# **Building a Data Pipeline for a Real World Machine Learning Application**

Moses THIGA1, Pamela KIMETO2, Daisy KIPTOO3

Kabarak University, P.O. Box Private Bag, Kabarak, 20157, Kenya

1Tel: +254 0720 780 468, Email: mthiga@kabarak.ac.ke

2Tel: +254 0722 548 895, Email: [p\_kimeto@kabarak.ac.ke](mailto:p_kimeto@kabarak.ac.ke)

3Tel: +254 0714 875 094, Email: daisydrift.dk@gmail.com

**Abstract**: Predicting hypertension among individuals using historical blood pressure (BP) readings has been the focus of recent studies in Artificial Intelligence in health. This has been occasioned by the growing prevalence of hypertension in the general population as well as individuals’ desire to manage their health. The BP data used for predictions is collected during clinical visits and does not capture fluctuations in between clinic visits. Additionally, current prediction approaches rely solely on the BP readings without regard for the individual’s lifestyle and activities, which are known to affect individuals’ BP. This study developed a system that used agile methodology The system developed comprised of (i) a smartwatch with a Photolethysmograph (PPG) heart rate sensor for detecting the BP and (ii) a mobile phone application for receiving the BP readings and to collect data on participants age, weight, height and other health conditions. This system was piloted for regular collection of BP and related activity data for use in monitoring and prediction of an individual’s BP. Activities at the time of BP reading (sleep, exercise, chores) were also recorded. An alert was sent to the participant if the BP reading was abnormal. The pilot unearthed the following challenges: inability of the smartwatch to take readings on dark-skinned persons, the short time interval (30 mins) duration for data collection caused inconveniences, missing of readings during device charging, lack of complete integration between smartwatch and mobile application for the automatic transmission and recording of readings, inability to take readings in some locations due to security concerns for devices, inability to take readings at night because the smartwatch required light to function, and cases of forgetfulness by the participants in wearing smartwatch and/or entering the data. Recommendations from pilot include (i) an increase in the time interval to four hourly (ii) automate the process of taking and recording BP reading (iii) identify a smart watch that uses both PPG and ECG, and (iv) explore ways to for those not so literate to use the application.

**Keywords**: Blood Pressure, Hypertension, Photoplethysmograph, Smartwatch, Mobile Application

# Introduction

The continuous monitoring of individual health parameters such as physical activities as well as physiological and biochemical parameters has in the recent past been greatly enhanced through significant advances in wearable devices. These wearable devices are being adopted for the prevention of disease, maintenance of healthy habits, as well as individuals monitoring of (Wu et al., 2019). The most commonly measured data from these wearable devices include heart rate, blood pressure, body temperature, blood oxygen saturation, posture, and physical activities. This is done through the use of Photoplethysmograph (PPG), electrocardiogram (ECG), ballistocardiogram (BCG) sensors and other devices (Wu et al., 2019). Explain how PPG, ECG and BCG take these measurements briefly

A number of research activities around the collection of data from patients using a combination of wearable and mobile devices tend to only collect sensor based data and do not include provisions for users to input data on their activities and circumstances at the time the specified data is being detected from the individual. Examples of these studies include “*A healthcare real-time monitoring system for multiple sensors data collection and correlation*” by Romano *et al* (2009) that used a variety of sensors to collect real-time biomedical and environmental parameters from a user. A similar study by Nita, Cretu and Hariton (2011) “*System for remote patient monitoring and data collection with applicability on E-health applications*” also relies purely on sensors and does not provide for any user input data to complement the sensor data.

# Problem

The use of wearable devices and mobile phones to collect personal health data presents a great opportunity for the acquisition of data for use in enhancing personal health care through remote monitoring. However these current applications collect only physiological and preset activity data from individuals. They do not provide for user input of day to day or hour to hour activities, which can be subsequently harnessed for Machine learning or other data analytics.

# Study Objective

This study sought to develop and pilot a novel approach for the collection of BP readings and activity data from users in order to establish the practicality of this new approach in continuous monitoring of blood pressure reading and its fluctuations based on individuals day today activities.

