497 participants to validate our qualitative results from the first study. Our study results show that a majority of users do not show strong motivation of sharing PHR on SNS due to several concerns such as misuse/abuse and security issues of shared PHR. In particular, participants are highly concerned about sharing diseases and diagnostic test results than other types of PHRs (e.g., details of hospital visits and medical interviews). However, we found about 55.13% of the participants have the experiences of sharing their health-related information on SNS. Also, we learned that users' sharing behavior for disease data can be significantly influenced by the severity/type of the disease as well as preferred recipients.

Based on our findings, we propose a practical privacy setting method to automatically determine whether users' posts can be shared with everyone on SNS by analyzing the keywords frequently occurred in health-related posts. Our implementation using Random Forest achieved an F-measure of 98.7%, indicating that PHR can be restrictively shared with high accuracy, when sharing health-related posts on SNS.

KEYWORDS

Personal Health Record, Social Networking Service, User Privacy

1 INTRODUCTION

In recent years, online social networking services (SNS) have often been used as a powerful platform for people to freely discuss and seek help for questions relating to their personal interests in health and medicine [8, 21]. People can use SNS to seek answers to specific questions about health conditions and symptoms because they can directly and/or anonymously interact with experts and peers who have experienced similar health problems. Patients also often post their personal experiences and situations to help and empathize others with similar health conditions [19, 21, 28]. For example, Reddit community has provided a place (https://www.reddit.com/r/mentalhealth/) for the exchange of information about mental health disorders because questions and answers can be posted anonymously (see Appendix A). However, since health information is inherently private and sensitive, it is of utmost importance to carefully share and obtain health information via SNS. Several previous studies [25, 26, 29] demonstrated that SNS users are highly concerned about unwanted leakage of their personal health information to others via SNS.

In this paper, we extend those previous studies [25, 26, 29] with a deeper analysis of users' privacy concerns about personal health record (PHR), which is defined as electronic resource of health information maintained by individuals [7] shared on SNS. While previous research has mainly studied general and broad health-related information and topics, we specifically investigate users' sharing on medical data (e.g., diseases and diagnostic test results) categorized in Hospital Information System (HIS). We focus on medical data in HIS because it is a widely used comprehensive information system that manages various types of medical records, administrative and financial data for hospitals [11] and provide a de-facto standard for categorizing medical data, which health professionals and patients can easily understand. The overarching goal of our research is to elicit and gather users' concerns and perceptions on practical requirements collected from HIS for online health care services and provide practical suggestions for designers of SNS systems to improve the user experience of sharing PHR on SNS.

To analyze users' perceptions, reasons, and behaviors related to sharing PHR on SNS, we conducted two different user studies approved by the Institutional Review Board (IRB) at our university. In the first user study, we performed semi-structured interviews with 16 participants to find out what concerns users specifically have regarding sharing private health information via SNS. We further investigated specific factors that influence users' decisions to share PHR with other online users. In the second study, we performed a quantitative large-scale online study with 497 participants using Amazon MTurk to complement the first qualitative study. To conduct a more ecologically valid user study, we used a real-world PHR related terms from HIS [11] and health topic examples from Reddit rather than more general or broad terms from prior research.

From our study, we find that generally participants were primarily concerned with several issues when disclosing their PHRs on SNS. The main reasons are due to the following: (1) no need to share PHR, (2) misuse/abuse of shared PHR, and (3) security issues. However, more than half of participants, 274 out of a total 497

participants (55.13%), still shared PHR on their SNS with these perceived concerns. Therefore, to improve PHR sharing on SNS, SNS system should be designed for users to effectively manage the control of their private PHRs on SNS so that it can protect shared PHR from unwanted disclosure to SNS users. In addition, participants rated the "diseases and medical test results" as the most sensitive information to be shared on SNS among various PHR categories (e.g., details of hospital visits, medical interviews, and prescription drugs). In addition, interestingly users' sharing behavior for disease data can be strongly influenced by the severity/type of the disease. Furthermore, we found that users are more likely to share their health information with preferred recipients only.

Based on our findings, we propose a practical privacy setting method to help SNS users determine whether a given health-related post can be shared with everyone on SNS by automatically analyzing the post in order to reduce the risk of unwanted PHR exposures on SNS. Our implementation using a Random Forest classifier achieved the F-measure of 98.7%. These results signify that PHR can be restrictively shared with high accuracy based on the severity of PHR.

2 RELATED WORK

HIPAA and privacy rule: The Health Insurance Portability and Accountability Act (HIPAA) was established in 1996 to protect the privacy of patient health information and improve the effectiveness of the healthcare system [20]. The standards for privacy of individually identifiable health information (privacy rule) issued by the department of Health and Human Services (HHS) provides a set of standards for the protection of personal health information. The rule applies to health plans, health care providers, and health care clearinghouses. According to the privacy rule, all individually identifiable health information must be protected, including demographic data, physical or mental health, provision of health care and payment history. Therefore, users' personal health information on SNS should be securely protected in accordance with HIPAA guidelines. In this paper, we follow the terms defined by HIPAA to develop our survey questionnaires (in Appendix B and C) for our user studies.

Personal Health Record (PHR): PHR is defined as *"an electronic application through which individuals can access, manage and share their health information, and that of others for whom they are authorized, in a private, secure, and confidential environment"*, where this is one of the most widely accepted definitions by the Markle Foundation [7]. PHR can be contrasted with the physician's records on patient information, which is managed by the health care service provider [24]. PHRs can contain a variety of health-related data, including allergies, adverse drug reactions, family medical history, illnesses, hospitalizations, imaging and laboratory test results, vaccinations and observations of daily living, etc.

However, as people have started sharing PHR online publicly, it has become a serious issue [6, 13, 17]. Therefore, several previous studies [2, 3, 23] were conducted to review the privacy policies of online systems. Carrión et al. [3] evaluated the privacy policies of 22 websites to check whether the privacy of patients' data was preserved in accordance with the HIPAA guidelines. In this research, we extend their work by analyzing users' risk perceptions about

PHR shared among users to elicit practical security and privacy requirements. We further provide the practical implementation for desingers of SNS systems to better support PHR sharing features.

PHR sharing on SNS: SNS has often been used as a place for people to seek help for questions related to their personal interest in health and medicine [8, 25, 26]. The primary motivations for patients using health related SNS are to enhance their knowledge, exchange ideas and opinions, and engage in social support [1]. Oh et al. [22] showed the significance of perceived social support from SNS, especially emotional support, in enhancing one's self-efficacy. Also, SNS can make a meaningful contribution to the health status management aspect for people with chronic diseases [18]; Second Life (https://secondlife.com/) could be used to reduce the anxiety and stress of patients with chronic diseases [12]. Reddit community provides a practical platform to discuss, vent, support, and share information about mental health, illness, and wellness where questions and opinions about mental illness can be posted anonymously [4]. However, people have become increasingly concerned about sharing health information on SNS as the awareness of information security has increased [25, 26, 29]. Specifically, SNS users were afraid of the possibility of misuse of their health information and were fear of the social stigma [5, 25]. In particular, Torabi and Beznozov [26] examined health information sharing practices to analyze the main factors influencing users' motivation to share health information on Facebook. They found that the benefits obtained from previous health information sharing experiences and users' overall attitudes toward privacy were correlated with their willingness to disclose health information on Facebook. However, they do not dive deep into specific security and privacy concerns for PHR on SNS. In this paper, we extend their work by analyzing the main concerns and reasons for sharing PHR on SNS in more detail.

To mitigate users' privacy concerns, most SNS platforms such as Facebook, Twitter, Instagram, and Snapchat typically provide some types of access control mechanisms, allowing users to control what user groups can access to their profile data and contents they create on the SNS. Johnson et al. [14] found that the Facebook's privacy controls allow users to effectively manage the outsider threat, but they are not suitable for mitigating concerns over the insider threat. Hartzler et al. [10] introduced a tool called HealthWeaver which enables patients to create, manage, and share personal health information with selected users of their social networks and evaluated its usability. Even though participants were highly confident with the tool, several participants made errors in determining what information was shared with whom. These study results demonstrate that it is very challenging to develop a privacy control mechanism that enables granular control over what personal health information is shared and with whom within social networks. In this paper, we propose a *dynamically configurable* privacy control mechanism that helps users determine which medical information can be shared with other SNS users using machine learning algorithms.

3 EXPLORATORY STUDY: INTERVIEW

We conducted a semi-structured interview study with 16 participants to better understand the practices and concerns about sharing

PHR behaviors on SNS. This qualitative study was designed to identify main factors that influence users' PHR sharing practices. We also measured users' risk perceptions, and reasons and concerns by investigating the specific target recipients whom participants are willing to share their PHRs with, through our interview process. Then, the interview results were utilized as the basis for the next quantitative online conformation study.

3.1 Participants recruitment

We recruited 16 participants from two online communities (*Health World* and *Sansamo*) related to healthy living. We chose these two different communities in an effort to recruit people with more diverse background, and health care and SNS usage experience. In *Health World*, people mainly share the health management and treatment information, while *Sansamo* provides several forums to share medical counseling, and local hospital and healthcare provider information. The qualifications to participate in this user study were those over 18 years old and used SNS at least once a month. Data on demographic background as well as health related information (e.g., existing health conditions, medical history and healthcare environments) were also collected through our interviews. After initial screening, qualified individuals met directly with our researchers. Each participant received a gift card worth $30 as a compensation for participating in our user study. Ethical perspective of our interview study was validated through the IRB at our institution.

