Sentiment analysis on social media for identifying public awareness of type 2 diabetes

Yuezhong Liu[1], Rudi Stouffs[2] and Yin Leng Theng[3]
[1, 3] Nanyang Technological University, Singapore
{yz.liu[1], TYLTheng[3]}@ntu.edu.sg
[2] National University of Singapore, Singapore
stouffs[2]@nus.edu.sg

Abstract: Many studies have investigated how Type 2 Diabetes Mellitus (T2DM) is impacted by environmental characteristics. However, it is unclear whether the public is aware of this relationship. This paper aims to infer public awareness based on sentiment analysis of environmental characteristics from social media (news) as context-based clues for supporting planners and designers to achieve a health-supportive environment. We apply natural language processing (NLP) techniques to past years' news (from 1992 to 2020) to understand the public awareness on the relationship between the prevalence of T2DM and environmental characteristics. The results not only verify that the public aware little that prevalence of T2DM correlates with environmental characteristics in Singapore (1% news related to both diabetes and environment), but also present a hypothetical urban design study to demonstrate how this approach can be adapted to identify the health-related urban issues at the early design stage. This study can offer urban planners and designers an innovative design perspective to incorporate residents' health into the design process.

Keywords: Urban design; Sentiment Analysis, Type 2 Diabetes Mellitus, Health-supportive Environment.

1. Introduction

With the increasing prevalence of Type 2 Diabetes Mellitus (T2DM) among Singapore residents, there is a need to better understand the factors that influence the prevalence of T2DM in the design of a healthy urban environment. The influence of the neighbourhood and environmental characteristics on health including T2DM is increasingly recognised and studied in recent years (Renalds, Smith, & Hale, 2010). Many studies have investigated how the prevalence of T2DM is impacted by environmental characteristics from urban design solutions. However, it is unclear whether the public acknowledges this relationship between environmental characteristics and T2DM. We propose to deduce public awareness of this kind of relationship from sentiment analysis in news as a scalable approach.

In particular, we applied natural language processing (NLP) techniques to past years' news (opinions from 1992 to 2020) in Singapore to understand the influence of lifestyle choices as impacted by urban design activities on the prevalence of T2DM. By analysing the words used in the opinions and editorials,

Imaginable Futures: Design Thinking, and the Scientific Method. 54th International Conference of the Architectural Science Association 2020, Ali Ghaffarianhoseini, et al (eds), pp. 956–965. © 2020 and published by the Architectural Science Association (ANZAScA).

this research extracts the good and bad comments associated with urban design related news associated with the public attitudes and lifestyles as impacted by design activities and correlates these with the annual prevalence of T2DM from public datasets (MOH, 2018). The proposed method includes four steps: 1) Integrate Opinions and Comments News Corpus: Integrating the opinion news from Factiva (https://www.dowjones.com/products/factiva/) and comments extracted by a news web- scraper in JavaScript Object Notation (JSON) format. From these opinions and comments, we identify sentences that describe urban issues, such as food choices, open spaces, mobility or accessibility of transportation. 2) Apply NLP analysis to simplify the news and reviews, and extract and clean root words by using the Natural Language Processing Toolkit (NLTK) tokenizer. Public awareness can be revealed from the sentiment analysis of each news and review in time series. The multi-class categorization (positive/neutral/negative) of sentiment analysis linked with urban design activities is counted in terms of frequency, location, and time interval in each year (Liu, 2012). 3) Conduct correlation analysis between the sentiment results from step 2 and the annual prevalence of T2DM to identify the trends of public awareness. The years with a slower rise in diabetes prevalence were the focal point for the correlation analysis process. 4) A hypothetical urban design case of East Coast 2050 demonstrates how to integrate these influence variables within the urban design process and visualize the solutions to designers & public.

The remainder of the paper is structured as follows: Section 2 describes the background context, as well as related work. Section 3 presents the proposed data corpus and method. Section 4 illustrates and discusses the sentiment analysis studies between public health and online opinions and comments. Section 5 demonstrates the urban design case study for how to identify health-related urban issues (T2DM) during the design process. Finally, Section 6 summarizes the conclusions and discusses future lines of research.