# Literature Review

Significant growth in the internet, mobile technologies, cloud computing and the wearable internet of things (Wearable IoT) incorporating a great number and variety of sensors has led to great improvements in the ability to collect more complex and larger data sets that can be used to improve personalized health interventions. Wearable devices can be used to collect data about movements, physical activity, sleep and physiological response such as (Wu et al., 2019). Mobile phones have also been used to collect individual behaviours and experiences in real time otherwise known as ecological momentary assessments (EMA) (Shiffman et al., 2008)

The data collected through wearable and mobile technologies can be described as Big Data because of its great volume, variety and velocity at which it is collected. The ability to collect and store this data is a great achievement but with this development has come the challenge of extracting actionable insights from the data within reasonable time and accuracy limits. Big Data analytics has been the traditional approach to the analysis of this data (Belle et al., 2015) but the approach has its fair share of challenges especially when it comes to querying and reporting (Bresnick, 2017) (Dash et al., 2019). These and other challenges associated with the traditional approaches to big data analytics have led to the consideration of other viable options. One of these emerging approaches that have showed great promise in handling big and unstructured data effectively is Artificial Intelligence (AI) (Dash et al., 2019).

Artificial Intelligence refers to computing techniques that are used to ‘train’ computers to accomplish specific tasks by processing large amounts of data and recognizing patterns in the data. This makes it possible for computers to learn from experience, adjust to new inputs and perform human-like tasks. (SAS Institute Inc, 2020). There are six major branches of AI, namely; Machine Learning, Neural Networks, Expert Systems, Robotics, Natural Language Processing and Fuzzy Logic. While all these techniques have found some level of adoption in healthcare, neural networks, machine learning and deep learning stand out as key approaches that show the greatest promise for future solutions development in healthcare (Davenport & Kalakota, 2019).

The proliferation of wearable and mobile devices for the collection of vast amount of health data presents an opportunity for the greater utilization of AI in enhancing personal healthcare. Some examples include the use of continuous glucose meters (CGMs) that can collect glucose readings every five minutes as well as continuous blood pressure monitors that are able to show an individual how their blood pressure fluctuates with daily activities (Hadad, 2018).

# Methodology

The study was undertaken in two phases;

1. System development that followed an agile methodology. A mobile application for both blood pressure and activity data was developed and integrated with a smartwatch for blood pressure reading.
2. System pilot whose objective was to establish the usability and potential challenges with the system setup. Three participants were identified. After explaining to them what the pilot was meant to achieve and what was required of them, they gave their consent to participate in the pilot study

# Results

The following are the key results of the study.