3.2 Data collection and analysis

We developed an interview guide to collect qualitative responses from participants. The interview questionnaire with 24 questions are provided in Appendix B. We developed and extended our questions based on similar research by Torabi and Beznozov [26] to investigate factors of sharing health information. Our interview questions are structured in the following four categories to assess: (1) SNS usage and activity (3 Qs), (2) PHR sharing on SNS (11 Qs), (3) Health care environment (5 Qs), and (4) Health status of participants (5 Qs). The first 'SNS Usage and Activity' questions are to gauge whether users are active SNS users. The second 'PHR Sharing on SNS' questions are to find details of users' sharing behaviors, perceptions, and concerns, and we further asked specific reasons for sharing or not sharing those. The last 'Health care environment' and 'Health status' questions are to examine participants' health related information for medical institutions and health care providers. We present the actual in Appendix B.

The data collection period was from November 2017 to January 2018. We conducted a semi-structured interview that lasted an average of 40 minutes per visit. The interviews were recorded under the consent from participants, and the recordings were transcribed during the entire interview. The Nvivo11 software was as the main data analysis tool for coding and analyzing the transcripts to perform the grounded theory analysis [9]. The first researcher conducted the open coding to identify and categorize interesting phenomena in the data, and the second researcher checked and reviewed the coded results. Next, two researchers analyzed and discussed about the results, in order to establish explicable relationships among the open codes. We obtained a total of 187 codes and 44 unique codes. After interviewing and analyzing 12 participants, we did not

find any new codes, yielding code saturation. Hence, further interviews did not contribute to new finding. An average of 21.88 codes were obtained per interview, with a standard deviation of $\sigma = 4.87$, where the two coders reached approximately 90% agreement.

3.3 Results

Demographics and SNS usage: The age range of the participants was between 24 and 55 years (mean: 33.25, σ: 7.54) and 12 were females. Except one participant, all participants had at least a bachelor's degree. While 5 participants were unemployed, other people worked for various industry sectors. All participants except one was daily active SNS users and 12 participants posted and shared information on SNS at least once a week. 12 and 4 participants were active users on Instagram and Facebook, respectively.

Health care environment and health awareness: Only 3 participants answered that their health condition was not good at the time of the interview. P4 was a breast cancer patient; P15 had hepatitis B; P11 did not have any specific illnesses, but she recently felt that her health was not good. Also, 14 participants reported to have below average level of knowledge of health. All the participants knew the definition of basic health and medical terms from HIS but they expressed they did not have clear or enough understanding of clinical knowledge for specific diseases, symptoms, etc.

Only one healthy participant P5 had a high level of health knowledge because a family member had experienced a serious health condition. In terms of health care management, most participants did not back up or stored their PHR. Instead, participants relied mainly on hospital information system such as HIS, where they were able to acquire their past medical records, if needed.

Reasons for sharing PHR on SNS: 9 out of 16 participants had experience with sharing PHR on SNS. We examined their motivations.

First of all, 6 participants wanted to receive comfort, attention, and encouragement from others. For example, P4 used SNS mainly for this purpose.

> "I wanted to seek attention and sympathy from others." – P5 (M, 30)

Some participants expressed that they used SNS as a medium to inform and update their health status.

> "When I was a little sick, I used SNS to inform friends that I had recovered." – P6 (F, 30)

P4 did not mind disclosing her experience and knowledge if it helped others.

> "I had a breast cancer surgery in 2014. I think I have a duty to inform patients like me, so I'm posting a lot of health information on Facebook." – P4 (F, 55)

Also, SNS was used as a platform to receive opinions, information and advice from others. P14 used SNS to follow up helpful diet and health related information.

> "I used SNS to receive more opinions about cosmetic surgery from different people. Also, I opened up a new Instagram account to receive helpful comments for diet and health management." – P14 (F, 24)

Additionally, participants mentioned the following reasons for sharing: managing and tracking PHR, and habitually sharing daily life. From these user responses, we constructed the following 7 reasons (R1–R7) for sharing PHR on SNS in Table 1. In particular, we highlight key factors in bold (in Table 1–5), which appear to be strong reasons indicated by participants.

On the other hand, 7 participants never shared PHR on SNS because they considered it is unnecessary to inform others of their health conditions. One participant, P13, did not desire to make others worry about him.

> "I think SNS is a space to hold my own story through empathy, so I personally like to share good news or enjoyable things. But I think it is unnecessary to post health problems, even if I can get attention or comfort." – P13 (M, 32)

However, 11 participants (68.8%) mentioned that possible privacy risk was the main reason that they were afraid of sharing their PHR on SNS. In summary, we found that more than half of participants perceived sharing PHR is risky, even though 9 of them actually still did share their PHR on SNS. Therefore, we need a better mechanism to mitigate and address users' various concerns, which we examine next.

Table 1: Decomposed reasons for sharing PHR.

Reasons for sharing PHR (Acronym)
R1: To inform my health issues (Inform)
R2: To share healthcare experience and knowledge with others (Share Exp.)
R3: I have been habitually sharing my daily life (Habitual)
R4: To know others' opinion/comments about my health issues (Opinion)
R5: To gain others' sympathy or attention with my health issues (Sympathy)
R6: To be motivated on health care and inspire myself
R7: To manage my PHR history (Manage)

Users' concerns about sharing PHR: In order to look deeper into users' PHR sharing behaviors and reasons, we asked users' concerns about sharing PHR on SNS. Most participants provided several interesting concerns about sharing PHRs, where they were allowed to provide multiple responses. 1 participants stressed social concerns such as discrimination, prejudice, and bias caused by sharing PHR on SNS. And most of participants were afraid of the prejudices or disadvantages that might be caused from disclosing their health condition online.

> "Because I might get a job someday, I'm afraid I would be disadvantaged by my health problems." – P11 (F, 37)

In the case of P15, who had a chronic infection with hepatitis B, suffered from those who misunderstood the disease.

> "Some people do not know that hepatitis B cannot be infected from other people. Therefore, when I said that I am a hepatitis carrier, they were afraid of being infected with hepatitis B from me." – P15 (F, 48)

In addition, 4 participants said that they did not want to be considered as unhealthy or attention seekers by others. P3 was reluctant to receive excessive attention or sympathy from the surroundings due to their health issues.

> "Except for mild sickness like cold and flu, I do not want to be seemed as a person who are begging for comfort. I

> do not want someone to feel sorry for me because of my health problem." – P3 (F, 29)

Also, 5 participants worried that their PHR on SNS can be misused or abused. For instance, P14 joined the community for weight management and was contacted by a variety of drug companies such as selling weight loss supplements.

> "I have a lot of contacts from H company and such an advertising company." – P14 (F, 24)

Moreover, one participant mentioned the possibility that the exposed PHR can be misused for cybercrimes.

> "As the personal data is exposed, I worry that voice phishing could happen to my family." – P12 (F, 34)

Another reported major privacy concern by participants is that PHR might be found by search engines such as Google. While P6 had a positive attitude towards sharing her PHR on SNS, she was afraid that her information could be possibly accessible by other online users through search engines:

> "I am a little concerned that my health information would be searched. I'm trying not to contain sensitive terms in hashtags to prevent searching." – P6 (F, 30)

On the other hand, 6 participants mentioned that it seems unnecessary and did not express the strong motivation to share PHR.

> "It seems unnecessary because it is personal for me. Also, I am concerned out others' attention? I can share it if anonymized." – P2 (F, 29)

> "Seems unnecessary. I do not want to reveal my location and status. Also, I am concerned that I might be perceived as a unhealthy person." – P10 (F, 30)

However, from those participants' responses of "It seems unnecessary · · · " (see P2 and P10), we find that other concerns were also frequently mentioned together indicating that there are likely to be complex and multiple reasons behind their answers. To validate our findings through a large scale online user study, we divided collected concerns about sharing PHR into the following C1–C9 in Table 2.

Table 2: Decomposed concerns about sharing PHR.

Concerns about sharing PHR (Acronym)
C1: Perceived attention seeker (Attention)
C2: Perceived unhealthy person (Unhealthy)
C3: People might be overly concerned about me (Over-concern)
C4: Concerns about disadvantages (Disadvantages)
C5: My PHR can be searched on the Internet (Search)
C6: My PHR can be abused or misused (Misuse)
C7: My PHR is known to people who I dislike (Dislike)
C8: Concerns about security issues (Security)
C9: It seems unnecessary (No need)

Degree of concerns by PHR categories: Since PHR contains a diverse range of data, we first organize the PHR to following seven categories (U1–U7) from HIS as shown in Table 3. Then, we investigate and identify degree of users' concerns with respect to specific PHR categories in sharing on SNS.

Through interviews, we learned that information about "U7: Hospital/clinic appointment" did not matter much with respect to privacy. On the other hand, information about "U4: Laboratory and

imaging test results" as well as "U1: Disease and illnesses" turned out to be the the most reluctant information that users wanted to share on SNS. As the importance of U1 and U4 was discovered throughout the interviews, we expanded and constructed more specific questions in U1 and U4 to better understand user's perception of sharing behavior. These are further evaluated through a large scale online user study.

Table 3: De-factor standard PHR categories.

PHR category (Acronym)
U1: Diseases and illness (Diseases)
U2: Sign and symptom (Symptom)
U3: Medical interview (Interview)
U4: Laboratory and imaging test results (Test Results)
(e.g. blood test, urine test, X-ray, MRI) (Test Results)
U5: Prescription drug information (Drug Info)
U6: Treatments (Treatments)
(e.g. Surgery, chemotherapy, medication)
U7: Hospital/clinic appointment (Appointment)

Preferred PHR recipients: Next, we examine whom users are likely to share their PHRs with on SNS. In our study, 15 participants (93.8%) were willing to share their PHRs with not only partner/spouse, but also family members. 9 participants mentioned that they were able to share PHR with close friends and acquaintances. However, more than half of the participants did not want to share their PHR with colleagues and online friends.