2. Literature review

2.1. Health-supportive design: Diabetes

Diabetes mellitus (DM) is characterized as a group of metabolic diseases characterized by hyperglycemia resulting from defects in insulin secretion and/or insulin action (American Diabetes Association, 2009). The pathophysiology of T2DM has not been completely elucidated until now, but influencing factors include genetic susceptibility, environmental and behavioural factors such as sedentary lifestyle, nutrition and gut metagenome (Wu, Ding, Tanaka, & Zhang, 2014; Zimmet, Alberti, & Shaw, 2001). A variety of lifestyle factors such as sedentary lifestyle (Zimmet et al., 2001), smoking (Manson, Ajani, Liu, Nathan, & Hennekens, 2000) and alcohol consumption (Cullmann, Hilding, & Östenson, 2012), are of importance to the development of T2DM (Wu et al., 2014). The latest diabetes prevention programs in the United States have demonstrated the efficacy of lifestyle factors interventions that could reduce the risk of progressing from impaired glucose tolerance (IGT) to diabetes by 58%. Hence, with this basic knowledge of diabetes, achieving a supportive built environment for human health is an interdisciplinary research area involving evidence-based urban planning and design, and public health and related practice. Renalds et al. (2010) identified that aspects of the built environment can play an important role in supporting lifestyle choices that result in either beneficial or adverse health consequences for the individual and at the community level. Kent & Thompson (2012) concluded that professionals in health and the built environment (designers and policy-makers) should work together to translate health knowledge into effective policy and practice. Lowe et al. (2014) provided an overview of the evidence of the association between the built environment and chronic diseases, highlighting progress and future challenges for health promotion. Focusing on diabetes, this paper proposes an integrated approach to investigate the diabetes factors (lifestyle choices) with respect to the built environment for inclusion in the urban planning/design process.

2.2. Sentiment analysis

Liu (2012) defines sentiment analysis as “the field of study that analyses people’s opinions, sentiments, evaluations, appraisals, attitudes, and emotions towards entities such as products, services, organizations, individuals, issues, events, topics, and their attributes”. Unlike traditional content analysis, sentiment analysis is a more automated process, using a sentiment dictionary and a computer program to analyse large data sets (Hollander & Renski, 2015; Yuan, Cong, Ma, Sun, & Thalmann, 2013). Sentiment analysis of news and social media data has been used to detect social/environmental issues in many studies. Bollen et al. (2011) conducted a sentiment analysis on Twitter and calculated a daily mood for their corpus and correlated that with external notable events that took place in the period. Bertrand et al. (2013) applied sentiment analysis to Tweets and found that the public mood is generally the highest in public parks and lowest at transportation hubs, and located other areas of strong sentiment such as cemeteries, medical centres, a jail, and a sewage facility. Balahur et al. (2010) pointed out the difference in tasks for sentiment analysis between social media and news and distinguished three different possible views on newspaper articles: author, reader and text. Salas-Zárate et al. (2017) applied sentiment analysis into the health domain for diabetes, but unrelated to the urban environment. Chen et al. (2017) and Psyllidis et al. (2015) both proposed to apply sentiment analysis as part of urban data analysis for decision-making during the urban planning and design process. Hence, there still lacks clear research that could link public health issues (individual level) with sentiment analysis of news for a health-supportive urban planning and design process (community/urban level).

3. Data and methods

3.1. Opinion and comments corpus

The focus of this research lies in opinion news and their comments. Previous studies (Balahur et al., 2010; Kolhatkar et al., 2018) have demonstrated that the opinion news/articles tend to receive more interesting comments, which are more subjective than general news/articles. To understand the public awareness on the relationship between T2DM and environmental characteristics, the proposed build-up method for this corpus is modified from an existing solution by Kolhatkar et al. (2018). The proposed data sources contain two parts: 1) news searched and downloaded from Factiva, and 2) parts of comments scraped using Python. After integrating the two results, this research created a Singapore context-based opinion and comments corpus for the last 29 years (1992-2020). Currently, the corpus contains 23,301 opinions together with their comments.

3.2. Research method

Within sentiment analysis, the texts belong to either positive or negative classes or are multi-valued like positive, negative and neutral (or irrelevant). This research followed the existing lexical analysis approach to analyse the opinion and comments data corpus (Medhat et al., 2014; Thakkar & Patel, 2015). There are five steps during the lexical and comparative analysis.

