Applied Soft Computing 116 (2022) 108324

Contents lists available at ScienceDirect

Applied Soft Computing

journal homepage: www.elsevier.com/locate/asoc

COVID-19 Symptoms app analysis to foresee healthcare impacts: Evidence from Northern Ireland

José Sousa [a,b], João Barata [c,*], Hugo C van Woerden [d], Frank Kee [e]

[a] Personal Health Data Science, SANO-Centre for Computational Medicine, Krakow, Poland
[b] Faculty of Medicine, Health and Life Sciences, Queen's University of Belfast, Belfast, Northern Ireland, United Kingdom
[c] Centre for Informatics and Systems, Department of Informatics Engineering, University of Coimbra, Coimbra, Portugal
[d] Public Health Agency, Belfast, Northern Ireland, United Kingdom
[e] Centre for Public Health, Queen's University of Belfast, Belfast, Northern Ireland, United Kingdom

ARTICLE INFO

Article history:
Received 30 November 2020
Received in revised form 20 October 2021
Accepted 13 December 2021
Available online 20 December 2021

Keywords:
COVID-19
SARS-COV-2
Strong structuration theory
Semantic networks
Mobile app
Location analytics
Symptoms assessment

ABSTRACT

Mobile health (mHealth) technologies, such as symptom tracking apps, are crucial for coping with the global pandemic crisis by providing near real-time, in situ information for the medical and governmental response. However, in such a dynamic and diverse environment, methods are still needed to support public health decision-making. This paper uses the lens of strong structuration theory to investigate networks of COVID-19 symptoms in the Belfast metropolitan area. A self-supervised machine learning method measuring information entropy was applied to the Northern Ireland COVIDCare app. The findings reveal: (1) relevant stratifications of disease symptoms, (2) particularities in health-wealth networks, and (3) the predictive potential of artificial intelligence to extract entangled knowledge from data in COVID-related apps. The proposed method proved to be effective for near real-time in-situ analysis of COVID-19 progression and to focus and complement public health decisions. Our contribution is relevant to an understanding of SARS-COV-2 symptom entanglements in localised environments. It can assist decision-makers in designing both reactive and proactive health measures that should be personalised to the heterogeneous needs of different populations. Moreover, near real-time assessment of pandemic symptoms using digital technologies will be critical to create early warning systems of emerging SARS-CoV-2 strains and predict the need for healthcare resources.

1. Introduction

The COVID-19 pandemic was recognised by the World Health Organisation on 30th January 2020 and became a prominent line of research across disciplinary boundaries. Globally, as of 19 October 2021, there were 240,940,937 cases of COVID-19, with the number of deaths reported to the World Health Organisation scaling to near 5 million (4,903,911 deaths), across virtually all countries. In Northern Ireland, as of 18th October 2021, there were 2629 reported deaths and 260,974 individuals with a positive laboratory confirmed test, with a total of 4,654,280 tests undertaken for COVID-19. Predicting the pandemic's spread, forecasting its severity in regions or groups, and the effects on healthcare systems have been the focus of many research teams around the globe [1,2]. However, this is a challenging task that, in practice, is primarily supported by test and trace systems and models of non-pharmacological interventions.

Patients with the new coronavirus (SARS-CoV-2) require a range of remote and/or face-to-face medical care, depending on the severity of the symptoms and the need for palliative or healthcare interventions. Non-pharmacological interventions (e.g., quarantine, economic aid, and regulations) have been and will be necessary to adjust the pandemic curve to support the healthcare response. It has become clear that information is one of the most valuable assets in dealing with the heterogeneous nature of COVID-19, particularly data emerging from the mobile health (mHealth) ecosystem, facilitating real-time epidemiology research [3,4].

Recently, mobile technologies and apps have become valuable data sources for advancing our knowledge of COVID-19 [5,6]. These have included, for example, mobile applications to trace COVID-19 infections or monitor the symptoms of suspected/infected patients [4,7]. One such example is the Northern Ireland COVIDCare app, which the Department of Health launched in Northern Ireland (DoHNI), used in this paper.

* Corresponding author.
E-mail addresses: J.Sousa@sanoscience.org (J. Sousa), barata@dei.uc.pt (J. Barata), hugo.vanwoerden@uhi.ac.uk (H.C.v. Woerden), F.Kee@qub.ac.uk (F. Kee).

https://doi.org/10.1016/j.asoc.2021.108324
1568-4946/

As mobile data for understanding COVID-19 symptoms have accumulated at an increasing pace, so too have the opportunities to use artificial intelligence (AI) [8]. The uncontrolled spread of COVID-19 has quickly expanded outside of medical settings. In fact, the "*impact of pandemics is beyond imagination and not limited to the loss of human lives but can threaten the economic stability and existence of affected countries*" [9]. AI should be included in the toolbox to deal with the ongoing pandemic crisis "*establishing the natural history of infection, including incubation period and mortality rate; identifying and characterising the causative organism; and, in some instances, epidemiological modelling to suggest effective prevention and control measures*" [8]. However, AI modelling of symptoms progression has been poorly utilised thus far and mobile app data to help fight COVID-19 requires greater exploration. These constraints can be attributed to the nature of how the data is collected. Self-reported symptoms of COVID-19 on mobile platforms can be integrated with location analytics to advance more effective, and in-situ, public health measures.

We formulated the following research objective: Create AI semantic networks of COVID-19 symptoms from mobile data and analyse health-wealth implications at the meso level (specific population groups) of a metropolitan area. Strong Structuration Theory [10,11] provided the theoretical basis for analysis. Semantic networks were created using self-supervised machine learning with node-significance measured as betweenness, interdependence as weight connectivity, and network structural change entropy based on betweenness [12,13].

The remainder of this paper is presented as follows: The following section presents background theory, namely, mobile health adoption in COVID-19 pandemic management and the need to address local contexts of disease progression for effective public measures. Subsequently, the method is explained, followed by the results obtained by modelling COVID-19 data for Belfast and three evaluation episodes. A discussion follows, including the implications for theory and practice. Finally, the conclusions, limitations, and opportunities for future research are presented. This structure follows the publication schema suggested by [14] for the design science research (DSR) paradigm [15].

2. Background

2.1. The role of mHealth in COVID-19

Mobile technologies adopted in COVID-19 pandemics have followed two trends. One focuses on contact tracing and the other on remote monitoring/assistance of patients. Such apps emerged in all corners of the world [5,6], collecting large amounts of data that the general population can now use (e.g., information or real-time warnings about contact with infected individuals) and researchers worldwide [3,4]. Nevertheless, a digital health "*implementation process is likely to be challenging and resource-intensive*" [16], and public health authorities are not yet utilising the full potential of near real-time data to support decision-making, for example, modelling mobile data when lab tests are unavailable or scarce.

Disease symptoms are essential to understand the severity of COVID-19 [17], but COVID-19 is not a socially neutral disease [18]. For example, "[o]*lder, age, male sex, comorbidities*, [and specific health related symptoms] *predicted critical care admission and mortality. Non-white ethnicity predicted critical care admission but not death*" [19]. Moreover, "*people with complex needs, vulnerable populations, and marginalised groups are at increased risk from covid-19 and the health effects of containment strategies*" [20]. As multiple outbreaks and waves reveal, this "*syndetic pandemic*" [18] is highly dynamic and challenging to contain. New methods that integrate diverse data [21] are necessary to allow near real-time monitoring of COVID-19 at the individual and at the community or social group level.