> "I don't think I can share my health information because I feel like it's hard to trust people on SNS." – P3 (F, 29)

In addition, 15 participants mentioned that the details of PHR to be shared depended on preferred PHR recipients. In the event that specific PHR contains a serious health condition or excessive information, participants intended to share their PHR with only very close people such as family members and close friends.

Also, participants' decision is also related to emotion and empathy as described by P4 and P5.

> "If I have a fatal disease like cancer, I don't feel like telling people about it. But it's okay for family." – P5 (M, 30)

> "I think I can only share images (taken by scanning medical imaging devices such as CT and MRI) to family members or friends. But, for example, cholesterol level is high in blood tests. It's okay to share these kind of things." – P4 (F, 55)

Overall, participants' decision to share is strongly influenced by trust and closeness to PHR recipients.

Factors related to sharing of disease and illness: As the importance of U1. Diseases and illness was revealed from the previous question, we asked more details. 13 participants pointed out that the severity of the disease is one of the most important factors to consider when sharing PHRs on SNS. For instance, P1 remarked as follows.

> "I'm more likely to post common illnesses such as cold and stomach on my SNS, but I never tell about serious disease like cancer, infectious disease and mental disorder." – P1 (M, 34)

Moreover, injured body parts related to disease was another major factor influencing users' sharing decision. Particularly, female participants were afraid of sharing about female-related diseases.

> "If the disease was related to genital, I will never open this to anyone" – P6 (F, 30)

> "I mean there are diseases related to women. If I have a female-specific disease, I can't post it on the SNS." – P10 (F, 30)

In addition, the characteristics of disease seemed to strongly affect decision to share PHR on SNS. For example, participants were concerned about the misconceptions and prejudices for infectious or contagious diseases, such as HIV or hepatitis B virus. However, illnesses such as flu and food poisoning were not of concern. We noticed that participants were reluctant to share incurable or chronic diseases among infectious or contagious diseases.

Based on users' concerns, we constructed the following D1–D7 factors in Table 4 to investigate further about users' concerns through a large scale online user study, where D1 and D5 appear to be strong influential factors in sharing U1.

Table 4: Detailed derived factors related to sharing of U1: Disease and illness.

Influential Factors for U1. disease and illness (Acronym)
D1: Location of disease or damaged body part (Location)
D2: Stages of disease progression (Stages)
D3: Communicable disease (Communicable)
D4: Rareness of disease (Rareness)
D5: Severity of disease (Severity)
D6: Treatment period (Period)
D7: Type of disease (Type)

Factors related to sharing of diagnostic test results: Diagnostic test result reflects individual's accurate health status. Therefore, it is regarded as one of the most sensitive personal information. 4 participants were not reluctant to share diagnostic test results, if no abnormal signs were found in the results. On the other hand, 12 participants were unwilling to disclose if the test result indicated a fatal disease or a serious health condition. They tended to avoid posting bad news on SNS in order not to have a negative psychological impact to others.

> "I guess it depends on the results rather than the type of test. Because I think it's important what the problem is with the test results. In my case, disclosing to SNS will be depended on the stage of the disease." – P16 (F, 29)

Moreover, participants showed significant differences in sharing diagnostic test results depending on their sensitivity. For instance, P7 remarked as follows.

> "If I get a health check and the results are good, I think I can post it on my SNS. But the result shows that I have a serious health problem, or if I have a disease and I'm on treatment for it, I do not think I will share it with anyone. It is a little sensitive issue." – P7 (F, 32)

We found that 3 participants were sensitive to diagnostic tests related to the reproductive system, while they were not concerned about sharing certain types of diagnostic test (e.g., blood test, X-ray, and MRI) results. On the other hand, 7 participants mentioned that they were likely to share PHR restrictively based on diagnosed parts

of body. Also, 2 participants were less likely to share the results of imaging rather than the results of in vitro diagnostics, which are performed for screening purposes.

From these user responses, we constructed the following 8 factors (T1–T8) affecting decision to share the diagnostic test results on SNS in Table 5. Again, we tested T1–T8 with online participants and validate our findings.

Table 5: Detailed derived factors related to sharing of U4: diagnostic test results.

Factors for U4. diagnostic test results (Acronym)
T1: Diagnosed parts of body (Parts)
T2: General Health checkup (General)
T3: Sign and symptom of health problem (Symptom)
T4: Post-treatment monitoring (Post)
T5: Accidents and injuries (Accidents)
T6: Amount and detail of diagnostic test results (Amount)
T7: Results of diagnostic test (+/- results)
T8: Type of diagnostic test (Type) (e.g., Blood test, X-ray, MRI)

4 CONFIRMATORY STUDY: ONLINE SURVEY

To verify our findings derived from the first exploratory study, we conducted a large-scale online user survey, where survey questions are provided in Appendix C. The goals of the online study are to confirm our findings in qualitative study and to further investigate users' different sharing behavior on specific PHR categories vs. specific groups of recipients. We describe details of participants recruitment, data collection, and results in this section.

4.1 Participants recruitment

Participants were recruited from Amazon MTurk, which is the most popular crowdsourcing platform. Since MTurk workers are likely to use SNS [15], it seems suitable to collect high-quality data for our research. We recruited participants with at least 100 completed human intelligence tasks (HITs) and > 95% HIT acceptance rate. All our participants were the residents of the United States, and at least 18 years old. Participants who accepted our consent form were only allowed to participate the survey and were rewarded with $1.00 through MTurk, if they successfully completed the entire survey.

4.2 Data collection

Our survey consisted of 15 questions (see Appendix C) in two phases and designed to be completed within 30 minutes. The questionnaires contained the following items: (1) demographics and SNS usage, (2) reasons for sharing PHR, (3) privacy and security concerns about sharing PHR, (4) rank of various PHR types measured by its sensitivity, (5) sharing PHR details vs. preferred PHR recipients, (6) factors related to sharing of diseases and illnesses, and (7) factors related to sharing of diagnostic test results.

4.3 Data analysis methods

The survey results were analyzed using R software. The statistical confidence in the reasons/concerns for sharing and not sharing were tested using Mann-Whitney U (MWU) test. We also performed Chi-Square test to compare the proportions of PHR sharing and PHR non-sharing groups. Post-hoc comparisons were corrected for multiple-testing using Bonferroni correction when appropriate.

4.4 Results

Demographics: 515 participants completed the survey from March 16 to 23 in 2018. 18 of them were disqualified, since they took less than five minutes to complete the survey. On average, participants took the survey 12.68 minutes with a standard deviation of 5.23 minutes. Therefore, we only use a total of 497 responses in the rest of paper. The majority was in the 31–40 (35%) and 18–30 (31%) age groups, where about 51% were female. Approximately, 60% had a bachelor's degree or higher. For the occupation, the majority were in management (12.5%), education (11.7%) and finance (11.3%). The details of the demographics are presented in Appendix D.

We also asked participants about their health insurance. A total of 39 participants (7.8%) said that they did not have health insurance or did not know whether they have or not. On the other hand, 458 participants (92.2%) had one or more health insurances and 62% of them were covered by health insurance at work. As for SNS usage, 85.1% of the participants checked their SNS at least once a day and 64.2% of participants posted at least once a week. Especially, 19.3% of participants marked that they posted daily on SNS.

Reasons for sharing PHR on SNS: Out of the total 497 participants, 274 (55.1%) shared PHR on their SNS, which was quantitatively similar to the result (51.4%) in the previous study [26]. Those who share PHR, 92.3% used Facebook (253), followed by Instagram (107) and Twitter (84).

To identify users' reasons for sharing PHR on SNS, we asked participants to rate seven (R1–R7) reasons derived from our exploratory study in Section 3 on a 5-point Likert scale from 1 (strongly disagreed) to 5 (strongly agreed), where 3 being neutral. The R1–R7 reasons were randomly placed and presented to participants during the survey.
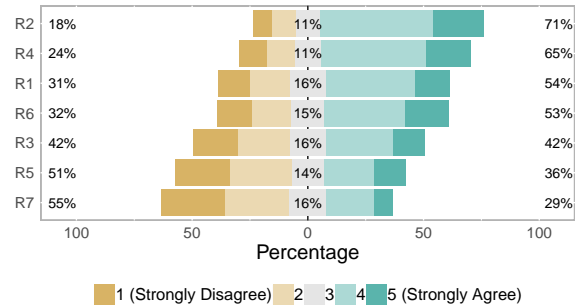

Figure 1: Reasons for sharing PHR on SNS (R1: Inform, R2: Share Exp., R3: Habitual, R4: Opinion, R5: Sympathy, R6: Motivate, and R7: Manage, as defined in Table 1).

As shown in Figure 1, "R2: To share healthcare experience and knowledge with others" and "R4: To know others' opinion/comments about my health issues" were top two reasons chosen by MTukers. We found a statistically significant difference between the top two reasons (R2 and R4) and all other remaining reasons (all $p \leq 0.03$, Bonferroni-corrected MWU) except R4 vs. R6 ($p = 0.1378$). This showed clear motivations for obtaining health-related knowledge or opinions from others. Interestingly, despite "R5: To gain others' sympathy or attention with my health issues" was one of the frequently

mentioned reasons for sharing PHR on SNS in the exploratory study, more than 51% of survey online participants disagreed with it.

Users' concerns about sharing PHR: To identify users' concerns about PHR sharing on SNS, we asked participants to rate nine (C1–C9) reasons derived from our exploratory study in Section 3. using the same 5-point Likert scale.

