Achievements, Milestones and Challenges in Biomedical and Health Informatics
J. Mantas et al. (Eds.)

doi:10.3233/SHTI220949

Temporal Phenomics – A Powerful Approach Using AI to Achieve "Earlier Medicine"

Yu-Chuan (Jack) LI[a,1]
[a] *Graduate Institute of Biomedical Informatics, Taipei Medical University, Taiwan*

Abstract. The resurgence of machine learning AI has triggered the importance of collecting "personal big data" over a long period of time from wearable devices and EHRs. Collecting data from this large number of variables over a significant period of time has further induced the study on "Temporal Phenomics", which can be a powerful approach to achieve pre-emptive and "earlier medicine". The paper presents a methodology to make studying "Temporal Phenomics" more feasible and convenient without limitations on the number of variables and the length of time periods.

Keywords. Decision making, temporal phenomics, artificial intelligence

1. Background

Machine learning (ML) has been a very hot topic in all research fields equipped with large amounts of data. ML-based Artificial Intelligence (AI) resurged as the most promising approach to solve difficult problems after the rise of Internet in the 21st century which, in turn, accumulated immense amounts of user data (and hence, Big Data). In order to store and process big data, cloud, high-speed network and high-density computing units like GPU (Graphic Processing Unit) followed suit and evolved with the speed of "Moore's law" [1]. All these technology advancements led us into an age of democratized "deep learning", where everyone can harness extremely complex and powerful ML tools and use them to process, analyze and "learn" from very large data sets with billions of tuples and hundreds/thousands of dimensions without having to worry about the intricate and expensive infrastructure. We are in a brave new world that almost anything is possible as long as you know a little about ML and have access to enough data.

Like every new and exciting technology, people want to use it to solve the eternal problems of human health. Many of the medical problems are high-value and also highly complicated since we do not fully understand how the human body actually works. Furthermore, we do not know how to properly manage the individual differences of metabolism and patho-physiology and why everyone is distinctively different on things

[1] Corresponding author, Yu-Chuan (Jack) Li, M.D., Ph.D., FACMI, FACHI, FIAHSI Distinguished Professor, Graduate Institute of Biomedical Informatics, Taipei Medical University, Taiwan, Dermatology Dept, Wan Fang Hospital, President, International Medical Informatics Association (IMIA), Editor-in-Chief, BMJ Health and Care Informatics; E-mail: jack@tmu.edu.tw.

like drug-response and disease risk. The conception of "precision medicine" intends to mitigate the individuality problem in medical practices [2]. However, in order to really address the problem, we will need a way to capture all the personal health data elements (including genome, microbiome, environmental factors and those from the EHRs and wearable devices) over a sufficient period of time and an algorithm that can process such a "personal big data (PBD)". In the PBD, genomic data do not change over time, while microbiome and environmental factors could change slowly. Only the data in the EHRs and the vital signs/activity data from the wearable devices would change dynamically across the whole lifetime with a granularity from seconds to years.

As we know, the current healthcare system spends a great deal of resources on the later stage of diseases. Up to 10 percent of the healthcare spending went to the end-of-life care [3]. This trend is increasing the rocketing healthcare costs even more and contributes little to the quality of life. That is the reason why a promising branch of the "precision medicine" called "precision health" would, instead, focus on preventing the diseases from happening or worsening by properly managing each individual's health risk factors before the health- threatening events. The traditional approach to identifying risk factors typically involves an experts' nomination of candidate variables (factors) followed by a statistical selection and often dichotomization of these final variables to enter a fitting model such as logistic regression. Two major problems plaguing this approach are:

(1) we often have to eliminate a lot of variables so that the final fitted model can be explainable with a "manageable" number of variables (usually less than twenty), and
(2) we often have to ignore the ups and downs of disease progression and dichotomize a patient as just having a disease or not.

For example, a patient who had been through a difficult-to-treat twenty-year history of diabetes would be labelled as "positive" for diabetes, exactly like another patient who was recently diagnosed as diabetic but can be well controlled by merely a low-dose metformin. It is not hard to imagine how much of the "individualities" over the time dimension was actually thrown away by this process and how this overly simplified model can have inaccurate and even misleading conclusions.

2. The temporal phenomic map model

To amend these problems, we proposed a "Temporal Phenomic Map (TPM)" model that will inclusively take into consideration all the possible variables and their change over time [4,5,6]. Using data elements from the EHRs as examples, we can model all the diagnoses and medications that happened in a specific period of time before the initial diagnostic date of the target health threat into a two-dimensional map for each patient. For this map, all the disease and medication variables will be listed on the Y-axis and the time as the X-axis with months or weeks as the basic unit. Each patient will then be represented as a 2-D map with either 1 or 0 in each cell, where 1 can be disease or medication present on that specific time unit and 0 as the absence. Now we have numerous TPMs representing our sample patients and then we label each map as having the target health threat (positive) or not (negative) according to the EHRs. Using a widely-deployed ML algorithm specific for recognizing images called CNN (Convolutional Neural Networks), we can easily train a ML model to recognize a positive

TPM from the sample. Now the trained CNN can be readily used to spot the target health threat out of any patient. We have used the TPM to accurately predict the risk of several major cancers including non-melanoma skin cancer, lung cancer and liver cancer using 36 months of EHRs data [4,5,6].

3. Discussion and conclusions

As described above, current healthcare systems invest heavily on the later stage of care and tend to under-invest the early stage of diseases. Driven by the ever-evolving AI technology, further study into Temporal Phenomics may lead to a future where we can always spot a major health threat for each individual earlier than we possibly can now. Through achieving this “Earlier Medicine” vision, we may one day catch most of the cancer patients at stage zero and cure them, spot a heart failure at stage A or predict a stroke and avoid long-term disabilities, etc. [7]. The possibilities are endless. Is this not a future worth investing into? Is this not a future we all want to live in? I believe that the answer is crystal clear.

References

[1] Moore GE. Cramming More Components onto Integrated Circuits. Electronics, 1965; 38(8).

[2] Collins FS, Varmus H. A new initiative on precision medicine. N Engl J Med. 2015 Feb 26;372(9): 793-5.

[3] Aldridge MD, Kelley AS. The myth regarding the high cost of end-of-life care. Am J Public Health. 2015 Dec;105(12):2411-5.

[4] Wang H, Wang Y, Liang C, Li Y. Assessment of Deep Learning Using Nonimaging Information and Sequential Medical Records to Develop a Prediction Model for Nonmelanoma Skin Cancer. JAMA Dermatol. 2019;155(11):1277–1283.

[5] Yeh M, Wang Y, Yang H, Bai K, Wang H, Li Y. Artificial Intelligence–Based Prediction of Lung Cancer Risk Using Nonimaging Electronic Medical Records: Deep Learning Approach. J Med Internet Res. 2021;23(8):e26256.

[6] Liang C, Yang H, Islam M, Nguyen P, Feng Y, Hou Z, Huang C, Poly T, Li Y. Predicting Hepatocellular Carcinoma With Minimal Features From Electronic Health Records: Development of a Deep Learning Model. JMIR Cancer. 2021;7(4):e19812. URL: https://cancer.jmir.org/2021/4/e19812.

[7] Iqbal U, Celi LA, Li YJ. How Can Artificial Intelligence Make Medicine More Preemptive? J Med Internet Res. 2020 Aug 11;22(8):e17211.