1. **The system development process**

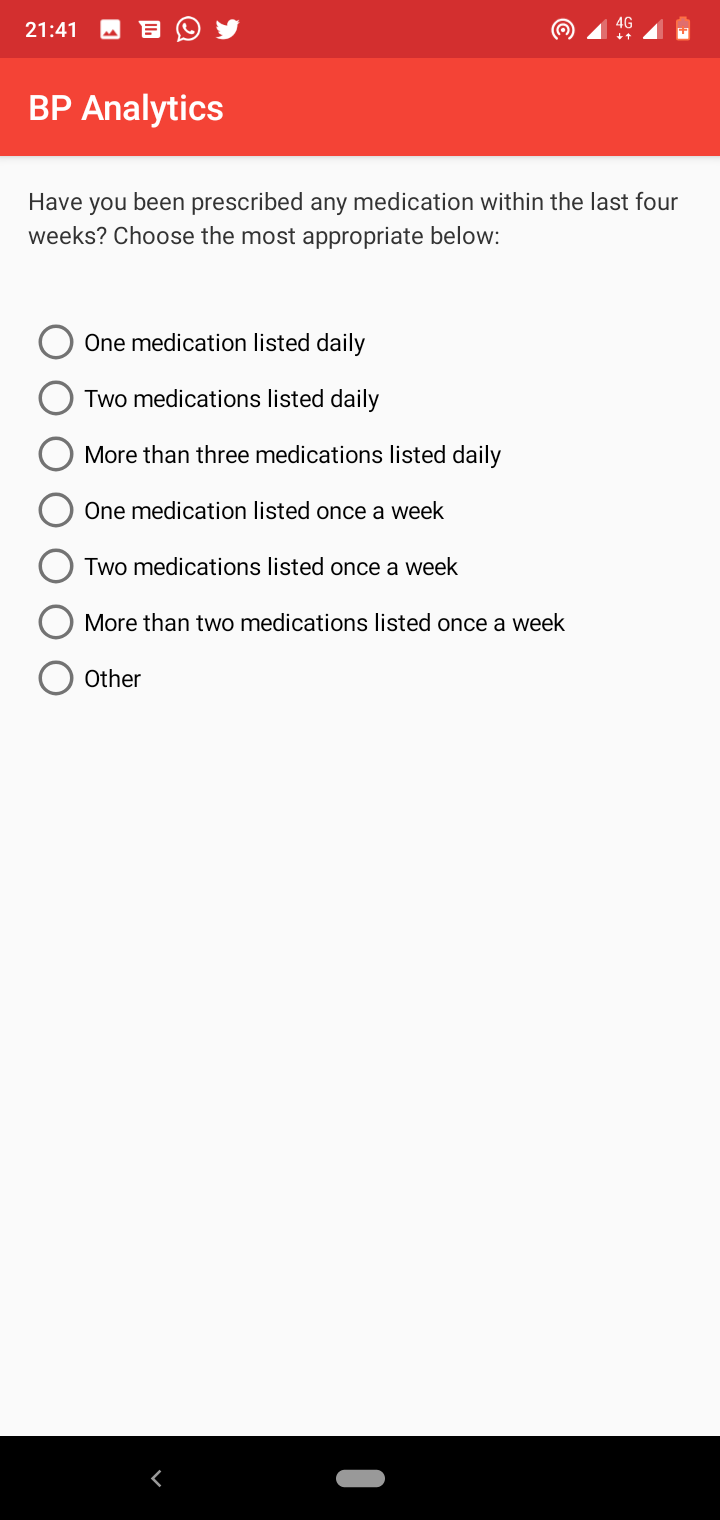
***Smartwatch Identification -*** The watch identified was the F1 Wristband Heart and Heart Rate Monitor. The watch uses a Photolethysmography (PPG) heart rate sensor for detecting the BP and sends the readings to the mobile application using a Bluetooth connection. The PPG heart rate sensor works by casting a green or red light onto the skin and using a photodetector at the surface of skin to measure the intensity of the non absorbed light. The varying intensities of the non absorbed light are then used to computer the volume of blood flowing in the arteries (Castaneda et al., 2018). The F1 Wristband Heart and Heart Rate Monitor uses green light.

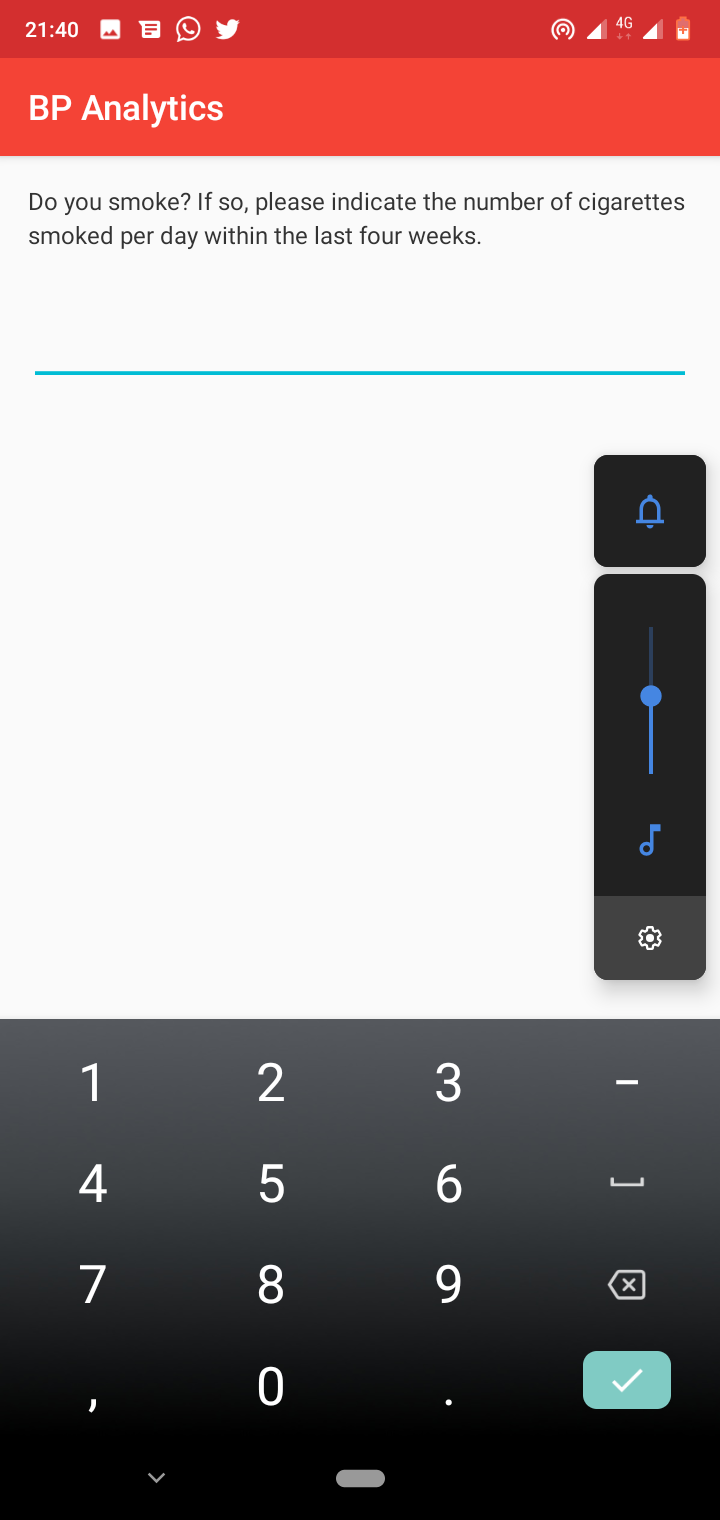
*Figure 1: The F1 Smartwatch and Heart Rate Monitor*