The first step converts the corpus into tokens (broken into single words) by the Natural Language Processing Toolkit (NLTK) Tokenizer (http://www.nltk.org). Then the second step cleans and transforms the unnecessary excess words according to pre-tagged lexicons. The following step checks and marks every new token encountered from the lexicon in the dictionary. If there is a positive match, the total pool of scores for the input corpus will increase. For instance, if "impressive" is a positive match in the dictionary then the total score of the text is incremented. Otherwise, the score is decremented. At the fourth step, the lexical analysis results are prepared and ready for the last comparative analysis by Pearson correlation.

4. Sentiment analysis of opinions and comments

This study filters the commentaries/opinions news by the existing identified environmental characteristics (Dendup et al., 2018): "amenities", "facilities", "walkability", "urban sprawl", "public transport" and "green space". Eight searches are conducted using eight search terms (Table 1): 1) combined keyword 1: (amenities or facilities or walkability or (urban sprawl) or (public transport) or (green space)); 2) combined keyword 2: diabetes and (amenities or facilities or walkability or (urban sprawl) or (public transport) or (green space)); 3) amenities; 4) facilities; 5) walkability; 6) urban sprawl;
7) public transport; 8) green space.

Table 1: The total annual number of news (Commentaries/Opinions) by each search terms.

Year	Combined keyword 1	Combined keyword 2	amenities	facilities	walkability	urban sprawl	public transport	green space
1992	4	0	3	3	0	0	0	0
1993	1	0	1	0	0	0	0	0
1994	4	0	0	4	0	0	0	0
1995	5	0	0	5	0	0	0	0
1996	5	0	1	3	0	0	1	0
1997	2	0	1	1	0	0	0	0
1998	3	0	0	2	0	0	1	0
1999	2	0	1	1	0	0	0	0
2000	35	0	3	33	0	0	0	1
2001	20	0	0	19	0	0	1	0
2002	13	0	1	7	0	0	6	0
2003	13	0	0	13	0	0	0	0
2004	10	0	1	9	0	0	0	0
2005	25	0	3	20	0	0	4	0
2006	94	3	8	69	0	0	22	0
2007	63	1	1	54	0	0	10	0
2008	46	0	5	33	0	0	12	0
2009	65	1	4	52	0	0	9	0
2010	83	1	10	62	0	0	13	0
2011	27	0	1	23	0	0	3	0
2012	17	0	1	14	0	0	3	0
2013	25	0	0	20	0	0	5	0
2014	37	1	2	30	0	1	5	0
2015	44	0	3	33	0	0	11	1

Year	Combined keyword 1	Combined keyword 2	amenities	facilities	walkability	urban sprawl	public transport	green space
2016	18	0	0	14	0	0	4	0
2017	28	0	2	18	0	0	7	1
2018	62	0	8	55	0	0	5	0
2019	81	1	10	53	0	0	23	0
2020	164	2	11	120	0	1	39	2

4.1. Correlation analysis

Firstly, Table 1 indicates that the number of news that includes both environmental characteristics and T2DM (combined keyword 2) is far below the number of news with environmental characteristics (combined keyword 1). In other words, the opinions and comments related to T2DM are less linked with the aspects of environmental characteristics. And table 1 also reveals the public attention about environmental characteristics is changed annually. Depends on the number, environmental characteristics can be ranked from high to low to public attention: facilities, public transport, amenities, urban sprawl and green space, walkability. According to each characteristic, the basic word frequency analysis processes in word cloud chart. Figure 1 presents the most popular keywords for each environmental determinant. The top keywords for amenities are "hawker centres" & "asian infrastructure". The facilities contain "global city", & "regional hull market". The keywords within urban sprawl are "recent news" & "week visit". And the words "barrels per day (bbl day)" and "nuclear future" are related to public transport. And green space includes something like "winter trip", "tiny island" and "urban jungle". However, there is no opinions and comments from news related to walkability.

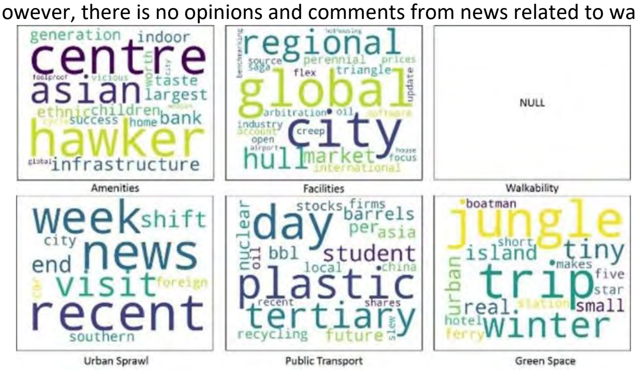


Figure 1: High-frequency keywords from the news for environmental characteristics from 1992 to 2020

