**Context and motivation**

Anemia is a serious global public health problem that particularly affects young children and pregnant women. WHO estimates that 42% of children less than 5 years of age and 40% of pregnant women and 20% of the older population worldwide are anemic. The diagnosis of anemia and its causes is important for improving human well-being.

The traditional diagnosis of anemia requires laboratory-based measurements of a venous blood sample, which could bring trauma and pain to patients, even wound infection. The diagnosis process involves complex equipment that requires professional operators and a fixed test site. To address these issues, researchers keep looking for novel anemia detection methods.

Novel anemia detection methods based on AI data analysis and advanced sensing technology. For the first part, the ML/DL models can be built for classifying patients' fingernails, fundu, conjunctival, or other bio-images into anemia and normal condition. However, these models are not able to give the cause of the anemia or predict the anemia in an early stage. For the second part, researchers have developed sensors based on the principle of multispectral analysis, transmission spectrum analysis, etc. Similarly, these methods can not explain the reason for the anemia and it is unable to iterate and update the algorithm with new biomarkers.

To address the issues in novel methods, in other words, to diagnose the anemia with higher accuracy and provide the cause, we would like to combine the EHR data with biomarker image/ sensor signal and process these data by machine learning model for anemia diagnosis.

The collection of data sets is the basis of all programs. The data collection methods, equipment, object, timelines, and evaluation metrics of the collected datasets will be discussed in this proposal.

**Target population and implementer**

Here we divide the collection method into two kinds according to the target population and implementer.

Method A: target population: people around us (students/colleagues)

Implementer: Zhuo ZHI or some MSc students

Method B: target population: patients in NHS

Implementer: medical staff in NHS

**Collection method**

we would like to collect the test taker’s fingertipPPG signal, conjunctival image, EHR information and the true value of the hemoglobin. Each of these data will be introduced next.

1. **Fingertip PPG signal**

**1.1 The principle of the PPG signal**

First of all, we would like to introduce the principle of the PPG signal.The sequence of events that happen between the beginning and the end of a heartbeat is called the cardiac cycle. The cardiac cycle is composed of two main phases: the ventricular diastole and the ventricular systole. In the diastole or relaxation phase, the blood flows to the auricles, causing pressure to decrease in the blood vessels. In the systole, or contraction phase, the blood is pumped out of the ventricles and distributed throughout the body, causing pressure to increase in the blood vessels. PPG sensors measure the amount of infrared light absorbed or reflected by blood. Volume changes are caused by pressure changes in blood vessels, which occur throughout the cardiac cycle [1]. PPG sensors can be divided into two types according to the way of light transmission, as shown in Fig. 1.