The recent advances in epidemiology using mobile data and artificial intelligence are significant. For example, Menni et al. [4] suggest "*that loss of sense of smell and taste could be included as part of routine screening for COVID-19 and should be added to the symptom list currently developed by the World Health Organisation*". Going beyond the relevance of a specific symptom, Menni et al. [4] further state that a "*combination of symptoms, including anosmia, fatigue, persistent cough and loss of appetite,* (…) *together might identify individuals with COVID-19*", which is consistent with [3] who found that "*individuals with complex or multiple (3 or more) symptomatic presentation perhaps should be prioritised for testing*". However, these authors also conclude that additional research is necessary to combine symptoms and predict COVID-19 incidence and progression.

mHealth solutions are supporting significant epidemiologic studies. For example, in the UK [22], one of the studies using mobile app data found six main clusters of COVID-19 symptoms predictive of different probabilities of intensive care need. According to the authors, the need for respiratory support ranged from 1,5% in the less severe cluster of symptoms to 19,8% in cluster 6, the most dangerous condition. The first two clusters are similar to flu and have little risk for health care support, while cluster 3 adds a new combination of symptoms: loss of smell, headache, loss of appetite, diarrhoea, chest pain, and sore throat. This cluster has a 5x higher probability of a hospital visit and a consequent impact on public service response [22]. This work inspired new studies exploring the role of near real-time symptoms monitoring and the efficient management of the healthcare response.

However, existing results are not conclusive, and few studies have combined self-reported symptoms and the characteristics of the population at the (meso) city level, which could allow a more fine-grained perspective on the disease's social determinants (e.g., localities with similar health-wealth indicators).

2.2. Theoretical lens for meso-analysis of COVID-19

COVID-19 is being extensively studied with AI at the micro-level (e.g., individual diagnosis [23,24]), and many studies provide a macro vision of symptoms, country-level performance, and dynamics [3,25,26]. However, examining COVID-19 mobile data at a meso level, that combines the social characteristics of a specific region and its complex interrelations or entanglements (e.g., hot spots, healthcare capacity constraints, quarantine efficacy) needs near real-time data and knowledge visualisations able to assist professionals (e.g., public health staff). Mobile apps are emerging as a *quasi-testing* tool.

Structuration theory [27] suggests that structures and agents are inseparable, and both are necessary to understand a social phenomenon. According to this social theory, although the micro (e.g., an individual) and macro (e.g., country or continent) levels of analysis are essential, the meso level of analysis is equally important [11]. According to strong structuration theory (SST), the four elements that must be considered are: (1) the external structures (context where the action takes place), (2) the internal structures represented by conjunctural networks of agents (humans and technology), (3) the actions, and (4) outcomes of the action [11]. This theory has helped us understand "conjunctures" and their application in healthcare, particularly regarding technology adoption in practice [11]. Therefore, we considered it a suitable lens to understand the influence of local networks, linking position and practice concerning COVID-19.

Modelling complex systems in uncertain environments requires a capacity for data-driven learning within the system [28],

Problem	***Research Process***	***Solution***
Tests are essential to assist public health decisions but do not evaluate severity. Healthcare facilities need tools to predict patient admissions. Understand the possible relationship between regional deprivation and COVID-19 severity.	Delimitation to the (mobile app) self-reported symptoms class of problems. Modelling Belfast regions. Evaluation.	Complex network symptoms modelling at the meso-level of analysis – a possible solution to foreseen healthcare impact.
Input Knowledge	***Concepts***	***Output Knowledge***
mHealth in COVID-19 COVID-19 mobile app data.	Complex networks. Structuration theory.	Modelling approach to foreseen COVID-19 severity at a regional scale. Evaluate health-wealth implications.

Fig. 1. The DSR grid for COVID-19 symptoms app analysis.
Source: Adapted from [33].

and the modelling and visualisation of significance and interdependence. Network concepts and tools are a vital part of addressing such problems [29]. A complex network is a structure of connected (linked) elements (nodes) that allows the development of knowledge representations of the behaviour of techno-social systems [30]. Therefore, creating complex networks to represent meso-level relationships of COVID-19 offers a promising framework for advancing our knowledge of symptom prevalence and the creation of new tools to support public health interventions.

3. Method

Our work follows the design science research approach, including the activities of building the artifacts (models), evaluating the results according to different metrics, and producing relevant and justified theoretical knowledge [31,32]. Fig. 1 outlines the work according to the DSR grid proposed by [33].

The solution presented in Fig. 1 is relevant to support the decisions about: (1) resource allocation in local healthcare facilities (e.g., human resources planning, COVID-19 bed occupancy projection), (2) early warning of an exceptional pandemic spreading, and (3) open data available for COVID-19 research. The primary user of the proposed system at this stage is the public health department. Nevertheless, an interface for healthcare facilities could be engaging for future work.

The data used in our research was obtained via the Northern Ireland COVIDCare app, an IT solution developed by the Northern Ireland Department of Health, on behalf of the Public Health Agency for Northern Ireland, in collaboration with a private company to inform the public and track the symptoms of symptomatic individuals [34]. The data modelling process uses semantic networks to visualise the interdependence of complex sociotechnical structures [29]. This research has adopted a self-supervised statistical machine learning methodology to develop data abstractions modelled as a network of significance and interdependence, generating semantic knowledge. The approach identifies relevant nodes of a complex network and their relationships, using information entropy based on betweenness [12,13].

The steps to create the network models include:

(1) Contextualisation:
Selecting the environment to be modelled.

(2) Abstraction
Creation of data abstractions or data reduction. Each variable considered in the network is reduced to a composition of the feature name and its qualitative value or to the most usual value (mode) and classified as up (U) when it is above and down (D) when it is below (1).

(3) Learning
Production of statistical learning models from the data reduction process presented in step 2. The resulting structure is visualised as a network of significance interdependence using betweenness, communities, and connectivity weights.

(4) Inference
Measurement of the entropy of the network and filtering the most relevant variables using their betweenness value [13,35].

The methodology was developed using R and Python in the Zeppelin framework and integrated with Gephi [36] to provide the structural measurements and network visual representation. The context is described by producing an edge list created by concatenating the string name with its value as described by the following algorithm:

IF event = discrete

Node = event_name + discrete_value

ELSE

IF event = continuous

Select Centrality Measure (Mode, Mean, Median)

M=measurement(event)

IF event_value > M *(Algorithm 1)*

Node = event_name+U

ELSE

IF event_value = M

Node = event_name+M

ELSE

Node = event_name+D

The result is a network of nodes with colours representing communities and lines (thicker lines represent a more relevant interaction between nodes) that can be interpreted by non-experts in mathematics or statistics (as happens in more complex representations) and support a near real-time visualisation and attention structure [37] for COVID-19 analysis.