From the result of sentiment analysis in Figure 2 also shows that most of the results of the news are negative (positive: 7 and negative: 887). The annual prevalence rates of Diabetes in Singapore are 8.6% in 1992, 9.0% in 1998, 8.2% in 2004, 11.3% in 2010 and 11.4% in 2014 (MOH, 2018). The correlation is calculated between sentiment analysis and annual prevalence rate of Diabetes by Pearson correlation (SPSS 2.5). And the strength of the correlation using the guide: (0.00, 0.19): “very weak”; (0.20-0.39): “weak”; (0.40-0.59): “moderate”; (0.60-0.79): “strong”; (0.80-1.0): “very strong” (Evans, 1996). Here are the findings:

- The news collected by combined keyword 2: diabetes and (amenities or facilities or walkability or (urban sprawl) or (public transport) or (green space)) (R2 = 0.965) have a strong correlation with prevalence of T2DM.
- The news collected by combined keyword 1: (amenities or facilities or walkability or (urban sprawl) or (public transport) or (green space)) (R2 = 0.697), facilities (R2 = 0.715) and public transport (R2 = 0.709) have moderate correlation with prevalence of T2DM.
- The news searched by amenities (R2 = 0.361), urban sprawl (R2 = 0.384) walkability (no related news) and green space (no related news) have a weak correlation with prevalence of T2DM.

From the first finding, the news includes diabetes match with the prevalence of T2DM, it means that public care for the prevalence of T2DM. However, the total number 1% (10 articles in 996 articles between 1992 and 2020) of news filtered by combined keyword 2 is very small compared to other topics. It means that public awareness of the relationship between environmental characteristics and the prevalence of T2DM in Singapore is little. The second finding reveals the fluctuation of the sentiment analysis result on the issues from facilities, public transport match with the prevalence of T2DM. The second finding indicates the news filtered by keywords “facilities” and “public transport” may reflect the prevalence of T2DM in Singapore.

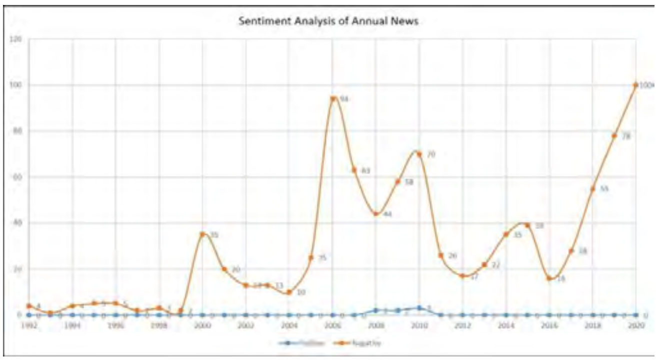


Figure 2: Sentiment analysis of news including environmental characteristics without T2DM from 1992 to 2020

5. Case study and discussion

This hypothetical case study is based on an actual team-based design studio project. The proposed design rethinks the T2DM challenges at an urban scale: walkability with amenities, green space, clusters of buildings and transportation. The site is located near East Coast Park in Singapore (total population of 5.64 million). Considering the high risk of T2DM and if nothing were to be done, the number of diabetics under 70 in Singapore is expected to rise to 670,000 by 2030. Figure 3 displays the site survey of lifestyles near East Coast Park.

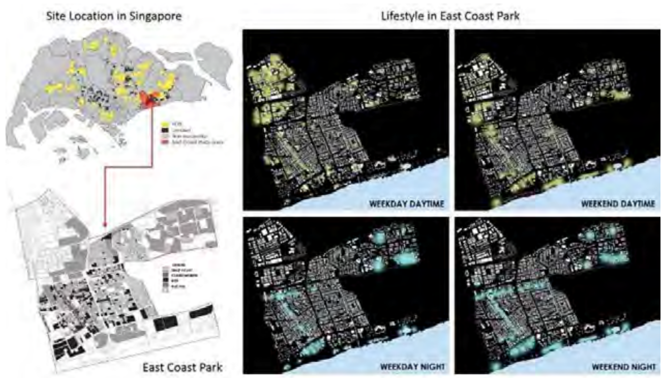
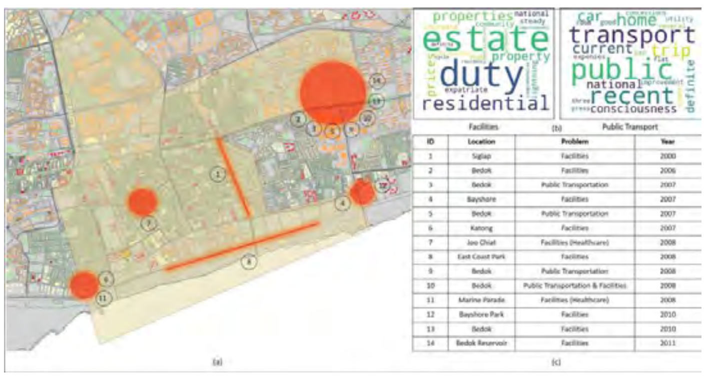