We analyzed those results for both groups who shared (PHR sharing group) and did not share PHR on SNS (PHR non-sharing group), respectively. Figure 2 shows the percentage of participants who selected "strongly agree" or "agree" on C1–C9.

(a) PHR sharing group
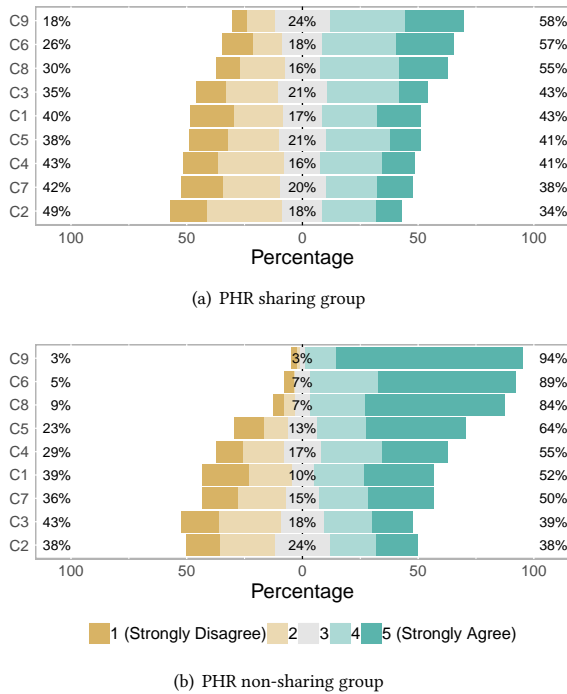
(b) PHR non-sharing group

Figure 2: Concerns about sharing PHR (C1: Attention, C2: Unhealthy, C3: Over-concern, C4: Disadvantage, C5: Search, C6: Misuse, C7: Dislike, C8: Security, and C9: No need, as shown in Table 2).

For PHR sharing group, "C9: It seems unnecessary", "C6: My PHR can be abused or misused" and "C8: Concerns about security issues" were top three concerns chosen because there was a statistically significant difference between the top three concerns (C9, C6 and C8) and all other remaining concerns (all $p \leq 0.004$, Bonferroni-corrected MWU) except C8 vs. C3 ($p = 0.053$) and C6 vs. C3 ($p = 0.114$). For PHR non-sharing group, there was a statistically significant difference between the top reason (C9) and all other remaining concerns (all $p \leq 0.002$, Bonferroni-corrected MWU). Overall, MTukers rated security concerns (C6, C8, and C5) higher than social concerns (C1, C2, C3, C4 and C7) directly.

Not surprisingly, in all concerns, participants who have never shared PHR had a higher percentage of "agree" rather than "disagree". In PHR non-sharing group, about 94% of participants considered that sharing PHR with others was unnecessary. There were

significant differences between PHR sharing group and PHR non-sharing group in C9, C6, C8 and C5 (all $p \leq 0.004$, Bonferroni-corrected Chi-Square), while we failed to show significant difference in C1, C2, C3, C4 and C7. This is somewhat expected. PHR non-sharing group participants did not share their PHRs on SNS because they were likely to be more concerned about some security or social concerns. Therefore, it would be important to address their top concerns in order to increase the adoption rate of SNS for PHR sharing.

Degree of concerns by PHR categories: Because we found that participants PHR sharing concerns were changed by PHR categories in the exploratory study in Section 3, we asked participants to rank the PHR categories (U1–U7) as defined in Table 3 , where 1 is the most sensitive information to disclose to SNS and 7 being the least sensitive. Similarly, we analyzed the ranking results for both PHR sharing group and PHR non-sharing group, respectively. The results are shown in shown in Figure 3.
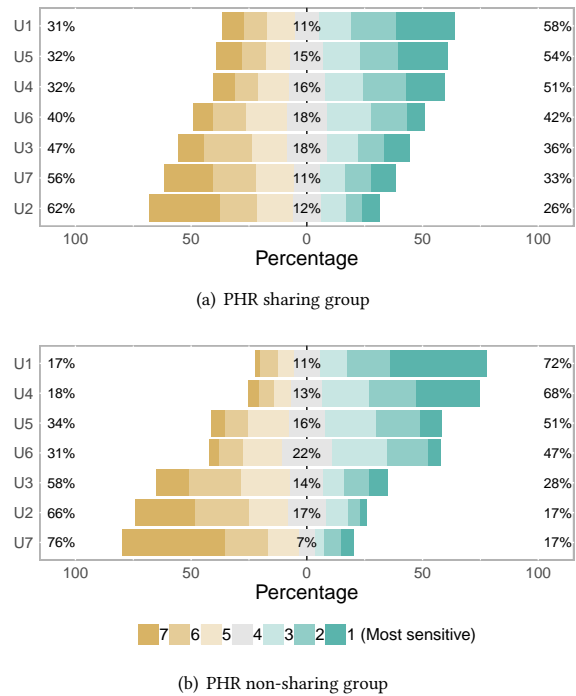
(a) PHR sharing group

(b) PHR non-sharing group

Figure 3: Rank order of unwillingness to share PHR on SNS by specific PHR categories (U1: Diseases, U2: Symptom, U3: Interview, U4: Test Results, U5: Drug Info, U6: Treatments , and U7: Appointment, as defined in Table 3).

For PHR sharing group, "U1: Diseases and illness", "U5: Prescription drug information" and "U4: Laboratory and imaging test results" were chosen as the most sensitive PHR categories because there was a statistically significant difference between the top three categories (U1, U5, and U4) and all other remaining categories (all $p \leq 0.006$, Bonferroni-corrected MWU) except U4 vs. U6 ($p = 0.1724$). For PHR non-sharing group, U1 and U4 were chosen because there was a statistically significant difference between the top two categories (U1 and U4) and all other remaining categories (all $p \leq 0.0003$,

Bonferroni-corrected MWU). This result is consistent with our exploratory study, where interviewees also strongly agreed that U1 and U4 were the top two reasons in PHR categories in affecting their sharing decisions. On the other hand, "U5: Prescription drug information" was highly rated by MTurkers for PHR sharing group.

Interestingly, we did not find significant statistical difference between PHR sharing group and PHR non-sharing group in all PHR categories except U2 ($p = 0.009$, Bonferroni-corrected Chi-Square).

Preferred PHR Recipients vs. PHR Category: Next, we examined how users' preferences were affected by PHR categories (U1–U7) and PHR recipient groups (Family, Friends, Acquaintances and Others). This allows us to measure and compare the sharing ratio of specific PHR category to the preferred recipient (e.g., U1: Diseases and illness to family vs. U3: Medical Interview to family). Figure 4 shows the sharing ratio of preferred recipients per each PHR category (U1–U7).
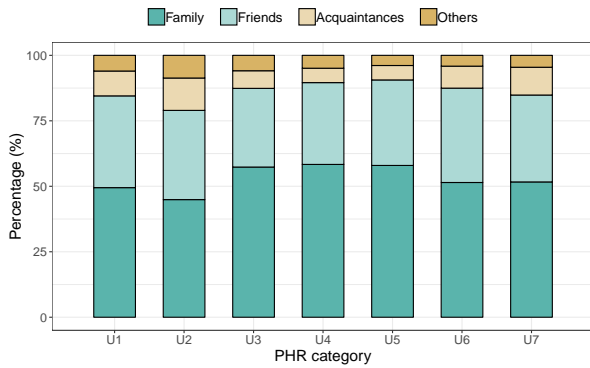

Figure 4: Distribution of preferred recipients by PHR category (U1: Diseases, U2: Symptom, U3: Interview, U4: Test Results, U5: Drug Info, U6: Treatments, and U7: Appointments, as defined in Table 3).

Overall, across all PHR categories, preference to share PHR to family members (63.17%) and close friends (40.39%) were high, while less than 11% of participants were willing to share PHR with acquaintances (10.55%) or anonymous SNS users (6.69%). Despite the fact that U1 and U4 were the top ranked PHR categories, preferred recipients have a similar sharing ratio distribution across all PHR categories, as shown in Figure 4. Interestingly, we found that there was no statistical difference in selecting preferred recipients across different PHR categories. This result shows that access control according to specific categories may not be the best way because users would share PHR over their preferred recipients without significant changes.
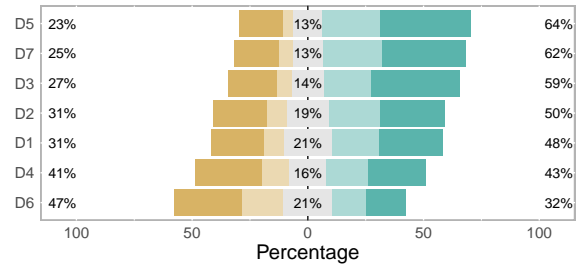
Factors related to sharing of diseases and illnesses: To identify main factors that influence the sharing of "U2: Diseases and illnesses on SNS", we asked participants to rate seven factors (D1–D7) obtained from the exploratory study in Section 3 on a 5-point Likert scale from 1 (not influential at all) to 5 (extremely influential).

We present the factors ranked by the highest to lowest influential factor for PHR sharing and non-sharing groups in Figure 5. (a) and (b), respectively. Interestingly, the rank of influential factors rated by two groups were almost the same (except the reverse order of D1 and D4), as shown in Figure 5. In fact, we did not find statistically significant difference between PHR sharing group and PHR non-sharing group in all factors (all $p \geq 0.24$, Bonferroni-corrected Chi-Square).
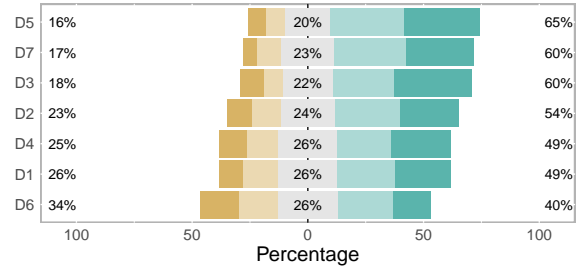
(a) PHR Sharing Group

(b) PHR Non-Sharing Group

Figure 5: Factors affecting sharing of diseases and illnesses (D1: Location, D2: Stages, D3: Communicable, D4: Rareness, D5: Severity, D6: Period, and D7: Type, as defined in Table 4).