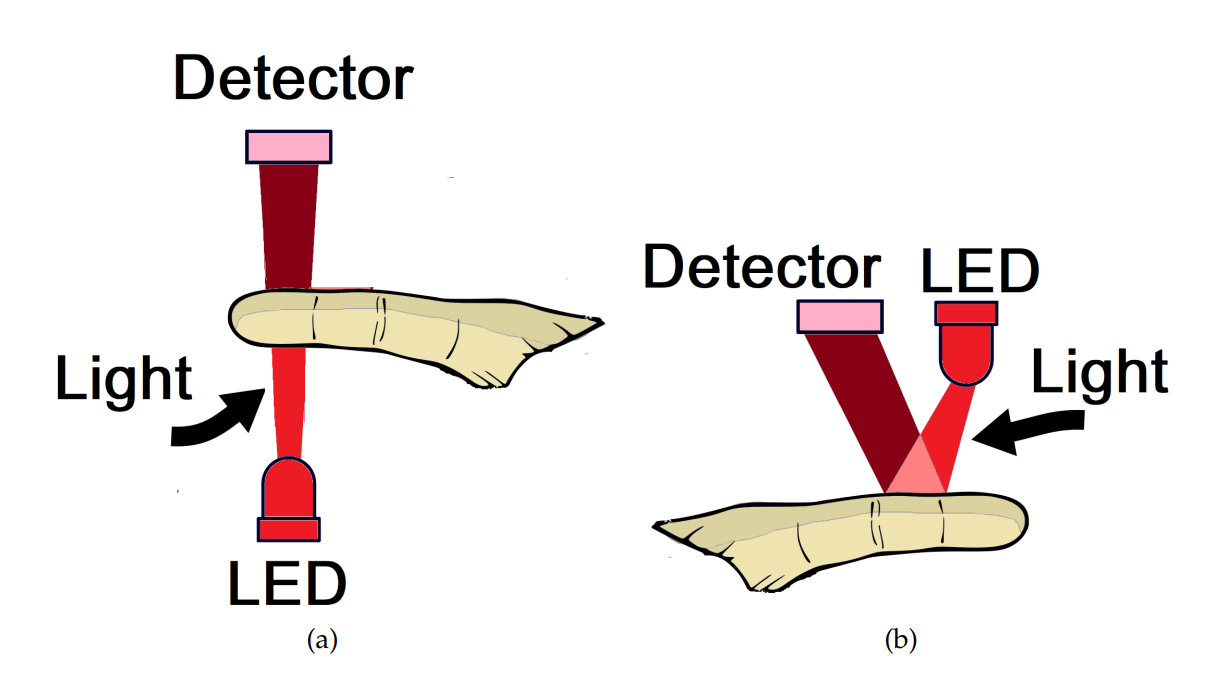


Fig. 1. The operation of PPG sensors for finger application (a) transmission mode (b) reflection mode

As we can find in Fig. 1, With a PPG sensor in transmission mode, the LED light passes through absorbent substances, such as the skin pigmentation, bone and arterial and venous blood, and is then received by the detector and quantified by filters and converters. In contrast, a PPG sensor in reflection mode reflects the LED light on the skin, which is received by the detector, and quantified in a similar fashion through the use of filters and converters. Nonetheless, this mode is applied mainly in the body parts too thick to allow the transmission of light (for example, the wrist and forehead). Therefore, the PPG sensors can assume varied shapes, for example, a band, a wristwatch, or a patch.

PPG makes use of low-intensity infrared (IR) light. When light travels through biological tissues it is absorbed by bones, skin pigments and both venous and arterial blood. Since light is more strongly absorbed by blood than the surrounding tissues, the changes in blood flow can be detected by PPG sensors as changes in the intensity of light. The voltage signal from PPG is proportional to the quantity of blood flowing through the blood vessels. Even small changes in blood volume can be detected using this method, though it cannot be used to quantify the amount of blood.

The photoplethysmographic wave describes changes in the attenuation of light energy in its pathway when transmitted or reflected in tissues and the bloodstream. This waveform is totally related to the systole and the diastole of the cardiac cycle, as can be observed in Fig. 2. The PPG signal series represented in Fig. 2 was extracted from a healthy person. One cycle of the waveform is taken out and shown in Fig. 3.

