Deriving A Novel Health Index Using A Large-Scale Population Based Electronic Health Record With Deep Networks

Chen-Ying Hung, Huan-Yu Chen, Lawrence JK Wee, Ching-Heng Lin, and Chi-Chun Lee

Abstract—**Health indexes are useful tools for monitoring the health condition of a population and can be used to guide healthcare policy of governments. However, most health indexes are constructed by using statistical methods to summarize recent adverse events (e.g., mortality). Information from these tools may reflect merely the impact of prior health policy holistically and can hardly indicate the most recent dynamics and its impact on future health conditions. As the advancements in medications and medical techniques rapidly evolve, there is a need of new health indexes that can reflect the most recent predictive health condition of a population and can easily be summarized with respect of any sub-population of interest. In this work, we develop a novel health index by using deep learning technique on a large-scale and longitudinal population based electronic health record (EHR). Three deep neural network (DNN) models were trained to predict 4-year event rates of mortality, hospitalization and cancer occurrence at an individual-level. Platt calibration approach was used to transform DNN output scores into estimated event risks. A novel health index is then constructed by weighted scoring these calibrated event risks. This individual-level health index not only provide a better predictive power but can also be flexibly summarized for different regions or sub-populations of interest - hence providing objective insights to develop precise personal or national policy beyond conventional health index.**

I. Introduction

Severe health-related events (such as mortality, hospitalization or cancer occurrence) significantly impact healthcare costs and lives of patients. A key function of governments is to develop informed healthcare policy to prevent these events and improve their citizen's health. Population-based health indexes are useful tools for monitoring the health of a society and are used to assist healthcare policy makers to better understand the current health conditions [1, 2]. Health indexes are quantifiable evidence to objectively describe the health conditions of a population. In the past, researchers usually use a survey-based methodology to collect event statistics that generalize to the target population. There are several health indices that have been developed for these purposes, i.e. the Health Status Index [3] and the Healthcare Quality and Access Index [1]. Each of these health indicators have already been used for evaluation, comparison, resource allocation and decision making [4].

Although health indexes have been extensively used by the governments to make health related policy, traditional health index metric, i.e. annual mortality data [5], may only reflect the summary impact of the past health policies which have been in place for several years. However, the rapid progress of medicines and medical technologies in this era has affected healthcare costs and public health dynamically over time. Information from these tools is therefore no longer a sensitive measurement when being used to help adapt the current health policy for the future [6]. It is reasonable to assume that these global health indicators suffer from major limitations when using it in the current complexly evolved society. A health index consists of future health events prediction may better reflect the current health status of a population and could offer better predictive decision support for policy makers. However, the challenge in reliably predicting health events has hindered the development of such a prediction-based health index.

Artificial intelligence (AI) techniques, e.g., those based on deep learning algorithms, have already demonstrated impressive power for predicting clinical events in several works [7]. Deep learning can model complex non-linear relationships between predictive variables without prior statistical assumptions [8]. Our recent works have shown that by using deep networks on large-scale EHRs, it can achieve a higher disease predictive accuracy than other machine learning methods [9, 10]. Deep learning has also been successfully applied in disease identification and outcome prediction in conventionally challenging clinical prediction tasks, such as young stroke prediction [11]. Furthermore, population based EHRs are collected non-obtrusively in a large-scale long-term follow-up manner that include several important yet diverse aspects of health-related information at an individual-level. These characteristics make population based EHR an especially valuable data source for constructing a prediction health index with deep learning techniques.

In this work, we propose a novel health index developed by using deep learning technique with a large-scale population based EHR. The health index incorporates 3 important health predictive indicators (mortality, hospitalization and cancer occurrence). There are 4 steps to develop such a health index: (1) training 3 DNN models to predict the 4-year event rate of mortality, hospitalization and cancer occurrence for each individual, (2) using Platt calibration approach to transform the DNN outputs into estimated 4-year event risks, (3) calculating the individual health index value by scoring and weighting the impact of each indicator, (4) summarizing health index for selected population (e.g., people lives in different regions) to provide intuitive insights (e.g., using map data visualization) for government to develop health

CCL is the corresponding author for this work. He is with the Department of Electrical Engineering, National Tsing Hua University, Hsinchu, Taiwan. (phone: +88635162439. e-mail: cclee@ee.nthu.edu.tw)

CCL, CYH, HYC are with the MOST Joint Research Center for AI Technology and All Vista Healthcare, Taiwan.

CYH, HYC are with the Department of Electrical Engineering, National Tsing Hua University, Hsinchu, Taiwan.

