Using signal processing techniques to predict PPG for T2D

Gerald C Hsu, Email: g.hsu@eclaireMD.com
eclaireMD Foundation, USA

Abstract

The author has collected an entire set of PPG and lifestyle data for a period of 994 days with 2,982 meals (6/11/2015-3/1/2018). This paper discusses the methodology and accuracy of his developed PPG prediction model using signal processing techniques for type 2 diabetes. Materials & Methods: thanks to his academic background in mathematics, physics, and engineering, he views these biomedical and lifestyle data as a set of nonlinear signal waves. He applied signal processing to decompose this time-series measured PPG signal into multiple (>10 lifestyle factors) single-sourced composite waveforms, examined each composite signal, then recombined them into a predicted PPG curve. Finally, he compared this predicted signal against the measured signal to calculate its accuracy and correlation. He further improved his model via a trial-and-error "curve-fitting" method. Results: The PPG's major creation source, corresponding glucose, and contribution level are as follows: Carbs/Sugar: 14.5 mg/dL, 37%; Post-meal Exercise-15.7 mg/dL, 41%; Weather 3.8 mg/dL, 10%; Measurement delay -2.4 mg/dL, 7% et al. -1.9 mg/dL, 5%. During this era , his average PPG values are: Predicted 119.16 mg/dL Measured 119.88 mg/dL with 99.4% linear accuracy and a high correlation of 70%. Conclusion: The quantitative results from the developed PPG prediction model reflect the accuracy and applicability for type 2 diabetes control via a guided lifestyle management. The use of signal processing from electronics engineering and computing is additionally proven quite effective for this investigation

Regardless the argument on glucose testing method's accuracy via either lab-tested A1C or finger piercing and testing strips, the author has collected a complete set of PPG data using lab-tested A1C and finger prick testing strips plus his created lifestyle data during a period of 1,075 days with 3,225 meals (6/1/2015 - 5/11/2018). This PPG-related data set, size of ~400,000 data, is only a small portion of his entire ~1.5 million data. Due to his mathematics and engineering background, he views these data curves related to biomedical conditions and lifestyle management as a collection of various nonlinear input and output signal waves of the human body. At first, he applied "Finite Element Method" of engineering modeling to convert this "analog" human system into a "digitized" mathematical system in order to get an approximate solution of the real human system. He sees each digitized sub-wave as representing a single-source created contribution element of the PPG wave. Therefore, he applied signal processing techniques to decompose this measured PPG signal into more than 10 single-sourced sub-waves. He carefully checked each sub-signal waveform for its completeness, accuracy, and correlation with other curves, using time-series, spatial, and frequency domain analyses, etc. Over the past three years, he continuously explored and added some missing influential factors into the formation of the PPG signal. His purpose was trying to improve the predicted PPG waveform's contents and accuracy while maintaining high correlation with the measured PPG waveform. For example, by the fall of 2016, the accuracy of his predicted PPG reached ~95%. In September of 2017, he identified that weather temperature also had an impact on glucose value. Therefore, he selected a 2-year period (6/2015 - 7/2017) to examine his travel schedule in detail and also entered each day's local ambient temperature of the city where he stayed. In this way, he was able to generate a new temperature sub-wave which brought the accuracy of the predicted PPG from ~95% to ~98%.

Another factor was that his glucose was quite high when he was sick with flu for a month at the end of 2017. After that experience, he further enhanced his prediction model with the inclusion of "physical sickness or wellbeing" which finally brought the prediction accuracy to 99.8%. The author used his measured data as the base for data comparison. He has safeguarded the integrity of his data and has never altered its original content or influenced its integrity. All data was collected in its entirety from one patient only, via customized software, over an extended period of time. Therefore, the author needed very little "data cleaning" before starting his research work, which included data analysis and data interpretation. This project does not have to be concerned with problems such as data interference and data contamination due to different genetic conditions, various lifestyles, and contradicting interpretations. These data come from a consistent sample source, making it much easier for the author to dive into one variable and extract the buried information. After analyzing each sub-wave in detail, he was ready to reintegrate these sub-waveforms into another nonlinear predicted PPG waveform.

He further improved his model via a “curve-fitting” trial-and-error engineering method. He has continuously compared these two sets of data and improved the accuracy until it reached a very high linear accuracy while still maintaining high correlation. High correlation means the trend of predicted curve moves along with the measured curve like its “twin”. For A1C estimation, he specifically added in a “safety margin”, e.g. +15%, on top of his originally predicted A1C value for the purpose of providing a numerical buffer which can serve as an “early warning” to T2D patients. Both the Adjusted Glucose and Estimated A1C models also utilized “self-adjusting” machine-learning algorithms in order to correct or compensate for the built-in “error” from chemical process of various lab tests and glucometers. The quantitative results from the developed PPG prediction model reflect the accuracy and applicability for type 2 diabetes control via a guided lifestyle management. The utilization of signal processing from electronics engineering is also proven quite effective for this investigation. As shown in (Figure 5), Health Data Comparison Between 2010 and 2017, the author’s health condition has been improved significantly due to the control of his glucose, especially PPG.

Biography

Gerald C Hsu has received his PhD in Mathematics and majored in Engineering at MIT. He has attended different universities over 17 years and studied seven academic disciplines. He has spent 20,000 hours in T2D research. First, he studied six metabolic diseases and food nutrition during 2010-2013, then conducted research during 2014-2018. His approach is “Math-Physics and Quantitative Medicine” supported mathematics, physics, engineering modeling, signal processing, computing , big data analytics, statistics, machine learning, and AI. His main focus is on preventive medicine using prediction tools. He believes that the higher the prediction, the more control you've got.

This work is partly presented at International Conference on Metabolomics and Diabetology

May 23-24, 2018 | New York, USA