Toward Population Health Intelligence: When Artificial Intelligence Meets Population Health Research

Wang, J., Chen, L., Lycett, D., Vernon, D. & Zheng, D.

Original citation & hyperlink:
Wang, J, Chen, L, Lycett, D, Vernon, D & Zheng, D 2024, 'Toward Population Health Intelligence: When Artificial Intelligence Meets Population Health Research', Computer, vol. 57, no. 6, pp. 62-72.
https://dx.doi.org/10.1109/mc.2023.3283857

DOI 10.1109/mc.2023.3283857
ISSN 0018-9162
ESSN 1558-0814

Publisher: Institute of Electrical and Electronics Engineers

Towards Population Health Intelligence: When Artificial Intelligence Meets Population Health Research

Jiangtao Wang
Centre for Intelligent Healthcare, Coventry University, Coventry, United Kingdom, ad5187@coventry.ac.uk

Long Chen *
Centre for Intelligent Healthcare, Coventry University, Coventry, United Kingdom, ad8579@coventry.ac.uk

Deborah Lycett
Centre for Intelligent Healthcare, Coventry University, Coventry, United Kingdom, ab5042@coventry.ac.uk

Duncan Vernon
Warwickshire County Council and South Warwickshire University NHS Foundation Trust, Warwick, United Kingdom, Duncan.Vernon@swft.nhs.uk

Dingchang Zheng
Centre for Intelligent Healthcare, Coventry University, Coventry, United Kingdom, ad4291@coventry.ac.uk

***Abstract*—**
Artificial intelligence (AI), has been increasingly adopted to improve human's health and wellbeing, but the influential articles in this field mainly focus on the individual-level clinical predictions. To date there has been little consideration of AI to model health and disease at population level. To this end, this article aims to provide multidisciplinary researchers with a deep, comprehensive, and insightful vision of how AI can empower research in population health. Specifically, this article summarizes the state-of-the-art research agenda of AI-based population health in a logical way, we review how AI can be integrated to different tasks and stages of population health to solve unmet healthcare needs. We go on to describe our recent research project called Compressive Population Health (CPH) as a case study. Finally, we discuss opportunities of AI-empowered population health to inspire future research in this promising area.

Keywords: Artificial intelligence (AI), Environmental health, Epidemiology, Population health

1. Introduction

Artificial intelligence (AI), such as machine learning, has been increasingly adopted in health-

care, enabled by the availability of various types and modalities of digitalized health or health-related data. The combination of AI and Health Data Science generates a new vision called Health Intelligence[1], which can be further classified as applications at levels of the individual and the population. Individual-level health intelligence mainly focuses on the prediction and management of individual health and well-being, which is referred to as personal health intelligence, while population-level health Intelligence aims to understand and promote public health through multiple disciplines including environmental health, epidemiology, social sciences, and economics, which is referred to as Population Health Intelligence here.

Take the topic of deep learning for Electronic Health Record (EHR) analysis as an example of personal health intelligence, there are a lot of surveys, vision, and position papers such as[2,3], which make a very comprehensive summary and insightful outlook of its research landscape. For example, diverse deep learning methods and frameworks have been applied to various clinical applications, which includes information extraction, outcome forecasting, and phenotyping. However, although as important as personal health intelligence, there are relatively much fewer literature reviews or vision articles to investigate how AI can empower the research agenda of population health. The research paper including [4,5,6] do focus on the population health intelligence, but they mainly list some interesting applications on AI health inequalities and social determinants with an absence of a deeper exploration of the key technologies and their relationship with unmet needs in population health research such as epidemiology in Noncommunicable Diseases, A narrative review [7] focuses on the combination of AI and geospatial techniques on population health, for example Infectious diseases, environmental health, and genetics but this only shows part of the potential of population health intelligence.

The core contribution of the article lies in presenting a first-of-its-kind in-depth view point regarding how AI can be effectively adopted in the context of population health: (1) we provide multidisciplinary researchers with insight into the state-of-the-art research agenda of population health intelligence where AI can be integrated to different stages/tasks of population health, and to solve which category of unmet needs (**Section 3**). (2) We Introduce one of our recent research projects called Compressive Population Health (CPH), showcasing it as a case study with joint consideration of different stages of population health intelligence (**Section 4**). (3) We describe the existing gaps and opportunities in this area with ideas to inspire future research that crosses multidisciplinary boundaries of public health and AI. (**Section 5**). Here, please note that this is a vision article rather than a comprehensive survey paper.

2. Population Health: Preliminaries

To provide AI researchers with basic landscape of population health research, the following terms regarding different sub-disciplines and research tasks are introduced as preliminaries.

Population Health has been defined as "the health outcomes of a group of individuals, including the distribution of such outcomes within the group" [8] , which aims to improve the health of an entire human population. It has been described as consisting of three components. These are "health outcomes, patterns of health determinants, and policies and interventions".

The research discipline for population health is also named as Epidemiology, that is, the branch of medicine which studies the distribution and determinants of health-related states among specified populations and the application of that study to the control of health problems. Epidemiologists are public health researchers who analyze population health data (for instance, to study how often diseases occur in different groups of people / different areas, and why), and monitor disease progression (e.g., infectious diseases such as COVID-19).

There are also several terms regarding the sub-disciplines of population health, including environmental health, spatial epidemiology, health geography, and so on. Environmental health is the branch of public health concerned with all aspects of the natural and built environment affecting human health. Environmental health focuses on the natural and built environments for the benefit of human health. Spatial epidemiology is a subfield of epidemiology focused on the study of the spatial distribution of health outcomes.