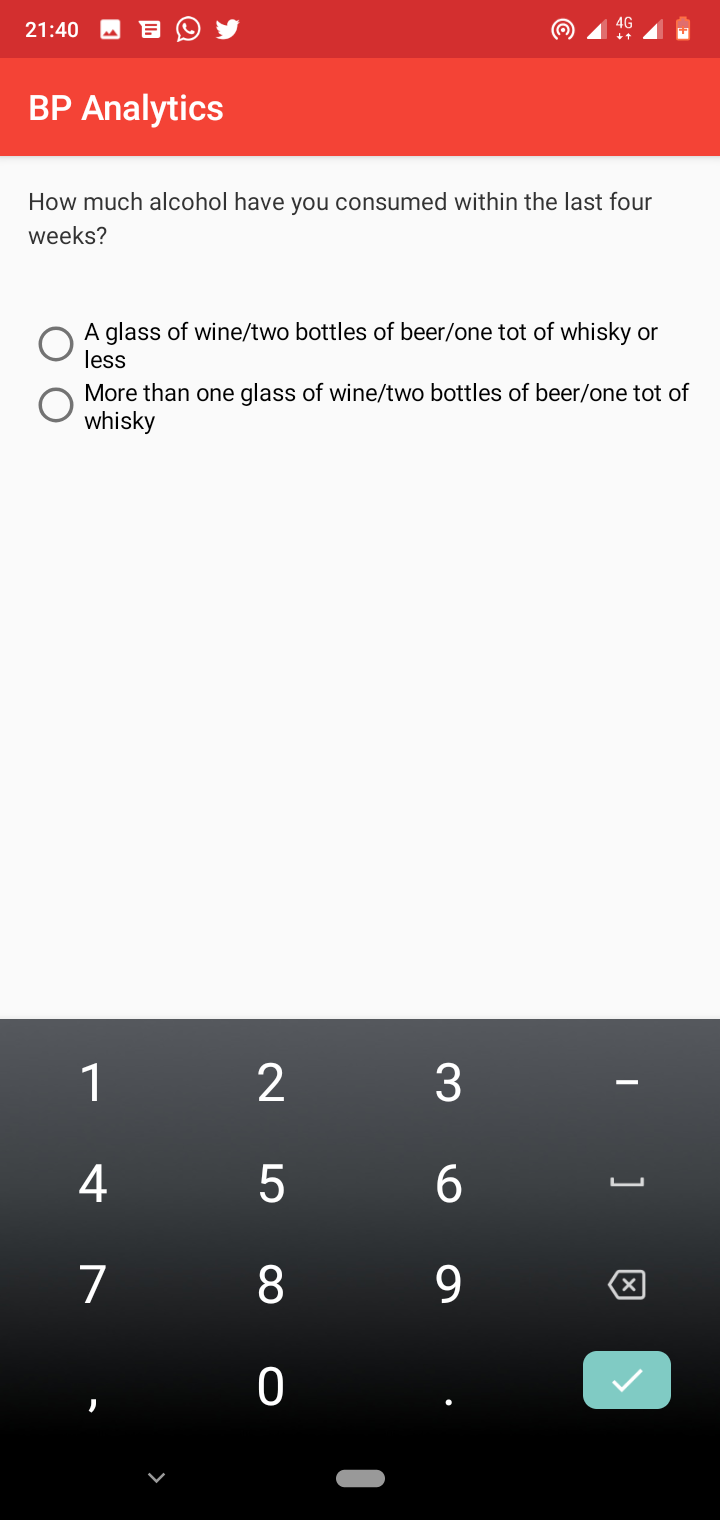
***Mobile Application Development*** – for this aspect of the study a mobile application was identified as the method of choice. The application was developed to receive BP readings from the smartwatch and to also collect activity data from the user. The specific data collected by the application is as follows;

