Applying Artificial Intelligence in Healthcare Social Networks to Identity Critical Issues in Patients' Posts

Giacomo Fiumara[1], Antonio Celesti[2,4], Antonino Galletta[2,3],
Lorenzo Carnevale[2,3] and Massimo Villari[2,3]

[1]*Department of Mathematics and Computer Science, Physics and Earth Sciences, University of Messina, Contrada di Dio 1 - 98166, Messina, Italy*

[2]*Department of Engineering, University of Messina, Contrada di Dio 1 - 98166, Messina, Italy*

[3]*IRCCS Centro Neurolesi "Bonino Pulejo", Contrada Casazza, SS113 - 98124, Messina, Italy*

[4]*Alma Digit S.R.L. Research Labs, Contrada di Dio 1 - 98166, Messina, Italy*

Keywords: Healthcare, Social Network, Artificial Intelligence, Machine Learning.

Abstract: Nowadays, the possibility of using social media in the healthcare field is attracting the attention of clinical professionals and of the whole healthcare industry. In this panorama, many Healthcare Social Networking (HSN) platforms are emerging with the purpose to enhance patient care and education. However, they also present potential risks for patients due to the possible distribution of poor-quality or wrong information. On one hand doctors want to promote the exchange of information among patients about a specific disease, but on the other hand they do not have the time to read patients' posts and moderate them when required. In this paper, we propose an Artificial Intelligence (AI) approach based on a combination of stemming, lemmatization and Machine Learnign (ML) algorithms that allows to automatically analyse the patients' posts of a HSN platform and identify possible critical issues so as to enable doctors to intervene when required. In particular, after a discussion of advantages and disadvantages of using a HSN platform, we discuss in detail an architecure that allows to analyse big data consisting of patients' posts. In the end, real case studies are discussed highlighting future challenges.

1 INTRODUCTION

Nowadays, there is an increasing interest of clinical operators in social media, big data analytics and Cloud computing. All over the world, the number of investments in Information and Communication Technology for health and wellbeing (eHealth) is rapidly increasing. Global eHealth market is expected to reach USD 308.0 billion by 2022, according to a new report by Grand View Research Inc. In particular, the transition of the healthcare industry into digital healthcare system for management and analysis of patients' health is expected to be the most vital driver of the market (www.grandviewresearch.com, 2016). The European Commission's eHealth Action Plan 2012-2020 has already provided a roadmap to empower patients and healthcare workers, to link up devices and technologies, and to invest in research towards the personalised medicine of the future (ec.europa.eu, 2012). In February 2017, the European Commission set up an internal task force bringing together technology and health policy makers to examine EU policy actions to ensure the transformation of health care into a Digital Single Market (DSM) bringing benefits for people, health care systems and the economy (ec.europa.eu, 2014). Guaranteeing access to high-quality health care is a key objective of social protection systems in European countries and it represents the second largest social expenditure item after pensions.

In this panorama, social media represent a tempting opportunity for healthcare operators for improving the patients' well-being. Many social media tools are available over the Internet such as social networking, professional networking, media sharing, content production including blogs (e.g., Tumblr) and microblogs (e.g., Twitter), knowledge/information aggregation (e.g., wikipedia), virtual reality and gaming environments (e.g., second life). In particular, many Healthcare Social Networking (HSN) platforms have

Fiumara G., Celesti A., Galletta A., Carnevale L. and Villari M.
Applying Artificial Intelligence in Healthcare Social Networks to Identity Critical Issues in PatientsâĂŹ Posts.
DOI: 10.5220/0006750606800687
In *Proceedings of the 11th International Joint Conference on Biomedical Engineering Systems and Technologies (HEALTHINF 2018)*, pages 680-687
ISBN: 978-989-758-281-3

emerged with the purpose to enhance patient care and education. Popular HSB platforms include Sermo, Doximity, Orthomind, QuantiaMD, WeMedUp, Digital Healthcare and so on. However, these social networks require the massive action of medical professional acting as moderators. In fact, healthcare social networks present potential risks for patients due to the possible distribution of poor-quality or wrong information. On one hand clinical operators want to promote the exchange of information among patients about a specific disease, but on the other hand they do not have the time to read patients' posts and moderate them when required.