Specifically, spatial epidemiology is concerned with the description and examination of disease and its geographic variations. Health geography is the application of geographical information, perspectives, and methods to the study of health, disease, and health care. From the above definitions, spatial epidemiology and health geography are more similar, where the location is an important index within the study, while the scope of environmental health is wider.

The lifecycle of population health research usually consists of three sequential stages/tasks: (1) population health data collection (health outcome profiling): collect health or health-related data from a target group of population, and then pre-process the data to form the required profile for study; (2) population data analytics (modeling and understanding): analysis on the collected data to model the pattern of certain health outcomes so that a understanding of knowledge can be obtained. (3) population health intervention: stakeholders including public health administrators, epidemiologists, health policy makers, and built environment planners, co-develop evidence-based strategies to improve public health and wellbeing.

3. AI-Empowered Population Health: Research Agenda

When talking about the adoption of AI in the certain domains, we usually refer to the tasks such as health data analytics and prediction. However, for population health research, the power of AI has already been integrated in the full lifecycle. In this section, we will provide an overview research agenda of AI-empowered population health based on a stage-aware and challenge-driven structure, where we will introduce the key challenges/needs in each stage and the how the AI technologies can be embedded to tackle them.

3.1. AI for Population Health Data Collection

As Fig 1 shows, conventional population health (or health-related) data collection mainly relies on specific staff, surveys, and devices, which is of high cost in terms of the consumed time and money. Moreover, because of the expensive cost and uncertainty of participation willingness, the spatial-temporal coverage or scale of the data collection and profiling is usually very limited (or with missing records), and moreover, the obtained health profiles are usually coarse-grained. Other forms of surveillance include disease registers, which either aggregate information from multiple sources impacting on timeliness, or are dependent on patients interacting with health professions to receive a diagnosis. These barriers in data collection inevitably hinders further knowledge extraction and understanding, especially in understanding inequalities in health outcomes.

In recent years, the combination of AI and other key technologies, such as Internet of Things (IoT), cloud computing, mobile social network, etc., has jointly enabled a new paradigm for crowdsourced population health data collection (see Fig 5). For example, urban green space may be important to mental health, but the association between long-term green space exposures and depression, anxiety, and cognitive function in adults remains unknown. To address current limitations regarding the lack of green space measurements at the highly granular street scale, one recent study in [9] calculated the street-level measures from Google Street View images in Portland, Oregon, US. Another typical example is the collection of a multitude of data sources for outbreak detection, including electronic health records and non-traditional public health data sources such as Twitter feeds [10] .

3.2. AI for Population Health Data Analytic

After the population health data have been collected and preprocessed (cleaned, linked, and calibrated), health data scientists will move the analytic phase, which aims to extract knowledge for better understanding about the public health measures and its determinants. There are two primary types of data analytic tasks for population health research: cause-and-effect analysis and predictive analysis.

Cause-and-effect Analysis. This has overlaps with the developing field of Population Health Management. A key task of population health research is to investigate a potential cause-and-effect relationship between the health outcome and its possible determinants. The main challenge of this type of task is the existence of confounding variables, that are, unmeasured variables that