Figure 3: Site survey of lifestyles near East Coast Park (image adapted from Integrated Project Studio)

Following the proposed sentiment analysis approach, word cloud and annual sentiment analysis chart related to East Coast Park are showed in Figure 4. All the annual sentiment analysis is negative. As concluded above, the news filtered by keywords “facilities” and “public transportation” related to East Coast Park will help to reflect the prevalence of T2DM in this region. To help designers to understand the news context, the sentiment analysis results are mapped into open source GIS software (QGIS), the mentioned locations, times and topics from the news are mapping into the site. From Figure 4, it is useful for designers to figure out the design issues from public attitudes and concerns. The public complains about both facilities and public transportation mainly come from Bedok and facilities concern come from Bayshore, Joo Chia and Katong.



ID	Location	Problem	Year
1	Siglap	Facilities	2000
2	Bedok	Facilities	2006
3	Bedok	Public Transportation	2007
4	Bayshore	Facilities	2007
5	Bedok	Public Transportation	2007
6	Katong	Facilities	2007
7	Joo Chiat	Facilities (Healthcare)	2008
8	East Coast Park	Facilities	2008
9	Bedok	Public Transportation	2008
10	Bedok	Public Transportation & Facilities	2008
11	Marine Parade	Facilities (Healthcare)	2008
12	Bayshore Park	Facilities	2010
13	Bedok	Facilities	2010
14	Bedok Reservoir	Facilities	2011

Figure 4: (a) Negative news related to East Coast Park mapping in QGIS; (b) High-frequency keywords of facilities and public transport news from 1992 to 2020; (c) the details of each negative news

The hypothetical urban design process provides several insights which we believe are important for adequacy and environment equity assessment of public health (T2DM) in cities. Environmental characteristics provision indicators during the design process by top-down approach tend to have skewed distributions, and therefore the choice of the sentiment analysis of news (bottom-up approach from public concern) will perform as a supplement the outcome of public health (T2DM) adequacy and equity assessment. In particular, the hotspot complained from the public should be more considered as urgent issues for favourable and equitable health-supportive environment provision than others. This demonstrated design process have implications on considering public concerns by sentiment analysis of social media for a health-supportive environment in cities. We further discuss the implications, possible methods contributing to the health-supportive design process and highlight areas needing further research.

- Is there inequality of environmental characteristics distribution in Singapore and does the spatial scale of study make a difference related to public health (T2DM)?

It is well-documented in different cities that public health is associated with environmental characteristics distribution (Balti, Echouffo-tcheugui, Yako, & Kengne, 2014; Dendup et al., 2018; Kauhl, Schweikart, Krafft, Keste, & Moskwyn, 2016; Tan & Samsudin, 2017). However, research knowledge has not been recognized by the public widely. The public concerns about environmental characterise (particularly facilities and public transport) could help planners/designers to propose the health-supportive solution progressively. Tan & Samsudin (2017) have pointed out spatial scale has a large impact on equity assessment in Singapore and unequal distribution of parks will lead to uneven wealth and income. Hence, as the same, public health also will be impacted by inequality of environmental characteristics distribution and spatial scale. This also indirectly shows the importance of sentiment analysis from the public by different region/precinct. The planning/design process should consider/include this proposed sentiment analysis workflow.

- How is the proposed sentiment analysis dealing with other public health issues (coronavirus disease, allergies, food safety, etc.) during the design process in Singapore?