The main contribution of this paper is to propose an Artificial Intelligence (AI) technique aimed at enhancing the well-being of patients participating in a HSN platform. The basic idea is to allow an automatic analysis of patients' posts in order to identify possible critical issues so as to enable doctors to intervene when required. In particular, our architecture includes a Patients' Posts Analysis System (PaPAS) acting as a filter that is connected with a Complex Event Processing (CEP) that sends warning and alert messages with different level of risks to the medical personnel who can reply to critical patients' posts. In particular, in this paper, we focus on the PaPAS specifically analysing the adopted AI approach based on a combination of stemming, lemmatization and Machine Learning (ML) algorithms among others.

The remainder of the paper is organized as follows. Section 2 provides a brief overview of the major recent initiatives in the fields of AI and social media for eHealth. The advantages and disadvantages of adopting a HSN platform along with our reference architecture are discussed in Section 3. Section 4 focuses on the Patients' Posts Analysis System specifically focusing on the adopted AI technique. A discussion of possible application scenarios and future challenges is provided in Section 5. Section 6 concludes the paper along with lights to the future.

2 RELATED WORK

Social media and healthcare quality improvement is an emerging topic (Ranney and Genes, 2016). Recently many scientific works have been proposed facing different aspects of social media for healthcare purposes.

Regarding the social implication of such systems, the benefits, best practices, risks and ethical issues of applying social media to healthcare professionals are discussed in (Lee Ventola, 2014), (Hors-Fraile et al., 2016), (Pinho-Costa et al., 2016), (Aboelmaged et al., 2016), (Abbas et al., 2016). Social media can be used to enhance professional networking and education, organizational promotion, patient care, patient education, and public health programs, but they can also present several potential risks for patients regarding the distribution of poor-quality information, damage to professional image, breaches of patient privacy, violation of personal-professional boundaries, and licensing or legal issues. However, social media are also changing the healthcare industry (Opel, 2016) and marketing (Malvey et al., 2015), (Koumpouros et al., 2015). Evolution of social media in scientific research in the domain of ICT and healthcare professionals in Saudi Arabia Universities is discussed in (Abdullatif et al., 2017), whereas a similar study performed on young people in Russia is available in (Bugrezova, 2016). Studies on the effectiveness of social media data in healthcare communication involving both medical personnel and patients is proposed in (Saqib Nawaz et al., 2017a), (Huby and Smith, 2016), (Smailhodzic et al., 2016), (Benetoli et al., 2017). The role of social media in menopausal healthcare is discussed in (Short, 2017). There is also a strong correlation among online data coming from search engines and social media in the healthcare domain. In this regard, in (Saqib Nawaz et al., 2017b) it is discussed an approach for collecting twitter data by exploring contextual information gleaned from Google search queries logs.

Due to the huge amount of information to be analysed and processed, many scientific works have faced the need of decisional support systems. In this regard, a recommendation approach helping social media users to identify topics of interests is discussed in (Li and Zaman, 2014). Such an approach was also used for the assessment of user similarity in heterogeneous network with the purpose to look for people that can give informational and emotional support in a more efficient way is discussed in (Jiang and Yang, 2017). Another user similarity study in healthcare social media using content similarity and structural similarity is presented in (Jiang and Yang, 2015). A Study on healthcare social media aimed at underserved communities based on a mobile decision support system (MDSS) providing information dissemination is discussed in (Miah et al., 2017). The need of pervasive decision support systems in healthcare using intelligent robots in social media is discussed in (Samad-Soltani et al., 2017).

All aforementioned scientific works consider the benefits of using healthcare social media, but in many cases require a strong interaction of the medical personnel. For the best of our knowledge a system that filter for doctors only the patients' posts that raises