CYH is with the Department of Internal Medicine, Taipei Veterans General Hospital, Hsinchu Branch, Hsinchu, Taiwan.

LJW is with Allianz SE Asia-Pacific, Singapore.

CHL is with the Department of Medical Research, Taichung Veterans General Hospital, Taichung, Taiwan.

978-1-7281-1990-8/20/$31.00 ©2020 IEEE

TABLE 1. INFORMATION OF THE STUDY DATASETS

	Training dataset (TR-2003)	Testing dataset 1 (TE-2003)	Testing dataset 2 (TE-2007)
No of records	4,638,196	1,168,102	1,303,337
No of people	383,322	95,746	102,625
Age (year)	43.3	43.3	44.6
Male (N, %)	183,412 (47.8%)	45,914 (47.9%)	49,517 (48.2%)
4-year events			
Mortality	14,592 (3.8%)	3,680 (3.8%)	4,041 (3.9%)
Hospitalization	38,274 (10.0%)	9,551 (10.0%)	10,387 (10.1%)
Cancer	11,639 (3.0%)	2,855 (3.0%)	3,440 (3.3%)

improvement policy. Importantly, since the method is based on individual-level prediction, one can flexibly derive relevant index for different sub-population of interest for the policy maker to assess the societal health condition with variable granular precision. In this study, we detail our approach and show that the method indeed better predicts the risk of events than traditional health indexes.

II. METHODS

A. Database and study population

The National Health Insurance program in Taiwan has been operated for more than 20 years and covers about 99% of entire population. The routinely collected data from the insurance program can ideally be used for formulation of an overall population-based health index that measures the health status of Taiwan. The National Health Research Institute (NHRI) established the National Health Insurance Research Database (NHIRD) and provided the data to researchers. The database was confirmed by the NHRI to be representative of the general population of Taiwan [12]. It contains de-identified health-care information of over 900,000 patients from 2000 to 2011. Ethical approval for this study was granted by the Institutional Review Board of National Tsing Hua University.

In this study, we developed 3 predictive models to estimate 4-year risk of 3 important health-related events (mortality, hospitalization and cancer occurrence) and used these models to construct a novel health index. In order to develop and evaluate the DNN-based algorithms, subjects in the data were assigned into two groups, training group (~80% of total patients) and testing group (~20% of total patients). Patients aged 18 to 90 years in 2003 were identified from these 2 groups and formed the training dataset (TR-2003) and testing dataset 1 (TE-2003). Patients aged 18 to 90 years in 2007 were identified from the testing group and formed the testing dataset 2 (TE-2007). Following this inclusion criteria, our final dataset includes a total number of 383,322 patients in TR-2003, 95,746 patients in TE-2003, and 102,625 patients in TE-2007. (see Table 1)

B. Feature engineering and events definition

In our previous work, we have developed a feature engineering method for the NHIRD database [9, 10, 11]. We utilized data from outpatient departments (within 3 years prior to the enrollment) to generate features and data from inpatient departments (within 4 years after the enrollment) to retrieve target event labels. We gathered several measurements (demographic information, medication use, and disease diagnosis) from records of an individual subject. The information of medication use was converted to Anatomical Therapeutic Chemical (ATC) code (a total of 582 drug class) and the disease diagnosis was converted to ICD-10-CM code (a total of 1,676 disease class) by using the code-converting sheet provided by the National Health Insurance Bureau of Taiwan. We utilized time stamp (a total of 5 time period: 0.25, 0.5, 1, 2, 3 year) for these measurements to generate the final feature vector that captures these clinical variables temporal information. Finally, we extracted a total of 11,292 features (age, gender, 2,910 drug and 8,380 disease information) from the dataset. We additionally used simple Pearson correlation method to perform feature selection to identify the most discriminative 256 features before training DNN algorithms. The outcome event was defined as any mortality, hospitalization, and cancer occurrence recorded in the hospital discharge diagnoses (International Classification of Diseases, Tenth Revision, Clinical Modification, [ICD-10-CM] code: C00~D49) in the inpatient database within 4 years after patients being enrolled.