This methodology also facilitates machine learning enabled measurement of change. While the concept of entropy has been used and studied in different network contexts, the current use is significantly different as we considered a data-driven emergent network describing symptoms as a complex adaptive system [38]. We consider network entropy from the formula given in [18,19] integrating normalised betweenness values (Eqs. (1) and (2)).

$$H(A) = -\sum_{i=1}^{n} P(a_i) log_2 P(a_i) \tag{1}$$

$$H(B) = \left(-\sum_{i=1}^{n} b_i log_2 [b_i] \right) /N \tag{2}$$

Information entropy in complex adaptive systems [35,38] can measure the system complexity and is an important measurement to describe the structural change of a complex network. It has been used in different contexts [13,39]. The novelty of this approach resides in the data-driven self-supervised learning of the emergent network and its encoded knowledge, based on significance and interdependence expressed by betweenness values and connectivity, which are used as measurements of change and quantifiers for AI inference.

Our research evolved in three phases. First, we adapted our method to develop semantic networks exclusively to the Belfast area, combining health-related inputs (e.g., symptoms, comorbidities) and social characteristics of the app user as inputs (e.g., demographic data). The visualisation of COVID-19 patterns in specific localities is vital to public health authorities to: (1) anticipate the possible impact on specific hospitals; (2) understand the severity of symptoms in particular areas (e.g., residential/industrial, deprived areas); and (3) its potential for near real-time monitoring of COVID-19 dynamics in specific population sub-groups. We have thus analysed the conjunctures found in the COVID-19 models supported by the mobile data. Finally, we have evaluated the models' accuracy, comparing models based on entropy and betweenness centrality and previous findings reported in the literature [17].

4. Results

The initial dataset relates to the period between 22nd March and 15th April 2020, i.e., aligning with the early adoption of the app (more daily inputs), with a total of 1702 updates. Figs. 2 and 3 present the models obtained with COVIDCare data for each region of Belfast. The models are ordered according to the deprivation measurements, and the first model (West region) includes a short legend to describe the most relevant nodes.

The health-wealth analysis is presented in the following subsections.

4.1. Self-reported symptoms

West Belfast was the only region that did not report breathing difficulties (node ***DBREATHING_NO*** has the higher significance in the symptoms network). As we found in clusters 4–6 presented in [22] and in the evaluation phase using the data from [17], breathing difficulties are one of the most dangerous symptoms. Moreover, this model also reveals that most users did not report fever or cough. The most significant population of confirmed patients (***EILLNESS_YES***) seems asymptomatic (no breathing difficulties and did not stop most of their activities). Therefore, during this phase of the pandemic, West Belfast would not be expected to put pressure on local healthcare facilities. In this paper, we are not accounting for public health measures implemented in that region (for example, if a quarantine intervention was widespread, the model could indicate success of the intervention). Nevertheless, as asymptomatic patients have the capacity to spread the infection, such regions should still be closely monitored.

The North Belfast network is more complex in the number of nodes and symptoms identified, although it does not reach the complexity of the South and East areas. The North Belfast district demonstrates more patients with more significant breathing difficulties (Community integrating ***DBREATHING_YES*** and ***EILLNESS_YES***) and other less known symptoms, such as loss of smell and taste (more recently included in the list of relevant symptoms in the UK), and muscle and joint pain. This region could therefore be expected to put more pressure on healthcare facilities (when compared to the West).

East Belfast shows more complex combinations of symptoms, including breathing and fever, and users who present with symptoms for more extended periods (over 14 days). This region reveals an ongoing outbreak pattern, as the symptoms that began during the previous six days are also relevant to the model. However, cough is the prevailing symptom in this community, with more sick cases (***EILLNESS_YES***) in green.

South Belfast presents a more complex scenario when compared to West and North regions, but less concerning than East Belfast. The most severe symptoms are weakly connected (red and purple communities). The analysis of symptoms revealed an interdependence between two particular symptoms: loss of taste/smell, and muscle/joint pain — a pattern that was particularly evident in three of the regions.

4.2. Socioeconomics

We could not find a uniform relationship between symptoms and social factors in each region, which suggests that different aspects, such as age or comorbidities may be more related to the

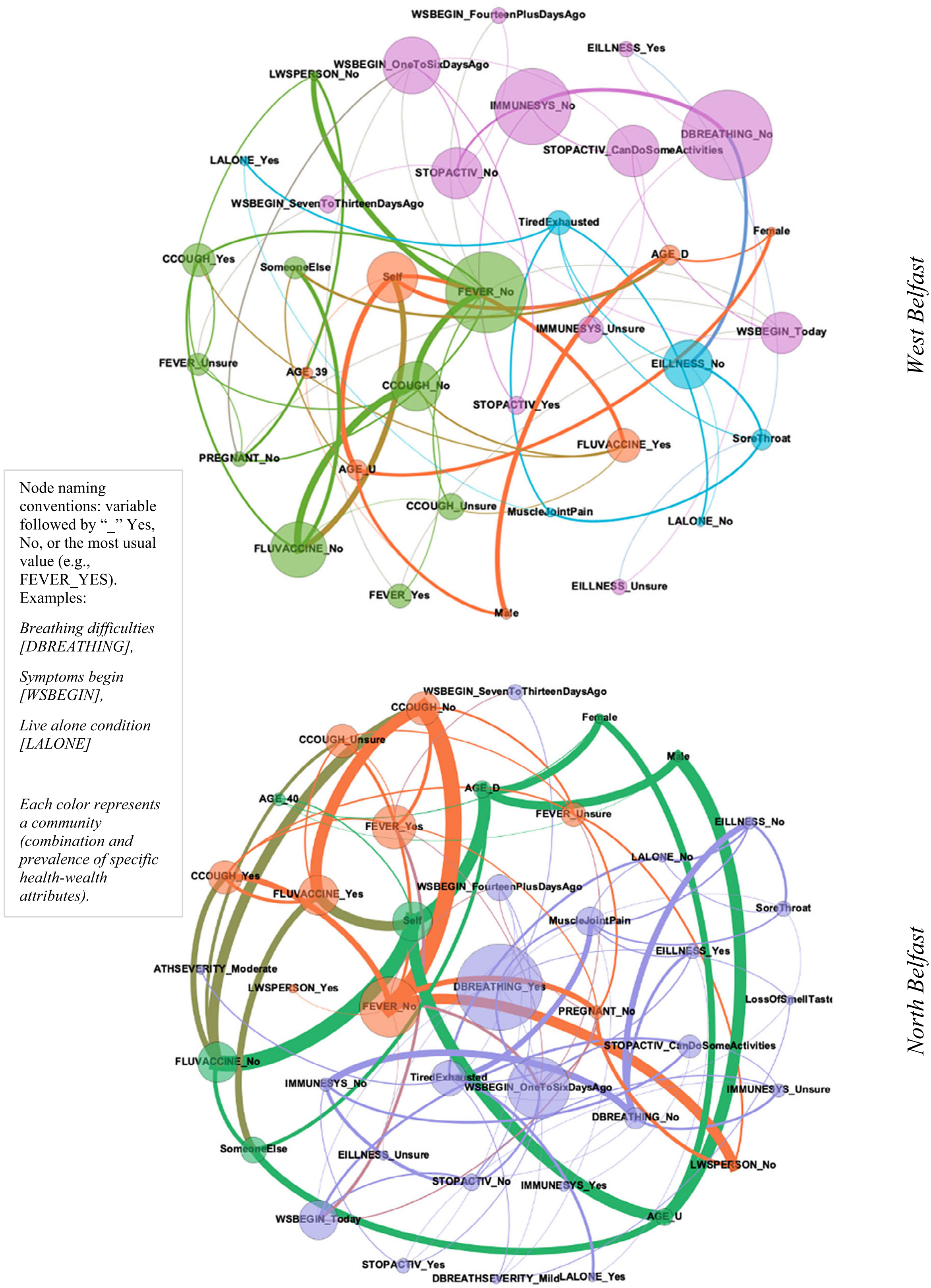


Fig. 2. West Belfast and North Belfast area modelling results on the NI APP data.