From the literature, data derived from social media has been successfully used for capturing diverse trends about public health issues(Paul & Dredze, 2014; Salathé & Khandelwal, 2011; Samuel et al., 2020; Zunic et al., 2020). Most of these approaches are based on natural language processing and share a similar workflow which involves data collection, human annotation, classification, and subsequent sentiment analysis. The approach has proven promising in many cases, but it also remains gaps to apply to the design process. Considering urban design as a broad and complex discipline, a broad range of research has confirmed that the design process requires urban designers to have the ability to tackle ill- defined problems through both problem-solving and reflective practice, including problem structuring and formulating (Cross, 2006; Lawson, 2005). Hence, the gaps mainly come from how to retrieve right & useful social media information linked with design parameters. To include public health, we need further narrow down social media by topic with spatial information. As a pilot study and concluded in part 4.1, we identified that the news filtered by keywords “facilities” and “public transport” may reflect the prevalence of T2DM in Singapore. This sentiment analysis enhanced by correlation analysis also has the potential ability to address the other public health issues. And the precondition for designers is to understand the impact of public health by which physical design parameters. Then we could link the public opinions and concerns on design parameters and public health topic. Hence, the proposed sentiment analysis workflow could help to perform as a bridge to cover the gaps between public health and the design process.

6. Conclusion

The environmental characteristics are everyday facets of urban living and individually account for most of the variations in people's lifestyle choices. This research shows public awareness of environmental characteristics related to the prevalence of T2DM and how urban designers can consider the prevalence of T2DM from a design perspective. Comparing historical news with sentiment analysis, we conclude that the public awareness of the relationship between environmental variables and diabetes is little in Singapore (1% news for both diabetes and environment). However, the news filtered by keywords "facilities" and "public transportation" may help to reflect the prevalence of T2DM in Singapore. Based on this finding, an urban design case study is proposed to demonstrate how and where this sentiment analysis can be applied to the urban design process to support design from a public health perspective. There are still many limitations to this research. The sentiment classifier by lexicons needs further improvement for news articles and the validation of the suggested points should be narrowed down to the neighbourhood level. For the future, we will enrich the current corpus by enabling analysis of multiple languages by Google translation API and seek to evaluate these design hypothesis with context-based real data.

Acknowledgements

The authors would like to acknowledge the infrastructure project "Ageing-related Data Repository" supporting the data collection and analysis process. We thank all the other professors and students from the Ageing Research Institute for Society and Education (ARISE), NTU, who have involved in the discussion to provide the ideas and methods.

References

American Diabetes Association. (2009). Diagnosis and classification of diabetes mellitus. Diabetes Care, 32(SUPPL. 1).

Balahur, A., Steinberger, R., Kabadjov, M., Zavarella, V., Goot, E., Halkia, M., Belyaeva, J. (2010). Sentiment Analysis in the News. In Proceedings of the 7th International Conference on Language Resources and Evaluation (pp. 2216–2220). Valletta, Malta.

Balti, E.V., Echouffo-tcheugui, J.B., Yako, Y.Y., & Kengne, A.P. (2014). Air pollution and risk of type 2 diabetes mellitus : A systematic review and meta-analysis. Diabetes Research and Clinical Practice, 106(2), 161–172. https://doi.org/10.1016/j.diabres.2014.08.010

Bertrand, K. Z., Bialik, M., Virdee, K., Gros, A., & Bar-Yam, Y. (2013). Sentiment in New York City: A High Resolution Spatial and Temporal View. ArXiv Preprint, 1–12.

Bollen, J., Mao, H., & Pepe, A. (2011). Modeling Public Mood and Emotion: Twitter Sentiment and Socio-Economic Phenomena. In Proceedings of the Fifth International AAAI Conference on Weblogs and Social Media (Vol. 36, pp. 194–205).

Chen, N.C., Zhang, Y., Stephens, M., & Nagakura, T. (2017). Urban Data Mining with Natural Language Processing: Social Media as Complementary Tool for Urban Decision Making. In the 17th International Conference on Computer Aided Architectural Design Futures (pp. 101–109). Istanbul, Turkey.

Cross, N. (2006). Designerly ways of knowing. London: Springer-Verlag.

Cullmann, M., Hilding, A., & Östenson, C.G. (2012). Alcohol consumption and risk of pre-diabetes and type 2 diabetes development in a Swedish population. Diabetic Medicine, 29(4), 441–452.

Dendup, T., Feng, X., Clingan, S., & Astell-burt, T. (2018). Environmental Risk Factors for Developing Type 2 Diabetes Mellitus : A Systematic Review. International Journal of Environmental Research and Public Health, 15(1), 78.