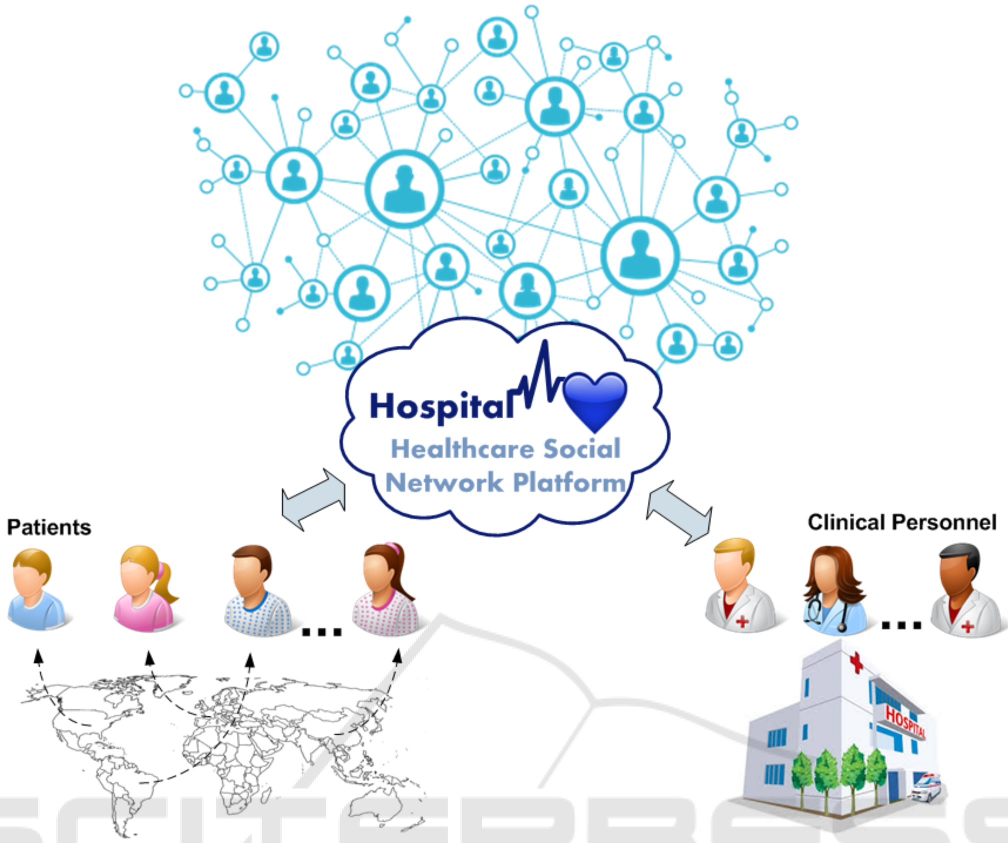

Figure 1: General healthcare social network scenario including both clinical personnel and patients.

critical issues has not been proposed yet. In this paper, we aim to overcome such a gap.

3 MOTIVATION

Social media applied in an healthcare context represent a tempting opportunity to improve the patients' well-being promoting patient care and education. Figure 1 shows a general healthcare social network scenario including both clinical personnel and patients who interact by means of a HSN platform available over the internet. Commonly, the major HSB platforms require the massive action of medical professional who reply to patient's queries also acting as moderators on specific topics when it is required.

Healthcare social networks present several benefits including:

- promoting networking and information exchange enabling self-education among patients about particular diseases.
- sharing patients' experiences that can be helpful for other ones;
- supporting the treatment process;
- reducing the patient's stress when he/she is waiting for a diagnosis or when he/she discovers to be affected of a particular disease;
- promoting information gathering and prevention campaign regarding specific diseases;
- optimizing the work of the clinical personnel who interact with patients skilled on their diseases;
- promoting knowledge management;
- promoting research and monitoring activities.

All the aforementioned benefits can potentially improve the whole world healthcare education system. On the other hand healthcare social networks present several disadvantages including:

- possible distribution of poor-quality or wrong information among patients;
- the need of qualified medical personnel who promptly read patients' posts and who reply them;
- often the medical personnel do not have the time to read patients' posts and to reply them;
- the medical personnel do not want the responsibility of the consequences on patients (worsening, risk of death or death) when they do not reply in time.

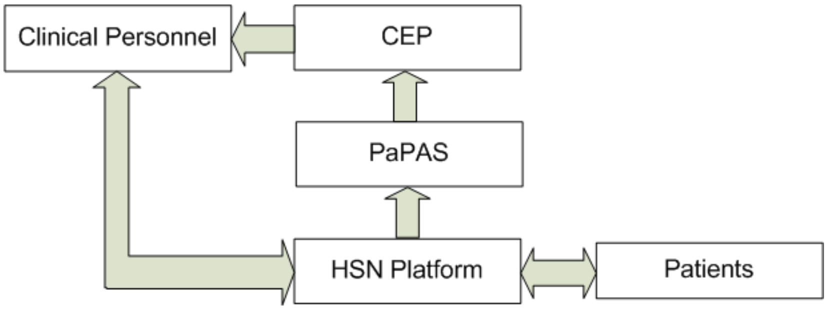

Figure 2: Patients' Posts moderation architecture.