severity of the disease. This type of visualisation may be helpful both to public health authorities and general practitioners as it provides greater detail regarding localities that may pose a risk to service capacity.

The results were compared with the Northern Ireland Statistics and Research Agency report [40], which presents the characteristics of the population in each of Northern Ireland regions. West Belfast has the lowest percentage of people over

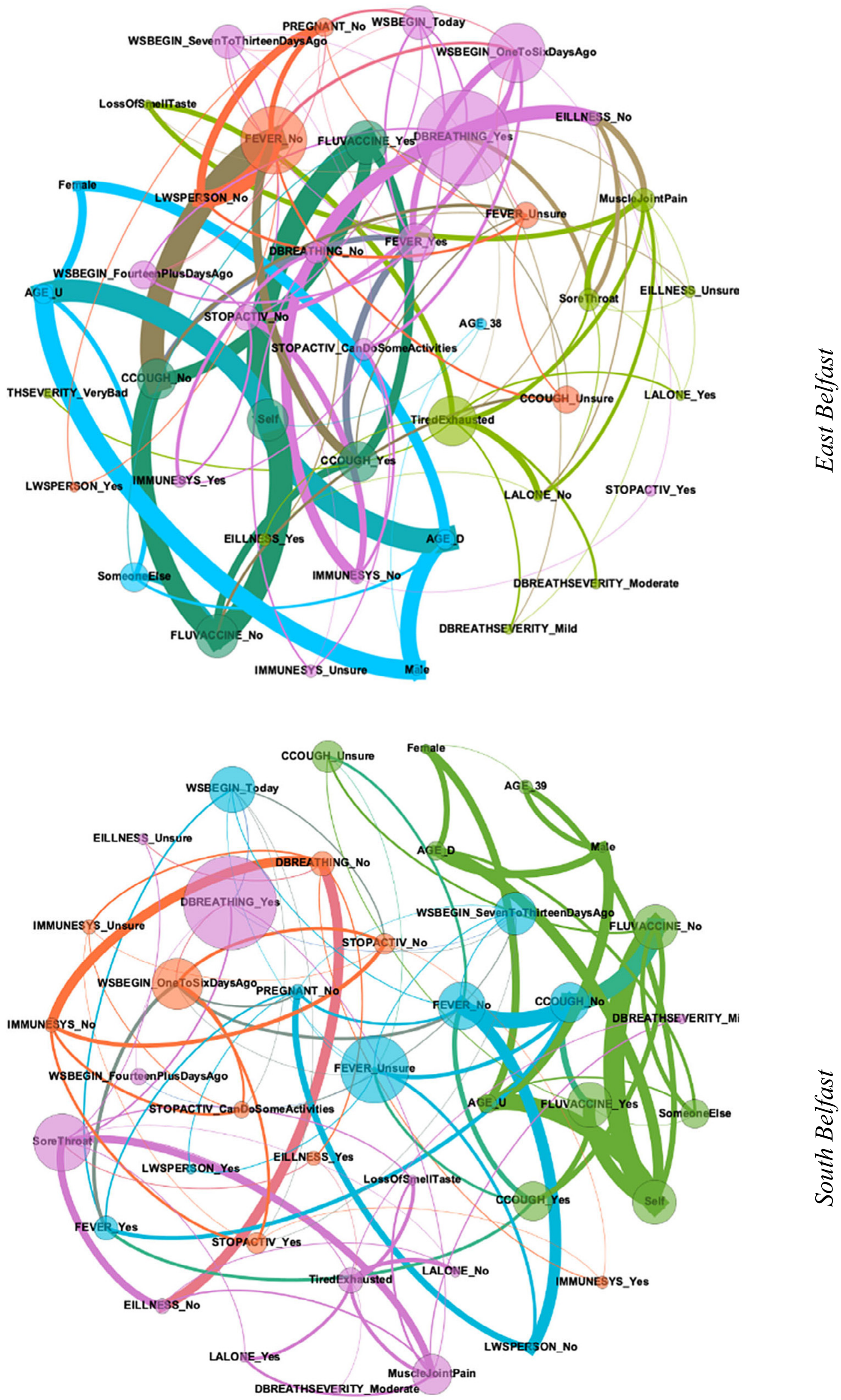


Fig. 3. East Belfast and South Belfast area modelling results on the NI APP data.

Table 1
Deprivation in Belfast area.

	West	North	East	South
Population (number)[a]	94 445	103 834	94 905	114 065
Population 0–15 (rank)	1	2	3	4
Population 16–39 (rank)	2	3	4	1
Population 40–64 (rank)	3	2	1	4
Population over 64 (rank)	4	2	1	4
Multiple deprivation measure	46%	31%	8,7%	5,2%
Income domain	10%	29,6%	2,2%	1,7%
Employment (18–64)	58%	31%	4,3%	1,7%
Health	56%	29,3%	13%	8,6%
Education	40%	34,5%	19,6%	15,5%
Living environment	30%	19%	8,7%	29,3%
Crime & Disorder	22%	24,1%	15,2%	17,2%

[a]Belfast West and Belfast South have the lowest percentage of population over 65% in all NI. Conversely, both have the highest percentage of population in the interval 16–39. Belfast West is the most deprived area in NI, followed closely by North.

65 years old, followed by Belfast South (below 15%). On the other hand, Belfast East (with the highest percentage between 40 and 65 years) and Belfast North has the highest percentage of its population over 65 yrs. Deprivation measurements on the four regions are presented in Table 1.

Table 1 also shows high deprivation levels in Belfast West (which is also the region with a younger population), followed closely by the North (both over 30% of the population living in social deprivation), East, and South (both below 10% of the population living in social deprivation) respectively.

Although previous research points to a relation between past/current pandemic periods and deprivation measures in specific localities (e.g., where there is a higher prevalence of comorbidities or where there may be difficulties in social distancing within more deprived communities), such as the work of [18,20], other studies are inconclusive in that regard. For example, another study [19] states that "*social deprivation was not predictive of outcome* [critical care admission and death]", while [41] found that "[s]*ocioeconomic deprivation and having no qualifications were consistently associated with a higher risk of confirmed infection*". Our findings may contribute to this debate as we found less severe symptoms in two of the most deprived regions of Northern Ireland (East and North Belfast), which theoretically support the notion that social deprivation may not be directly associated with more severe outcomes. Other factors such as the population's age distribution and access to healthcare services (many may be asymptomatic) may be relevant. Lower usage of the symptom app in areas of social deprivation could also be a confounding factor.