TABLE 2. SCORING FOR EACH HEALTH INDEX (HI) AND ITS CORRESPONDING 4-YEAR EVENT RISK

Scores	100	80	60	40	20
Mortality HI	<0.4%	0.4-0.8%	0.8-4%	4-8%	>8%
Hospitalization HI	<2.4%	2.4-4%	4-8%	8-24%	>24%
Cancer HI	<0.8%	0.8-2.4%	2.4-4%	4-8%	>8%

TOTAL HEALTH INDEX = 50% MORTALITY HI + 25% HOSPITALIZATION HI + 25% CANCER HI

C. Development of the health index

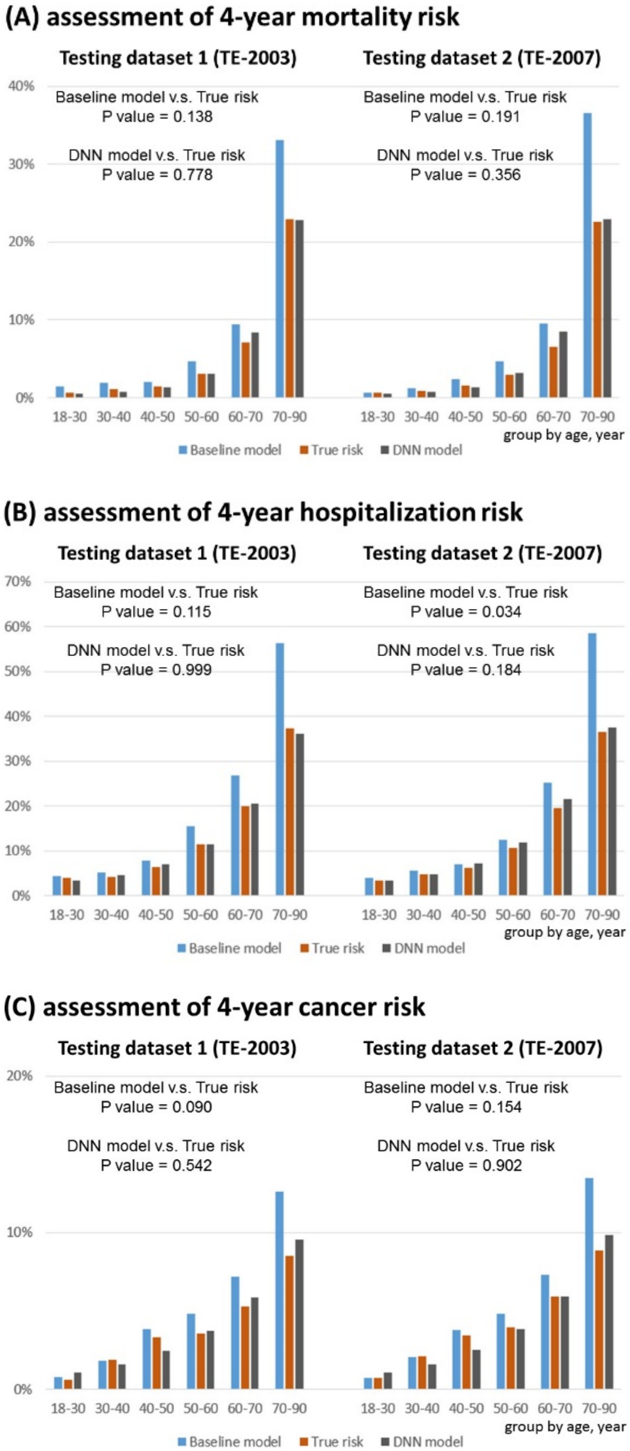
Step 1: Training three DNN models

In this step, we trained 3 DNN models to predict the 4-year event rate of mortality, hospitalization and cancer occurrence. We used a multilayered feed-forward neural network as our main architecture. The structure of our DNN model was composed of 5 fully connected layers, including an input layer, 3 hidden layers (each layer had 256 neurons), and an output layer with sigmoid function. The activation function was hyperbolic tangent and the optimization algorithm was stochastic gradient descent. We also applied a simple normalization approach by scaling the feature values to a range between 0 and 1 to speed up the training process. The algorithm was implemented using the Keras (2015, GitHub) toolbox. Performance of the DNN prediction models was examined on the testing datasets (TE-2003 and TE-2007) using area under the receiver operative cure (AUROC) values.

Step 2: Estimating 4-year event risk using Platt calibration

We employed the use of Platt calibration [13] as the method for calibrating the scores of machine learning models by fitting a logistic transformation to the model's outputs in order to estimate event risk accurately [10, 14]. With this risk calibration, we then performed a comparison of the risk assessment ability between traditional health index and our

Figure 1. Comparing true 4-year (A) mortality risk, (B) hospitalization risk, and (C) cancer risks with risks estimated by baseline models and DNN models.