1. **Background information –** such as name, age, gender, height, medication, family history of hypertension, smoking history, alcohol consumption and exercise levels.
2. **History of illness –** such as Pheochromocytoma, Hyperthyroidism, Acromegaly, Obstructive sleep apnea, Diabetes, Kidney disease, Hyperaldosteronism, Scleroderma, Cushing Syndrome and Lupus
3. **Blood Pressure Readings -** Systolic and Diastolic readings
4. **Additional Data taken during Blood Pressure Readings –** such as the heart rate, participant’s mood, activity and time of measurements.
5. **Data collected after four week of use such –** weight, exercise levels, alcohol consumption, smoking frequency, sleep patterns and use of medication.

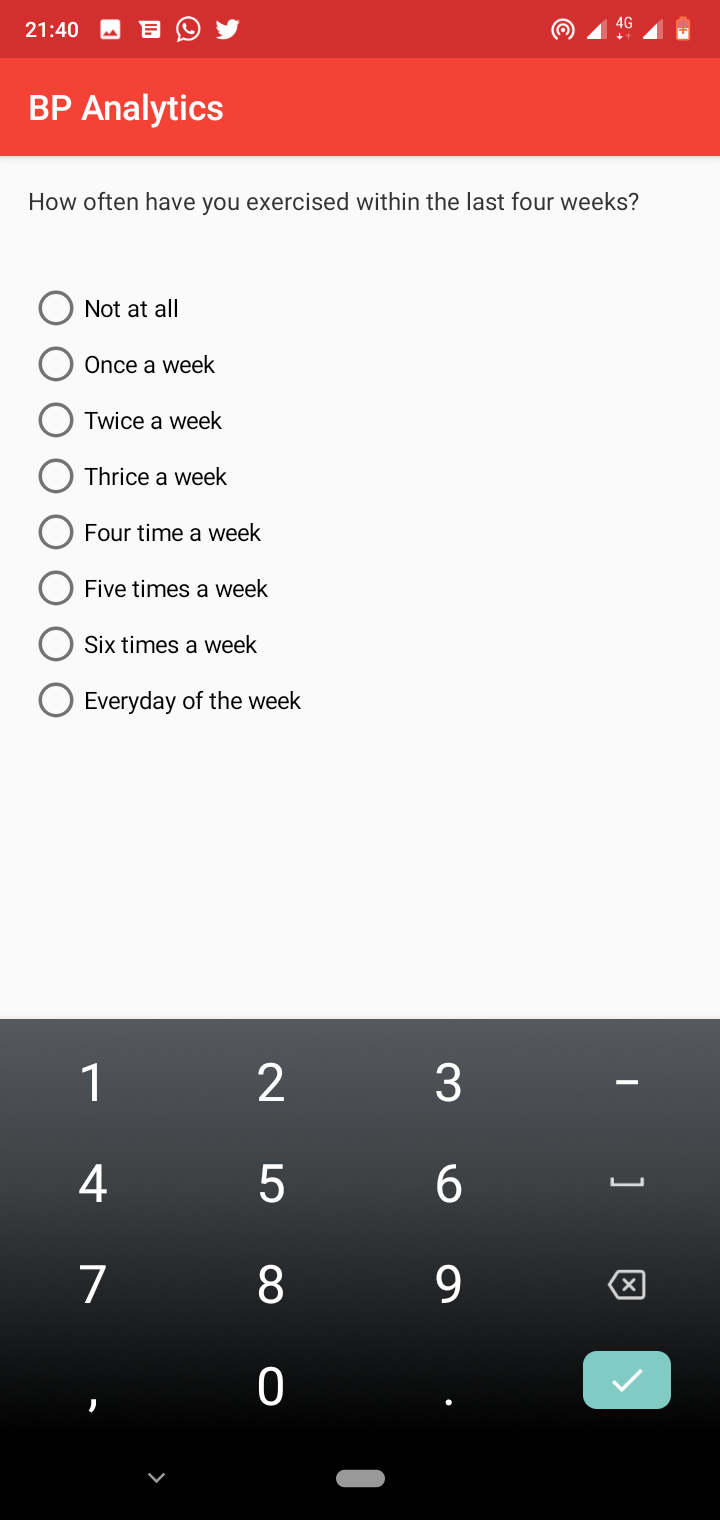
Some sample data collection screens are presented below;