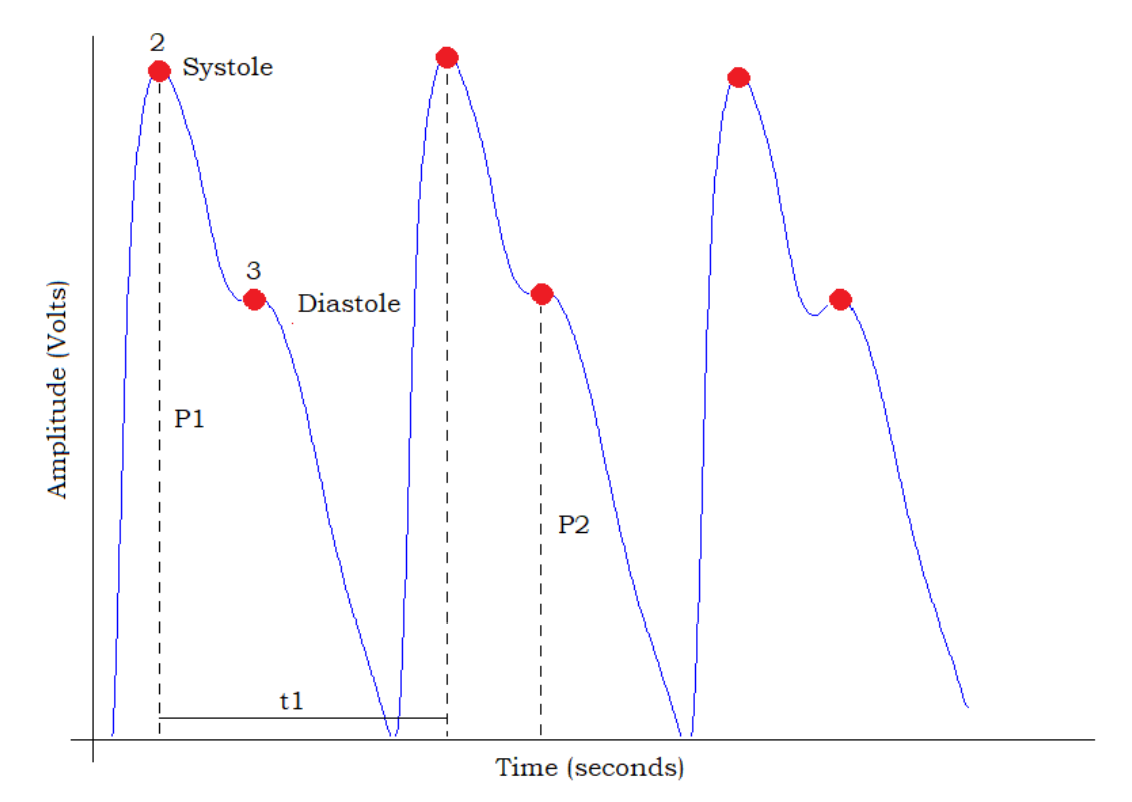


Fig. 2. Example of the PPG signal from a healthy person



Fig. 3. One cycle of the PPG signal

Based on the analysis in Fig. 3, it is possible to estimate some parameters, such as the systolic peak, an indicator of the pulsatile changes in blood volume caused by arterial blood flow, is generated by the direct pressure wave coming from the left ventricle to the periphery of the body. The systolic amplitude, representing pulsatile changes in blood volume, can lead to a machine-learning algorithm to correlate the pulsatile changes with blood constituent levels. The dicrotic notch is a small downward deflection between the systolic and diastolic points of a PPG cycle. The pulse interval represents the relationship between the contribution that the wave reflection makes to the systolic arterial pressure and the reflected wave coming from the center.

**1.2 Popular PPG collector used for estimating hemoglobin level**

As introduced before, the LED light and the detector are the most important part of the PPG collector. It is worth noting that the single-wavelength LED can not meet the requirement for hemoglobin prediction, which is explained in [1][2][3]. The reason can be summarized as:

*In living tissues, hemoglobin and melanin absorb in the 400nm-650nm spectrum, and other proteins absorb light in the violet and ultraviolet regions. A dominant tissue chromophore is water, which absorbs photons strongly around the 1000nm wavelength of light. The spectrum range from 650nm to 1100nm is known as the tissue optical window or NIR region. In the NIR region, light absorption is lower than the other spectral areas for oxy-hemoglobin and deoxy-hemoglobin. That is why the light sources including 850nm, 940nm, and camera light are considered as hemoglobin responsive. On the other hand, 1070nm wavelength NIR LED light is selected as plasma sensitive.*

After summarizing the existing research, we find that two kinds of PPG collectors are commonly used for hemoglobin prediction.

1. Record video with the phone camera and attach various LEDs

PPG signal can be extracted from the video, the principle is that when the finger is close to the surface of the camera, the camera can record the color change of the LED light passing through the finger during the heartbeat cycle, as shown in Fig. 4. The advantage of this method is that the phone can be fully utilized and only a few accessories need to be added.