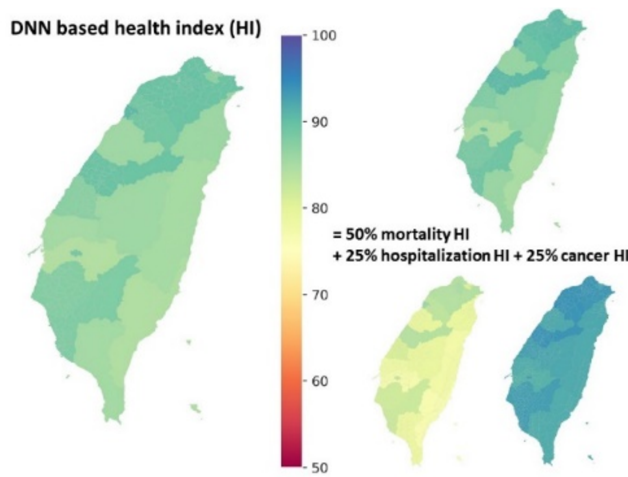


estimated risk. We used paired samples t-test to exam the difference between risk calculated by traditional model and true risk, as well as the difference between our DNN-based estimated risk and true risk. (see Figure 1)

Step 3: Calculating health index of each individual

We designed a converting sheet (Table 2) according to the distribution of events in the database for calculating the final

Figure 2. Map data visualization for the novel health index



health index score. Three different scores were calculated separately for the 3 indicators to come up with the final health index value. The total health index was calculated as a summation of 50% mortality health score, 25% hospitalization health score, and 25% cancer health score.

Step 4: Summarizing health index of a selected population

After calculating health index for each person, we summarized the average values for total population or each sub-population of interest (e.g., those lives in different counties of Taiwan) to provide a general view of the different health condition enabling comparison between different groups. This methodology could be integrated further with map-based approach to visualize the region-specific health condition information.

III. RESULTS

A total of 383,322 patients aged 18 to 90 years were included in the training dataset (TR-2003). The mean age of the development dataset population was 43.3 years, with 47.8% men. The 4-year event rate is 3.8% for mortality, 10% for hospitalization, 3% for cancer occurrence. Detailed demographics and characteristics of the training and testing datasets are summarized in Table 1. As mentioned above, a total of 256 features were selected out of 11,292 generated features for developing DNN model. Three DNN algorithms were constructed for predicting the 3 health related events: mortality, hospitalization, and cancer occurrence. The trained DNN model for mortality prediction achieved AUROC values of 0.872 and 0.884 in TE-2003 and TE-2007. The trained DNN model for hospitalization prediction achieved AUROC values of 0.779 and 0.785 in TE-2003 and TE-2007. The trained DNN model for cancer occurrence prediction achieved AUROC values of 0.737 and 0.731 in TE-2003 and TE-2007. These models while achieving high discriminatory power (high accuracy and AUROC) for event prediction but with poor calibrating power (poorly related to the true event risk if no further calibration process was performed [10, 14]). We therefore applied Platt calibration [13] to the DNN model outputs for estimating the true event risks.

Figure 1 compares true 4-year risks with risks estimated by baseline methods and our DNN models in TE-2003 and

TE-2007. The baseline methods use the annual event rate at the time of database being constructed (2003 for TE-2003, and 2007 for TE-2007) as a direct estimate to the 4-year future event rate. This method represents most of the traditional surveillance tool by which many current health indexes are constructed based on. True event risks are the actual 4-year event rate for each event. The risks estimated by our DNN models (blue bar) showed a closer match to the true 4-year risks (orange) than baseline methods (gray bar). While paired samples t-test showed no significant differences between baseline/DNN models and true risk (except for hospitalization risk estimation in TE-2007, in which risk estimated by baseline model was borderline different from true risk with a p value of 0.034), the difference between true risk and DNN models showed a larger p value than the difference between true risk and baseline models.

We calculated the health index for each person by using the score conversion sheet showed in Table 2. The total health index is composed by 50% mortality health score, 25% hospitalization health score, and 25% cancer health score. For a clear view of regional health condition, we calculated the average value for population in each counties of Taiwan. Figure 2 showed the map visualization results and indicated an intuitive trend that urban area had an overall higher health index score than rural area. Because this novel health index is derived for each person, health condition in different subgroup can be easily accessed for policy maker; every person in Taiwan also would be able to know his/her own health index if enrolled in the National Health Insurance program, which could further be used to assist self-health management.

IV. DISCUSSION

Deep learning method has repeatedly achieved impressive predictive and recognition results across a variety of AI tasks in recent years [7, 8]. In this work, we used DNN model to construct a novel health index that can perform better risk estimation than traditional surveillance model in constructing health index. This approach makes the estimation of real time health condition of population more accurately and arguably provide better-informed analytics for policy maker with its future predictability. The similar validation results on TE-2003 and TE-2007 demonstrated that the superior predictive ability of DNN model change little over time. To the best of our knowledge, this population-based health index is one of the first studies for applying deep learning technique for developing health index based on predictive risk assessment. Our results show that DNN algorithms can reliably estimate health event risk in different age range by using information from the EHR source.