However, our work goes beyond previous research, as we reveal the dynamic evolution of COVID-19, permitting near real-time monitoring of the complex interactions in each region, surpassing previous "static" associations of symptoms and social characteristics. COVID-19 spread has no respect for borders between age groups, races, genders, or geographies. For example, if the population in some Belfast regions work outside their area, the disease will probably spread between regions. If unemployment is high, social interactions between older people and the more mobile segments of the population (students, mid-age persons) might lead to more severe outcomes in that region. Therefore, monitoring changes over time (e.g., contrasting the node significance and its interactions) could usefully reveal the impacts of policy interventions. For example, a more complex combination of symptoms in North Belfast could reveal a local outbreak that can potentially affect areas with more elderly residents and a greater possibility for spread to the regions where those residents work. In the South, it is anticipated that the risk of spreading to other regions may be lower, as social deprivation is less common in that area.

4.3. Evaluation

Three evaluation strategies were used following the FEDS framework suggested by [42] for studies that aim to design new artifacts (e.g., models) and support decision-making in sociomaterial contexts [43]. First, we conducted unstructured interviews with four public health experts in NI to discuss the possible impact of predicting COVID-19 infections using AI models, compared with lab testing. Our purpose was to evaluate: (1) the model's comprehensibility, (2) whether public health teams were already using similar techniques, and (3) the approach's potential. According to the feedback received, this model could be used to probe the evolution of the pandemic in specific locations lacking sufficient testing data, allowing more efficient use of testing and enhancing health evidence regarding disease progression and the emergence of new variants.

We have evaluated technical risk and efficacy [42] using a different approach. We wished to confirm that our results, using complex networks, would be equivalent to other techniques. Therefore, a modelling of the symptoms in disease stages using the proposed self-supervised modelling was conducted using the data published in [17]. The results are illustrated in Fig. 4 and confirm the capacity of the learning process to identify the most prevalent symptoms found by the authors of the study. Thus, pneumonia is an aggregator for a set of symptoms including: cough, difficulty breathing, high temperature, loss of appetite, and chest pain.

The modelling process revealed a strong interdependence between fever and cough, and between fever and fatigue, in the group of those who had not recovered. The interdependence between age and fever increases to the age of sixty years compared to the group of recovered patients. The complexity of symptoms increases significantly in this model, including: lack of appetite, difficulty walking, or muscle pain.

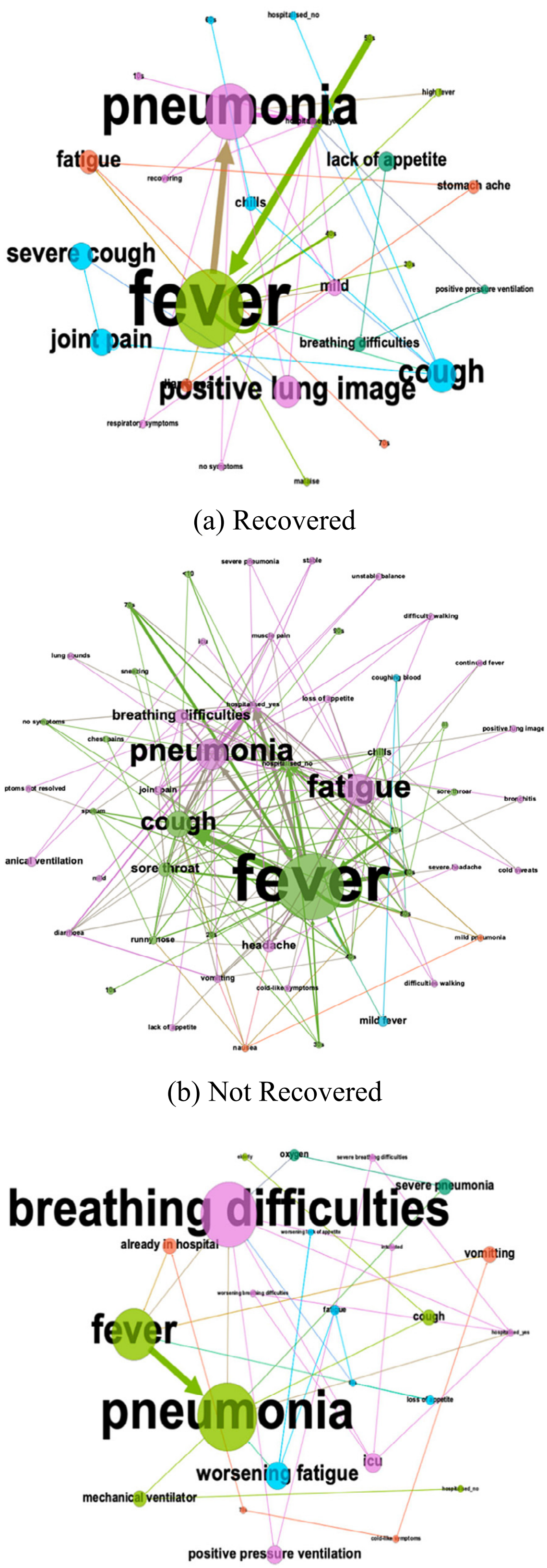
Finally, we evaluated the evolution of test results and the analysis of app data (Fig. 5).

Fig. 5 compares the evolution of positive reported cases between 12th April and 31st May (blue), and the app updates for severity 3 cluster using: (1) a seven-day rolling average (green), and (2) a daily figure (grey). The dataset was exported from the HSC NI COVID App data provided by the Social Media Observatory at Queen's University of Belfast based on HSC NI provided data. Over 24 000 app updates were used for this evaluation step (a 450 daily average).

Although insufficient to conclusively determine the accuracy of "digital testing" supported by mobile app data, there are interesting insights in the selected period. First, a similar trend between traditional test results and the mobile app index was obtained with the selected severity 3 cluster of symptoms. Second, the drastic increase of cluster 3 cases at the beginning of May, and on 17/05, (more evident for the daily cases, because the seven-day average reduces variations), parallels the high number of positive cases found with traditional tests. Third, the possibility arises of using mobile app data as quasi-test indicators when traditional test results are unavailable (lack of results or long delay in obtaining those results).

5. Discussion

Mobile technologies in a health ecosystem are precious tools that can play a central role in managing emerging threats, such as the actual COVID-19 pandemic. Firstly, they provide easy access to near real-time assessment of specific population segments. Secondly, they provide sources of data which are crucial to location analytics. Third, the use of symptom tracking apps could contribute to encouraging protective health behaviours. Although



(a) Recovered

(b) Not Recovered

(c) Deceased

Fig. 4. Modelling symptoms of non-recovered patients using the data of [17].

we did not find a relationship between disease severity and social deprivation (at an area level), it is important to acknowledge that factors such as the level of knowledge of COVID-19 symptoms, and adherence to preventative behaviours, are related to the risk of spread of infection [44].