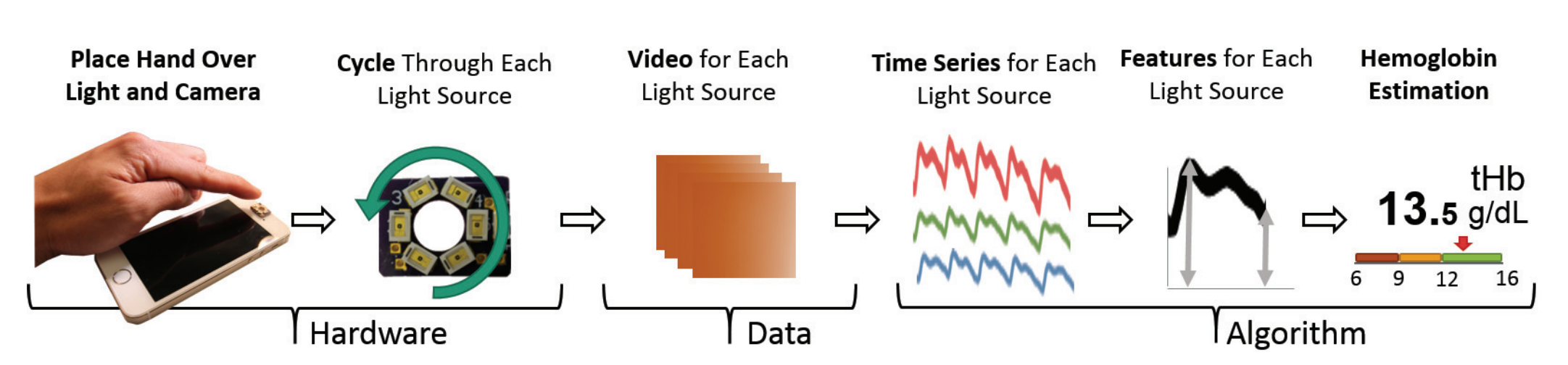


Fig. 4. Extract the PPG signal from the fingertip video

Take the method from [2] as an example, the system is shown in Fig. 5.



Fig. 5. The system design of the PPG collector in [2]. (LED1&LED2: 880nm IRLED, LED3&LED6: white LED, LED4&LED5: 970nm IRLED)

As shown in Fig. 5, an O-shaped LED board is mounted on the phone camera. Six LEDs are put on the board: two white LEDs, two IRLED with 880nm wavelength, and two IRLED with 97nm wavelength. During the collection process for each patient, LED6 and LED3 are always bright (provide the camera with enough light). LED1, LED2 and LED5, LED4 light up for 15 seconds in sequence. Finally, two videoes are collected for each patient and used for analysis.

1. Record PPG signal with PPG sensor

There are certain drawbacks in the method ‘recording video by the phone camera with attached various LEDs’. For example, the design of LED board has strict requirements for mobile phone models and does not have good generalization. In addition, it is difficult to control the change of environment light, which may bring noise and offset to the collected data. To address these issues, many researchers design the PPG collector with Dedicated components and molds. Take the method in [3] as an example, the block diagram of the system is shown in Fig. 6.

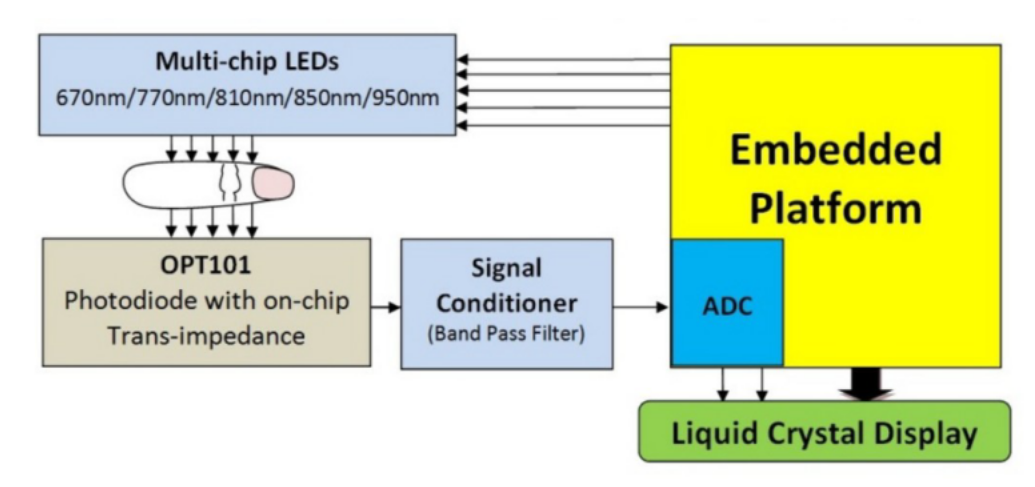


Fig. 6. The Block diagram of the system in [3]

The multi-chip LEDs and OPT101 are the LED model and detector model in the PPG collector. It is worth noting that both modules are available off the shelf, which is shown in Fig. 7.

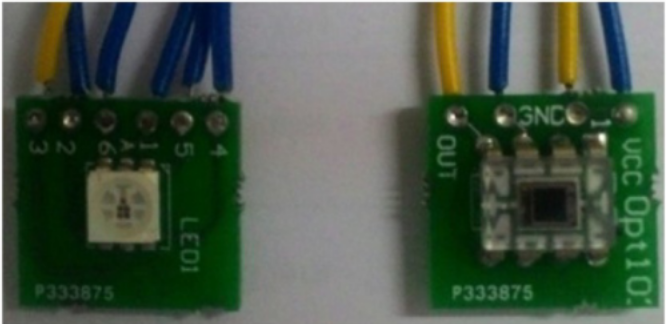


Fig. 7. Multi-chip Emitter MTMD6788594SMT6 and OPT101.