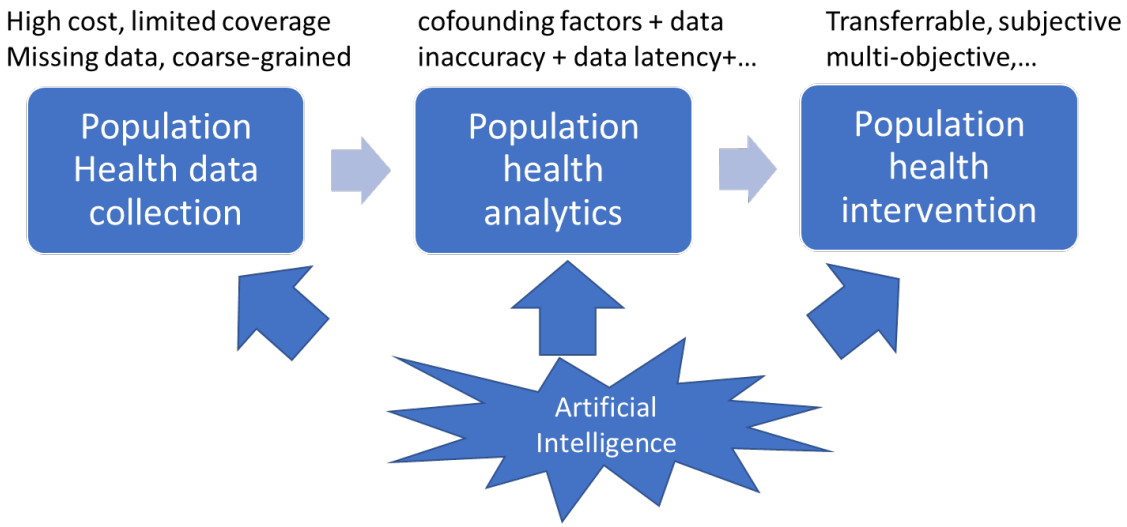

Figure 1: Integration of AI in multiple stages of population health

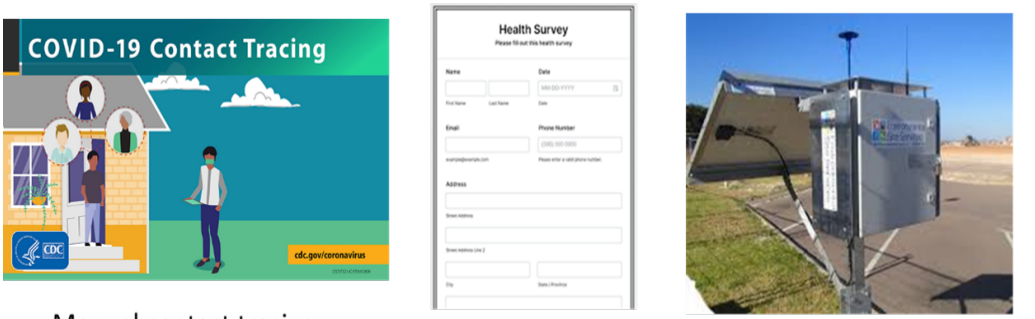

Figure 2: Population health data collection: conventional ways

influence both the supposed cause and the supposed effect. It's important to consider potential confounding variables and account for them in your research design to ensure the results are valid. The most used approach is to restrict your treatment group by only including subjects with the same or at least similar values of potential confounding factors. The study about the effectiveness of lockdown measures on the control of pandemic [11] is an example using this simple approach, where other confounding factors, such as density of population and aging features, have been successfully controlled. However, the problem of this simple method is that it will significantly reduce the number of samples, so that the qualified group may not be representative. Alternatively, you can select a comparison group that matches with the treatment group. For example, the authors in [12] studies the effect of sport venue presence on the prevalence of antidepressant prescriptions in over 600 neighborhoods in London over a

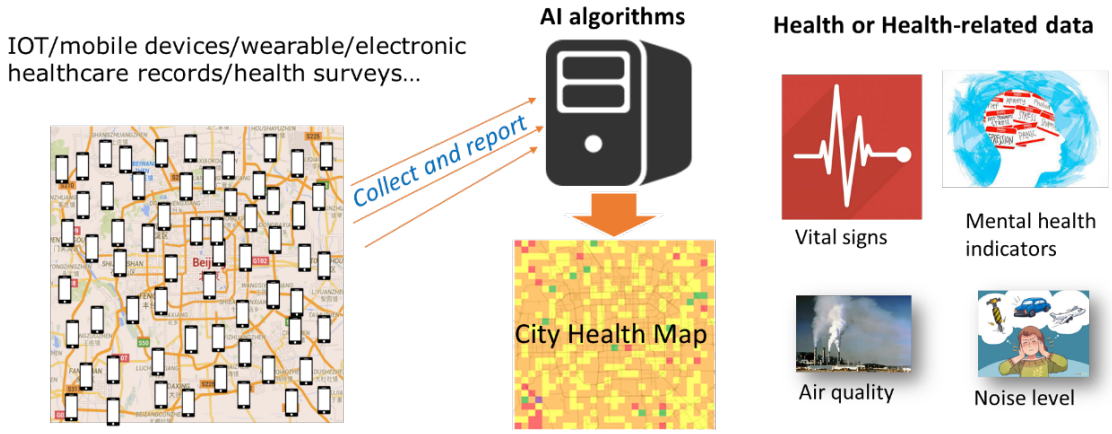

Figure 3: Crowdsourced population health data collection

period of three years, and find the distribution of effects is approximately normal, centered on a small negative effect on prescriptions with increases in the availability of sporting facilities, on average. The key technology of this research is how to construct the treated and control group by considering confounding factors including demographics (population structure), green space, and social-economical determinants such as housing benefit claims and job seekers allowance. To achieve this, the wards are matched in such a way that minimizes the difference between the confounding variables over all matched pairs and maximizes the difference between the treatment 'dose' e.g., the binned value of available sport venues in a ward, in each pair.

Predictive Analysis. an increasing number of machine learning algorithms (especially deep learning, including CNN, LSTM, RNN, etc.) have been proposed to predict/infer population-level health outcome such as the prevalence of diseases, ranging from non-communicable diseases to infectious diseases. For example, the authors in [13] use users' posts or tweets on social networks to predict large-scale flu evolution patterns. The work in [14] uses the population mobility patterns of metropolitan area residents to predict the prevalence of several chronic diseases in urban neighborhoods by looking at local human lifestyles. Rather than using traditional missing data recovery approaches for spatial data (e.g., interpolation, k-nearest neighbors, and matrix completion) which mainly focus on linear and single-view data correlation, the technical trend of these recent work is to construct a deep learning network to utilize correlations from multiple views, including both temporal and spatial ones. the temporal view aims to model the correlations between measures with near time periods via models such as Long Short-Term Memory (LSTM), while the spatial view characterizes local spatial correlation with neighboring grids via Convolutional Neural Network (CNN). Beyond general predictive analysis models, researchers recently also focus on some specific challenges closely linked to population health data and propose corresponding solutions to address them. One is the data inaccuracy, where integrating data from personal devices with healthcare services requires accurate and reliable data that can be used to make sound policy decisions. Personal health devices, including devices for in-home use, are well-known to be susceptible to errors unlike clinical equipment or environments. AI-based calibrations techniques that compensate these errors and increase validity of measurements can improve accuracy of the measurements [15] . Another challenge is the data latency issue, where population-level disease prediction estimates the number of potential patients of particular diseases in some location at a future time based on (frequently updated) historical disease statistics. Existing approaches often assume the existing disease statistics are reliable and will not change. However, in practice, data collection is often time-consuming and has time delays, with both historical and current disease statistics being

updated continuously. To address this challenge, research work such as [16] propose population-level disease prediction model which captures data latency and incorporates the updated data for improved predictions. Specifically, it models real-time data and updated data using two separate systems, each capturing spatial and temporal effects using hybrid graph attention networks and recurrent neural networks. Then it fuses the two systems using both spatial and temporal latency-aware attentions in an end-to-end manner. Uses of predictive analytics to compare observed and expected patterns of disease also have applications in identifying health inequalities, or cohorts of patients likely to have undiagnosed illness. More effective stratification of patients into risk categories will also have benefits for targeting other proactive healthcare interventions or predicting future healthcare costs, which has been systemetically reviewed in [41].