As the use of novel clinical intervention strategies and medication changes over time, the health condition of the population also changed rapidly, especially in the modern era [6]. While traditional surveillance model may reflect the holistic results of medical care system several years ago, our novel DNN based health index estimated the health condition by predicting into the future and therefore reflect the impact of current medical care system. Health policy makers may find this innovative index potentially very useful as it incorporates several health dimensions and reflect the result of current medical care system with its predictability into the future. Moreover, these health indicators are constructed from routinely collected data of Taiwan health care system. Ideally, this approach can even provide a real-time individual level health index for any specific person (or any sub-populations/region-specific cohort of interest) to evaluate and manage their health condition.

V. CONCLUSIONS

In this evaluation of applying DNN strategy in using EHRs for constructing a novel health index, we demonstrate that this method can achieve better performance for further risk estimation than traditional approaches. This work presents one of the first methods in applying DNN to achieve health condition recognition for a specific population. Further prospective research is necessary to determine the feasibility of applying this novel health index in real world practice and to see whether such a DNN based health index could improve health care system and assist health care policy making for the general population.

REFERENCES

[1] GBD 2017 Disease and Injury Incidence and Prevalence Collaborators, “Global, regional, and national incidence, prevalence, and years lived with disability for 354 diseases and injuries for 195 countries and territories, 1990-2017: a systematic analysis for the Global Burden of Disease Study 2017,” Lancet, vol. 392, pp. 1789-858, Nov 2018.

[2] E. Kaltenthaler, R. Maheswaran, and C. Beverley, “Population-based health indexes: a systematic review,” Health Policy, vol. 68, pp. 245-55, May 2004.

[3] N. Frohlich, and C. Mustard, “A regional comparison of socioeconomic and health indices in a Canadian province,” Soc Sci Med, vol. 42, pp. 1273-81, May 1996.

[4] C. Larson, and A. Mercer, “Global health indicators: an overview,” CMAJ, vol. 171, pp. 1199-200, Nov 2004.

[5] GBD 2015 Healthcare Access and Quality Collaborators, “Healthcare Access and Quality Index based on mortality from causes amenable to personal health care in 195 countries and territories, 1990-2015: a novel analysis from the Global Burden of Disease Study 2015,” Lancet, vol. 390, pp. 231-66, Jul 2017.

[6] H.P. Water, R.J. Perenboom, and H.C. Boshuizen, “Policy relevance of the health expectancy indicator; an inventory in European Union countries,” Health Policy, vol. 36, pp. 117-29, May 1996.

[7] R. Alvin, O. Eyal, C. Kai, M.D. Andrew, H. Nissan, et al., "Scalable and accurate deep learning with electronic health records," npj Digital Medicine, vol. 1, pp. 18, 2018.

[8] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," Nature, vol. 521, pp. 436-44, May 2015.

[9] C.Y. Hung, W.C. Chen, P.T. Lai, C.H. Lin, and C.C. Lee, "Comparing deep neural network and other machine learning algorithms for stroke prediction in a large-scale population-based electronic medical claims database," Conf Proc IEEE Eng Med Biol Soc, pp. 3110-3, July 2017.

[10] C.Y. Hung, C.H. Lin, T.H. Lan, G.S. Peng, and C.C. Lee, "Development of an intelligent decision support system for ischemic stroke risk assessment in a population-based electronic health record database," PLoS One, vol. 14, pp. e0213007, Mar 2019.

[11] C.Y. Hung, C.H. Lin, and C.C. Lee, "Improving young stroke prediction by learning with active data augmenter in a large-scale electronic medical claims database," Conf Proc IEEE Eng Med Biol Soc, pp. 5362-5, July 2018.

[12] CY Hsieh, CC Su, SC Shao, SF Sung, SJ Lin, et al, “Taiwan's National Health Insurance Research Database: past and future,” Clin Epidemiol, vol. 11, pp. 349-58, May 2019.

[13] JC Platt, “Probabilistic outputs for support vector machines and comparisons to regularized likelihood methods,” Advances in Large Margin Classifiers, vol. 10, pp. 61–74, 1999.

[14] M.J. Pencina, R.B. D'Agostino, R.B. D'Agostino, R.S. Vasan, “Evaluating the added predictive ability of a new marker: from area under the ROC curve to reclassification and beyond,” Stat Med, vol. 27, pp. 157-72, Jan 2008.