For PHR sharing group, "D5: Severity of disease", "D7: Type of disease" and "D3: Communicable disease" were chosen as the most important factors, because the test showed a statistically significant difference between the top three factors (D5, D7 and D3) and all other remaining categories (all $p \leq 0.02$, Bonferroni-corrected MWU) except D3 vs. D2 ($p = 0.2044$). For PHR non-sharing group, D5, D7 and D3 were still chosen, because there was a statistically significant difference between the top three factors (D5, D7 and D3) and all other remaining categories (all $p \leq 0.04$, Bonferroni-corrected MWU).

Interestingly, the major factor results are slightly different from our observation from the exploratory study in Section 3. In our exploratory study, some interviewees mentioned that they were reluctant to disclose female-specific diseases because of the area of the body the diseases occurred (D1: Location of disease or damaged body part). According to our confirmatory study results, however, D1 was ranked 5th, being less likely to be a major factor as shown in Figure 5. We surmise that underlying demographic differences may explain this. The 16 participants in the exploratory study were recruited from two healthy living related online communities. In fact, 12 (75%) of them were females, while the participants in the confirmatory study were recruited from a general crowdsourcing

platform and the proportion of female participants is 51.31%. Therefore, the female participants in the exploratory study were more highly concerned with locations of diseases or damaged body parts.

Factors related to sharing of diagnostic test results: To examine the factors that influence the decision to share "U4: Laboratory and imaging test results", we asked participants to rate 8 factors (T1–T8) derived from our exploratory study in Section 3. Participants were asked to rate 1 as "not influential at all" to 5 as "extremely influential" for each T1–T8. The results for PHR sharing and non-sharing groups are shown in Figure 6.(a) and (b), respectively.

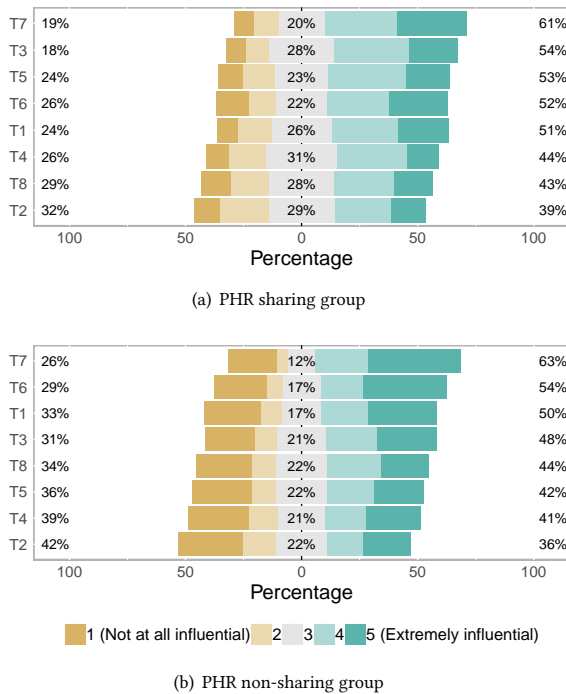(a) PHR sharing group

(b) PHR non-sharing group

Figure 6: Factors affecting sharing of diagnostic test results (T1: Parts, T2: General, T3: Symptom, T4: Post, T5: Accidents, T6: Amount, T7: +/-Results, and T8: Type, as shown in Table 5).

For PHR sharing group, "T7: Results of diagnostic test" was ranked as the highest among all factors, but the differences between T1–T8 were not significant. For PHR non-sharing group, "T7: Results of diagnostic test" and "T6: Amount and detail of diagnostic test results" were chosen as the important factors because there was a statistically significant difference between the top two factors (T7 and T6) and all other remaining factors (all $p \leq 0.0008$, Bonferroni-corrected MWU) except T6 vs. T1 ($p = 1.000$) and T6 vs. T3 ($p = 0.0949$). In fact, T7 was mentioned by more than half of the interview participants in our exploratory study. Therefore, this result was consistent and confirmed our findings from our exploratory study. However, interestingly we did not find a statistically significant difference between PHR sharing group and PHR non-sharing group in all factors except T4 ($p = 0.033$, Bonferroni-corrected Chi-Square).

5 DISCUSSIONS AND RECOMMENDATIONS

Reasons for sharing PHR: Our study results demonstrate that "R2: To share healthcare experience and knowledge with others" (71%) and "R4: To know others' opinion/comments about my health issues" (65%) were ranked as the top two reasons for sharing PHR on SNS. These results are quite different from the previous study [26], where they reported two main reasons as follows: "R5: To seek social support from friends and family" (42%) and "R2: To share healthcare experience and knowledge with others" (34%). The differences are likely due to the fact that our study has focused on medical data (e.g., U1: diseases and U4. diagnostic test results) categorized in HIS while previous study mentioned general health information in their studies. Medical data would be more preferred to share with other users because it requires others' expertise, knowledge, and experience. Our finding suggests better opportunities for SNS platforms to support PHR sharing, if they can be designed to help users to connect with others having medical experts or similar health experiences. Therefore, SNS systems can analyze user profiles, posts and connections to recommend new connections and contents using data mining techniques. However, because many users were also concerned about "C6: My PHR can be abused or misused", SNS system designers need to consider some types of privacy-preserving techniques to achieve this goal in a secure manner.

Concerns for sharing PHR: Our confirmatory study results demonstrate that "C9: It seems unneccessary" (58% for PHR sharing group vs. 94% for PHR non-sharing group), "C6: My PHR can be abused or misused" (57% for PHR sharing group vs. 89% for PHR non-sharing group), and "C8: Concerns about security issues" (55% for PHR sharing group vs. 84% for PHR non-sharing group) were ranked as the top three concerns for sharing PHR on SNS.

To mitigate C9 concern, it seems important to understand the true reasons behind this perceived concern. Based on the participants' responses in the exploratory study, we surmise that the true reasons for C9 are related to the other concerns (C1–C8).Probably, the most serious concerns seem to be related to security and privacy concerns such as C6 and C8.

To overcome C6 and C8 concerns, SNS systems need to protect PHR on SNS from unwanted leakage of information using a security mechanism, while providing valuable information. That is, SNS systems should be designed to support a secure access control mechanism, allowing users to share their PHR with preferred recipients only. Even though most SNS systems already support such an access control mechanism, previous studies [10, 16] demonstrated that SNS users are unable to correctly manage their privacy settings. Therefore, a more usable privacy setting for PHR sharing should be provided in practice.

To improve the benefits of PHR sharing on SNS, again, it seems important to connect those users to medical professionals or communities which can provide more useful contents, health advice and guidance, and/or help users feel better as shown in R2 and R4.

Importance of contextual information: Since interviewees showed sensitivity of sharing disease (U1) and diagnostic test results (U4) per each contextual PHR category, our survey was conducted to investigate the effects of underlying characteristics of diseases and test results. Especially, a number of participants reported that

the severity and contagion of the disease had a significant impact on the decision to disclose those on SNS.

This is consistent with the finding that Twitter is typically used to share information about general heath issues rather than information on serious medical conditions, disabilities, and conditions known to bear social stigma [5]. In the case of diagnostic test results, participants' decision were affected by whether the test result was positive or negative (T7). In general, the results that indicate negative health conditions were reluctant to be revealed on SNS. Therefore, a context-aware (e.g., specific PHR category) access control can be designed to effectively reduce the privacy risk of PHR on SNS with dynamically configurable privacy settings.

Limitations: We have a few limitations in our user studies. In the exploratory study, a small number of participants who were recruited from two online health communities may not be sufficient to enumerate all possible codes for our studies. Because online survey questionnaire in the confirmatory study was designed based our codes from the exploratory study, our analysis could be possibly impacted by the bias due to the limited codes. To address this issue, we tested whether code saturation was reached with two separate coders.

Moreover, the participants could have possibly misunderstood some of the questions/terms, even though we explained them in detail. For example, PHR is a broad and general term referring to various health data from hospital appointment to medical test results. To keep the chances of such misunderstanding low and ensure consistency, we had two researchers interviewing together in the exploratory study, and had conducted a pilot study prior to the confirmatory study in order to resolve the ambiguity and misconceptions surrounding the terms and questions.

Since our studies were designed to use self-reported data, our results inherently depend on the honesty and knowledge of participants. To collect only valid responses, we tried to exclude all invalid or unreasonable responses by systemically analyzing the response completion time or response choices made by participants.

6 PRIVACY SETTINGS FOR PHR SHARING

As discussed in Section 5, to address users' security and privacy concerns, we develop a real-time machine learning classifier to automatically determine whether a given post can be shareable with everyone on SNS. This classification method can be potentially used to recommend preferred recipients for health-related posts.

6.1 Data acquisition

To maintain the ecological validity of our experiments, we chose Reddit for collecting real-world health-related posts on SNS because it provides a wide range of several online bulletin boards to discuss specific health topics such as health anxiety, cancer, asthma, cardiology, diabetes, kidney stones and chronic pain.