3.3. AI for Population Health Intervention

AI has also been adopted in population health intervention in recent years. As the built and natural environment is a key environmental determinant of health and wellbeing, one typical aspect of population health intervention is spatial planning for health. In other words, the considerate design of spaces and places can help to promote good health; access to goods and services; and alleviate, or in some cases even prevent, poor health thereby having a positive impact on reducing health inequalities. Although there is a multitude of guidance supporting and advocating action on the built and natural environment to improve health outcomes, the evidence base underpinning these principles is still a matter of debate amongst the scientific and the practitioner communities. The subjective and individual nature of the built and natural environment make it difficult to develop evidence-informed approaches that can be universally applied, and successful practices in one community setting may not always be transferrable to another. To this end, recent studies such as [17] formulate the problem of spatial planning for health as a multi-objective optimization for spatial planning, and then develop machine learning and artificial intelligence approaches to address spatial planning issues. These models simulate the decision-making processes of multiple stakeholders in typical urban planning tasks such as land use allocation, the machine learning approach obtains the nonlinear behavioral rules of land use, and the artificial intelligence approach provides a flexible optimization framework that can incorporate agents' preferences.

More recently, some studies [38,39,40] proposed new health intervention scenarios in the advent of the 5G network, where they examined, categorized, grouped, and classified methods such as radio-frequency fingerprinting, mutual authentication, and IoT-5G device authentication.

4. Case Study: Compressive Population Health

For the current landscape of AI-empower population health research, they mainly integrated AI separately in each of the stages of population health. In this section, we will introduce our recent study of AI-based population health, which can be characterized as one of the pioneering works jointly considering the stages of health data collection and prediction in an end-to-end model.

4.1. Motivation and basic vision

Non-communicable diseases (NCDs), such as hypertension, diabetes, and obesity, are major causes of death globally, particularly in developed countries [22]. Uncovering the hidden patterns of NCDs is critical for health authorities to understand and address these issues. We can traditionally accomplish this by conducting population health surveillance in traditionally sensed-areas (TS-A), which refers to the profiling process that measures the health statistics of a population in the target area. However, it is a difficult task because private health data is sensitive, difficult to obtain, and often entails a high operating expense. Specifically, data protection[35], data anonymization[36], and data augmentation[37] are key barriers when dealing with sensitive healthcare data, such as medical records, clinical trials data, and genetic information. Data protection measures for sensitive healthcare data include implementing access controls, such as user authentication and role-based access control, to limit access to only authorized personnel. Encryption and other security measures can also be used to ensure that the data is protected in transit and at rest. Data anonymization techniques for healthcare

data include de-identification and pseudonymization. De-identification involves removing or altering any data elements that could potentially identify an individual, such as names, addresses, and Social Security numbers. Pseudonymization involves replacing identifying information with a pseudonym, which can be used to link different data sources without revealing the identity of the individual. Both techniques can help protect individual privacy while still allowing for the use of the data for research and other purposes. Data augmentation techniques for healthcare data can be used to generate synthetic data that can be used for training machine learning models. This can be particularly useful when dealing with small datasets, as it can help increase the size and diversity of the data without compromising individual privacy. Synthetic data can be generated using techniques such as generative adversarial networks (GANs) and variational autoencoders (VAEs).

As a result, the collected data is often reduced to a limited spatial coverage. Therefore, as shown in Fig. 4, our primary goal is to present an AI-based health paradigm called Compressive Population Health Profiling (CPHP), which aims to reduce the effort required for traditional prevalence profiling while maintaining data quality. Before a task starts, we need to determine fine-grained regions & surveyor recruitment. Then, each disease selects certain regions for profiling, and the surveyor uploads the corresponding prevalence. Lastly, CPH uses the TS-A data to impute the un-selected regions (called Inferred Areas, IF-A) by exploiting both intra-disease and inter-disease correlations. The use of leave-one-out bootstrapping ensures the quality of the imputed data.

4.2. Approaches

In this section, we present Compressive Population Health Profiling (CPHP), a deep active learning algorithm that seamlessly combines data imputation and active learning. CPHP is based on a cutting-edge data imputation algorithm known as Compressive Population Health (CPH) [20] , which uses epidemiology knowledge to exploit both inter-disease and intra-disease correlations among multiple NCDs (e.g., obesity, diabetes and hypertension), greatly reducing the profiling cost. In contrast to the classic Generative Adversarial Network, CPH is based on the Missing Data Imputation approach, particularly, Generative Adversarial Imputation Nets (GAIN) [21] . The intra-disease correlation refers to the adjacency effect on NCDs, in which diseases from neighbouring regions are more likely to have a similar prevalence rate than diseases from distant regions. The underlying logic of this type of correlation is that adjacent regions frequently have a similar demographic distribution. The inter-disease correlations, specifically multi-morbidity [19] , which is the co-occurrence of multiple NCDs. For example, areas with high obesity rates are more likely to have high hypertension rates as well[23] .