Furthermore, the author integrates these two models in a shell, which is shown in Fig. 8.



Fig. 8. The image of the system

The shell can protect the detector from environmental light. What’s more, through the programming of the MCU, the cycle time of the LED and the start and end of the detector collection can be accurately controlled. During the collection process for each patient, 5 LEDs light up for 4 seconds in sequence. Finally, 5 videoes are collected for each patient and used for analysis.

**1.3 Our Design of PPG Sensors**

After market research, we did not find the PPG collector that supports multi-wavelength LEDs. By comparing the advantages and drawbacks of the two kinds of collectors, we decide to design the PPG collector with a finger probe, multi-wavelength LED board, detector, MCU, and other components. Import models we choose are discussed as follows:

1. Finger probe

The finger probe is used as the shell to integrate the LED board and the detector. We can use the 3D print technology to make it which can meet our needs for adaptation of various modules (more suitable after initial validation), or, we can just buy one cheap pulse oximeter and replace the original electronic components with ours. For quick verification, we chose a pulse oximeter from Amazon, which is shown in Fig. 9. The purchase link is <https://www.amazon.co.uk/SIVLERS-Saturation-Oximeter-Portable-Batteries/dp/B08Y21THKF/ref=sr_1_7?crid=2K76QYLH2N7CG&keywords=pulse+oximeter&qid=1653771578&s=drugstore&sprefix=pulse+oximeter+%2Cdrugstore%2C84&sr=1-7>.



Fig. 9. A pulse oximeter from Amzon

1. MCU

MCU is used for signal acquisition and processing, due to the low sampling frequency and complexity of PPG sensors, we would like to choose the Arduino R3 as the MCU, which is shown in Fig. 10. Arduino UNO R3 is a microcontroller board based on the ATmega328P. It has 14 digital input/output pins (of which 6 can be used as PWM outputs), 6 analog inputs and a 16 MHz ceramic resonator. The purchase link is <https://www.amazon.co.uk/CubeplugTM-Development-ATMEGA328P-MEGA328P-Arduino/dp/B09S5YN5RV/ref=sr_1_21?keywords=arduino&qid=1653907925&sr=8-21>.



Fig. 10. Arduino UNO R3

1. Multi-wavelength LED board

We would like to choose the MTMD6788594SMT6 Multi-chip Emitter as the Multi-wavelength LED board, which is shown in Fig. 11. The purchase link is <https://www.digikey.co.uk/en/products/detail/marktech-optoelectronics/MTMD6788594SMT6/4746381>.



Fig. 11. MTMD6788594SMT6 Multi-chip Emitter

The parameter of the MTMD6788594SMT6 Multi-chip Emitter is shown as follows:

* Type Infrared (IR): Visible
* Current - DC Forward (If) (Max): 80mA, 80mA, 80mA, 50mA, 50mA
* Wavelength: 950nm, 850nm, 810nm, 770nm, 670nm
* Voltage - Forward (Vf) (Typ): 1.25V, 1.45V, 1.5V, 1.65V, 1.8V

1. Detector

We would like to choose OPT101 as the detector, which is shown in Fig. 12. The purchase link is <https://www.digikey.co.uk/en/products/detail/texas-instruments/OPT101P-J/301390>.



Fig. 12. OPT101 detector

The parameter of the OPT101 detector is shown as follows:

• Single Supply: 2.7 to 36 V

• Photodiode Size: 0.090 inch × 0.090 inch (2.29 mm × 2.29 mm)

• Internal 1-MΩ Feedback Resistor

• High Responsivity: 0.45 A/W (650 nm)

• Bandwidth: 14 kHz at RF = 1 MΩ

• LowQuiescent Current: 120 μA

The price of all modules is shown in Table 1.