We crawled 1,040 various health-related posts (about health condition, disease, medical interview, medicine, hospital information, medical knowledge and healthcare-related experience) from Reddit. With the crawled posts, we asked MTurkers in the confirmatory study to label sensitive posts that can be shared with everyone on SNS. Each MTurk participant was given 10 posts, which were randomly selected from the crawled 1,040 posts and asked to label

each of the posts as either *shareable* or *non-shareable*, based on their own opinions. To create the ground-truth dataset, we finally labeled each post again as either *shareable* or *non-shareable* according to the majority of participants' labels. However, final 46 posts were excluded because the related labeling tasks were too quickly completed. Consequently, we used 994 out of 1,040 posts, consisting of 201 shareble and 793 non-shareable posts.

Although Reddit is a platform for open forums, the lack of informed consent for SNS users might arise an ethical question. To avoid this issue, we comply with the guideline [27]. We anonymized the collected data and did not disclose data of individual units in the research results.

6.2 Classification

Our main idea for classification is to use specific keywords as features to identify non-shareable posts. That is, we need to analyze the frequency of words in shareable posts and non-shareable posts, respectively, to construct the keyword list consisting of specific words that frequently occurred in non-shareable posts, while rarely occurred in shareable posts. The overview of our classification process is presented in Figure 7.

In the training phase, we first performed preprocessing that consists of three steps: (1) removing punctuation from a given post, (2) tokenizing words, and (3) eliminating stop words (e.g., 'the', 'me' and 'very'), which could be unnecessary for classification because they are commonly used words in most statements. After performing preprocessing, we can obtain words from each post. To select keyword features, we counted the frequency of words in both shareable and non-shareable posts, respectively. Table 6 shows the top 10 most distinguishable words between shareable and non-shareable posts. Figure 8 depicts the percentage difference of frequently occurred words between two groups. The words occurred least frequently in non-shareable posts relative to shareable posts were: (1) 'back' (0.248%), (2) 'like' (0.194%), and (3) 'months' (0.188%), while the words occurred least frequently in shareable posts relative to non-shareable posts were: (1) 'get' (0.119%), (2) 'ive' (0.119%), and (3) 'infection' (0.113%).

Then, we constructed a classifier using *bag-of-words* model features, where a post is represented in the vector space using the dictionary consisting of a number of distinguishable words. Depending on the presence and the co-occurrence of distinguishable words in the post, we represent each post as a word vector (a sequence of 1 or 0) as shown in Figure 7. In a *bag-of-word* model, the number of features can significantly affect the performance of classifiers. To find the optimal number of features, we analyzed the performance of Random Forest with the different number of features. Figure 9 captures that F-measure of Random Forest rapidly increases toward 1 until the number of features is less than 50 and then tends to be smooth and flat. Therefore, we empirically chose 70 distinguishable words since it overall yielded the best classification performance.

In the testing phase, the preprocessing is the same as the training phase. We constructed a feature vector using the dictionary of distinguishable words (i.e., bag-of-words) from the preprocessed tokens. Finally, this vector is fed into the classifier built in the training phase to determine whether the given post is shareable or non-shareable.
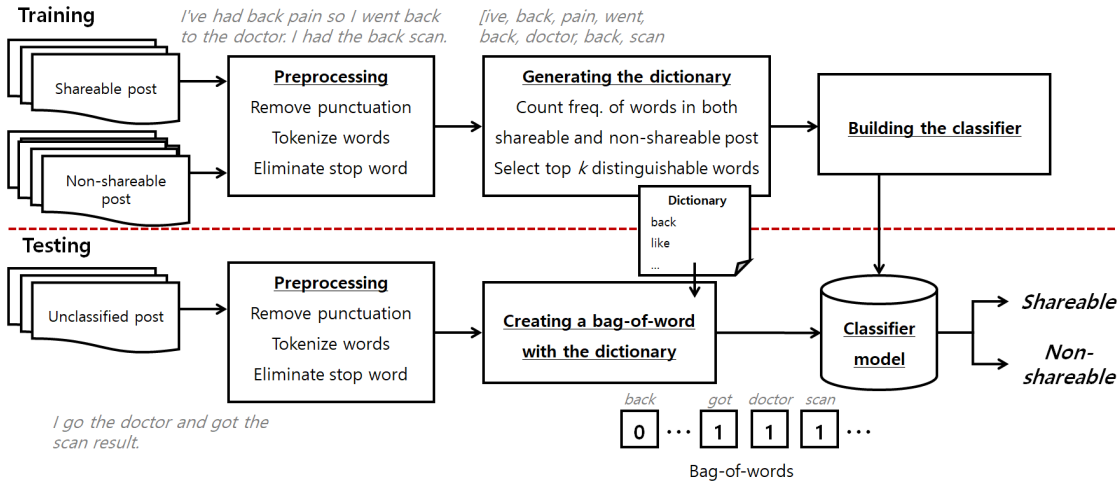
Figure 7: Overview of the proposed classification method for PHR sharing access control.

Table 6: Top 10 most distinguishable words between shareable and non-shareable posts (Diff.(%). is the percentage difference of words occurred in shareable and non-shareable posts).

Rank	Keyword	Share.(%)	Non-share.(%)	Diff.(%)
1	back	0.858%	0.611%	0.248%
2	like	0.798%	0.604%	0.194%
3	months	0.556%	0.368%	0.188%
4	feeling	0.351%	0.166%	0.185%
5	cancer	0.459%	0.301%	0.158%
6	got	0.641%	0.488%	0.153%
7	doctor	0.907%	0.758%	0.149%
8	scan	0.278%	0.150%	0.128%
9	stress	0.181%	0.058%	0.123%
10	get	0.592%	0.712%	0.119%

Figure 9: F-measure of Random Forest with the number of different distinguishable words (features).

automated lexical analysis. To evaluate the performance of classifiers, we use accuracy, precision, recall, and F-measures, where they are defined in Appendix E.

To reduce any bias in a dataset,we used a 10-fold (stratified) cross-validation, where the training samples are divided into 10 equal-sized groups. Table 7 shows the performance of the six classification algorithms. As shown in Table 7, 'Random Forest' (98.7% for F-measure) and 'MLP classifier' (97.7% for F-measure) significantly outperformed over the other classifiers.

Table 7: Performance of classification algorithms.

Classifier	Acc.	Prec.	Rec.	F-meas.
Naive Bayes	82.8%	97.5%	83.8%	90.0%
Logistic Regression	82.7%	98.7%	82.9%	90.0%
Random Forest	**98.0%**	**100.0%**	**97.5%**	**98.7%**
MLP classifier	96.4%	98.2%	97.3%	97.7%
Linear SVC	83.0%	82.7%	82.7%	89.9%
SGD classifier	82.9%	95.6%	85.1%	90.0%

We also measured the running time of the classifiers to show the relative efficiency of the classification methods. The classifiers
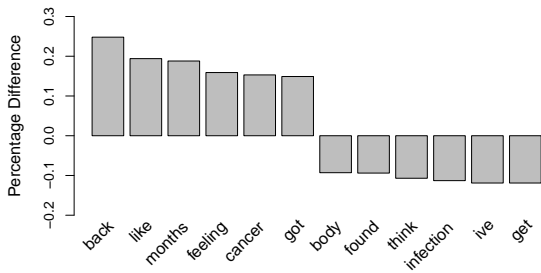

Figure 8: Differences (%) of the frequently occurred words in shareable and non-shareable posts, respectively.

6.3 Evaluation

To show the feasibility of the proposed method and the optimal settings for the classifier, we implemented several classification algorithms (Naive Bayes, Logistic Regression, Random Forest, Multi-Layer Perceptron (MLP), Linear Support Vector Classification (Linear SVC) and Stochastic Gradient Descent (SGD) classifier) using
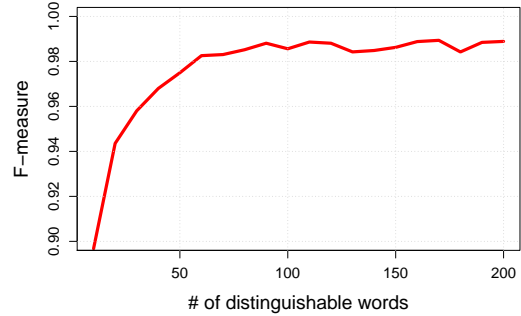
Table 8: Mean time (sec) for running each classification algorithms (μ: mean, σ :standard deviation).

Algorithms	Training		Testing	
	μ	σ	μ	σ
Naive Bayes	0.532	0.009	0.072	0.008
Logistic Regression	0.546	0.013	0.071	0.009
Random Forest	0.677	0.054	0.081	0.009
MLP classifier	1.600	0.152	0.072	0.007
Linear SVC	0.556	0.012	0.072	0.008
SGD classifier	0.552	0.011	0.071	0.008

were implemented using scikit-learn (https://scikit-learn.org/) in Python. The experiments were conducted with a dual core i7-4790 CPU (3.60GHz each core) and 16GB RAM. The running time for training and testing is presented as the mean time using the whole dataset in a group from the 10 independent experiments.

The results for times taken for running classification algorithms are presented in Table 8. The mean training time for each classifier took less than 0.7 second except for 'MLP classifier' which took 2.281 seconds. Moreover, the mean testing times for all classifiers are less than 0.1 seconds. Therefore, our recommendation is to use 'Random Forest' which is not only capable of achieving the highest F-measure (98.7%), but also is up to about 2.4 times more efficient than the second best 'MLP classifier' in the training time.

7 CONCLUSION

This paper investigates risk perceptions associated with PHR sharing on SNS through a qualitative study with 16 participants and a quantitative study with 497 participants.