Furthermore, in the profiling process, a Bayesian active learning scheme [18] is used to intelligently select the top TS-A regions with the highest uncertainty. Fig. 5 shows the overall architecture. In each cycle, CPHP selects the next salient TS-A for profiling and waits for the oracles to obtain the results in the target hospitals present in that TS-A. This process will be iteratively occurred until the estimated data quality satisfies the predefined error bound requirement. Then, the task allocation formally ends, and the missing data are filled with CPH in one profiling cycle.

4.3. Experimental results and findings

We evaluate the proposed methods using two real-world datasets, Ward Boundaries of London and Chronic Diseases Prevalence, both of which do not require any licences to access. Moreover, both datasets comply with the GDPR regulations for their processing, which includes implementing adequate security measures, e.g., data anonymization and obtaining the required authorizations from national authorities. The former is supplied by the UK's mapping agency with the government's arguably most accurate geographical data. The dataset includes 630 London wards, as well as their specific names, shapes, and codes. The latter was downloaded from the National Health Service between 01/04/2008 and 31/03/2017, and it contains three types of diseases: obesity, diabetes, and hypertension. The prevalence of each type is given as a percentage of all patients on the practise list. The two dataset are linked in terms of the location id at ward level. We use historical

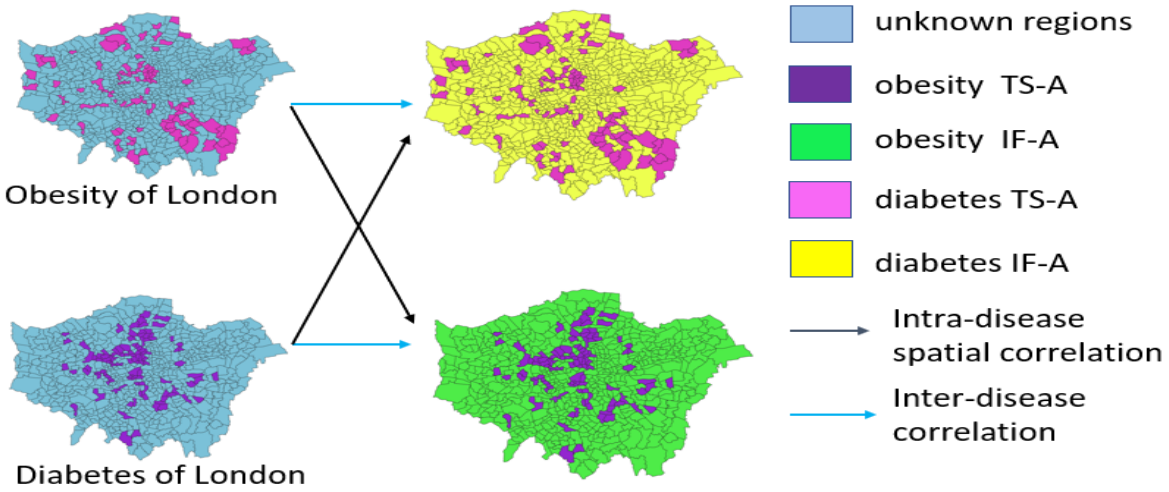

Figure 4: The basic vision of CPHP

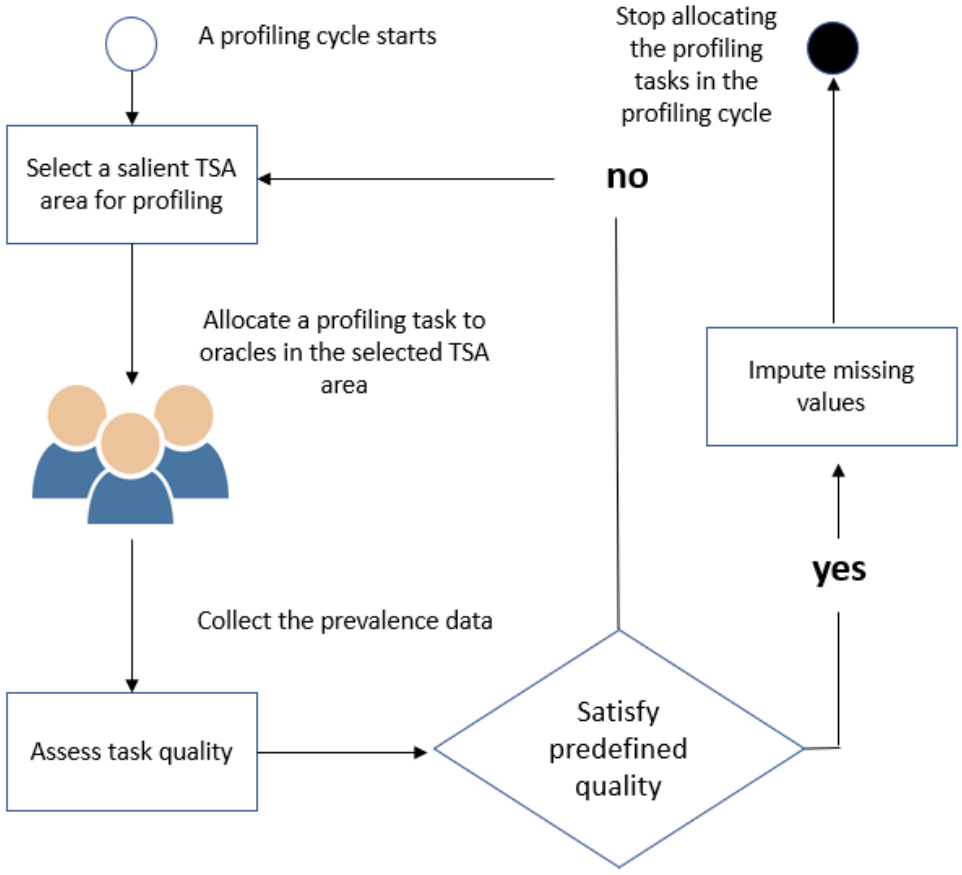

Figure 5: The Overall Workflow of CPHP

training data from 2008 to 2014 and test data from 2015 to 2017. For example, if we set the current year to 2015, we can test the performance of active learning in 2015 using data from 2008 to 2014. If the current year is set to 2016, we compare 2016 performance to data from 2008 to 2015.In addition, we compare CPHP with the following baselines:

- Entropy-based sampling strategy: The model assumes that higher values of entropy correspond to greater uncertainty.
- Random sampling strategy: The model that first shuffle all the data points, and then the acquired dataset is randomly drawn from a uniform distribution.

Table 1 shows the results of various active learning models for the prevalence rate deduction of diabetes in 2017. Note that the error bound is set at the levels of 10% and 5% respectively. The minimal proportion of sampling regions that meet the error bound criteria is used as a measure of performance. This metric's lower value indicates a lower cost of profiling with the same level of performance. It is clear that CPHP supersedes the other baselines because of its cutting-edge active learning technique and higher convergence speed, which in turn decreases the profiling cost significantly. Specifically, CPHP requires only 12.4 percent of the entire region to be sampled, whereas an entropy-based sampling system and a random sampling system require 14.9 percent and 15.8 percent of the entire region correspondingly to achieve the same error constraint. Even though the required sampling percentage is only reduced by 2.4% and 3.6%, it is really reduced by 19.8% and 26.2% relative to the baselines, which we claim is a substantial improvement. In a nutshell, CPHP that uses a Bayesian active learning technique to make considerable improvements over previous baseline algorithms, and the introduction of CPHP has an impact on enhancing health profiling practise by reducing costs and time for collecting data.

5. Conclusion and Future opportunities

In this section, we highlight the research gaps and future opportunities for AI-empowered population health, which may lead to novel solutions in this increasingly important field.

Uncertainty of Correlations. Motivated by the First Law of Geography: 'near things are more related than distant things', epidemiologists and health data scientists have found that many diseases also have spatial correlation, that is, the disease prevalence tends to be more similar between neighboring regions than distant ones. However, according to our recent findings in[24] , the First Law of Geography sometimes does not work for many diseases, where two girds could be spatially distant but are similar in their prevalence evolution pattern. For example, grid A is more distant from B than C, but the prevalence rates between A and B might be more similar than that between A and C. In this case, relying on the spatial correlation will undermine the data reconstruction accuracy. Besides, the correlations may be altered from time to time. Therefore, it is an urgent need to develop more robust population health inference/prediction algorithms to deal with uncertain correlations.

Interpretability. Although deep learning models are superior in many population health data analysis tasks compared to conventional Interpolation methods such as linear regression, they suffer from a drawback. That is, data consumers such as public health administrators and epidemiologists will hesitate to trust the predicted/inferred measures and take corresponding intervention actions, because the model only tells them about the inferred measure without giving explanations on why the decision has been made. Therefore, providing a certain level of explainability for the proposed model is also an important requirement, that is, our approaches need to provide a capability to enable data consumers to understand the evidence of the decisions made by AI models. One possible AI solution for this is mimic learning, where we can use a simple but interpretable model (e.g., multivariate regression or a shallow neural network model) to mimic a complex deep learning model.

Causal Representation Learning. The two fields of machine learning and causal modeling and are developed separately in the area of population health intelligence. However, there is, now, cross-pollination and increasing interest in both fields to benefit from the advances of the other. Most machine-learning-based population health prediction tasks focus on the predicted outcomes

Error bound	Entropy-based Sampling	Random Sampling	CPHP
10%	15.3%	14.5%	12.1%
5%	17.7%	17.3%	13.7%

Table 1: The performance of varying active learning models for diabetes in 2017

rather than understanding spatial or temporal causality, which are limited in their ability to generalize the patterns they find in a training data set for real-world public health practice. The ability to uncover the causes and effects of different phenomena in complex public health problems would help us build better solutions.

Decentralized Machine Learning. Existing AI-based population health data analysis commonly assume that health or health-related data from different regions/systems/stakeholders will be processed and analyzed in a centralized server. However, because of the region-dependent data protection and security regulation, this assumption may not always be satisfied. Take our case study of compressive population health as an example, the data collection of different diseases may not be handled by the same authority, and different data profiling organizers may not all be cooperative to upload the data to a centralized repository. Therefore, we need to develop an aggregation-free approach for the spatial-temporal health data analytics on the basis of AI technologies such as federated learning and transfer learning.

Multimodality and Multiview Learning. We can see that multiple sources of data have been associated with population health outcome, even for some new sources of data that public health researchers have not used before (e.g., citizens' mobility). However, in terms of the predictive analysis task, existing research mainly exploit the data with single modality. A possible research direction is to exploit multiple modalities at the same time with a multi-view and multi-modality learning strategy. Take the prediction of Covid-19 pandemic as an example, we may exploit both textual data on the web (e.g., social media) and sensory data (human mobility) at the same time to maximize their complementary power.