- possible legal issues for the medical personnel;
- risks for the reputation of the medical personnel.

A possible solution comes from the automatic analysis of patients' posts by means of Artificial Intelligence (AI) techniques. In fact AI opens towards various application scenarios of patients' posts analytics including the identification of critical posts that can lead toward dangerous situations for patients themselves. In order to address such an issue, in this paper, we propose a Patients' Posts Moderator (PPM) architecture whose basic flowchart is shown in Figure 2. Both patients and medical personnel interact by means of a HSN platform. A Patients' Posts Analysis System (PaPAS) works as batch process that continuously analyses patients' posts of a HSN platform. When a critical issue is detected, it generates an event that is caught by a Complex Event Processing (CEP) component that elaborates it and sends an alert message to the interested medical personnel who can intervene on the HSN platform, replying to critical patients' discussion groups and/or triggering medical interventions (doctors can directly contact the patient or send ambulance with a medical equipment if required.).

4 APPLYING ARTIFICIAL INTELLIGENCE IN HSN PLATFORMS

In this Section, we specifically discuss an AI approach on which PaPAS can be based. The main purpose of PaPAS is to analyze patients' posts and the evaluation of possible critical issues that may trigger clinicians' intervention. Figure 3 shows the PaPAS architecture. It includes the following components: i) *Extractor*, whose role is to extract patients' content from the HSN platform; ii) *Selector*, whose role is to select relevant keywords; iii) *Rank Generator*, whose role is to rank selected keywords; iv) *Categorisator*, whose role is to categorise the various levels of seriousness; v) *Classificator*, whose role is to classify patients' posts according to different categories; and vi) *Evaluator*, whose role is to assess results' quality. A detailed description of each components is provided in the following Subsections.

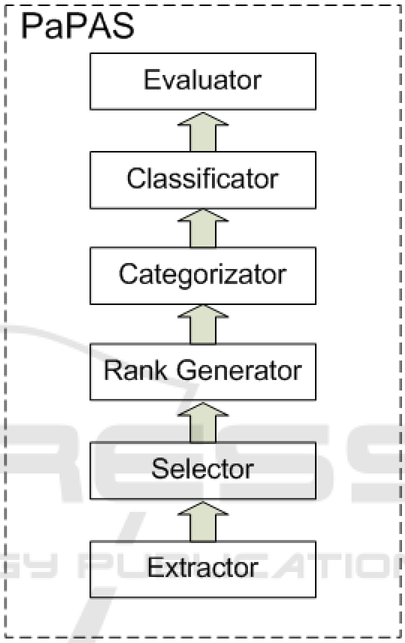

Figure 3: PaPAS architecture.

4.1 Extractor

The first component of our system deals with the extraction of patients' contents from HSN platform. The accomplishment of this task is not unique because it greatly depends on the architecture of the considered social network. In general, social networks provide APIs that help in the automatic extraction of specific information of interest. In some situations these APIs may not be publicly available, or some information (e.g. patients' posts) cannot be extracted. In these cases special-purpose pieces of software must be used, generically referred to as *wrappers*. Thor-

Table 1: Features measured in a post.

Feature	Definition
PostLength	The number of words
Pos	NumOfPos/PostLength, where NumOfPos is the number of positive words/emoticons
Neg	NumOfNeg/PostLength, where NumOfNeg is the number of negative words/emoticons
Name	NumOfName/PostLength, where NumOfName is the number of names mentioned
Slang	NumOfSlang/PostLength, where NumOfSlang is the number of Internet slangs
PosStrength	Positive sentiment strength
NegStrength	Negative sentiment strength
PosVsNeg	(NumOfPos+1)/(NumOfNeg+1)
PosVsNegStrength	PosStrength / NegStrength
Sentence	The number of sentences
AvgWordLen	The average length of words
QuestionMarks	The number of question marks
Exclamation	The number of exclamations

ough descriptions can be found in (Aggarwal and Zhai, 2012; Liu, 2006).