Table 1. The price of all modules

|  |  |  |
| --- | --- | --- |
| No. | Modules | Price |
| 1 | Finger probe: SIVLERS - Pulse Oximeter Oxygen | £10.97 |
| 2 | MCU: Arduino UNO R3 | £10.99 |
| 3 | Multi-wavelength LED board: MTMD6788594SMT6 Multi-chip Emitter | £32.19 |
| 4 | Detector: OPT101 detector | £11.09 |
| 5 | Others: Dupont Line, PCB, etc | £50 |
|  | Total | 115.24 |

**2. Conjunctival image**

* 1. **popular conjunctival image collecting method**

In most research related to hemoglobin estimation by conjunctival image, the image is taken by phone camera. How to eliminate the effects of different ambient Lighting is the most important. Two methods are commonly used for light correction. The first is to use colour calibration card, as shown in Fig. 13.

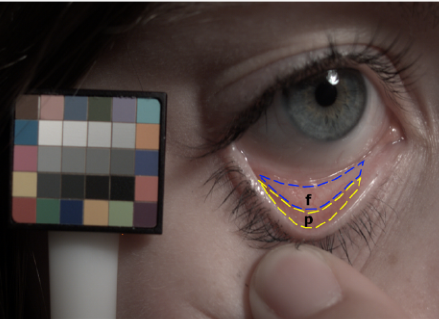


Fig. 13. Using a color correction card during image collection

The patient raises the color card with one hand and pulls down the eyelid with the other during the image collecting process. All images are calibrated by color charts after capture to keep color accuracy. The second method is to block out ambient light with additional equipment, as shown in Fig. 14.



Fig. 14. Using a box to block out ambient light during image collection

As we can see from Fig. 1, a plastic box is attached to the phone to keep the light consistent. However, it is difficult to make one device suitable for different mobile phones and such a device makes the operation complicated.

* 1. **The method we choose to collect the conjunctival image**

We do not want patients to add complex external devices during operation and consider the adaptation of different mobile phones, the first method is chosen. The conjunctival image collection steps are shown as follows.

* Turn off the phone flash and turn on autofocus
* The patient raises the color chart in one hand and places it close to the eye to be photographed
* The patient pulls down the eyelid with the other hand
* The photographer moves to the position where the patient's eyes and the colorimetric card almost fill the screen to take the image
* Perform the above steps to both eyes

1. **EHR information**

If the data collection is performed by NHS (Method B), we can get the patient's EHR data from the NHS system. If the data is collected by us (Method A), we would like to ask the test taker to fill out the questionnaire, including some simple vitals/demographic information that is listed as follows.

* Gender
* Race
* Age
* Smoking status
* Height
* Weight
* Marriage status
* Vegetarian or not

1. **The true value of hemoglobin**

If the data collection is performed by NHS (Method B), we can get the true value of the hemoglobin by a blood test of the patient. It is worth noting that although the EHR in the NHS system contains the hemoglobin test result in the past, the patient still needs to take the blood test to get the updated value (the previous value can be helpful in the prediction model). If the data is collected by us (Method A), we would like to use a high-performance non-invasive hemoglobin tester to get the hemoglobin value of the test taker (regard this value as the ground truth) because blood collection is unacceptable for many people. The hemoglobin tester involved is Pronto® Pulse CO-Oximeter®, shown in Fig. 15. This product has been used in much research to get the ground truth [2]. The purchase link is <https://cascadehealth.com/pronto-pulse-co-oximeter-by-masimo>. The latest price is not given on the web (around £500-1000). I have contacted the sales and I am waiting for their reply.



Fig. 15. Pronto® Pulse CO-Oximeter®

**Timeline**

The timeline of the main parts is as follows.

* Buy components and make PPG collector: 2weeks
* Buy Pronto® Pulse CO-Oximeter®: 1month
* Make the Color Calibration Card: 1week

[1] Moraes, J. L., Rocha, M. X., Vasconcelos, G. G., Vasconcelos Filho, J. E., De Albuquerque, V. H. C., & Alexandria, A. R. (2018). Advances in photopletysmography signal analysis for biomedical applications. Sensors, 18(6), 1894.

[1] Hasan, Md Kamrul. BEst (Biomarker Estimation): health biomarker estimation non-invasively and ubiquitously. Diss. Marquette University, 2019.

[2] Wang, Edward Jay, et al. "HemaApp: noninvasive blood screening of hemoglobin using smartphone cameras." Proceedings of the 2016 ACM International Joint Conference on Pervasive and Ubiquitous Computing. 2016.

[3] Pinto, Caje, Jivan Parab, and Gourish Naik. "Non-invasive hemoglobin measurement using embedded platform." Sensing and Bio-Sensing Research 29 (2020): 100370.