*Figure 2 : Collection of Medication Information*

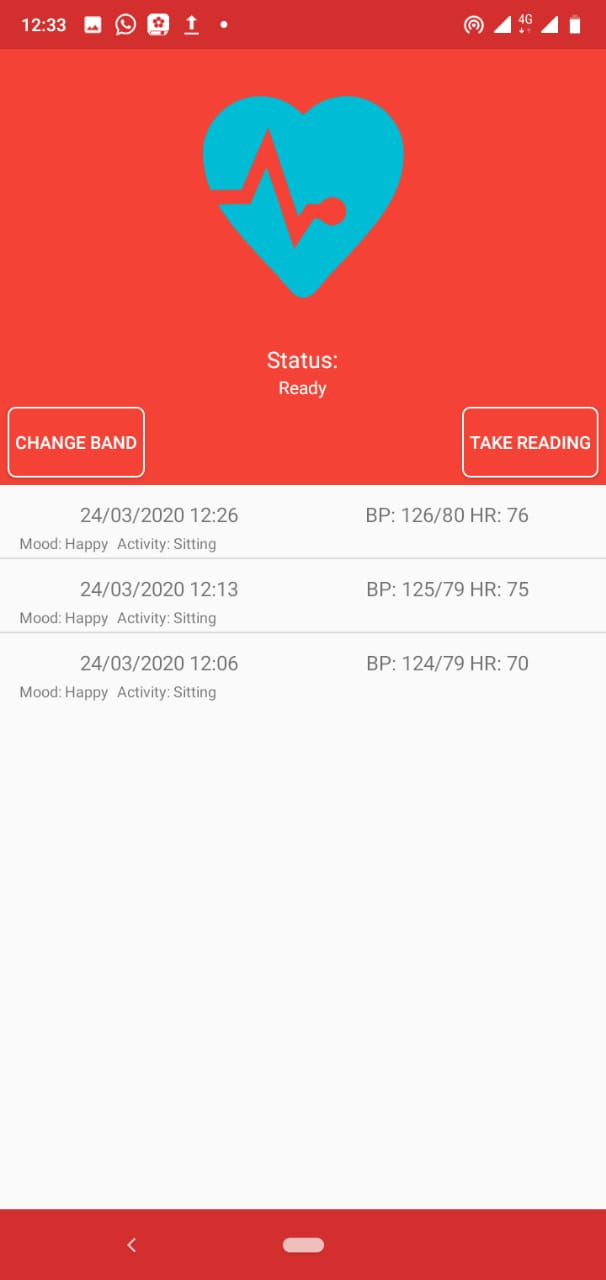
*Figure 3: Smoking History*



*Figure 4: Alcohol Consumption*



*Figure 5: Exercise History*

**

*Figure 6: A screenshot of the mobile application with readings*

1. **The System Pilot**

**Pilot study Duration –** The pilot study took a period of three months from May-August 2020

**Participant recruitment –** Three participants were recruited for the study and each of them used the smartwatch and application for at least 1 week. The participants were not able to get any reading because within the first few days it was realized that the smart watch that used PPG was not able to detect blood pressure or heart rate from these individual because of their dark skin tone. Another sample of persons with light skin tone were recruited. The smart watch with PPG was then able to detect BP and HR from these individuals for a period of four weeks.