AI and Precision Public Health. The vision of precision public health is that using better public health surveillance, laboratory investigations and geo-spatial modelling may allow more precise targeted population health interventions. It is essentially about delivering the right intervention at the right time, to the right population. To realize this vision, AI will play a very important role across multiple stages including health data collection, health data analytic, and health intervention. The above research of compressive population health is a typical example of how AI can be beneficial for reducing the cost of public health profiling by jointly considering both data collection and analytic phases. However, research in this area is still very insufficient, and many key applications and problems need to be studied urgently. For example, the majority of the mosquito-related infectious disease burden worldwide (e.g. Dengue, Chikungunya) could be addressed by focusing on the high risk geographic areas. We can combine AI and geo-spatial technology to accurately identify theses areas and then health interventions can be designed and implemented.

ACKNOWLEDGMENTS This work was supported by EPSRC New Investigator Award under Grant No EP/V043544/1.

References

[1] Shaban-Nejad, Arash, Martin Michalowski, and David L. Buckeridge. "Health intelligence: how artificial intelligence transforms population and personalized health." NPJ digital medicine 1.1 (2018): 1-2. (Reference of two types of health intelligence)

[2] Benjamin, et al. "Deep EHR: a survey of recent advances in deep learning techniques for electronic health record (EHR) analysis." IEEE journal of biomedical and health informatics 22.5 (2017): 1589-1604.

[3] Miotto, Riccardo, et al. "Deep learning for healthcare: review, opportunities and challenges." Briefings in bioinformatics 19.6 (2018): 1236-1246.

[4] Panch, Trishan, et al. "Artificial intelligence: opportunities and risks for public health." The Lancet Digital Health 1.1 (2019): e13-e14.

[5] Benke, Kurt, and Geza Benke. "Artificial intelligence and big data in public health." International journal of environmental research and

public health 15.12 (2018): 2796.

[6] Thiébaut, Rodolphe, and Frantz Thiessard. "Artificial intelligence in public health and epidemiology." Yearbook of medical informatics 27.01 (2018): 207-210.

[7] Kamel Boulos, Maged N., Guochao Peng, and Trang VoPham. "An overview of GeoAI applications in health and healthcare." International journal of health geographics 18.1 (2019): 1-9.

[8] Kindig D, Stoddart G (March 2003). "What is population health?". American Journal of Public Health. 93 (3): 380–3.

[9] Larkin A, Hystad P: Evaluating street view exposure measures of visible green space for health research. J Expo Sci Environ Epidemiol 2018.

[10] Osborne, Matthew T., et al. "Catch the tweet to fight the flu: Using Twitter to promote flu shots on a college campus." Journal of American College Health (2021): 1-15.

[11] K. Li, et al. "Does Our Collective Stringency Control the Virus? Investigating Lockdown Effectiveness on Community Mobility Data," 2021 IEEE 45th Annual Computers, Software, and Applications Conference (COMPSAC), 2021, pp. 608-617

[12] Hasthanasombat, Apinan, and Cecilia Mascolo. "Understanding the effects of the neighbourhood built environment on public health with open data." The World Wide Web Conference. 2019.

[13] S. Verma, Y. Park, and M. Kim, "Predicting flu-rate using big data analytics based on social data and weather conditions," Advanced Science Letters, 2017

[14] Wang, Yingzi, et al. "Predicting the Spatio-Temporal Evolution of Chronic Diseases in Population with Human Mobility Data." IJCAI, 2018.

[15] Benke, Kurt, and Geza Benke. "Artificial intelligence and big data in public health." International journal of environmental research and public health 15.12 (2018): 2796.

[16] Gao, Junyi, et al. "PopNet: Real-Time Population-Level Disease Prediction with Data Latency." arXiv preprint arXiv:2202.03415 (2022).

[17] Ding, Xiaoe, Minrui Zheng, and Xinqi Zheng. "The application of genetic algorithm in land use optimization research: A review." Land 10.5 (2021): 526.

[18] Houlsby N, Huszár F, Ghahramani Z, Lengyel M. Bayesian active learning for classification and preference learning. arXiv preprint arXiv:1112.5745. 2011 Dec 24.

[19] Prados-Torres A, Calderón-Larrañaga A, Hancco-Saavedra J, Poblador-Plou B, van den Akker M. Multimorbidity patterns: a systematic review. Journal of clinical epidemiology. 2014 Mar 1;67(3):254-66.

[20] Feng Y, Wang J, Wang Y, Helal S. Completing Missing Prevalence Rates for Multiple Chronic Diseases by Jointly Leveraging Both Intra-and Inter-Disease Population Health Data Correlations. InProceedings of the Web Conference 2021 2021 Apr 19 (pp. 183-193).

[21] Yoon J, Jordon J, Schaar M. Gain: Missing data imputation using generative adversarial nets. InInternational conference on machine learning 2018 Jul 3 (pp. 5689-5698). PMLR.

[22] Habib SH, Saha S. Burden of non-communicable disease: global overview. Diabetes and Metabolic Syndrome: Clinical Research and Reviews. 2010 Jan 1;4(1):41-7.

[23] Prados-Torres A, Calderón-Larrañaga A, Hancco-Saavedra J, Poblador-Plou B, van den Akker M. Multimorbidity patterns: a systematic review. Journal of clinical epidemiology. 2014 Mar 1;67(3):254-66.

[24] Williams, P.T., 1997. Relationship of distance run per week to coronary heart disease risk factors in 8283 male runners: the National Runners' Health Study. Archives of internal medicine, 157(2), pp.191-198.

[25] Wang, Zhiyuan, et al. "From personalized medicine to population health: a survey of mHealth sensing techniques." IEEE Internet of Things Journal (2022).