Recent studies, for example, [3], have shown how mobile technologies can assist in high-level analysis of epidemiologic patterns in Wales and Scotland. Our work adds to this research by revealing an approach to data-driven machine learning, which produces knowledge about symptom clusters and stratification as well as its context-specific significance and interdependence.

Global and national level research studies are essential to address the challenges of COVID-19. However, pandemic management also requires a more granular examination of population segments and specific geographical areas, namely, cities, small towns, or even more restricted communities and social groups. Complex network analysis offer an interesting tool to evaluate and visualise those combinations, as demonstrated in the Belfast region.

Our conclusions also confirm several findings from previous research, for example, the need to differentiate clusters (communities in the case of complex networks) with potential for more significant impact on local healthcare, including, for example, the impact of social conditions on COVID-19 management. Comorbidities affect the severity of COVID-19, but there are other factors to consider, such as the type of work, life habits, or types of activity common in the population [45].

The modelling process also revealed information about COVID-19 symptoms' interdependence, namely, the association between the loss of smell and taste, and muscle and articular pain. These associations extends the findings of [4], who suggested *"that loss of sense of smell and taste could be included as part of routine screening for COVID-19"*. Nevertheless, additional research is necessary to confirm that these symptoms are related or emerge as a combination of other factors, for example, among certain age groups. Our models did not reveal any association between the flu vaccine and COVID-19 symptoms.

Local health policies (e.g., travel restrictions, tests) must combine the explicit knowledge provided by mobile apps (e.g., symptoms) and local contextual "environmental" information. Moreover, COVID-19 outbreaks happen rapidly, requiring continuous monitoring of different variables that may reveal changes in the disease pattern in the population across different locations, as we found in our models. For example, mobile app data can be used to indicate COVID-19 progress when testing is not available and provide an early warning of a possible increase in hospital admission (identified by an increase of cases in clusters that demonstrate higher level of severity).

The mobile data revealed less complex patterns of self-reported symptoms of COVID-19 in regions with younger populations, more significant health deprivation, and higher unemployment rates. Over time, the analysis of complex symptom networks may provide insights into trends, particularly when tests are unavailable, and help highlight communities that should be prioritised for testing.

6. Conclusion

This paper presents a semantic network approach to model COVID-19 entanglement using mobile data to extract explicit (e.g., self-reported symptoms) and implicit knowledge (e.g., location, social factors, trends). Our proposal extends past research, including a health-wealth layer and AI self-supervised learning capacity to profile symptoms in specific contexts.

Modelling social and health-related data at the meso level is essential to understanding the dynamics of the virus in the community, complementing, or even replacing test results (e.g., Lateral Flow or PCR tests) when these are unavailable.

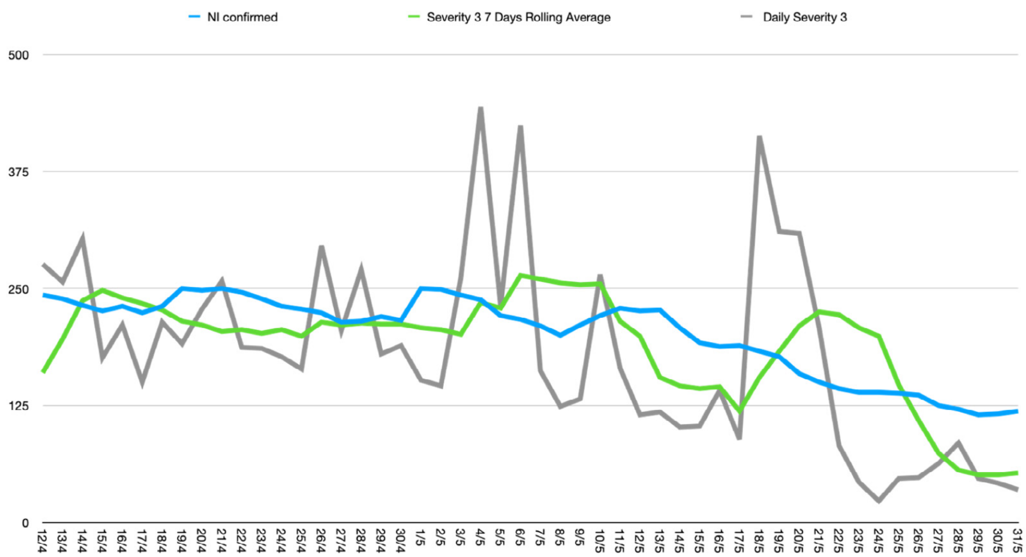


Fig. 5. Evolution of NI positive reported cases and daily severity modelling of mobile app data.

6.1. Limitations

Adopting artificial intelligence techniques to address the challenges of COVID-19 with mobile data is still evolving. Our work reveals that intelligent COVID-19 self-reported symptom data analysis can assist mounting an appropriate public health response and complements existing official data and lab test results. Until now, self-reported mobile data has mostly been helpful after the fact (e.g., tracking contacts or warning the user about symptoms that deserve attention). Our work sheds light on the value of understanding patterns of COVID-19 symptoms in association with the social characteristics of the target population. However, the data is only related to the Belfast region and does not represent the entire Northern Ireland population. We have only evaluated the most noticeable relationships between the network nodes in this dataset. Moreover, deprivation has been attributed at a regional level, based on the user's location, rather than based on the individual characteristics of each user (e.g., income and education).

A natural limitation of the approach that we have used is the existence of asymptomatic cases, which can be identified by laboratory tests. The adherence of the population to the app use is another crucial aspect to consider. Therefore, using AI techniques in mobile data is complementary and most valuable in evaluating the progression of symptoms severity (which is difficult to do with Lateral Flow or PCR tests) and providing projected utilisation of healthcare facilities.

The number of records is considered sufficient to (1) evaluate the accuracy of the visualisation approach and (2) reveal its capacity to represent COVID-19 entanglement based on mobile data. However, the predictive value of our model needs additional research. Despite the alignment of our results with other models [17,22] and the apparent interest in revealing trends of COVID-19 when lab tests are unavailable (or to inform testing strategies in the case of limited testing capacity), AI and machine learning methods should be used in parallel with traditional testing and epidemiological techniques to support public health decision-making.

6.2. Opportunities for future work

There is scope to add more elements to the semantic networks that we have developed besides demographic and deprivation measures, for example, a day-by-day modelling comparison with reported cases and/or mobility. Comparing symptom patterns across the most relevant economic sectors in the region (e.g., construction, retail) could also provide interesting results.

Inspired by the three replicability questions proposed specifically for DSR by [46], the following suggestions are put forward. First, "*Does the artifact provide utility?*" The same AI approach can be used in different datasets, exploring symptom patterns and providing longitudinal analysis, in conjunction with public health professionals. It is also possible to explore other data analysis techniques. Second, "*Is the design theory complete?*" It would be interesting to include other attributes in the model, such as personal data (social characteristics, other comorbidities, social habits), and new symptoms that would help do distinguish other specific viruses, such as influenza. Finally, "*What design theory components fit a larger context?*" One of the most relevant potential applications of our results [47] is the early identification and monitoring of circulating viral strains, opening new opportunities for future research.