**Pilot study findings** – The following are the key outcomes of the pilot study;

| **Aspect** | **Finding/s** |
| --- | --- |
| **Installation of the Application** | * 1. All the participants found the application easy to install   2. The application refused to run on Android 4.2 |
| **Usability of the application** | 1. The users remarked that the user interface was good looking and easy to use. 2. The system was also observed to be very responsive. |
| **Application power usage** | The application was observed to cause high power utilization on the mobile phone. |
| **Daily Data Input** | 1. The application at times retained the previous ‘Activity’ and ‘Mood’ between subsequent data entry sessions. 2. Problems were experiences with pairing the smartphone and the smartwatch sometimes. |
| **Time Interval for Data Collection** | The 30 minutes time interval set for the application was found to be too frequent by some respondents. |
| **The Smartwatch** | 1. The smartwatch was observed to give readings with a great variance even when measurements were taken within a short time interval. 2. The smartwatch did not have a ‘home button’ to allow a user to go back to the time display from the BP measurement function. |
| **Blood Pressure readings** | The manual process required to initiate the Blood pressure readings was found to be tedious for the users. |
| **Challenges in taking readings** | The participants remarked that the following circumstances prevented them from taking readings   1. Dark conditions such as during the night caused the smartwatch not to take readings. 2. The use of a bulky device such as a tablet, which made it cumbersome to take readings. Away from areas deemed to be safe to use the tablet 3. Engagements at workplaces or when doing manual work would prevent them taking readings. 4. Being on transit for example, walking in town and driving also prevented the participants from taking readings. |
| **Additional remarks on taking readings at night** | 1. The smartwatch was found not to take the readings in dark environments. 2. Taking readings when a user is asleep was not possible unless the process was automated. 3. The participants remarked that some people might not like to wear a watch when going to sleep. |
| **Daily activity disruption when taking readings** | The poor connection between the application and the smartwatch made it difficult for the participants to record the data at the times required, this caused inconvenience to their daily activities. |
| **Activities Data Entry** | Participants indicated that it would be good to allow for the entry of an activity that was not on the list provided on the application. |
| **Mood data entry** | Participants suggested the use of emoji for the data capture in the area of moods |

# Discussion

1. The development of the system comprising of the smartwatch and mobile application. - This aspect of the study was undertaken successfully as the smartwatch was able to send BP readings to the smartwatch.
2. The piloting of the system – The system pilot was successful and the participants involved were able to test the system under a variety of condition and use scenarios. They were also able to make valuable observations about the usability and potential value of the system for monitoring blood pressure.
   1. The smartwatch did not work well with persons who have a dark skin tone. The implication of this is that the system as developed would not find universal applicability in the Kenyan set up. This challenge was occasioned by the use of a Photolethysmography (PPG) heart rate sensor that depends on the readings taken from the absorption of green light. Green light has been found to be absorbed more on persons with dark skin pigment.
   2. The smartwatch was found to have difficulties in taking readings in the dark. This is because of the green light in the PPG sensor in the smart watch whose performance is influenced by ambient light.
   3. The integration between the smartwatch and mobile application was not very consistent leading to challenges when taking readings. This was occasioned by the power levels and Bluetooth settings in user devices that would at times cause the smartwatch not to connect to the mobile application.
   4. Users at times forgot to take the readings and to record their activities. This was occasioned by their busy schedules and the lack of reminders as well as automation of the process of taking readings. In addition the high frequency of the readings made it tedious to take the readings.