4.2 Selector

After the extraction of patients' posts, the next step consists in the selection of the relevant keywords that may trigger clinicians' intervention. At the beginning, stop-words are removed. Then, the Natural Language Toolkit (NLTK)[1] stemming and lemmatization algorithms are employed in order to reduce inflectional forms and avoid the various syntactical forms a word may have. Usually, the occurrence of syntactic variations of the same root form is less frequent in social media posts with respect to long texts. Nevertheless, the values of recall (see Subsection 4.6) is negatively influenced by the presence of duplicates and/or syntactic variations. Next, the identification of relevant keywords takes place. As usual, a Term Frequency - Inverse Document Frequency (TF-IDF) approach has been used (see for example (Liu, 2006)). Here, we only recall that the importance of a term in a text is evaluated according to the frequency with which it appears across multiple texts. Therefore, the Term Frequency tf_i of the i-th term in a text of n terms is given by

$$tf_i = \frac{f_i}{max\{f_1, f_2, \ldots, f_n\}}$$

The Inverse Document Frequency idf_i of the i-th term within N posts in which the i-th term appears at least once in df_i posts is defined as

$$idf_i = log\frac{N}{df_i}$$

The resulting TF-IDF term weight is given by:

$$w_i = tf_i \cdot idf_i$$

4.3 Rank Generator

Beside the relevant keywords, also a sentiment analysis of the posts is necessary in order to evaluate their seriousness and therefore trigger the clinicians's intervention. This is important in order to disambiguate, namely to discern situations in which a negative term is used within a positive context (e.g., 'a friend of mine suffered from a stroke but after a while has recovered') from situations in which a negative term is used within a negative context (e.g., 'a friend of mine is having a stroke'). To this aim, we resorted to rely on

[1] www.nltk.org

the work of (Qiu et al., 2011), in which the strength of emotions in a post are measured as in (Thelwall et al., 2010). All measures, as in (Qiu et al., 2011), are summarized in Table 1. Here we want to underline the importance of some measures such as, for example, PosVsNeg (the ratio of positive over negative strength) which describes the overall tone of a post.

4.4 Categorizator

A big question is: how serious is the message contained in a user's post? Another interesting question is: does it requires a clinician's immediate intervention? It would be highly desirable that only really serious posts should trigger an alert that may cause the intervention of a clinician. To this aim, the number of the levels of seriousness is a critical feature in our platform. In fact, a small number of levels (e.g., only two) should increase the percentage of interventions while a relatively high number of levels (e.g., four or five) should cause underestimates of the gravity of a situation. Moreover, it has important consequences on the Machine Learning (ML) algorithm we adopted (see Subsection 4.5), because only supervised algorithms allow to choose in advance the number of classes in which clusterize patients' posts. Therefore, we decided to adopt three levels of seriousness, that are: normal, warning, and critical.

4.5 Classificator

After having chosen the number of levels of seriousness, we classify patients' posts according to the categories described in the Subsection 4.4. We devised to adopt the Naïve Bayes Classifier (NBC). It is a simple yet powerful ML algorithm which embodies some desirable features such as: i) it is extremely fast for both training and prediction, ii) it produces simple probabilistic prediction, and iii) it performs very well in similar situations. As to the latter feature, in (Jain and Kumar, 2015) is reported a similar study in which posts extracted from Tweeter related to influenza pandemic are classified using various MLAs. From their results appear that (NBC) outperforms other MLAs in terms of precision, recall and F-Measure (see Subsection 4.6).

4.6 Evaluator

During the training phase, some measures are necessary in order to assess the quality of the results. We first defined the number of *true positives* (TP) as the number of posts correctly labeled as belonging to the positive class, the number of *false positives* (FP) as the number of posts incorrectly labeled as belonging to the postitive class. We also defined the number of *true negatives* (TN) as the number of posts correctly labeled as belonging to the negative class, and the number of *false negatives* (FN) as the number of posts incorrectly labeled as belonging to the negative class. Having this in mind, as customary, we introduced the precision p, the recall r and the F-Measure F as

$$p = \frac{TP}{TP+FP}$$

$$r = \frac{TP}{TP+FN}$$

$$F = \frac{2pr}{p+r}$$

5 DISCUSSION AND FUTURE CHALLENGES

AI-based PaPAS also enables other applications scenarios considering a combination with emerging ICT technologies such as Cloud computing, Edge computing, Cyber-Physical System, Internet of Things (IoT) and Big Data analytics technologies.