The limitations of our work may point to new research directions, for example, studying different cities and regions and comparing their patterns of COVID-19 symptoms. We hope this research may inspire other researchers working with mobile data to increase our understanding of the complex interactions of social and health factors in pandemic management.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

This work uses non-identifiable data provided through the use of the DoH NI app, 'COVID-19 NI'. The app was produced on behalf of the DoH by Digital Health and Care Northern Ireland (DHCNI) and the Public Health Agency for Northern Ireland, working with commercial partners Civica and BigMotive. We wish to acknowledge the access granted to the non-identifiable data, which led to this output. This work is partially funded by national funds through the FCT — Foundation for Science and Technology, I.P., Portugal, within the scope of the project CISUC — UID/CEC/00326/2020 and by European Social Fund, through the Regional Operational Program Centro 2020. The authors are grateful for the insightful suggestions made by the reviewers.

References

[1] N. Zheng, S. Du, J. Wang, H. Zhang, W. Cui, Z. Kang, T. Yang, B. Lou, Y. Chi, H. Long, M. Ma, Q. Yuan, S. Zhang, D. Zhang, F. Ye, J. Xin, Predicting COVID-19 in China using hybrid AI model, IEEE Trans. Cybern. 50 (2020) 2891–2904, http://dx.doi.org/10.1109/TCYB.2020.2990162.

[2] B. Rostami-Tabar, J.F. Rendon-Sanchez, Forecasting COVID-19 daily cases using phone call data, Appl. Soft Comput. (2020) 106932, http://dx.doi.org/10.1016/j.asoc.2020.106932.

[3] D.A. Drew, L.H. Nguyen, C.J. Steves, C. Menni, M. Freydin, T. Varsavsky, C.H. Sudre, M.J. Cardoso, S. Ourselin, J. Wolf, T.D. Spector, A.T. Chan, Rapid implementation of mobile technology for real-time epidemiology of COVID-19, Science (80-.) 0473 (2020) eabc0473, http://dx.doi.org/10.1126/science.abc0473.

[4] C. Menni, A.M. Valdes, M.B. Freidin, C.H. Sudre, L.H. Nguyen, D.A. Drew, S. Ganesh, T. Varsavsky, M.J. Cardoso, J.S. El-Sayed Moustafa, A. Visconti, P. Hysi, R.C.E. Bowyer, M. Mangino, M. Falchi, J. Wolf, S. Ourselin, A.T. Chan, C.J. Steves, T.D. Spector, Real-time tracking of self-reported symptoms to predict potential COVID-19, Nat. Med. 26 (2020) 1037–1040, http://dx.doi.org/10.1038/s41591-020-0916-2.

[5] Y.S. Bae, K.H. Kim, S.W. Choi, T. Ko, C.W. Jeong, B.L. Cho, M.S. Kim, E.K. Kang, Information technology-based management of clinically healthy COVID-19 patients: Lessons from a living and treatment support center operated by seoul national university hospital, J. Med. Internet Res. 22 (2020) e19938, http://dx.doi.org/10.2196/19938.

[6] M.N. Islam, I. Islam, K.M. Munim, A.K.M. Najmul Islam, A review on the mobile applications developed for COVID-19: An exploratory analysis, IEEE Access 4 (2020) 1, http://dx.doi.org/10.1109/access.2020.3015102.

[7] A. Kumar, P.K. Gupta, A. Srivastava, A review of modern technologies for tackling COVID-19 pandemic, Diabetes Metab. Syndr. Clin. Res. Rev. 14 (2020) 569–573, http://dx.doi.org/10.1016/j.dsx.2020.05.008.

[8] B. McCall, COVID-19 and artificial intelligence: protecting health-care workers and curbing the spread, Lancet Digit. Heal. 2 (2020) e166–e167, http://dx.doi.org/10.1016/S2589-7500(20)30054-6.

[9] S. Bobdey, S. Ray, Going viral – Covid-19 impact assessment: A perspective beyond clinical practice, J. Mar. Med. Soc. 22 (1) (2020) 9, http://dx.doi.org/10.4103/jmms.jmms_12_20.

[10] R. Stones, Structuration Theory, Palgrave-Macmillan, Basingstoke, 2005.

[11] T. Greenhalgh, R. Stones, Theorising big IT programmes in healthcare: Strong structuration theory meets actor-network theory, Soc. Sci. Med. 70 (2010) 1285–1294, http://dx.doi.org/10.1016/j.socscimed.2009.12.034.

[12] A.G. Nikolaev, R. Razib, A. Kucheriya, On efficient use of entropy centrality for social network analysis and community detection, Soc. Netw. 40 (2015) 154–162, http://dx.doi.org/10.1016/j.socnet.2014.10.002.

[13] S.M. Pincus, Approximate entropy as a measure of system complexity, Proc. Natl. Acad. Sci. USA 88 (1991) 2297–2301, http://dx.doi.org/10.1073/pnas.88.6.2297.

[14] S. Gregor, A.R. Hevner, Positioning and presenting design science research for maximum impact, MIS Q. 37 (2013) 337–355.

[15] A.R. Hevner, S.T. March, J. Park, S. Ram, Design science in information systems research, MIS Q. 28 (2004) 75–105.

[16] M.F. Alwashmi, The use of digital health in the detection and management of COVID-19, Int. J. Environ. Res. Public Health 17 (2020) http://dx.doi.org/10.3390/ijerph17082906.

[17] R. Verity, L.C. Okell, I. Dorigatti, P. Winskill, C. Whittaker, N. Imai, G. Cuomo-Dannenburg, H. Thompson, P.G.T. Walker, H. Fu, A. Dighe, J.T. Griffin, M. Baguelin, S. Bhatia, A. Boonyasiri, A. Cori, Z. Cucunubá, R. FitzJohn, K. Gaythorpe, W. Green, A. Hamlet, W. Hinsley, D. Laydon, G. Nedjati-Gilani, S. Riley, S. van Elsland, E. Volz, H. Wang, Y. Wang, X. Xi, C.A. Donnelly, A.C. Ghani, N.M. Ferguson, Estimates of the severity of coronavirus disease 2019: a model-based analysis, Lancet Infect. Dis. (2020) 669–677, http://dx.doi.org/10.1016/S1473-3099(20)30243-7.

[18] C. Bambra, R. Riordan, J. Ford, F. Matthews, The COVID-19 pandemic and health inequalities, J. Epidemiol. Community Health (2020) jech–2020–214401, http://dx.doi.org/10.1136/jech-2020-214401.

[19] J.B. Galloway, S. Norton, R.D. Barker, A. Brookes, I. Carey, B.D. Clarke, R. Jina, C. Reid, M.D. Russell, R. Sneep, L. Sugarman, S. Williams, M. Yates, J. Teo, A.M. Shah, F. Cantle, A clinical risk score to identify patients with COVID-19 at high risk of critical care admission or death: An observational cohort study, J. Infect. (2020) http://dx.doi.org/10.1016/j.jinf.2020.05.064.