# Recommendations

The following are the key recommendations from the study;

1. There is need to identify and test an alternative smartwatch for use in taking the BP readings from users with a dark skin tone. A potential solution lies in the use of smartwatches that combine both PPG and ECG sensors. ECG sensors work by measuring the electrical activity of the heart using sensors placed on the skin(Rashkovska et al., 2020). This approach therefore works on all skin tones.
2. There is need to further automate the process of taking the BP readings to take care of situations where the users may be busy or forgetful. The system can then prompt them to fill in the activity data later on or allow for the participants to access their calendars for this information.
3. The frequency of taking the readings can be reduced to once every four hours. This is a reasonable time period to allow for the BP level to stabilize in the event of strenuous activities that would cause it to fluctuate.
4. The participants in this study were all literate and had some understanding of the use and manipulation of a smart watch and mobile device. This made it easier for them to read and input their data into their mobile devices. There is need to explore way that can be put in place to help those that may not be literate in using the smart watch and mobile device for data collection and learning.

# Acknowledgements

This work has been funded by the Artificial Intelligence for Development (AI4D) Program of the International Development Research Centre (IDRC) through the Knowledge for all Foundation (K4A).

# References

Belle, A., Thiagarajan, R., Soroushmehr, S. M. R., Navidi, F., Beard, D. A., & Najarian, K. (2015). Big Data Analytics in Healthcare. *BioMed Research International*, *2015*, 370194. https://doi.org/10.1155/2015/370194

Bresnick, J. (2017, June 12). *Top 10 Challenges of Big Data Analytics in Healthcare*. HealthITAnalytics. https://healthitanalytics.com/news/top-10-challenges-of-big-data-analytics-in-healthcare

Castaneda, D., Esparza, A., Ghamari, M., Soltanpur, C., & Nazeran, H. (2018). A review on wearable photoplethysmography sensors and their potential future applications in health care. *International Journal of Biosensors & Bioelectronics*, *4*(4), 195–202. PubMed. https://doi.org/10.15406/ijbsbe.2018.04.00125

Dash, S., Shakyawar, S. K., Sharma, M., & Kaushik, S. (2019). Big data in healthcare: Management, analysis and future prospects. *Journal of Big Data*, *6*(1), 54. https://doi.org/10.1186/s40537-019-0217-0

Davenport, T., & Kalakota, R. (2019). The potential for artificial intelligence in healthcare. *Future Healthcare Journal*, *6*(2), 94–98. PubMed. https://doi.org/10.7861/futurehosp.6-2-94

Hadad, Y. (2018, December 13). *How Artificial Intelligence Will Help Transform Personal Health in 2019—Digital Health*. https://hitconsultant.net/2018/12/13/artificial-intelligence-transform-personal-health/

L. Nita, M. Cretu, & A. Hariton. (2011). System for remote patient monitoring and data collection with applicability on E-health applications. *2011 7TH INTERNATIONAL SYMPOSIUM ON ADVANCED TOPICS IN ELECTRICAL ENGINEERING (ATEE)*, 1–4.

Rashkovska, A., Depolli, M., Tomašić, I., Avbelj, V., & Trobec, R. (2020). Medical-Grade ECG Sensor for Long-Term Monitoring. *Sensors (Basel, Switzerland)*, *20*(6), 1695. PubMed. https://doi.org/10.3390/s20061695

Romano, L., Coppolino, L., Elia, I. A., & Spagnuolo, G. (2009). A healthcare real-time monitoring system for multiple sensors data collection and correlation. In E. Damiani, J. Jeong, R. J. Howlett, & L. C. Jain (Eds.), *New Directions in Intelligent Interactive Multimedia Systems and Services—2* (pp. 455–464). Springer Berlin Heidelberg. https://doi.org/10.1007/978-3-642-02937-0\_42

SAS Institute Inc. (2020). *Artificial Intelligence – What it is and why it matters*. https://www.sas.com/en\_us/insights/analytics/what-is-artificial-intelligence.html

Shiffman, S., Stone, A. A., & Hufford, M. R. (2008). Ecological momentary assessment. *Annual Review of Clinical Psychology*, *4*, 1–32. https://doi.org/10.1146/annurev.clinpsy.3.022806.091415

Wu, M., Luo, J., & Contributors, P. O. J. of N. I. (2019, November 25). *Wearable Technology Applications in Healthcare: A Literature Review*. https://www.himss.org/resources/wearable-technology-applications-healthcare-literature-review