In fact, AI applied to HSNs can trigger delivery of various kinds of Cloud-based healthcare services and applications over telecommunication networks and the Internet aimed at providing assistance to patients when warning and critical levels of seriousness occur. The benefit of adopting AI in telemedicine is twofold: on one hand it can push down clinical costs and on the other hand it can improve the quality of life of both patients and their families. Telemedicine solutions that can be triggered by AI mechanisms can be aimed at tele-nursing, tele-rehabilitation, tele-dialog, tele-monitoring, tele-analysis, tele-pharmacy, tele-trauma care, tele-psychiatry, tele-pathology, etc.

Feedback provided by AI applied to HSN can also control physical processes of patients considering Cyber-Physical System that on one side is connected with an Healthcare Cloud provider (e.g., managed by a Hospital or clinical centre) and on the other side is connected with the patient by means of a series of medical IoT devices deployed on him/her by means of a personal body network including medical sensors and actuators and/or deployed in medical devices placed in the patient's home on which the patient is attached or monitoring and controlling the surrounding environment.

Furthermore, AI applied to HSN opens towards various scenarios of big data analytics. In fact, it can allows researchers to study and understand the origins, causes and diffusion over a wide geographical area of a particular disease besides understanding their social implications.

6 CONCLUSION AND FUTURE WORK

In this paper, we proposed an architecture aimed at supporting the medical personnel in monitoring and moderating patients of participating to HSN platform.

In particular, we focused on a PaPAS architecture that implements the AI logic by means of a combination of stemming, lemmatization and Machine Learning (ML) algorithms among others. Specifically, the aim of such a system is to enhance the well-being of patients participating to the HSN platform. Basically, PaPAS analyses patients' posts in order to detect possible critical issues considering three levels of seriousness, that are: normal, warning, and critical. If the content of a post crosses the threshold of criticality, the clinical personnel may promptly intervene.

Although our work is in a preliminary state, some experiments have been carried out that demonstrate effectiveness of the considered algorithms considered alone. In particular, it was demonstrated that the adopted ML algorithm (Naïve Bayes Classifier) is fast and reliable enough to allow real-time applications as in critical environment.

In future work, we plan to analyse the performance of the whole PaPAS considering a concrete dataset including patients' posts coming from a HSN and to make a comparison with existing solutions based on Deep Learning algorithms such as *word-embeddings* and *n-grams*. Considering the huge amount of posts that our system must be able to analyse in a real HSN environment, in our implementation we plan to consider big data analytics solution based on Apache Hadoop and Spark. Future work will include the application of PaPAS in different healthcare contexts.

REFERENCES

Abbas, A., Ali, M., Shahid Khan, M., and Khan, S. (2016). Personalized healthcare cloud services for disease risk assessment and wellness management using social media. *Pervasive and Mobile Computing*, 28:81–99.

Abdullatif, A., Shahzad, B., and Hussain, A. (2017). Evolution of social media in scientific research: A case of technology and healthcare professionals in saudi universities. *Journal of Medical Imaging and Health Informatics*, 7(6):1461–1468.

Aboelmaged, M., Thomas, S., and Elsheikh, S. (2016). *Trends of social media applications in healthcare: A managerial perspective*.

Aggarwal, C. C. and Zhai, C. (2012). *Mining text data*. Springer Science & Business Media.

Benetoli, A., Chen, T., and Aslani, P. (2017). How patients' use of social media impacts their interactions with healthcare professionals. *Patient Education and Counseling*.

Bugrezova, E. (2016). The social media contribution into healthcare practices among russian young people. *Ekonomicheskaya Sotsiologiya*, 17(5):103–129.

ec.europa.eu (2012). European commission. ehealth action plan 2012-2020: Innovative healthcare for the 21st century [accessed on november 12th, 2017]. https://ec.europa.eu/digital-single-market/en/news/ehealth-action-plan-2012-2020-innovative-healthcare-21st-century.

ec.europa.eu (2014). European commission. transformation of health and care in the digital single market [accessed on november 12th, 2017]. https://ec.europa.eu/digital-single-market/en/policies/ehealth.