Evans, J. D. (1996). Straightforward statistics for the behavioral sciences. Pacific Grove: CA: Brooks/Cole Publishing.

Hollander, J.B., & Renski, H. (2015). Measuring Urban Attitudes Using Twitter: An Exploratory Study.
Kauhl, B., Schweikart, J., Krafft, T., Keste, A., & Moskwyn, M. (2016). Do the risk factors for type 2 diabetes mellitus vary by location? A spatial analysis of health insurance claims in Northeastern Germany using kernel density estimation and geographically weighted regression. International Journal of Health Geographics, 15(1).
Kent, J., & Thompson, S. (2012). Health and the built environment: Exploring foundations for a new interdisciplinary profession. Journal of Environmental and Public Health, 2012, 10.
Kolhatkar, V., Wu, H., Cavasso, L., Francis, E., Shukla, K., & Taboada, M. (2018). The SFU Opinion and Comments Corpus: A Corpus for the Analysis of Online News Comments.
Lawson, B. (2005). How designers thinks: the design process demystified. Architectural Press (Fourth). Elsevier.
Liu, B. (2012). Sentiment Analysis and Opinion Mining. Morgan & Claypool Publishers.
Lowe, M., Boulange, C., & Giles-Corti, B. (2014). Urban design and health: Progress to date and future challenges. Health Promotion Journal of Australia, 25(1), 14–18.
Manson, J.A.E., Ajani, U. A., Liu, S., Nathan, D.M., & Hennekens, C.H. (2000). A prospective study of cigarette smoking and the incidence of diabetes mellitus among US male physicians. American Journal of Medicine, 109(7), 538–542.
Medhat, W., Hassan, A., & Korashy, H. (2014). Sentiment analysis algorithms and applications: A survey. Ain Shams Engineering Journal, 5(4), 1093–1113.
MOH. (2018). The War On Diabetes. Singapore.
Paul, M. J., & Dredze, M. (2014). Discovering health topics in social media using topic models. PloS One, 9(8), e103408–e103408.
Psyllidis, A., Bozzon, A., Bocconi, S., & Titos Bolivar, C. (2015). A platform for urban analytics and semantic data integration in city planning. In the 16th International Conference on Computer Aided Architectural Design Futures (Vol. 527, pp. 21–36).
Renalds, A., Smith, T.H., & Hale, P. J. (2010). A systematic review of built environment and health. Family and Community Health, 33(1), 68–78.
Salas-Zárate, M.D.P., Medina-Moreira, J., Lagos-Ortiz, K., Luna-Aveiga, H., Rodríguez-García, M.Á., & Valencia- García, R. (2017). Sentiment Analysis on Tweets about Diabetes: An Aspect-Level Approach. Computational and Mathematical Methods in Medicine, 2017.
Salathé, M., & Khandelwal, S. (2011). Assessing Vaccination Sentiments with Online Social Media: Implications for Infectious Disease Dynamics and Control. PLOS Computational Biology, 7(10), e1002199.
Samuel, J., Ali, G.G.M.N., Rahman, M.M., Esawi, E., & Samuel, Y. (2020). COVID-19 Public Sentiment Insights and Machine Learning for Tweets Classification. Information.
Tan, P.Y., & Samsudin, R. (2017). Landscape and Urban Planning Effects of spatial scale on assessment of spatial equity of urban park provision. Landscape and Urban Planning, 158, 139–154.
Thakkar, H., & Patel, D. (2015). Approaches for Sentiment Analysis on Twitter: A State-of-Art study.
Wu, Y., Ding, Y., Tanaka, Y., & Zhang, W. (2014). Risk factors contributing to type 2 diabetes and recent advances in the treatment and prevention. International Journal of Medical Sciences, 11(11), 1185–1200.
Yuan, Q., Cong, G., Ma, Z., Sun, A., & Thalmann, N.M. (2013). Who, Where, When and What: Discover Spatio-temporal Topics for Twitter Users. Proceedings of the 19th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 605–613.
Zimmet, P., Alberti, K. G. M.M., & Shaw, J. (2001). Global and societal implications of the diabetes epidemic. Nature, 414(December 2001), 782–787.
Zunic, A., Corcoran, P., & Spasic, I. (2020). Sentiment Analysis in Health and Well-Being: Systematic Review. JMIR Medical Informatics, 8(1), e16023–e16023.