[20] G. Anderson, J.W. Frank, C.D. Naylor, W. Wodchis, P. Feng, Using socioeconomics to counter health disparities arising from the covid-19 pandemic, BMJ 369 (2020) m2149, http://dx.doi.org/10.1136/bmj.m2149.

[21] D. Alahakoon, R. Nawaratne, Y. Xu, D. De Silva, U. Sivarajah, B. Gupta, Self-building artificial intelligence and machine learning to empower big data analytics in smart cities, Inf. Syst. Front. (2020).

[22] C.H. Sudre, K. Lee, M. Ni Lochlainn, T. Varsavsky, B. Murray, M.S. Graham, C. Menni, M. Modat, R.C.E. Bowyer, L.H. Nguyen, D.A. Drew, A.D. Joshi, W. Ma, C.G. Guo, C.H. Lo, S. Ganesh, A. Buwe, J. Capdevila Pujol, J. Lavigne du Cadet, A. Visconti, M. Freydin, J.S. El Sayed Moustafa, M. Falchi, R. Davies, M.F. Gomez, T. Fall, M.J. Cardoso, J. Wolf, P.W. Franks, A.T. Chan, T.D. Spector, C.J. Steves, S. Ourselin, Symptom clusters in Covid19: A potential clinical prediction tool from the COVID Symptom study app, MedRxiv (2020) http://dx.doi.org/10.1101/2020.06.12.20129056, 2020.06.12.20129056.

[23] L. Sun, Z. Mo, F. Yan, L. Xia, F. Shan, Z. Ding, B. Song, W. Gao, W. Shao, F. Shi, H. Yuan, H. Jiang, D. Wu, Y. Wei, Y. Gao, H. Sui, D. Zhang, D. Shen, Adaptive feature selection guided deep forest for COVID-19 classification with chest CT, IEEE J. Biomed. Heal. Inform. (2020) 1, http://dx.doi.org/10.1109/JBHI.2020.3019505.

[24] M. Nour, Z. Cömert, K. Polat, A novel medical diagnosis model for COVID-19 infection detection based on deep features and Bayesian optimization, Appl. Soft Comput. J. (2020) 106580, http://dx.doi.org/10.1016/j.asoc.2020.106580.

[25] Y. Ye, S. Hou, Y. Fan, Y. Zhang, Y. Qian, S. Sun, Q. Peng, M. Ju, W. Song, K. Loparo, α-Satellite: An AI-driven system and benchmark datasets for dynamic COVID-19 risk assessment in the United States, IEEE J. Biomed. Heal. Inform. XX (2020) 1, http://dx.doi.org/10.1109/jbhi.2020.3009314.

[26] N. Aydin, G. Yurdakul, Assessing countries' performances against COVID-19 via WSIDEA and machine learning algorithms, Appl. Soft Comput. J. 97 (2020) 106792, http://dx.doi.org/10.1016/j.asoc.2020.106792.

[27] A. Giddens, The constitution of society: Outline of the theory of structure, 1984.

[28] B.B. Hazarika, D. Gupta, Modelling and forecasting of COVID-19 spread using wavelet-coupled random vector functional link networks, Appl. Soft Comput. J. 96 (2020) 106626, http://dx.doi.org/10.1016/j.asoc.2020.106626.

[29] M. Mitchell, Complex systems: Network thinking, Artificial Intelligence 170 (2006) 1194–1212, http://dx.doi.org/10.1016/j.artint.2006.10.002.

[30] A. Vespignani, Predicting the behavior of techno-social systems, Science (80-.) 325 (2009) 425–428, http://dx.doi.org/10.1126/science.1171990.

[31] S. Gregor, L. Chandra Kruse, S. Seidel, Research perspectives: The anatomy of a design principle, J. Assoc. Inf. Syst. 21 (2020) 1622–1652, http://dx.doi.org/10.17705/1jais.00649.

[32] S.T. March, G.F. Smith, Design and natural science research on information technology, Decis. Support Syst. 15 (1995) 251–266.

[33] J. vom Brocke, A. Maedche, The DSR grid: six core dimensions for effectively planning and communicating design science research projects, Electron. Mark. 29 (2019) 379–385.

[34] Digital Health and Care NI, Digital Health and Care NI, 2020. https://play.google.com/store/apps/details?id=net.hscni.covid19ni&hl=en_US.

[35] C. Guo, L. Yang, X. Chen, D. Chen, H. Gao, J. Ma, Influential nodes identification in complex networks via information entropy, Entropy 22 (2020) 1–19, http://dx.doi.org/10.3390/e22020242.

[36] D. Zinoviev, Complex Network Analysis in Python: Recognize - Construct - Visualize - Analyze - Interpret, Pragmatic Bookshelf, 2018.

[37] M. Shanahan, B. Baars, Applying global workspace theory to the frame problem, Cognition 98 (2005) 157–176.

[38] J.H. Holland, Studying complex adaptive systems, J. Syst. Sci. Complex. 19 (2006) 1–8, http://dx.doi.org/10.1007/s11424-006-0001-z.

[39] M. Prokopenko, C. Gershenson, Entropy methods in guided self-organisation, Entropy 16 (2014) 5232–5241, http://dx.doi.org/10.3390/e16105232.

[40] T.N.I.S. and R.A. NISRA, NISRA Annual Report and Accounts, 2017, pp. 2017–2018.

[41] C.L. Niedzwiedz, C.A. O'Donnell, B.D. Jani, E. Demou, F.K. Ho, C. Celis-Morales, B.I. Nicholl, F. Mair, P. Welsh, N. Sattar, J. Pell, S.V. Katikireddi, Ethnic and socioeconomic differences in SARS-CoV2 infection in the UK Biobank cohort study, MedRxiv (2020) http://dx.doi.org/10.1101/2020.04.22.20075663, 2020.04.22.20075663.

[42] J. Venable, J. Pries-Heje, R. Baskerville, FEDS: A framework for evaluation in design science research, Eur. J. Inf. Syst. 25 (2016) 77–89.

[43] R. Baskerville, A. Baiyere, S. Gregor, A. Hevner, M. Rossi, Design science research contributions: Finding a balance between artifact and theory, J. Assoc. Inf. Syst. 19 (2018) 358–376, http://dx.doi.org/10.17705/1jais.00495.

[44] R. Burger, C. Christian, B. Maughan-Brown, R. Rensburg, L. Rossouw, COVID-19 risk perception, knowledge and behaviour in South Africa, 2020, http://dx.doi.org/10.13140/RG.2.2.22121.31842.

[45] C. Wenham, J. Smith, R. Morgan, COVID-19: the gendered impacts of the outbreak, Lancet 395 (2020) 846–848, http://dx.doi.org/10.1016/S0140-6736(20)30526-2.

[46] A.B. Brendel, T.B. Lembcke, J. Muntermann, L.M. Kolbe, Toward replication study types for design science research, J. Inf. Technol. (2021) http://dx.doi.org/10.1177/02683962211006429.

[47] R. Baskerville, J. Pries-Heje, Projectability in design science research, J. Inf. Technol. Theory Appl. 20 (2019) 53–76.