Our study results revealed that users are concerned with (1) no need to share PHR on SNS, (2) misuse/abuse of shared PHR and (3) security issues, when sharing their PHR on SNS. We also identified that underlying characteristics of PHR (e.g., severity of disease) can affect the decision to disclose those on SNS. To address users' security concerns, we propose a privacy setting method to automatically determine whether a given post can be shared with everyone on SNS. Overall, our implementation using Random Forest achieved an F-measure of 98.7%.

For future work, we plan to develop other recommendation algorithms to effectively connect users to other users who can provide more useful knowledge, contents, and comments about PHR that the users are interested in. It would be also interesting to deploy the proposed privacy setting method on a real-world SNS system and evaluate the usability of the proposed method.

REFERENCES

[1] Marjolijn L Antheunis, Kiek Tates, and Theodoor E Nieboer. 2013. Patients' and health professionals' use of social media in health care: Motives, barriers and expectations. *Patient Education and Counseling* 92, 3 (2013), 426–431.

[2] Mariam Bachiri, Ali Idri, José Luis Fernández-Alemán, and Ambrosio Toval. 2018. Evaluating the Privacy Policies of Mobile Personal Health Records for Pregnancy Monitoring. *Journal of Medical Systems* 42, 8 (2018), 1–14.

[3] Inma Carrión, Jose L. Fernández-Alemán, and Ambrosio Toval. 2011. Usable Privacy and Security in Personal Health Records. In *Proceedings of the 13th IFIP Conference on Human-Computer Interaction.*

[4] Munmun De Choudhury and Sushovan De. 2014. Mental Health Discourse on reddit: Self-Disclosure, Social Support, and Anonymity. In *Proceedings of the Eighth International AAAI Conference on Weblogs and Social Media.*

[5] Munmun De Choudhury, Meredith Ringel Morris, and Ryen W. White. 2014. Seeking and sharing health information online: comparing search engines and social media. In *Proceedings of the 32nd ACM Conference on Human Factors in Computing Systems.*

[6] Robert G. Fichman, Rajiv Kohli, and Ranjani Krishnan. 2011. Editorial Overview—The Role of Information Systems in Healthcare: Current Research and Future Trends. *Information Systems Research* 22, 3 (2011), 419–684.

[7] Markle Foundation. 2003. Markle Connecting for Health. (2003).

[8] Susannah Fox. 2014. The social life of health information. (2014).

[9] Barney G Glaser and Anselm L Strauss. 2017. *Discovery of grounded theory: Strategies for qualitative research.* Routledge.

[10] Andrea Hartzler, Meredith M Skeels, Marlee Mukai, Christopher Powell, Predrag Klasnja, and Wanda Pratt. 2011. Sharing is caring, but not error free: Transparency of granular controls for sharing personal health information in social networks. In *Proceedings of the 14th AMIA Annual Symposium.*

[11] Reinhold Haux. 2006. Health information systems–past, present, future. *International journal of medical informatics* 75, 3-4 (2006), 268–281.

[12] Daniel B. Hoch, Alice J. Watson, Deborah A. Linton, Heather E. Bello, Marco Senelly, Mariola T. Milik, Margaret A. Baim, Kamal Jethwani, Gregory L. Fricchione, Herbert Benson, and Joseph C. Kvedar. 2012. The Feasibility and Impact of Delivering a Mind-Body Intervention in a Virtual World. *PLOS ONE* 7, 3 (2012).

[13] Patrick C. K. Hung. 2005. Towards a Privacy Access Control Model for e-Healthcare Services. In *Proceedings of the 3rd Conference on Privacy, Security and Trust.*

[14] Maritza Johnson, Serge Egelman, and Steven M Bellovin. 2012. Facebook and privacy: it's complicated. In *Proceedings of the 8th Symposium on Usable Privacy and Security.* ACM.

[15] Ruogu Kang, Stephanie Brown, Laura Dabbish, and Sara Kiesler. 2014. Privacy attitudes of Mechanical Turk workers and the US public. In *Proceedings of the 10th USENIX Symposium on Usable Privacy and Security.*

[16] Michelle Madejski, Maritza Johnson, and Steven M. Bellovin. 2012. A study of privacy settings errors in an online social network. In *IEEE International Conference on Pervasive Computing and Communications Workshops.*

[17] Carlos A. Meier, Maria C. Fitzgerald, and Joseph M. Smith. 2013. Health information systems–past, present, future. *International journal of medical informatics* 15, 1 (2013), 359–382.

[18] Mark Merolli, Kathleen Gray, and Fernando Martin-Sanchez. 2013. Health outcomes and related effects of using social media in chronic disease management: a literature review and analysis of affordances. *Journal of biomedical informatics* 46, 6 (2013), 957–969.

[19] Meredith Ringel Morris, Jaime Teevan, and Katrina Panovich. 2010. What do people ask their social networks, and why?: a survey study of status message q&a behavior. In *Proceedings of the 28th ACM Conference on Human Factors in Computing Systems.*

[20] US Department of Health, Human Services, et al. 2003. Health outcomes and related effects of using social media in chronic disease management: a literature review and analysis of affordances. *Washington, DC: Author. Retrieved December* 2, 6 (2003), 2007.

[21] Hyun Jung Oh, Carolyn Lauckner, Jan Boehmer, Ryan Fewins-Bliss, and Kang Li. 2013. Facebooking for health: An examination into the solicitation and effects of health-related social support on social networking sites. *Computers in Human Behavior* 29, 5 (2013), 2072–2080.

[22] Hyun Jung Oh, Carolyn Lauckner, Jan Boehmer, Ryan Fewins-Bliss, and Kang Li. 2013. Facebooking for health: An examination into the solicitation and effects of health-related social support on social networking sites. *Computers in Human Behavior* 29, 5 (2013), 2072–2080.

[23] Ali Sunyaev, Dmitry Chornyi, Christian Mauro, and Helmut Krcmar. 2010. Evaluation Framework for Personal Health Records: Microsoft HealthVault Vs. Google Health. In *Proceedings of the 43rd Hawaii International Conference on System Sciences.*

[24] Paul C Tang, Joan S Ash, David W Bates, J Marc Overhage, and Daniel Z Sands. 2006. Personal health records: definitions, benefits, and strategies for overcoming barriers to adoption. *Journal of the American Medical Informatics Association* 13, 2 (2006), 121–126.

[25] Sadegh Torabi and Konstantin Beznosov. 2013. Privacy Aspects of Health Related Information Sharing in Online Social Networks. In *Proceedings of the 4th USENIX Workshop on Health Information Technologies.*

[26] Sadegh Torabi and Konstantin Beznosov. 2016. Sharing Health Information on Facebook: Practices, Preferences, and Risk Perceptions of North American Users. In *Proceedings of the 12nd USENIX Symposium on Usable Privacy and Security.*

[27] Leanne Townsend and Daniel Z Wallace, Claire. 2016. Social media research: A guide to ethics. *University of Aberdeen* (2016), 1–16.

[28] Katherine White, Achamyeleh Gebremariam, Dana Lewis, Weston Nordgren, James Wedding, Josh Pasek, Ashley Garrity, Emily Hirschfeld, and Joyce M Lee. 2018. Motivations for Participation in an Online Social Media Community for Diabetes. *Journal of Diabetes Science and Technology* 12, 3 (2018), 712–718.

[29] Donna M Zulman, Kim M Nazi, Carolyn L Turvey, Todd H Wagner, Susan S Woods, and Larry C An. 2011. Patient interest in sharing personal health record information: a web-based survey. *Annals of Internal Medicine* 155, 12 (2011), 805–810.

Appendix

A EXAMPLE OF POSTS IN REDDIT

Recently, Reddit community has provided the mental health subreddit (https://www.reddit.com/r/mentalhealth/) that is the central forum to discuss, vent, support and share information about mental health, illness and wellness (see Figure 10).
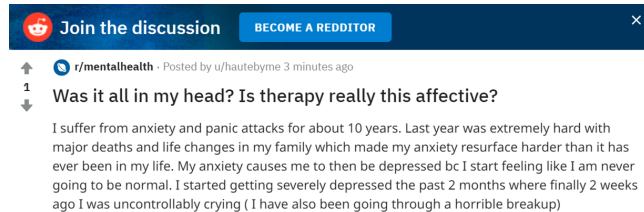

Figure 10: Example of the mental health related posts in Reddit.

B INTERVIEW SURVEY QUESTIONNAIRE

We used the following questions during the interview process.

(1) SNS Usage and Activity
- Which SNS platform do you use the most often?
- How often do you check your SNS feeds?
- How often do you post on your SNS?

(2) PHR Sharing on SNS
- Have you ever shared any Personal Health Records about you or your family members on SNS?
- What are the reasons for sharing or not sharing the Personal Health Records on SNS?
- What are your concerns about sharing the Personal Health Records on SNS?
- Did you have any unpleasant experiences related to sharing information about your Personal Health Records on SNS? What was it?
- Suppose you share your specific Personal Health Records on SNS (e.g., disease, hospital schedule, test result). Whom do you prefer to share those specific health record information ?
- What types of Personal Health Records do you prefer not to share on SNS?
- What are the main factors that influence your decision to disclose any diseases that you or your family members have?
- What are the main factors that influence your decision to disclose any health test results about you or your family members?
- Before or after going to the hospital, do you search for medical or health-related information on the Internet?
- Have you ever posted any health related questions on the Internet?
- If you can get benefit by sharing your or your family members' Personal Health Records on SNS, are you willing to disclose those PHRs on SNS?

(3) Health Care Environment
- What types of insurance do you currently have?
- How long do you usually wait to see a doctor?
- How long does it take to go to the primary and secondary health care center from your home?
- When you received medical services such as medical treatment and examination, how much did you understand explanations from the medical staffs?