Hors-Fraile, S., Atique, S., Mayer, M., Denecke, K., Merolli, M., and Househ, M. (2016). The unintended consequences of social media in healthcare: New problems and new solutions. *Yearbook of medical informatics*, (1):47–52.

Huby, K. and Smith, J. (2016). Relevance of social media to nurses and healthcare: 'to tweet or not to tweet'. *Evidence-Based Nursing*, 19(4):105–106.

Jain, V. K. and Kumar, S. (2015). An effective approach to track levels of influenza-a (h1n1) pandemic in india using twitter. *Procedia Computer Science*, 70:801–807.

Jiang, L. and Yang, C. (2015). Determining user similarity in healthcare social media using content similarity and structural similarity. *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 9105:216–226.

Jiang, L. and Yang, C. (2017). User recommendation in healthcare social media by assessing user similarity in heterogeneous network. *Artificial Intelligence in Medicine*, 81:63–77.

Koumpouros, Y., Toulias, T., and Koumpouros, N. (2015). The importance of patient engagement and the use of social media marketing in healthcare. *Technology and Health Care*, 23(4):495–507.

Lee Ventola, C. (2014). Social media and health care professionals: Benefits, risks, and best practices. *P and T*, 39(7):491–499.

Li, J. and Zaman, N. (2014). Personalized healthcare recommender based on social media. pages 993–1000.

Liu, B. (2006). *Web Data Mining: Exploring Hyperlinks, Contents, and Usage Data (Data-Centric Systems and Applications)*. Springer-Verlag New York, Inc., Secaucus, NJ, USA.

Malvey, D., Beardsley, A., Howse, P., and Ruff, S. (2015). *Healthcare marketing and social media*.

Miah, S., Hasan, N., Hasan, R., and Gammack, J. (2017). Healthcare support for underserved communities using a mobile social media platform. *Information Systems*, 66:1–12.

Opel, D. (2016). Ethical information flows: Working with/against the healthcare industry's fascination with social media.

Pinho-Costa, L., Yakubu, K., Hoedebecke, K., Laranjo, L., Reichel, C., Colon-Gonzalez, M., Neves, A., and Errami, H. (2016). Healthcare hashtag index development: Identifying global impact in social media. *Journal of Biomedical Informatics*, 63:390–399.

Qiu, B., Zhao, K., Mitra, P., Wu, D., Caragea, C., Yen, J., Greer, G. E., and Portier, K. (2011). Get online support, feel better – sentiment analysis and dynamics in an online cancer survivor community. *2011 IEEE Third Int'l Conference on Privacy, Security, Risk and Trust and 2011 IEEE Third Int'l Conference on Social Computing*, pages 274–281.

Ranney, M. and Genes, N. (2016). Social media and healthcare quality improvement: A nascent field. *BMJ Quality and Safety*, 25(6):389–391.

Samad-Soltani, T., Rezaei-Hachesu, P., and Ghazisaeedi, M. (2017). Pervasive decision support systems in healthcare using intelligent robots in social media. *Iranian Journal of Public Health*, 46(1):148–149.

Saqib Nawaz, M., Bilal, M., Ikram Ullah Lali, M., Mustafa, R., Aslam, W., and Jajja, S. (2017a). Effectiveness of social media data in healthcare communication. *Journal of Medical Imaging and Health Informatics*, 7(6):1365–1371.

Saqib Nawaz, M., Mustafa, R., and Lali, M. (2017b). *Role of online data from search engine and social media in healthcare informatics*.

Short, H. (2017). The role of social media in menopausal healthcare. *Post Reproductive Health*, 23(1):4–5.

Smailhodzic, E., Hooijsma, W., Boonstra, A., and Langley, D. (2016). Social media use in healthcare: A systematic review of effects on patients and on their relationship with healthcare professionals. *BMC Health Services Research*, 16(1).

Thelwall, M., Buckley, K., Paltoglou, G., Cai, D., and Kappas, A. (2010). Sentiment in short strength detection informal text. *J. Am. Soc. Inf. Sci. Technol.*, 61(12):2544–2558.

www.grandviewresearch.com (2016). Grad view research. ehealth market projected to reach usd 308.0 billion by 2022 [accessed on november 12th, 2017]. http://www.grandviewresearch.com/press-release/global-e-health-market.