[26] Vyas, Sonali, et al. "Integration of artificial intelligence and blockchain technology in healthcare and agriculture." Journal of Food Quality 2022 (2022).

[27] Alanazi, Abdullah. "Using machine learning for healthcare challenges and opportunities." Informatics in Medicine Unlocked (2022): 100924.

[28] Herath, H. M. K. K. M. B., and Mamta Mittal. "Adoption of artificial intelligence in smart cities: A comprehensive review." Interna-

tional Journal of Information Management Data Insights 2.1 (2022): 100076.

[29] Abubakar, Ibrahim, et al. "The Lancet Nigeria Commission: investing in health and the future of the nation." The Lancet 399.10330 (2022): 1155-1200.

[30] Vyas, Sonali, et al. "Integration of artificial intelligence and blockchain technology in healthcare and agriculture." Journal of Food Quality 2022 (2022).

[31] Murphy, Kathleen, et al. "Artificial intelligence for good health: a scoping review of the ethics literature." BMC medical ethics 22.1 (2021): 1-17.

[32] Mhasawade, Vishwali, Yuan Zhao, and Rumi Chunara. "Machine learning and algorithmic fairness in public and population health." Nature Machine Intelligence 3.8 (2021): 659-666.

[33] Wahl, Brian, et al. "Artificial intelligence (AI) and global health: how can AI contribute to health in resource-poor settings?." BMJ global health 3.4 (2018): e000798.

[34] Chen, Mei, and Michel Decary. "Artificial intelligence in healthcare: An essential guide for health leaders." Healthcare management forum. Vol. 33. No. 1. Sage CA: Los Angeles, CA: SAGE Publications, 2020.

[35] Kwon, Juhee, and M. Eric Johnson. "Health-care security strategies for data protection and regulatory compliance." Journal of Management Information Systems 30, no. 2 (2013): 41-66.

[36] Haddow, Gill, Ann Bruce, Shiva Sathanandam, and Jeremy C. Wyatt. "'Nothing is really safe': a focus group study on the processes of anonymizing and sharing of health data for research purposes." Journal of evaluation in clinical practice 17, no. 6 (2011): 1140-1146.

[37] Tan, Xuyan, Xuanxuan Sun, Weizhong Chen, Bowen Du, Junchen Ye, and Leilei Sun. "Investigation on the data augmentation using machine learning algorithms in structural health monitoring information." Structural Health Monitoring 20, no. 4 (2021): 2054-2068.

[38] Kumar, NK Senthil, V. Dhilip Kumar, M. Kavitha, Fayadh Alenezi, Kemal Polat, Adi Alhudhaif, and Majid Nour. "Implications of 5G Network on IoT-based Healthcare Systems Using Deep Learning: A Comprehensive Review." (2022).

[39] Peralta-Ochoa, Angélica M., Pedro A. Chaca-Asmal, Luis F. Guerrero-Vásquez, Jorge O. Ordoñez-Ordoñez, and Edwin J. Coronel-González. "Smart Healthcare Applications over 5G Networks: A Systematic Review." Applied Sciences 13, no. 3 (2023): 1469.

[40] Sodhro, Ali Hassan, Ali Ismail Awad, Jaap van de Beek, and George Nikolakopoulos. "Intelligent authentication of 5G healthcare devices: A survey." Internet of Things (2022): 100610.

[41] Rajpurkar, Pranav, Emma Chen, Oishi Banerjee, and Eric J. Topol. "AI in health and medicine." Nature medicine 28, no. 1 (2022): 31-38.

Author Bios

- **Jiangtao Wang** is an associate professor at Coventry University, Coventry, CV1 5RW, U.K. His research interests include AI, data science, and digital health. Dr Wang received a Ph.D. in computer science from Peking University. Contact him at jiangtao.wang@coventry.ac.uk.
- **Long Chen** is a lecturer of data science at Coventry University, Coventry, CV1 5RW, U.K. His research interests include Intelligent Healthcare and Cost-effective AI. Dr chen received a Ph.D. in computer science from the University of London. Contact him at ad8579@coventry.ac.uk.
- **Deborah Lycett** is Director of the Research Institute for Health and Wellbeing. Her research is in both the clinical and public health context, focusing on improving well-being and life quality of individuals living with dietary and nutrition-related conditions. Her approach is wholistic exploring biobehavioural, psychological, social, spiritual and digital solutions. Professor Lycett is a Registered Dietitian with a PhD in Behavioural Medicine, she is a member of the British Dietetic Association, on the Scientific Committee for European Conference of Religion, Spirituality and Health, and is an associate Non-executive Director for a National Health Service Trust. Deborah.Lycett@coventry.ac.uk
- **Duncan Vernon** is a Consultant in Public Health, jointly employed by South Warwickshire Universities Foundation Trust and War-

wickshire County Council. He is a Fellow of the Faculty of Public Health and also holds am MSc in Engineering Systems. Duncan's research interests are about the practical application of research evidence, and evaluation methodologies – as well as how data and intelligence can be used to promote system change to improve health and wellbeing. Duncan.Vernon@swft.nhs.uk

- **Dingchang Zheng** received the B.Eng. degree in Biomedical Engineering from Zhejiang University, Hangzhou, China, and the Ph.D. degree in Medical Physics from Newcastle University, Newcastle upon Tyne, U.K. He is currently a Professor of Healthcare Technology and the Director Of Research Center for Intelligent Healthcare with Coventry University, Coventry, U.K. He is leading research in innovative healthcare technology and solutions development with intelligent physiological measurements and advanced bio-signal processing.