(4) Health Status
- How do you rate your overall health status?
- How much do you care about your health?
- How do you rate your health and medical knowledge?
- Do you save your health records?
- If yes, why and how do you save and manage those data?

C ONLINE SURVEY QUESTIONNAIRE

We used the following questions for the online survey.

(1) How often do you check your SNS?
- More than 3 times a day
- 1-2 times a day
- At least once a week
- At least once a month
- Less than once a month
- Do not use any SNS platforms
- Choose not to answer

(2) How often do you post any information on SNS?
- At least once a day
- More than 3 times a week
- At least once a week
- At least once a month
- Less than once a month
- Never
- Choose not to answer

(3) Please select the SNS that you shared about your Personal Health Records (PHRs).
- Facebook
- Instagram
- … 7 additional choices hidden …
- Never shared my PHRs on SNS
- Choose not to answer

Please rate the level of agreement or disagreement on the following statements (each item was rated on a 5-point Likert scale from "Strongly disagree (1)" to "Strongly agree (5)".):

(4) Why did you share your PHRs on SNS? (skip this question if you choose "Never shared my PHRs on SNS" in Q.3)
- To inform my health issues (Likert scale (1–5))
- To share health care experience and knowledge with others (Likert scale (1–5))
- I have been habitually sharing my daily life (Likert scale (1–5))
- To know others' opinion/comments about my health issues (Likert scale (1–5))
- To get comfortable with my health issues (Likert scale (1–5))
- To motivate myself and improve my health status (Likert scale (1–5))
- To manage my PHR history (Likert scale (1–5))
- Other reason (Please specify)

(5) What concerns do you have about sharing PHRs on SNS?
- People would think of me as an attention seeker (Likert scale (1–5))
- People might think me as an unhealthy person (Likert scale (1–5))
- I'm worried about any caused disadvantage or misunderstanding from posting PHRs (Likert scale (1–5))
- My health records can be searched on the Internet (Likert scale (1–5))
- My health records can be abused or misused (e.g., targeted online ads or telemarketing) (Likert scale (1–5))
- People whom I dislike can follow me. I do not want to let them know about my health issues (Likert scale (1–5))

- I'm worried about security and privacy issues from posting (e.g., hacking personal account, phishing attack) (Likert scale (1−5))
- I do not see the need to share my health records on SNS (Likert scale (1−5))
- Other (Please specify)

(6) Please rank 1 to 8 for each of the following health record, where 1 being the most sensitive and 8 being the least sensitive information for disclosure.
- Diseases and illness (1−8)
- Sign and symptoms (1−8)
- Medical interviews (1−8)
- Laboratory and imaging test results (e.g. blood test, urine test, X-ray, MRI) (1−8)
- Prescription drug information (1−8)
- Treatments (e.g. surgery, chemotherapy, medication) (1−8)
- Visited hospitals/clinics (1−8)
- Hospital/clinic appointments (1−8)

(7) Suppose that you are going to share your Persona Health Records on your SNS. Do you think you can disclose each of the following health records to family members (FM), close friends (CF), acquaintances (AC), and SNS friends (SNSF)? Please select all that applies. (select "Yes", "No" for each type of person)
- Diseases and illnesses (Yes/No for (FM), (CF), (AC), (SNSF))
- Sign and symptoms (Yes/No for (FM), (CF), (AC), (SNSF))
- Medical interviews (Yes/No for (FM), (CF), (AC), (SNSF))
- Laboratory and imaging test results (e.g. blood test, urine test, X-ray, MRI) (Yes/No for (FM), (CF), (AC), (SNSF))
- Prescription drug information (Yes/No for (FM), (CF), (AC), (SNSF))
- Treatments (e.g. Surgery, chemotherapy, medication) (Yes/No for (FM), (CF), (AC), (SNSF))
- Visited hospitals/clinics (Yes/No for (FM), (CF), (AC), (SNSF))
- Hospital/clinic appointments (Yes/No for (FM), (CF), (AC), (SNSF))

(8) Suppose that you are diagnosed with a disease. How much does each of the following factors contribute to your concern in deciding whether to share your disease information on SNS? (each item was rated on a 5-point Likert scale from "Not at all concerned" to "Extremely concerned")
- Specific location of diseases or injured body parts (Likert scale 1−5)
- Stages of disease progression (early detection) (Likert scale 1−5)
- Communicable diseases (Likert scale 1−5)
- Rareness of diseases (Likert scale 1−5)
- Severity of diseases (Likert scale 1−5)
- Treatment periods (Likert scale 1−5)
- Types of diseases (Likert scale 1−5)

(9) Suppose you have a diagnostic test at the hospital and received the results. How much does each of the following factors contribute to your concern in deciding whether you share your test results on SNS? (Diagnostic test: CT scans, Ultrasound, Blood test, Genetic test, Endoscopy, etc.) (each item was rated on a 5-point likert scale from "Not at all concerned" to "Extremely concerned")
- Diagnosed parts of body (Likert scale 1−5)
- Purpose of diagnostic tests - general health checkup (Likert scale 1−5)
- Purpose of diagnostic tests - sign and symptoms of Health problem (Likert scale 1−5)
- Purpose of diagnostic tests - post-treatment monitoring (Likert scale 1−5)
- Purpose of diagnostic tests - accidents and injuries (Likert scale 1−5)
- The amount and detail of diagnostic test results (Likert scale 1−5)
- Positive/Negative results of diagnostic test (Likert scale 1−5)

- Type of Examination test (e.g. Blood test, X-Ray, MRI) (Likert scale 1−5)

(10) How long do you wait to see a doctor on average?
- On the same day
- 1-3 days
- A week
- 2-3 weeks
- More than a month
- I do not know

(11) Which health insurance do you currently have? Please select all that applies.
- Public insurance
- health insurance offered at work
- Individual health insurance
- No insurance
- I do not know
- Others (Please specify)

(12) What is your gender?
- Male
- Female
- Decline to answer

(13) What is your age?

(14) What is the highest level of education you have completed?
- Less than High school
- High school diploma
- Diploma (Post-secondary vocational, technical education)
- Some college, university courses
- Undergraduate University degree (e.g., B.S., B.A.)
- Graduate University degree (e.g., M.S., M.A., Ph.D., M.D., J.D.)
- Other (Please specify)
- Decline to answer

(15) What is the field of your occupation?
- Assistant, supporter
- Agricultural, Forestry and Fishery
- Art/Cultural (e.g., Artists, Actor, Dancer)
- Education (e.g., Professor, Teacher, Educator)
- Engineer (e.g., Mining Engineer)
- Financial (e.g., Accountant, banker)
- Health/medical (e.g., Doctor, Pharmacist)
- Legal (e.g., Lawyer, Judge)
- Manager (e.g., Sales Manager)
- Religious (e.g., Clergy, Nun)
- Sales/Marketing (e.g., Dealer, Distributor)
- Scientist/Researcher (e.g., Mathematician)
- Service (e.g., Clerk, Server, Guide, Cashiers)
- Sports/Tourism (e.g., Athletes, Coaches
- Student
- Skilled worker (e.g., Welder, Plumber)
- Writing (e.g., Author, Journalists)
- Unemployed
- Other (Please specify)
- Decline to answer

D   DEMOGRAPHICS FOR THE SURVEY

Participants' demographic information for the online confirmatory study is presented in Table 9.

Table 9: The demographics of the online survey (N = 497).

Gender		
Male	239	(48.09%)
Female	**255**	**(51.31%)**
No answer	3	(0.60%)
Age group		
18–30	154	(30.99%)
31–40	**176**	**(35.41%)**
41–50	89	(17.91%)
50+	73	(14.69%)
No answer	3	(1.01%)
Education		
Less than high school	3	(0.60%)
High school	51	(10.26%)
Diploma (post-secondary courses)	12	(2.41%)
Some college / university courses	133	(26.76%)
Undergraduate University (Bachelor's)	**199**	**(40.04%)**
Graduate University (Master's/PhD)	98	(19.72%)
Other	0	(0.00%)
No answer	1	(0.20%)
Occupation		
Assistant, supporter	34	(6.84%)
Agricultural, Forestry and Fishery	8	(1.61%)
Art/Cultural (e.g., Artists, Actor, Dancer)	15	(3.02%)
Education (e.g., Professor, Teacher)	58	(11.67%)
Engineer (e.g., Mining Engineer)	30	(6.04%)
Financial (e.g., Accountant, banker)	56	(11.27%)
Health/medical (e.g., Doctor, Pharmacist)	28	(5.63%)
Legal (e.g., Lawyer, Judge)	9	(1.81%)
Manager (e.g., Sales Manager)	**62**	**(12.47%)**
Religious (e.g., Clergy, Nun)	1	(0.20%)
Sales/Marketing (e.g., Dealer, Distributor)	41	(8.25%)
Scientist/Researcher (e.g., Mathematician)	16	(3.22%)
Service (e.g., Clerk, Receptionists, Guide)	40	(8.05%)
Sports/Tourism (e.g., Athletes, Instructor)	4	(0.80%)
Student	10	(2.01%)
Skilled worker (e.g., Welder, Plumber)	19	(3.82%)
Writing (e.g., Author, Journalists)	16	(3.22%)
Unemployed	41	(8.25%)
Decline to answer	9	(1.81%)

E   PERFORMANCE METRICS

To evaluate the performance of classifiers, we used the following metrics.

- **Accuracy**: the proportion of correctly classified posts;
- **Precision**: the proportion of posts classified as shareable that actually are shareable;
- **Recall**: the proportion of shareable posts that were accurately classified;
- **F-measure**: the harmonic mean of *precision* and *recall*.